# Principal Components Analysis in Finance
# Modeling the Dynamics of Interest Rates

Peter J. Kempthorne

March 29, 2016

The Treasury Yield Curve gives the annualized interest rate of US Treasury securities as a function of the *tenor* (time-to-maturity) of the security. The U.S. Federal Reserve maintains historical daily values for yields of Treasury securities. These data are used by finance professionals (risk managers, traders, investors) to model the dynamics of the yield curve. Such yield curve models are important tools used to value the price and risk of fixed-income securities (e.g., bonds). Treasury securities are guaranteed by the U.S. government so their yields provide benchmark interest rates for 'risk-free' bonds.

Figure 1 displays the Treasury Yield Curve for each of the first six (business) days of 2001. There is a curve for each day. Figure 2 displays how the yield curve changes from day to day by plotting the yield for each *tenor* over the six days.

**Figure 1. Yield Curve For 9 US Treasury Securities on 6 Days in 2001**

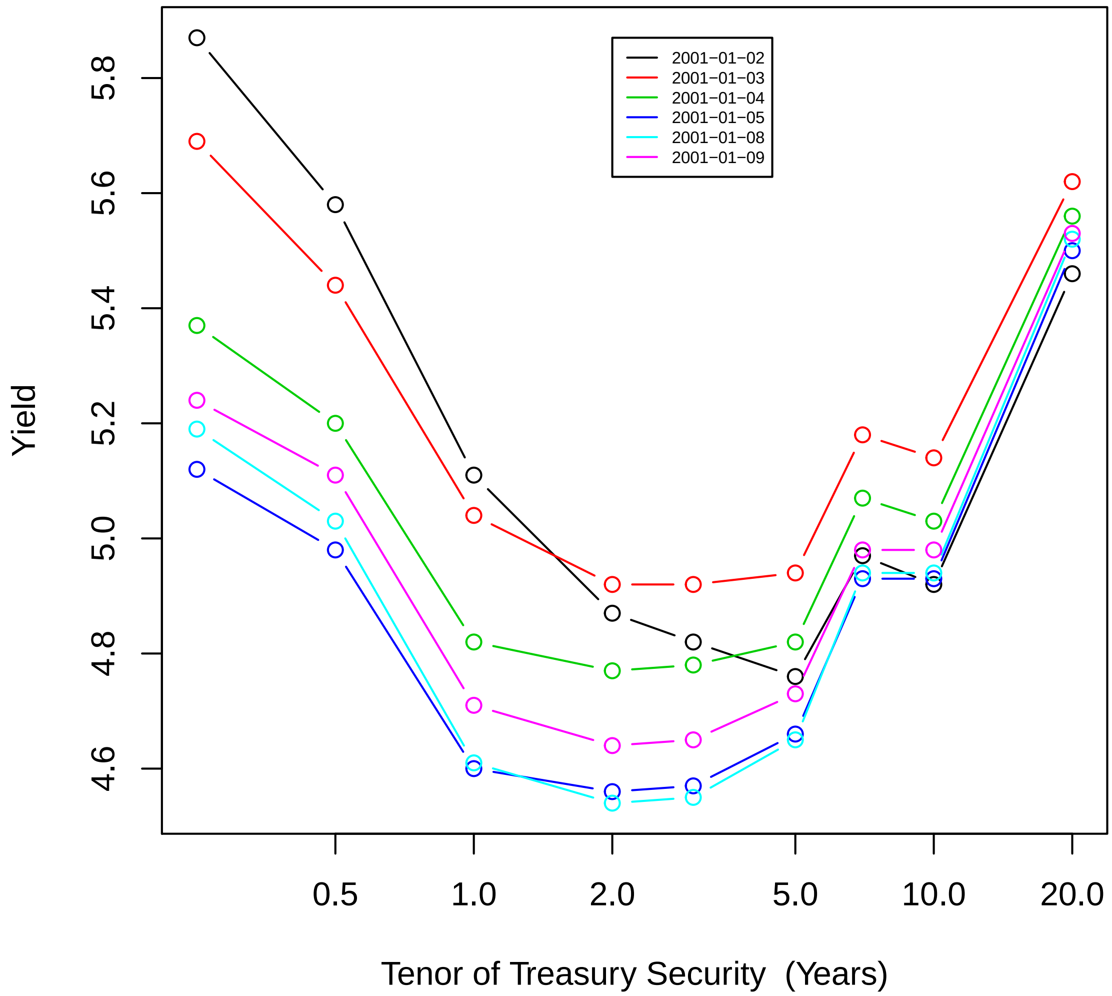
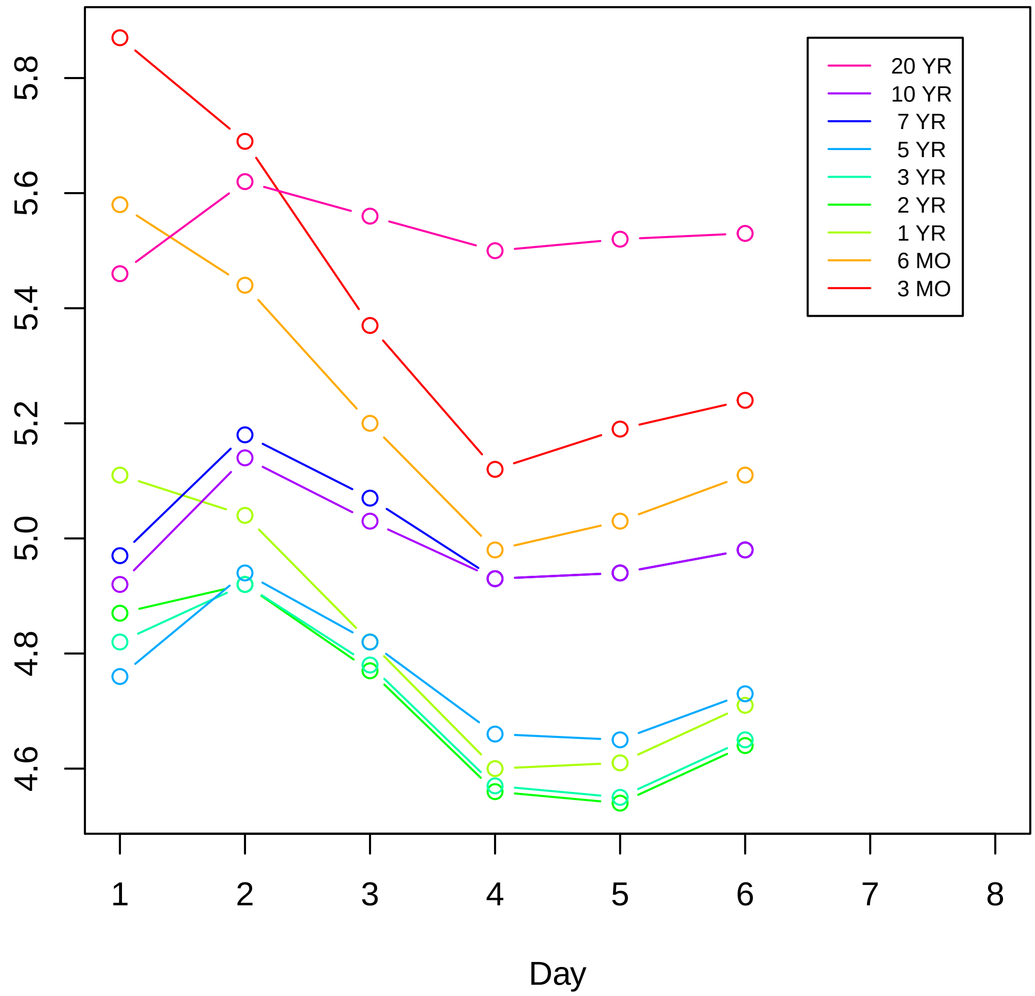


**Figure 2. Six Days of Yields for Nine Treasury Securities (https://research.stlouisfed.org)**



Principal Components Analysis provides a compelling model for the dynamics of the yield curve. First, daily changes in yield are computed, for each tenor. Second, the yield changes for each tenor are adjusted by subtracting their daily means. For the six yield curves in Figure 1, Table 1 gives the yield values for the nine tenors on each of the six days (columns 1-6), together with the five daily yield changes ($\times 100$) i.e., in Basis Points (BPS), adjusted to have zero mean. The $(9 \times 5)$ matrix of yield changes is the matrix $A$ for the Principal Components Analysis. The daily changes in the yields of the 9 tenors can be effectively modeled in terms of a small number of principal components variables corresponding to the largest singular values of $A$. As will be seen below, these principal components variables correspond to interpretable features of the yield curve (i.e., "level shifts" of the curve, changes in the "spread" between long-tenor and short-tenor yields, and changes in the "curvature" of yields).

**Table 1. U.S. Treasury Yields for 6 Days in 2001**

| | US Treasury Yields | | | | | | Matrix $A$ $(9 \times 5)$ for PCA Yield Changes ($BPS$) minus row-means | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Tenor | Jan 03 | Jan 04 | Jan 05 | Jan 06 | Jan 07 | Jan 10 | Jan 04 | Jan 05 | Jan 06 | Jan 07 | Jan 10 |
| 3 MO | 5.87 | 5.69 | 5.37 | 5.12 | 5.19 | 5.24 | -5.40 | -19.40 | -12.40 | 19.60 | 17.60 |
| 6 MO | 5.58 | 5.44 | 5.20 | 4.98 | 5.03 | 5.11 | -4.60 | -14.60 | -12.60 | 14.40 | 17.40 |
| 1 YR | 5.11 | 5.04 | 4.82 | 4.60 | 4.61 | 4.71 | 1.00 | -14.00 | -14.00 | 9.00 | 18.00 |
| 2 YR | 4.87 | 4.92 | 4.77 | 4.56 | 4.54 | 4.64 | 9.60 | -10.40 | -16.40 | 2.60 | 14.60 |
| 3 YR | 4.82 | 4.92 | 4.78 | 4.57 | 4.55 | 4.65 | 13.40 | -10.60 | -17.60 | 1.40 | 13.40 |
| 5 YR | 4.76 | 4.94 | 4.82 | 4.66 | 4.65 | 4.73 | 18.60 | -11.40 | -15.40 | -0.40 | 8.60 |
| 7 YR | 4.97 | 5.18 | 5.07 | 4.93 | 4.94 | 4.98 | 20.80 | -11.20 | -14.20 | 0.80 | 3.80 |
| 10 YR | 4.92 | 5.14 | 5.03 | 4.93 | 4.94 | 4.98 | 20.80 | -12.20 | -11.20 | -0.20 | 2.80 |
| 20 YR | 5.46 | 5.62 | 5.56 | 5.50 | 5.52 | 5.53 | 14.60 | -7.40 | -7.40 | 0.60 | -0.40 |

We reproduce here the steps used to compute the matrix $A$ from the original matrix $A00$ of yield curves: 9 tenors/rows for six days/columns.

$A00$ : $(9 \times 6)$ matrix of Yields of 9 US Treasury Securities on 6 Days in 2001.

$$A00[i,j] = \text{Yield on } j\text{-th day of Treasury Security with } i\text{-th tenor.}$$

| | Jan 02 | Jan 03 | Jan 04 | Jan 05 | Jan 08 | Jan 09 |
|---|---|---|---|---|---|---|
| 3 MO | 5.87 | 5.69 | 5.37 | 5.12 | 5.19 | 5.24 |
| 6 MO | 5.58 | 5.44 | 5.20 | 4.98 | 5.03 | 5.11 |
| 1 YR | 5.11 | 5.04 | 4.82 | 4.60 | 4.61 | 4.71 |
| 2 YR | 4.87 | 4.92 | 4.77 | 4.56 | 4.54 | 4.64 |
| 3 YR | 4.82 | 4.92 | 4.78 | 4.57 | 4.55 | 4.65 |
| 5 YR | 4.76 | 4.94 | 4.82 | 4.66 | 4.65 | 4.73 |
| 7 YR | 4.97 | 5.18 | 5.07 | 4.93 | 4.94 | 4.98 |
| 10 YR | 4.92 | 5.14 | 5.03 | 4.93 | 4.94 | 4.98 |
| 20 YR | 5.46 | 5.62 | 5.56 | 5.50 | 5.52 | 5.53 |

$A0$ : $(9 \times 5)$ matrix of Daily Yield Changes $(\times 100)$

| | Jan 03 | Jan 04 | Jan 05 | Jan 08 | Jan 09 |
|---|---|---|---|---|---|
| 3 MO | -18 | -32 | -25 | 7 | 5 |
| 6 MO | -14 | -24 | -22 | 5 | 8 |
| 1 YR | -7 | -22 | -22 | 1 | 10 |
| 2 YR | 5 | -15 | -21 | -2 | 10 |
| 3 YR | 10 | -14 | -21 | -2 | 10 |
| 5 YR | 18 | -12 | -16 | -1 | 8 |
| 7 YR | 21 | -11 | -14 | 1 | 4 |
| 10 YR | 22 | -11 | -10 | 1 | 4 |
| 20 YR | 16 | -6 | -6 | 2 | 1 |

$A$ : $(9 \times 5)$ adjusted $A0$ matrix with row-means subtracted

|  | Jan 03 | Jan 04 | Jan 05 | Jan 08 | Jan 09 |
|---|---|---|---|---|---|
| 3 MO | -5.4 | -19.4 | -12.4 | 19.6 | 17.6 |
| 6 MO | -4.6 | -14.6 | -12.6 | 14.4 | 17.4 |
| 1 YR | 1.0 | -14.0 | -14.0 | 9.0 | 18.0 |
| 2 YR | 9.6 | -10.4 | -16.4 | 2.6 | 14.6 |
| 3 YR | 13.4 | -10.6 | -17.6 | 1.4 | 13.4 |
| 5 YR | 18.6 | -11.4 | -15.4 | -0.4 | 8.6 |
| 7 YR | 20.8 | -11.2 | -14.2 | 0.8 | 3.8 |
| 10 YR | 20.8 | -12.2 | -11.2 | -0.2 | 2.8 |
| 20 YR | 14.6 | -7.4 | -7.4 | 0.6 | -0.4 |

The singular-value decomposition (SVD) of $A$ has 3 components:

```
$d
[1] 7.277085e+01 3.985864e+01 1.169534e+01 2.389852e+00 2.121659e-15

$u
          [,1]       [,2]        [,3]        [,4]        [,5]
 [1,] 0.3830183  0.5291979 -0.47897581  0.06049368  0.02049152
 [2,] 0.3364652  0.4368233 -0.04602992  0.21038865  0.12024911
 [3,] 0.3581801  0.2634263  0.22596793 -0.49184577 -0.38877792
 [4,] 0.3527232 -0.0286657  0.46042904  0.09663509  0.53210504
 [5,] 0.3713278 -0.1311673  0.43012237  0.25840071  0.04259306
 [6,] 0.3498874 -0.2934614  0.11737198 -0.18857534 -0.52684967
 [7,] 0.3234104 -0.3657703 -0.22870020  0.45955691 -0.17195535
 [8,] 0.2975810 -0.3780225 -0.35157337 -0.57926041  0.48357804
 [9,] 0.1841935 -0.2809029 -0.36154940  0.22750774 -0.08963872

$v
           [,1]       [,2]       [,3]        [,4]      [,5]
[1,]  0.3740211 -0.7944806 -0.1673538 -0.03002457 0.4472136
[2,] -0.5162273 -0.1131721  0.4823220  0.53671867 0.4472136
[3,] -0.5634134  0.0765316 -0.3174224 -0.61314865 0.4472136
[4,]  0.2361191  0.4643168 -0.5638385  0.45906841 0.4472136
[5,]  0.4695005  0.3668043  0.5662926 -0.35261385 0.4472136
```

The total variance in the data is the sum of the sample variances for each of the 9 tenor yield changes. The following table shows that the Total Variance is 1756.70. The table includes the standard deviations ($S_j$) which are in the same units as the entries in the matrix $A$ (BPS = (1/100)th of 1 percent). For these data, the shortest tenor (3 MO) had the highest standard deviation of 17.70 BPS while the longest tenor (20 Yr) had the smallest standard deviation of 8.99 BPS.

| Tenor | $S_j^2$ | $S_j$ |
|---|---|---|
| 3 MO | 313.30 | 17.70 |
| 6 MO | 225.80 | 15.03 |
| 1 YR | 199.50 | 14.12 |
| 2 YR | 172.30 | 13.13 |
| 3 YR | 195.80 | 13.99 |
| 5 YR | 196.80 | 14.03 |
| 7 YR | 193.70 | 13.92 |
| 10 YR | 178.70 | 13.37 |
| 20 YR | 80.80 | 8.99 |
| Total Variance | 1756.70 | – |

The Total Variance is also the sum of eigen values of $S = \frac{AA^T}{n-1}$. These eigen values ($\lambda_k$) are related to the singular values of $A$ ($d_i$) by $\lambda_k = d_k^2/(n-1)$, where $n$ is the number of columns (days) in $A$. These eigen values are the variances of the PC Variables.

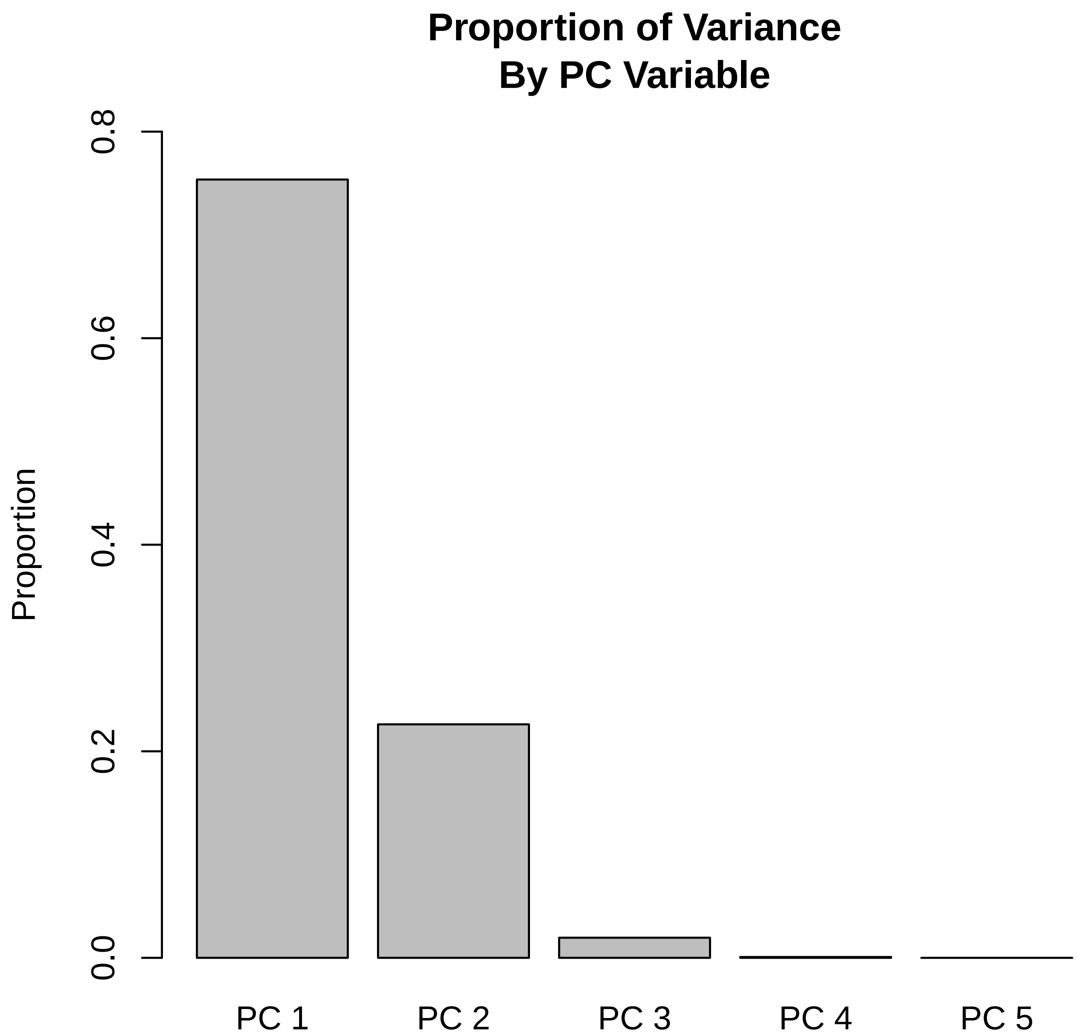
| PC Variable | $\lambda_k = Variance(PC\ k)$ | St. Dev |
|---|---|---|
| PC 1 | 1323.90 | 36.39 |
| PC 2 | 397.18 | 19.93 |
| PC 3 | 34.20 | 5.85 |
| PC 4 | 1.43 | 1.19 |
| PC 5 | 0.00 | 0.00 |
| Total Variance | 1756.70 | – |

The Principal Components Variables are an orthogonal transformation of the original coordinate axes (rows) of $A$ that corresponds to a simple 'rotation' of the data about the origin (data mean). Principal Component Variable PC 1 is the axis of the data with the greatest variability (highest standard deviation), and Principal Component Variable PC 2 is the axis of the data orthogonal to the first with next greatest variability. Note that the standard deviation of PC 1 (36.39 $BPS$) is larger than all the standard deviations of the original axes (tenors/rows of $A$).

The importance of principal components variables can be measured by the proportion of total variability of each component $k$ and the cumulative proportion of the first $k$ components Principal Component Variable PC 1 represents 75.4% of the total variability while the first 3 Principal Component Variables represent 99.9% of the total variability.

| Component | Stand. Dev. | Proportion | Cumulative Proportion |
|---|---|---|---|
| PC 1 | 36.3854 | 0.7536 | 0.7536 |
| PC 2 | 19.9293 | 0.2261 | 0.9797 |
| PC 3 | 5.8477 | 0.0195 | 0.9992 |
| PC 4 | 1.1949 | 0.0008 | 1.0000 |
| PC 5 | 0.0000 | 0.0000 | 1.0000 |

Note that in this example, there are only 5 data points in the 9-dimensional space of changes in the yield curve, so there can be only 5 Principal Component Variables. Also, the $(9 \times 5)$ matrix $A$ was defined by subtracting the row-means which eliminates one dimension of variability; this is why PC 5 has zero variability. Real-world application of PCA to modeling the dynamics of interest rates would use many data points (e.g., $\sim 252$ for business days in one year). (We note the curious fact that the structure and interpretation of principal components variables using larger data sets conforms closely to those illustrated here with just 5 data points!)


Proportion of Variance
By PC Variable
Proportion
0.0
0.2
0.4
0.6
0.8
PC 1
PC 2
PC 3
PC 4
PC 5

Loadings of Principal Components Variables:

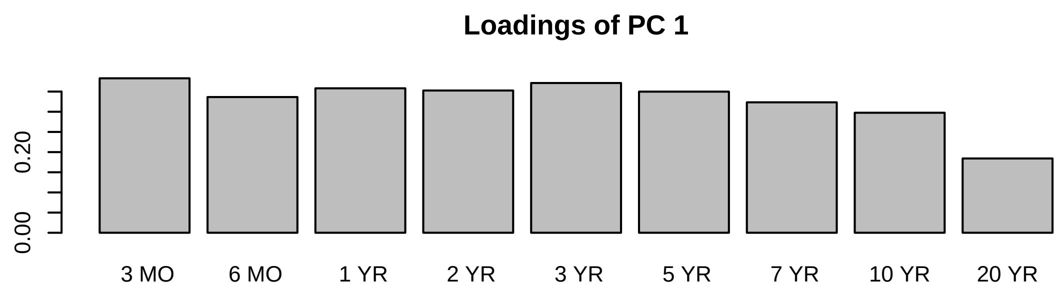
The $j$-th column of matrix $u$ in the Singular-Value Decomposition (SVD) of $A$ contains the "loadings" of the $j$-th Principal Component Variable. Each PC variable is a weighted sum of yield changes in different rows (tenors) of $A$ and the loadings are the weights applied for each PC variable.

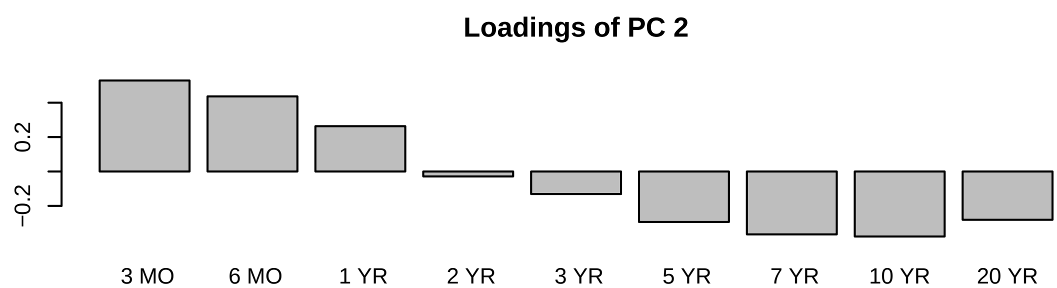
The next table presents the loadings for the first 3 PC variables and the stacked plots display and compare their loadings on different tenors. Note that

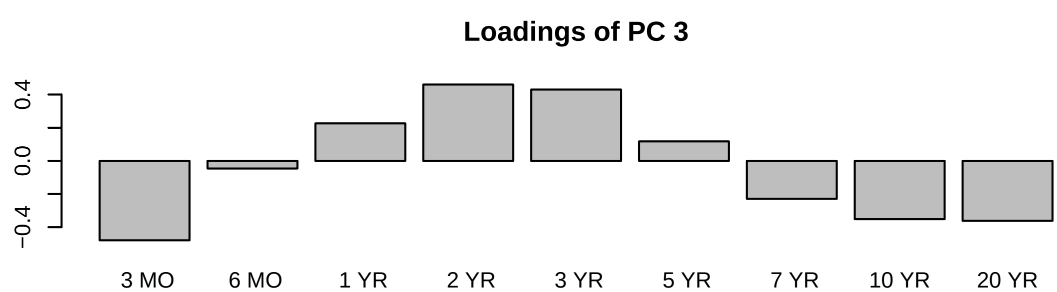
- PC 1 measures level shifts in the yield curve – its daily values are a weighted sum of all 9 yield changes.
- PC 2 measures the spread in the yield curve between short and long maturities – its daily values are a difference between a weighted sum of the shorter-tenor yield changes and a weighted sum of the longer-tenor yield changes.
- PC 3 measures the curvature in the yield curve – it corresponds to a second difference of yield changes: the difference between two first differences (mid-tenor minus short-tenor, and long-tenor minus mid-tenor).

| | PC 1 | PC 2 | PC 3 |
|---|---|---|---|
| 3 MO | 0.38 | 0.53 | -0.48 |
| 6 MO | 0.34 | 0.44 | -0.05 |
| 1 YR | 0.36 | 0.26 | 0.23 |
| 2 YR | 0.35 | -0.03 | 0.46 |
| 3 YR | 0.37 | -0.13 | 0.43 |
| 5 YR | 0.35 | -0.29 | 0.12 |
| 7 YR | 0.32 | -0.37 | -0.23 |
| 10 YR | 0.30 | -0.38 | -0.35 |
| 20 YR | 0.18 | -0.28 | -0.36 |

Graphical display of loadings for the top 3 PC variables.


Loadings of PC 1
0.00
0.20
3 MO
6 MO
1 YR
2 YR
3 YR
5 YR
7 YR
10 YR
20 YR


Loadings of PC 2
−0.2
0.2
3 MO
6 MO
1 YR
2 YR
3 YR
5 YR
7 YR
10 YR
20 YR


Loadings of PC 3
−0.4
0.0
0.4
3 MO
6 MO
1 YR
2 YR
3 YR
5 YR
7 YR
10 YR
20 YR

Principal Components Analysis allows us to approximate daily changes in the Treasury yield curve with changes in the first and second Principal Components Variables. The first plot shows how the level of the curve (PC1) shifted up (day 1) then down (days 2 and 3) and finally up (days 4 and 5). The second plot shows how the spread of the curve (shorter-tenor yields minus longer-tenor yields) decreased (days 1 and 2) before increasing (days 3-5).