HW: week 1

Question 2.1

2.1 Consider the following five observations. You are to do all the parts of this exercise using only a calculator.

x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$
3	4	$3-1=2$	$2^2=4$	$4-2=2$	$2\times2=4$
2	2	$2-1=1$	$1^2=1$	$2-2=0$	$1\times0=0$
1	3	$1-1=0$	$0^2=0$	$3-2=1$	$0\times1=0$
–1	1	$-1-1=-2$	$-2^2=4$	$1-2=-1$	$-2\times-1=2$
0	0	$0-1=-1$	$-1^2=1$	$0-2=-2$	$-1\times-2=2$
$\sum x_i = 5$	$\sum y_i = 10$	$\sum(x_i-\overline{x}) = 0$	$\sum(x_i-\overline{x})^2 = 10$	$\sum(y_i-\overline{y}) = 0$	$\sum(x_i-\overline{x})(y_i-\overline{y}) = 8$

Question a.

Given the data we first find the means:

$$\bar{x} = \frac{3+2+1+(-1)+0}{5} = 1$$

$$\bar{y} = \frac{4+2+3+1+0}{5} = 2$$

Question b.

Slope Coefficient:

$$b_2 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{8}{10} = 0.8$$

Intercept:

$$b_1 = \bar{y} - b_2\bar{x} = 2 - 0.8 \times 1 = 1.2$$

Question c.

Given data:

$$x = [3, 2, 1, -1, 0]$$

$$y = [4, 2, 3, 1, 0]$$

$$\sum_i x_i^2 = 3^2 + 2^2 + 1^2 + (-1)^2 + 0^2 = 15$$

$$\sum_i x_i y_i = 3*4 + 2*2 + 1*3 + (-1)*1 + 0*0 = 18$$

let number of observation = N then $N=5$

$N\bar{x}^2 = 5\times 1^2 = 5$ $\qquad$ $N\bar{x}\bar{y} = 5\times(1)\times(2) = 10$

$\sum_1 (x_i - \bar{x})^2 = 10$

$\sum_1 x_i^2 - N\bar{x}^2 = 15 - 5 = 10$ hence $\sum(x_i-\bar{x})^2 = \sum_1 x_i^2 - N\bar{x}^2$

$\sum_1 (x_i - \bar{x})(y_i - \bar{y}) = 8$

$\sum_1 x_i y_i - N\bar{x}\bar{y} = 18 - 10 = 8$ hence $\sum_1 (x_i-\bar{x})(y_i-\bar{y}) = \sum_1 x_i y_i - N\bar{x}\bar{y}$

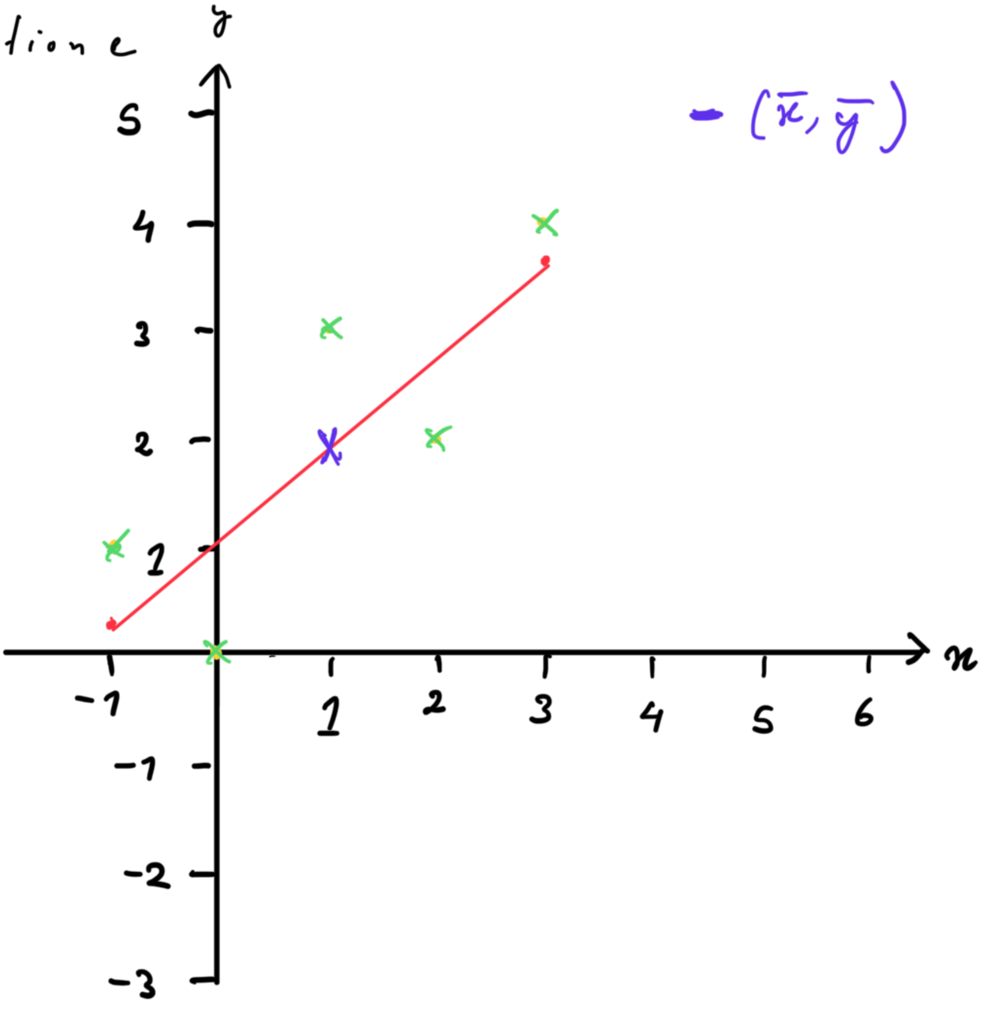
d. Use the least squares estimates from part (b) to compute the fitted values of y, and complete the remainder of the table below. Put the sums in the last row.
Calculate the sample variance of y, $s_y^2 = \sum_{i=1}^N (y_i - \overline{y})^2/(N-1)$, the sample variance of x, $s_x^2 = \sum_{i=1}^N (x_i - \overline{x})^2/(N-1)$, the sample covariance between x and y, $s_{xy} = \sum_{i=1}^N (y_i - \overline{y})(x_i - \overline{x})/(N-1)$, the sample correlation between x and y, $r_{xy} = s_{xy}/(s_x s_y)$ and the coefficient of variation of x, $CV_x = 100(s_x/\overline{x})$. What is the median, 50th percentile, of x?

x_i	y_i	$\hat{y}_i$	$\hat{e}_i$	$\hat{e}_i^2$	$x_i\hat{e}_i$
3	4	3.6	0.4	0.16	1.2
2	2	2.8	−0.8	0.64	−1.6
1	3	2.0	1.0	1	1.0
–1	1	0.4	0.6	0.36	−0.6
0	0	1.2	−1.2	1.44	0
$\sum x_i =$ 5	$\sum y_i =$ 10	$\sum \hat{y}_i =$ 10	$\sum \hat{e}_i =$ 0	$\sum \hat{e}_i^2 =$ 3.6	$\sum x_i\hat{e}_i =$ 0

$\hat{y} = 1.2 + 0.8x$; median value of $\cancel{x}$ $= 1$ because

$x = [-1, 0, 1, 2, 3]$

Question e



$x = -1$

$\hat{y} = 1.2 + 0.8(-1) = 0.4$

$x = 3$

$\hat{y} = 1.2 + 0.8(3) = 1.2 + 2.4 = 3.6$

Question G.

$\bar{x} = 1$ $\quad$ $b_1 = 1.2$

$\bar{y} = 2 \qquad b_2 = 0{,}8$

$\bar{y} = b_1 + b_2 \bar{x} = 1.2 + 0{,}8 \times 1 = 2$

Question h.

$\bar{y} = 2$

$$\bar{\hat{y}} = \frac{\sum_i \hat{y}_i}{N} = \frac{3{,}6 + 2{,}8 + 2{,}0 + 0{,}4 + 1.2}{5} = 2$$

Question i

$$\sigma^2 = s_y^2 = \frac{\sum_i (y_i - \bar{y})^2}{N-1} = \frac{4+0+1+1+4}{5-1} = \frac{10}{4} = 2.5$$

Question j:

$$\operatorname{var}(\hat{b}_2 | x) = \frac{\hat{\sigma}^2}{\sum_i (x_i - \bar{x})^2} = \frac{2.5}{10} = 0.25$$

$$se(\hat{b}_2) = \sqrt{\operatorname{var}(\hat{b}_2 | x)} = \sqrt{0{,}25} = 0{,}5$$

Question 2.14

Question a.

$$\text{Elasticity} = \frac{\partial \widehat{WAGE}}{\partial EDUC} \times \frac{EDUC}{\widehat{WAGE}} = b_2 \times \frac{\overline{EDUC}}{\overline{WAGE}}$$

Where $b_2 = 1.80$, $\widehat{WAGE} = 19.74$ hence

$$19.74 \approx -4.88 + 1.80\,\overline{EDUC} \Rightarrow \overline{EDUC} = \frac{19.74 + 4.88}{1.80} = 13.68$$

$$\text{Elasticity} = 1.80 \times \frac{13.68}{19.74} \approx 1.25$$

Question b.

$$SE(E) = SE(\beta_2) \times \frac{\overline{EDUC}}{\overline{WAGE}}$$

$\beta_2 = 2.46$ $\quad \overline{EDUC} = 13.68$

$SE(\beta_2) = 0.16$ $\quad \overline{WAGE} = 19.74$

$$SE(E) = 0.16 \times \frac{13.68}{19.74} \approx 0.111$$

Question c.

12 years of education:

Urban:

$$\overline{WAGE} = -10.76 + 2.46 \times 12 = -10.76 + 29.52 = 18.76$$

RURAL:

RURAL:

$$\hat{WAGE} = -4.88 \ 1.80 \times 12 = 16.72$$

16 years of education:

Urban: $\hat{WAGE} = -10.76 + 2.46 \times 16 = 28.60$

Rural: $\hat{WAGE} = -4.88 + 1.80 \times 16 = 23.32$

Question 2.16

Part a.

I think the simplest explanation is that the CAPM model looks exactly like a simple regression model. For example the CAPM model clearly has one independent variable (rm - rf). According to the reading materials, simple regressionj models assume a linear relationship between a dependent varibale and only one independent variable. Additionally the relation between the dependent variable (rj - rf) and the independent variable (rm - rf) is linear in parameters alpha j and beta j, making it a simple linear regression model. The model includes an error term ej to account for deviations of observed values from predicted values, which is essential for any regression model as noted in the principles from your project documents.

Part b.

Results of the regression models:

(I guess General Electric):

```
Call:
lm(formula = ge_excess ~ mkt_excess, data = capm4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.161385 -0.037402 -0.003939  0.034422  0.182421

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.001167   0.004759  -0.245    0.807
mkt_excess   0.899260   0.098782   9.104 1.33e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05468 on 130 degrees of freedom
Multiple R-squared:  0.3893,    Adjusted R-squared:  0.3846
F-statistic: 82.87 on 1 and 130 DF,  p-value: 1.325e-15
```

Disney regression model:

```
Call:
lm(formula = dis_excess ~ mkt_excess, data = capm4)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.184659 -0.029342 -0.007377  0.029038  0.277836 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.001149   0.005956  -0.193    0.847    
mkt_excess   0.897838   0.123627   7.262 3.11e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06843 on 130 degrees of freedom
Multiple R-squared:  0.2886,    Adjusted R-squared:  0.2831 
F-statistic: 52.74 on 1 and 130 DF,  p-value: 3.11e-11
```

General Motors:

```
> summary(gm_model)

Call:
lm(formula = gm_excess ~ mkt_excess, data = capm4)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.40485 -0.05473 -0.00517  0.06035  0.29157 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.011550   0.009743  -1.185    0.238    
mkt_excess   1.261411   0.202223   6.238 5.77e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1119 on 130 degrees of freedom
Multiple R-squared:  0.2304,    Adjusted R-squared:  0.2244 
F-statistic: 38.91 on 1 and 130 DF,  p-value: 5.773e-09
```

IBM:

```
> summary(ibm_model)

Call:
lm(formula = ibm_excess ~ mkt_excess, data = capm4)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.266910 -0.039943 -0.002883  0.036759  0.265579 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.005851   0.006091   0.961    0.339    
mkt_excess  1.188208   0.126433   9.398 2.52e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06998 on 130 degrees of freedom
Multiple R-squared:  0.4045,	Adjusted R-squared:    0.4 
F-statistic: 88.32 on 1 and 130 DF,  p-value: 2.518e-16
```

MSFT:

```
> summary(msft_model)

Call:
lm(formula = msft_excess ~ mkt_excess, data = capm4)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27541 -0.05240 -0.00780  0.04239  0.35082 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.006098   0.007747   0.787    0.433    
mkt_excess  1.318947   0.160790   8.203 1.98e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.089 on 130 degrees of freedom
Multiple R-squared:  0.3411,	Adjusted R-squared:  0.336 
F-statistic: 67.29 on 1 and 130 DF,  p-value: 1.979e-13
```

XOM:

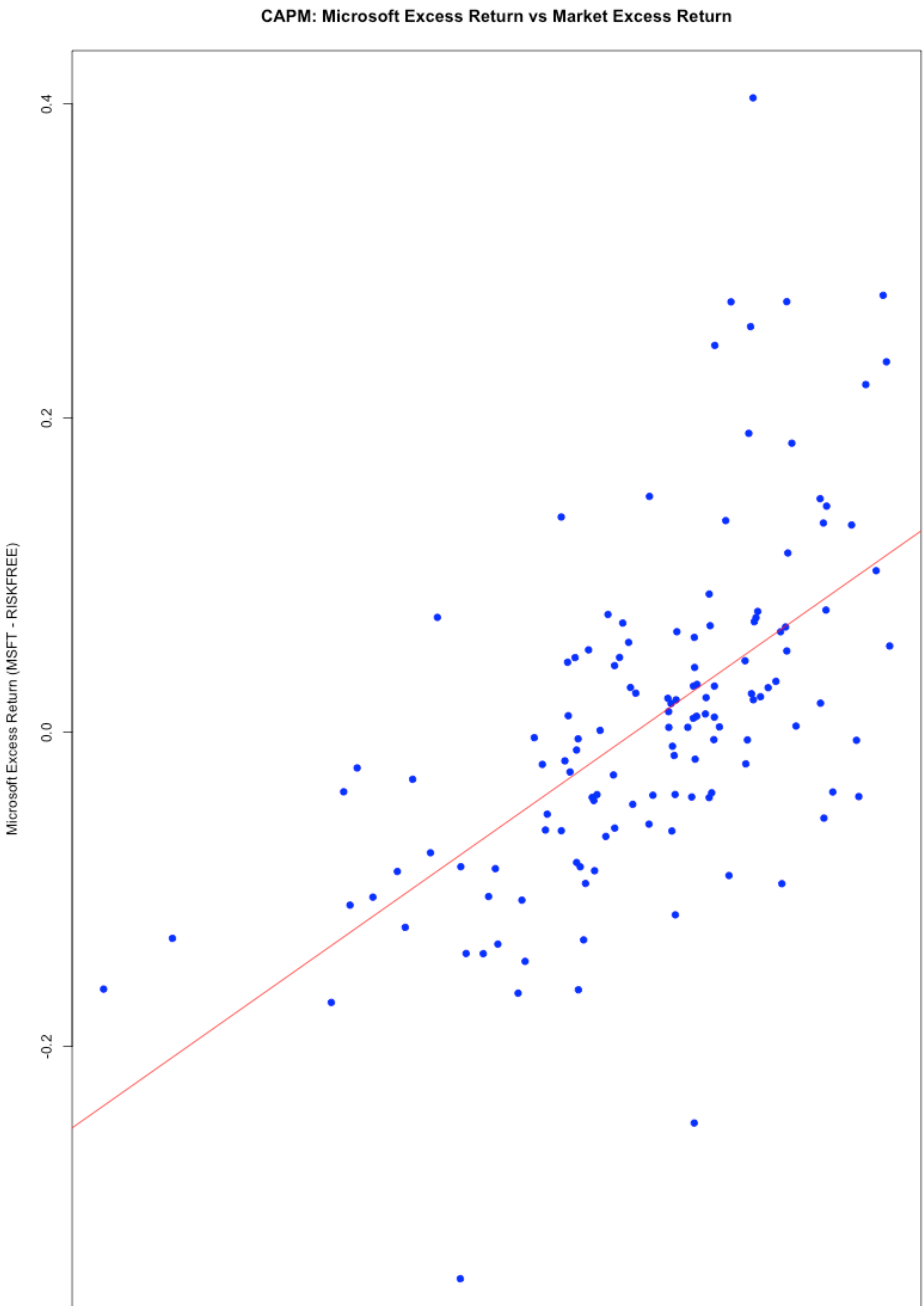
```
> summary(xom_model)

Call:
lm(formula = xom_excess ~ mkt_excess, data = capm4)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.117543 -0.029814 -0.001812  0.025259  0.213676 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.007880   0.004322   1.823   0.0706 .  
mkt_excess  0.413969   0.089713   4.614 9.33e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04966 on 130 degrees of freedom
Multiple R-squared:  0.1407,    Adjusted R-squared:  0.1341 
F-statistic: 21.29 on 1 and 130 DF,  p-value: 9.332e-06
```

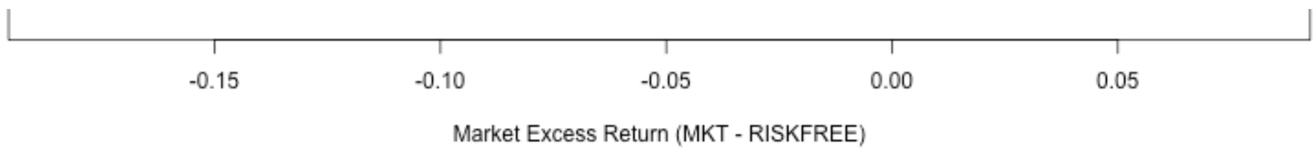
To make interpreting more simple I asked ChatGPT to prepare the following table:

```
Firm      Intercept (Estimate)   Beta (Estimate)   Adj. R-Squared
GE        -0.0012                0.8993            0.3846
Disney    -0.0011                0.8978            0.2831
GM        -0.0116                1.2614            0.2244
IBM       0.0059                 1.1882            0.4000
MSFT      0.0061                 1.3189            0.3360
XOM       0.0079                 0.4140            0.1341
```

As we can see from the table above Microsoft displays the largest beta and hence the strongest reaction to changes to market movements, making it the most aggressive firm in this sample. ExxonMobil exhibits a notably smaller beta, which means it is much less sensitive to overall market fluctuations and is therefore the most defensive. General Motors, IBM, General Electric, and Disney all have betas closer to one, but still vary in how closely their excess returns track those of the market.

Part c.


CAPM: Microsoft Excess Return vs Market Excess Return
Microsoft Excess Return (MSFT - RISKFREE)
0.4
0.2
0.0
-0.2



The scatter plot shows that Microsoft's excess returns generally move in tandem with the market's excess returns, with a fitted regression line that has a slope exceeding one. That slope suggests that the stock amplifies the market's movements in both directions. At the same time, the intercept of this line is very close to zero, indicating that after adjusting for the market effect, there's little sign of persistent outperformance or underperformance. According to CAPM, that intercept should indeed be zero for a fairly priced stock. The data here supports that principle: Microsoft's returns exhibit higher sensitivity to market swings, but overall remain in line with CAPM's prediction once you factor out those broader market influences.

Part d.

I created models under the assumption that alpha j = 0. Based on the output results of the code we get the following table:

Firm	Intercept	Beta	Adj. R-Squared
GE	-0.0012	0.8993	0.3846
Disney	-0.0011	0.8978	0.2831
GM	-0.0116	1.2614	0.2244
IBM	0.0059	1.1882	0.4000
MSFT	0.0061	1.3189	0.3360
XOM	0.0079	0.4140	0.1341

So nothing much really changed - either I messed up the code or the intercepts were by default closer to zero. I lean more into the assumption that actually everything is correct. Alpha in the context of CAPM basically represents the difference between an investments realized returns and the returns predicted by CAPM. In short it's the returns which couldn't be explained by broader market movements, and which can be attributed to pretty much anything outside of market forces.