Seminar paper

Term Structure Analysis using PCA

based on
"An Introduction to Machine Learning in Quantitative Finance"
by Hao Ni, Xin Dong, Jinsong Zheng and Guangxi Yu

written under the guidance of

Dipl.-Ing. Dr.techn. Stefan Gerhold

by

Aleksei Malinovskii
Matriculation Number: 01627726

Vienna,
March 29, 2022

Contents

1 Introduction

The following seminar paper is based on the book ”Introduction to Machine Learning in Quantitative Finance” written by Hao Ni and Xin Dong, Jinsong Zheng, and Guangxi Yu [1].

In the book, the authors reveal the basic principles of classic machine learning algorithms and provide many practical examples from the field of quantitative finance. The particular chapter I focus on shows the application of PCA to term structure analysis. The Python codes contained in this book have been made publicly available on the author's GitHub [2]. In my analysis, I used different timelines and bonds to see how they would affect the results; for example, I looked at the timeline of the financial crisis caused by Covid-19. In addition to the main book, I also used alternative sources to understand the topic better.

2 Principal Component Analysis

2.1 Dimension Reduction

In statistics, dimensionality reduction is the transformation of data such that the number of variables is reduced by obtaining principal variables. It can be helpful in the further processing of the data using standard supervised or unsupervised machine learning algorithms, especially those that are dimensionality sensitive (e.g. k-nearest neighbours algorithm). Moreover, in some cases, dimensionality reduction can help us see the key structure of original data. For example, if the data lies in a low dimensional subspace, it is conceivable that one could restrict our learning problem to this low dimensional subspace and thereby simplify it.

There are two techniques of dimensionality reduction that can be applied to the feature space:

- **feature selection**

  is the process of selection of a subset of relevant features, which can be done manually based on domain knowledge or using statistical tools.

- **feature projection**

  converts data from high-dimensional space to low-dimensional space. The data transformation can be linear or nonlinear.

The basic linear technique for dimensionality reduction, **the principal component analysis (PCA)**, performs a linear mapping of the data into a smaller space such that the variance of the data in the low-dimensional representation is maximized[1]. In practice, a covariance matrix (and sometimes a correlation matrix) of the data is constructed, and the eigenvectors of this matrix are calculated. The eigenvectors corresponding to the largest eigenvalues (principal components) can now be used to recover most of the variance of the original data. Moreover, the first few eigenvectors can often be interpreted in terms of the large-scale physical behaviour of the system. The original space (with a dimension equal to the number of points) is reduced (with loss of data, but with the hope that the most important variance remains) to a space stretched over a few eigenvectors.

2.2 Principal Component Analysis

The goal of PCA is to find an alternative representation of data X in the transformed space so that X can be approximated by variables of smaller dimensions while maintaining a given level of the original information.

Let us denote a zero mean p-dimensional **feature space** $X = (X_1, ..., X_p)$. We aim to obtain a linear transformation V such that the p-dimensional transformed feature variable $Z := XV$ consists of components orthogonal to each other, and the variances of those components are in descending order. Alternatively, we are aiming to find the representation of X: $X = ZV^{-1}$ can be seen as a projection onto a new feature space, on which a dimensionality reduction approach can be built.

Figure 2.1 illustrates the main idea of PCA: to find such orthogonal vectors (principal components) Z_k for $k = 1, ..., p$, and to reduce the dimensionality of the original data X by mapping this data to the space generated by first $l, l = 1, ..., p-1$ vectors (on this figure $p = 3$ and $l = 2$).

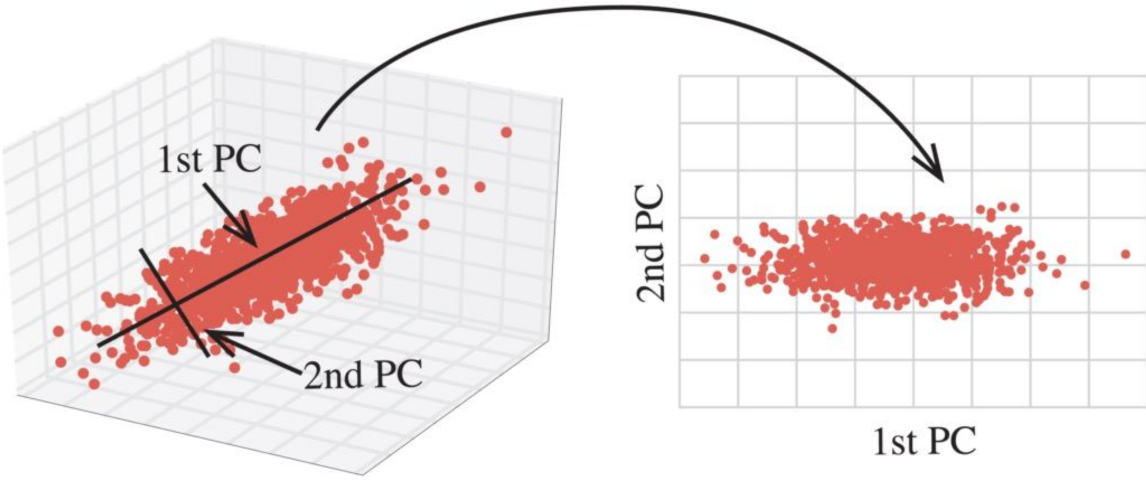


Figure 2.1: Dimensionality Reduction with PCA
[https://jermmy.github.io/images/2017-12-15/pca.jpeg]

2.3 Linear transformation

Let $V = (v_1, ..., v_p)$ be the **loading matrix** with $v_k = (v_{1k}, ..., v_{pk})^T$ then the linear transformation of X with V is

$$Z = XV. \tag{2.1}$$

and V should satisfy the following:

1. Euclidean norm: $\|v_k\|^2 = 1, \forall k \in 1, ..., p$;
2. The direction of V is chosen such that the transformed variables $\{Z_k\}_{k=1}^p$ should be all orthogonal to each other;
3. The variances are in a descending order: $\mathbb{V}(Z_1) \geq \mathbb{V}(Z_2) \geq ... \geq \mathbb{V}(Z_p)$.

Definition 2.3.1. Z_k with above properties is called the k^{th} **principal component** or k^{th} **PC score** for $k = 1, ..., p$.

The following algorithm can be used to find the loading matrix V :

Algorithm 1 PCA: Loading Matrix V

1. Find v_1:
$$v_1 = \underset{\|v\|=1}{\operatorname{argmax}} \|Xv\|^2,$$
then $Z_1 := Xv_1$ has the largest variance among all possible projected variables, i.e., $\|Z_1\|^2 = max_{v,\|v\|^2=1}\|Xv\|^2$.
2. For $k = 2, ..., p$, find v_k:
$$v_k = \underset{\|v\|=1}{\operatorname{argmax}} \|\tilde{X_k}v\|^2,$$
where $\tilde{X_k}$ is the residual of X after substracting $k-1$ components reconstructed with PCs found from previous steps:
$$\tilde{X_k} = X - \sum_{j=1}^{k-1} \tilde{X_j} = X - \sum_{j=1}^{k-1} (Xv_j)v_j^T.$$

2.4 Singular value decomposition

Definition 2.4.1. The singular value decomposition (SVD) is a factorization of a real or complex matrix. A real $N \times p$ dimensional matrix X can be decomposed as
$$X = UDV^T, \tag{2.2}$$
where

- U is an $N \times p$ orthonormal matrix: rows and columns of U are orthogonal, and $U^TU = I_p$, where I_p is the $p \times p$ dimensional identity matrix.
- V is a $p \times p$ orthonormal matrix: rows and columns of V are orthogonal, and $V^TV = I_p$.
- D is a $p \times p$ diagonal matrix with real entries $(d_i)_{i=1,...p}$ called **singular values** and $d_1 \geq d_2 \geq ... \geq d_p \geq 0$.

The principal components Z can be obtained by
$$Z = XV = (UDV^T)V = UD. \tag{2.3}$$

2.5 Principal components and covariance

Let's recall some basic defenitions from probability theory and statistics [3]:

- for the vector $X = (X_1, X_2, ..., X_k)^T$ of k jointly distributed random variables with finite second moments, its auto-covariance matrix K_{XX} (also denoted as Σ_X) is defined as follows:
$$K_{XX} = \operatorname{Cov}(X, X) = \mathrm{E}\big[(X - \mathrm{E}[X])(X - \mathrm{E}[X])^T\big] = \mathrm{E}[XX^T]–\mathrm{E}[X]\mathrm{E}[X]^T. \tag{2.4}$$

- let X be a random vector with covariance matrix Σ_X, and let A be a matrix that can act on X on the left. The covariance matrix of the matrix-vector product AX is:

$$\begin{aligned}\mathrm{Cov}(AX, AX) = \mathrm{E}[(AX)(AX)^T] – \mathrm{E}[AX]\mathrm{E}[(AX)^T] = \\ = \mathrm{E}[AXX^TA^T] – \mathrm{E}[AX]\mathrm{E}[A^TX^T] = \\ = A\mathrm{E}[XX^T]A^T - A\mathrm{E}[X]\mathrm{E}^TA^T = \\ = A\big(\mathrm{E}[XX^T] – \mathrm{E}[X]\mathrm{E}[X]^T\big)A^T = A\Sigma_X A^T. \end{aligned} \tag{2.5}$$

- The empirical covariance matrix X can be written as

$$\Sigma_X = \frac{1}{N-1} XX^T. \tag{2.6}$$

- For every $n \times n$ real symmetric matrix, the eigenvalues are real and the eigenvectors can be chosen real and orthonormal. Thus a real symmetric matrix A can be decomposed as

$$A = Q\Lambda Q^T \tag{2.7}$$

  where Q is an orthogonal matrix whose columns are (the above chosen, real and orthonormal) eigenvectors of A, and Λ is a diagonal matrix whose entries are the eigenvalues of A.

By using the decomposition (2.2) we can now this empirical matrix (2.6):

$$\Sigma_X = \frac{1}{N-1} XX^T = \frac{1}{N-1}(UDV^T)(UDV^T)^T = V\Big(\frac{1}{N-1}D^2\Big)V^T. \tag{2.8}$$

Additionaly as the covariance matrix Σ_X is a real nonnegative symmetric matrix we can use an eigenvalue decomposition (2.7):

$$\Sigma_X = Q\Lambda Q^T. \tag{2.9}$$

It is clear that the right singular vectors in equation (2.8) are the eigenvectors in (2.9):

$$V = Q,$$

and the diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ is $\Big(\frac{1}{N-1}D^2\Big)$, i.e componentwise:

$$\lambda_j = \frac{d_j}{N-1}. \tag{2.10}$$

That means that the new coordinates V are given by the eigenvectors of the covariance matrix.

Now finally the empirical covariance matrix of the principal components Z can be rewritten:

$$\Sigma_Z = \frac{1}{N-1} Z^T Z = \frac{1}{N-1}(XV)^T(XV) = V^T\Sigma_X V = \frac{D^2}{N-1} = \Lambda. \tag{2.11}$$

To summarize [1]:

1. The principal components $Z_1, ..., Z_p$ are orthogonal (i.e. $Z^{-1} = Z^T$), and the variances of $Z_1, ..., Z_p$ are the eigenvalues of the covariance matrix of X in descending order.
2. With the representation of the covariance of the data using SVD, we can apply PCA to reduce the dimensionality of the data.

2.6 Summary

Once we have the principal components and the loading matrix, the representation of X becomes

$$X = ZV^{-1} = ZV^T = \sum_{j=1}^{p} Z_j v_j^T. \tag{2.12}$$

Thus an approximating sequence $\left\{\tilde{X}^{(k)}\right\}_{k=1,...,p}$ of feature variables X is obtained. In section (2.5) we have shown that $\tilde{X}^{(k)}$ including the first k principal components of X:

$$\tilde{X}^{(k)} = \sum_{j=1}^{k} Z_j v_j^T \tag{2.13}$$

covers such proportion

$$\frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \tag{2.14}$$

of the variance of X. From the equation above, it is obvious that if we take $k = p$, we recover all the original data of X.

Assuming, therefore, that the information is incorporated in the covariance, given the data set X and the information preservation criterion in the form of the percentage of variance explained, we can find an approximation of X with reduced dimensionality.

The **residual** from the approximation using k PCs is

$$R^{(k)} = X - \tilde{X}^{(k)} = \sum_{j=k+1}^{p} Z_j v_j^T. \tag{2.15}$$

2.7 Practical problems

1. **PCA is not invariant to scaling.** Before applying PCA, we should consider whether or not to standardize the data X, besides removing of the mean from the data before applying PCA. In practice, it depends on the context of the data and the objective, but there are two general tips regarding this problem[1]:
   - If features are measured in different units that are not comparable, we should first standardize them before applying PCA.
   - If features are measured in comparable units, then we can keep them unscaled to preserve the original variability information.

2. **PCA can not be used if the data is not Gaussian.** PCA is based on the assumption that the data is normally distributed, that is why we can represent all the information from the covariance. In the case of non-Gaussian data, independent component analysis (ICA) is often used; for more details, see [4] and [5].

3 Application: Term Structure Analysis Using PCA

3.1 Introduction to fixed income term structure

Bonds are one of the many financial instruments with a specified maturity date available in the market. A bond is a fixed-income instrument representing a loan made by an investor to a borrower (typically corporate or governmental). They are used by companies, municipalities, states, and sovereign governments to finance projects and operations. Owners of bonds are debtholders, or creditors, of the issuer [6]. The main risk that comes with corporate bonds is credit risk. They depend on the creditor's ability to repay the debt, so there is always the possibility of defaulting on payment.

Treasury bonds are government-issued bonds that are deemed safe and secure. While corporate bonds have some level of default risk, Treasury bonds are guaranteed if held to maturity and can be used as a benchmark to measure the performance of other fixed-income investments. *With lower risk comes lower reward:* in comparison to corporate bonds, Treasury bonds earn a lower interest rate.

3.1.1 Bond yield and yield curve

A bond's yield is the interest income received by an investor from investing in debt securities. Interest income on them is generated from two sources. On the one hand, bonds with a fixed coupon, like deposits, have an interest rate that accrues on the face value. On the other hand, bonds, like stocks, have a price, which can change depending on market factors and the situation in the company. Although, price changes are less significant with bonds than with stocks. Current yield is the ratio of annual coupon payments to the current market value of the bond:

$$\text{Current Yield} = \frac{\text{Annual Coupon Payment}}{\text{Bond Price}}.$$

Having the definition of bond yield, we define the yield curve. The yield curve represents the time structure of interest rates and shows the relationship between the yields of financial instruments and their maturity. Using this tool, an investor gets an idea of several market properties of traded bonds and can also predict the potential behaviour of the security's price under the influence of market factors. By analysing graphical and tabular data, it is possible to assess the current state of the market, calculate fair premiums and calculate bond prices under forecasted interest rate movements.

3.1.2 Treasuries bond yield curve and recession

The Treasuries bond yield curve is not only used as a benchmark to assess the value of other debt instruments but is also considered one of the most important economic indicators to watch as a precursor to a future cyclical downturn in the US economy.

Usually, yields of longer-dated securities are higher than those of shorter-dated securities (this is a compensation for risk). However, the short-term yield sometimes exceeds the long-term yield, and the spread turns negative. The result is a concave curve.

If short securities turn out to be more profitable than long ones, investors do not believe in the long term, i.e. they are afraid to hold long bonds.

The inversion (concavity) of the Treasuries curve in the US is considered one of the proxies for an imminent recession or downturn in the economy. At least, as history shows, in seven out of ten cases, it has indeed become a clear harbinger of an American recession [7].

In the following example, we analyse the data of Treasury bonds between 1st January 2017 and 31st December 2018 taken using the Federal Reserve Bank of St. Louis API[8]. The US economy has been growing steadily in the period under review since after a rebound in 2016, it grew by 3.1% in 2018 after climbing by 2.5% in 2017 [9][10].

3.2 Factor Model

We will start with the **mathematical model.**
Let X be the yield curve consisting of bonds of p maturities with N observations. We will model X as a k factor model.

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,p} \end{pmatrix}$$

The k factor model for the yield curve X can then be presented in the following way

$$X = \mu_X + Zf + e, \tag{3.1}$$

where $\mu_x = (\mu_1, ..., \mu_p)$ is the mean vector of $X, e = (e_1, ..., e_N)$ is the residuals and

$$Z = \begin{pmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,k} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N,1} & z_{N,2} & \cdots & z_{N,k} \end{pmatrix} \text{ consist of } k \text{ factors,} \tag{3.2}$$

$$f = \begin{pmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,p} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ f_{k,1} & f_{k,2} & \cdots & f_{k,p} \end{pmatrix} \text{ the factor loading matrix;} \tag{3.3}$$

i.e., the observation i of feature j is

$$x_{i,j} = \mu_j + \sum_{l=1}^{k} z_{i,l} f_{l,j} + e_{i,j}. \tag{3.4}$$

How to determine the factors?

One natural solution would be to simply take macroeconomics factors, including economic outputs, unemployment rates, gross domestic product (GDP), inflation, etc. Bond traders and financial analysts often cite monetary policy as a significant factor in the movement of the term structure. The bond market reacts instantly to the release of macroeconomic news. Conversely, movements in the term structure carry essential information about the future state of the macroeconomy [11].

Macroeconomic factors are indeed a good solution for explaining the behaviour of the curve over a long period, such as five years, but this approach has its limitations; in particular, it is not as effective for explaining shorter-term behaviour. There are several simple explanations for this: first, we do not closely observe the key macroeconomic factors; second, a highly correlated set of macroeconomic factors may be unable to provide a stable result in the fitting process.

Therefore, instead of identifying external factors, we will focus on self-contained independent drivers, i. e., factors related to the curve itself, such as slope and curvature.

Through this approach, we will recognize the key structure of the yield curve dynamics and will be able to answer how 2Y, 10Y, and 30Y bond yields move together. Moreover, PCA will help us to naturally reduce the dimensionality of the factors based on the coefficient of explained variance. In our final steps, we will focus on hedging based on risk exposure representation in terms of the factors.

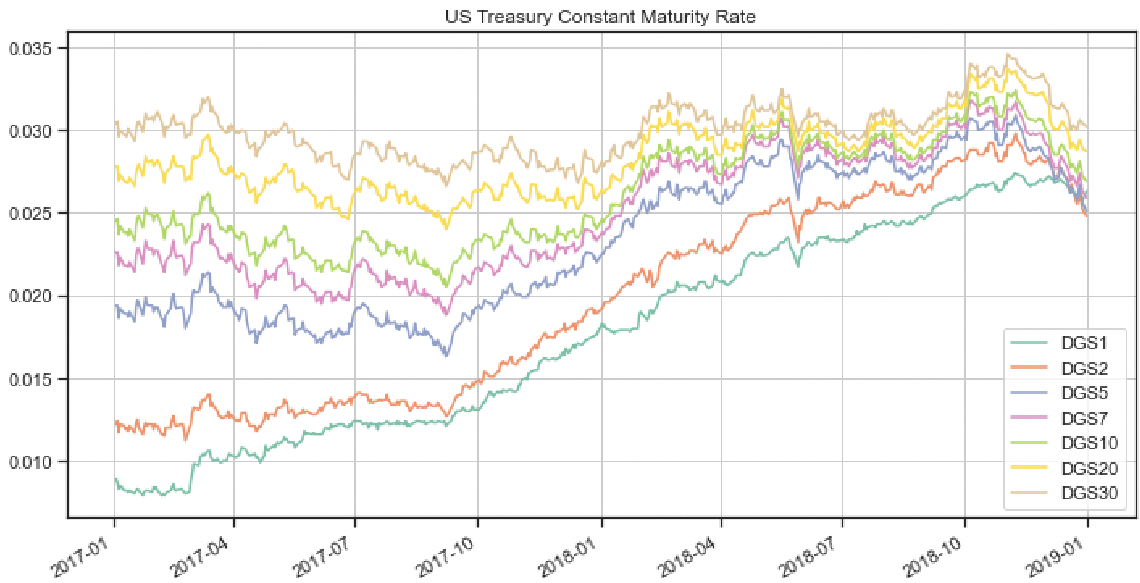


Figure 3.1: US Treausry constant maturity rates

3.3 Data and observation

The yield/rate time series of different maturities are shown in Figure 3.1. It is clear that all curves are moving in the same direction, which indicates the idea that some common factors drive the yield dynamic.

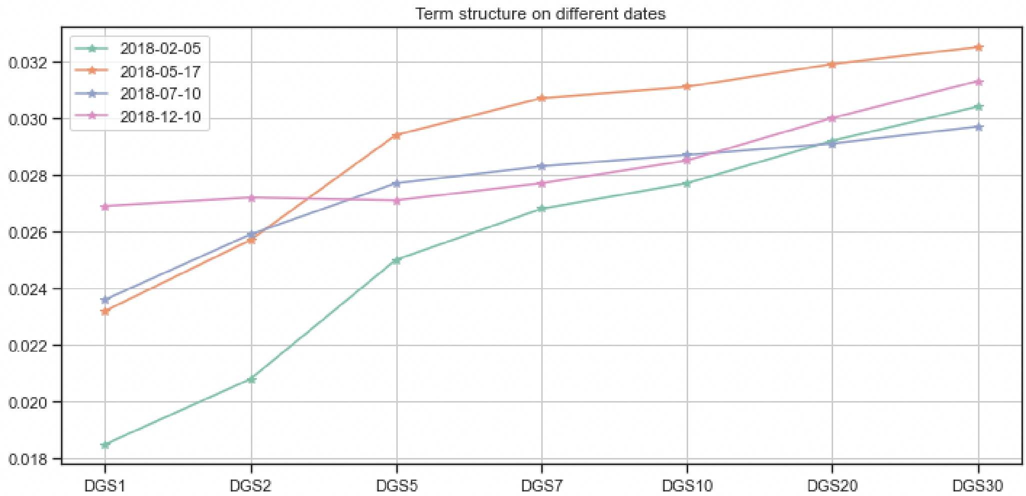


Figure 3.2: Term structure of interest rates

In Figure 3.2 we can see the yield curve term structure on different dates. In contrast to the previous figure, here one can distinctly see the inconsistencies in the dynamics and it is obvious that in some periods the movement is quite the opposite.

Let us consider in more detail the different types of changes between the dates:

1. **Level change**

   The level shock modifies the interest rates of all maturities in nearly equal amounts, inducing a parallel shift that changes the level of the entire yield curve [12].

   An example of level change can be seen by observing the green (2018-02-05) and orange (2018-05-17) curves. Both are moving in the same direction with almost identical slopes, and the only difference is the level: the green curve is shifted compared to the orange one.

2. **Slope change** The shock to the slope factor increases short-term interest rates by much more than long-term interest rates so that the yield curve becomes less abrupt and its slope decreases [12].

   For example, from 2018-05-17 (orange) to 2018-12-10 (pink), the slope of the curve changed as the short-term yield increased while the long-term yield decreased (DGS2 to DGS5).

3. **Curvature change**

   The change in curvature reflects how the difference between the medium-long-term premium (i.e., the long-term minus the medium-term rate level) and the short-medium-term premium (i.e., the medium-term minus the short-term rate level) changes from each day. In other words, the "belly" of the curve shifts relatively to variations in short-term and long-term levels daily [1].

Moreover, we can say that the orange (2018-05-17) and green (2018-02-05) curves show us a classic example of the so-called normal yield curve, a concave curve with a positive slope. Such a yield curve is considered "normal" because the market usually requires more compensation for more risk under normal conditions. Long-term bonds are subject to greater risks, such as changes in interest rates and increased potential default risk.

3.4 PCA on term structure

We have already noticed that the dynamics of the yield curve are determined by some common factors, and the PCA allows us to investigate these factors.

Consider X as the de-meaned daily yield change of p maturities on N days. From Equation 2.3 in Section 2.4, we get the representation of X as

$$X = ZV^T,$$

where Z consists of the principal components and V is the loading matrix. Now we are definitely prepared to analyse the PCA results.

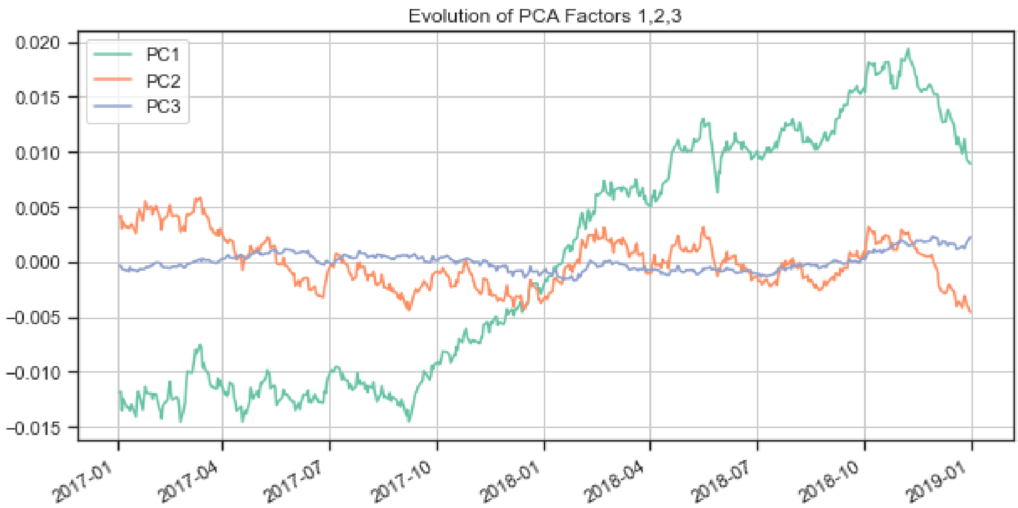
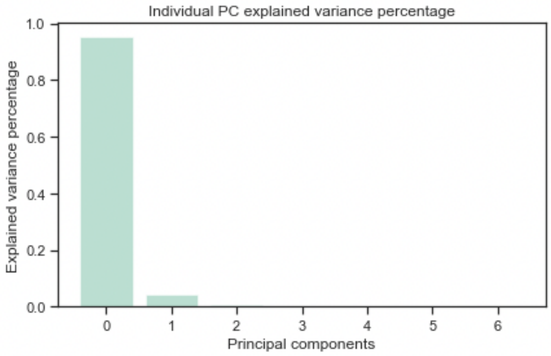


Figure 3.3: The first 3 principal components

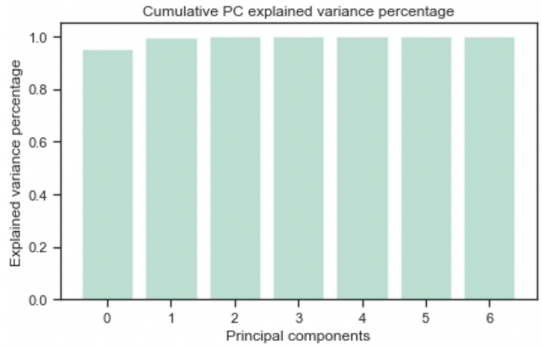
3.4.1 Principal components (Z)

Figure 3.3 shows the time series of the first three principal components. It is clear that the variances of these components are in descending order: the PC1 (green) varies a lot, while PC3 (blue) fluctuates around zero.

Figure 3.4 indicates that almost 95% of the variation in the data is described by the first principal component, and combined with the second and third components, it is almost 100%, which means that all of the dynamics of the yield curve can be described by the first three components. In the next section, we will take a closer look at the loading matrix V.



(a) Individual PC explained variance



(b) Cumulative PC explained variance

Figure 3.4: PCA-explained variance ratio

3.4.2 Loading matrix (V)

V is the loading matrix that represents X in the new coordinates. The matrix itself is presented in Figure 3.5 (a) and visualized in Figure 3.5 (b). We can use the loading matrix to determine which PCs have the greatest effect on each bond. A loading close to 0 indicates a weak impact of the variable, and a loading close to ± 1 shows a high impact. Let us now take a closer look at the behaviour of each PC:

- **PC1** (green)

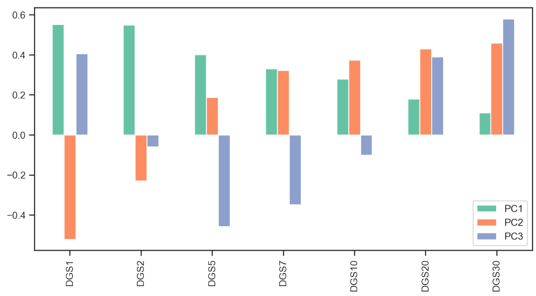
  All loads are positive, implying that the PC1 factor causes the movement of the yield curve in the same direction. Therefore, this factor is responsible for the **level change**. This reflects that short-term yields tend to move more than long-term yields since the loads on the short end are larger. Moreover, this figure demonstrates that most of the varience is accumulated in PC1.

- **PC2** (orange)

  The loading increases from a negative value at the short end to a positive value at the long end. Furthermore, it crosses zero one time between two (2Y) and five (5Y)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
DGS1	0.553153	-0.521945	0.406891	0.503280	-0.014236	0.049874	-0.007363
DGS2	0.548462	-0.229310	-0.058753	-0.784664	0.142051	-0.033247	0.078571
DGS5	0.401411	0.187341	-0.458251	0.105166	-0.664757	-0.058601	-0.370652
DGS7	0.330703	0.322376	-0.348364	0.263626	0.146838	0.041370	0.756690
DGS10	0.279229	0.373706	-0.101193	0.142498	0.626985	0.304434	-0.515790
DGS20	0.180465	0.429733	0.390784	0.010568	0.062438	-0.789335	-0.054686
DGS30	0.110360	0.460045	0.580245	-0.173307	-0.345178	0.524912	0.121570

(a) PCA loading matrix entries



(b) PCA loading matrix visualization

Figure 3.5: PCA loading matrix

years bonds. This indicates that short-term yields and long-term yields tend to move in different directions, so it describes the **change in slope** of the curve. Note that a zero crossing point of the curve is an anchor point, and as we can see from the figure, it is around five years bonds.

- **PC3** (blue)

  The loadings from short to long-term cross zero twice, which means that the very short-term and long-term yields move in the same direction while the "belly" part of the curve has the tendency to move in the opposite direction. This factor represents the **curvature** of the curve in a natural way [1].

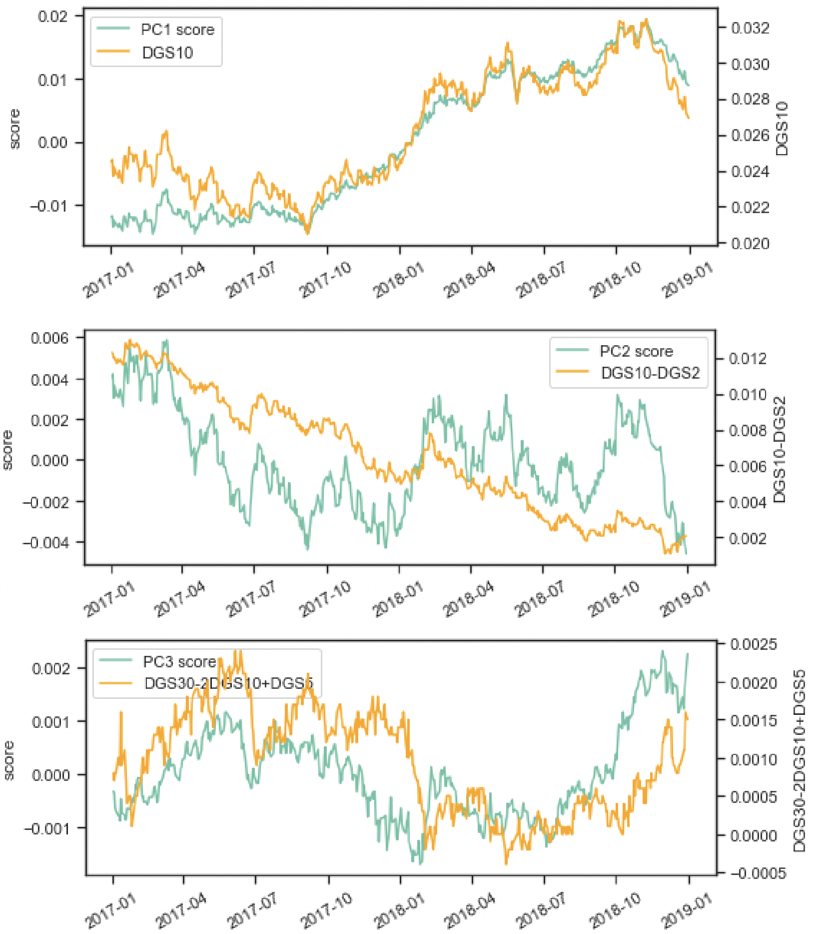


Figure 3.6: What do PC scores represent?

As we discussed in Section 3.3, the first three principal components represent the internal characteristics of the curve: level, slope, and curvature. To confirm our reasoning, we will choose combinations of bonds that reflect the above factor observations.

We can choose any maturity bond to represent the **level (PC1)**, since all bonds have essentially the same change in the level. In the first part of Figure 3.6 we use the 10Y bond to see how close PC1 fits, indeed the green (PC1) and orange (DGS10) curves are almost a match.

Since the **slope factor (PC2)** expands short-term bonds by a much larger amount than long-term bonds, we can represent it as the difference between long-term and short-term bonds: 10Y−2Y. The second part of Figure 3.6 shows that although the error is larger than in the previous case, the tendency of the curves is similar.

The coefficient **curvature (PC3)** describes the fact that changes in short-term and long-term bonds go in one direction while medium-term bonds move in another, so let us represent it in the following way: 30Y − 2 × 10Y + 5Y. Indeed, the third part of Figure 3.6 shows that our assumption is very close.

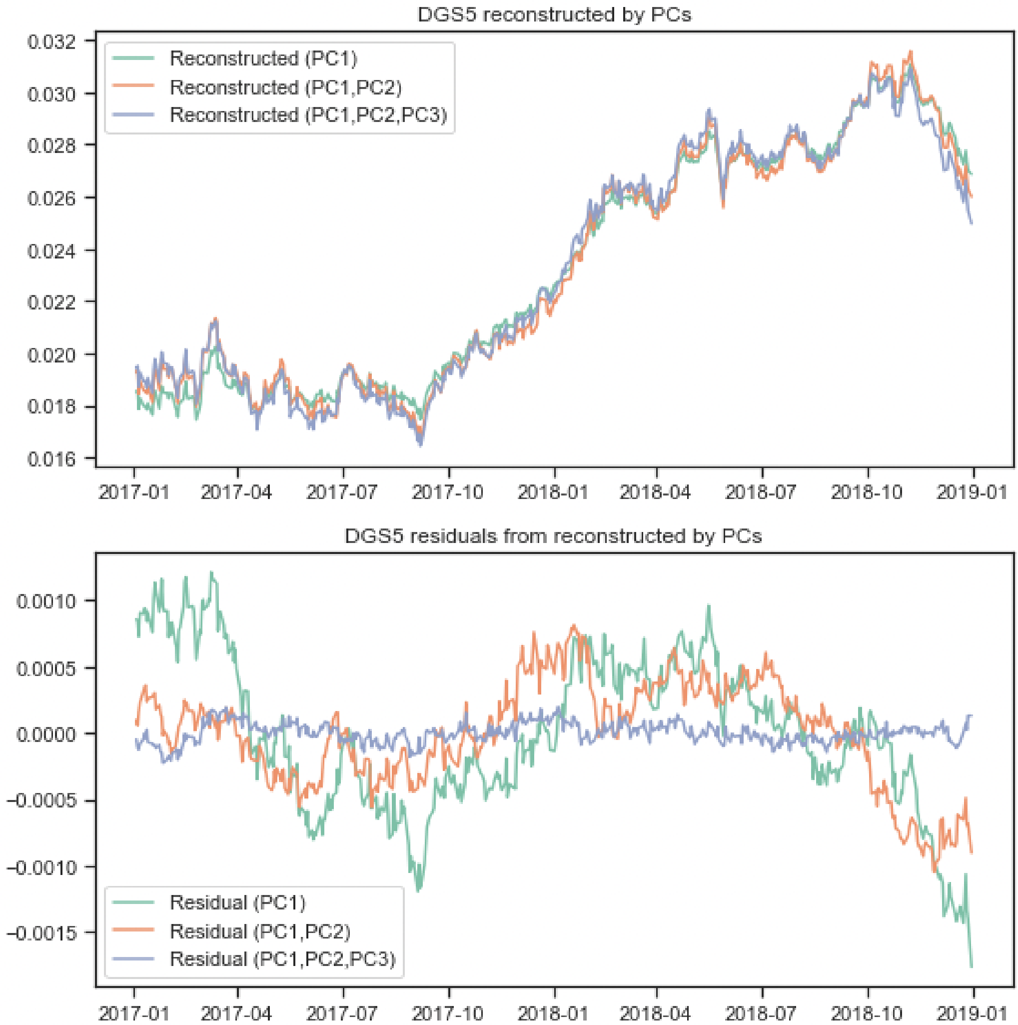


Figure 3.7: Reconstructed series from PC:DGS5

3.4.3 Representation of X with PCs

When the underlying factors have been found from PCA, we can finally use the first three principal components to approximate the original data X. As we observed above (Figure 3.4), almost 100% of the variance can be explained by the first three components therefore, we will use those components to construct $\tilde{X}^{(3)}$ using the equations (2.12) and (2.13). The corresponding residual $R^{(3)}$ could be calculated via equation (2.15).

The upper figure 3.7 shows the approximation of bond yield 5Y by all PCs in aggregate, and the lower figure 3.7 shows the corresponding residuals. It can be seen that even just PC1 is quite adequate to obtain a very close result with an error of the order of 10^{-3}.

Figure 3.8 shows the 10Y bond approximation: PC1 has the dominant explanatory power, its residuals are about 10^{-3}. Adding PC2 improves the reconstruction so that the residuals are roughly 10^{-4}, and including PC3, in addition, makes no appreciable difference. This means that PC1 and PC2 are sufficient to provide an explanation of the dynamics.

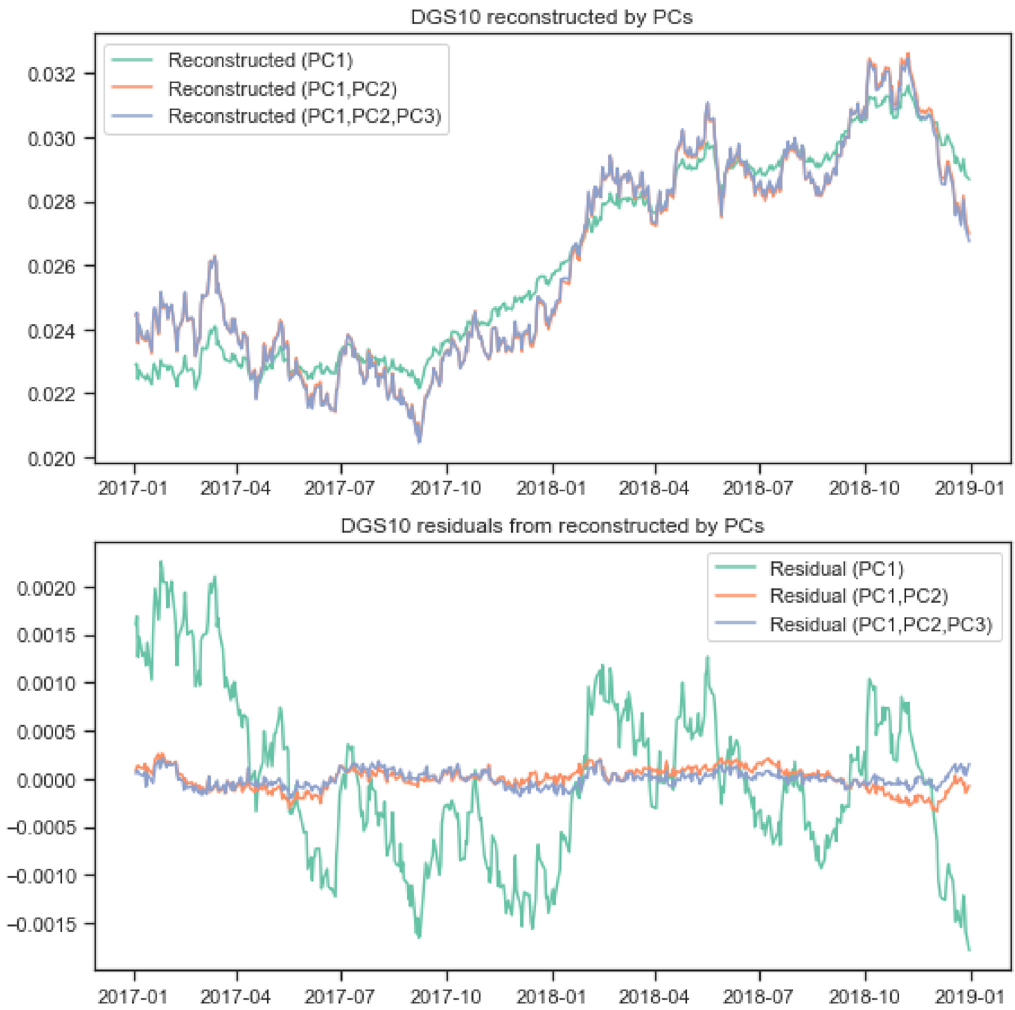


Figure 3.8: Reconstructed series from PC:DGS10

Our last approximation is for 30Y bonds, and it is presented in Figure 3.9. In comparison to the 10Y curve, PC2 is more significant in explaining the 30Y trends. Moreover, in this case, the contribution of PC3 is much more substantial.

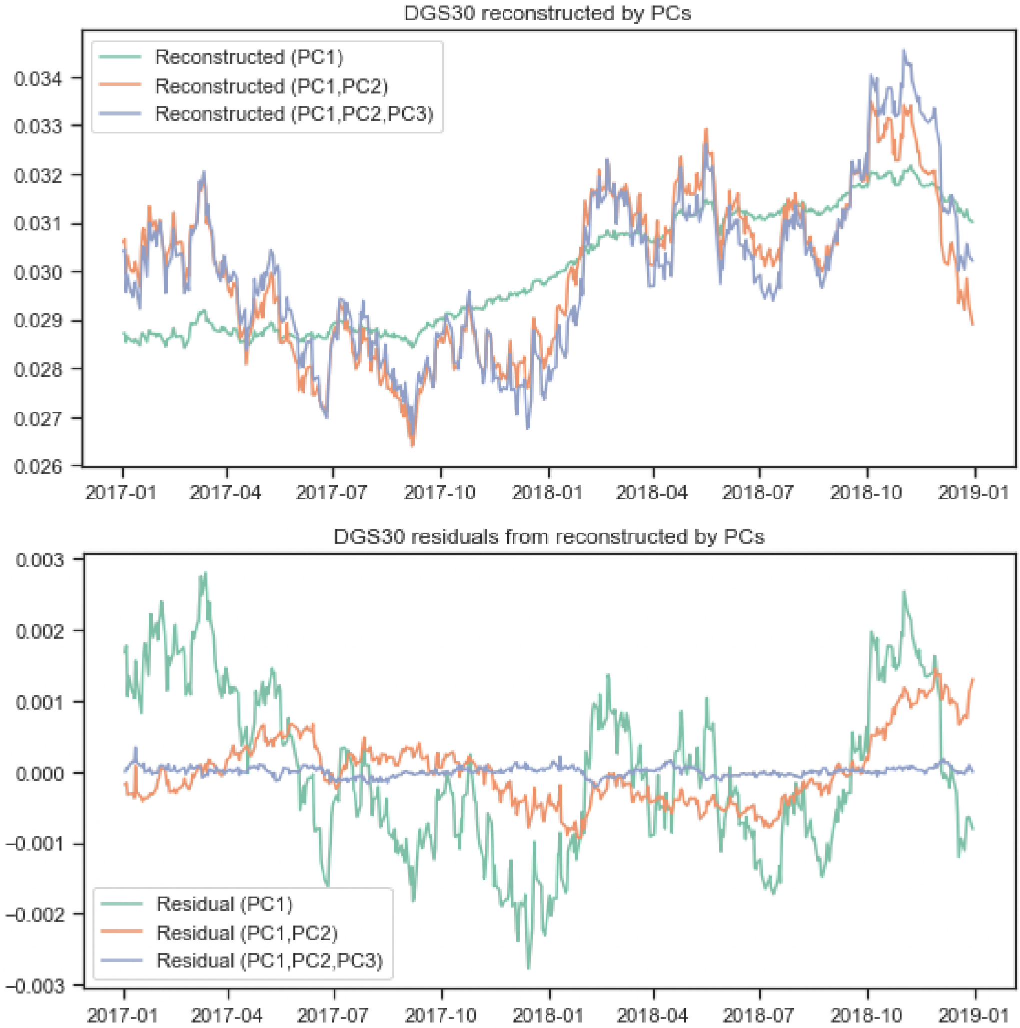


Figure 3.9: Reconstructed series from PC:DGS30

3.5 Term structure analysis during COVID-19 crisis

3.5.1 Why this timeline?

The goal was to find a timeline illustrating how US Treasury bonds behave in times of global recession. Therefore, the choice of the year 2020 was natural, as the COVID-19 pandemic caused an unprecedented human resource and health crisis. The measures required to contain the spread of the virus triggered an economic downturn. Prices of risky assets have fallen precipitously since the onset of the pandemic, with prices of risky assets at the lower end of the recent surge selling off at half or more of their declines observed in 2008 and 2009. For instance, many equity markets, both large and small, have recorded price declines of 30 per cent or more at the low end. The *S&P Index* covers approximately 80% of US market capitalisation and indicates key market sentiment and trends. The provider itself, Standard & Poor's Ratings Services [13], calculates that the S&P 500 Index tracks portfolios totalling more than 4.6 trillion US Dollars. Indeed, Figure 3.10 shows a dramatic fall in February and March. In Figure 3.11, one can see that US Treasuries also took a fall in the same period, although not as big as the S&P index.

Figure 3.10: S&P Index

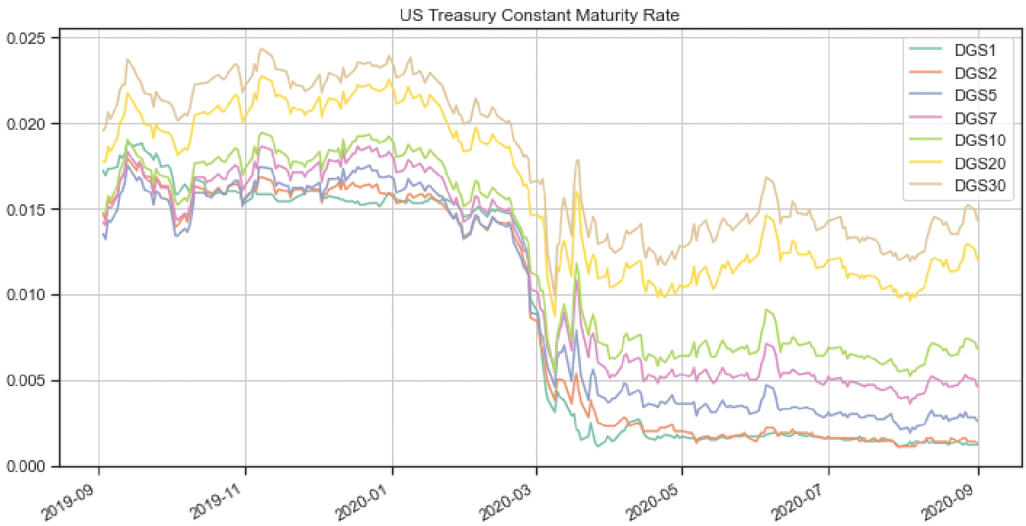


Figure 3.11: US Treausry constant maturity rates

3.5.2 Data and Results

The dynamics of the term structure, shown in Figure 3.12, are slightly different in this period than in our previous case, see Figure 3.2. We definitely see a change in the level, but all the curves are much closer to the slope: we can see a more evident difference in this period.

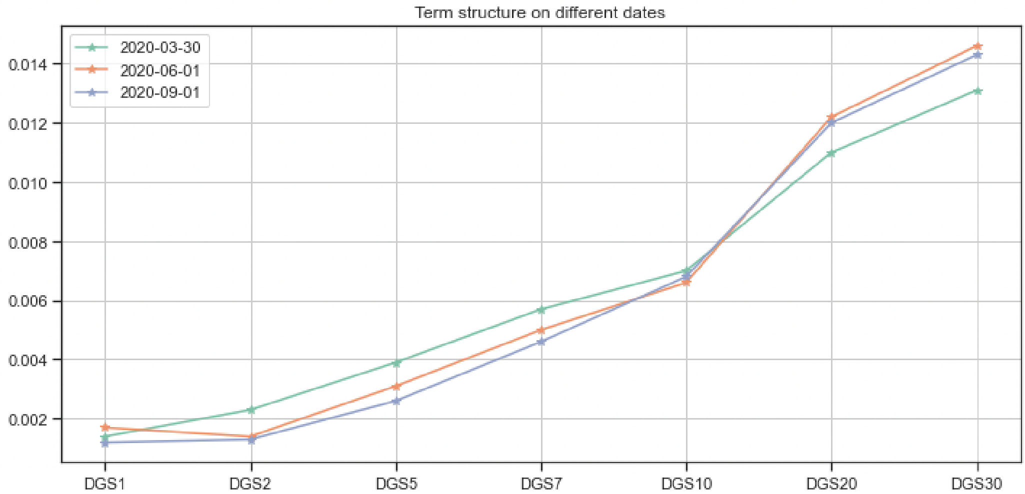


Figure 3.12: Term structure of interest rates

Note that all the curves are still positively sloped, but they are now convex instead of concave. It can be read as a tendency of the yield curve to flatten at long maturities. The flat curve suggests that short- and long-term bonds offer equivalent returns, usually little benefit from holding the longer-term instrument; the investor receives no additional compensation for the risks of holding long-term securities. A flat yield curve usually indicates that investors and traders are concerned about the macroeconomic outlook.

Figure 3.13 shows that PC1 plays an even more significant role here. It explains about 98% of the variance, so the other PCs contribute less than in the first example.

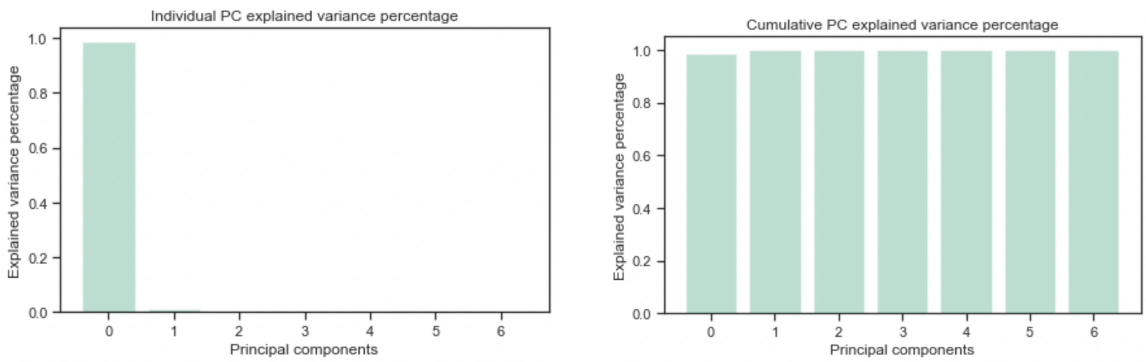


Figure 3.13: PCA-explained variance ratio

We can finally observe the PC approximation for the 10Y and 30Y bonds. In the top Figure 3.14, we can see that even just the first PC gives a good approximation. Although if we look at the residuals, it is evident that during the period of maximum decline in March 2020, it is difficult to reconstruct this spectacular fall by the only PC1. We can also remark that PC2 improves the result, while PC3 has almost no effect.

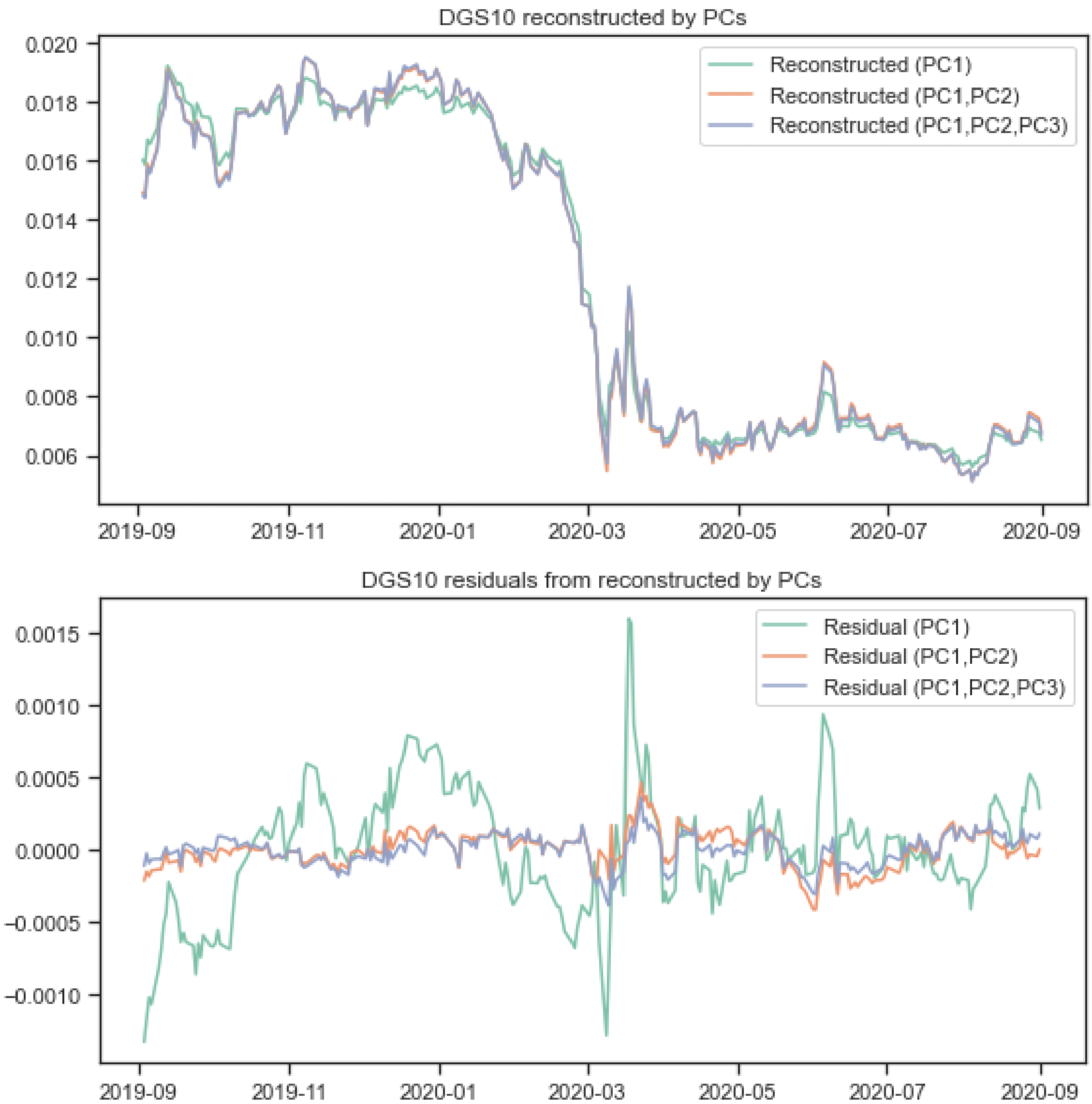


Figure 3.14: Reconstructed series from PC:DGS10

For 30Y bonds, the picture is even worse: the top Figure 3.15 shows that PC1 closely recovers the general direction, but even small picks are lost. The lower Figure 3.15 shows a larger residual PC1 value than for 10Y bonds. PC3 plays a much more crucial role in this case, only when using the three PCs together do we see a residual value close to zero.

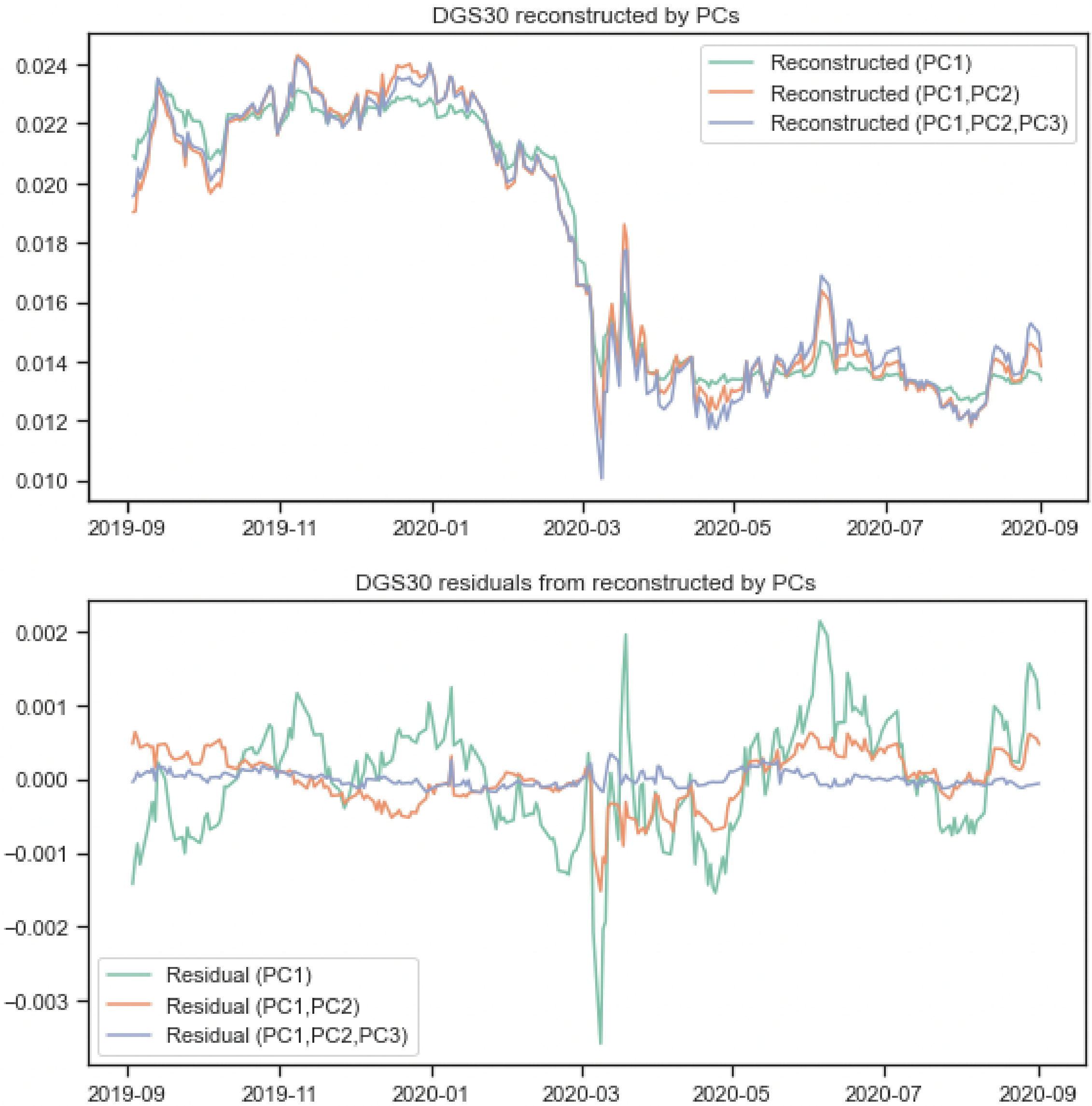


Figure 3.15: Reconstructed series from PC:DGS30

4 PCA for hedging

In this chapter, we will concentrate on the application of PCA for hedging portfolio risk.

4.1 Hedging

Hedging is the way to ensure financial risks by taking an opposite position on an asset in the market. Hedging against investment risk means strategically using financial instruments or market strategies to offset the risk of any adverse price movements. In other words, investors hedge one financial instrument by making a trade with another. A hedging instrument is a financial instrument that protects against potential risks, and it could be currencies, securities, deposits, fixed-term contracts, etc.

Technically, in order to hedge an asset, one would have to make multidirectional trades in securities and financial instruments with a negative correlation. Of course, nothing in the financial world is free, so there will be some form of payment for this type of insurance. Reducing risk will always mean reducing potential profits. Thus, hedging is a tool designed not to increase profits but to decrease potential losses. If the assets an investor hedges continue to grow, it reduces his potential gain, but if the assets lose in value, the hedge reduces the size of the loss.

One needs to answer the following questions in order to make a hedging decision [1]:

- **What is the current risk? Can we quantify it?**

  When the yield curve moves, the value of the bond portfolio shifts accordingly. From the previous sections, we know that by PCA, the yield curve can be decomposed into three components: level, slope and curvature.

- **What types of risk can be taken? How often is hedging planned?**

  It depends on two factors: the strategy of the individual investor and macroeconomic factors: the more volatile the situation, the more often hedging is necessary.

- **Which instruments are chosen for hedging?**

  The main hedging instruments are financial derivatives, i.e. forward exchange contracts (futures and options). Liquid products are generally a good hedging instrument.

One easy way to use PCA for hedging is to simply have a look at the loading matrix (Figure 3.5), we can find those few bonds with the highest correlations with PC1 and PC2 and choose them for further hedging [14].

In this section, we will observe a different method.

4.2 Risk representation

Let us take a yield term structure involving bonds with maturities p. At time t, we define the yield of maturity T_m as r_t^m and the value of the respective bond as P_t^m. Then the whole yield term structure is $r_t = (r_t^1, ..., r_t^m)$, and its daily change dr of N entries is an $N \times p$ matrix with the following PC representation:

$$dr = (UD)V^T = ZV^T, \tag{4.1}$$

where Z is the $N \times p$ PCs and V is the $p \times p$ loading matrix.

The portfolio risk from the yield curve motion is called the **delta risk**.The delta risk of bond m with respect to maturity k is

$$\delta_t^{m,k} := \frac{\partial P_t^m}{\partial r_{t,k}},$$

therefore the price change is given as

$$dP_t^m = \sum_{t=1}^{k} \frac{\partial P_t^m}{\partial r_{t,k}} dr_{t,k} = \sum_{t=1}^{k} \delta_t^{m,k} dr_{t,k}. \tag{4.2}$$

Now we will rewrite the price change in terms relevant to the PC algorithm:

$$dP_t^m = \sum_{t=1}^{p} \frac{\partial P_t^m}{\partial Z_{t,j}} dZ_{t,j} \tag{4.3}$$

where $\frac{\partial P_t^m}{\partial Z_{t,j}}$ is jth PC risk exposure showing the sensitivity of the price of bond m to the jth PC. From Equation (4.1), we get

$$\frac{\partial P_t^m}{\partial Z_{t,j}} = \sum_{k=1}^{m} \frac{\partial P_t^m}{\partial r_{t,k}} \frac{\partial r_{t,k}}{\partial Z_{t,j}} = \sum_{k=1}^{m} \delta_t^{m,k} v_{k,j}. \tag{4.4}$$

Thus, if we assume that the price change can be largely explained by the first three PCs, then from equations (4.2) and (4.3) we have

$$dP_t^m = \sum_{j=1}^{p} \left(\sum_{k=1}^{m} \delta_t^{m,k} v_{m,j} \right) dZ_{t,j}. \tag{4.5}$$

Equations (4.2) and (4.5) give two different representations of risk.

4.3 Hedge level and slope with PCA

Let us select a bond T_k to hedge the exposure of the first PC (level) and solve the hedge quantity $w_{t,1}^k$ using the following equation:

$$\sum_{m=1}^{p} \frac{\partial P_t^m}{\partial Z_{t,1}} = w_{t,1}^k \frac{\partial P_t^m}{\partial Z_{t,1}}. \tag{4.6}$$

This equation can be easily generalized to hedge two PCs (level and slope):

$$\sum_{m=1}^{p} \frac{\partial P_t^m}{\partial Z_{t,1}} = w_{t,2}^{k_1} \frac{\partial P_t^{k_1}}{\partial Z_{t,1}} + w_{t,2}^{k_2} \frac{\partial P_t^{k_2}}{\partial Z_{t,1}}, \tag{4.7}$$

$$\sum_{m=1}^{p} \frac{\partial P_t^m}{\partial Z_{t,2}} = w_{t,2}^{k_1} \frac{\partial P_t^{k_1}}{\partial Z_{t,2}} + w_{t,2}^{k_2} \frac{\partial P_t^{k_2}}{\partial Z_{t,2}}. \tag{4.8}$$

A similar technique can be used to hedge higher-order PCs.

5 Conclusion

In this paper, we examined the PCA method itself, the mathematical principles underlying it, and after applied it to a practical example: an analysis of U.S. Treasures with different maturities and over different time horizons.

PCA is a mathematical method for dimensionality reduction, which converts correlated input data into an uncorrelated output data set whose variance is maximized; as a result, the loss of information is reduced [14]. The idea behind PCA in the context of the term structure is that most of the yield curve movements can be expressed as a set of two or three independent driving factors, the principal components. More precisely, the PCs are constructed to explain the largest fraction of the total variance of the data set without any overlap. One can understand them as axes that form a new coordinate system (see Figure 2.1). The first PC contains the maximum amount of the variance and, in our application, has a meaning of the yield curve level; it explains most of the movements of the term structure, up to 90%. The second and the third components typically only add another 5 - 10% to the explanatory power of the method; they represent the slope and the curvature, respectively.

Another powerful application of PCA is presented in the last chapter, where we introduce the hedging mechanism. We show the representation of the risk in terms of PCs and calculate the hedge level for our components. Thus the method is beneficial for searching and correcting investors' trading strategies.

Nowadays, in an era of big data, PCA is a powerful tool for removing surplus information from a large data set, resulting in further simplified analysis, interpretation and visualization. Moreover, it can be used as a pure analytical instrument, as illustrated in the last section.

6 Appendix: Python code

6.1 Jupyter Notebook Code

```python
import matplotlib
matplotlib.use('TkAgg')
import seaborn as sns
sns.set(style="ticks", color_codes=True)
import matplotlib.cm as cm
import matplotlib.pyplot as plt
import matplotlib as mpl
%matplotlib inline

import pandas as pd
import numpy as np
from datetime import datetime
from sklearn.preprocessing import scale
from sklearn.decomposition import PCA
#mpl.rcParams['figure.dpi'] = 200

from IPython.core.display import HTML
HTML("<style>.container {width:98% !important; }</style>")
plt.rcParams['axes.facecolor'] = 'white'
import seaborn as sns
plt.rcParams['axes.facecolor'] = 'white'
sns.set_palette("Set2")

pip install fredapi

from fredapi import Fred
fred = Fred(api_key='cd9a27c9afcd4ee82ec0be135fb8b223')

# get data
startDate = '2019-09-01'
endDate = '2020-09-01'
df = []
ids = ['DGS{}'.format(i) for i in [1,2,5,7,10,20,30]]
for s in ids:
    df.append(fred.get_series(s, observation_start=startDate,observation_end=endDate)/100)

df = pd.concat(df,axis=1)
df.columns = ids
```

```python
df = df.dropna()

# curve dynamic
fig,(ax,ax2)=plt.subplots(nrows=2,ncols=1,figsize=(10,5*2))
df.plot(grid=True, title='US Treasury Constant Maturity Rate', ax=ax)
x = df.loc[df.index.intersection([datetime(2019,9,1),datetime(2019,12,15),datetime(2020,
x.index = [t.date() for t in x.index]
ax2.plot(x.T.index,x.T,marker='*')
ax2.legend(x.index)
ax2.grid(True)
ax2.set_title('Term structure on different dates')
fig.tight_layout()

# contruct pca object
from pca import PCABase
pcab = PCABase(df)

# loading matrix (direction may change but doesn't matter)
V = pd.DataFrame(pcab.pca().components_,index=pcab.pc_names(pcab.n_features),columns=pca
V.T.iloc[:,0:3].plot(figsize=(10,5),kind='bar')
pcab.cps()

# resconstruction and residuals
r = 'DGS10'
fig,(ax,ax2)=plt.subplots(figsize=(8,4*2),ncols=1,nrows=2)
ax.plot(pcab.x_projected(1)[r])
ax.plot(pcab.x_projected(2)[r])
ax.plot(pcab.x_projected(3)[r])
ax.legend(['Reconstructed (PC1)','Reconstructed (PC1,PC2)','Reconstructed (PC1,PC2,PC3)'])
ax.set_title('{} reconstructed by PCs'.format(r))

ax2.plot(pcab.residuals(1)[r])
ax2.plot(pcab.residuals(2)[r])
ax2.plot(pcab.residuals(3)[r])
ax2.legend(['Residual (PC1)','Residual (PC1,PC2)','Residual (PC1,PC2,PC3)'])
ax2.set_title('{} residuals from reconstructed by PCs'.format(r))
fig.tight_layout()

# resconstruction and residuals
r = 'DGS30'
fig,(ax,ax2)=plt.subplots(figsize=(8,4*2),ncols=1,nrows=2)
```

```
ax.plot(pcab.x_projected(1)[r])
ax.plot(pcab.x_projected(2)[r])
ax.plot(pcab.x_projected(3)[r])
ax.legend(['Reconstructed (PC1)','Reconstructed (PC1,PC2)','Reconstructed (PC1,PC2,PC3)'])
ax.set_title('{} reconstructed by PCs'.format(r))

ax2.plot(pcab.residuals(1)[r])
ax2.plot(pcab.residuals(2)[r])
ax2.plot(pcab.residuals(3)[r])
ax2.legend(['Residual (PC1)','Residual (PC1,PC2)','Residual (PC1,PC2,PC3)'])
ax2.set_title('{} residuals from reconstructed by PCs'.format(r))
fig.tight_layout()

# PC Scores
fig,(ax1,ax2,ax3)=plt.subplots(nrows=3,ncols=1,figsize=(8,3*3))
l1=ax1.plot(pcab.scores()['PC1'])
ax12 = ax1.twinx()
l2=ax12.plot(pcab.X['DGS10'],color='orange')
ax1.tick_params('x',rotation=30)
ax1.legend(l1+l2,['PC1 score','DGS10'])
ax1.set_ylabel('score')
ax12.set_ylabel('DGS10')

l1=ax2.plot(pcab.scores()['PC2'])
ax22 = ax2.twinx()
l2=ax22.plot(pcab.X['DGS10']-pcab.X['DGS2'],color='orange')
ax2.tick_params('x',rotation=30)
ax2.legend(l1+l2,['PC2 score','DGS10-DGS2'],loc='best')
ax2.set_ylabel('score')
ax22.set_ylabel('DGS10-DGS2')

l1=ax3.plot(pcab.scores()['PC3'])
ax32 = ax3.twinx()
l2=ax32.plot(pcab.X['DGS30']-2*pcab.X['DGS10']+pcab.X['DGS5'],color='orange')
ax3.tick_params('x',rotation=30)
ax3.legend(l1+l2,['PC3 score','DGS30-2DGS10+DGS5'],loc='best')
ax3.set_ylabel('score')
ax32.set_ylabel('DGS30-2DGS10+DGS5')
fig.tight_layout()

# PC
fig,ax=plt.subplots(figsize=(10,5))
l1=pcab.scores().iloc[:,0:3].plot(ax=ax)
```

```python
ax.grid(True)
ax.set_title('Evolution of PCA Factors 1,2,3')

# PCA-explained variance ratio

fig,(ax,ax2) = plt.subplots(figsize=(6*2, 4),ncols=2,nrows=1)
ax.bar(range(pcab.n_features), pcab.cumsum_expvar_ratio()[0], alpha=0.5,
    align='center')
ax.set_ylabel('Explained variance percentage')
ax.set_xlabel('Principal components')
ax.set_title('Individual PC explained variance percentage')

ax2.bar(range(pcab.n_features), pcab.cumsum_expvar_ratio()[1], alpha=0.5,
    align='center')
ax2.set_ylabel('Explained variance percentage')
ax2.set_xlabel('Principal components')
ax2.set_title('Cumulative PC explained variance percentage')

fig.tight_layout()
```

6.2 PCA Class Code

```python
import pandas as pd
import numpy as np
from datetime import datetime
from sklearn.preprocessing import scale
from sklearn.decomposition import PCA

class PCABase(object):
    def __init__(self, X, adjust_sign=True):
        self.X = X
        self.n_features = X.shape[1]
        self.dates = X.index
        self.Xc = self.X - self.X.mean()  # centered
        self.pc_names = lambda n: ['PC' + str(i) for i in np.arange(1, n +
            1)]
        self.adjust_sign = adjust_sign


    def pca(self, n_pc=None):
        '''
        fit pca model
        n_pc: number of pcs to fit, take total feature numbers if not
    specified
```

```
        '''
        if n_pc:
            model = PCA(n_components=n_pc).fit(self.Xc)
        else:
            model = PCA().fit(self.Xc)
        return model

    def cps(self):
        '''
        loading matrix => principal axes in feature space
        '''
        cps = self.pca().components_.T
        cps = self.to_df_pc(cps, is_loading=True)
        if self.adjust_sign:
            cps.loc[:, 'PC1'] = np.sign(cps.loc[:, 'PC1'].values[0]) * cps.loc[:, 'PC1']
            cps.loc[:, 'PC2'] = -np.sign(cps.loc[:, 'PC2'].values[0]) * cps.loc[:, 'PC2']
        return cps

    def cumsum_expvar_ratio(self):
        var_exp = self.pca().explained_variance_ratio_
        var_exp_cumsum = np.cumsum(var_exp)
        return var_exp, var_exp_cumsum

    def scores(self):
        '''
        PC scores:
        '''
        scores = self.pca().transform(self.Xc)
        scores = self.to_df_pc(scores)
        if self.adjust_sign:
            cps = self.cps()
            scores.loc[:, 'PC1'] = np.sign(cps.loc[:, 'PC1'].values[0]) * scores.loc[:, 'PC1']
            scores.loc[:, 'PC2'] = -np.sign(cps.loc[:, 'PC2'].values[0]) * scores.loc[:, 'PC2']
        return scores

    def scores2(self):
        '''
        equivalent to the sklearn transform function
        '''
        scores = self.Xc.dot(self.cps())
        scores = self.to_df_pc(scores)
```

```
        if self.adjust_sign:
            cps = self.cps()
            scores.loc[:, 'PC1'] = np.sign(cps.loc[:, 'PC1'].values[0]) * scores.loc[:, 'PC1']
            scores.loc[:, 'PC2'] = -np.sign(cps.loc[:, 'PC2'].values[0]) * scores.loc[:, 'PC2']
        return scores

    def x_projected(self, p, centered=False):
        xp = self.scores().iloc[:, 0:p].dot(self.cps().T.iloc[0:p, :])
        if not centered:
            xp = xp + self.X.mean()
        return xp

    def residuals(self, p):
        residuals = self.X - self.x_projected(p, centered=False)
        return residuals

    def covX(self):
        return self.X.cov()

    def eigenv(self):
        eig_vals, eig_vecs = np.linalg.eig(self.covX())
        return eig_vals, eig_vecs

    def to_df_pc(self, data, is_loading=False):
        cols = self.pc_names(self.n_features)
        idx = self.X.columns if is_loading else self.dates
        return pd.DataFrame(data, columns=cols, index=idx)
```

Bibliography

[1] H. Ni, X. Dong, J. Zheng, and G. Yu, *An Introduction to Machine Learning in Quantitative Finance*. World Scientific, 2021.

[2] https://github.com/deepintomlf/mlfbook.

[3] https://en.wikipedia.org/wiki/Covariance.

[4] T. H. João Moreira, Andre Carvalho, *A General Introduction to Data Analytics*. Wiley, 2019.

[5] M. Ivanov, "Comparison of pca with ica from data distribution perspective," 2017.

[6] https://www.investopedia.com/terms/b/bond.asp.

[7] L. Benzoni, O. Chyruk, and D. Kelley, "Why does the yield-curve slope predict recessions?," 2018.

[8] https://fred.stlouisfed.org/docs/api/fred.

[9] W. A. Strauss and T. Haasl, "Economic outlook symposium: Summary of 2017 results and 2018 forecasts," 2018.

[10] W. A. Strauss and K. Sarussi, "Economic outlook symposium: Summary of 2018 results and 2019 forecasts," 2019.

[11] T. Wu, "Macro factors and the affine term structure of interest rates," January 2005.

[12] T. Wu, "What makes the yield curve move?," October 2003.

[13] https://www.spglobal.com/ratings/en.

[14] S. Kumar, "Effective hedging strategy for us treasury bond portfolio using principal component analysis," January 2022.