Using PCA to De-Noise Treasury Yields Data

CS305: Machine Learning
Victoria Lu

Problem and Motivation

Assuming that the shape of the Treasury yield curve is primarily determined by a couple of factors, one may use principal components analysis (PCA) to identify these factors. Further, one can use the difference between the actual yield and the PCA-implied yield (the residual) for each maturity as a proxy for whether a tenor is overvalued or undervalued.

Data Description and Cleaning

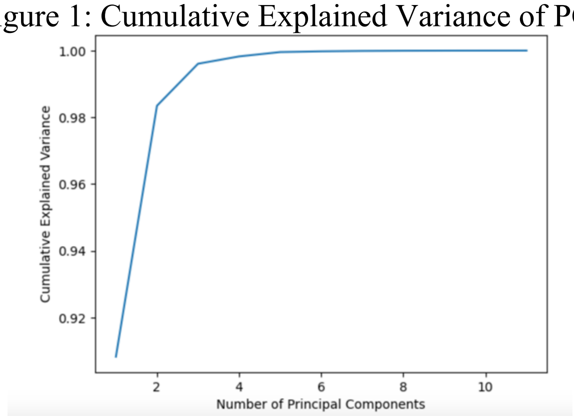
The raw dataset is from the US Treasury website and contains daily Treasury rates across the yield curve from 2016-01-01 to 2023-11-10. It has 1750 rows, each of which represents a day, and 14 columns, which include a date column and 13 columns that each represent a tenor (1 Mo, 3 Mo, 6 Mo, etc.). Due to the large number of missing values for the 2 Mo and 4 Mo tenors, I opted to drop these columns.

Methodology

PCA is a dimensionality reduction technique is used to transform high-dimensional data with d features into a lower-dimensional space with k features while retaining as much of the original variability as possible.

With my dataset, I first standardized the numeric columns and calculated the 10x10 covariance matrix, which represents how the tenors are correlated with each other. Then, I performed eigen-decomposition. The eigenvectors yielded from this step represent directions in the original feature space along which the data varies the most and the eigenvalues represent the amount of variance captured by each eigenvector where a bigger eigenvalue implies that the corresponding eigenvector explains more variance in the original feature space. By looking at the plot of cumulative explained variance (see Figure 1), one can see that the first three eigenvectors/principal components (PCs) sorted by eigenvalues explain over 99% of the variance in the original feature space. Given this, I opted to project the original data onto these three eigenvectors, thereby reducing the dimensionality of the dataset from 10 to 3.

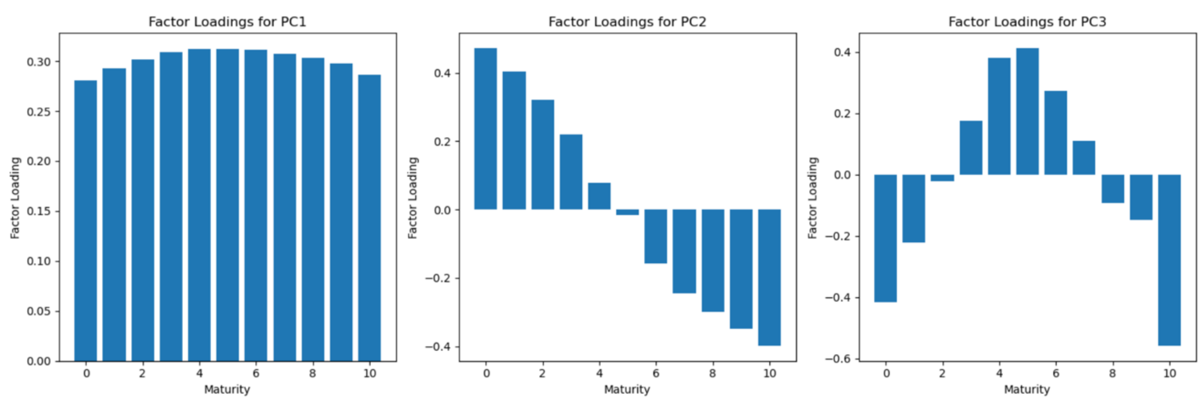
Figure 1: Cumulative Explained Variance of PCs



Model Results and Interpretation
By taking the dot product between the original data and the matrix of the first three principal components, I was able to arrive at the loadings matrix, which shows the weights of the original features in the linear combination that makes up each principal component and can offer some useful intuition (see Figure 2).

One can see that for PC1, the original features load positively and equally onto it. Even though the PCs cannot be precisely interpreted, one can infer from the loadings that PC1 represents the overall level of yields. For PC2, the loadings for shorter-tenors are positive, whereas those for longer-tenors are proportionately negative, meaning that an increase in short-term yields contributes positively to PC2 and an increase in long-term yields contributes negatively to PC2. Based on this, one can infer that PC2 represents the spread between short-term and long-term yields. Finally, for PC3, the loadings for the very short and long tenors are negative, whereas those for the intermediate tenors are positive. This means that PC3 likely captures the curvature of the yield curve.

Figure 2: Loadings for the First Three PCs



Discussion
Using the PCA-transformed data, one can further identify relative value opportunities in the Treasury market. By projecting the PCA-transformed data back into the original feature space, one can reconstruct the original data using only the first three PCs, and by subtracting this reconstructed data from the original data, one can obtain the residuals or "noise" in the original data. Figure 3 shows the residuals of all tenors over time, where a residual > 0 indicates a potential undervaluation (i.e., actual yield is higher than that implied by PCA) and a residual < 0 indicates a potential overvaluation. Overall, there does appear to be a mean-reversion pattern for the residual of all tenors, though there are no discernible patterns in how fast this reversion occurs. Taking any cross-sectional data, one can also easily visualize the relative value of different tenors across the curve at a point in time. For example, based on the PCA results, on 2023-11-01, the 1 Mo Treasury looks cheaper than the 3 Mo Treasury (see Figure 4).

Figure 3: Residuals of Different Tenors from 2016-2023

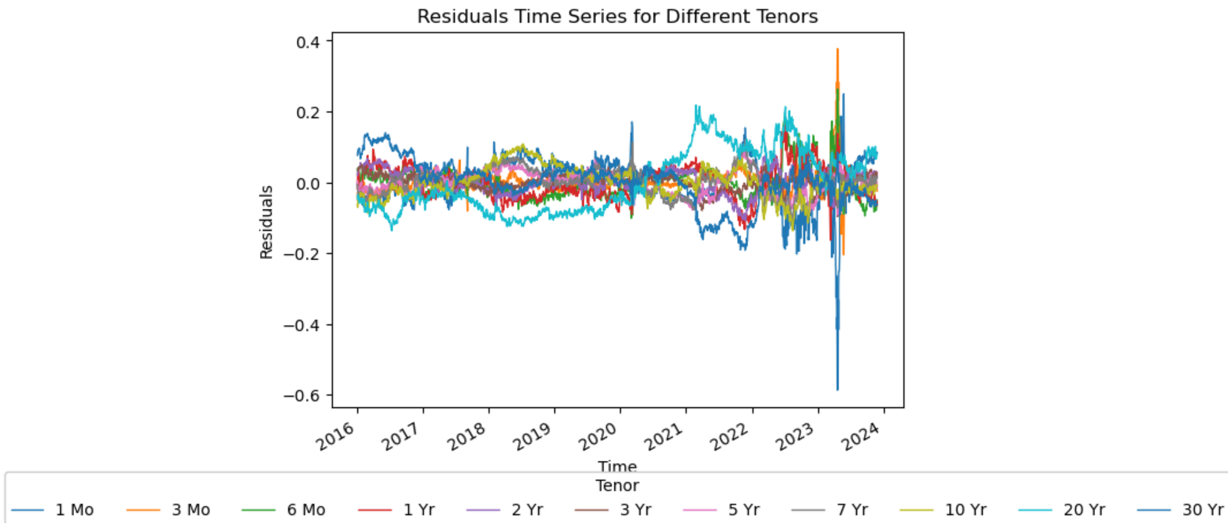
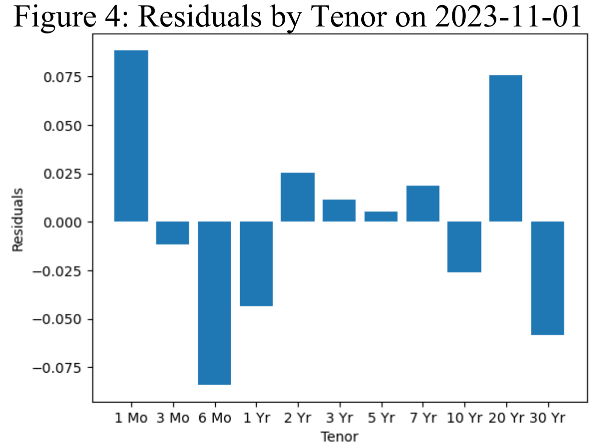


Figure 4: Residuals by Tenor on 2023-11-01



Limitations and Extensions

PCA makes a few assumptions that may not be satisfied in real life. Since the PCs are linear combinations of the original features, PCA does not consider any non-linear associations between the original features. This can be especially problematic during times of market stress, when psychological factors such as risk aversion, flight to safety, and sudden changes in market expectations come into play. To account for non-linear relationships, one could apply kernel PCA instead, but with this approach, interpretability would be lost completely.

PCA also assumes that the data is stationary, which in this case, means that the statistical properties of the yield curve does not change over time. This assumption can become violated when there are structural shifts in the economy. There are several potential ways to mitigate against this issue. One could choose a shorter timeframe when building the PCA model, but a smaller sample size could also mean less reliable results. Alternatively, one could also first use time-series techniques to filter out temporal dependencies and then apply PCA.

Finally, it is worth noting that while this project focuses on treasury yields, the same approach can be, in theory, used to de-noise any data related to rates curves (e.g., any swap curves or the yield curve of different countries). It would be interesting to compare the performance of the technique when applied to different scenarios.

Works Cited

Crump, R. K., & Gospodinov, N. (2019). Deconstructing the yield curve. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3368414

U.S. Department of the Treasury. (n.d.). *Daily Treasury Par Yield Curve Rates*. https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value_month=202312https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value_month=202312