STA437Project

Thanh Huy Dang, Hojung Kim

2025-03-25

Introduction

Credit default swap (CDS) spreads are valid and readily available indicators of creditworthiness of individual companies and the economy as a whole. Spreads at different tenors are quoted daily.

Multivariate methods are attractive options to analyse this data set because we expect they can extract the historical behaviour of the CDS spreads at different tenors (considering them as variables), as well as the correlation structure among the CDS spreads for different pairs of companies.

In this report, we address two main questions.

First, we investigate how market volatility affects the aggregate CDS curve of all companies (and the curve of Exxon Mobil Corp in particular) and whether this effect is different for the periods before and during COVID.

Second, we expect stronger correlation in daily CDS spreads changes between companies within the same sector. In particular, we investigate whether there is a difference in how changes in CDS spreads of Exxon Mobil Corp relate to that of Chevron Corp (energy sector) and that of American Express (finance sector).

The multivariate methods employed in this report are Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA).

PCA is applied to the daily changes in CDS spreads to investigate the behaviour of the CDS curve. In particular, we would like to observe how the spreads at different tenors move relative to each other when there is a change in the direction of each significant principal component.

CCA is applied to the daily changes in CDS spreads for the two pairs of companies (Exxon Mobil Corp vs Chevron Corp, and Exxon Mobil Corp vs American Express) to investigate the degree in which they are correlated.

Data and Preprocessing

Data Description

This dataset contains CDS spreads for various companies across 10 different tenors, labeled PX1 to PX10. Each row represents the CDS spreads for a specific company on a specific date.

There is a summary table of the dataset, which includes key statistics such as mean and range for each tenor. Relevant histograms are also included to illustrate the distribution of CDS spreads as well as their differences. For histogram of CDS spreads frequency within each tenor, I took normalization and excluded extreme CDS spreads due to the visibility.

Data Cleaning

There are no missing values in the dataset, so no imputation or row removal was necessary.

Transformations

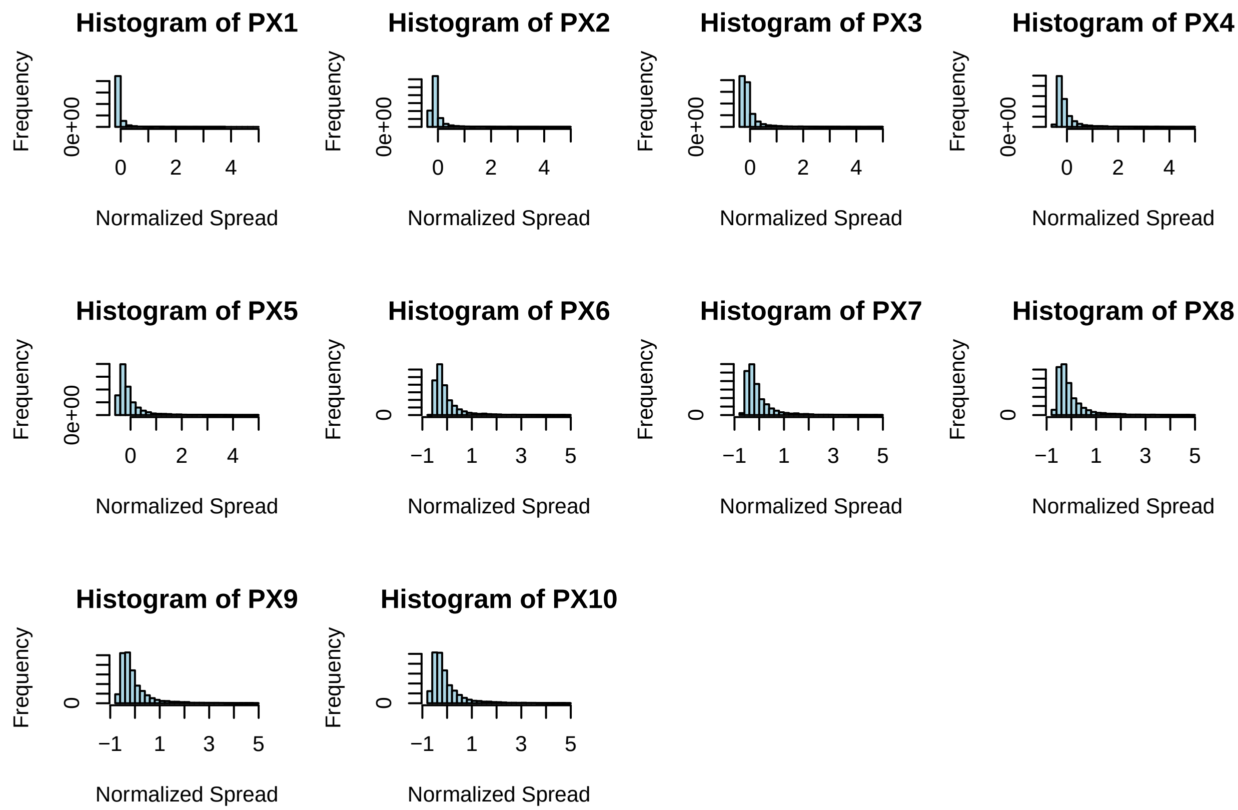
Several transformations were applied during preprocessing.

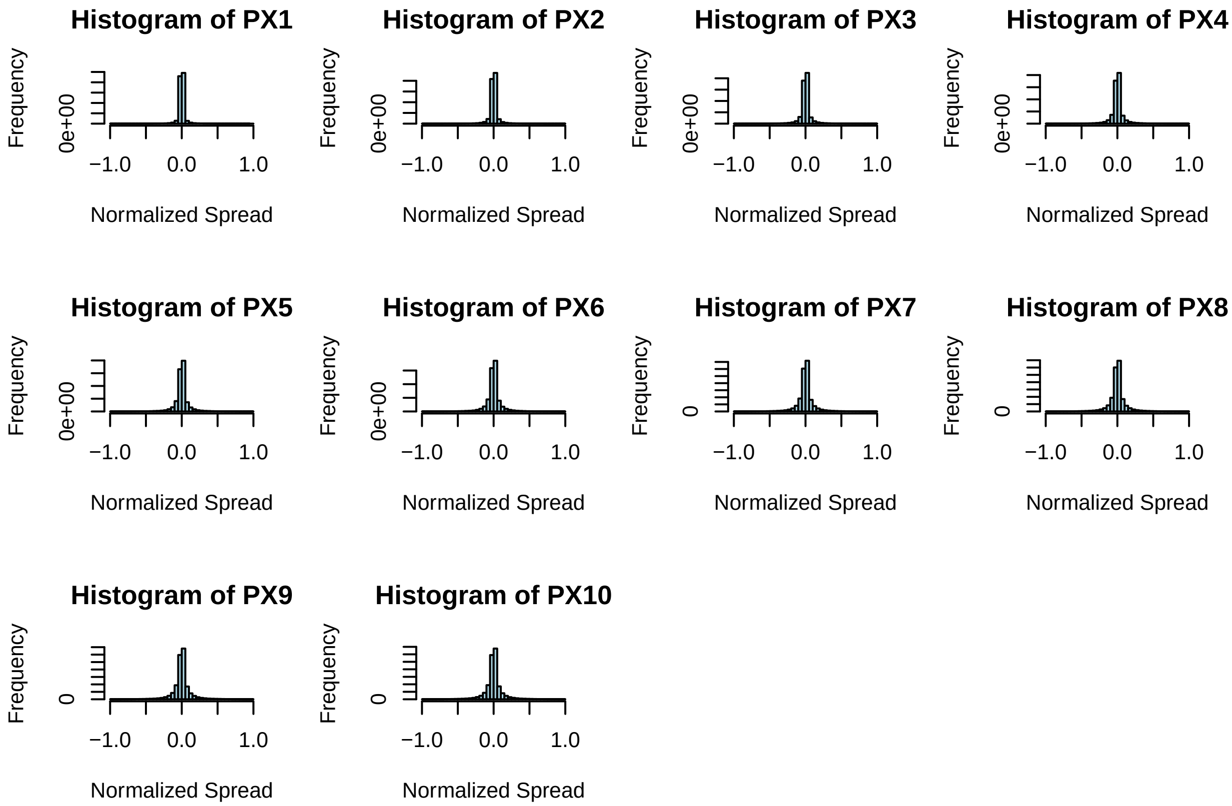
First, we computed the first difference of each tenor series within each company. This transformation captures daily changes in credit risk, allowing us to focus on the dynamics of risk rather than absolute CDS levels.

Second, we applied normalization to the differenced data before performing PCA. This step ensures that each tenor contributes equally to the analysis regardless of scale.

```
##       PX1                 PX2                 PX3                 PX4
##  Min.   :    1.15   Min.   :    2.07   Min.   :    3.35   Min.   :    4.802
##  1st Qu.:    8.95   1st Qu.:   15.18   1st Qu.:   23.08   1st Qu.:   32.624
##  Median :   14.87   Median :   24.68   Median :   36.75   Median :   50.920
##  Mean   :   41.20   Mean   :   53.68   Mean   :   68.91   Mean   :   85.408
##  3rd Qu.:   28.91   3rd Qu.:   45.19   3rd Qu.:   64.44   3rd Qu.:   86.570
##  Max.   :48115.24   Max.   :34522.49   Max.   :32246.85   Max.   :30048.361
##       PX5                  PX6                  PX7                 PX8
##  Min.   :    5.796   Min.   :    9.479   Min.   :   10.95   Min.   :   12.68
##  1st Qu.:   42.130   1st Qu.:   51.620   1st Qu.:   58.40   1st Qu.:   63.65
##  Median :   64.990   Median :   77.756   Median :   86.89   Median :   93.50
##  Mean   :  101.520   Mean   :  114.916   Mean   :  124.58   Mean   :  131.12
##  3rd Qu.:  107.975   3rd Qu.:  124.630   3rd Qu.:  137.29   3rd Qu.:  145.47
##  Max.   :28017.639   Max.   :26671.760   Max.   :25561.03   Max.   :24590.86
```

```
##       PX9                 PX10
##  Min.   :   14.26   Min.   :   14.91
##  1st Qu.:   67.70   1st Qu.:   71.02
##  Median :   98.69   Median :  102.89
##  Mean   :  136.22   Mean   :  140.27
##  3rd Qu.:  151.84   3rd Qu.:  157.06
##  Max.   :23763.96   Max.   :23052.64
```

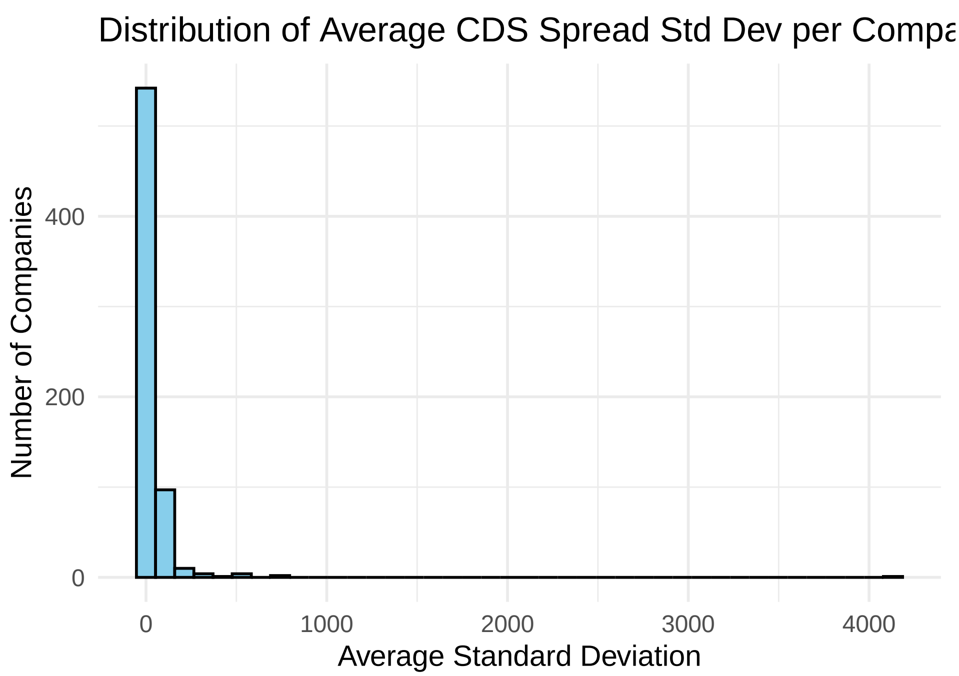



Histogram of PX1
Histogram of PX2
Histogram of PX3
Histogram of PX4
Histogram of PX5
Histogram of PX6
Histogram of PX7
Histogram of PX8
Histogram of PX9
Histogram of PX10
Frequency
0e+00
0
−1.0
0.0
1.0
Normalized Spread

Methodology

PCA Across All Companies

To ensure that our PCA reflects the collective behavior of CDS spreads and is not overly influenced by any single firm, we examined the distribution of average standard deviation across companies. The histogram shows a tightly concentrated distribution, with most companies exhibiting low to moderate spread variability. While a few companies have notably higher variability, the bulk of the distribution suggests that no company overwhelmingly dominates the CDS data. This supports the validity of applying PCA across the entire dataset.



We perform a first difference on the CDS spreads of all companies to find the daily changes at all ten tenors. The daily changes are normalized to account for differences in scales. We then conduct PCA on this data to find the principal components (i.e. the eigenvectors of the sample variance matrix). A scree plot is created to inspect how much of the variance is explained by the principal components, and how many principal components we should investigate more closely.

We add (and subtract) the scaled (by a factor of 20) principal components to the mean CDS curve to study the effect on the CDS curve from a change in the direction of each principal component. The scaling is to magnify the effects for better visualization in our CDS curve plots.

Finally, the above procedure is repeated separately for the data from before the COVID pandemic (before March 2020) and the data during the COVID pandemic to observe if the behaviour of the CDS curve is different in these two periods.

PCA on Exxon Mobil Corp

In the next part of our analysis, we place our focus on Exxon Mobil Corp (Exxon). The steps above are repeated for this company to observe if its CDS curve behaves differently from the results we obtain from the whole group of companies.

We also investigate the case when there is a change in all the significant principal directions (proportioned by the percentage of variance explained).

Canonical Correlation Analysis (CCA)

Next, we obtain the daily changes for Chevron and American Express and pairing them with the Exxon data to conduct CCA.

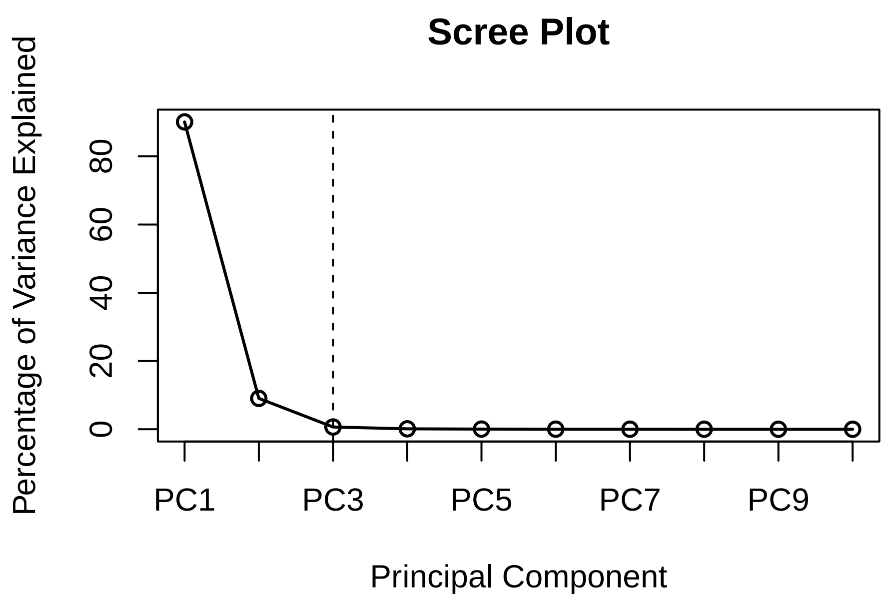
We plot two curves of canonical correlations for the two pairs respectively to investigate whether there are differences in how the spreads in each pair move together. Specifically, we want to see if the first few canonical correlations appear very different for the two pairs. Larger canonical correlations indicate that changes in CDS spreads of that particular pair of companies are more closely related.

Results

PCA on All Companies

With regards to daily changes of CDS spreads of all companies, 90% of the total variance is explained by PC1, and 9% by PC2. The remaining principal components contribute minimally to the variance. This is apparent from the Scree plot.

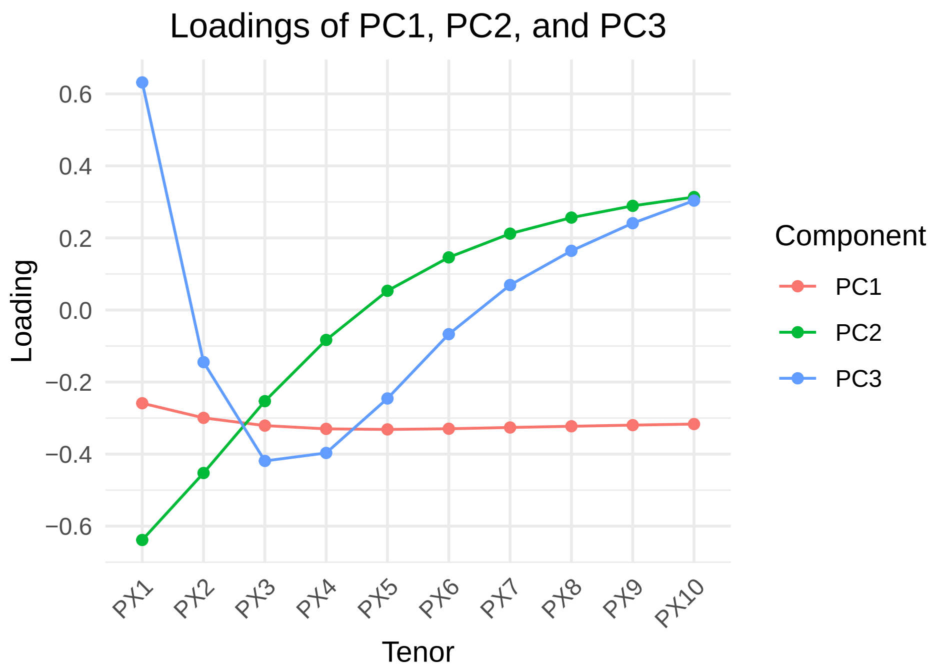
```
## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     3.0014 0.95141 0.25641 0.10813 0.06946 0.04893 0.04042
## Proportion of Variance 0.9008 0.09052 0.00657 0.00117 0.00048 0.00024 0.00016
## Cumulative Proportion  0.9008 0.99135 0.99793 0.99909 0.99958 0.99982 0.99998
##                            PC8     PC9     PC10
## Standard deviation     0.01250 0.00593 0.003711
## Proportion of Variance 0.00002 0.00000 0.000000
## Cumulative Proportion  1.00000 1.00000 1.000000
```



The figure below shows the first 3 PCs across all tenors. For PC1, we can see that all tenors share similar small negative values. It means that if there is a change in the direction of PC1, CDS spreads will be increased or decreased by a similar amount across all maturities. Since PC1 explains 90% of the total variance, we can safely assume that the dominant movement in CDS spreads is a parallel shift.

Observe the signs of the entries in the second principal component — they change from negative to positive at PX5. It means that short-term and long-term tenor CDS spreads move in opposite directions.

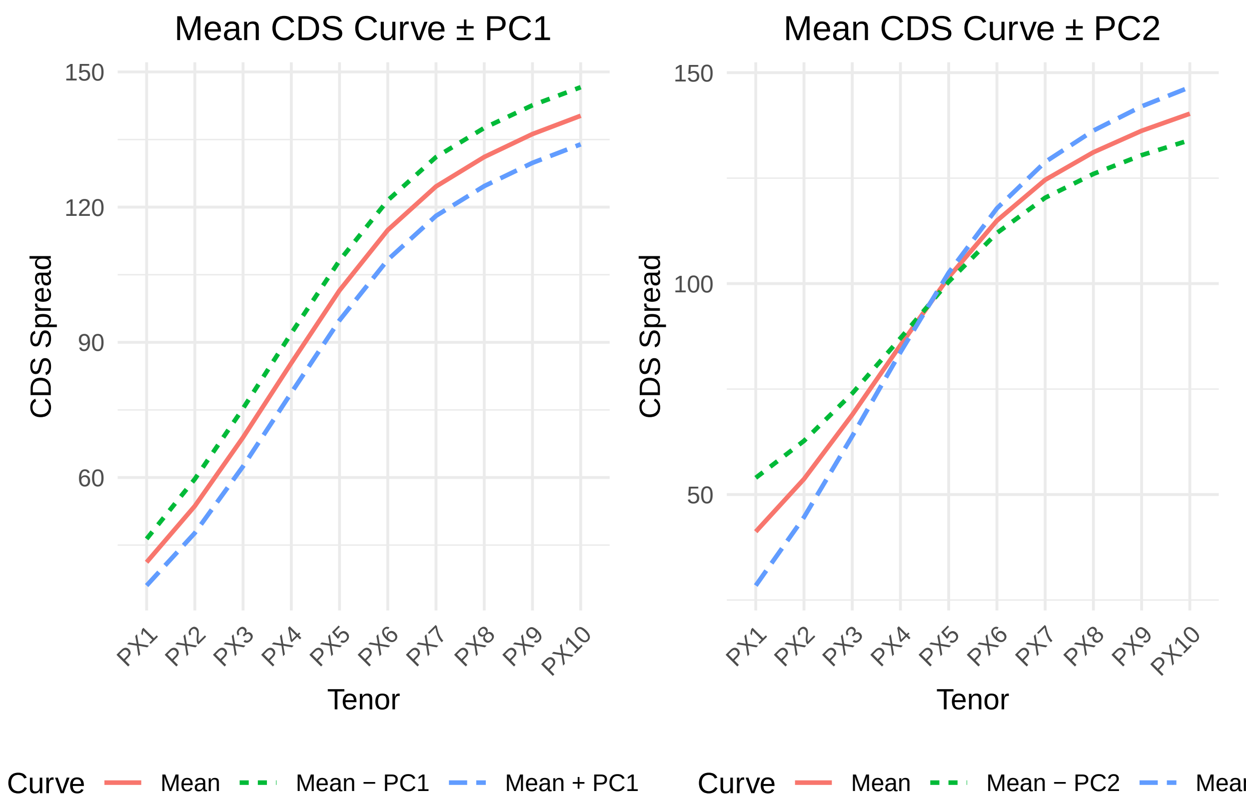
This might explain why during a financial crisis, short-term risk increases and so does the cost of protection for immediate risk of default. On the contrary, stable market conditions should imply higher cost of protection for potential default further in the future.



By adding and subtracting PC1 and PC2 from the mean CDS curve, we can visualize how each component affects the shape of the CDS term structure. This reinforces our interpretation of the loadings for each tenor.

As shown in the graph, adding or subtracting PC1 causes a uniform upward or downward shift across all tenors. This reflects a parallel shift, which aligns with our earlier interpretation of PC1 as capturing overall market-wide movements in CDS spreads.

When we add or subtract PC2, we observe that an increase in PC2 raises long-term spreads while lowering short-term spreads. Conversely, a decrease in PC2 increases short-term spreads and lowers long-term spreads. This demonstrates that PC2 drives the steepening or flattening of the CDS curve, consistent with the interpretation of PC2 representing changes in the slope of the term structure.



Comparison: Before vs During COVID

Now we attempt to compare and contrast the CDS spreads behaviour in the periods before and during the COVID pandemic.

We observe that during the second period, the first principal component contributes more to the total variance. It means the CDS spreads at all tenors shift together more strongly during COVID. The second principal component consequently contributes less, signifying that relative movements of short- and long-term risks are secondary to overall general risk.

In the graph comparing pre-COVID and during-COVID time's eigenvectors of PC1, PC2, and PC3, we can also check that PC1 has higher loadings during COVID, which means CDS scores increased in a broad fashion.

PC2's slope change is also notable:

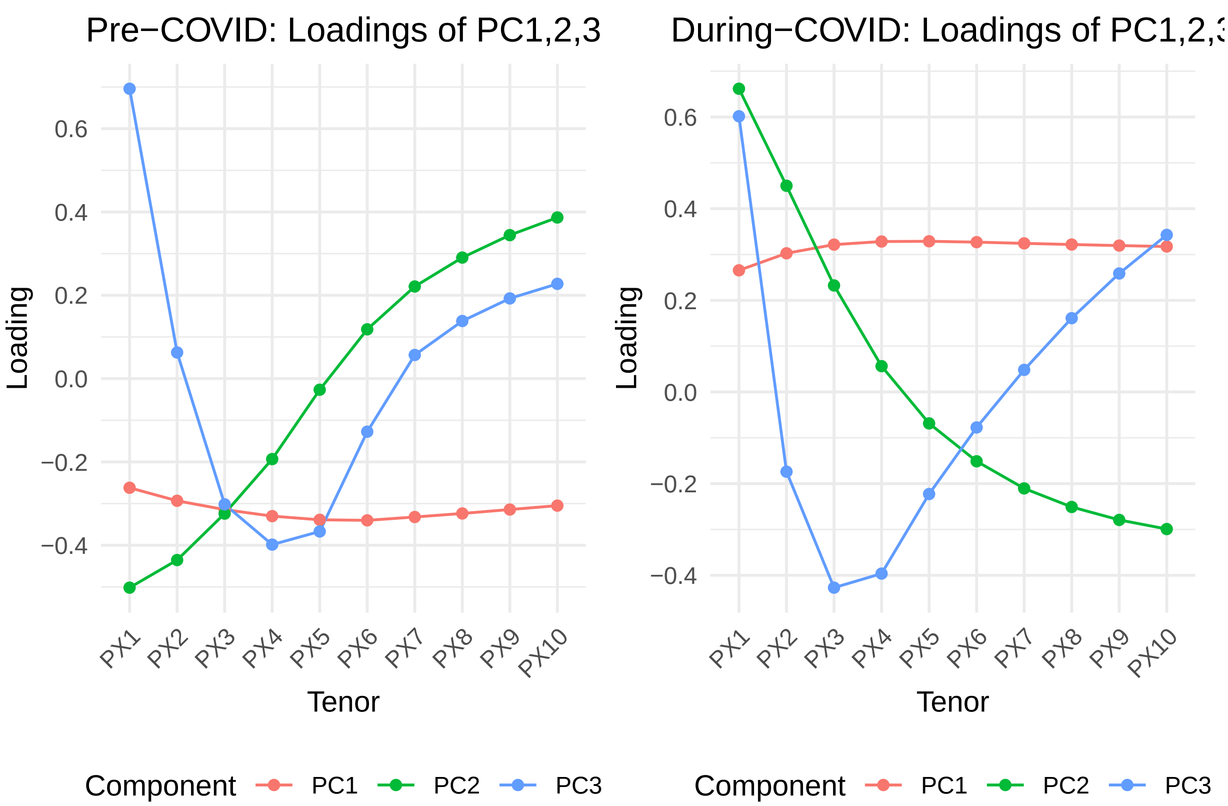
- During COVID, investors thought short term was riskier and expected long-term stability.
- This contrasts with the pre-COVID period, where investors were confident in short-term risk but more uncertain about the long term.

As shown in the graph, adding or subtracting PC1 causes a uniform upward or downward shift across all tenors. This reflects a parallel shift, which aligns with our earlier interpretation of PC1.

When we add or subtract PC2, we observe:

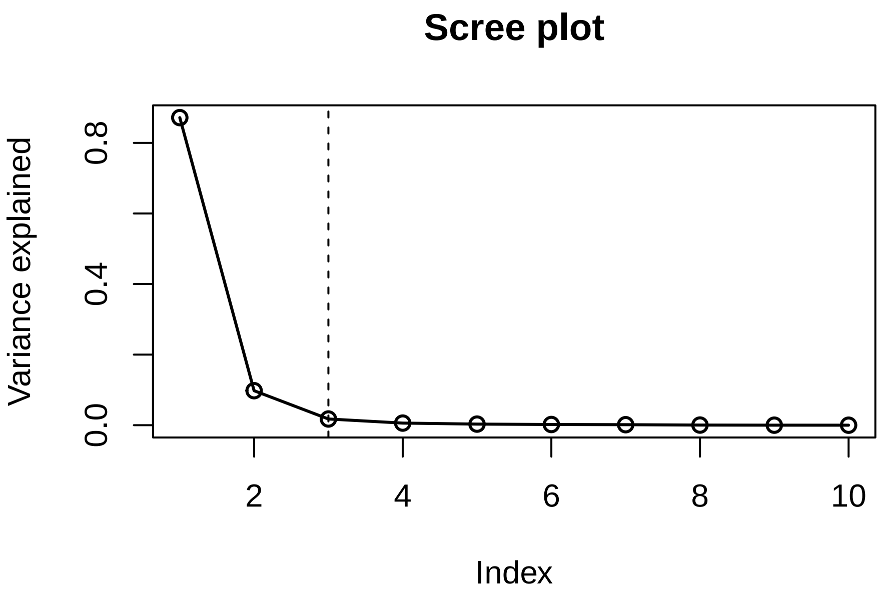
- An increase in PC2 raises long-term spreads while lowering short-term spreads.
- A decrease in PC2 increases short-term spreads and lowers long-term spreads.

This demonstrates that PC2 drives the steepening or flattening of the CDS curve, consistent with it representing changes in the slope of the term structure.

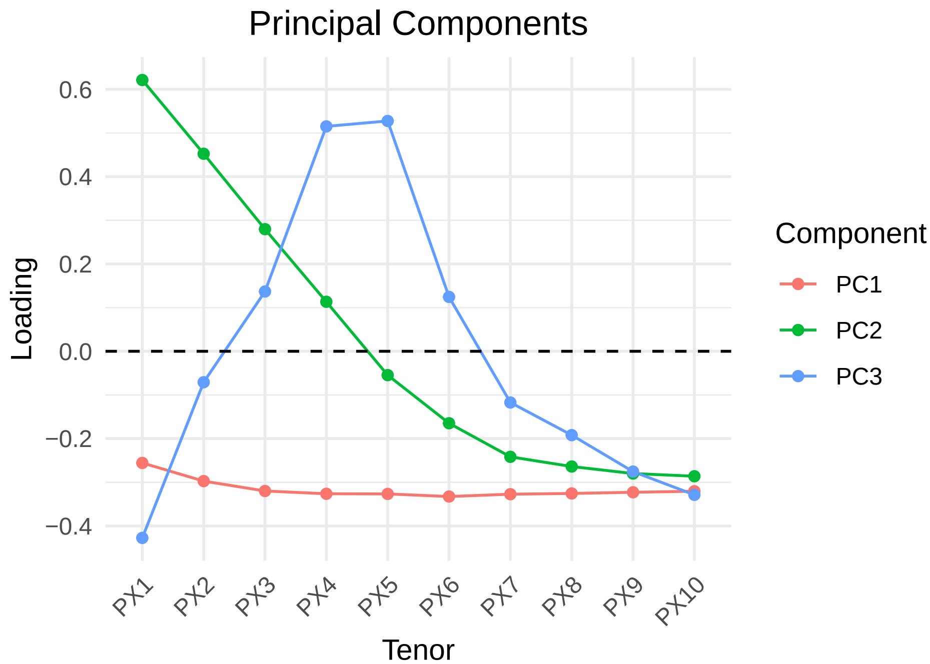


PCA on Exxon Mobil Corp

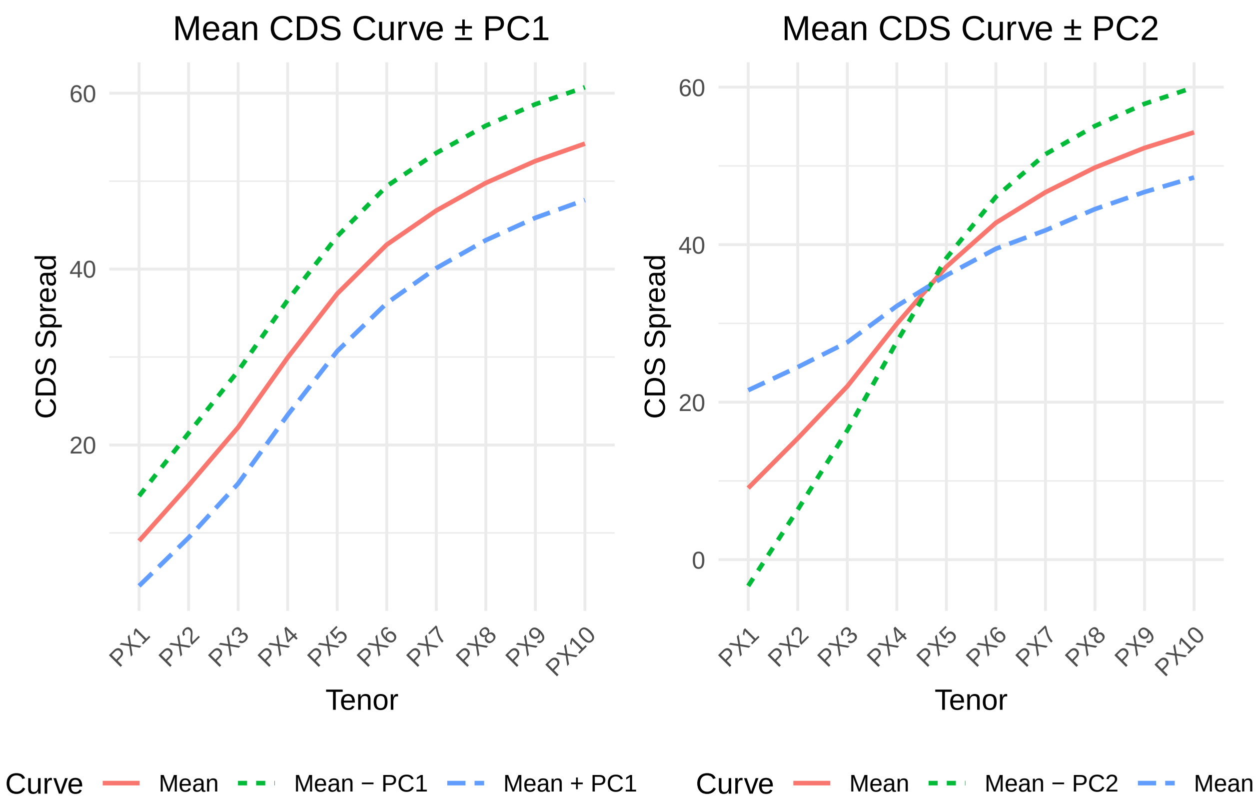
With regards to the case of Exxon, the Scree plot indicates that most of the variance in the daily changes of CDS rates is explained by the first three principal components, so we decide to focus on them.



- The first principal component has negative values for all entries, ranging between -0.25 and -0.35 approximately. Hence, a change in this direction may increase or decrease the CDS rates at all tenors by roughly the same amount. Figure shows an approximately parallel shift in the CDS curve.
- The entries of the second principal component are decreasing. A change in this direction may increase or decrease the slope of the CDS curve. This is because the rates are adjusted to a larger extent (and in opposite directions) at the shorter and longer tenors, but are only adjusted minimally at the medium tenors. Figure illustrates this behaviour.
- The entries of the third principal component are increasing and then decreasing. A change in this direction may increase or decrease the convexity of the CDS curve. The resulting curve after the change will intersect the original curve at two points, as illustrated in Figure.



When effecting changes in all three principal directions at the same time (in proportion with the respective percentages of variance explained), we see that the CDS curve shifts in a parallel manner. This is expected because the Scree plot indicates close to 90% of the variance is explained by the first principal component. We note, however, that the CDS rates at shorter tenors appear to be shifted by a smaller extent compared to the longer tenors.



Exxon: Before vs During COVID

We separate the data for Exxon into the “Before COVID” and “During COVID” portions.

The results obtained are largely similar to the case where we investigate it as a whole. The first three principal components explain most of the variance, with the first one dominating the others (close to 90% in both datasets). The behaviour of the CDS curve in response to changes in the principal directions are as described above.

However, the change in the convexity of the CDS curve is not as obvious with the “During COVID” data. These results are summarized in Figure and Figure.

Canonical Correlation Analysis

The first three canonical correlations for the CDS rates of Exxon and Chevron are 0.34, 0.19, 0.17

These numbers for Exxon and American Express are 0.29. 0.15, 0.10 which are smaller in magnitude.

This is not very surprising since Exxon and Chevron are from the energy sector while

American Express is from the finance sector.

However, as seen in the figure, the differences are not very substantial. So it is likely that macroeconomic factors at large may have very strong influence on the CDS rates of individual companies (especially with regards to these three mature successful companies), relative to sector-specific factors.