

Estimating Systematic Risk Using the Capital Asset Pricing Model, CAPM

Application of CAPM to IBM and Microsoft Stocks, 2000-2022

John Karuitha

12/6/22

Contents

1	Background	2
2	Critique of the CAPM	3
3	Data	4
4	Exploring the Data	6
4.1	Missing Values	6
4.2	Summary Statistics for Stock Prices	9
4.3	Trends in Stock Returns	9
4.4	Summary Statistics - Stock Returns	11
5	Regression Model	13
5.1	Regression for GM Stock	13
5.2	Regression for Microsoft Stock	14
5.3	Are the β 's for Both Stocks Significantly Different from 1 at the 95% Confidence Level?	15
5.3.1	GM	15
5.3.2	Microsoft	15
6	Checking Regression Model Assumptions	17
6.1	GM Model	17
6.2	Microsoft Model	19
7	Conclusion	21
	References	22

Chapter 1

Background

In this analysis, I use stock prices for Microsoft and General Motors (GM) to estimate the systematic risk of the stocks using the [Capital Asset Pricing Model \(CAPM\)](#). CAPM, developed by William Sharpe, Jack Treynor, John Lintner and Jan Mossin ([Perold, 2004](#)) quantifies the systematic risk and the expected return on an asset, particularly stocks.

[Systematic risk](#) is inherent to the entire market or market segment. The risk affects the entire economy. Hence, investors cannot diversify away systematic as they can unsystematic risk. For this reason, systematic risk is also referred to as **undiversifiable risk**, **market risk** or **volatility risk**.” In finance terminology, β , a symbol that we define in the next section before we encounter it in the section on regression analysis represents systematic risk.

CAPM quantifies the relationship between the required rate of return of an investment and its risk. The model utilizes return on the given asset r_a (in our case the return on GM or Microsoft), the risk free rate r_f usually proxied by the return on the 90 day treasury bills, and the return on the market r_m taken as the return a market index like the Dow Jones Industrial Index, the S&P 500, or other similar indices ([Bodie, 2020](#)). The relationship is as follows;

$$r_a - r_f = \beta (r_m - r_f)$$

A stock with a β of 1 is as volatile as the market. However, stocks with $\beta > 1$ are more volatile than the market and hence should command a higher return. Including such a stock in your portfolio will raise the volatility of your portfolio relative to the market portfolio. On the contrary, stocks with a $\beta < 1$ have lower risk relative to the market and correspondingly, a lower rate of return. The stocks with a $\beta < 1$ reduce the volatility of the portfolio relative to the market. In other words, such assets have a diversification benefit.

Chapter 2

Critique of the CAPM

CAPM is a simple and pragmatic to utilize. However, this simplicity also gives rise to its major weaknesses, the unrealistic assumptions. Specifically, the model rests on the following [assumptions about markets, financial assets, and market actors](#) (Dumas, 1996; Simmet, 2020).

- Investors are risk-averse, utility-maximizing, rational individuals.
- Markets are friction-less, devoid of transaction costs and taxes.
- Investors plan for the same, single holding period.
- Investors have homogeneous expectations or beliefs.
- All investments are infinitely divisible.
- Investors are price takers.

As a result we have several alternatives to CAPM. One such model is based on the Arbitrage Pricing Theory (APT) (Roll, 1980) that, even when more [theoretically solid](#), are harder to apply (Huberman, 1982).

More recent developments are the models by Fama and French that seek to improve on CAPM. For instance, the [Fama and French 3 factor model](#) extends CAPM by factoring in the proportion of the value stocks and small cap stocks in a portfolio. The Fama and French model has seen further extensions by adding more factors (Dirkx, 2020; Mosoeu, 2020).

Chapter 3

Data

I use data from Yahoo! Finance for the period 2002-2022. The *Tidyquant* package in R (courtesy of [Matt Dancho](#)) ([R Core Team, 2022](#); [Dancho, 2022](#)) provides a convenient way to get this data (see the code chunk below). I collect data for the period from January 2, 2000 to October 25, 2022. Further, I only work with adjusted prices (*adjusted*). The adjustments take into account events like stock splits and share buy-backs. The data also includes the 90 day US treasury bills rate that proxies the risk free rate.

```
## uncomment this code to download the data
# tq_get(c("GM", "MSFT", "^GSPC"),
#       get = "stock.prices",
#       from = "2000-01-01",
#       to = "2022-10-25") %>%
#   select(date, symbol, adjusted) %>%
#   mutate(symbol = case_when(
#     symbol == "^GSPC" ~ "SP500",
#     TRUE ~ symbol
#   )) %>%
#   write_csv(., "stocks.csv")
# #####
# tq_get("DTB3", from = '2000-01-01', to = '2022-10-25', get = "economic.data") %>%
#   select(date, symbol, price) %>%
#   rename(adjusted = price) %>%
#   mutate(symbol = if_else(symbol == "DTB3", "us.tbill", symbol)) %>%
#   write_csv(., "tbills.csv")
```

We also get the [90 day treasury bills rate here](#) and combine this with the stock prices data collected

above.

```
stock_prices <- read_csv("stocks.csv") %>%
  pivot_wider(names_from = symbol, values_from = adjusted)

#####
tbill_rate <- read_csv("FRB_H15.csv", skip = 5,
  na = c("", "NA", "ND")) %>%
  clean_names() %>% mutate(time_period = mdy(time_period)) %>%
  select(date = time_period,
    adjusted = riflgfcm03_n.b) %>%
  mutate(symbol = "tbill") %>%
  relocate(symbol, .before = adjusted) %>%
  filter(date > '2000-01-01')

#####
final_data <- stock_prices %>%
  left_join(tbill_rate, by = 'date') %>%
  select(-symbol) %>% rename(tbill = adjusted)
```

The result is a data set with 5740 and 5 columns. The table below defines each of the variables.

```
tribble(~ Variable, ~ Description,
  "Date", "The date corresponding to the stock price.",
  "GM", "The stock for GM, General Motors.",
  "MSFT", "The stock for Microsoft.",
  "SP500", "The S&P 500 Index.",
  "tbill", "The US 3 months treasury bills rate.") %>%
  formatting_function(caption = "Description of Variables")
```

Table 3.1: Description of Variables

Variable	Description
Date	The date corresponding to the stock price.
GM	The stock for GM, General Motors.
MSFT	The stock for Microsoft.
SP500	The S&P 500 Index.
tbill	The US 3 months treasury bills rate.

Chapter 4

Exploring the Data

4.1 Missing Values

The data set has a few missing values for treasury bills. However, for GM, data is only available in the Yahoo! Finance database from November 18, 2010 and hence the notable gap.

```
Amelia::missmap(final_data)
```

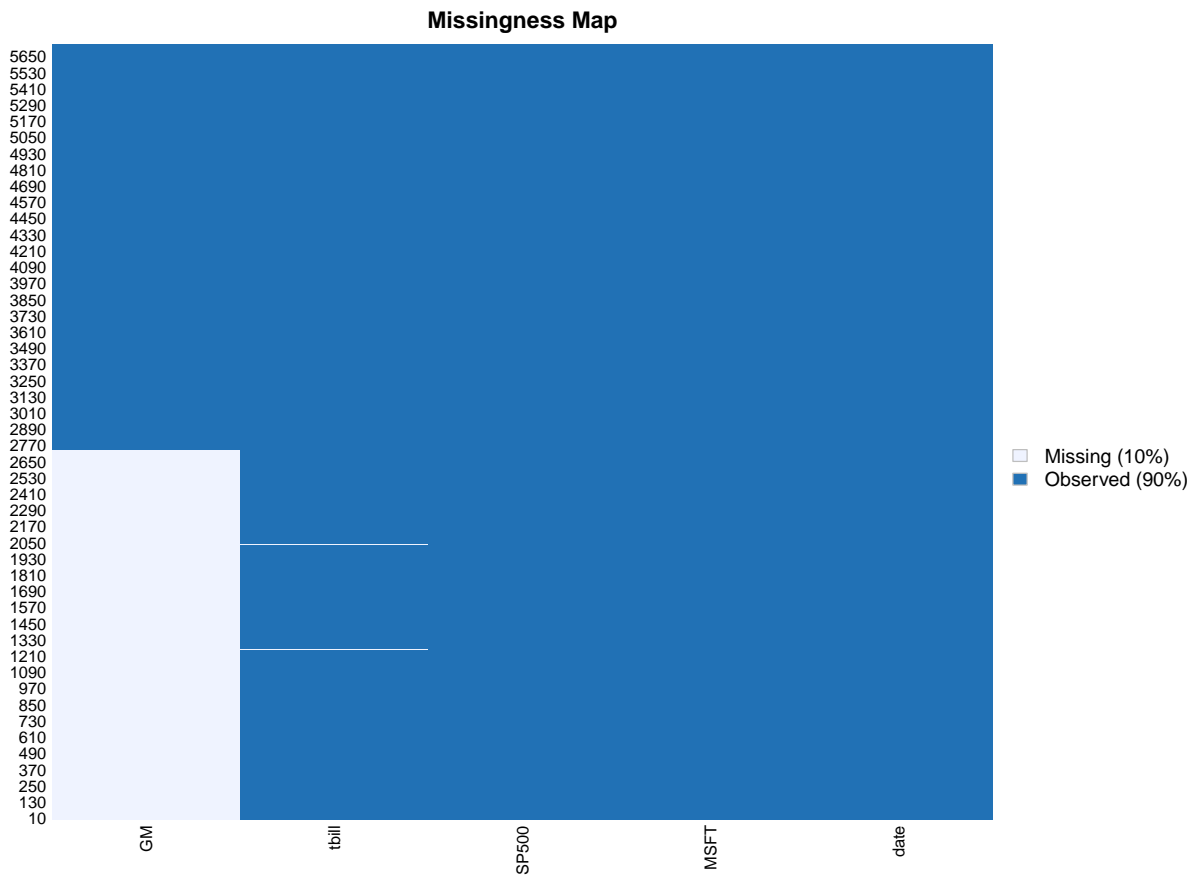


Figure 4.1: Missing values

Hence I interpolate the missing values with the median.

```
stock_prices <- stock_prices %>%

  mutate(MSFT = replace_na(MSFT, median(MSFT, na.rm = TRUE)),

         SP500 = replace_na(SP500, median(SP500, na.rm = TRUE)))
#####
tbill_rate <- tbill_rate %>%
  mutate(adjusted = replace_na(adjusted, median(adjusted, na.rm = TRUE)))

#####

final_data <- stock_prices %>%
  left_join(tbill_rate, by = 'date') %>%
```



```
select(-symbol) %>% rename(tbill = adjusted)
```

We then examine the trend in stock prices for GM and Microsoft in Figure () below.

```
final_data %>%
```

```
  pivot_longer(cols = -date, names_to = "stock",
               values_to = "value") %>%
  ggplot(mapping = aes(x = date, y = value, color = stock)) +
  geom_line() + theme_minimal() + labs(x = "", y = "Adjusted Prices",
                                     title = "Trends in GM and Microsoft Stock Prices",
                                     caption = "John karuitha, 2022 \n Data source: Yahoo! Fin
  theme(legend.position = 'bottom', legend.title = element_blank()) +
  facet_wrap(~stock, scales = "free_y")
```

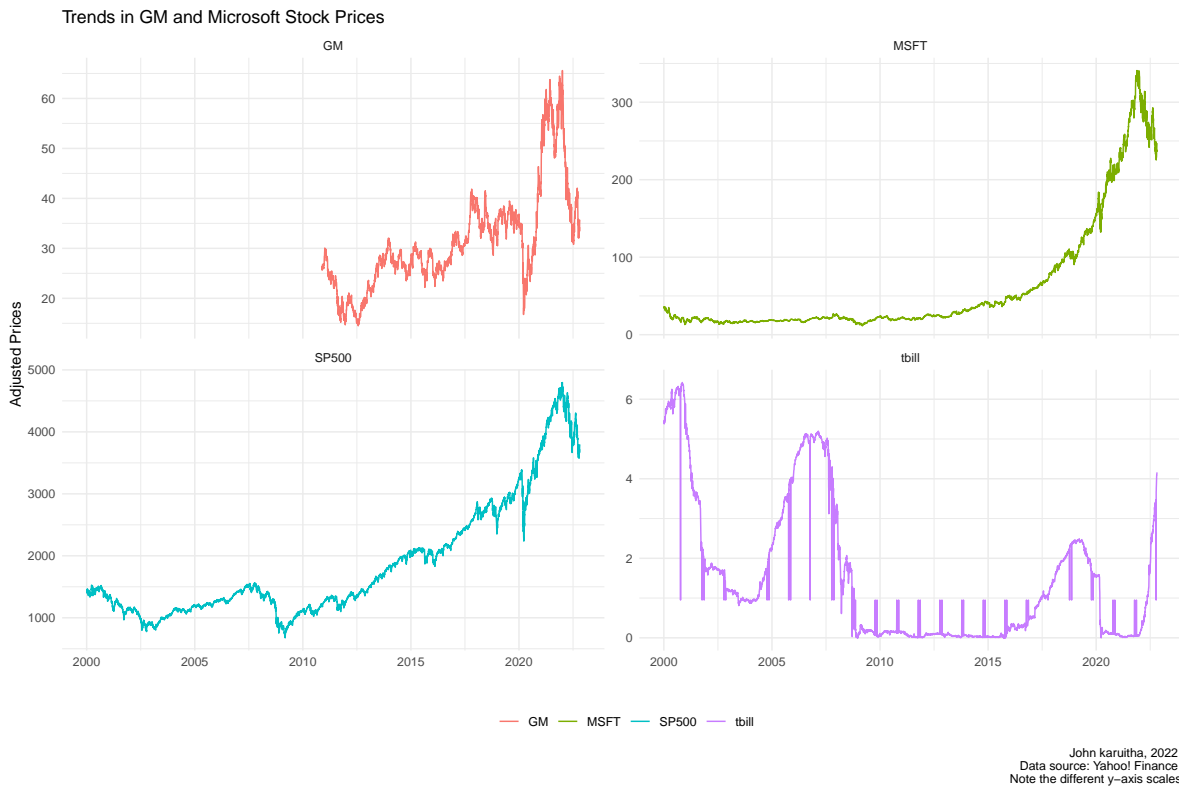


Figure 4.2: Trends in GM and Microsoft Stock Prices

All the stocks have had a steep increase in prices. The treasury bill rate has been relatively flat compared to the stock with notable occasional spikes.

4.2 Summary Statistics for Stock Prices

Next, I summarise the prices of stocks.

```
#head(final_data)
final_data %>%
  select(-date) %>%
  skimr::skim_without_charts() %>%
  mutate(across(-c(skim_type, skim_variable), .fns = round, 2)) %>%
  select(-skim_type) %>%
  formatting_function(caption = "Summary Statistics for Full Data")
```

Table 4.1: Summary Statistics for Full Data

skim_variable	n_missing	complete_rate	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	n
GM	2737	0.52	31.37	10.17	14.50	25.26	29.28	
MSFT	0	1.00	60.90	75.87	11.44	19.10	23.01	
SP500	0	1.00	1857.13	951.66	676.53	1182.11	1426.67	
tbill	0	1.00	1.53	1.77	0.00	0.09	0.95	

4.3 Trends in Stock Returns

In this section I compute the returns of the assets and visualize them. I compute the return of each stock as the price of the stock today less the price of the stock in the previous day divided by the price of the stock in the previous day.

$$return = \frac{price_t - price_{t-1}}{price_{t-1}}$$

```
returns <- final_data %>%
  pivot_longer(cols = -date, names_to = "stock",
               values_to = "value") %>%
  mutate(value = value + 0.00000001) %>%
  group_by(stock) %>%
  summarise(return = (value - dplyr::lag(value)) / dplyr::lag(value)) %>%
  drop_na(return) %>%
  mutate(day = row_number()) %>%
  ungroup() %>% relocate(day)

#####
returns %>% ggplot(mapping = aes(x = day, y = return, color = stock)) +
  geom_line() + theme_minimal() + labs(x = "", y = "Returns",
                                       title = "Trends in GM and Microsoft Stock Returns",
                                       caption = "John karuitha, 2022 \n Data source: Yahoo! Fin
```

```
theme(legend.position = 'bottom', legend.title = element_blank()) +
facet_wrap(~stock, scale = "free")
```

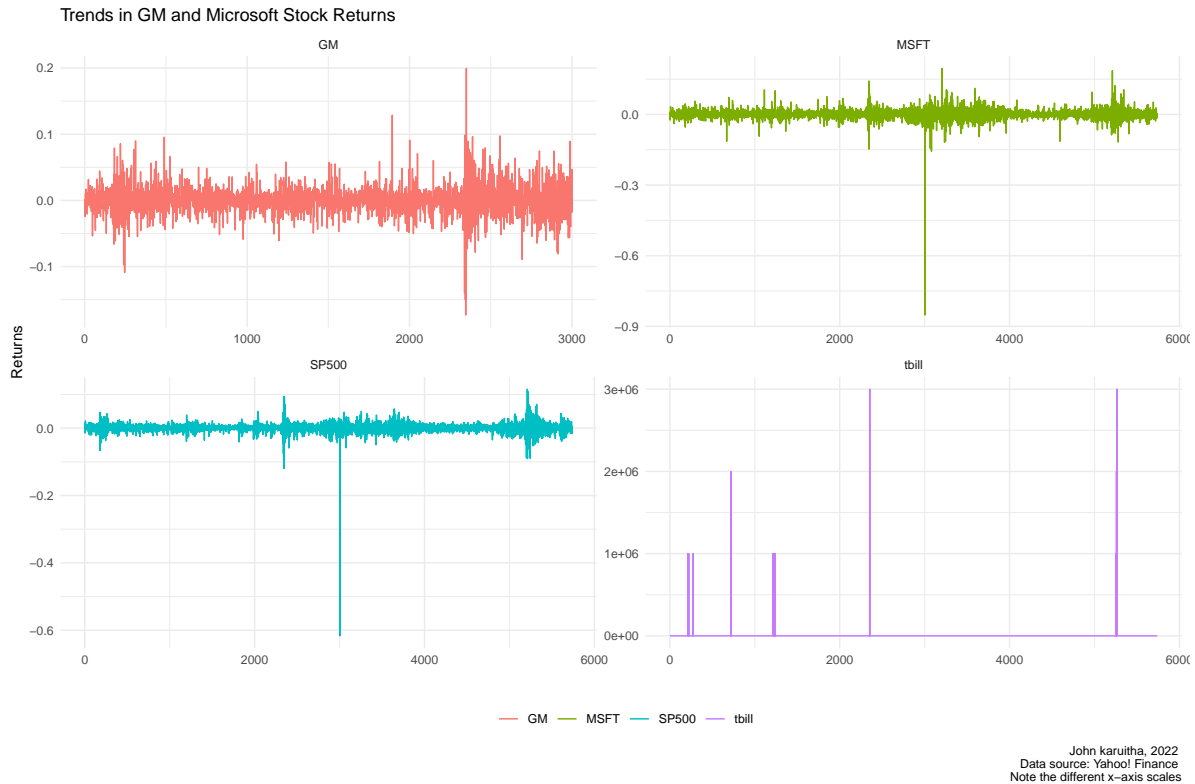


Figure 4.3: Trends in GM and Microsoft Stock Returns

The stocks and the market exhibit mean reversion although the GM stock appears the most volatile. The returns from the stock of Microsoft are surprisingly stable and closely resembles the pattern of the market portfolio (S&P500). The treasury bill has an erratic pattern with notable swings. We shall revisit the significance of this perceived difference in volatility in the next section. We also look at the distribution of the variables in the dataset.

```
returns %>%
  drop_na(return) %>%
  ggplot(mapping = aes(x = return, fill = stock)) +
  geom_histogram(alpha = 0.6) +
  theme_minimal() + labs(x = "% Return", y = "", title = "Distribution of Returns") +
  theme(legend.title = element_blank(),
        legend.position = "bottom") + xlim(-1, 1.5) +
```

```
scale_y_continuous(trans = 'log10',
  labels = scales::comma_format())
```

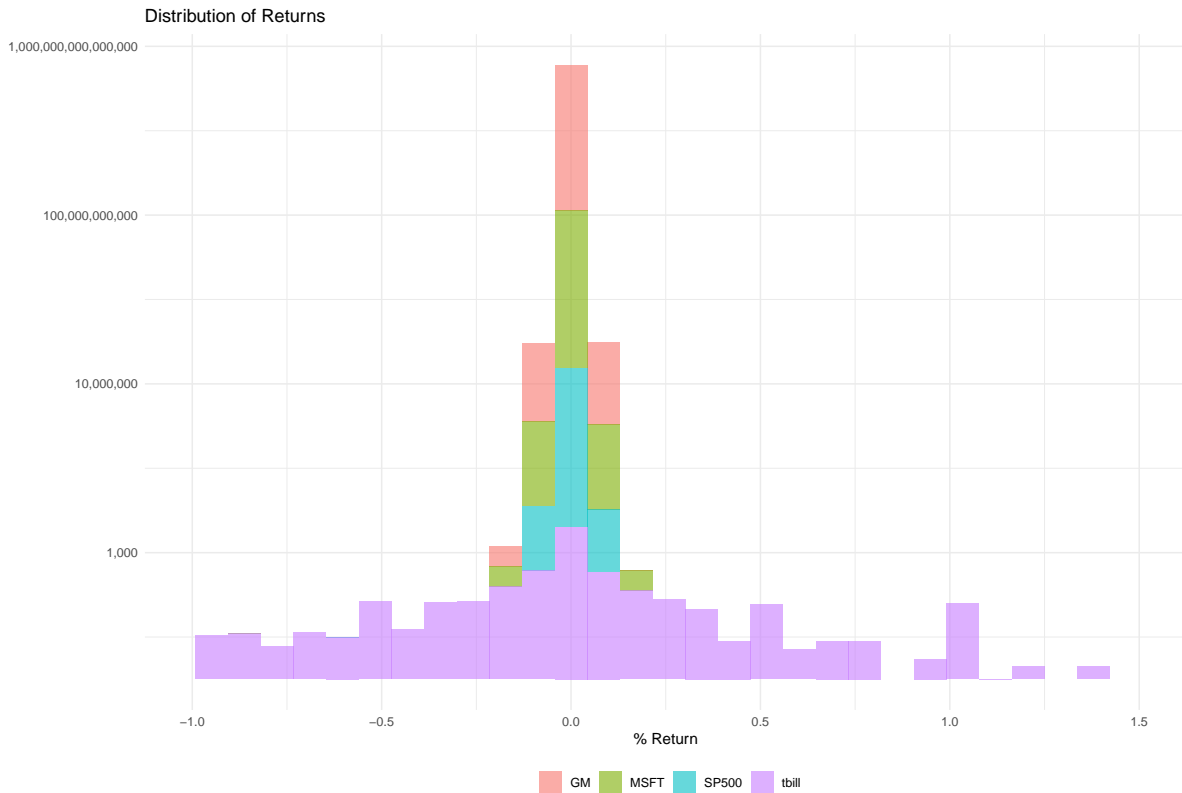


Figure 4.4: Distribution of Stock Returns

4.4 Summary Statistics - Stock Returns

In this section, we summarise the data on stock returns.

```
returns %>%
group_by(stock) %>%
  skimr::skim_without_charts() %>%
  filter(skim_variable != "day") %>%
  select(-skim_type, -skim_variable,
    -complete_rate) %>%
  kbl(booktabs = TRUE,
    caption = "Summary Statistics") %>%
  kable_classic(latex_options = c("hold_position", "striped"))
```

Table 4.2: Summary Statistics

stock	n_missing	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	numeric.p75	numeric.p100
GM	0	3.37400e-04	2.17414e-02	-0.1732284	-0.0103114	0.0002797	0.0108732	1.9943
MSFT	0	3.69500e-04	2.23683e-02	-0.8521522	-0.0082206	0.0003478	0.0093472	1.9565
SP500	0	1.35100e-04	1.49082e-02	-0.6167791	-0.0048761	0.0005877	0.0058958	1.1580
tbill	0	3.65929e+03	8.02175e+04	-0.9999998	-0.0087083	0.0000000	0.0091743	3.00000

Chapter 5

Regression Model

To fit the regression model we compute the excess returns of the stocks (GM and Microsoft) and the market portfolio (S&P 500) over return on the US 90 day Treasury bills rate. For each stock, we fit the excess return (premium) against the excess return of the market.

5.1 Regression for GM Stock

The β for GM is negative but not statistically significant. However, the α is significant but very close to 1 (at 1.0012069).

```
gm_reg <- final_data %>%
  select(GM, SP500, tbill) %>%
  drop_na(GM) %>%
  mutate(tbill =
    replace_na(tbill, median(tbill, na.rm = TRUE))) %>%
  summarise(gm_return = diff(log(GM)),
            market_return = diff(log(SP500)),
            risk_free = diff(log(tbill + 0.01))
  ) %>%

  transmute(gm_premium = gm_return - risk_free,
            market_premium = market_return - risk_free) %>%
  lm(gm_premium ~ market_premium, data = .)

summary(gm_reg)
```

Call:

```
lm(formula = gm_premium ~ market_premium, data = .)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.137871	-0.009037	-0.000312	0.007969	0.128284

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0002831	0.0003090	-0.916	0.36
market_premium	1.0006233	0.0009257	1080.906	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01693 on 3000 degrees of freedom
Multiple R-squared: 0.9974, Adjusted R-squared: 0.9974
F-statistic: 1.168e+06 on 1 and 3000 DF, p-value: < 2.2e-16

5.2 Regression for Microsoft Stock

Like in the case of GM, the `tbill` for Microsoft is not statistically significant. However, the `market_premium` of 1.002 is significant but close to 1.

```
msft_reg <- final_data %>%
  select(MSFT, SP500, tbill) %>%
  drop_na(MSFT) %>%
  mutate(tbill =
    replace_na(tbill, median(tbill, na.rm = TRUE))) %>%
  summarise(msft_return = diff(log(MSFT)),
            market_return = diff(log(SP500)),
            risk_free = diff(log(tbill + 0.01))
  ) %>%

  transmute(msft_premium = msft_return - risk_free,
            market_premium = market_return - risk_free) %>%
  lm(msft_premium ~ market_premium, data = .)

summary(msft_reg)
```

Call:

```
lm(formula = msft_premium ~ market_premium, data = .)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.94980	-0.00581	-0.00005	0.00604	0.14447

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.070e-07	2.456e-04	0.003	0.997

```
market_premium 1.002e+00 8.682e-04 1154.306 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01861 on 5737 degrees of freedom
Multiple R-squared:  0.9957,    Adjusted R-squared:  0.9957
F-statistic: 1.332e+06 on 1 and 5737 DF,  p-value: < 2.2e-16
```

5.3 Are the β 's for Both Stocks Significantly Different from 1 at the 95% Confidence Level?

To evaluate the riskiness of an asset, we examine 2 values β and α . In both cases the alpha is not statistically significant. The threshold for β is 1. An asset with a β of 1 is as volatile as the market portfolio. For $\beta > 1$, the asset is more volatile than the market portfolio and commands higher returns. The opposite is true for $\beta < 1$.

In this section we examine whether the computed β differ significantly from 1. The hypotheses are as follows.

H_0 : The β is not equal to 1. H_1 : The β is equal to 1.

We compute the t-statistic as follows.

$$t = \frac{\text{hypothesised}_{\beta}}{\text{standarderror}}$$

From the t-statistic, we then compute the p-value using the `pt` function in R.

5.3.1 GM

The β for the GM stock does not differ significantly from 1.

```
hypothesised_value = 1
tgm = (gm_reg$coefficients["market_premium"] - hypothesised_value) / summary(gm_reg)$sigma
pt(tgm, df.residual(gm_reg))
```

```
market_premium
0.5146824
```

5.3.2 Microsoft

Likewise, the β for the GM stock does not differ significantly from 1.

```
tmsft = (msft_reg$coefficients["market_premium"] - hypothesised_value) / summary(msft_reg)$sigma
pt(tmsft, df = df.residual(msft_reg))
```



```
market_premium
0.5456123
```

Chapter 6

Checking Regression Model Assumptions

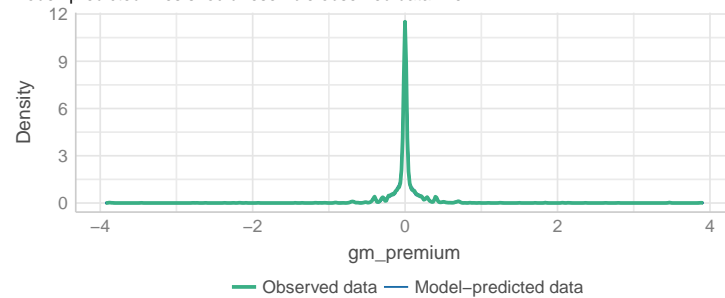
In this section, we check for violations of the regression model assumptions. Note that the model has one independent variable so that multicollinearity is not an issue. The model evaluation focuses on linearity, homoscedasticity, outliers, and normality of residuals.

6.1 GM Model

Figure () below shows that the model does not significantly violate these assumptions but for normality of residuals. Even then the model does not display a high degree of departure from this assumption.

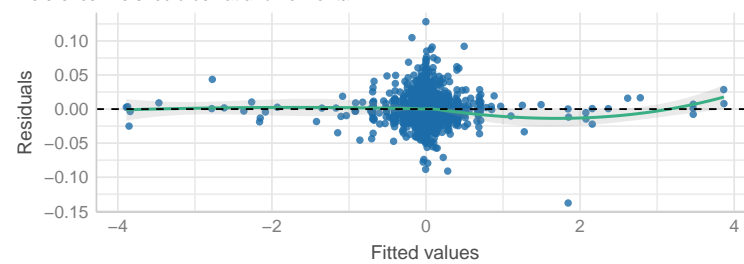
Posterior Predictive Check

Model-predicted lines should resemble observed data line



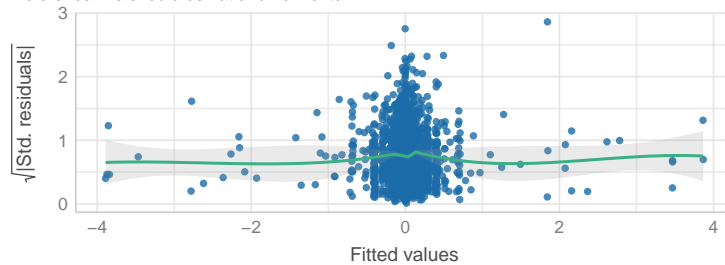
Linearity

Reference line should be flat and horizontal



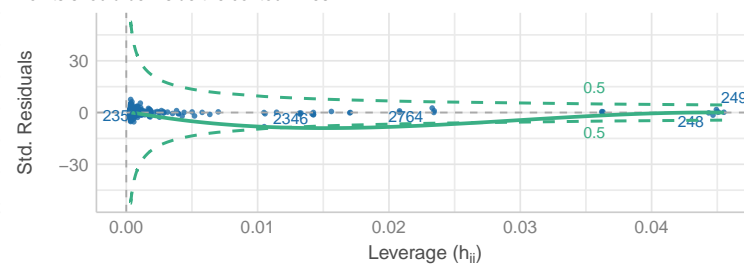
Homogeneity of Variance

Reference line should be flat and horizontal



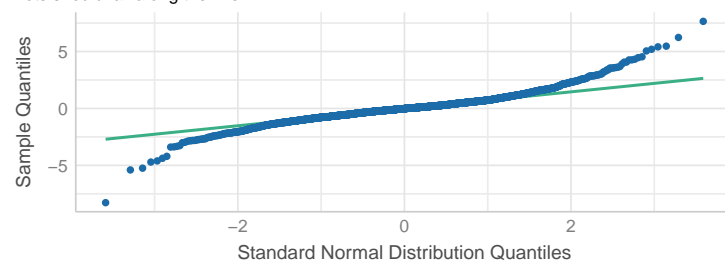
Influential Observations

Points should be inside the contour lines



Normality of Residuals

Dots should fall along the line

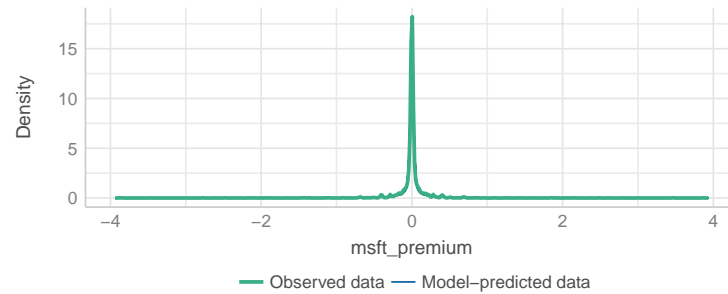


6.2 Microsoft Model

Figure () below shows that the model for Microsoft satisfies all the assumptions. As a robust model, it could be useful for prediction.

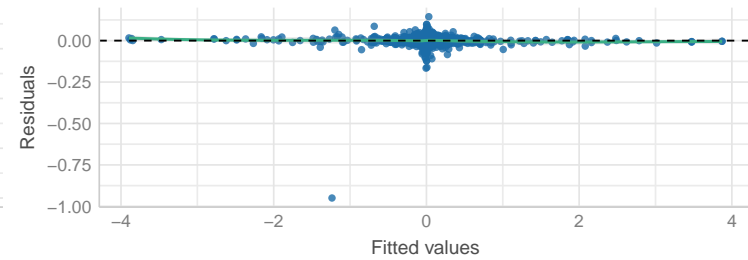
Posterior Predictive Check

Model-predicted lines should resemble observed data line



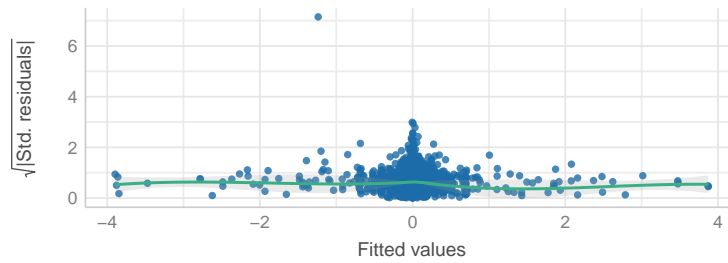
Linearity

Reference line should be flat and horizontal



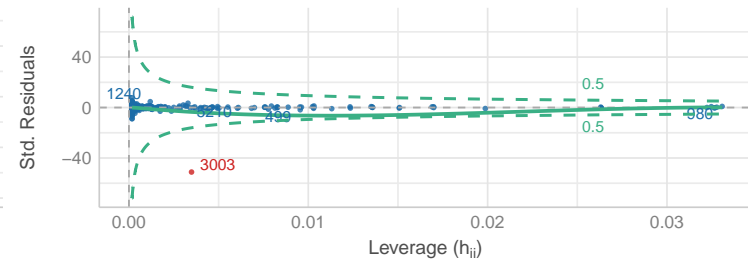
Homogeneity of Variance

Reference line should be flat and horizontal



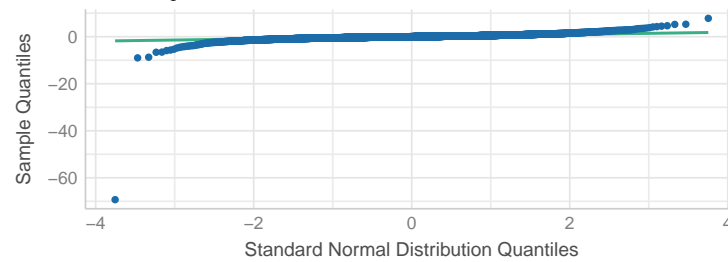
Influential Observations

Points should be inside the contour lines



Normality of Residuals

Dots should fall along the line



Chapter 7

Conclusion

In this analysis, we used the CAPM to estimate the systematic risk for GM and Microsoft stocks. The main findings are that both stocks are as volatile as the market portfolio. The implication is that in the long run, these stocks may not provide diversification benefits for investors. The results also suggest that while investors may outperform the market in the short run, they may not sustain this performance for extended periods.

References

- Bodie, A.A.M., Zvi AND Kane (2020) *Investments*. Boston: McGraw-Hill-Irwin.
- Dancho, D., Matt AND Vaughan (2022) *Tidyquant: Tidy quantitative financial analysis*. Available at: <https://CRAN.R-project.org/package=tidyquant>.
- Dirkx, F.J., Philipp AND Peter (2020) ‘The fama-french five-factor model plus momentum: Evidence for the german market’, *Schmalenbach Business Review*, 72(4), pp. 661–684.
- Dumas, B., Bernard AND Allaz (1996) ‘Critique and evaluation of the CAPM’, in *Financial securities*. Springer, pp. 113–149.
- Huberman, G. (1982) ‘A simple approach to arbitrage pricing theory’, *Journal of Economic Theory*, 28(1), pp. 183–191.
- Mosoeu, O., Selebogo AND Kodongo (2020) ‘The fama-french five-factor model and emerging market equity returns’, *The Quarterly Review of Economics and Finance* [Preprint].
- Perold, A.F. (2004) ‘The capital asset pricing model’, *Journal of Economic Perspectives*, 18(3), pp. 3–24.
- R Core Team (2022) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Roll, S.A., Richard AND Ross (1980) ‘An empirical investigation of the arbitrage pricing theory’, *The Journal of Finance*, 35(5), pp. 1073–1103.
- Simmet, W., Anastasia AND Pohlmeier (2020) ‘The CAPM with measurement error: “there’s life in the old dog yet!”’, *Jahrbücher für Nationalökonomie und Statistik*, 240(4), pp. 417–453.