The R data HW08.RData has 2 time series objects, FF5 and Rt, along with 3 string vectors, syb, industry, by_industry and a numerical vector Ns .

The Daily Return Data Rt

Stock Returns Rt contains daily returns (in %) of 27 stocks from January 1, 2016 to August 31, 2024. As usual, syb is the ticker symbols of the 27 stocks from the following 4 industries.

Food (1–7): Conagra Brands (CAG), Campbell Soup (CPB), Danone SA (DANOY), General Mills (GIS), Hormel Foods (HRL), Kellogg (K), McCormick & Co. (MKC). These 7 companies are in the "Packaged Foods" Industry of the "Consumer Defensive" sector.

Entertainment (8-14): AMC Entertainment (AMC), Liberty Global (LBTYA), Lions Gate Entertainment (LGF-A), Live Nation Entertainment (LYV), Netflix (NFLX), Paramount (PARA), Warner Brothers Discovery (WBD). These 7 companies are in the "Entertainment" Industry of the"Communication Services" Sector.

Health Care (15–20): Centene Co. (CNC), Humana Inc. (HUM), Pediatrix Medical Group (MD), Molina Healthcare (MOH), Tenet Healthcare (THC), UnitedHealth Group (UNH). These 6 companies are in the " Healthcare Plans" Industry of the "Healthcare" Sector.

Tech (21–28): Autodesk (ADSK), Advanced Micro Devices (AMD), CPS Technologies (CPSH), DXC Technology (DXC), Etsy, Inc. (ETSY), Microsoft (MSFT), Western Digital (WDC). These 7 tech companies are the " Technology" Sector.

Fama-French 5 factors FF5 is downloaded from Kenneth French's website. It is daily data of the same period as Rt with excess return of market portfolio in % (Mkt.RF), small firm effect variable (SMB), value premium variable (HML), robustness in profitability variable (RMW), aggressiveness in investment variable(CMA) and risk-free rate in % (RF)

industry – A length-4 string vector of the 4 industries.

by_industry – A length-27 string vector of indicating which industry asset j belongs to, $j = 1, \ldots, 27$.

The R **Functions**

The file FM_Functions.R contains several R functions for making plots that required for this homework. You will need to source the file:

```
> source("/path_you_save_the_file/FM_Functions.R")
```

The file FM_Tests.R contain several R functions for testing the models and covariance matrices. Similarly, you need to source the file before using them,

```
> source("/path_you_save_the_file/FM_Tests.R")
```

Both files give brief instructions about how to use these functions. Please Read them. It will be convenient to "attach" the Fama French 5 factor data with the function `attach()`.

Homework Questions

All the models you are going to fit do NOT include industry/sector as a factor. When you give your answers, please always take this factor into account if applicable and explain any pattern or difference across industries.

1. Fit the Fama-French 3 factor model to the 27 asset daily return data.

   (a) Summarize the R Squared values and betas of the 27 assets by plotting them with the R function `coef.plot()` from `FM_Functions.R`.

   ```
   > coef.plot(R.Sq, betas)
   ```

   The first argument required the vector of R Squared values, the second argument is simply `coef(fit)[-1,]`, where `fit` is the returned object of `lm()`.

   Describe what you see from the R-Squared and the betas of the excess market returns. How many of them are considered "aggressive" asset? Give the numbers for each industry and comment on your findings.

   (b) Identify those individual assets that do not follow the FF-3-factor model and their sectors.

   (c) Test if the FF-3-factor model holds for ALL the 27 assets. Use both the Wald and the LRT.

   (d) Test if FF-3-factor model holds for each industry by testing the assets from the same industry. You should have 4 tests. Please give both the Wald and the LRT tests.

   (e) In estimating the covariance matrix of the asset returns, the reduction in the number of parameters is substantial with the FF-3-factor model when $N = 27$. First, give the numbers of parameters of the sample covariance and that of the model based.

   The model based estimation relies on the assumption that the covariance ε_t is diagonal. Check this assumption with the graph and individual correlation tests by applying the `resid.summary()` function from `FM_Functions.R`,

   ```
   > resid.summary(res)
   ```

   The argument `res` is simply the residuals from the regression `resid(fit)`. The function plots the heatmap of the residual covariance and the number of significant pairs. Carefully describe what you see from the graph and its implications.

   (f) Give formal structure tests on $\mathbf{\Sigma}_\varepsilon$, the variance matrix of errors ε_t. Test if the matrix is diagonal (independence test) as well as if the matrix is block-diagonal (independence among industries).

2. Apply the factor analysis to the return data. For the sake of comparison, use 3 factors. Compute the factors, R-squared and residuals. The residual $\hat{\varepsilon}_t$ is calculated by the formula (8.18) for each t. For the data of n observations, the residual matrix can be obtained by

$$\underset{n\times N}{\hat{\mathbf{E}}} = \underset{n\times N}{\tilde{\mathbf{Z}}} - \underset{n\times 3}{\mathbf{F}}\,\underset{3\times N}{\hat{\mathbf{B}}}$$

where $\tilde{\mathbf{Z}}$ is standardized then de-meaned excess returns (See the R code below on page 207). Please do not display the residuals and R Squared vector.

(a) Examine the loading estimate $\hat{\mathbf{B}}$, can you find interpretation about these coefficients?

(b) Check the covariance structure of the errors with the `resid.summary()` function. Describe your findings.

3. Apply the principal component analysis to the return data. Please use $\tilde{\mathbf{Z}}$.

(a) Plot the scree plot. If `pca` is your returned object from `prcomp()`, the plot can be gotten by

```
> plot(pca)
```

According to the plot how many principal components would you choose?

(b) Approximate the factor model with $p = 3$ (only for the purpose of comparison, this is not necessary the best choice). Compute the estimate factors and R Squared. Please do not display them.

(c) Examine the wights of principal components, these are the estimates of loadings, can you find interpretations form them?

(d) Check the covariance structure of the errors with the `resid.summary()` function. Describe your findings.

4. (a) Plot the R Squared values of the 3 models using the `RSq.plot()` function from `FM_functions.R`,

```
> RSq.plot(RSq.all)
```

There is only one argument which is a $N \times 3$ matrix of all R squared values from 3 models. You can use `cbind()` to combine the three vectors. Please use the default ordering, (1) Fama-French model; (2) factor analysis; then (3) principal components.

(b) Compare and comment the three approaches in pricing modeling from all the analysis you have done.