PCA on Credit Default Swaps Spreads Across Maturities

Cross-Sectional Risk Patterns from One-Day Data (2020-01-02)

Names:

Yunzhao Li

Yujie Sun

Instructor: Piotr Zwiernik

STA437H1 S2025

Methods for multivariate data

Contents

1 Introduction

1.1 Motivation

The Credit Default Swap (CDS) dataset provides a unique view into how financial markets perceive the credit risk of hundreds of companies. The CDS spreads for over 600 companies across 10 different maturities reflect how markets assess this risk and naturally form a multivariate structure. Principal Component Analysis (PCA) is well-suited for identifying hidden patterns in how these spreads move together across maturities in high-dimensional data.

1.2 Research Question

Can PCA help identify and interpret the latent factors that summarize the term structure of CDS spreads across companies at a given point in time (One-Day Data)?

1.3 Brief Summary of Methods

We apply Principal Component Analysis (PCA) to the CDS spread term structures (PX1 to PX10) of over 600 companies on a single date. PCA reduces the features dimensionality while preserving the main variation of the data. We interpret the key patterns in the CDS curves through the loadings and scores of the first two principal components.

1.4 Background

We referenced several popular applications of PCA on curve data, such as in Litterman and Scheinkman (1991), who showed that bond yields could be summarized by a level, slope, and curvature component. We further examined whether a similar structure applies to CDS term spreads. We found PCA has also been applied to credit markets, where CDS spreads serve as market-based measures of default risk (Blancoet al., 2005). The structured nature of CDS spreads across maturities makes multivariate approaches like PCA an appropriate choice to uncover common components in credit risk profiles (Alexander & Kaeck, 2008).

2 Data and Preprocessing

2.1 Description of the Dataset

The dataset includes CDS spreads for over 600 companies across 10 different maturities. To show the general pattern while reducing the complexity, we filtered the data on 2020-01-02 to conduct our study, with separate maturities data from 1 to 10 years. We visualize the term structure with the summarized average and variability of CDS spread.

2.2 Summary table and diagram

Table 1: Mean and SD of CDS Spreads (PX1–PX10)

PX	mean	sd
PX1	22.84002	65.88765
PX2	32.84101	74.52450
PX3	45.18720	81.91038
PX4	59.16060	85.89405
PX5	73.33506	90.36014
PX6	86.80179	93.60937
PX7	96.40466	96.26178
PX8	103.20683	97.47536
PX9	108.51127	98.43832
PX10	112.74878	99.25095

Table 1 shows the mean and standard deviation of CDS spreads for each maturity (PX1–PX10). Spreads increase steadily from nearly 23 bps at PX1 to 113 bps at PX10. The variability also shows the same trend. This suggests the increasing trend of the structured term curve's risk over time.

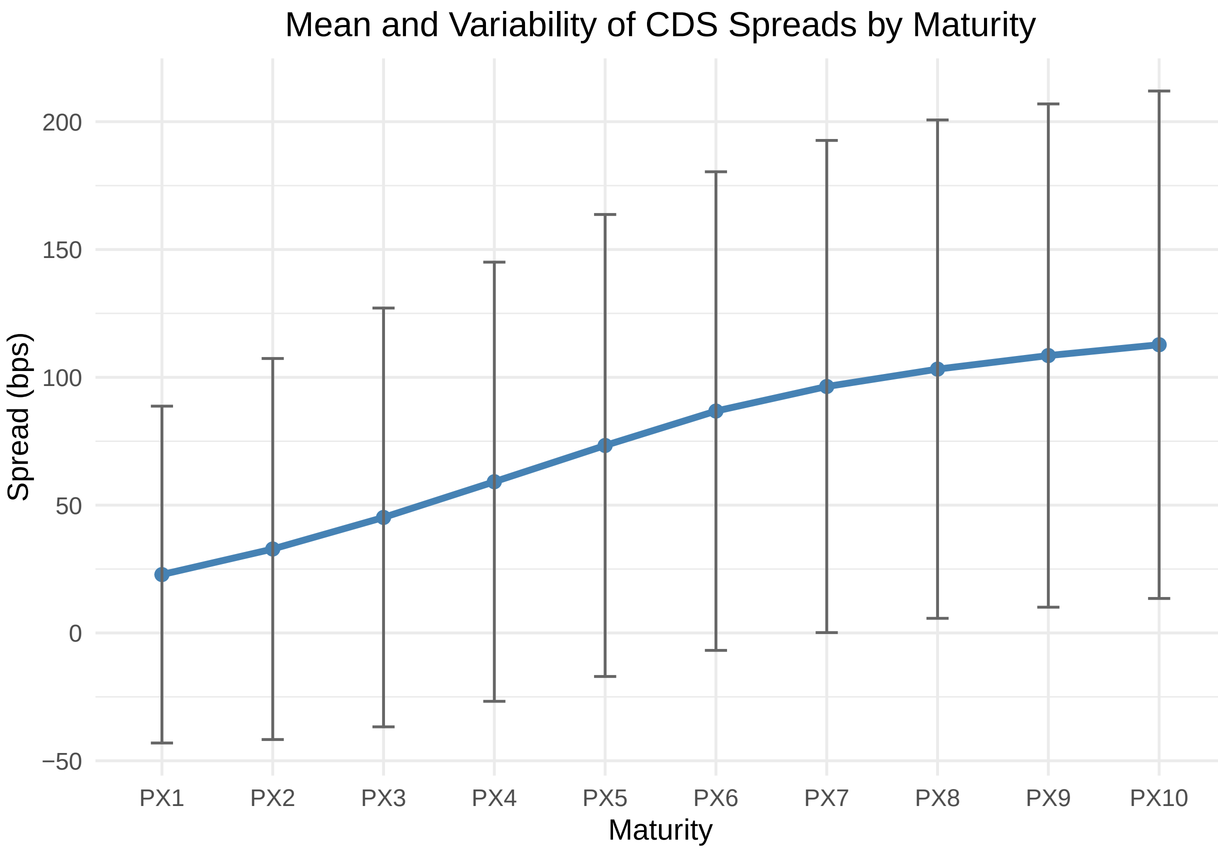


Figure 1 also shows an increasing trend in average spreads and variability (±1 SD) across maturities, consistent with the tendency for credit risk to grow over time.

2.3 Handling of Missing Data

Any rows with missing values in PX1 to PX10 were removed to ensure PCA works well.

2.4 Transformations

We subtracted the mean from each maturity column to center the data, as the preparation for PCA analysis.

3 Methodology

To analyze the structure of CDS spreads across maturities, we use Principal Component Analysis (PCA), a dimensionality reduction technique designed to identify principal components, the uncorrelated directions that explain the maximum variance in a dataset. PCA simplifies the CDS spreads across 10 maturities by reducing to a few dimensions with key features.

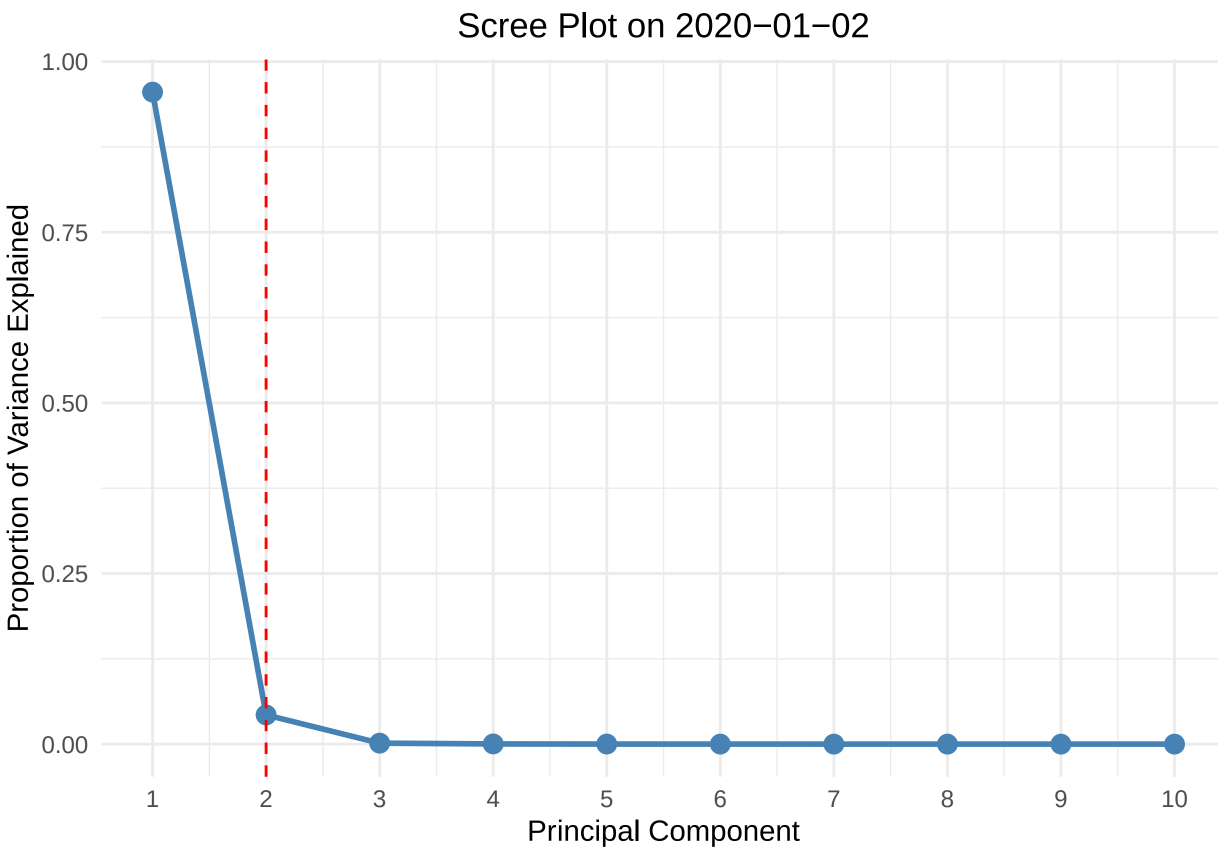
We treat each company's CDS data as a 10-dimensional vector, each corresponding to a maturity from PX1 to PX10. Then, we applied the prcomp() function in R to conduct the PCA on the mean-centered data after filtering by the particular date we chose. This preserves the relative magnitudes of CDS.

The loadings show what each component represents, and the scores represent how companies differ. These help us understand what kind of variation each component captures. A uniform positive loading indicates a "level" shift across all maturities, while opposite signs between short and long maturities suggest a "slope". Secondly, we look at the scores, corresponding to each company along the directions of PC1 and PC2. This visualizes how companies differ based on the principal components.

PCA works well in this case because the 10 CDS spreads are correlated, and the first few components capture almost all of the variation. This allows us to reduce the dataset to two dimensions while still preserving its essential information. These findings closely align with those of Litterman and Scheinkman (1991), who found similar patterns when analyzing bond yield curves.

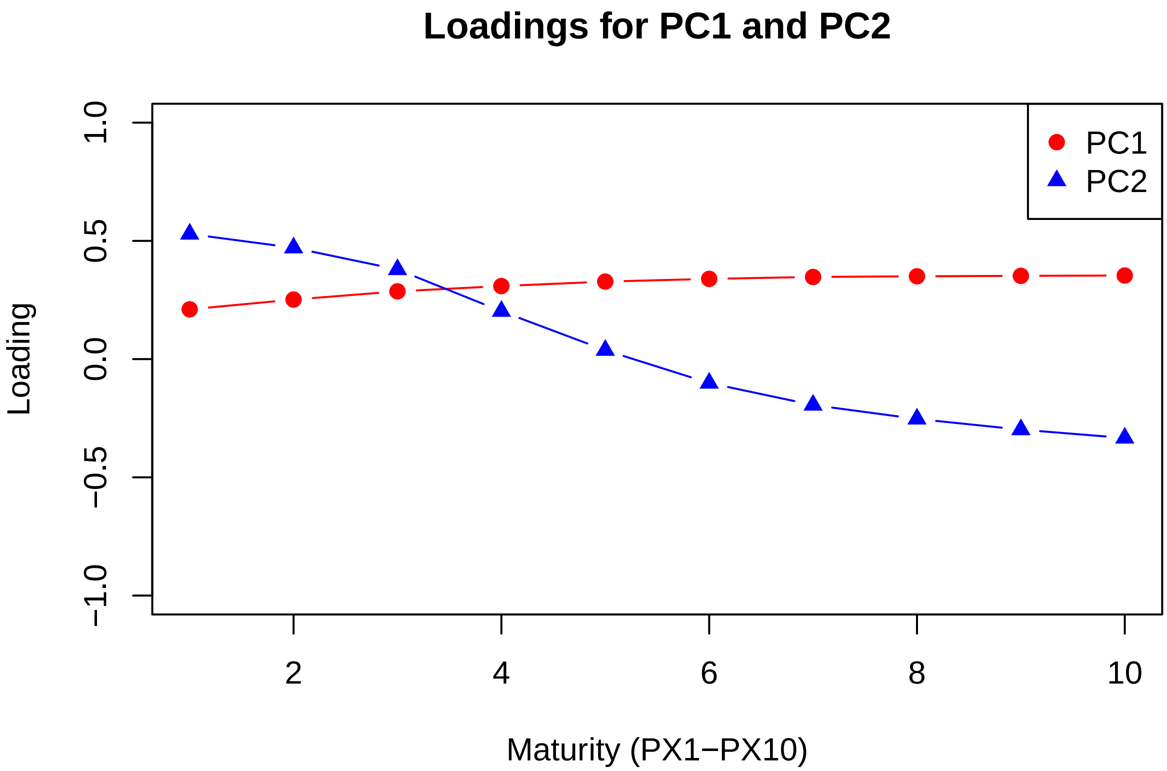
4 Results

4.1 Scree Plot



The scree plot shows that the first two principal components explain almost all of the variation in the data: PC1 accounts for over 98% of the variance, while PC2 contributes nearly 1.5%. All other components explain almost no additional variation. This justifies our decision to interpret only the first two components.

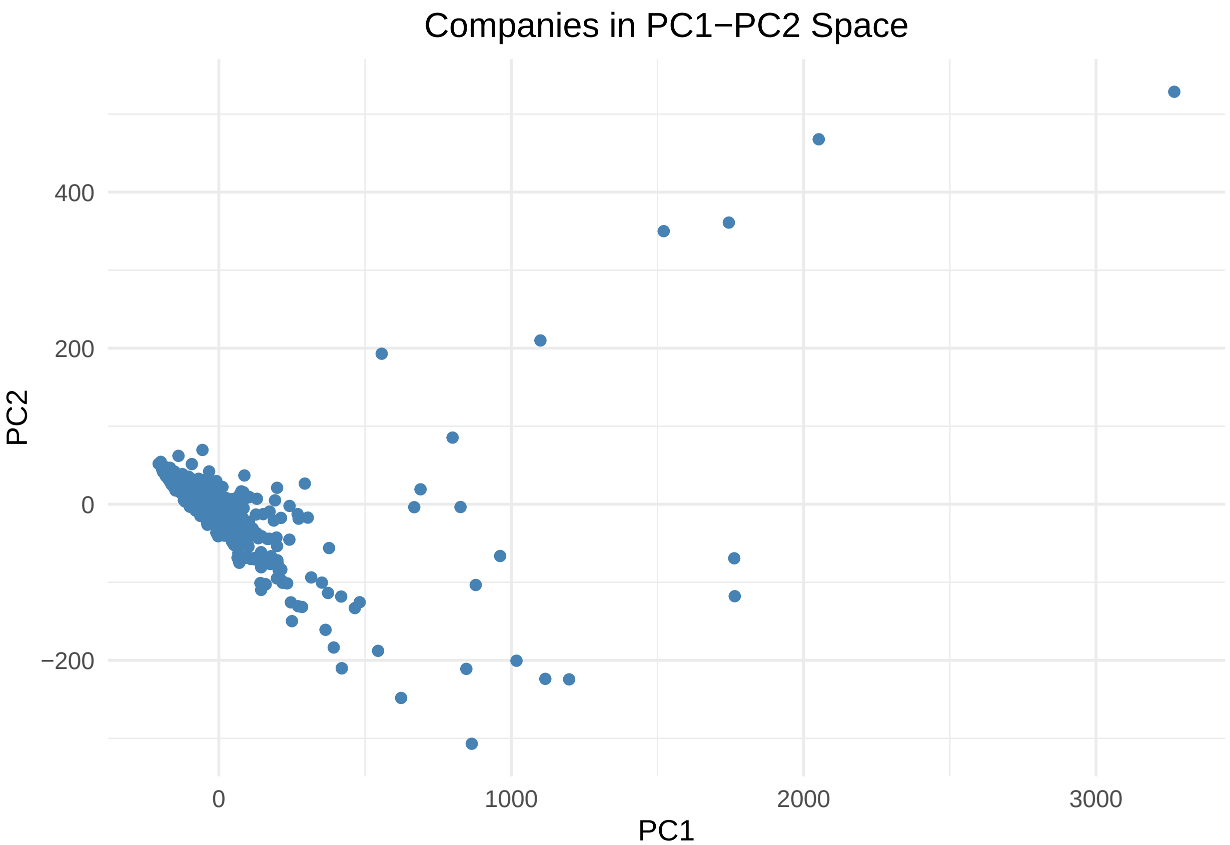
4.2 Loading Plot



The loading plot visualizes the weights of each CDS maturity (PX1 to PX10) for the first two principal components. For PC1, all loadings are positive and very close to each other, indicating PC1 is essentially the average level of CDS spreads across maturities. This suggests PC1 captures the overall level of credit risk for a company.

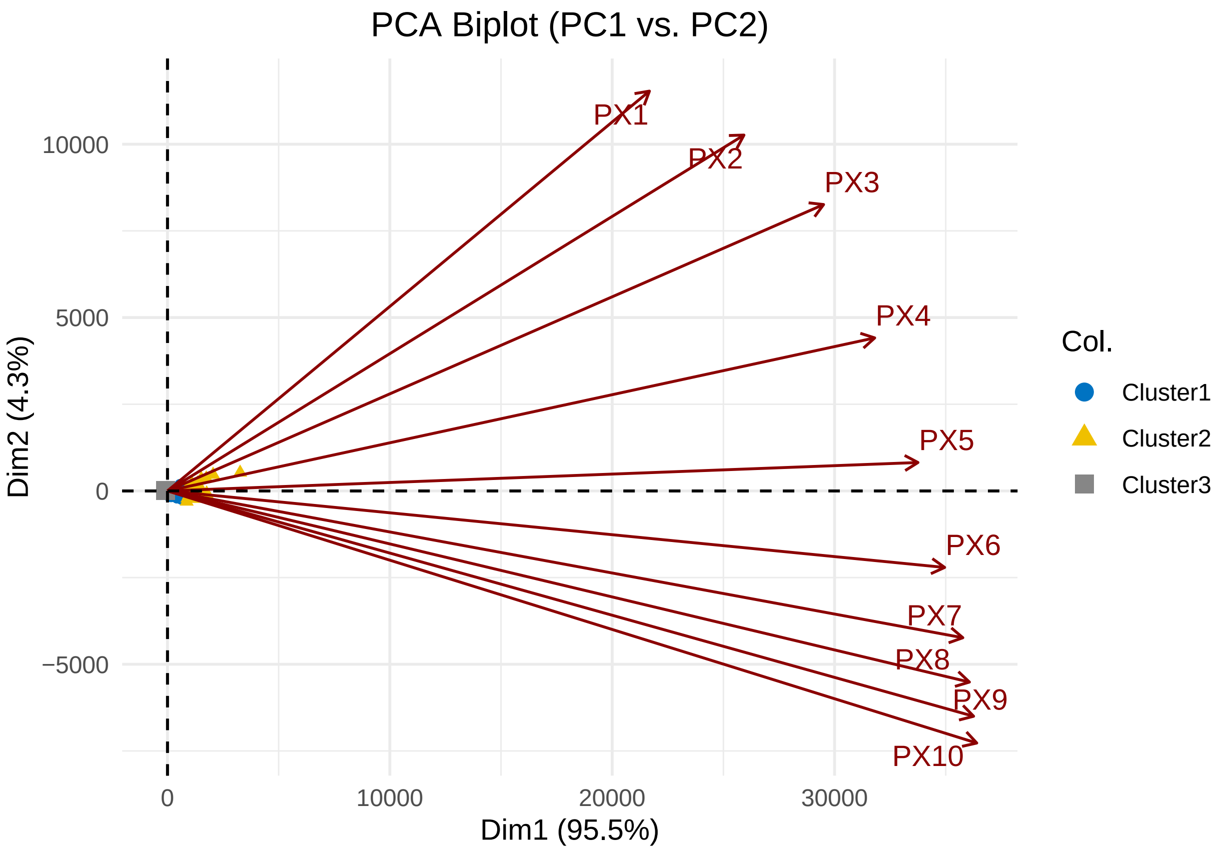
For PC2, the loadings start high for short-term spreads (PX1–PX3), decrease gradually, and become negative for long-term spreads (PX7–PX10). This pattern shows that PC2 captures the slope of the CDS curve: whether a company's short-term credit risk is higher or lower relative to its long-term risk.

4.3 PCA Score Plot



The score plot positions each company in a 2D space defined by PC1 and PC2. Most companies are clustered near the center, but a few appear as outliers. Companies with high PC1 scores have uniformly high CDS spreads (higher credit risk), while companies with low PC1 scores have lower spreads. Companies with high positive PC2 scores tend to have inverted curves (higher short-term spreads), while those with strongly negative PC2 scores have steep curves (higher long-term spreads).

4.4 PCA Biplot



The biplot combines both scores and loadings. Each point represents a company, colored by cluster, and the arrows represent the contribution of each original variable (PX1–PX10) to the PCs. The alignment of the arrows confirms the interpretations above: all arrows pointing in a similar direction along PC1 support the level factor, while the spread from top-left to bottom-right along PC2 reflects the difference between short and long maturities of the slope factor.

4.5 Summary of Findings

We find that the CDS term structures across companies can be effectively summarized using two principal components, which well supports our research question. PC1 reflects the overall level of credit risk, represented by a parallel shift across maturities. While PC2 captures the slope of the term structure, indicating short-term versus long-term risk (Litterman and Scheinkman, 1991). Most companies have CDS spreads that are consistently high or low across all maturities, mainly differ in how risky they are overall. Only a few companies show unusual patterns, such as higher short-term risk and lower long-term risk.

This result is easy to understand and helps compare credit risk between companies. For example, companies with high PC1 scores tend to have high CDS spreads at all maturities — meaning the market sees them as generally riskier. In contrast, companies with very high or very low PC2 scores have unusual patterns in their spread curves — like much higher short-term risk than long-term, or the reverse. This may reflect how the market views their short-term vs. long-term stability.

These findings were expected and are consistent with prior work on term structure analysis, particularly the study by Litterman and Scheinkman (1991), which shows that most changes in bond yield curves can be explained by just a few main patterns, like overall level and slope. Our analysis finds the same kind of structure in CDS data, which supports using PCA to better understand and compare credit risk across companies.

5 References

Alexander, C., & Kaeck, A. (2008). *Regime dependent determinants of credit default swap spreads. Journal of Banking & Finance*, *32*(6), 1008–1021. https://doi.org/10.1016/j.jbankfin.2007.08.002

Blanco, R., Brennan, S., & Marsh, I. W. (2005). *An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. The Journal of Finance*, *60*(5), 2255–2281. https://doi.org/10.1111/j.1540-6261.2005.00798.x

Litterman, R., & Scheinkman, J. (1991). *Common factors affecting bond returns. The Journal of Fixed Income*, *1*(1), 54–61. https://doi.org/10.3905/jfi.1991.692347

6 Appendix

6.1 Codes

```
# Load libraries

library(tidyverse)

library(factoextra)

library(ggplot2)

library(dplyr)

library(tidyr)

library(knitr)

library(kableExtra)


# Load and process data

load("CDS_data.RData")

data <- data


# Filter one date (snapshot)

target_date <- "2020-01-02"

data_day <- data %>%

  filter(Date == target_date) %>%

  select(Ticker, PX1:PX10) %>%

  distinct() %>%

  na.omit()


# Matrix of companies × 10 maturities

X <- as.matrix(data_day[, -1])

rownames(X) <- data_day$Ticker


# Center data (no scaling)

X_centered <- scale(X, center = TRUE, scale = FALSE)


# Perform PCA
```

```
pca_result <- prcomp(X_centered, center = TRUE, scale. = FALSE)

# Variance explained
var_explained <- pca_result$sdev^2 / sum(pca_result$sdev^2)
scree_data <- data.frame(PC = 1:length(var_explained), Variance = var_explained)

# Loadings and scores
loading_df <- data.frame(Maturity = 1:10,
                         PC1 = pca_result$rotation[, 1],
                         PC2 = pca_result$rotation[, 2])

score_df <- data.frame(Ticker = rownames(X),
                       PC1 = pca_result$x[, 1],
                       PC2 = pca_result$x[, 2])

# Summary statistics table
stats_df <- data_day %>%
  summarise(across(PX1:PX10, list(mean = mean, sd = sd))) %>%
  pivot_longer(everything(),
               names_to = c("PX", "stat"),
               names_pattern = "(.+)_(mean|sd)") %>%
  pivot_wider(names_from = stat, values_from = value)

kable(stats_df, caption = "Mean and sd of CDS Spreads (PX1-PX10)")

stats_df$PX <- factor(stats_df$PX, levels = paste0("PX", 1:10))
# Plot: Mean ± SD
ggplot(stats_df, aes(x = PX, y = mean, group = 1)) +
  geom_line(color = "steelblue", size = 1.2) +
  geom_point(color = "steelblue", size = 2) +
  geom_errorbar(aes(ymin = mean - sd, ymax = mean + sd),
                width = 0.2, color = "gray40") +
```

```r
  labs(title = "Mean and Variability of CDS Spreads by Maturity",
       y = "Spread (bps)", x = "Maturity") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

# Scree plot
ggplot(scree_data, aes(x = PC, y = Variance)) +
  geom_point(size = 3, color = "steelblue") +
  geom_line(linewidth = 1, color = "steelblue") +
  geom_vline(xintercept = 2, linetype = "dashed", color = "red") +
  labs(title = paste("Scree Plot on", target_date),
       x = "Principal Component",
       y = "Proportion of Variance Explained") +
  scale_x_continuous(breaks = 1:ncol(X)) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

# Loadings plot
loadings <- pca_result$rotation
plot(loadings[, 1], type = "b", pch = 19, col = "red", ylim = c(-1, 1),
     main = "Loadings for PC1 and PC2", xlab = "Maturity (PX1-PX10)", ylab = "Loading")
lines(loadings[, 2], type = "b", pch = 17, col = "blue")
legend("topright", legend = c("PC1", "PC2"), col = c("red", "blue"), pch = c(19, 17))

# PCA score plot
ggplot(score_df, aes(x = PC1, y = PC2)) +
  geom_point(color = "steelblue") +
  labs(title = "Companies in PC1-PC2 Space",
       x = "PC1", y = "PC2") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

```
# PCA biplot
set.seed(23)
clusters <- kmeans(pca_result$x[, 1:2], centers = 3)$cluster
cluster_labels <- paste0("Cluster", clusters)

fviz_pca_biplot(pca_result,
                axes = c(1, 2),
                geom = "point",
                col.var = "darkred",
                col.ind = cluster_labels,
                palette = "jco",
                repel = TRUE,
                title = "PCA Biplot (PC1 vs. PC2)") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```