

**Part1:** The data was queried using the Fred API: <https://github.com/mortada/fredapi> to obtain 1, 2, 3, 5, 7 and 10 year constant maturity treasury data series. Data of 1Y, 2Y, 5Y and 10Y go back to April 1953. 3Y goes back to June 1976 and 7Y goes back to July 1969.

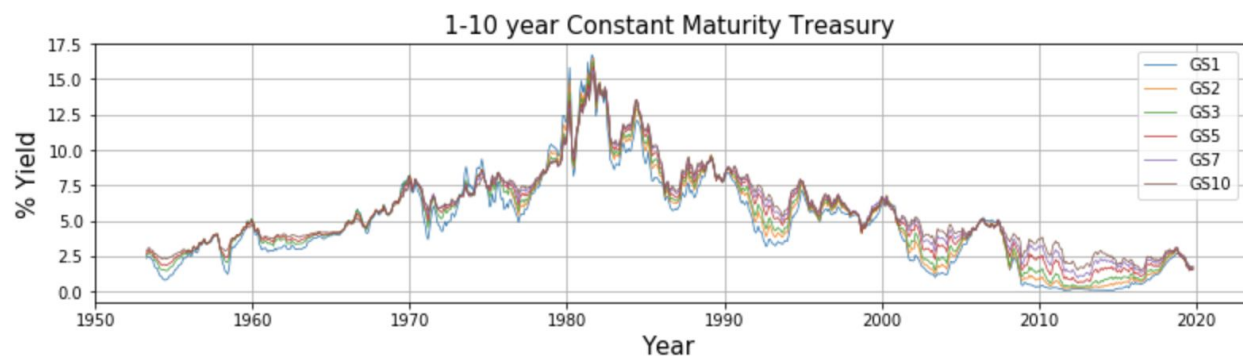


Figure 1: Constant maturity treasury data for all 6 data series

Initial analysis of the dataset shows a high correlation in the 6 data series. It is observed that a higher year Constant Maturity Treasury rate increases the percentage yield.

**Part 2:** Principal Component Analysis (PCA) was selected to reduce the dimensionality of the dataset. PCA transforms the dataset, reducing it to a selected number (3) of dimensions. As there is data for all 6 series from 1976 onwards, we initially perform PCA on all 1976 - Present.

The 1st principal component explains 99% of the variance, the 2nd component: 1% and the 3rd: a negligible amount. The stability of PCA is typically determined by the similarity of eigenvalues, near-equal eigenvalues indicating instability. As this PCA produces non-similar eigenvalues evident from the difference in explained variance %, this PCA is very stable. It follows that the first principal component of highly correlated datasets should explain a high proportion of the variance as the orthogonal transformation PCA uses leaves linearly uncorrelated components.

Plotting the 3 principal components against time in Figure 2, component 1 follows a very similar trend to that of Figure 1 (for the years included). Note the y-axis for reduced dimension data does not have physical meaning.

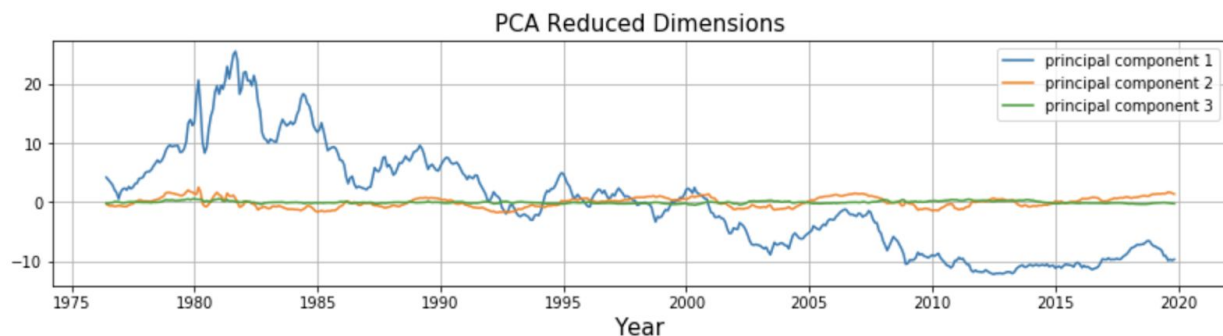


Figure 2: Three Principal Components over time

To look at the stability of the reduced components over time, the proportion of explained variation for each principal component is calculated using a 12 month rolling covariance matrix. This was calculated for 1953 - 1969 from 1Y, 2Y, 5Y, 10Y data, for 1969 - 1976, 7Y was incorporated and for 1976 - 2019, 3Y included. A time of 12 months was selected as plots using fewer numbers of months have higher noise due to a small number of data points. In Figure 3, the stability of the eigenvalues is represented by the fluctuation of principal components. Interpreting the results, component 1 explains the majority (70 - 100%) of the variation and during periods where this reduces, component two almost symmetrically increases to sum to nearly 100% of the variation. Component 3 generally explains less than 1% of the variation. This plot indicates that the reduced dimensions are reasonably stable over time however from 2004 - 2006 it is slightly more unstable.

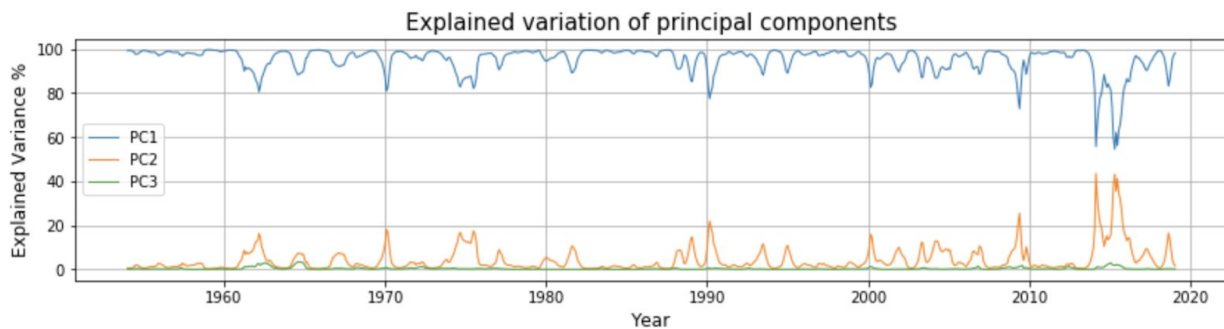


Figure 3: Explained variation proportion using a 12 month rolling covariance matrix

**Part 3:** Evaluating similarity for the historical data all 6 series. The first component of the reduced dimension dataset was used in Figure 1 as it is successful in explaining 99% of the variance. A suitable metric of average euclidean distance from each month in 2019 (Jan - Nov) to the corresponding month in each previous year was chosen. This was selected as there are just 11 data points to measure similarity on. This measure incorporates both trend and actual value of the 6 data series. Other time series similarity measures were considered such as dynamic time warping and Fourier transforms, however they were not deemed necessary for this data with so few data entries. Based on euclidean distance, the most similar year to 2019 is 2003. An additional check to see if any years prior to 1976 where the data was incomplete for all 6 series was carried out. PCA on the 4 data series spanning back to 1953 shows 2003 was still computed as the most similar year to 2019. Figure 4 compares of the first principal component of 2003 compared to 2019, note that because the dimensions have been reduced the y axis does not have any meaningful value and is an arbitrary scale.

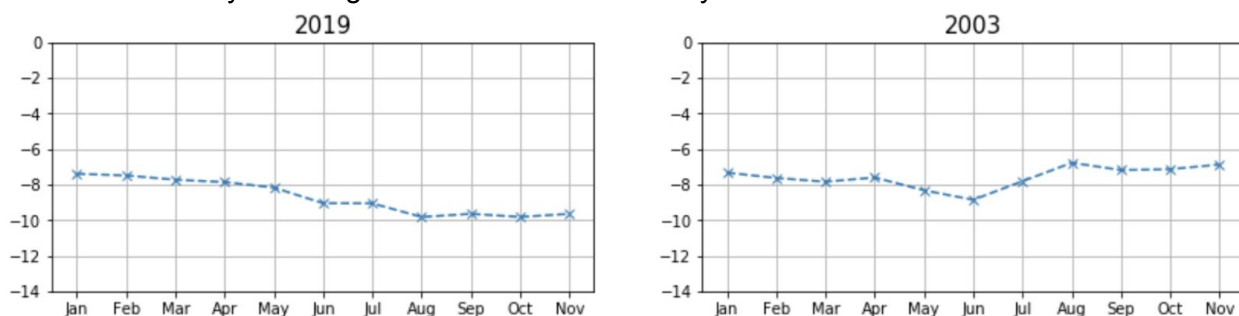


Figure 4: 1st principal component of 2019 compared with 2003