Yield Curve Modelling with PCA

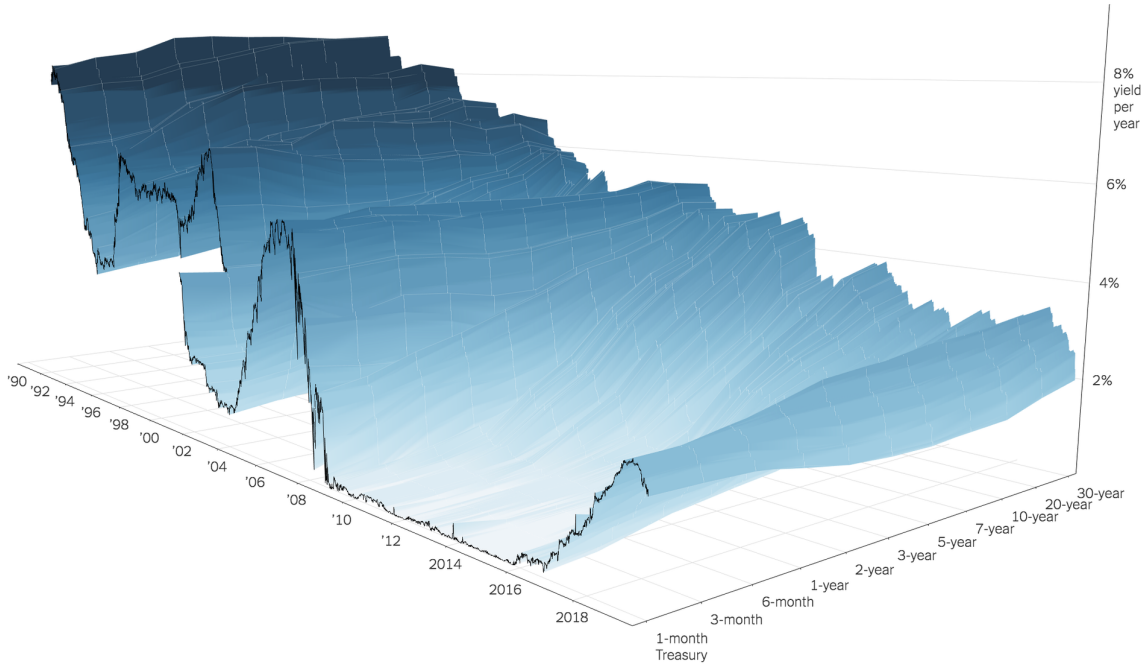
Math 585 - Applied Linear Statistical Modeling

Eunice Ofori-Addo

Introduction

A principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. Its general objectives are; data reduction and interpretation.
Principal Component Analysis (PCA) technique is used to reduce the dimensionality of a data set, finding the causes of variability and sorting them by importance. It reduces the number of variables that are correlated to each other into fewer independent variables without losing the essence of these variables. As PCA is useful to reduce information contained in a big dataset, it makes it easier to interpret by finding the least amount of variables that explain the largest proportion of the data.It does this by transforming the data from a correlation/covariance matrix onto a subspace with fewer dimensions, where all explanatory variables are orthogonal (perpendicular) to each other, i.e there is no multicollinearity.

Principal components analysis (PCA) is a way to analyze the yield curve. It makes use of historical time series data and implied covariances to find factors that explain the variance in the term structure. The yield curve shows the interest rates the government must pay to borrow money across different maturites. In other words, it can be thought of as “price of funds” and because of time value of money it is expected to be upwards sloping. Below is a great representation from New York Times (source).



Government bonds are said to have negligible default risk, as the government can simply borrow more money to finance their repayments. According to the expectations theory of interest rates, the yield curve is made up of two aspects:

1. An average of market expectations concerning future short-term interest rates.
2. The term premium — the extra compensation an investor receives for holding a longer-term bond. This is essentially because of the time value of money — $100 is worth more today than it is worth tomorrow, due to its potential earning capacity due to interest. Therefore, for a fixed-income investment to be worth the extra time the investor must part with their cash, the bond issuer must pay the investor some extra amount.

We can model these aspects of the yield curve using principal components decomposition.

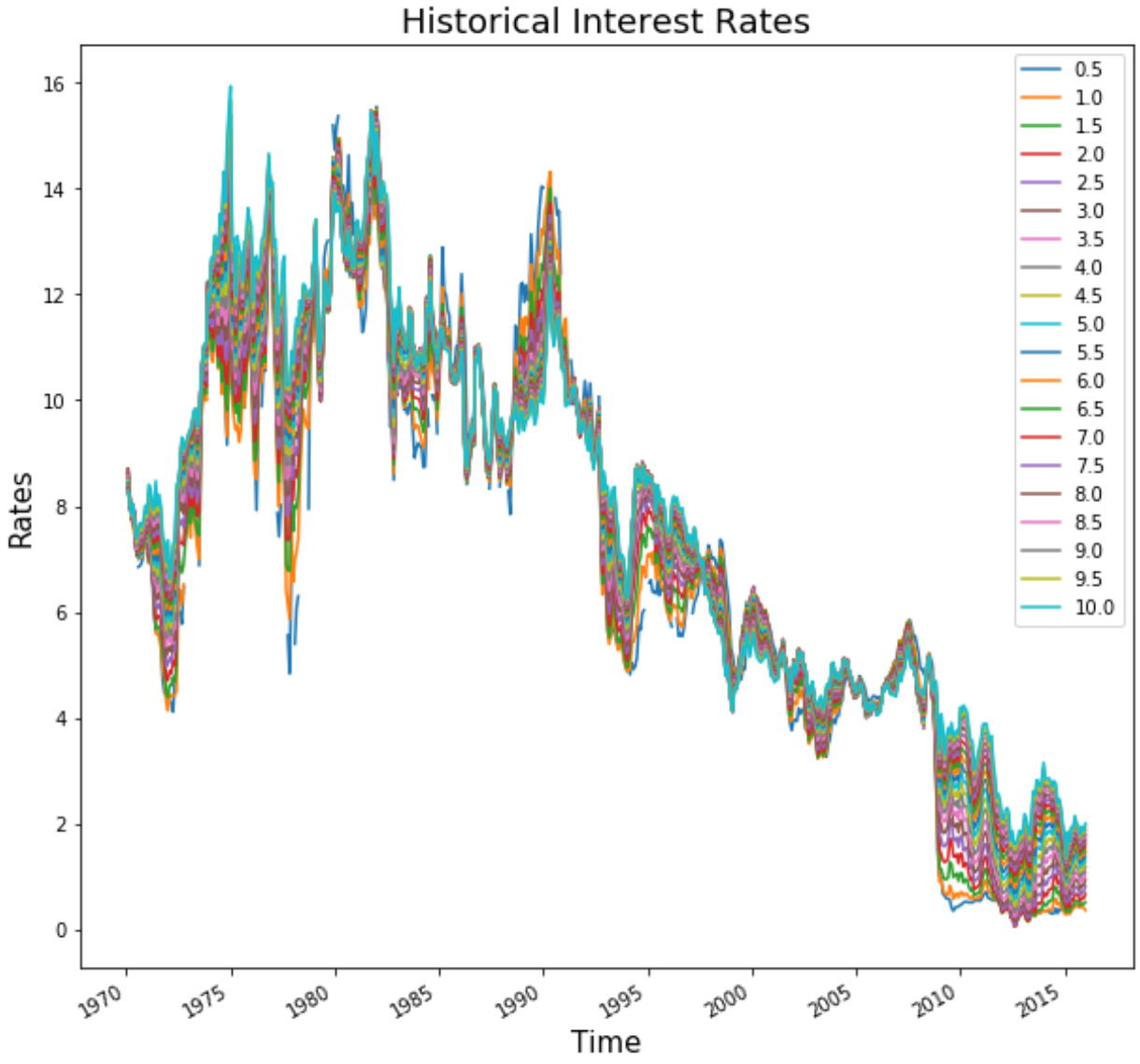
A major problem of estimating the term structure is the high dimensionality of the yield curve. In this project, the aim is to find the basis shape of the yield curve. Using PCA to capture variability in the movement of interest rates along the term structure.
Understanding the variability allows for creation of stressed interest rate term structures that can be applied in risk management(whenever uncertainty of interest rates is a concern).
What makes this underlying problem interesting is that PCA results can help investors identify trade opportunities. For example the cheapest point to go long on a curve, get the signal on when it is time to take profit on a position and its applications in risk management.

Data

Data

Data used are UK government bond spot rates from 0.5 years up to 10 years maturity and is collected from Bank of England database. It has 552 observations from January 1970 to December 2015 (Source:Bank Of England). Below is a table containing variable descriptions, means and standard deviations.

	Mean	St. Deviation	Min	Max
0.5 yrs	6.6682	3.959592	0.195146	15.363369
1 yrs	6.9216	3.911558	0.172541	14.955572
1 yrs	7.0077	3.860881	0.084630	15.024028
2 yrs	7.1004	3.827454	0.066103	15.116724
2.5 yrs	7.1859	3.797050	0.096886	15.190487
3 yrs	7.2629	3.769072	0.161504	15.264897
3.5 yrs	7.3327	3.743917	0.249812	15.400012
4 yrs	7.3967	3.721857	0.352506	15.485957
4.5 yrs	7.4562	3.702944	0.460420	15.529717
5 yrs	7.5118	3.687053	0.572102	15.540962
5.5 yrs	7.564	3.673927	0.685439	15.528246
6 yrs	7.6129	3.663216	0.798795	15.498448
6.5 yrs	7.6586	3.654535	0.911001	15.456901
7 yrs	7.701	3.647520	1.021170	15.447631
7.5 yrs	7.7402	3.641855	1.128612	15.462246
8 yrs	7.7762	3.637281	1.224624	15.471598
8.5 yrs	7.8091	3.633595	1.267301	15.474833
9 yrs	7.839	3.630644	1.308345	15.471284
9.5 yrs	7.8659	3.628322	1.348058	15.664855
10 yrs	7.8901	3.626557	1.386691	15.930716



Key Math Principles of PCA

Any nxn symmetric matrix Q can be decomposed

$$Q = AIA^T$$

where

- I is the diagonal matrix of eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$
- A is an orthogonal matrix whose columns $a_1, a_2, \cdots, a_n$ are the normalized eigenvectors of $Q(Qa_1 = \lambda_i a_i)$, which form an orthonormal basis of R^n.

Decomposition of Covariance Matrix

Let X be an n-dimensional random vector with

- mean $\mu = E[X]$
- covariance matrix $Q = cov[X]$ (symmetric and positive semi-definite)

Spectral theorem: $Q = AIA^T$ with

- eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$
- eigenvectors $a_1, \cdots, a_n$

Principal Components and Loadings of X

The Principal components of X is given by

$$Y = A^T(X - \mu)$$

where

- $Y_i = a_i^T(X - \mu)$ is the ith principal component of X
- a_i is the ith vector of loadings of X

Principal component decomposition
$X = \mu + AY = \mu + \sum_{i=1}^{n} Y_i a_i$

Properties of Principal Components

The Principal components Y are uncorrelated and have variances $Var[Y_i] = \lambda_i$:

$$E[Y] = 0$$

$$Cov[Y] = A^T Q A = A^T A I A^T A = I$$

Y_1 has maximal vairiance among all standardized linear combinations of X:

$$Var[a_1^T X] = max\{Var[b^T X] | b^T b = 1\}$$

Variance Explained

Observe that

$$\sum_{n}^{i=1} Var[\lambda_i] = trace(Q) = \sum_{n}^{i=1} \lambda_i = \sum_{n}^{i=1} Var[Y_i]$$

Hence the amount of variability in X explained by the first k principal components $Y_1, Y_2, \cdots, Y_k$ is given by

$$\frac{\sum_{k}^{i=1} \lambda_i}{\sum_{n}^{i=1} \lambda_i}$$

PCA of Yield Curve

When finding the principal components of the yield curve, the main theory held in financial economicsis that:

- ***PC1*** represents constant ≈ long term interest rate ≈ R*
- ***PC2*** represents slope ≈ term premium
- ***PC3*** represents curvature

Computing PCA

Import data and clean data. Dropping all missing values they can affect PCA results.

Step 1 To avoid the problem due to scale of variable value, standardize the data using:

$$Z = \frac{X - \mu}{\sigma}$$

Converts the data to mean of 0 and standard deviation of 1.

Step 2 Find the covariance matrix Σ, with eigenvalue-eigenvector pairs $(\lambda_1, e_1), (\lambda_2, e_2), \cdots (\lambda_p, e_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$, the principal components are given by:

$$Y_i = e_i z$$

This is eigendecomposition on our standardized data. Eigenvalues are scalars of the linear transformations.
Find attached excel sheet containing eigenvalues and eigen vectors

Step 3 We can use eigenvalues to find the proportion of the total variance that each principal component explains using the formula:
Total population variance:

$$\sigma_{1,1} + \sigma_{2,2} + \cdots + \sigma_{p,p} = \lambda_1 + \lambda_2 + \cdots + \lambda_p$$

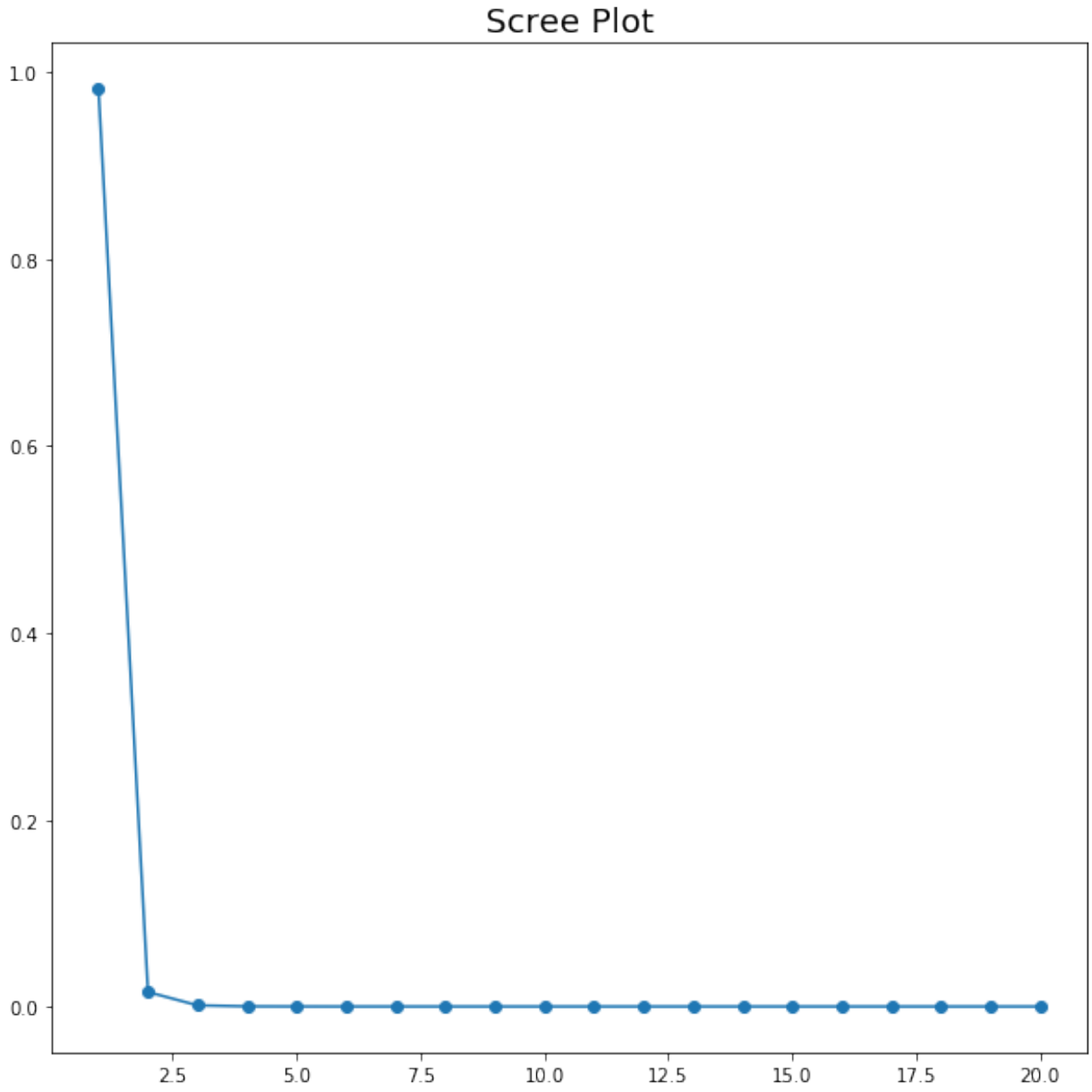
Proportion:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \quad k = 1, 2, \cdots, p$$

Step 4 Select principal components using the Scree plot.

	Eigenvalues	Explained proportion
1	19.6608	98.30%
2	0.309852	1.55%
3	0.0248899	0.12%
4	0.00349701	0.02%
5	0.000805609	0.00%
6	0.000102751	0.00%
7	8.59904e-06	0.00%
8	1.20913e-06	0.00%
9	9.34863e-08	0.00%
10	1.4878e-08	0.00%
11	2.83478e-09	0.00%
12	5.90107e-10	0.00%
13	1.3502e-10	0.00%
14	3.38837e-11	0.00%
15	7.75823e-12	0.00%
16	2.55909e-12	0.00%
17	1.11736e-12	0.00%
18	2.91278e-13	0.00%
19	6.58629e-14	0.00%
20	1.03114e-14	0.00%

Eigenvectors are the coefficients of these linear transformations, leaving the direction unchanged.



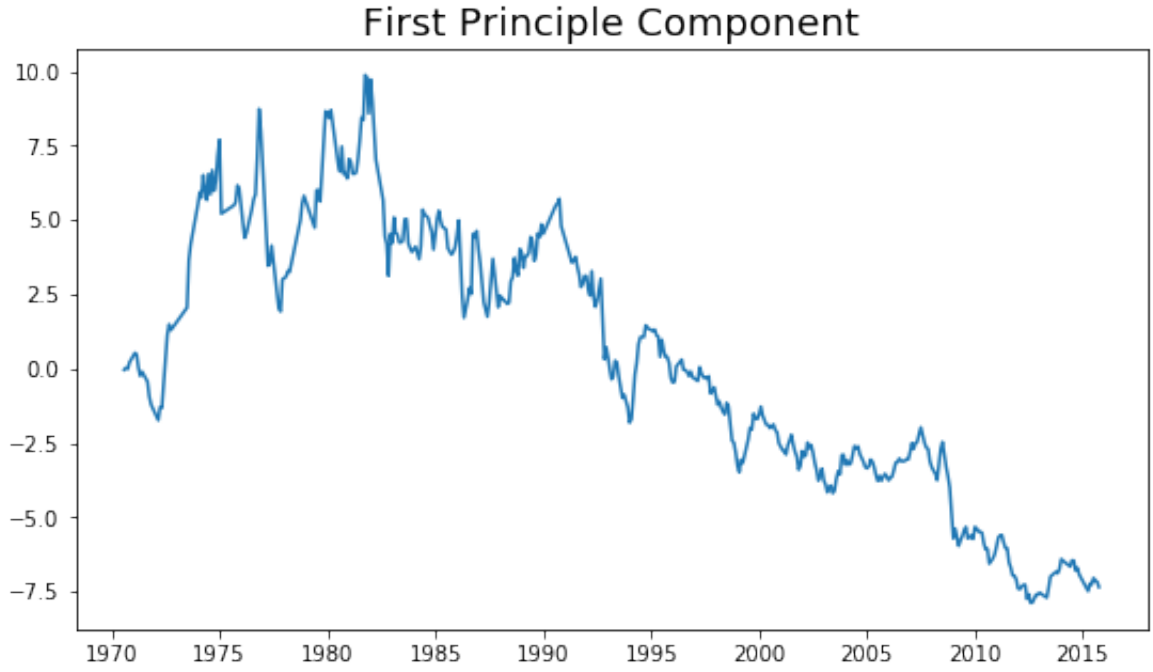
To form a time series for the principal components, we simply need to calculate the product between the eigenvectors and the standardized data. Below is the extract of the first five rows principal components time series.

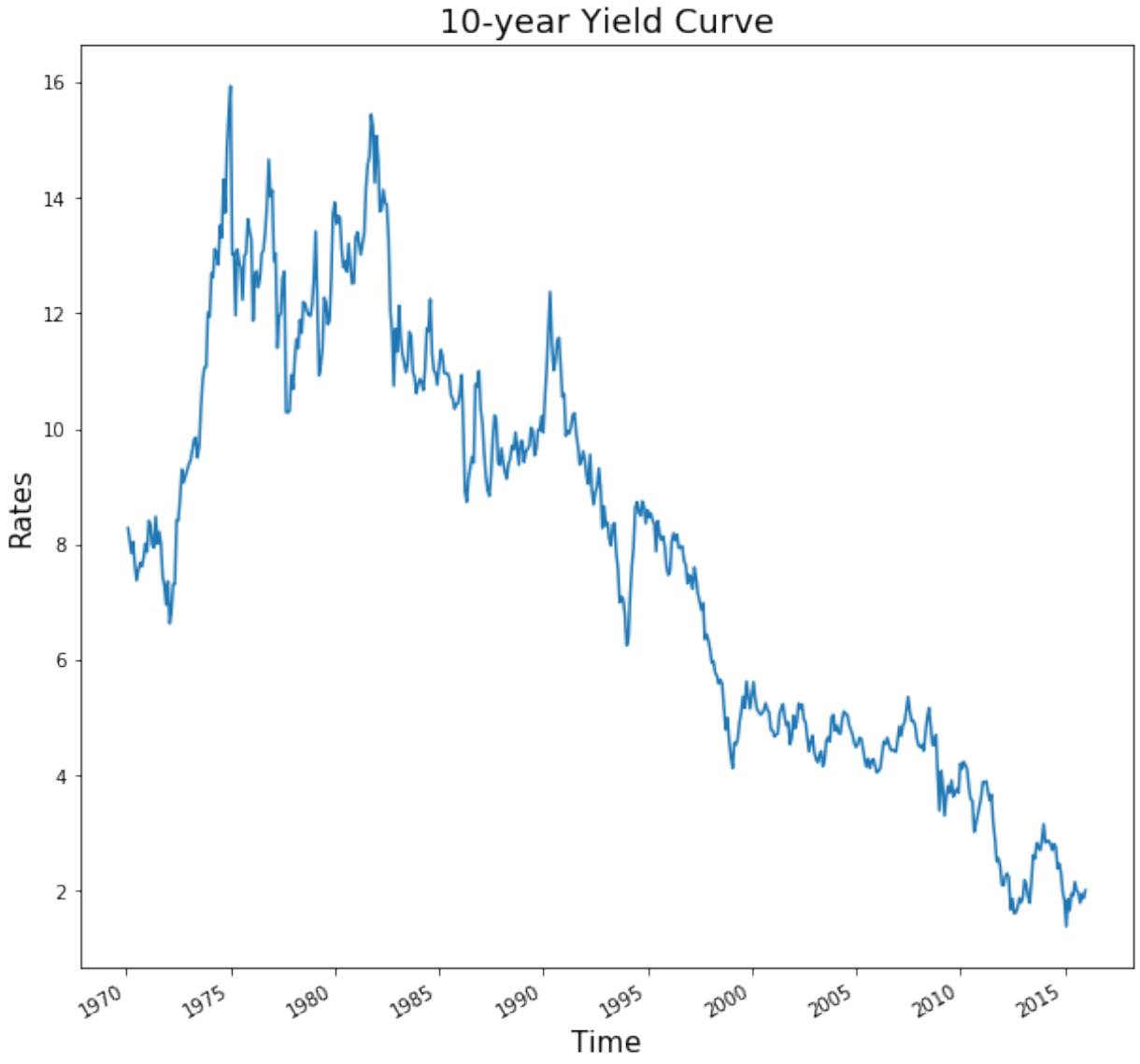
years:	0	1	2	3	4	5	
1970-07-31 00:00:00	-0.0328382	-0.204918	0.0674817	-0.146593	-0.00455011	-0.0217402	0.00196
1970-08-31 00:00:00	0.0318765	-0.126833	0.0629232	-0.10532	-0.0316634	-0.00818222	0.003
1970-09-30 00:00:00	-0.00932986	-0.150376	0.078643	-0.0893154	-0.0248819	-0.0107034	0.00202
1970-10-31 00:00:00	0.219963	-0.213589	0.0484864	-0.144802	-0.0375147	-0.0048405	0.0044
1971-01-31 00:00:00	0.533525	0.219927	0.126723	-0.0311346	-0.0110244	-0.0105896	0.0029

9	10	11	12	13	14	15	
0.000108448	4.99954e-05	-1.06893e-06	-2.26235e-07	-5.34679e-06	1.98404e-06	1.98052e-06	7.0759
-5.71552e-05	1.64879e-05	-1.30017e-06	3.75793e-06	-3.42992e-06	-2.01703e-07	2.31671e-06	-5.0625
-2.37896e-05	1.69164e-05	-3.53117e-06	7.74797e-06	-1.95356e-06	-7.39444e-07	1.2564e-06	-1.5585
-0.000109853	-7.67813e-06	-1.35302e-05	8.05235e-06	-6.09684e-06	-4.11785e-07	2.86023e-06	-7.6723
-2.02443e-06	1.16212e-05	9.63833e-06	8.79285e-06	4.26889e-06	-3.54167e-07	3.57151e-07	-1.3687

Results

When we plot the first principal component, we can see that it looks very similar to the actual 10-year yield curve. This makes sense, as according to our eigenvalues, the first principal component explains 98% of the data.

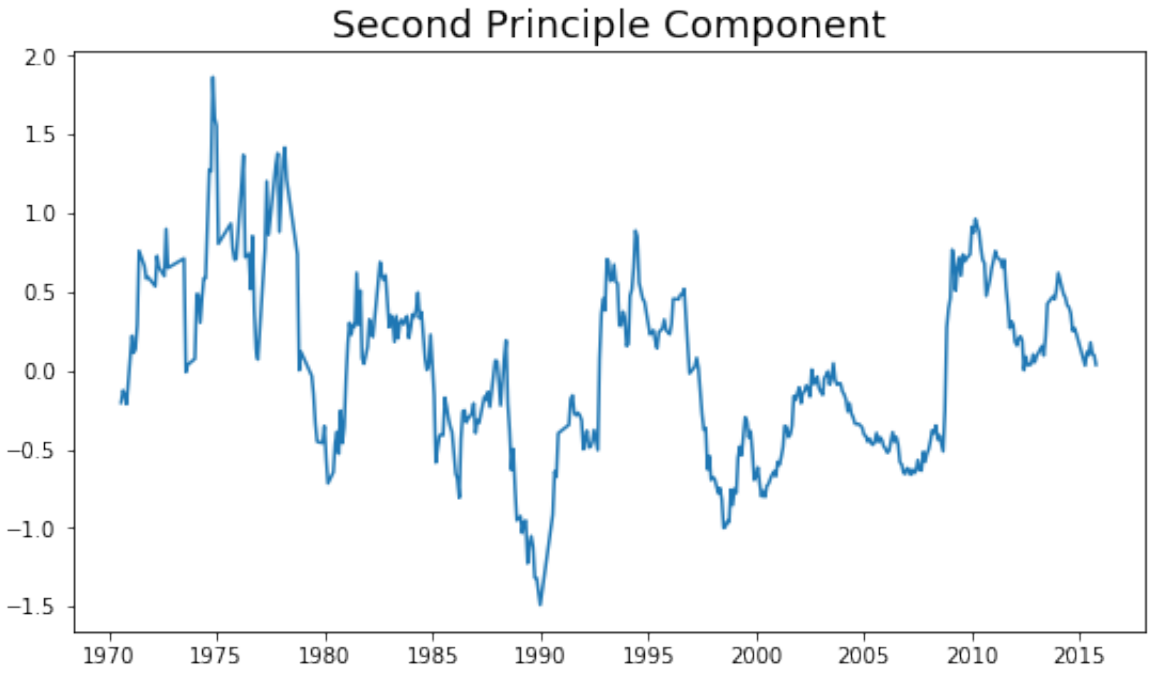


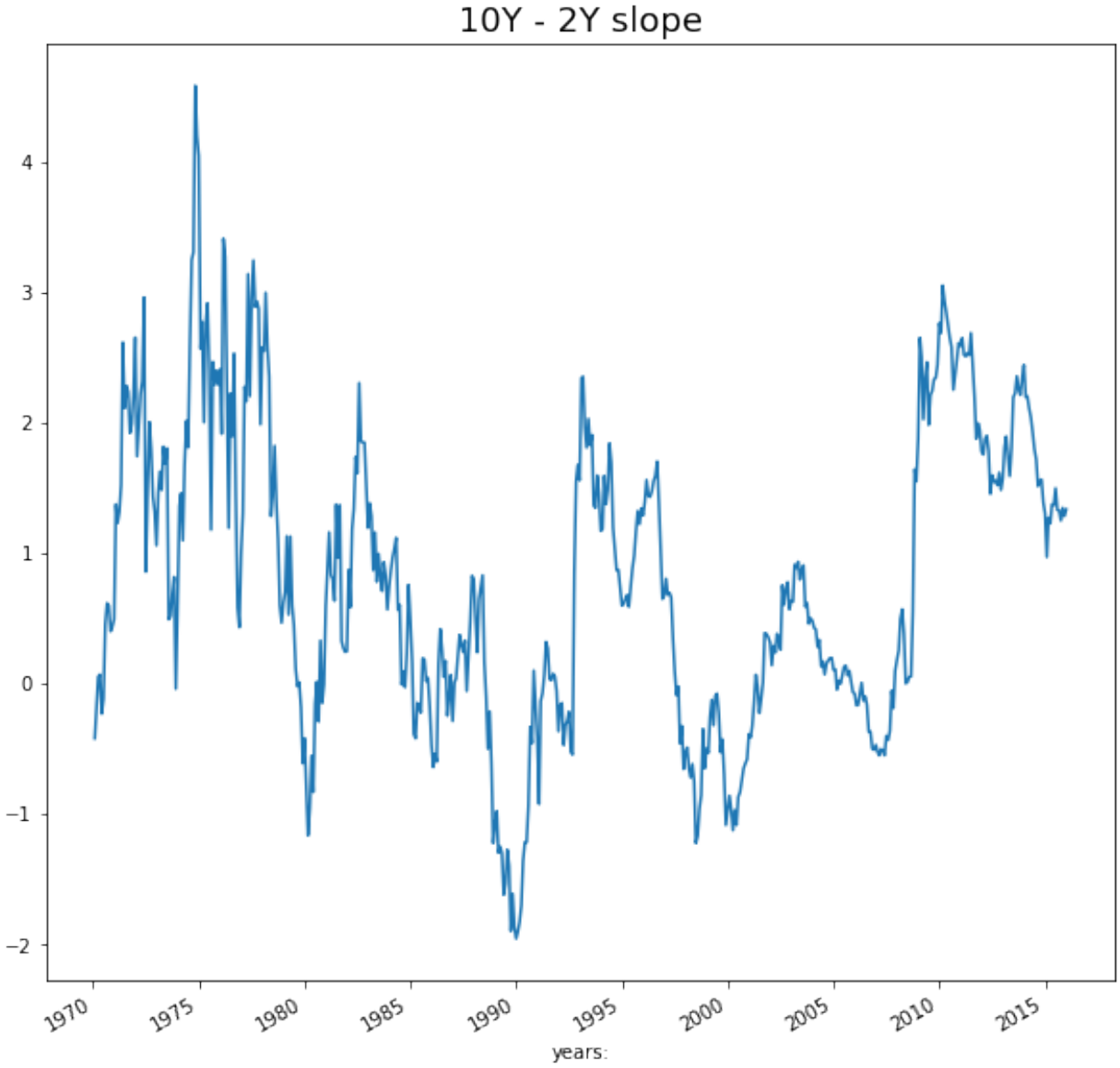


The second principal component represents the slope — this should have a correlation with the slope of the actual yield curve. One way we can calculate the slope is the 10-year spot minus the 2-year spot rate.

Years	2.0	10.0	Slope
1970-01-31 00:00:00	8.70073	8.27969	-0.421035
1970-03-31 00:00:00	7.79502	7.84522	0.0502035
1970-04-30 00:00:00	7.97352	8.0416	0.0680792
1970-05-31 00:00:00	7.86218	7.63017	-0.232015
1970-06-30 00:00:00	7.49362	7.37365	-0.119963

We can see that the slope looks almost identical to our second principal component from visual inspection.

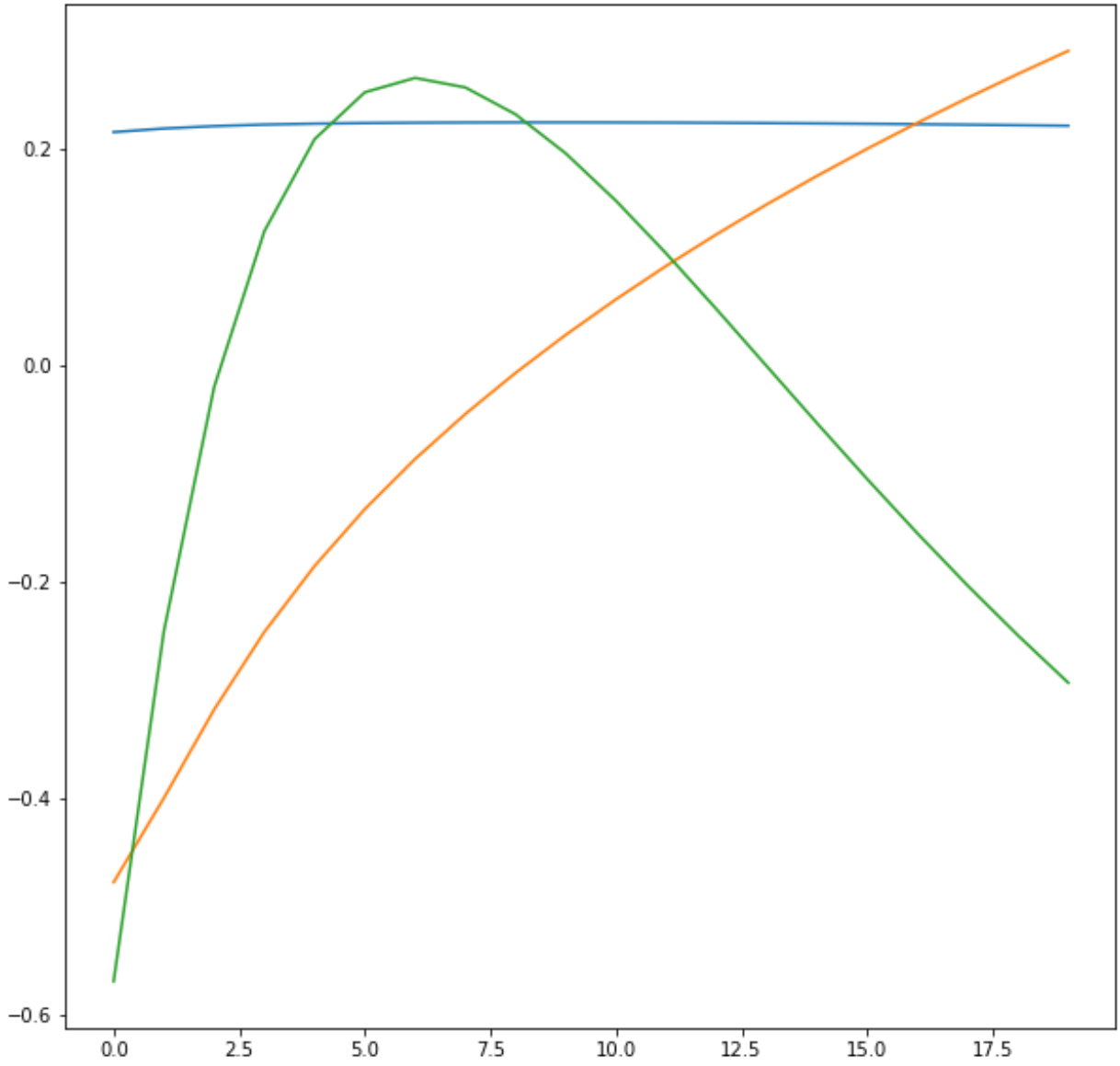

Second Principle Component
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015


10Y - 2Y slope
4
3
2
1
0
−1
−2
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
years:

When running the correlation between the second principal component and the 10Y-2Y slope of the yield curve, the high correlation of 0.96 shows us that the slope represent the slope.

$$\begin{bmatrix} 1 & 0.95856134 \\ 0.95856134 & 1 \end{bmatrix}$$

To finalize my conclusion about the interpretability of the derived Principal Components, each can be compared to a proxy of the traditional factors “level”, “slope” and “curvature”.



This is what the eigenvectors look like. The blue flat line is the first principal component(level), and corresponds to a parallel move up and down in the level of the entire yield curve. The orange line is the second principal component(slope) and is a steepening and flattening of the curve. The third component(convexity) is a twist. For this data the first principal component captures a staggering 98.30%! The second component manages to capture 1.55% and the third component 0.12%, beyond that only 0.02% of variance remains.

References

- Alexander, Carol, (2008). "Market Risk Analysis II, Practical Financial Econometrics".
- Adongo, Felix Atanga et al., 2018. Principal Component and Factor Analysis of Macroeconomic Indicators. Available at: https://pdfs.semanticscholar.org/8736/5855217edbc53e5e29c5c5872db7efb907cc.pdf
- Applied Multivariate Statistical Analysis by Dean W. Wichern, Richard A. Johnson, and Richard Johnson
- PCA Unleashed by Credit Suisse. Available at: https://plus.credit-suisse.com/rpc4/ravDocView?docid=kv66a7

In [ ]:

```
## Load DataFrame into R
df<-read.table(file="/Users/euniceofori-addo/Downloads/Math 585/GLC.xlsx")

## Remove any month with missing data so series have no NA values
df<- df[!is.na(apply(df,1,sum)),]

## Standardize the data
scaled.df <- scale(df)

## Eigendecomposition on our standardized data
pca <- eigen(scaled.df)

## Save eigenvalues and eigenvectors
save(pca$values, file = "df_eigval.xlsx")
save(pca$vectors, file = "df_eigvec.xlsx")

## Proportion of Explained Variation
PC <- data.frame(Eigenvalues = pca$values, VAR = (pca$sdev)^2/sum((pca$sdev

## Scree Plot
plot(var, type='b', main='Scree plot')
```