U.S. Treasury Yield Curve Data Analysis and Dimensionality Reduction using Principal Component Analysis

Lu Yang

1. U.S. Treasury Yield Curve Data

A yield curve is a graphical representation that shows the relationship between the interest rates (or yields) of bonds with different maturities but similar credit quality. It is a crucial tool in finance and economics, providing insights into future interest rate changes, economic activity, and investor sentiment. The U.S. Treasury yield curve plots the yields of U.S. Treasury securities across various maturities.

For this research project, we construct the yield curve using only a select few maturities: 1, 2, 3, 5, 7, and 10 years. The monthly data series are extracted from FRED using its API in Python with respective tickers: GS1, GS2, GS3, GS5, GS7 and GS10. Data is compared with those on FRED website to ensure the API works correctly.

The data utilized is the Constant Maturity Treasury (CMT) Yield, which are derived through interpolation from the yields of actual Treasury securities with maturities close to the respective target maturities. Although data is available on FRED starting from April 1953, the 2-year yield did not exist until June 1976, and the 7-year yield until July 1969. Consequently, the analysis uses data beginning from June 1976, when all six series are available.

Figure 1 illustrates the trend of the data over time. In Figure 2, yield curves for several selected dates are plotted. The recent observation of inverted yield curves, as shown, suggests an impending economic recession.

Figure 1 Market Yield on U.S. Treasury Securities (%)

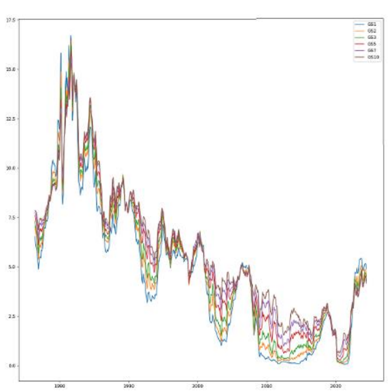


Figure 2 U.S. Treasury Securities Yield Curve (%)

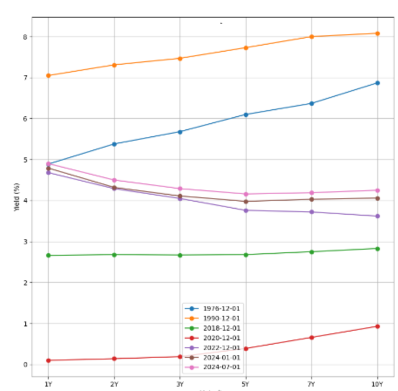
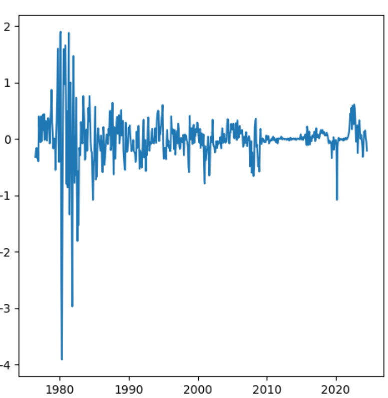


Figure 3 GS1 MoM change (%)



Differencing is a common method to detrend time series data and make non-stationary data more stationary. Monthly changes of the 6 series are derived by subtracting the previous month number from the contemporary number of each month. Figure 3 illustrates the GS1 MoM change as an example.

$$\Delta r_{maturity,t} = r_{maturity,t} - r_{maturity,t-1}$$

Where maturity ∈ {1, 2, 3, 5, 7, 10}; t ∈ {'1976-07-01', '1976-08-01', … …, '2024-07-01'}

2. Dimensionality Reduction

In order to reduce the dimensions of the dataset including 6 monthly change variables, Principal Component Analysis (PCA) is applied. PCA transforms the original variables (correlated) into a new set of uncorrelated variables called principal components (PCs), ordered by the amount of variance they capture from the data. In general, the first 2 or 3 PCs are able to capture greater than 95% of the total variance. Therefore, we are able to reduce the dimensionality from 6 to 2 for this dataset without losing much information.

By performing PCA, 6 principal components are generated. Figure 4 shows the first 2 PCs. As shown in the Scree plot of variance explained (Figure 5), the first PC explains 94% of the total variance, while the second explains 5%. In other words, the first 2 PCs accounts for 99% of the variance in the data.

Figure 4 Principal Components

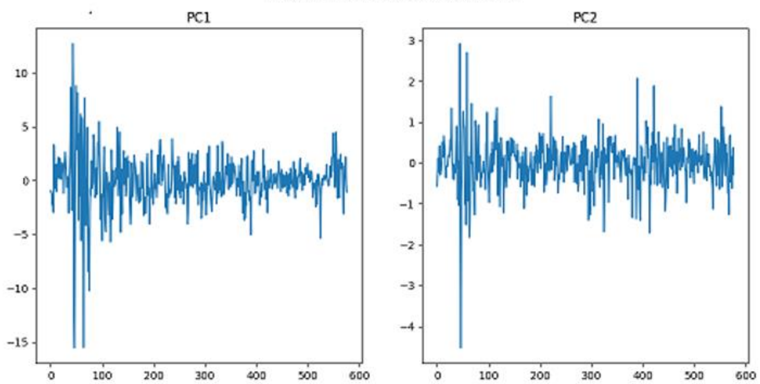
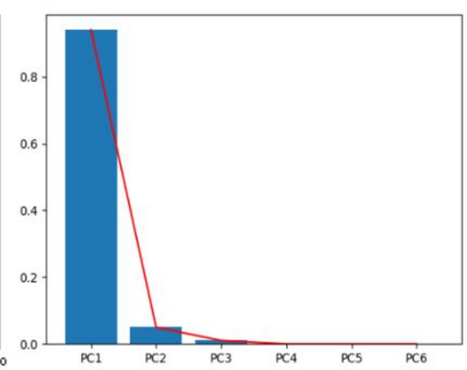
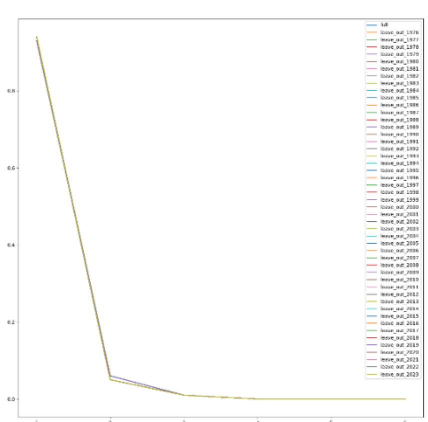


Figure 5 Variance Explained



Besides the mathematical interpretation of the first 3 PCs, in the context of analyzing Treasury yield, they have their business interpretation (with adjustment to the differencing transformation) – PC1, PC2 and PC3 represents, respectively, the position of the yield curve, slope and convexity.

Figure 6 Variance Explained from LOYO CV



Stability of PCA results is checked by performing a leave-one-year-out cross validation. Specifically, from 1976 to 2024, one year of data is left out and the rest is used to calculate PCs. I then look at the variances explained from all iterations, and the correlation between corresponding PCs. As shown in Figure 6, variances explained are almost the same. The correlation between PC1s are all close to 1. Both evidences suggest that the PCA result is stable.

As stable PCs obtained, we should be able to use them for other purposes such as building Factor Models, generating scenarios for stress testing and sensitivity analysis, estimating VaR and CVaR of an investment portfolio if using together with Monte-Carlo simulation, and portfolio immunization (hedging).