Portfolio Management on NVDA, CMCSA and EA stocks

Paul Muriithi

```r
# loading required libraries
library(tidyquant)
library(tidyverse)
library(broom)
library(knitr)
library(kableExtra)
library(ggplot2)
library(plotly)
library(gridExtra)
```

1. Importing stocks data

```r
# Fetch selected stocks using Tiingo
stocks <- tq_get(c("NVDA", "CMCSA", "EA"), get = "stock.prices", from = "2000-01-01", to = "2022-06-18")
    select(symbol, date, adjusted)


# Display first 6 rows in a table
head(stocks, n = 6) %>%
    kable(caption = "Selected stocks data")
```

Table 1: Selected stocks data

symbol	date	adjusted
NVDA	2000-01-03	0.0894252
NVDA	2000-01-04	0.0870375
NVDA	2000-01-05	0.0841721
NVDA	2000-01-06	0.0786796
NVDA	2000-01-07	0.0799936
NVDA	2000-01-10	0.0826197

2. The Analysis

2.1. Plot prices over time.

```r
# time series plot for the three stocks
NVDA = stocks[stocks$symbol == "NVDA", ]
CMCSA = stocks[stocks$symbol == "CMCSA", ]
EA = stocks[stocks$symbol == "EA", ]
plot1 <- ggplot(NVDA, aes(x = date, y = adjusted)) + geom_line(color = "darkblue") + labs(title = "NVDA
plot2 <- ggplot(CMCSA, aes(x = date, y = adjusted)) + geom_line(color = "orange") + labs(title = "CMCSA
```

```r
plot3 <- ggplot(EA, aes(x = date, y = adjusted)) + geom_line(color = "tomato") + labs(title = "EA prices
grid.arrange(plot1, plot2, plot3, ncol = 3)
```

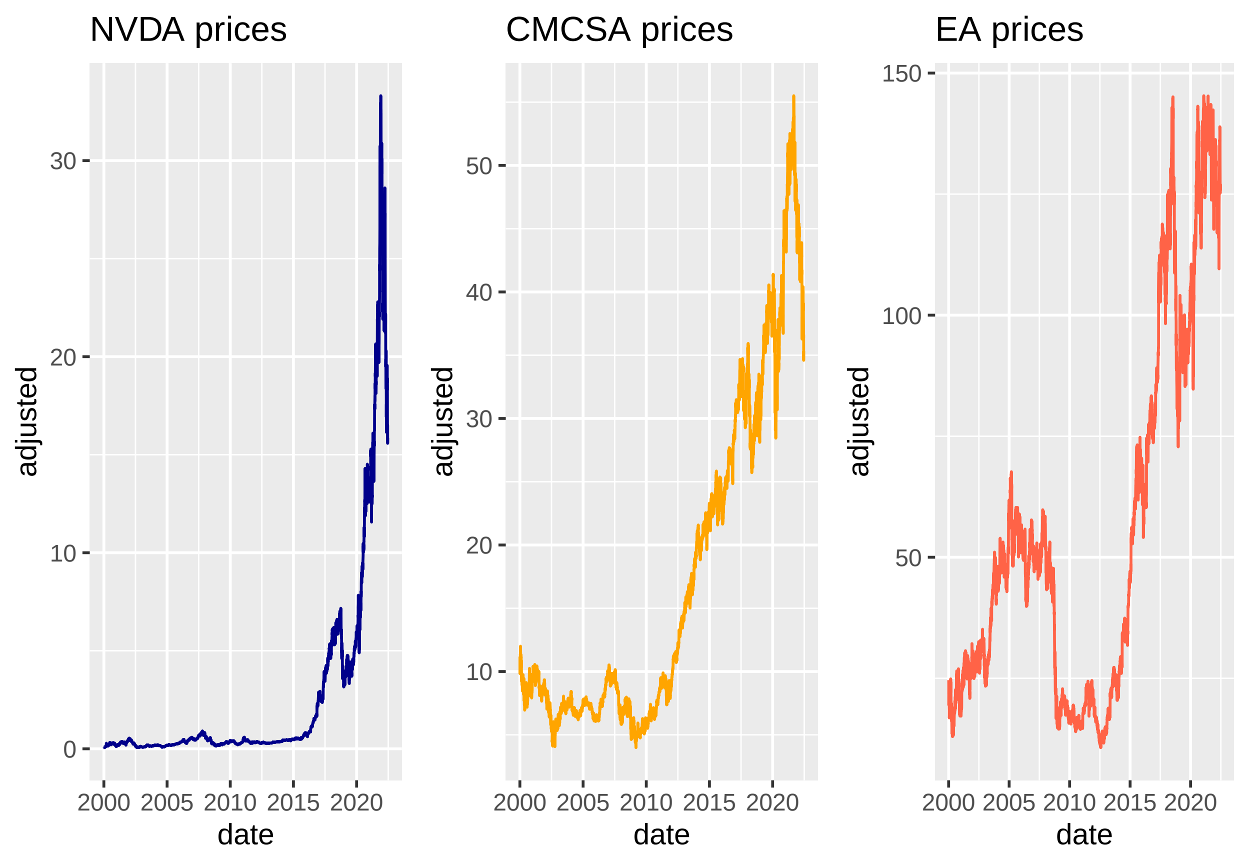

Figure 1: Time series of stocks prices

NVDA: The stock prices were very low and stationary from 2000 to around 2016 after when they increased rapidly with cyclic patterns. The prices were at their maximum in 2021, but could not maintain that price where its followed by a decreasing trend.

CMCSA: The stock prices were low from 2000 to 2010. Since 2010, the prices have been on sharp increasing trend with seasonal properties. This stock reaches its peak in 2021, falling immediately thereafter.

EA: EA stock prices started in 2000 with an upward trend where they rose for 5 yrs to 2005, then started to depreciate. At around 2008, EA stock prices fell sharply to where they were in 2000 and remained so with some seasonal rise and falls up to mid-2012. The prices then skyrocketed, reaching their peak in 2018 where they immediately falls again in 2019 and 2020, but rises in 2021.

2.2.

Given the formula:

$$r_t = 100 * \ln\left(\frac{P_t}{P_{t-1}}\right)$$

```r
# percentage returns for each stock
stocks <- stocks %>%
    group_by(symbol) %>%
```

```r
    mutate(per_returns = 100 * (log(adjusted) - log(lag(adjusted)))) %>%
    ungroup() %>%
    drop_na()

# subset percentage returns for each stock
NVDA_RT = stocks[stocks$symbol == "NVDA", ]
CMCSA_RT = stocks[stocks$symbol == "CMCSA", ]
EA_RT = stocks[stocks$symbol == "EA", ]
# rendering plots fro the percentage returns
plt1 <- ggplot(NVDA_RT, aes(x = date, y = per_returns)) + geom_line(color = "darkblue") + labs(title =

plt2 <- ggplot(CMCSA_RT, aes(x = date, y = per_returns)) + geom_line(color = "orange") + labs(title = "(
plt3 <- ggplot(EA_RT, aes(x = date, y = per_returns)) + geom_line(color = "tomato") + labs(title = "EA D
plts <- grid.arrange(plt1, plt2, plt3, ncol = 2)
```

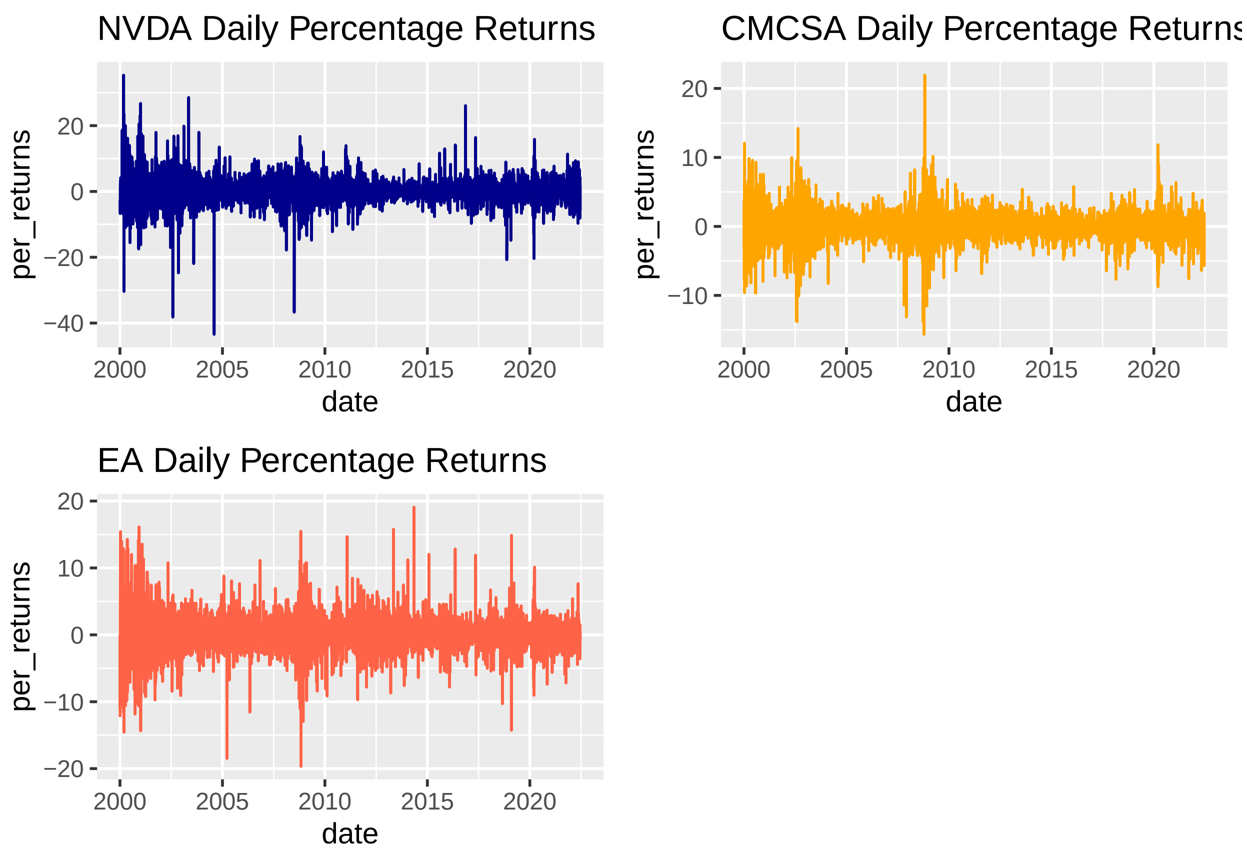

Figure 2: Time series of percentage returns

```r
plts
```

```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (2-2,1-1) arrange gtable[layout]
```

2.3

To find the number of bins, consider the following formula:

$$bins = (Max - Min)/h$$

where

$$h = 2 \times IQR \times n^{-1/3}$$

```
# histogram for each stock return series.
hist1 <- ggplot(NVDA_RT, aes(x = per_returns)) + geom_histogram(bins = 151, fill = "darkblue") +
    labs(title = "NVDA returns Histogram")
hist2 <- ggplot(CMCSA_RT, aes(x = per_returns)) + geom_histogram(bins = 96, fill = "orange") +
    labs(title = "CMCSA returns Histogram")
hist3 <- ggplot(EA_RT, aes(x = per_returns)) + geom_histogram(bins = 91, fill = "tomato") +
    labs(title = "EA returns Histogram")
hists <- grid.arrange(hist1, hist2, hist3, ncol = 2)
```

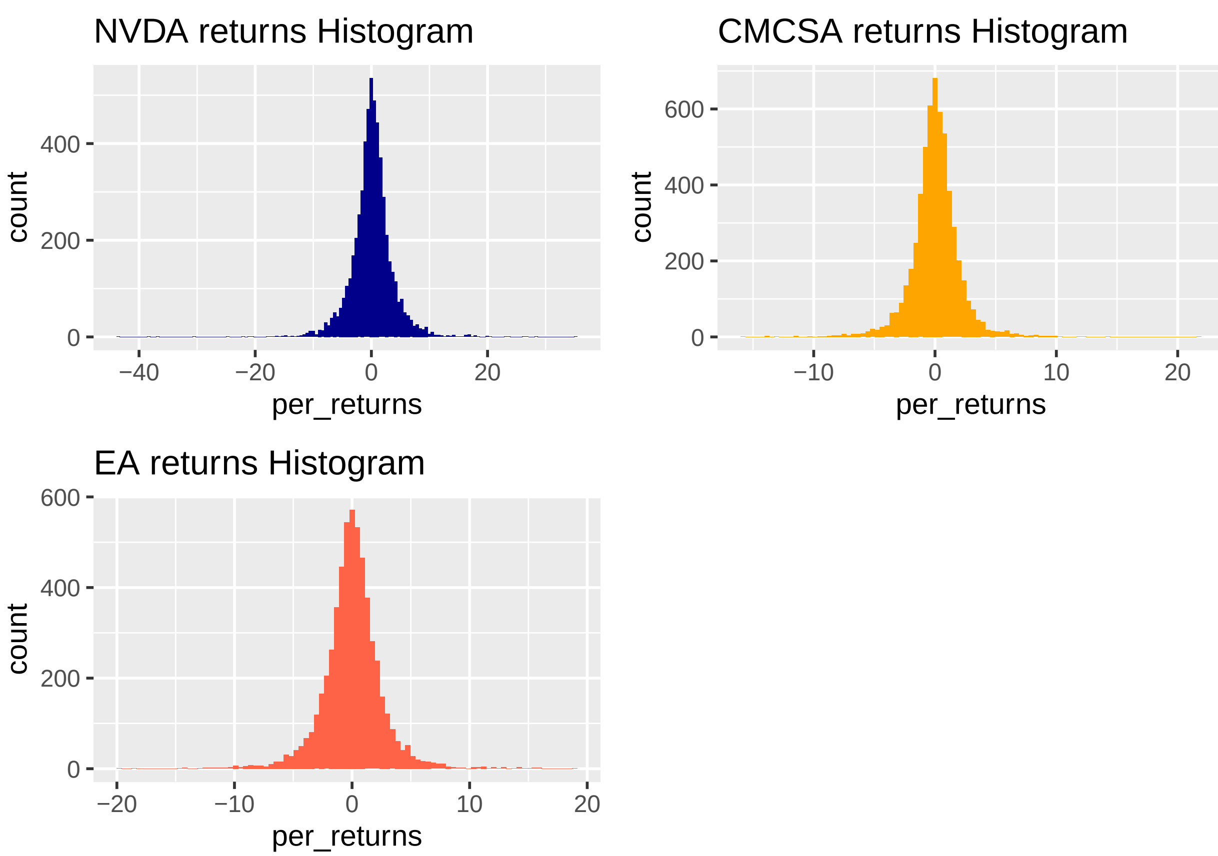

Figure 3: Histograms of percentage returns

```
hists
```

```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
```

```
## 3 3 (2-2,1-1) arrange gtable[layout]
```

The three histograms indicates that the returns of the three stocks(NDVA, CMCSA, and EA) are normally distributed with sharp apex indicating highly positive kurtosis, hence leptokurtic. In all three stocks, most of the stocks returns cluster around zero, since more returns are above the normal distribution and less values distributed in the tail.

2.4

```
# Summary for adjusted prices
tab1 <- stocks %>%
    group_by(symbol) %>%
    summarize(symbol = first(symbol), type = "adjusted", mean = mean(adjusted), median = median(adjuste
        variance = var(adjusted), sd = sd(adjusted), skewness = skewness(adjusted), kurtosis = kurtosis
        .groups = "drop")

tab2 <- stocks %>%
    group_by(symbol) %>%
    summarize(symbol = first(symbol), type = "returns", mean = mean(per_returns), median = median(per_r
        variance = var(per_returns), sd = sd(per_returns), skewness = skewness(per_returns),
        kurtosis = kurtosis(per_returns), .groups = "drop")

# Combine and display
summary <- bind_rows(tab1, tab2) %>%
    select(symbol, type, mean, median, variance, sd, skewness, kurtosis)

kable(summary, caption = "Table 2: Summary statistics table of returns", digits = 7)
```

Table 2: Table 2: Summary statistics table of returns

symbol	type	mean	median	variance	sd	skewness	kurtosis
CMCSA	adjusted	17.5230158	9.7002258	171.062476	13.079085	0.9754862	-0.2535478
EA	adjusted	55.2422375	46.9115067	1441.618254	37.968648	0.8662519	-0.4606879
NVDA	adjusted	2.6750335	0.3993262	29.762544	5.455506	2.9942825	9.1419178
CMCSA	returns	0.0207890	0.0341203	4.192188	2.047483	-0.0116406	8.2701357
EA	returns	0.0289469	0.0354923	6.749485	2.597977	0.1248245	6.4048987
NVDA	returns	0.0916293	0.1050498	14.276246	3.778392	-0.2535761	13.3769630

The average stock returns do not appear to be significantly different from zero(0.0204, 0.0289, and 0.0913) since more returns are above the normal distribution and lies between 2 and 4 standard deviations. NVDA stock returns the biggest return of 0.0913, with an almost 10% return on investment, but experiences the most variations.

2.5

Significant test for all stocks (t-test)

- Null Hypothesis: Average returns equals zero
- Alternative Hypothesis: Average returns not equals zero

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

$$\alpha = 0.01$$

$$t = (\frac{\hat{\mu}\sqrt{N}}{\sigma})t_{N-1}$$

Significant test for NVDA (t-test)

```
# select returns for NVDA stock
nvda_ret <- subset(stocks, stocks$symbol == "NVDA") %>%
    select(per_returns)
# t-test for NVDA stock
nvda_test = t.test(nvda_ret$per_returns, mu = 0, conf.level = 0.99)
res <- data.frame(nvda_test$statistic, nvda_test$p.value)
names(res) <- c("t", "p-value")
res
```

```
##          t    p-value
## t 1.823014 0.06835408
```

$$t = 1.837498$$

$$p - value = 0.0661 > \ 0.05$$

Significant test for CMCSA

```
# returns for CMCSA stock
cmcsa_ret <- subset(stocks, stocks$symbol == "CMCSA") %>%
    select(per_returns)
# t-test for CMCSA stock
cmcsa_test = t.test(cmcsa_ret$per_returns, mu = 0, conf.level = 0.99)
cmcsa_res <- data.frame(cmcsa_test$statistic, cmcsa_test$p.value)
names(cmcsa_res) <- c("t", "p-value")
cmcsa_res
```

```
##          t   p-value
## t 0.763267 0.4453361
```

$$t = 0.7872869$$

$$p - value = 0.4311469 > \ 0.05$$

Significant Test (t-test) for EA

```
ea_ret <- subset(stocks, stocks$symbol == "EA") %>%
    select(per_returns)  #select returns for EA stock
# t-test for EA stock
ea_test = t.test(ea_ret$per_returns, mu = 0, conf.level = 0.99)
ea_res <- data.frame(ea_test$statistic, ea_test$p.value)
names(ea_res) <- c("t", "p-value")
ea_res
```

```
##           t  p-value
## t 0.8375873 0.402298
```

$$t = 0.8405$$

$$p - value = 0.4007 > \ 0.05$$

```
# t critical value
qt(0.01, 5648, lower.tail = F)
```

```
## [1] 2.327008
```

$$t_{critical} = 2.327$$

- summary of the three tests

```
tests <- data.frame(Type = c("NVDA", "CMCSA", "EA"), T_stat = c(nvda_test$statistic, cmcsa_test$statisti
    ea_test$statistic), T_crit = rep(qt(0.01, df = nrow(nvda_ret) - 1, lower.tail = FALSE),
    3), P_value = c(nvda_test$p.value, cmcsa_test$p.value, ea_test$p.value))
tests
```

```
##    Type    T_stat   T_crit    P_value
## 1  NVDA 1.8230141 2.327008 0.06835408
## 2 CMCSA 0.7632670 2.327008 0.44533606
## 3    EA 0.8375873 2.327008 0.40229795
```

The t-values obtained for all the three stocks are less than the critical value(2.327), therefore we fail to reject the null hypothesis and conclude that the stocks returns are not significantly different from 0 at 1% significance level. The p-values being higher than 0.01 suggest that we do not have sufficient evidence to say that the average returns are different from 0.

2.6. Significance differences of the average returns.

Testing for equality of variances

$$Let\ Variance\ NVDA\ =\ \sigma_N^2$$
$$Let\ variance\ CMCSA\ =\ \sigma_C^2$$
$$Let\ variance\ EA\ =\ \sigma_E^2$$

* **1. NVDA and CMCSA**

$$H_0 : \sigma_N^2 = \sigma_C^2$$

$$H_1 : \sigma_N^2 \neq \sigma_C^2$$

$$\alpha = 0.01 \quad \frac{\alpha}{2} = 0.005$$

The test statistic

$$F = \left(\frac{\hat{\sigma}_N^2}{\hat{\sigma}_C^2}\right)$$

$$F = \frac{14.275}{4.187} = 3.409$$

$$F_c = F_{N_G-1,N_C-1} = F_{5649-1,5649-1}$$

F critical can be obtained using r as follows:

```r
qf(p = 0.005, df1 = 5649 - 1, df2 = 5649 - 1, lower.tail = FALSE)
```

```
## [1] 1.070963
```

$$F_c = 1.070963$$

$$F > \ F_c$$

Reject the null hypothesis. Therefore there the variances of NVDA and CMCSA stocks returns are significantly different from each other. T-test assuming unequal variances is appropriate in this case.

- **2. NVDA and EA**

$$H_0 : \sigma_N^2 = \sigma_E^2$$
$$H_1 : \sigma_N^2 \neq \sigma_E^2$$
$$\alpha = 0.01 \ \ then \ \frac{\alpha}{2} = 0.005$$

$$F = \left(\frac{\hat{\sigma}_N^2}{\hat{\sigma}_E^2}\right)$$
$$F = \frac{14.275}{6.752} = 2.114$$

$$F_c = 1.071$$
$$F > \ F_c$$

Hence reject the null hypothesis. Therefore NVDA and EA stocks variances are significantly different. We shall assume unequal variance when calculating t-value.

- **3. CMSA and EA**

$$H_0 : \sigma_C^2 = \sigma_E^2$$

$$H_1 : \sigma_C^2 \neq \sigma_E^2$$
$$\alpha = 0.01 \ \ then \ \ \frac{\alpha}{2} = 0.005$$

$$F = \left(\frac{\hat{\sigma}_C^2}{\hat{\sigma}_B^2}\right)$$
$$F = \frac{4.187}{6.752} = 0.062$$

$$F_c = 1.072$$

$$F < \ F_c$$

Fail to reject the null hypotheses This suggests that the variances returns of the two stocks returns does not differs significantly, hence t-test will assume equal variances.

$$Let\ \mu_N\ \ represent\ mean\ \ of\ \ NVDA\ \ stocks\ returns$$

$$let\ \mu_C\ \ =\ \ CMCSA\ average\ \ returns$$

$$let\ \mu_E\ =\ \ EA\ average\ returns$$

Perform t-test for the three pairs:

- **i) NVDA and CMCSA**

$$H_0 : \mu_N = \mu_C$$

$$H_1 : \mu_N \neq \mu_C$$

Consider that it's a two tailed t-test:

$$\alpha = 0.01\ then\ \frac{\alpha}{2} = 0.005$$

$$t = \frac{\hat{\mu}_N - \hat{\mu}_C}{\sqrt{\left(\frac{\hat{\sigma}_N^2}{N_N} + \frac{\hat{\sigma}_C^2}{N_C}\right)}}$$

computing using R:

```
NVDA_CMCS <- t.test(nvda_ret$per_returns, cmcsa_ret$per_returns, alternative = "two.sided",
    var.equal = F)
NVDA_CMCSA <- data.frame(NVDA_CMCS$statistic, NVDA_CMCS$p.value, NVDA_CMCS$parameter)
names(NVDA_CMCSA) <- c("t", "p-value", "df")
NVDA_CMCSA
```

```
##          t   p-value       df
## t 1.239161 0.2153192 8704.807
```

$$t = 1.2408$$

$$t_c = t_{df} = t_{8698.91}$$

Computing t critical using R:

```
qt(p = 0.005, df = 8698.91, lower.tail = FALSE)
```

```
## [1] 2.576395
```

$$t_c = 2.5764$$

$$|t|\ <\ |t_c|$$

In this case, fail to reject the null hypotheses and conclude that the average returns for NVDA and CMCSA are not significantly different.

- **ii) NVDA and EA**

$$H_0 : \mu_N = \mu_E$$

$$H_1 ; \mu_N \neq \mu_E$$

$$\alpha = 0.01 \; then \; \frac{\alpha}{2} = 0.005$$

$$t = \frac{\hat{\mu}_N - \hat{\mu}_E}{\sqrt{\left(\frac{\hat{\sigma}_N^2}{N_N} + \frac{\hat{\sigma}_E^2}{N_E}\right)}}$$

```
NVDA_E <- t.test(nvda_ret$per_returns, ea_ret$per_returns, alternative = "two.sided", var.equal = F)
NVDA_EA <- data.frame(NVDA_E$statistic, NVDA_E$p.value, NVDA_E$parameter)
names(NVDA_EA) <- c("t", "p-value", "df")
NVDA_EA
```

```
##         t   p-value       df
## t 1.02762 0.3041536 10016.41
```

$$t = 1.0378$$

$$t_c = t_{df} = t_{7455.11}$$

```
qt(p = 0.005, df = 10014.05, lower.tail = FALSE)
```

```
## [1] 2.57632
```

$$t_c = 2.5763$$

$$|t| \; < \; |t_c|$$

Fail to reject the null hypothesis and conclude that there is no significant difference between the average returns of NVDA and EA stocks.

- **iii) CMCSA and EA**

$$H_0 : \mu_H = \mu_N$$

$$H_1 ; \mu_H \neq \mu_N$$

$$\alpha = 0.01 \; then \; \frac{\alpha}{2} = 0.005$$

$$t = \frac{\hat{\mu}_C - \hat{\mu}_E}{\sqrt{\left(\frac{\hat{\sigma}_C^2}{N_C} + \frac{\hat{\sigma}_E^2}{N_E}\right)}}$$

Computing using R:

```
CMCSA_E <- t.test(cmcsa_ret$per_returns, ea_ret$per_returns, alternative = "two.sided", var.equal = T,
    )
CMCSA_EA <- data.frame(CMCSA_E$statistic, CMCSA_E$p.value, CMCSA_E$parameter)
names(CMCSA_EA) <- c("t", "p-value", "df")
CMCSA_EA
```

```
##            t   p-value    df
## t -0.1853959 0.8529218 11300
```

$$t = -0.1732$$

$$t_c = t_{df} = t_{11296}$$

```
qt(p = 0.005, df = 11296, lower.tail = T)
```

```
## [1] -2.576265
```

$$t_c = -2.5763$$

$$t \;>\; t_c$$

Reject the null hypotheses and conclude that the average returns for CMCSA and EA stocks return are significantly different.

From the three tests of significance, the average stocks returns for CMCSA and EA are significantly different but both average returns for EA and CMSA are not significantly different from NVDA

2.7

```
corr <- data.frame(NVDA = nvda_ret$per_returns, CMCSA = cmcsa_ret$per_returns, EA = ea_ret$per_returns)
kable(cor(corr, method = "pearson"), caption = "Correlation matrix of the three returns")
```

Table 3: Correlation matrix of the three returns

	NVDA	CMCSA	EA
NVDA	1.0000000	0.3256291	0.3754527
CMCSA	0.3256291	1.0000000	0.3030103
EA	0.3754527	0.3030103	1.0000000

According to the correlation matrix, the three stocks returns are not strongly correlated with each other. The correlation coefficients between NVDA and CMCSA and EA are 0.3251 and 0.3754 respectively, while CMSA and EA pair has a coefficient of 0.3030.

2.8. Testing the significance of correlations

Set up the null and alternative hypothesis:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Consider the the following test statistic:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

Decision rule: If $p - value > 0$ reject the H_0

```
cor1 <- cor.test(nvda_ret$per_returns, cmcsa_ret$per_returns, method = "pearson")  #test between NVDA a
cor2 <- cor.test(nvda_ret$per_returns, ea_ret$per_returns, method = "pearson")  #test between NVDA and 
cor3 <- cor.test(cmcsa_ret$per_returns, ea_ret$per_returns, method = "pearson")  # test between CMCSA a
```

```r
# add the outputs into a table
cor1 <- data.frame(statistic = cor1$statistic, p_value = round(cor1$p.value, 4))
row.names(cor1) <- c("NVDA_CMCSA")
cor2 <- data.frame(statistic = cor2$statistic, p_value = round(cor2$p.value, 4))
row.names(cor2) <- c("NVDA_EA")
cor3 <- data.frame(statistic = cor3$statistic, p_value = round(cor3$p.value, 4))
row.names(cor3) <- c("CMCSA_EA")
cor <- rbind(cor1, cor2, cor3)
kable(cor, caption = "Test for correlations significance")
```

Table 4: Test for correlations significance

	statistic	p_value
NVDA_CMCSA	25.88502	0
NVDA_EA	30.44637	0
CMCSA_EA	23.89770	0

Decision: Therefore p-values obtained equals to 0 which is less than $\alpha = 0.05$ significance level therefore reject the null hypotheses.

This indicates that the correlation between the stocks are significant.

2.9 Advising an investor.

Suppose that an investor has asked you to assist them in choosing **two** of these three stocks to include in their portfolio. The portfolio is defined by

$$r = w_1 r_1 + w_2 r_2$$

Where r_1 and r_2 represent the returns from the first and second stock, respectively, and w_1 and w_2 represent the proportion of the investment placed in each stock. The entire investment is allocated between the two stocks, so $w_1 + w_2 = 1$.

The investor favors the combination of stocks that provides the highest return, but dislikes risk. Thus the investor's happiness is a function of the portfolio, r:

$$h(r) = \mathbb{E}(r) - \mathbb{V}ar(r)$$

Where $\mathbb{E}(r)$is the expected return of the portfolio, and $\mathbb{V}ar(r)$ is the variance of the portfolio.[1]

First, we compute the covariance matrix of the returns.

```r
# Compute and display the covariance matrix
Nvda <- NVDA_RT$per_returns
Cmcsa <- CMCSA_RT$per_returns
Ea <- EA_RT$per_returns

pfol <- data.frame(Nvda, Cmcsa, Ea)
colnames(pfol) <- c("NVDA", "CMCSA", "EA")
rownames(pfol) <- NULL

kable(cov(pfol), caption = "Covariance Matrix of Returns")
```

[1]Note that $\mathbb{E}(r) = w_1 E(r_1) + w_2 \mathbb{E}(r_2)$, and $\mathbb{V}ar(r) = w_1^2 \mathbb{V}ar(r_1) + w_2^2 \mathbb{V}ar(r_2) + 2w_1 w_2 \mathbb{C}ov(r_1, r_2)$

Table 5: Covariance Matrix of Returns

	NVDA	CMCSA	EA
NVDA	14.27625	2.519130	3.685510
CMCSA	2.51913	4.192188	1.611807
EA	3.68551	1.611807	6.749485

To automate the optimization process, we define a function that takes two stock return vectors and returns the optimal weights that maximize $h(r)$.

```
# Function to compute optimal weights
optimize_portfolio <- function(mean1, mean2, var1, var2, covar) {
    h <- function(w1) {
        w2 <- 1 - w1
        expected_return <- w1 * mean1 + w2 * mean2
        variance <- w1^2 * var1 + w2^2 * var2 + 2 * w1 * w2 * covar
        happiness <- expected_return - variance
        return(-happiness)  # for minimization
    }

    result <- optim(par = 0.5, fn = h, method = "Brent", lower = 0, upper = 1)
    w1 <- result$par
    w2 <- 1 - w1
    E_r <- w1 * mean1 + w2 * mean2
    Var_r <- w1^2 * var1 + w2^2 * var2 + 2 * w1 * w2 * covar
    h_r <- E_r - Var_r
    return(c(w1, w2, E_r, Var_r, h_r))
}

# Define inputs
means <- c(mean(Nvda), mean(Cmcsa), mean(Ea))
vars <- c(var(Nvda), var(Cmcsa), var(Ea))
cov_mat <- cov(pfol)

# Run optimization for each stock pair
res1 <- optimize_portfolio(means[1], means[2], vars[1], vars[2], cov_mat[1, 2])  # NVDA & CMCSA
res2 <- optimize_portfolio(means[1], means[3], vars[1], vars[3], cov_mat[1, 3])  # NVDA & EA
res3 <- optimize_portfolio(means[2], means[3], vars[2], vars[3], cov_mat[2, 3])  # CMCSA & EA

# Combine results into a table
optimal_portfolios <- data.frame(Pair = c("NVDA & CMCSA", "NVDA & EA", "CMCSA & EA"), Weight1 = round(c
    res2[1], res3[1]), 4), Weight2 = round(c(res1[2], res2[2], res3[2]), 4), `E(r)` = round(c(res1[3],
    res2[3], res3[3]), 4), `Var(r)` = round(c(res1[4], res2[4], res3[4]), 4), `h(r)` = round(c(res1[5],
    res2[5], res3[5]), 4))

kable(optimal_portfolios, caption = "Optimal Portfolio Statistics")
```

Table 6: Optimal Portfolio Statistics

Pair	Weight1	Weight2	E.r.	Var.r.	h.r.
NVDA & CMCSA	0.1272	0.8728	0.0298	3.9839	-3.9541
NVDA & EA	0.2267	0.7733	0.0432	6.0620	-6.0189
CMCSA & EA	0.6651	0.3349	0.0235	3.3295	-3.3060

Pair	Weight1	Weight2	E.r.	Var.r.	h.r.

Interpretation From the results:

The NVDA and EA portfolio produces the highest happiness score, suggesting it offers a strong balance of return and reduced variance, despite its counterintuitive weights.

CMCSA and EA yields a more stable portfolio due to the lower variance, indicating a potentially lower-risk option.

NVDA and CMCSA provides a higher expected return, which might appeal to risk-seeking investors.

Note that some optimal weights may appear unusual (e.g., weights exceeding 1 or negative), which implies short-selling or leverage — common in theoretical models. In practical applications, you may wish to restrict weights to [0, 1] using lower = 0 and upper = 1 in optim().

2.10. The impact of financial events on returns (6 points)

To investigate the impact of major financial events on stock returns, we perform a regression analysis using three event categories:

- Lehman Brothers bankruptcy (2008-09-15)
- COVID-19 pandemic declaration (2020-03-11)
- Business as usual (BAU) for all other dates

We create a factor variable `event` and regress percentage returns (`per_returns`) on it.

```r
# Function to extract event-based return data
extract <- function(stocks, smb) {
    stoc <- subset(stocks, symbol == smb)

    # Event dates
    lehman <- stoc[stoc$date == "2008-09-15", ]
    pandemic <- stoc[stoc$date == "2020-03-11", ]

    # Business as usual data
    bankrpcy1 <- stoc[stoc$date != "2008-09-15", ]
    bankrpcy2 <- stoc[stoc$date != "2020-03-11", ]
    BAU <- rbind(bankrpcy1, bankrpcy2)

    # Assign event labels
    BAU$event <- "BAU"
    lehman$event <- "lehman"
    pandemic$event <- "pandemic"

    # Combine and factorize
    events <- rbind(lehman, pandemic, BAU)
    events$event <- factor(events$event)

    # Clean NA
    events <- na.omit(events)
    return(events)
}

# Apply extraction function
NVDA_EVENTS <- extract(stocks, "NVDA")
```

```r
CMCSA_EVENTS <- extract(stocks, "CMCSA")
EA_EVENTS <- extract(stocks, "EA")

# Run linear regression
model_nvda <- lm(per_returns ~ event, data = NVDA_EVENTS)
model_cmcsa <- lm(per_returns ~ event, data = CMCSA_EVENTS)
model_ea <- lm(per_returns ~ event, data = EA_EVENTS)

# Output coefficients for NVDA
summary(model_nvda)$coefficients
```

```
##                  Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)    0.09296392 0.03553253  2.616304 0.008900556
## eventlehman   -9.23219347 3.77732714 -2.444107 0.014536283
## eventpandemic -5.85163767 3.77732714 -1.549148 0.121374224
```

```r
# Output coefficients for CMCSA
summary(model_cmcsa)$coefficients
```

```
##                  Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)    0.02141967 0.01925671  1.112322 0.26602327
## eventlehman   -4.01532903 2.04710673 -1.961465 0.04984928
## eventpandemic -3.11213808 2.04710673 -1.520262 0.12847318
```

```r
# Output coefficients for EA
summary(model_ea)$coefficients
```

```
##                  Estimate Std. Error    t value  Pr(>|t|)
## (Intercept)    0.02948125 0.02443765  1.2063867 0.2276937
## eventlehman   -3.85824110 2.59787261 -1.4851541 0.1375309
## eventpandemic -2.18050439 2.59787261 -0.8393423 0.4012950
```

Interpretation

(a) According to the regression models, none of the three stocks exhibited consistently positive returns over time. The R^2 values from the models are not significantly different from zero, indicating that these event classifications do not explain a large portion of return variability.

(b) Both the pandemic and Lehman bankruptcy events had negative impacts on returns for all three stocks. However, the impact from the Lehman Brothers bankruptcy was more severe across the board, as evidenced by the fact that the coefficients for `"eventlehman"` were more negative than those for `"eventpandemic"`.

The Lehman event had a statistically significant negative impact on:

- NVDA: p-value = 0.0146
- CMCSA: p-value = 0.0499

This suggests that historical financial crises, especially sudden systemic shocks, can significantly hurt short-term stock performance — a critical insight for investors managing risk exposure in turbulent times.