Portfolio Creation and Stock analysis via Capital Asset Pricing Model and Fama-French 3 Factor Model

April 2021

Hayden McAlpin

Abstract:

The goal of this project is to utilize the Capital Asset Pricing Model (CAPM) and the Fama-French 3 Factor Model (FF3) to analyze a set of stocks and determine which ones would be investment worthy. Once the stocks have been assessed and chosen, they will be placed in a portfolio with a $100,000 investment amount and the overall estimated annual return will be calculated. These two models will provide an opportunity to see how stocks work and evaluate several factors in formulating a strong portfolio. R will be utilized to use the models properly and make any calculations or analytical analysis.

Data:

The data for the following stocks was downloaded from finance.yahoo.com. The assortment of stocks was initially chosen from a Kaggle data set but to get the most up to date numbers the data was extracted from Yahoo Finance. The data covers the five year range of 2016-2021. This will give an ample amount of data to determine how the stocks have been performing and it gives the models more information to work with. This will potentially give a more accurate assessment of each stock's risk factors and investment worthy qualities. The 24 stocks are from 4 different sectors in hopes that it will lead to a diverse portfolio.

Note: four stocks will be used to formulate the portfolio (One from each sector).

Aviation	Finance	Pharma/Healthcare	Technology
AAL – American Airlines Group Inc	BCS – Barclays PLC	BHC – Bausch Health Companies Inc	AAPL – Apple Inc
ALGT – Allegiant Travel Company	CS – Credit Suisse Group AG	JNJ – Johnson & Johnson	AMZN – Amazon.com, Inc
ALK – Alaska Air Group, Inc	DB – Deutsche Bank Aktiengesellschaft	MRK– Merck & Co., Inc	FB – Facebook, Inc
DAL – Delta Air Lines, Inc	GS – The Goldman Sachs Group, Inc	PFE – Pfizer Inc	GOOG – Alphabet, Inc
HA – Hawaiian Holdings, Inc	MS – Morgan Stanley	RHHBY – Roche Holding AG	IBM – International Business Machines Corporation

LUV – Southwest Airlines Co	WFC – Wells Fargo & Company	UNH – UnitedHealth Group Incorporated	MSFT – Microsoft Corporation

The stock data includes the following variables:

Date, Open Price, High Price, Low Price, Close Price, Adjusted Close Price and Volume

In order to assess how the market as a whole performed during the same time period, the data for the S&P 500 Index will be used. This data has the same variables as the stocks data and was also downloaded from Yahoo Finance. The adjusted close price will be the variable used to make any predictions and calculations.

The data needed to utilize the Fama-French 3 Factor Model was downloaded from Dartmouth.edu as it has a massive data library for their various models dating back to the early-mid 1900's. This data file also has the risk-free rate of return used in both models. The variables for the Fama-French model in this file are SMB and HML (more on these in the models section).

Exploring Stocks

Appendix A shows time series plots for the prices of stocks and normalized prices of the stocks. These allow the viewer to compare the trends in the prices of the stocks in the same sector. Since some of the stocks have a much higher price point than others in their sector i.e. Allegiant Travel Company, Amazon, and Google, the prices were normalized using the formula:

$$Normalized\ Price = \frac{Price - Minimum}{Maximum - Minimum}$$

This allows the stocks to be compared in a clearer plot.

In each graph there is a massive dive in price in the middle-late months of 2020. Due to the COVID-19 pandemic this is to be expected. Most of the stocks have made their claim back to their prices before the fall and have been trending upwards.

Models:

The first model used to evaluate the stocks' performance is the Capital Asset Pricing Model. It has the Time Series equation:

$$\left(R_i - r_f\right)_t = \alpha_i + \beta_i\left(R_m - r_f\right)_t + \epsilon_{it}$$

And its equation for estimated stock return is:

$$ER_i = r_f + \beta_i\left[ER_m - r_f\right]$$

Where:

$$\left(R_i - r_f\right)_i = Excess\ return\ of\ stock\ i\ at\ time\ t$$

$$\alpha_i = risk\ adjustment\ for\ stock\ i$$

$$\beta_i = beta\ of\ investment$$

$$\left(R_m - r_f\right)_t = Excess\ market\ return$$

$$\epsilon_{it} = \ standard\ error\ of\ stock\ i\ at\ time\ t$$

The CAPM focuses on comparing the stocks rate of return to the return of the market as a whole. Its risk assessment is that of systemic risk. Once the alpha and beta from the time series regression are calculated the stock can be evaluated in the following ways:

Alpha:

$$\alpha < 0\ indicates\ the\ stock\ didn't\ perform\ well\ enough\ for\ its\ risk$$

$$\alpha = 0\ indicates\ the\ stock\ performed\ evenly\ for\ its\ risk$$

$$\alpha > 0\ indicated\ the\ stock\ performed\ better\ than\ expected\ for\ its\ risk$$

Beta:

$$\beta < 1\ indicates\ the\ stock\ is\ less\ volatile\ than\ the\ market$$

$$\beta = 1\ indicates\ the\ stock\ is\ equally\ volatile\ as\ the\ market$$

$$\beta > 1\ indicates\ the\ stock\ is\ more\ volatile\ than\ the\ market$$

The goal is to maximize the beta value for a riskier investment, which is what this project will be focusing on.

The second model used is the Fama-French 3 Factor Model. This builds off the CAPM and adds two factors to consider. SMB (Small Minus Big) is the size premium which focuses on the size of the firm. It takes into consideration the publicly traded stocks that generate higher returns and have small market caps. HML (High Minus Low) is the value premium and focuses on the book to market values. This takes into account the stocks that give higher returns than the market that also have high book to market ratios. The time series equation is:

$$\left(R_i - r_f\right)_t = \ \alpha_i + \beta_1\left(R_M - r_f\right)_t + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_{it}$$

And formula for expected return is:

$$ER_i = r_f + \beta_1(ER_M - r_f) + \beta_2 ESMB + \beta_3 EHML$$

Where:

$$SMB = Size\ Premium\ (small\ minus\ big)$$

$$HML = value\ premium\ (high\ minus\ low)$$

$$(and\ the\ other\ variables\ follow\ the\ CAPM)$$

Alpha and Beta1 are evaluated in the same way as that of the CAPM and Beta2, Beta3 are evaluated with respect to their corresponding variables in the similar way Beta1 is with the market.

Process:

All price plots were made using the ggplot2 package in R.

The following is the function used to normalize the prices in R:

```
#Function to normalize data
normal <- function(df) {
  minimum = vector("numeric", length(df))
  maximum = vector("numeric", length(df))

  minimum[1] = 0
  maximum[1] = 0

  for(i in 2:length(df)) {
    minimum[i] <- min(df[,i])
    maximum[i] <- max(df[,i])
  }

  for(j in 2:length(df)) {
    for(i in 1:length(df[,1])){
      df[i,j] <- (df[i,j] - minimum[j])/(maximum[j]-minimum[j])
    }
  }
  return(df)
}
```

In order to make the calculations for the CAPM, the stocks were evaluated using simple linear regression in R. The following snapshot is the function used to create the model and return the alpha and beta values for the stocks:

```r
#Function for CAPM model coefficients alpha and beta
CAPM.model <- function(df) {
  alpha <- list(length(df)-3)
  beta <- list(length(df)-3)

  for(i in 2:(length(df)-2)) {
    model <- lm(df[,i] ~ SP500, data = df)
    alpha[i-1] <- coefficients(model)[1]
    beta[i-1] <- coefficients(model)[2]
  }

  ab <- data.frame()
  for(i in 1:length(alpha)) {
    ab[1, i] <- alpha[i]
    ab[2, i] <- beta[i]
    names(ab)[i] <- names(df)[i+1]
  }

  row.names(ab) <- c("alpha", "beta")
  return(ab)
}
```

Once the alpha and beta values were calculated, the stock's expected return could be calculated using the formula mentioned previously. This value is then the daily return rate as a decimal. Therefore in order to get the annualized return the values were multiplied my 25200 (252 for annual and 100 to convert to a percentage).

To get the alpha and beta values for the FF3 model, multiple linear regression was used. The following snapshot shows the function used to create the model and return the corresponding alpha and beta values for each stock:

```r
#Function for Fama-French 3 factor model coefficients alpha, beta1, beta2, beta3
FF.model <- function(df) {
  alpha <- list(length(df)-6)
  beta1 <- list(length(df)-6)
  beta2 <- list(length(df)-6)
  beta3 <- list(length(df)-6)

  for(i in 2:(length(df)-5)) {
    model <- lm(df[,i] ~ SP500 + SMB + HML, data = df)
    alpha[i-1] <- coefficients(model)[1]
    beta1[i-1] <- coefficients(model)[2]
    beta2[i-1] <- coefficients(model)[3]
    beta3[i-1] <- coefficients(model)[4]
  }

  ab <- data.frame()
  for(i in 1:length(alpha)) {
    ab[1, i] <- alpha[i]
    ab[2, i] <- beta1[i]
    ab[3, i] <- beta2[i]
    ab[4, i] <- beta3[i]
    names(ab)[i] <- names(df)[i+1]
  }

  row.names(ab) <- c("alpha", "beta1", "beta2", "beta3")
  return(ab)
}
```

Once these values were calculated, estimated return and annualized return was calculated in the similar fashion in that of the CAPM but with the formula for the FF3 as mentioned earlier.

Results:

The following show tables for the stocks in the same sector and their associated results from the CAPM. The Sharpe ratio is used to evaluate the risk of the stock using the formula:

$$\frac{ER - r_f}{\sigma}$$

Also note that the estimated of return of the market is: 14.57%. So, it would be ideal to select stocks above this threshold.

Aviation

	AAL	ALGT	ALK	DAL	HA	LUV
Alpha	-0.0061	0.00004	-0.00045	-0.00024	-0.00052	-0.00009
Beta	1.37856	1.31029	1.31249	1.27407	1.31570	1.06132
ER (annualized)	19.68%	19.57%	18.79%	18.27%	18.83%	15.40%
Sharpe Ratio	2.04233	2.44785	2.51621	2.48608	1.89080	2.43941

Here the selection was simple as the other 5 stocks had an alpha value less than zero indicating they didn't perform well enough for their risk. Therefore, ALGT is the clear choice.

Finance

	BCS	CS	DB	GS	MS	WFC
Alpha	-0.00023	-0.00027	-0.00051	0.00016	0.00048	-0.00056
Beta	1.25253	1.30360	1.24257	1.26841	1.42544	1.25281
ER (annualized)	17.98%	18.67%	17.85%	18.20%	20.31%	17.98%
Sharpe Ratio	2.5773	3.11924	2.44121	3.39880	3.53010	3.12578

Similarly, 4 stocks were eliminated due to their alpha values. MS had a higher return and with only a slightly higher beta value and Sharpe ratio, the slightly higher risk is worth the extra return.

Pharma/Healthcare

	BHC	JNJ	MRK	PFE	RHHBY	UNH
Alpha	-0.00061	0.00009	0.00002	-0.0010	0.00003	0.00041
Beta	1.36440	0.66323	0.69578	0.70034	0.62717	1.03676
ER (annualized)	19.49%	10.01%	10.47%	10.53%	9.54%	15.07%
Sharpe Ratio	1.90094	2.81990	2.66936	2.68080	2.45259	3.09082

BHC and PFE are immediately eliminated due to their negative alpha values. Since UNH is the only stock with a higher return than that of the market, it is the clear choice.

Technology

	AAPL	AMZN	FB	GOOG	IBM	MSFT
Alpha	0.00078	0.00097	0.00028	0.00035	-0.00043	0.00072
Beta	1.17531	0.92960	1.06679	1.06561	0.97502	1.19271
ER (annualized)	16.94%	13.62%	15.47%	15.46%	14.24%	17.17%
Sharpe Ratio	3.30430	2.67926	2.81415	3.39801	3.17478	3.67111

IBM was eliminated due to its alpha value and Amazon had a low beta and a return lower than the market. Of the remaining stocks MSFT had the highest return. Even though it has a slightly higher risk, it was chosen for the portfolio.

The first portfolio considered is one where the investment account is evenly distributed among the four stocks. Then the return follows:

$$0.25 * ER_{ALGT} * 100000 + 0.25 * ER_{MS} * 100000 + 0.25 * ER_{UNH} * 100000$$
$$+ 0.25 * ER_{MSFT} * 100000$$
$$= \$18{,}030$$

The second portfolio will distribute the funds proportionally based on the stocks' beta values. Then the payout follows:

$$0.2639 * ER_{ALGT} * 100000 + 0.2871 * ER_{MS} * 100000 + 0.2088 * ER_{UNH} * 100000$$
$$+ 0.2402 * ER_{MSFT} * 100000$$
$$= \$18{,}266.31$$

The additional risk in putting more money in the accounts with higher betas isn't worth the additional $200 a year in return. Therefore, it would be best to just invest an equal amount in each stock according to the CAPM.

The following show the results for the Fama-French model:

Aviation

	AAL	ALGT	ALK	DAL	HA	LUV
Alpha	-0.00019	0.00025	-0.00016	0.00007	-0.00011	0.00014
Beta1	1.20048	1.25032	1.17592	1.13591	1.11557	0.96628
Beta2	0.56322	0.69735	0.55129	0.49478	0.88722	0.28079
Beta3	1.44006	0.86713	1.06570	1.09847	1.53595	0.77499
ER (annualized)	9.1%	14.34%	11.39%	10.41%	8.46%	9.64%
Sharpe Ratio	0.88085	1.75534	1.46468	1.34911	0.78612	1.45756

AAL, ALK, and HA have negative alpha values and are therefore removed from consideration. DAL and LUV have far lower returns that the market so ALGT is chosen again.

Finance

	BCS	CS	DB	GS	MS	WFC
Alpha	0.00012	-0.00005	-0.00027	0.00041	0.00076	-0.00017
Beta1	1.11300	1.20331	1.12427	1.17642	1.33236	1.12242
Beta2	0.37249	0.36252	0.51251	0.11718	0.03231	0.04205
Beta3	1.15069	0.79634	0.91184	0.80021	0.83779	1.17450
ER (annualized)	9.28%	12.98%	11.64%	11.69%	13.22%	8.03%
Sharpe Ratio	1.25065	2.11066	1.53809	2.10782	2.22821	1.28508

CS, DB, and WFC are again eliminated due to their negative alpha values. MS has the highest expected return and is therefore selected.

Pharma/Healthcare

	BHC	JNJ	MRK	PFE	RHHBY	UNH
Alpha	0.00016	0.00012	0.00001	-0.00007	-0.00006	0.00042
Beta1	1.31092	0.67989	0.71169	0.70891	0.65790	1.04451
Beta2	1.31092	0.67989	0.71169	0.70891	0.65790	1.04451
Beta3	0.93983	-0.33647	-0.19721	-0.24593	-0.00621	-0.12338
ER (annualized)	20.91%	9.38%	10.54%	9.73%	11.91%	14.94%
Sharpe Ratio	2.04755	2.62051	2.68892	2.45503	3.13891	3.06223

PFE and RHHBY are eliminated from consideration due to negative alpha values. JNJ and MRK have too low of returns for what is ideal. BHC has a higher estimated return and a lower Sharpe ratio than that of UNH. Therefore, with les risk and a higher estimated return BHC is selected for the portfolio.

Technology

	AAPL	AMZN	FB	GOOG	IBM	MSFT
Alpha	0.00062	0.00069	0.00008	0.00021	-0.00034	0.00055
Beta1	1.23629	1.02202	1.13200	1.11221	0.94747	1.26031
Beta2	-0.11874	-0.03551	-0.01357	-0.03131	-0.03369	-0.17494
Beta3	-0.51708	-0.83069	-0.58983	-0.41441	0.26194	-0.55927
ER (annualized)	21.00%	20.65%	20.50%	18.92%	11.88%	21.42%
Sharpe Ratio	4.15105	4.17924	3.79619	4.21569	2.60598	4.64004

IBM is the only stock eliminated due to a negative alpha value. Of the remaining stocks, MSFT has the highest Sharpe ratio indicating more risk. AAPL and AMZN have the highest alpha values and AAPL has a slightly higher Beta1 and Expected Return. Therefore, AAPL will be chosen.

The first portfolio will again be one where the account is evenly distributed between the four stocks. The annual return will follow:

$$0.25 * ER_{ALGT} * 100000 + 0.25 * ER_{MS} * 100000 + 0.25 * ER_{BHC} * 100000$$
$$+ 0.25 * ER_{AAPL} * 100000$$
$$= \$17{,}367.50$$

The second portfolio will proportionally distribute the account into riskier stocks based on the stocks' Shape ratio. The annual return would then follow:

$$0.1724 * ER_{ALGT} * 100000 + 0.2188 * ER_{MS} * 100000 + 0.2011 * ER_{BHC} * 100000$$
$$+ 0.4077 * ER_{AAPL} * 100000$$
$$= \$18{,}131.22$$

In this situation, it would be worth the additional risk for a potential $800 increase in expected annual return. Therefore, based on the Fama-French results it would be better to invest more in the riskier stocks for a higher expected return.

Conclusion:

The two models used gave an opportunity to assess the performance of the stocks using different criteria. The ability of the Fama-French model to build off of the Capital Asset Pricing Model seems to give more accurate alpha and beta values to predict returns because there is more data to compare and asses. Once the individual model results were calculated and the portfolios were created it was more beneficial to invest more in the riskier stocks of the Fama-French model than it was in the CAPM. The was only a slight increase in additional return for the CAPM but it would take on more risk. The Fama-French model also had a significant drop in expected returns in comparison to the CAPM model. The negative correlation with the size and value premiums for some of the stocks must have contributed to that in a significant way. These

two models are some fundament techniques to assess stocks and build portfolios. It was a good stepping stone into the stock market and finding ways to track how well or poorly stocks perform. This experience can be built upon to learn new strategies and develop the best way to assess a portfolio's potential performance.

There were only two main struggles I had with the project. The first being just learning the Fama-French 3 Factor model and understanding what the additional variable represented and how they contributed to the model. The second was getting all the data to match up by date. The data from yahoo gave the date as a date data type however the Fama-French data listed it as an integer of the form YYYYMMDD. I then created the following function to change the values of the yahoo data to match the integer form:

```
55  #Function to alter Date in stock data files
56  DateConvertor <- function(df) {
57    df <- separate(data = df,col = "Date", into = c("Month", "Day", "Year"))
58    df <- as.data.frame(apply(df, 2, as.numeric))
59    df$Date <- df$Year*10000 + df$Month*100 + df$Day
60    return(data.frame(df))
61  }
```

This allowed me to line up the data easier and make things work smoothly.

All in all this was an interesting project to assess stock performance and I can hopefully apply some of what I learned to build my own portfolio when I have access to some investment funds.

Sources

Finance.Yahoo.com

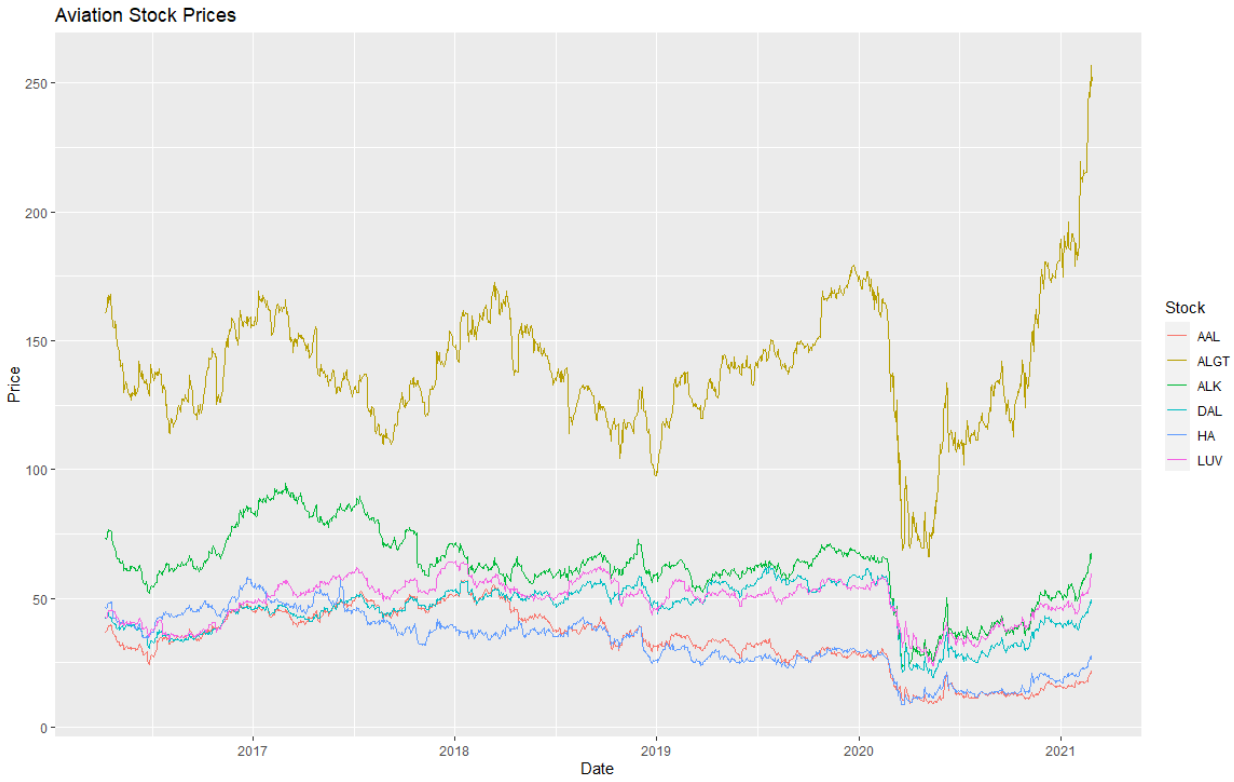
Hayes, A. (2020, September 16). Fama and French three factor model definition. Retrieved April 19, 2021, from https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp

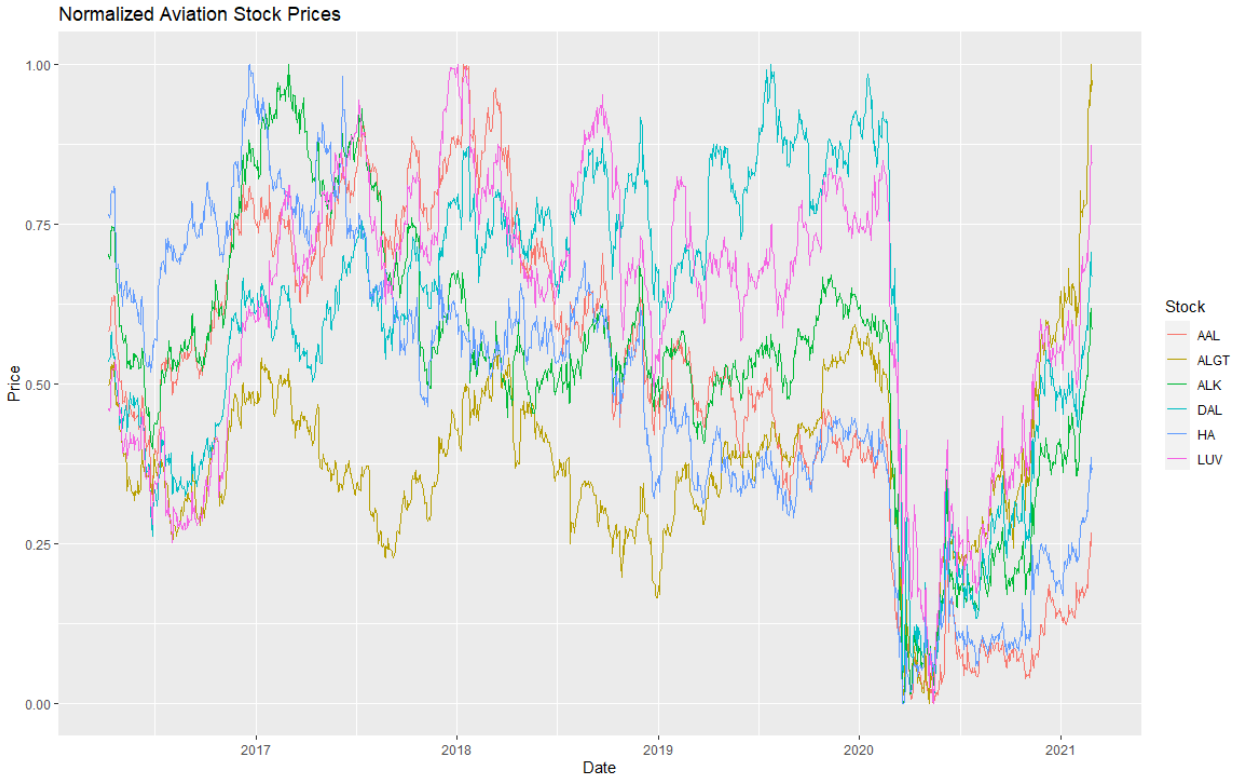
Investment Science 2nd Edition, David G. Luenberger Oxford University Press, 2013

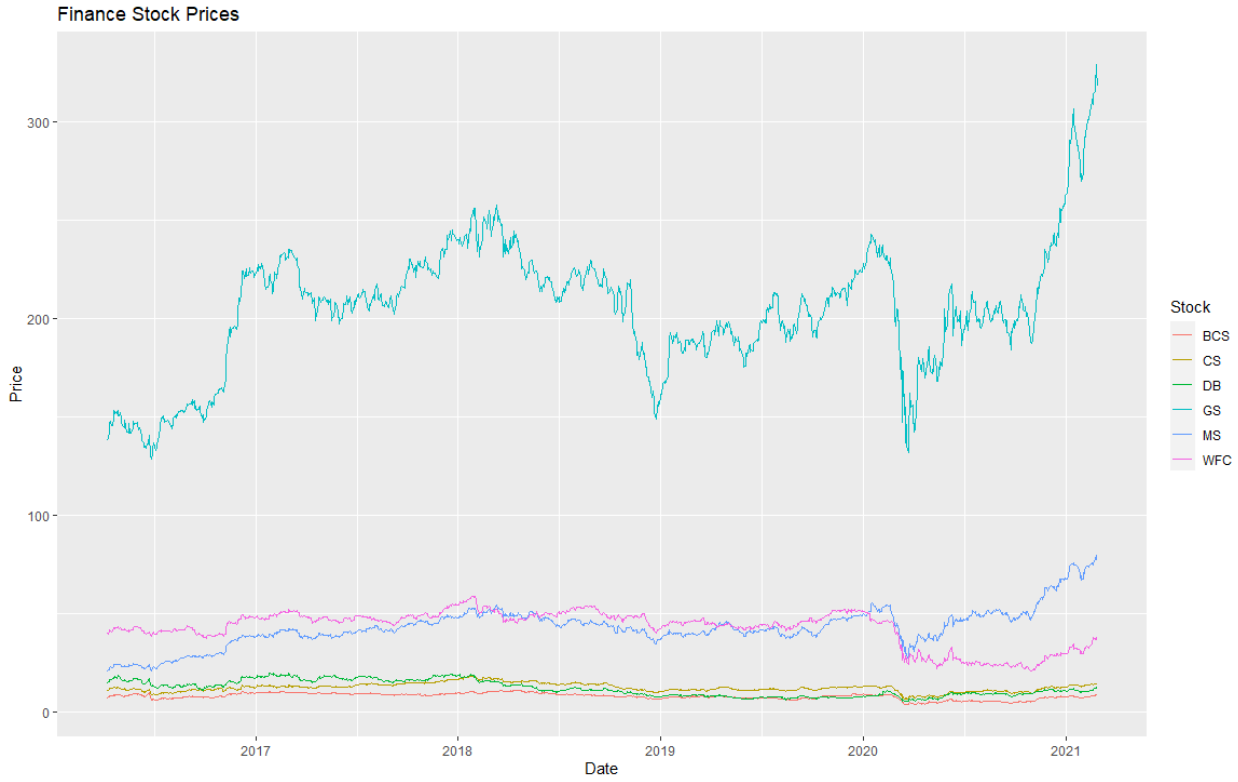
mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

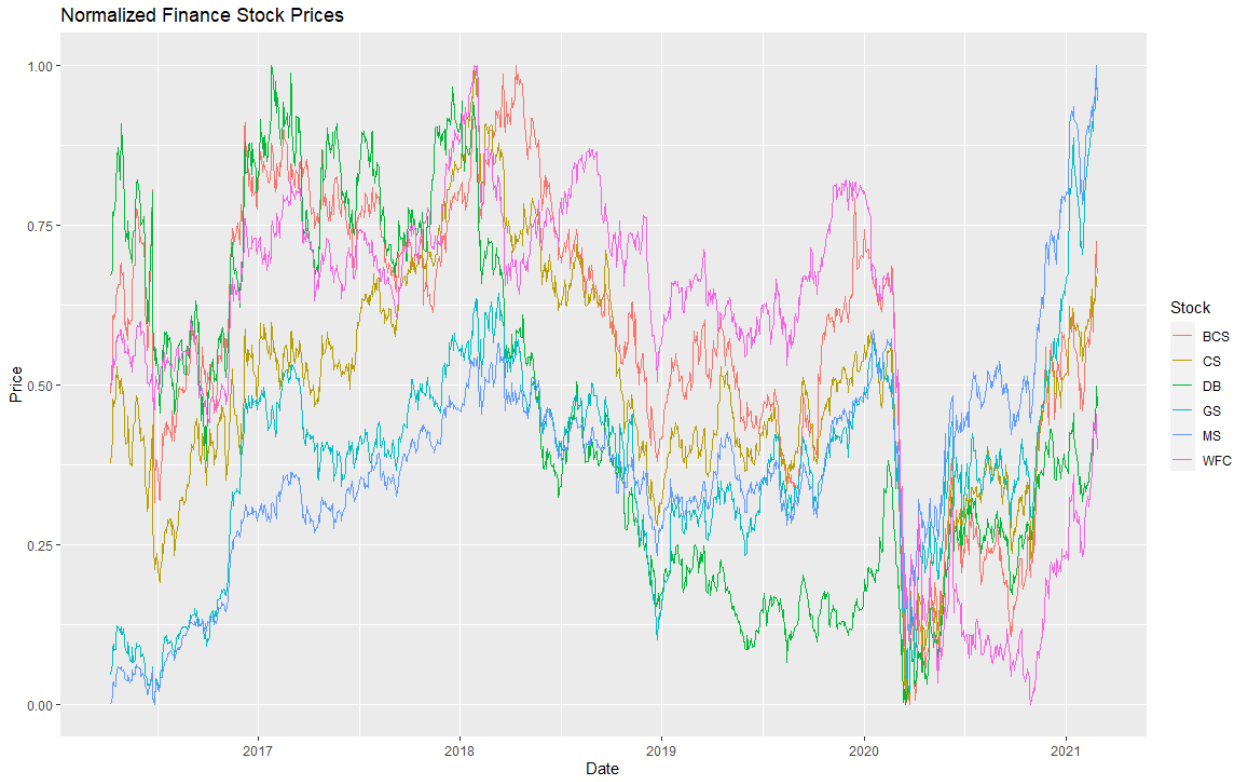
Hayes, A. (2020, September 16). Fama and French three factor model definition. Retrieved April 19, 2021, from https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp

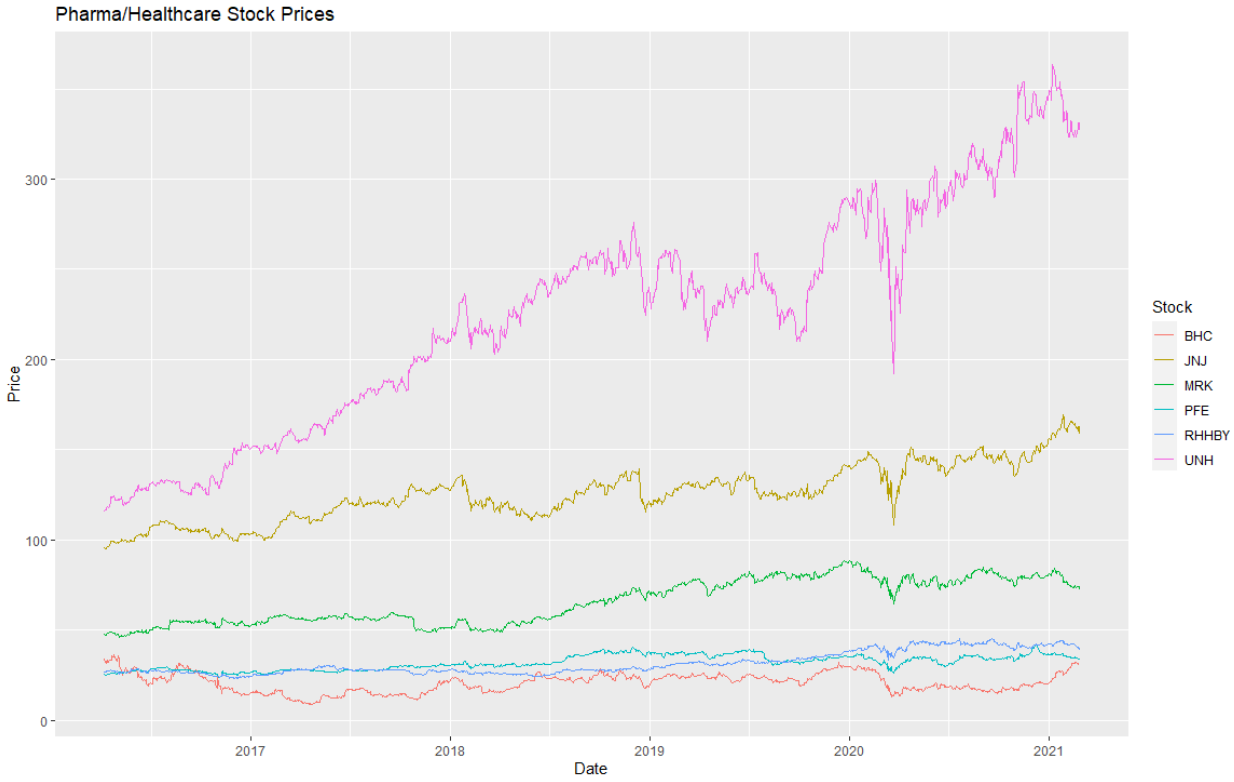
Appendix A

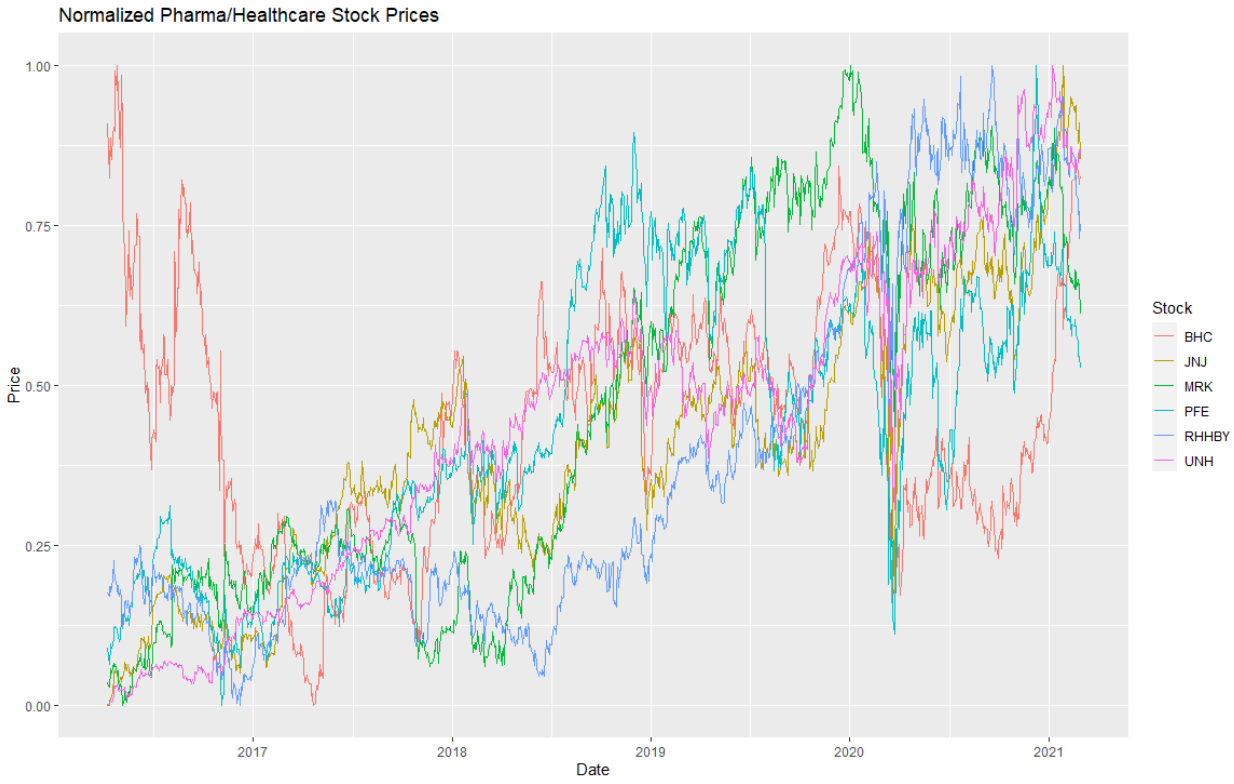

Aviation Stock Prices
Price
250
200
150
100
50
0
2017
2018
2019
2020
2021
Date
Stock
AAL
ALGT
ALK
DAL
HA
LUV

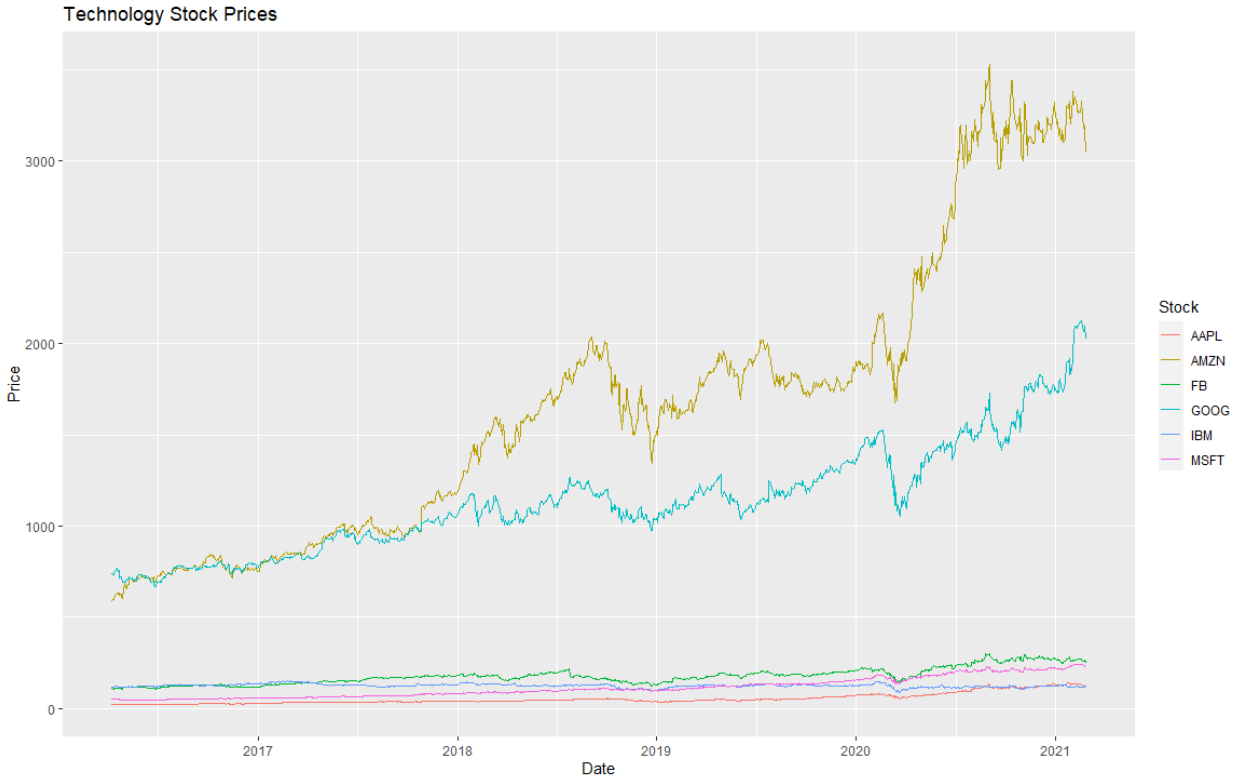

Normalized Aviation Stock Prices
Price
1.00
0.75
0.50
0.25
0.00
2017
2018
2019
2020
2021
Date
Stock
AAL
ALGT
ALK
DAL
HA
LUV

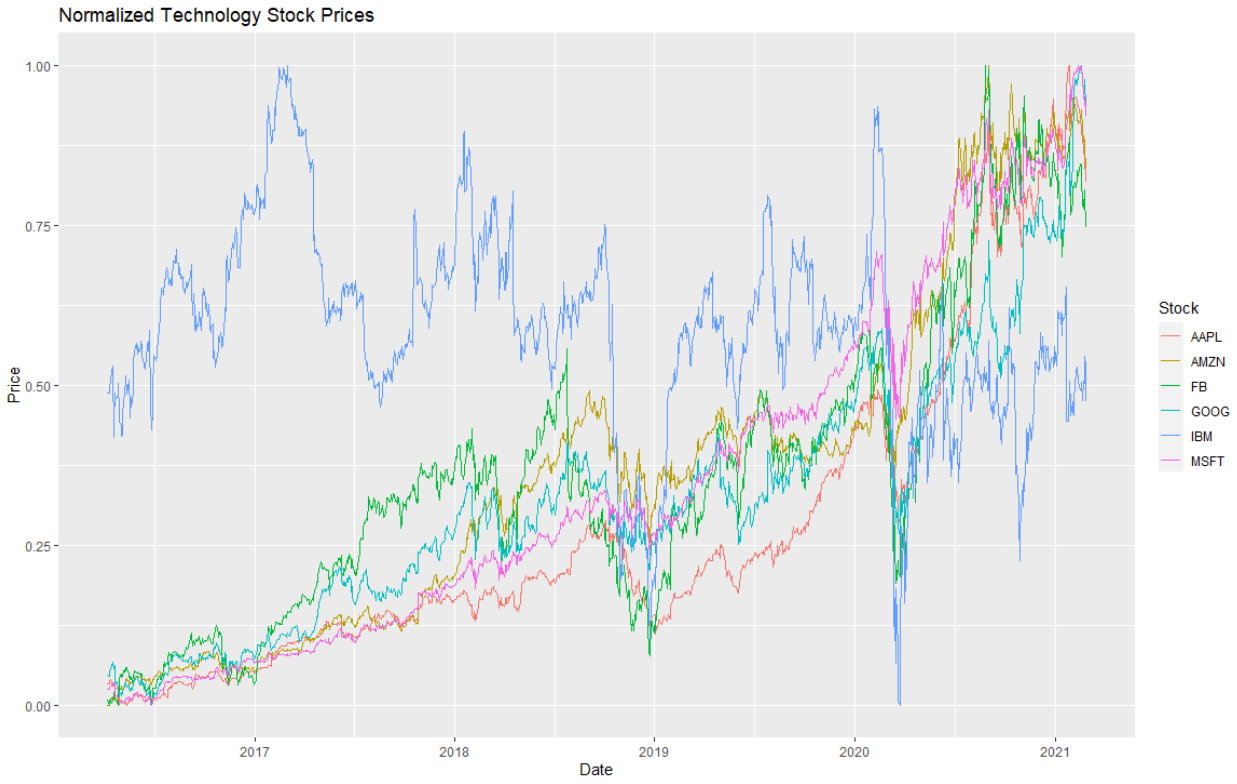

Finance Stock Prices
300
200
100
0
Price
2017
2018
2019
2020
2021
Date
Stock
BCS
CS
DB
GS
MS
WFC


Normalized Finance Stock Prices
1.00
0.75
0.50
0.25
0.00
Price
2017
2018
2019
2020
2021
Date
Stock
BCS
CS
DB
GS
MS
WFC


Pharma/Healthcare Stock Prices
Price
300
200
100
0
2017
2018
2019
2020
2021
Date
Stock
BHC
JNJ
MRK
PFE
RHHBY
UNH


Normalized Pharma/Healthcare Stock Prices
Price
1.00
0.75
0.50
0.25
0.00
2017
2018
2019
2020
2021
Date
Stock
BHC
JNJ
MRK
PFE
RHHBY
UNH


Technology Stock Prices
Price
3000
2000
1000
0
2017
2018
2019
2020
2021
Date
Stock
AAPL
AMZN
FB
GOOG
IBM
MSFT


Normalized Technology Stock Prices
Price
1.00
0.75
0.50
0.25
0.00
2017
2018
2019
2020
2021
Date
Stock
AAPL
AMZN
FB
GOOG
IBM
MSFT