Group Assignment

Group Members:

- Bryan Bastian Sambada – 33410003
- Cedric Rennard Anggawan – 33371857
- Gen Sheng Goh – 31927874
- Junbin Li – 33032556
- Zihao Qi – 31941885

ETF3500/5500 – High Dimensional Data Analysis

20 September 2024

Data Description

The portfolio consists of monthly returns for 100 stocks spanning January 2005 to December 2019, each identified by a PERMNO code and an industry letter. The industries represented include Mining (B), Manufacturing (D), Transportation and Public Utilities (E), Wholesale Trade (F), Retail Trade (G), Finance, Insurance, and Real Estate (H), and Services (I). By applying Principal Component Analysis (PCA) and Factor Analysis, we can better understand the covariance structure among these stocks, which provides valuable insights for identifying diversification opportunities and selecting stocks based on their shared risk factors.

Preliminary Analysis

Before conducting our analysis, we ensured there were no missing values, verified the data was in a tidy format, and checked for any obvious outliers. In our initial exploration, we identified a few outliers when plotting the portfolio's stock returns over time. However, we concluded that these extreme values aligned with typical financial return distributions. As expected, the majority of stock returns were clustered around zero, with some outliers reflecting the expected volatility in financial markets.

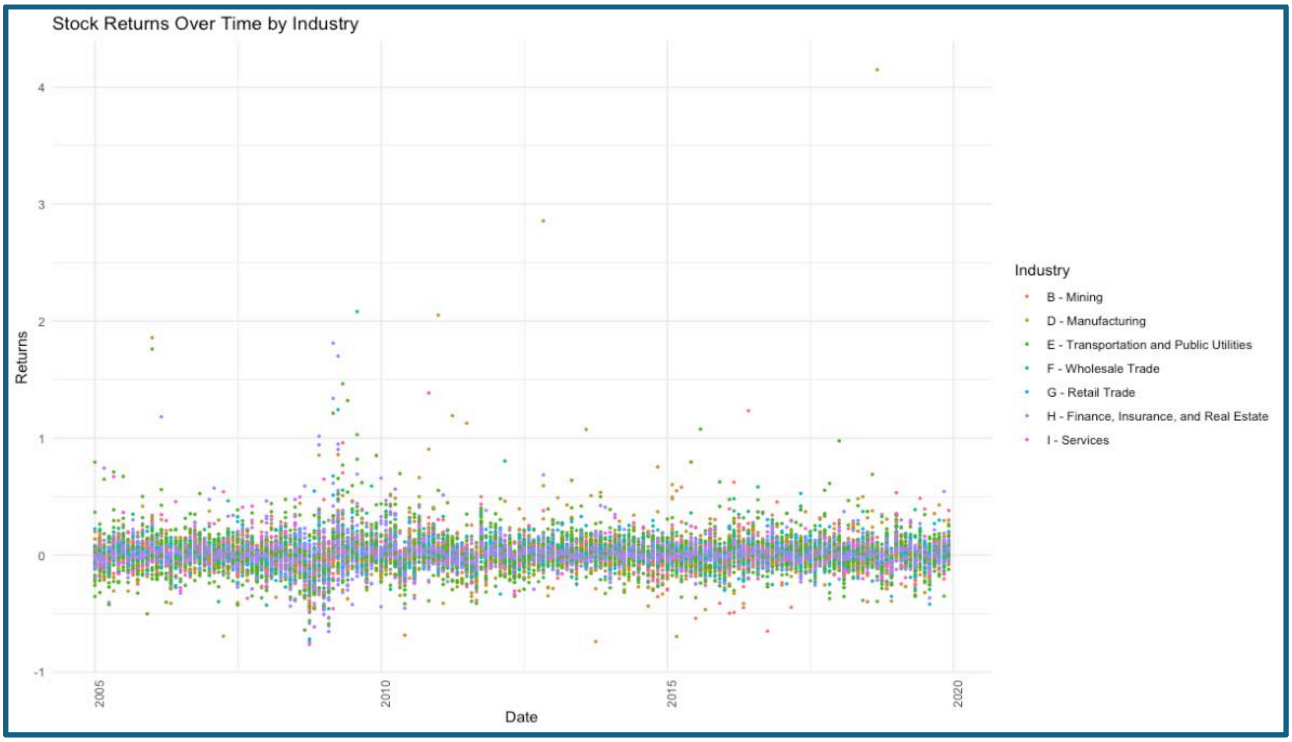

Fig 1. Individual Stock Returns Time Plot

Due to the dataset's high dimensionality, basic summary statistics fail to reveal common patterns or covariation between individual stocks. To simplify, we group stocks by industry and use mean returns for each group. However, this introduces variability disparities between industries with differing numbers of stocks, potentially skewing results. Thus, this analysis is only a preliminary step to understand the dataset’s structure before performing PCA and Factor Analysis, providing a basic overview through visualizations and summary statistics despite its limitations.

Mean Returns over Time

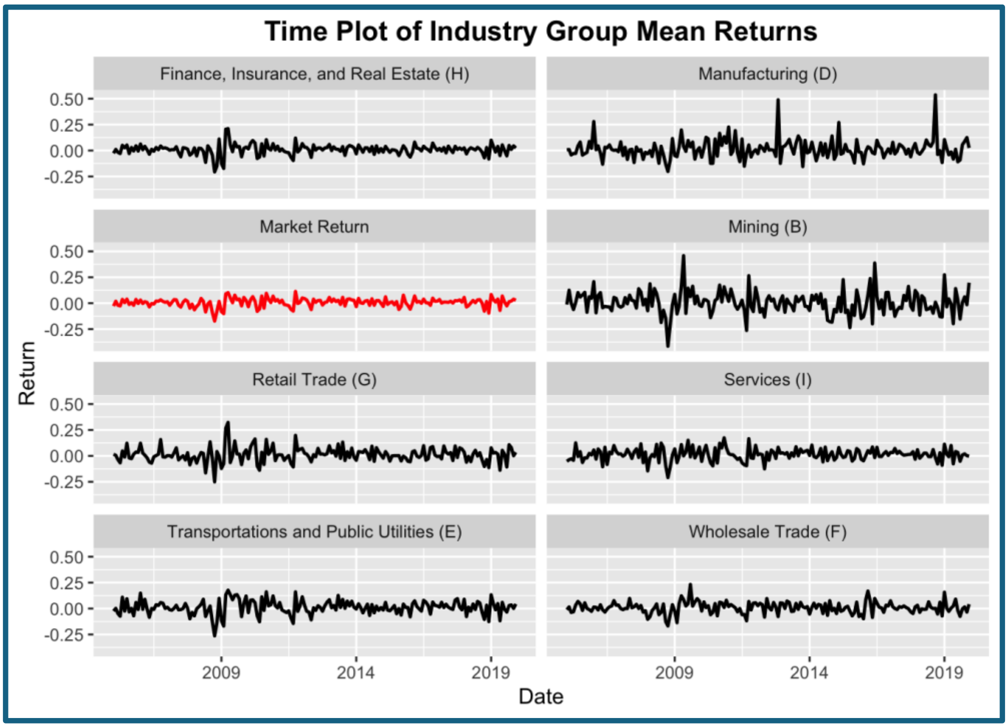

Fig. 2 Industry Group Time Plots

The plot illustrates the average stock returns by industry group over time. All industry groups show greater variability compared to market movements, with industries B and D having the highest volatility. Additionally, industries E, G, and H align more closely with market trends, supported by their higher correlation coefficients (refer to Appendix A).

Principle Component Analysis (PCA)

Principal components analysis creates linear combinations of variables to capture the majority of the data's variation. We standardize the data to account for potential errors across stocks. By using two principal components, we can visualize stock correlations in a biplot, where the angles between vectors indicate the relationship between stock movements.

PC number Selection

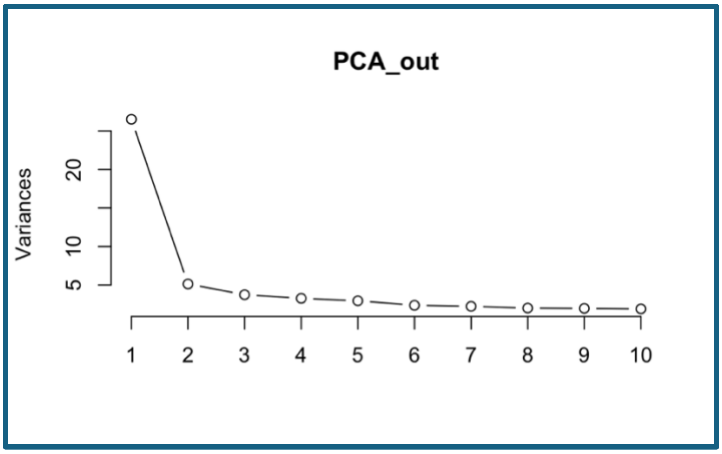

Fig 3. Scree Plot

To determine the optimal number of principal components (PCs), we applied both the elbow and Kaiser's rules. The elbow rule indicated retaining two PCs based on the scree plot's flattening, while Kaiser's rule suggested retaining up to 27 PCs with variances above 1. We initially used two PCs and considered adding more, but found no significant benefits beyond the first two.

Comparison of PCs

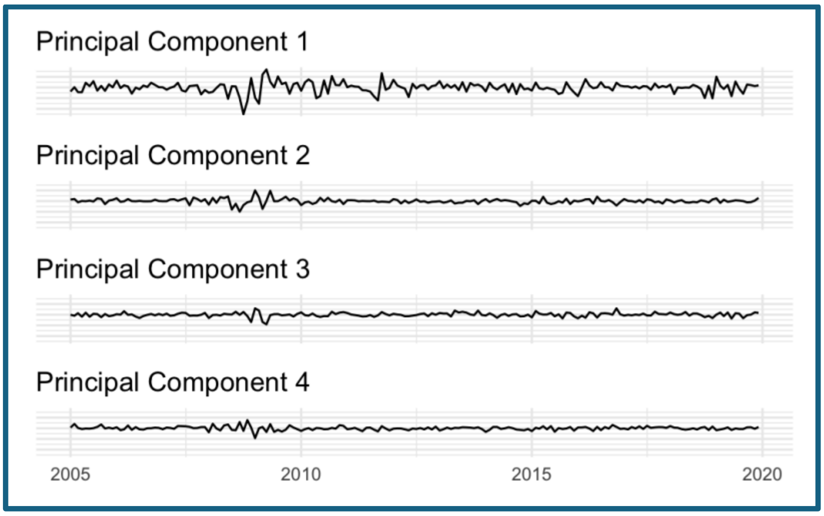

Fig 4. Principle Components Time Plots

When plotting the PCs over time, PC1 captures the most variation, followed by PC2. In contrast, PC3 and PC4 show minimal variation with flatter plots, indicating they add little information beyond the first two. Together, PC1 and PC2 account for 32% of the dataset's variation (see Appendix B), and adding further components likely won't yield significant insights.

PC1 vs Market Return

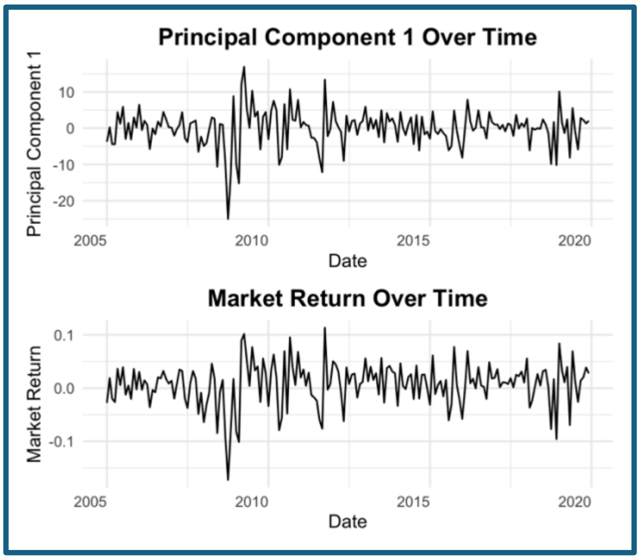

Fig. 5 PC1 vs Market Return

When analysing the time plot of PC1 against market returns, we see that PC1 closely mirrors overall market behaviour, including volatile periods like the 2008-2009 financial crisis. The high correlation coefficient of 0.9482 between PC1 and market returns further emphasizes this strong positive relationship, making PC1 a reliable indicator of broader market trends.

PCA Correlation Biplot

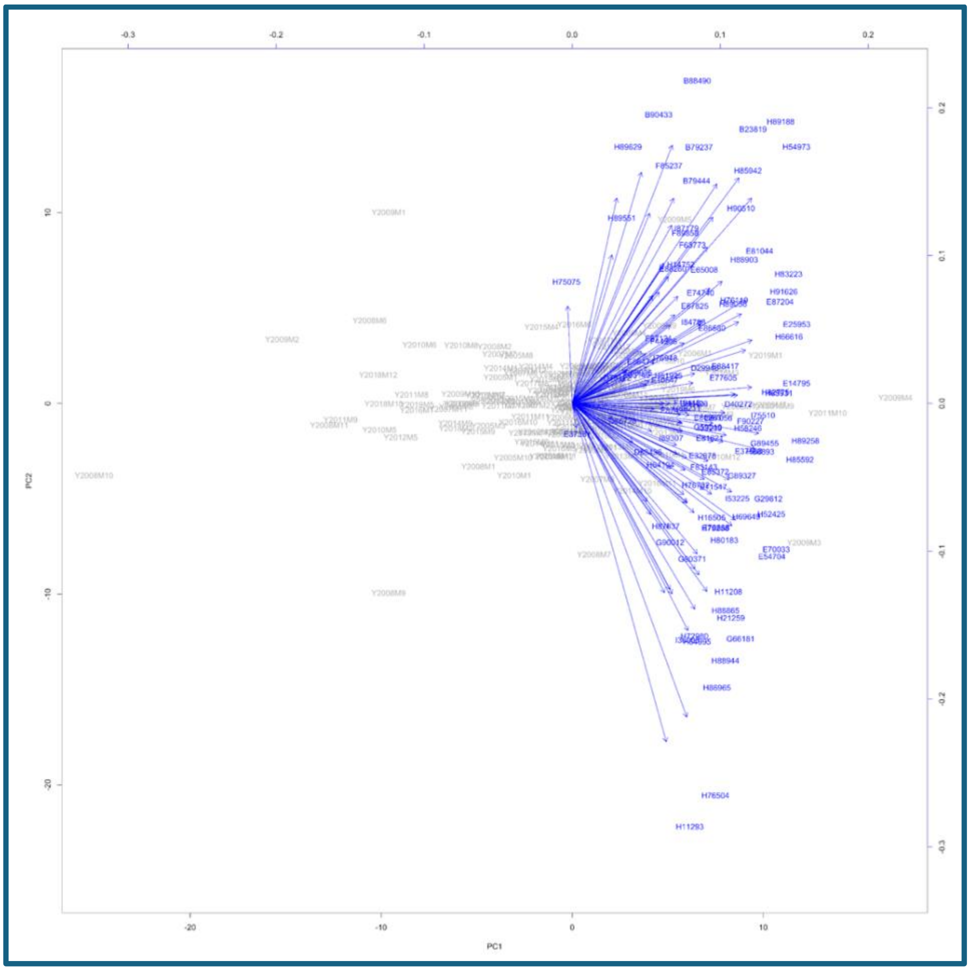

Fig. 6 Correlation Biplot

We examine the correlation biplot to understand relationships between stocks in the portfolio. Most stocks are positively aligned with PC1, reflecting their similar correlation with market movements. However, PC2 reveals mixed correlations, with some stocks being positively correlated and others negatively correlated. By analysing the angles between variable vectors, we can determine the level of covariation between individual stocks.

Pairs of stocks with **strong positive correlations** are identified by the small angles between their arrow vectors in the biplot, indicating that they follow similar directions along both PC1 and PC2.

- Stocks H72980 and H74995 both show a positive relationship with PC1 and a negative relationship with PC2, while stocks H14752 and E88260 move positively along both PC1 and PC2. These pairs of stocks exhibit high positive correlations**.**

Pairs of stocks with **strong negative correlations** are identified by the large angles between their arrow vectors in the biplot, indicating significant differences in their relationships with PC1 and PC2.

- Stocks H89629 and H11293 both move positively along PC1 but show substantial differences in their movement with PC2, resulting in a large angle between them, indicating a strong negative correlation.

Pairs of stocks with **little to no correlation** are identified by a near 90-degree angle between their arrow vectors in the biplot, indicating minimal or no relationship between their movements.

- E70033 and I87179 form an almost right angle between their vectors, suggesting they are not correlated.

Factor Models Analysis

Factor analysis is a technique used to simplify complex datasets by linking the covariation between observations to unobserved common factors. In this analysis, we will examine the factor loadings and scores to interpret how the stocks' movements align with these underlying factors, providing insights into their shared behaviours.

Choice of Number of Factors

We began our analysis with 2 factors, corresponding to the principal components from earlier, and gradually increased to 7 factors to match the number of industries in the dataset. This was done to determine if additional factors would improve the analysis. However, we found that 2 factors were sufficient, as adding more did not uncover significant new patterns or provide better insights into stock behaviour.

Evaluating Rotations

We considered using rotations like Varimax and Promax because they can make factors easier to interpret. These rotations help simplify the structure, with Varimax aiming for orthogonality and Promax allowing correlation between factors. By applying rotations that produce more zero loadings, we can better identify distinct patterns and improve the clarity of factor interpretations.

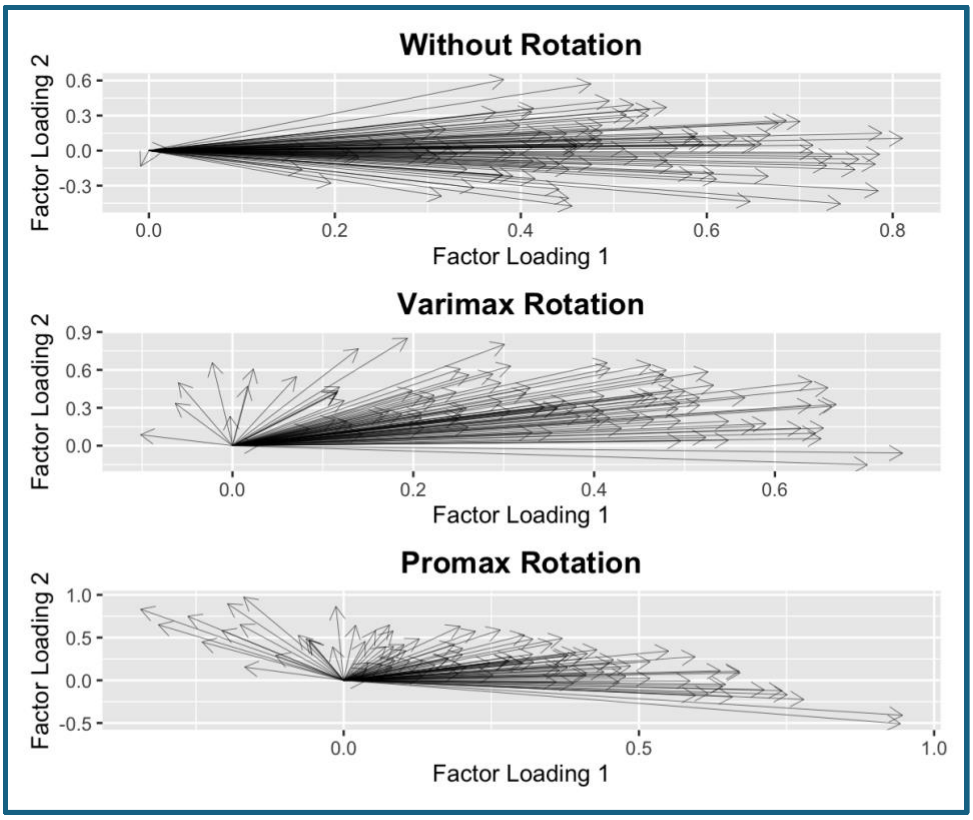

Fig. 7 Factor Analysis Rotations

Analysing the graphs, we find that while both rotations allow more values to align with the vertical axis, they also result in many stocks positioned diagonally between the two axes, rather than near or on the axes. This increases variation and makes interpreting the data more complex, making it harder to draw clear conclusions from the factor loadings. Therefore, using no rotation is preferable. In the unrotated model, approximately 30.4% of the covariation in stocks is explained by the common factors (Appendix C).

Factor 1 vs Market Return

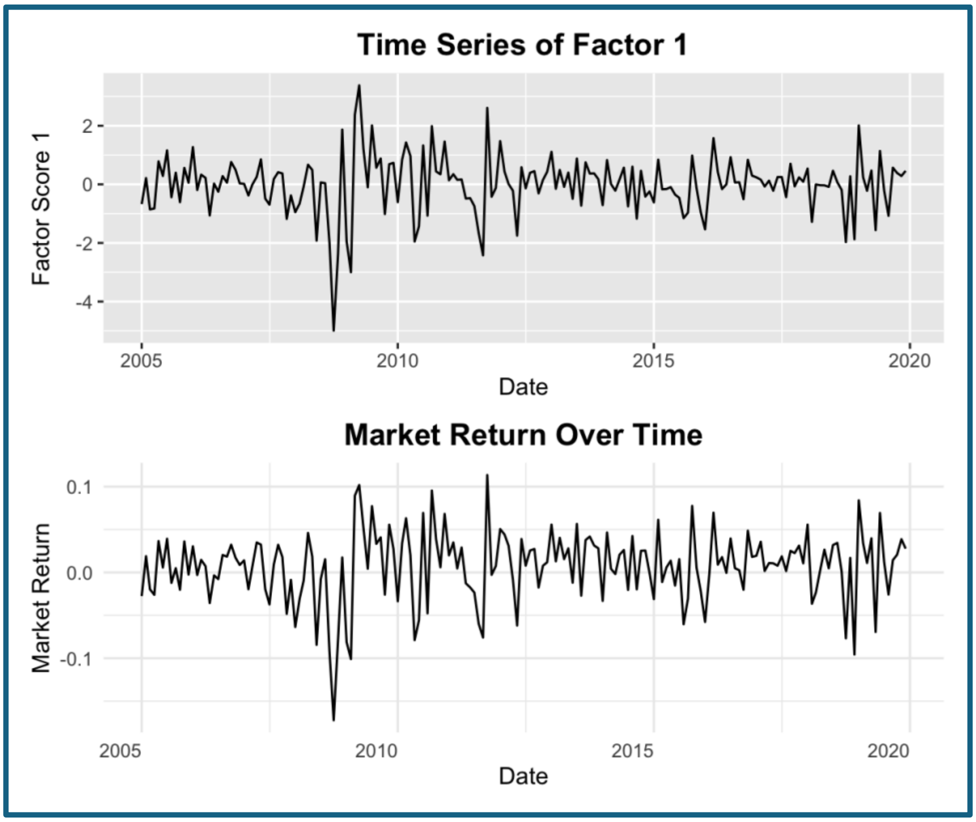

Fig. 8 Factor 1 and Market Return Time Plot

Comparing the factor scores of factor 1 with market returns over time reveals a close alignment, supported by a high correlation coefficient of 0.9455 (Appendix C). This indicates a strong positive correlation, making factor 1 a reliable indicator of market movements.

Stocks Code	Factor Loading 1	Correlation with Market Return
E37381	0.019	0.026
H75075	-0.001	-0.032
E14795	0.786	0.732
H89551	0.165	0.137

Table 1

This is further supported when we look analyse factor loadings of individual stocks and their respective correlation coefficients with market returns over time. Stocks with loadings near zero and high absolute values show negligible and strong correlations with market returns respectively, Strengthening the argument that factor 1 is highly likely to be closely related to market return.

Factor Loading Interpretations

Loadings:

	Factor1	Factor2
B23819	0.647	-0.437
B79237	0.451	-0.408
B79444	0.441	-0.332
B88490	0.455	-0.476
B90433	0.315	-0.391
D29946	0.457	
D40272	0.569	
D79122	0.157	
D86728	0.167	
D87056	0.488	
D88436	0.250	
D89935	0.225	
D90423	0.413	
E10547	0.301	
E11547	0.479	0.106
E14795	0.786	
E25953	0.781	-0.116
E32678	0.434	0.114

Fig. 9 Factor Loadings

After fitting a 2-factor model to the dataset, clear patterns emerge. Factor 1 shows mostly positive loadings, with only a few stocks showing no correlation. This supports previous analysis that Factor 1 is closely related to general market movements, as stocks tend to follow market trends. However, the magnitude of these loadings varies across industries, and no further insights can be drawn from this factor alone.

Factor 2 reveals more distinct industry-specific differences. All mining stocks show negative loadings, indicating an inverse relationship with Factor 2. In contrast, manufacturing stocks exhibit zero loadings, suggesting they are unaffected by it. This shows that the mining sector consistently moves in opposition to Factor 2, while manufacturing remains insulated from its influence. Other industries display a variety of loadings without clear patterns, reflecting different sensitivities to Factor 2 across sectors.

The only industries showing consistent patterns across both factors are mining and manufacturing. Mining stocks are positively related to Factor 1 and negatively to Factor 2, while manufacturing stocks are positively related to Factor 1 and show no correlation with Factor 2. This insight is valuable for sector-specific investment strategies, as it helps predict how these industries might respond to broader market trends (Factor 1). Further analysis may gradually reveal the influences of Factor 2, aiding in risk management or sector-focused opportunities.

Uniqueness
In a 2-factor model, we aim to explain stock movements using two common factors that represent market risk or overall market trends. The uniqueness of each stock reflects the portion of its price movement that cannot be attributed to these factors, indicating idiosyncratic or stock-specific risk. Stocks with high idiosyncratic risk are driven by unique factors unrelated to market movements, offering potential diversification benefits but also greater volatility and unpredictability.

Top 5 Highest Uniqueness

Stock	E37381	H75075	D79122	D86728	E83149
Uniqueness	0.9992	0.9821	0.9741	0.9708	0.9537

Top 5 Lowest Uniqueness

Stock	H89188	H54973	H89258	H85592	E25953
Uniqueness	0.2389	0.2656	0.3320	0.3556	0.3763

Stock E37381 has the highest uniqueness, with an idiosyncratic error of 0.9992, meaning 99.92% of its return variation is not explained by the common factors. In contrast, stock H89188 has the lowest idiosyncratic error at 0.2389, meaning only 23.89% of its variation remains unexplained by common factors.

Conclusions

Assumptions
In this analysis, we made several important assumptions regarding both the data and the methods used for PCA and Factor Analysis. First, we assume that any outliers in the dataset are not due to recording errors, as the data source is unavailable for verification. Additionally, we rely on key assumptions underpinning PCA and Factor Modelling, such as uncorrelated idiosyncratic errors, each with its own variance. For PCA, we also adopt the optimization constraint that the sum of the squared component loadings equals one. These assumptions are essential for the validity and interpretability of our results.

Limitations of Analysis
Any dimension reduction technique, such as principal component analysis (PCA), results in some loss of information. In this case, the first two principal components explain 31.65% of the overall variation, which is represented in the biplot. Similarly, the first two factor scores account for only 30.4% of the variation. Determining the optimal number of PCs and factors is challenging, as better configurations may exist. It's possible that using only 2 factors for both PCA and factor analysis is insufficient, leaving some covariation unexplained, and extending the models might capture more insights we didn't consider.

Summary
Based on the Factor Modelling and Principal Component Analysis, several conclusions can be made. The largest principal component and factor are both tied to general market movements. The PCA biplot can be further examined to identify stock pairs with favourable correlation patterns for building a diversified portfolio. Additionally, factor loadings help assess the portfolio’s alignment with unobserved market factors. The Mining and Manufacturing industries stand out, with stocks in these sectors consistently moving together. However, certain stocks, like E37381, provide little insight into covariation patterns, as it has a high idiosyncratic error of 99.92%, meaning it is almost entirely unexplained by the factors.

References

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

Pedersen T (2024). _patchwork: The Composer of Plots_. R package version 1.2.0, <https://CRAN.R-project.org/package=patchwork>.

Robinson D, Hayes A, Couch S (2024). _broom: Convert Statistical Objects into Tidy Tibbles_. R package version 1.0.6, <https://CRAN.R-project.org/package=broom>.

Load the necessary dataset and library:

```
library(tidyverse)
library(MASS)
library(patchwork)
library(broom)

Market <- read_csv('Market.csv') %>%
  mutate(MarketReturn = MarketReturn / 100)
Portfolio <- read_csv('SampleE.csv')
Combined_Portfolio <- inner_join(Portfolio, Market, by = "Date")
```

Appendix A: Preliminary Data Analysis

```
#Data Cleaning
sum(is.na(Portfolio))
```

```
#Preliminary Data Analysis
df_long <- Combined_Portfolio %>%
  pivot_longer(cols = -Date, names_to = "Stock", values_to = "Value")

df_grouped <- df_long %>%
  mutate(Group = substr(Stock, 1, 1)) %>%
  group_by(Date, Group) %>%
  summarise(MeanValue = mean(Value, na.rm = TRUE)) %>%
  ungroup()

df_grouped <- df_long %>%
  mutate(Group = substr(Stock, 1, 1)) %>%
  mutate(Group = case_when(
    Group == "B" ~ "Mining (B)",
    Group == "C" ~ "Construction (C)",
    Group == "D" ~ "Manufacturing (D)",
    Group == "E" ~ "Transportations and Public Utilities (E)",
    Group == "F" ~ "Wholesale Trade (F)",
    Group == "G" ~ "Retail Trade (G)",
    Group == "H" ~ "Finance, Insurance, and Real Estate (H)",
    Group == "I" ~ "Services (I)",
    Group == "M" ~ "Market Return",
    TRUE ~ Group
  )) %>%
  group_by(Date, Group) %>%
  summarise(MeanValue = mean(Value, na.rm = TRUE)) %>%
  ungroup()

df_grouped <- df_grouped %>%
  mutate(Date = as.Date(paste0(gsub("Y", "", gsub("M", "-", Date)), "-01"), format = "%Y
-%m-%d"))
```

```
plot1 <- ggplot(df_grouped, aes(x = Date, y = MeanValue, group = Group)) +
  geom_line(aes(color = ifelse(Group == "Market Return", "Market Return", "Other")), siz
e = 0.8) +
  labs(title = "Time Plot of Industry Group Mean Returns", x = "Date", y = "Return") +
  scale_x_date(date_breaks = "5 years", date_labels = "%Y") +
  scale_color_manual(values = c("Market Return" = "red", "Other" = "black"))+
  facet_wrap(~Group, ncol = 2)+
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))

plot1
```

```
df_wide <- df_grouped %>%
  pivot_wider(names_from = Group, values_from = MeanValue)

cor_matrix <- cor(df_wide[, -1], use = "complete.obs")

market_return_correlations <- cor_matrix["Market Return", ] %>%
  as.data.frame() %>%
  round(4)
colnames(market_return_correlations) <- ("Market Return")

print(market_return_correlations)
```

```
# Create the new column for Industry by extracting the first letter of each stock name
Portfolio_long <- Portfolio %>%
  pivot_longer(cols = -Date, names_to = "Stock", values_to = "Returns")

# Extract the first letter of each Stock to create an Industry column
Portfolio_long <- Portfolio_long %>%
  mutate(Industry = substr(Stock, 1, 1))

# Map the first letters to the corresponding industry names
Portfolio_long$Industry <- recode(Portfolio_long$Industry,
                                  "B" = "B - Mining",
                                  "D" = "D - Manufacturing",
                                  "E" = "E - Transportation and Public Utilities",
                                  "F" = "F - Wholesale Trade",
                                  "G" = "G - Retail Trade",
                                  "H" = "H - Finance, Insurance, and Real Estate",
                                  "I" = "I - Services")

# Convert the Date column to Date format
Portfolio_long$Date <- as.Date(paste0(gsub("Y", "", gsub("M", "-", Portfolio_long$Dat
e)), "-01"), format = "%Y-%m-%d")

# Create the time series plot with color based on Industry
ggplot(Portfolio_long, aes(x = Date, y = Returns, color = Industry)) +
  geom_point(cex = 0.5) +
  labs(title = "Stock Returns Over Time by Industry",
       x = "Date",
       y = "Returns",
       color = "Industry") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Appendix B: Principal Component Analysis

```
##Principal Component Analysis
PCA_out <- Portfolio %>%
  column_to_rownames('Date') %>%
  prcomp(scale. = TRUE)

summary(PCA_out)
```

```
screeplot(PCA_out,type = 'l')

biplot(PCA_out, scale = 0, cex = 0.5, arrow.len = 0.1, xlab = "PC1", ylab = "PC2", col =
c("lightgrey", "blue"))
```

```
cor(Market$MarketReturn, PCA_out$x)

#### Market Return and PC1 overtime
PC_df <- as.data.frame(PCA_out$x)
# Re-attach the Date column to the PCA results
PC_df$Date <- rownames(PC_df)
PC_df$Date <- as.Date(paste0(gsub("Y", "", gsub("M", "-", PC_df$Date)), "-01"), format =
"%Y-%m-%d")
Market$Date <- as.Date(paste0(gsub("Y", "", gsub("M", "-", Market$Date)), "-01"), format
= "%Y-%m-%d")
# Determine the y-axis limits based on PC1 with a buffer of +/- 5
y_limits <- range(PC_df$PC1)
```

```
# Create PC1 with consistent y-axis scale and title
p1 <- ggplot(PC_df, aes(x = Date, y = PC1)) +
  geom_line() +
  labs(title = "Principal Component 1", y = NULL) +  # Title for PC1
  theme_minimal() +
  theme(axis.text.x = element_blank(),  # Remove x-axis labels
        axis.title.x = element_blank(),
        axis.text.y = element_blank(),  # Remove y-axis labels
        axis.ticks.y = element_blank()) +
  ylim(y_limits)

# Create PC2 with consistent y-axis scale and title
p2 <- ggplot(PC_df, aes(x = Date, y = PC2)) +
  geom_line() +
  labs(title = "Principal Component 2", y = NULL) +  # Title for PC2
  theme_minimal() +
  theme(axis.text.x = element_blank(),  # Remove x-axis labels
        axis.title.x = element_blank(),
        axis.text.y = element_blank(),  # Remove y-axis labels
        axis.ticks.y = element_blank()) +
  ylim(y_limits)

# Create PC3 with consistent y-axis scale and title
p3 <- ggplot(PC_df, aes(x = Date, y = PC3)) +
  geom_line() +
  labs(title = "Principal Component 3", y = NULL) +  # Title for PC3
  theme_minimal() +
  theme(axis.text.x = element_blank(),  # Remove x-axis labels
        axis.title.x = element_blank(),
        axis.text.y = element_blank(),  # Remove y-axis labels
        axis.ticks.y = element_blank()) +
  ylim(y_limits)

# Create PC4 with consistent y-axis scale and title
p4 <- ggplot(PC_df, aes(x = Date, y = PC4)) +
  geom_line() +
  labs(title = "Principal Component 4", y = NULL) +  # Title for PC4
  theme_minimal() +
  theme(axis.title.x = element_blank(),
        axis.text.y = element_blank(),  # Remove y-axis labels
        axis.ticks.y = element_blank()) +
  ylim(y_limits)

# Combine the four plots
combined_plot <- p1 / p2 / p3 / p4

# Display the combined plot
print(combined_plot)
```

```
cor(Market$MarketReturn, PC_df$PC1)
```

Appendix C: Factor Analysis

```
FA <- Portfolio %>%
  column_to_rownames('Date')%>%
  scale() %>%
  factanal(factors = 2,rotation = 'none',scores = 'none')

# Extract loadings directly from the factanal object
loading <- FA$loadings

# Convert the loadings to a dataframe
loading_df <- as.data.frame(unclass(loading))
loading_df_sorted <- loading_df[order(rownames(loading_df)), ]
loading_df_sorted[loading_df_sorted > -0.1 & loading_df_sorted < 0.1] <- 0
loading_df_sorted <- loading_df_sorted %>%
  mutate(across(everything(), ~ round(., 3))) %>%
  mutate(across(everything(), ~ ifelse(. == 0, "-", .)))


cor(Portfolio$E37381, Market$MarketReturn)
cor(Portfolio$H75075, Market$MarketReturn)
cor(Portfolio$E14795, Market$MarketReturn)
cor(Portfolio$H89551, Market$MarketReturn)

FA$uniquenesses
sorted_uniqueness <- sort(FA$uniquenesses, decreasing = TRUE)


# Getting the 5 lowest values
lowest_5 <- head(sorted_uniqueness, 5)

# Getting the 5 highest values
highest_5 <- tail(sorted_uniqueness, 5)

# Display the results
lowest_5
highest_5

FA_v <- Portfolio %>%
  column_to_rownames('Date')%>%
  scale() %>%
  factanal(factors = 2,rotation = 'varimax',scores = 'none')

FA_p <- Portfolio %>%
  column_to_rownames('Date')%>%
  scale() %>%
  factanal(factors = 2,rotation = 'promax',scores = 'none')
```

7 Factors

```
FA_7 <- Portfolio %>%
  column_to_rownames('Date')%>%
  scale() %>%
  factanal(factors = 7,rotation = 'none',scores = 'none')

# Extract loadings directly from the factanal object
loading_1 <- FA_7$loadings

# Convert the loadings to a dataframe
loading_df_1 <- as.data.frame(unclass(loading_1))
loading_df_sorted_1 <- loading_df_1[order(rownames(loading_df)), ]
loading_df_sorted_1[loading_df_sorted_1 > -0.1 & loading_df_sorted_1 < 0.1] <- 0
loading_df_sorted_1 <- loading_df_sorted_1 %>%
  mutate(across(everything(), ~ round(., 3))) %>%
  mutate(across(everything(), ~ ifelse(. == 0, "-", .)))
loading_df_sorted_1
```

```
FA_v$loadings
FA_p$loadings
fa_df<-tidy(FA)
```

```
NoRot<-ggplot(fa_df,aes(x=fl1,y=fl2,
                        label=variable))+
  geom_segment(aes(xend=fl1,
                   yend=fl2,x=0,y=0),
               arrow = arrow(length = unit(0.1, "inches"),
               ),size = 0.1)+
  labs(title = "Without Rotation",
       x = "Factor Loading 1",
       y = "Factor Loading 2")+
  theme(  plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))



fa_df_1<-tidy(FA_v)
VRot <-ggplot(fa_df_1,aes(x=fl1,y=fl2,
                          label=variable))+
  geom_segment(aes(xend=fl1,
                   yend=fl2,x=0,y=0),
               arrow = arrow(length = unit(0.1, "inches"),
               ),size = 0.1)+
  labs(title = "Varimax Rotation",
       x = "Factor Loading 1",
       y = "Factor Loading 2") +
  theme(  plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))


fa_df_2<-tidy(FA_p)
ProRot <- ggplot(fa_df_2,aes(x=fl1,y=fl2,
                             label=variable))+
  geom_segment(aes(xend=fl1,
                   yend=fl2,x=0,y=0),
               arrow = arrow(length = unit(0.1, "inches"),
               ),size = 0.1)+
  labs(title = "Promax Rotation",
       x = "Factor Loading 1",
       y = "Factor Loading 2") +
  theme(  plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))

Merged <- NoRot/VRot/ProRot
Merged
```

```
FA_scores <- Portfolio %>%
  column_to_rownames('Date')%>%
  scale() %>%
  factanal(factors = 2,rotation = 'none',scores = 'Bartlett')
FA_scores$scores <- scale(FA_scores$scores)

df <- as.data.frame(FA_scores$scores)
df$Date <- rownames(df)
df_subset <- df[, c("Date", "Factor1")]
df_subset$Date <- as.Date(paste0(gsub("Y", "", gsub("M", "-", df_subset$Date)), "-01"),
format = "%Y-%m-%d")
```

```
# Extract the Date and Factor1 columns into a new dataframe

p5 <- ggplot(df_subset, aes(x = Date, y=Factor1))+
  geom_line() +
  labs(title = "Time Series of Factor 1",
       x = "Date",
       y = "Factor Score 1") +
  theme(  plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))

combined_plot_1 <- p5 / p1
combined_plot_1
```

```
cor(df_subset$Factor1, Market$MarketReturn)
```

```
cor(Portfolio$E37381, Market$MarketReturn)
cor(Portfolio$H75075, Market$MarketReturn)
cor(Portfolio$E14795, Market$MarketReturn)
cor(Portfolio$H89551, Market$MarketReturn)

FA$uniquenesses
sorted_uniqueness <- sort(FA$uniquenesses, decreasing = TRUE)



lowest_5 <- head(sorted_uniqueness, 5)

highest_5 <- tail(sorted_uniqueness, 5)

lowest_5
highest_5
```