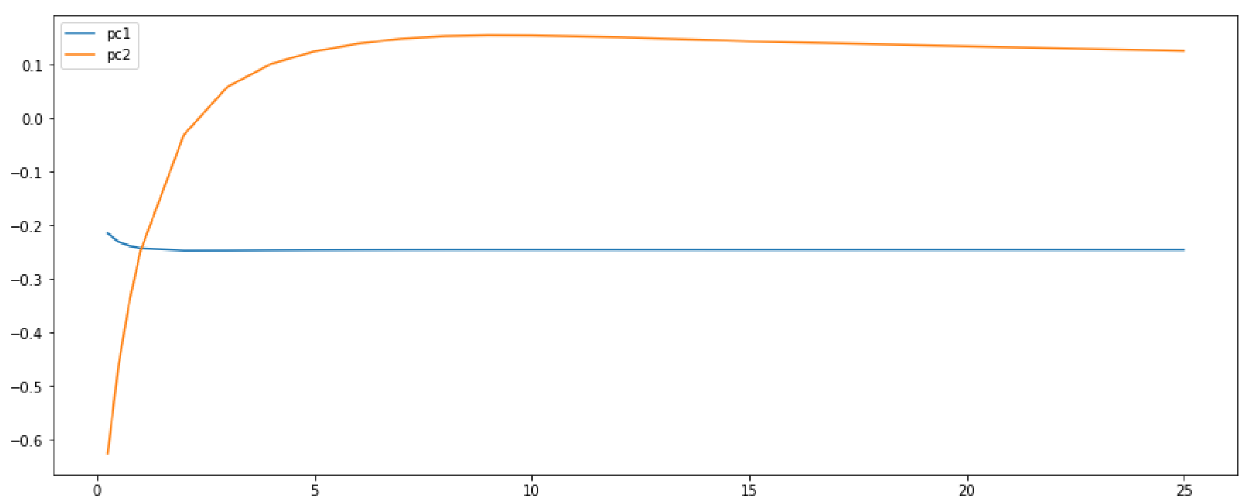
Assignment 2

Statement: We want to understand and quantify the major movements in the daily change of the yield curve. Movements may include, but not limited to things like the parallel shift of curve up or down, steeping/flattening of the curve, etc. To understand these “drivers”, we want to do a PCA on the data set. Let’s say we have three important aspects to look at – eigenvectors, eigenvalues, and proportion of variance. What inferences can one draw looking at the shapes (of eigenvectors) and other associated numbers? What are reasonable assumptions to make while doing this analysis?

Some observations:

1. The data has been cleaned manually by clearing two rows as there was a mismatch in the dtype of the columns. The columns have been renamed and the summary of the data has been presented in the ipynb file attached.
2. The correlation map shows a significant correlation among longer maturities and shorter maturities show less correlation with the longer ones.
3. The plot of the time series shows very similar movements in all the maturity IRS.
4. Also, on extensive analysis, most of the volatility was observed in March 2020, which is very obvious owing to the coronavirus pandemic.
5. Then, the data has been scaled and preprocessed for the PCA procedure.
6. From PCA, we observe that the first two components explain 99.8% of the variance (96% for 1st and 3.5% for 2nd).
7. The shape of the eigenvectors are given below:



The major observations here are:

1. The Eigenvalues corresponds to the amount of variance explained by the principal components.
2. The Eigenvectors for the first two principal components show different movements in the yield curve.

Principal Component 1:

The first component of the PCA decomposition explains 96% of the variation. This has the same sign in each of the maturity and thus reflects the parallel move in swap rates of all the maturity. It must be noted that the signs in the PCA decomposition are arbitrary and thus same sign in all the loadings represents the **directional movement**. This is consistent with the previous literature in the domain. Also, given the shape of the PC-1, the longer end is likely to move more than the shorter end.

Principal Component 2:

This component explains an additional 3.6% variation. This shows a negative sign in the shorter maturities (<=2 years) and a positive sign in the longer maturities. Again, since the signs are arbitrary, we can conclude the shorter and middle + longer maturities move in opposite directions, thereby representing a **slope/tilt** component. Thus, this can be very helpful to structure non-directional trades hedged against directional impacts.

Also, the analysis post-COVID also give very similar results. It must be noticed that Principal Component 3, which in our case however doesn't explain much variance, represents the curvature i.e., the shorter and longer maturity swap rates move in the same direction but the middle maturities move in the opposite direction, as discussed in various practitioner's works. This gives us signs to set up a butterfly trade in order to pocket the spread.

Also, the caveats here to look at the fact that the factors/drivers can become correlated in shorter intervals, and thus the assumptions of uncorrelated factors may break down. Thus, any strategy built upon the PCA components needs to be verified over shorter spans before proceeding. Also, the approach which utilizes only co-movements between maturities and ignores the volatility of the individual interest rates in the term structure can breakdown at times of major macroeconomic events. Also, PCA being a linear setup ignores the non-linear correlations which may exist.