MATH 584: Homework 1

Zachary Denis André Scialom
A20497400

08 February 2022

Contents

1 Problem 1

1.1 Question a

```
# Question 1a
price = stock_data['Adj Close'].values
ret = price[1:]/price[:-1] - 1
#print(price)
#print(len(ret[0]))
mu = np.mean(ret,axis=0)*250
sigma = np.cov(np.transpose(ret))*250
print(mu)
print()
print(sigma)
```

```
  0.33010831 0.36315891 0.09958138 0.19231409 0.11999506 0.12307365
  0.43144896 0.16072591 0.31297592 0.16617837 0.35193887 0.40950981
  0.25226074 0.32641948 0.18396519 0.63515061 0.14157472 0.30754092
  0.21307605 0.21696507 0.30254047 0.2468484  0.21097283 0.26757318]

 [[0.07888222 0.04827682 0.04601489 0.03116737 0.05621221 0.04423489
   0.04806016 0.04476252 0.02596319 0.03709853 0.03581335 0.02943105
   0.04497436 0.04361556 0.0441886  0.02542603 0.0544915  0.05885149
   0.05313928 0.04757556 0.04089112 0.06318821 0.03134419 0.04182681
   0.03767237 0.0400411  0.06477842 0.0446976  0.03452715 0.04345159]
  [0.04827682 0.09011503 0.05006362 0.03983992 0.06077523 0.05603833
   0.05412596 0.05625823 0.03400288 0.03870891 0.04092094 0.03731119
```

Figure 1: Estimated vector of mean returns and the beginning of the estimated covariance matrix

1.2 Question b

The Lagrangian can be used to identify the minimal-variance portfolio. By differentiating the Lagrangian, we obtain the following equation:

$$2\sum_{j=1}^{d}\Sigma_{ij}\alpha_j - \lambda = 0, i = 1, ..., d.$$

Moreover, the sum of all the weights must be equal to 1. The weights of the minimal-variance have been computed thanks to two different methods, both giving the same results (see the code). They only differ from the way the system $Ax = b$ is resolved.

```
[ 0.10913201 -0.10223863  0.0791295   0.02440598 -0.01039202 -0.05852095
  0.09320065 -0.02982642  0.26880867  0.08139421  0.01798879  0.11094012
 -0.04014705  0.02974064 -0.04708641  0.30246313 -0.04038867 -0.09756655
 -0.06704234 -0.02038566  0.00059143 -0.07822779  0.13729796  0.13669196
  0.02578848  0.01179626 -0.08245137  0.08673047  0.08895499  0.06921859
  0.05148106]
```

Figure 2: Weights of the minimal-variance portfolio

1.3 Question c

The weights of the optimal mean-variance portfolio with the coefficient of risk aversion $\gamma = 2$ can be computed by differentiating the Lagrangian. The following equation must be satisfied:

$$2\gamma\sum_{j=1}^{d}\Sigma_{ij}\alpha_j + \lambda = \mu^i, i = 1, ..., d.$$

The higher the risk-aversion coefficient, γ is, the more the investor dislikes risk. We can see that adding a risk aversion parameter utterly changes the whole trading strategy.

```
[ 0.43432442 -0.03994145 -0.34077097 -0.11072233 -0.14005743 -0.4482761
  0.12238648  0.50745363 -0.4379874   0.05854187 -0.3160308  -0.22803356
  0.48150713 -0.48835789  0.21478918  0.23125543  0.0775225   0.39235092
 -0.32079564  0.74150048 -0.02401557  0.61637477 -0.18722248  0.10913182
 -0.00122768  0.05628944 -0.12546055  0.24406012 -0.00658771 -0.07200062]
```

Figure 3: Weights of the optimal mean-variance portfolio

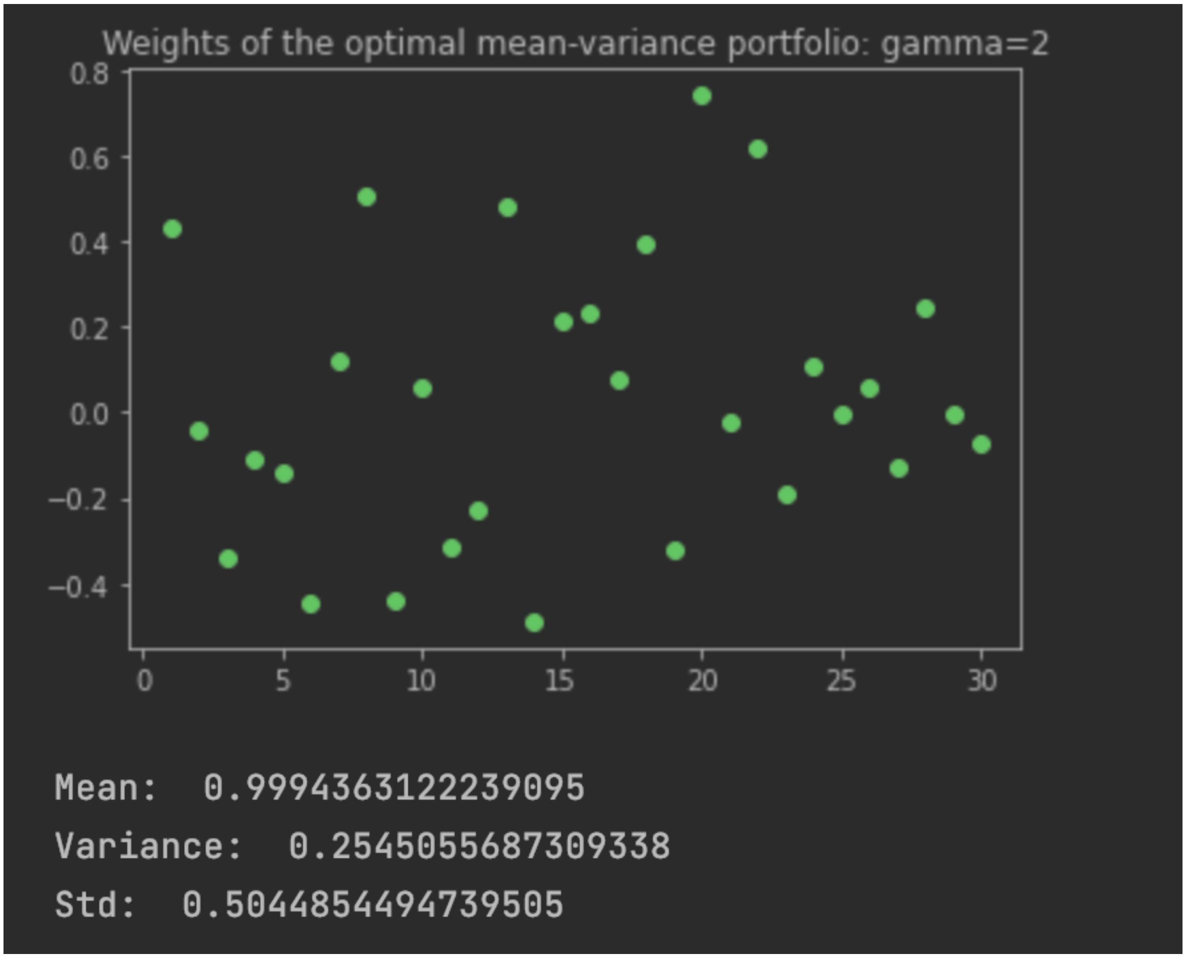

Figure 4: Weights of the optimal mean-variance portfolio with coefficient of risk aversion $\gamma = 2$.

Regarding the mean of the resulting portfolio, we can say that it has been a very good period for tech stocks as it is equal to 100%.

1.4 Question d

Adding robustness to the resulting portfolio can be done by penalizing the expected mean return in the process of maximization. In other words, it can be described as imagining a worst case scenario for the maximization. The weights of the optimal mean-variance portfolio in the robust setting can be derived by resolving the following problem:

$$max \sum_{i=1}^{d} (\mu^i \alpha^i - \epsilon^i |\alpha^i|) - \gamma \sum_{i,j=1}^{d} \alpha^i \alpha^j \Sigma_{ij}$$

subject to the sum of the weights equal to 1.
This optimization problem has been resolved using the following $scipy.optimize.minimize$ and the right constraints:

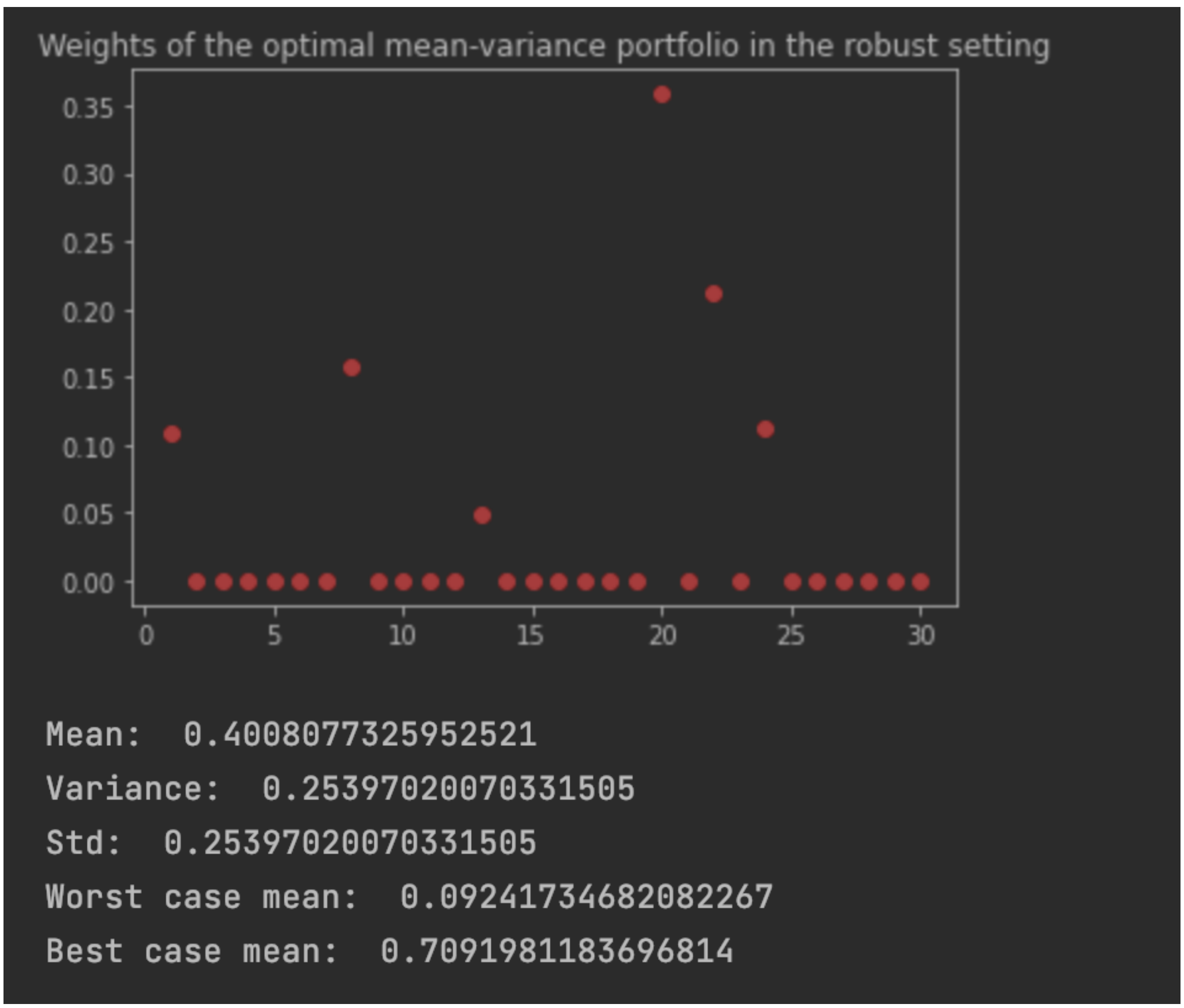

Figure 5: Weights of the optimal mean-variance portfolio in the robust setting

As expected, $\gamma = 2$, therefore the investor is not willing to take huge risks of investment. A high level of risk usually leads to high returns or big losses and vice-versa. This can be observed as the mean of the portfolio diminishes as it is equal to 0.40 (which is still good but far lower than 1.00).The best case scenario gives a mean equal to 0.70 and the worst case scenario mean is equal to 0.09.

What is interesting to see is among the thirty assets analyzed, only six of them are needed in the robust setting. Indeed, all the other weights are equal to zero. Potentially, a financial interpretation would be that with fewer traded assets, we have less possibilities to offset the risk in all states (by all states, I mean the randomness of the stock price (like a multi-state toy model)). The goal here is clearly to take into account the risk dimension, first with gamma but also with $\epsilon^i - |\alpha^i|$. Also, maybe there is a strong correlation between certain assets that lead to make their weight equal to zero. From what I have seen in Math.Finance I last semester, approximately 20 assets are needed to describe the S&P 500. Therefore, in a robust setting, we may want to not include all the assets to reinforce the idea of worst case scenario.

1.5 Question e

The efficient frontier can be computing by differentiating the Lagrangian and resolving the following equation:

$$2\sum_{j=1}^{d}\Sigma_{ij}\alpha_j - \lambda^1\mu^i - \lambda^2 = 0, i = 1, ..., d.$$

subject to $\sum_{i=1}^{d}\alpha^i\mu^i = \mu$ and the sum of the weights equal to 1.

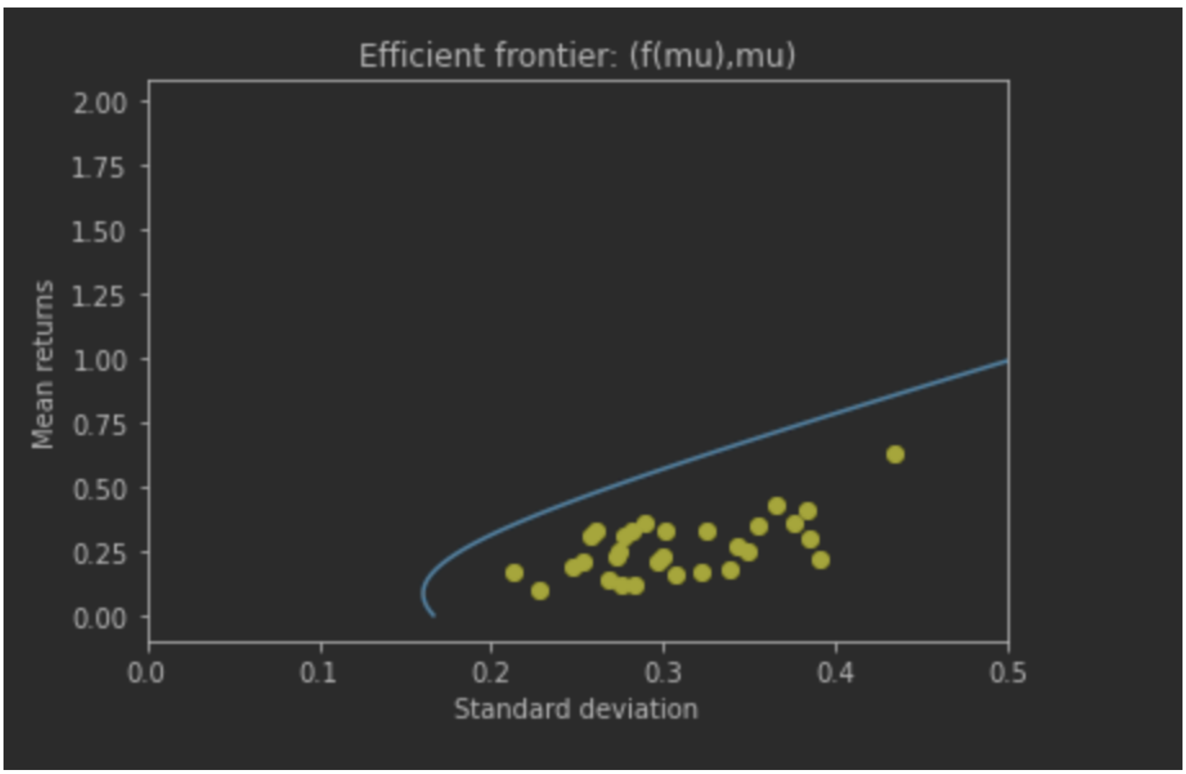

Figure 6: Efficient frontier: set $(\mathrm{f}(\mu), \mu)$

The efficient frontier in blue has a hyperbolic shape as expected. the pairs $(\sqrt{\Sigma_{ii}}, \mu_i)$ corresponding to the standard deviations and the means of the returns of individual basic assets are all lower than this frontier. This makes sense as if this was not the case, there would be a way to improve the mean and/or the standard deviation of the resulting portfolio.

1.6 Question f

A riskless asset has been added with a riskless return, $R = 0.01$. To compute the weights of the market portfolio (optimal fund theorem), the following system must be resolved:

$$2\sum_{j=1}^{d} \Sigma_{ij}\alpha_j = \lambda^1(\mu^i - R), i = 1, .., d$$

subject to the sum of the weights equal to 1.
I have added this riskless asset as the 31th asset. Therefore, new 31th row of zeros and 31th column of zeros have been added to the covariance matrix. The presence of a riskless asset implies that the efficient frontier can be described as a cone, and we only study the upper part of that one.

We need to construct:

$$\mu = \alpha R + (1 - \alpha)\mu^*$$

And,

$$\sigma = |1 - \alpha|\sigma^*$$

```
[ 0.5593099  -0.01599795 -0.50215683 -0.16265797 -0.18989345 -0.59807582
  0.13360385  0.71395353 -0.70963962  0.04975873 -0.44440894 -0.35831578
  0.68200134 -0.68748551  0.31543926  0.20388723  0.12284085  0.58064734
 -0.41832399  1.03432619 -0.03347311  0.8833405  -0.31194969  0.09853926
 -0.01161115  0.07339009 -0.14199083  0.30452869 -0.04330889 -0.12627725
  0.        ]
Mean:  1.3511334136225939
Std:  0.6812835033576166
Sharpe Ratio:  1.9685393922104284
```

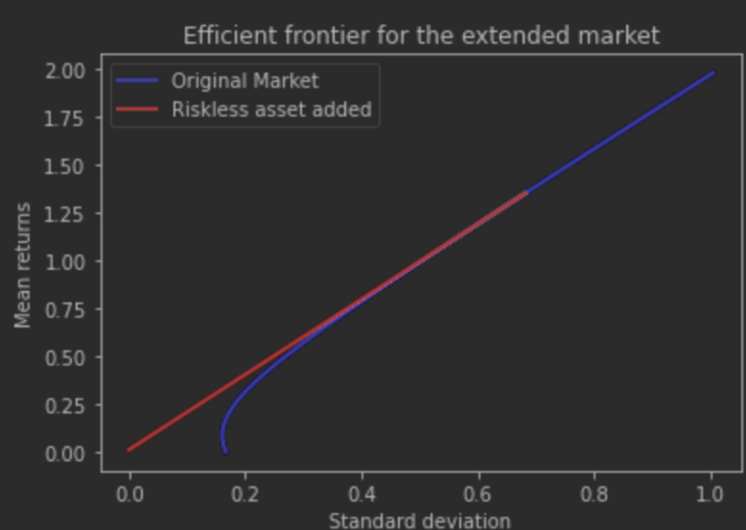

Figure 7: Efficient frontier for the extended market: riskless asset added

The mean of the market portfolio is equal to 1.35. This increase was expected as a riskless asset has been added. The standard deviation is equal to 0.68 and the Sharpe ratio, defined by $\frac{\mu^i - R}{\sqrt{\Sigma_{ii}}}$ is equal to 1.98. It is tough to comment on the value of the Sharpe ratio as another portfolio is needed to make a comparison to draw relevant conclusions.
The efficient frontier with a riskless asset is tangent with the efficient frontier without the riskless asset.

2 Problem 2

2.1 Question a

The beta of each asset have been computed according to the following CAPM formula:

$$\beta^i = \frac{\sum_{j=1}^{d} \Sigma_{ij} \alpha_{*,j}}{\sum_{k,j=1}^{d} \Sigma_{kj} \alpha_{*,k} \alpha_{*,j}}$$

Generally, β describes the reaction of an asset regarding the market portfolio. It enables to compare the covariance between the market and an asset with the volatility of this asset.

```
[0.23647673 0.23521741 0.16069592 0.11644725 0.25945452 0.16271178
 0.23868491 0.26332869 0.06679528 0.13594031 0.08201649 0.08431201
 0.31424835 0.11238696 0.22591035 0.11645252 0.2549626  0.29788969
 0.1806388  0.23593438 0.12971505 0.46613603 0.09810711 0.22185781
 0.15142121 0.15432102 0.21812928 0.17660316 0.14985298 0.19205634]

[<matplotlib.lines.Line2D at 0x124eb9ed0>]
```

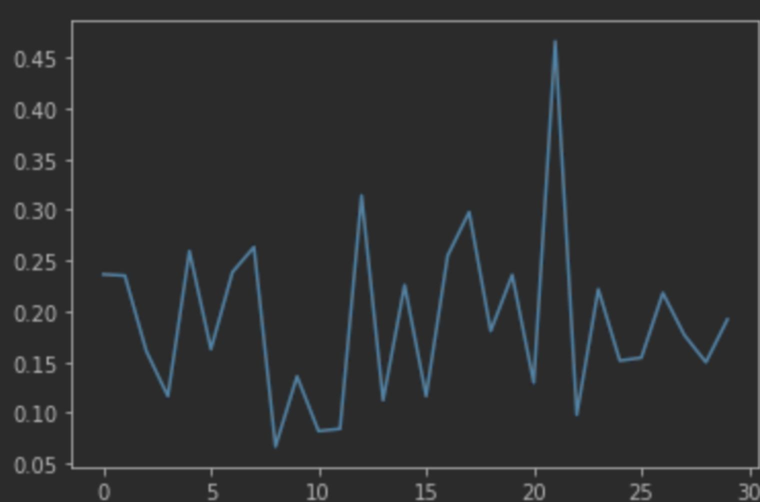

Figure 8: β_i for each basic asset, according to the CAPM formula

2.2 Question b

To regress the excess returns of the individual basic assets on the excess return of the market portfolio, a linear regression model has been used thanks to the function $scipy.stats.linregress()$. The linear regression model is the following:

$$r^i = a^i + r_M b^i + \epsilon^i, i = 1, ..., d$$

where $r^i = R^i - R$.
For a right regression, it was important to divide R by 250 while computing the excess return of the market portfolio as well as the one of each basic asset.

What can be seen is that the values of b^i are exactly equal to the values of β^i previously obtained. On the other hand, the values for alpha are close to zero (10^-19). This is the case when CAPM model is satisfied, a^i.

Question: However, among the requirements made in the course for CAPM, we need the weights to be positive, which is not the case here but CAPM is still verified (so I do not know if it is normal).

```
[0.23647673029058344, 0.23521741236842286, 0.1606959223633139, 0.11644724939760104, 0.25945452109292366, 0.162711780961113, 0.23868490780782814,
0.26332869016378296, 0.06679527628752935, 0.13594030672091834, 0.08201649265024835, 0.08431200943591637, 0.314248348738918, 0.11238696197492676,
0.22591035345196658, 0.11645252346967454, 0.25496260280937877, 0.2978896878902315, 0.1806388034372293, 0.235934379762815, 0.12971504867817762, 0
.4661360304357576, 0.09810710896854077, 0.2218578069031663, 0.1514212130621719, 0.1543210162548198, 0.21812928358186312, 0.1766031605553382, 0
.1498529766473073, 0.19205634300269572]

[4.336808689942018e-19, 6.505213034913027e-19, 1.3010426069826053e-18, 5.421010862427522e-19, 2.168404344971009e-19, 8.673617379884035e-19, 2
.168404344971009e-19, 8.673617379884035e-19, -1.0842021724855044e-19, 3.2526065174565133e-19, 5.421010862427522e-19, 3.2526065174565133e-19, -8
.673617379884035e-19, 6.505213034913027e-19, 0.0, 4.336808689942018e-19, -6.505213034913027e-19, 1.3010426069826053e-18, 4.336808689942018e-19,
6.505213034913027e-19, 2.168404344971009e-19, 1.3010426069826053e-18, 1.0842021724855044e-18, 6.505213034913027e-19, 5.421010862427522e-19, 6
.505213034913027e-19, -4.336808689942018e-19, 2.168404344971009e-19, -5.421010862427522e-19, 4.336808689942018e-19]
```

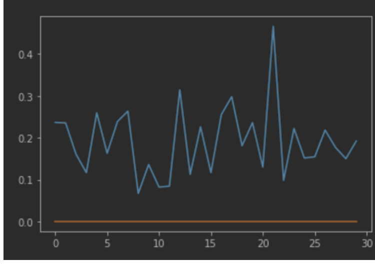

Figure 9: Results obtained with the linear regression model

2.3 Question c

For this question, two methods have been realized.
The first one consists in deriving the annual returns with the following expression: $\frac{Adjusted price at the end of the year}{Adjusted price at the beginning} - 1$.

To do so, I managed to operate on each year separately. Then, the linear regression has been realized:

$$\tilde{R_s^i} = b^i + c^i F_s^i + \tilde{\epsilon_s^i}$$

By considering the predictive factors F_s^i, the PE ratio of asset i in year $s-1$ minus its PE ratio in year $s-2$, for s = 2014,...,2021 and $\tilde{R_s^i}$, the annual returns of asset i over year s.
Here are the results obtained with this method:

```
[0.9222040522969832, 0.10527798073416837, 0.29663714484042775, 0.5024556540264988, 0.2739529988523719, 0.7165387243983417, 0.5532308555441467,
0.5432615290264134, 0.28867060932381294, 0.737747069183409, 0.612660643072991, 0.7525539381528386, 0.023833117913823712, 0.5099735972040633, 0
.8902126905106493, 0.4240440056973373, 0.9201132512533259, 0.5506164336597379, 0.052324609568533145, 0.7750498533986887, 0.036635524559044175,
0.16675009869376345, 0.3719064441630569, 0.7274699240101872, 0.12354536402113488, 0.36607889564651686, 0.8242267221809151, 0.4542180868559281,
0.889491208030532, 0.3861362620383647]

[0.04153893051223012, -0.6141263678439053, -0.4228165692845655, 0.2795850135288229, -0.44110250664337947, -0.15357709877616446, 0.24829287206037545,
0.25435100318506704, -0.4291581267110601, 0.14175580952279468, 0.21292944834203306, -0.13355072507780058, 0.7751815897852433, -0.27487998717572976,
-0.0586878501085586, -0.330443213805677, -0.042657987685441885, -0.2498777987379935, -0.7018531004378693, 0.12115373917606094, -0.7378823732440803,
0.5403841132533983, -0.3664734726926574, 0.14747364080196032, -0.5901490373997874, 0.37063535329513697, -0.09430335744301828, 0.31045516619396935,
-0.0590753228848282, 0.35643164926469384]
```

Figure 10: Method 1: p-values and R-values

```
[0.9222040522969832, 0.10527798073416837, 0.29663714484042775, 0.5024556540264988, 0.2739529988523719, 0.7165387243983417, 0.5532308555441467,
0.5432615290264134, 0.28867060932381294, 0.737747069183409, 0.612660643072991, 0.7525539381528386, 0.023833117913823712, 0.5099735972040633, 0
.8902126905106493, 0.4240440056973373, 0.9201132512533259, 0.5506164336597379, 0.052324609568533145, 0.7750498533986887, 0.036635524559044175,
0.16675009869376345, 0.3719064441630569, 0.7274699240101872, 0.12354536402113488, 0.36607889564651686, 0.8242267221809151, 0.4542180868559281,
0.889491208030532, 0.3861362620383647]

[0.04153893051223012, -0.6141263678439053, -0.4228165692845655, 0.2795850135288229, -0.44110250664337947, -0.15357709877616446, 0.24829287206037545,
0.25435100318506704, -0.4291581267110601, 0.14175580952279468, 0.21292944834203306, -0.13355072507780058, 0.7751815897852433, -0.27487998717572976,
-0.0586878501085586, -0.330443213805677, -0.042657987685441885, -0.2498777987379935, -0.7018531004378693, 0.12115373917606094, -0.7378823732440803,
0.5403841132533983, -0.3664734726926574, 0.14747364080196032, -0.5901490373997874, 0.37063535329513697, -0.09430335744301828, 0.31045516619396935,
-0.0590753228848282, 0.35643164926469384]
```

Figure 11: Method 1: b^i and c^i coefficients

Obviously, to see if the model fits, the p-values must be as low as possible (close to zero) and the R-squared values must be close to 1. Here, it is difficult to draw a general conclusion as the model does not fit well for some stocks but it does for others.
This is why I decided to do a second method which is essentially inspired from the file "UseFactors" available on Blackboard.

Here are the results obtained with this method:

```
P-values:
[0.6198514318832913, 0.5128107807235596, 0.32041654343769915, 0.7800240646561727, 0.9531706258730998, 0.20041144152188423, 0.46064543721296003,
0.9694014607703788, 0.738870090542086, 0.2782467523206106, 0.11032601470278415, 0.3108779577144735, 0.17090487973251284, 0.8159675545473654, 0
.8113903785745862, 0.810491006166814, 0.017709722337945866, 0.4283904153832262, 0.24528729501080435, 0.9359512224307707, 0.23631139178487237,
0.7440568320115344, 0.7115324126912522, 0.942740618990563, 0.1110512515126783, 0.5303629944029553, 0.16145035057032378, 0.5806271435443262, 0
.4352140818681612, 0.37410933361466514]

R-values:
[0.20872926018212462, -0.27311121458284554, 0.4043593767114439, 0.11842301788050966, -0.024986063509088868, 0.5063300590694774, 0.3062689821645321,
0.016322119625322557, -0.14113216043206148, 0.437585975340732, -0.6072789575605836, 0.4116817909095975, -0.5359927583286518, 0.09879154397757625,
0.10128230966068716, 0.10177201854037143, -0.7975988489755065, -0.32752829958172475, -0.4653177811105892, -0.034185973692575096, -0.4731902725736892,
-0.13825468684303482, 0.15638016841728503, 0.030557353199560662, -0.6063099690069003, -0.26224902453412763, -0.5460725879552911, 0.2318390298349844,
```

Figure 12: Method 2: p-values and R-values

```
bi coefficients:
[0.3038230585515005, 0.33911487326543077, 0.1592518569240448, 0.18496018438153014, 0.344765830868132, 0.5492662243280325, 0.30049243607670806,
0.252513804655289 7, 0.3277837094362147, 0.46262047079865665, 0.25458615934126033, 0.4692115114354413, 0.12510233281042127, 0.33760397711388507,
0.2536600840461327, 0.264640763949938, 0.22749682609641345, 0.26574476494424404, 0.32036543607671997, 0.31260064759241024, 0.38676166853453214,
0.42395211279581774, 0.08688189585553319, 0.20554847116601366, 0.20424207428582275, 0.10282469023511313, 0.15381759976843815, 0.18341756905364737,
0.14660533421461813, 0.1696300775164264]

ci coeffifients:
[0.7512000850363888, -0.3175926245185605, 2.073393106812312, 0.02086513942132238, -0.007465039553161561, 1.0523653688149874, 0.7343167652859967,
0.06016719273111301, -0.1855375427397, 0.8148104175516743, -0.7114035984862449, 1.1873611961831143, -1.2695152773931189, 0.4228671104026754, 0
.004659742534915876, 0.14529201468171832, -1.4472959689689864, -0.5786076719289011, -0.30861989018032643, -0.1689084047531726, -0.6836847333811517,
-0.00988632106057315, 0.7928876501646701, 0.093997340667704, -0.4036313268673068, -0.1294901136319213, -1.267152675512355, 0.3543212904511257,
```

Figure 13: Method 2: b^i and c^i coefficients

Thus, the results seem to get better as more stocks display a lower p-value and a higher R-square value but still, for some stocks, the model can be not convincing.
For annual returns, I would expect the regression to work pretty well (as it does for some stocks) especially because both variables $\tilde{R}_s^i$ and F_s^i are observed on the same frequency (annual).

2.4 Question d

At last, the goal was to compute the weights of the market portfolio by using predictive factors (P/E ratio) and daily returns. A linear regression cannot be done directly because the changes of $F^i_{s(t)}$ are done at a lower frequency than the changes of R^i_t(daily). This is why residual daily returns have been used:

$$\epsilon^i_t = R^i_t - (b^i + c^i F^i_{s(t)})/250$$

Then, by using the sample covariance matrix of the vectors $(\epsilon^1, ..., \epsilon^{30})_t$ as well as the vector of predicted mean returns in year 2014:

$$\tilde{\mu^i} = b^i + c^i F^i_{2014}$$

The weights of the market portfolio have been derived. This leads to a new mean equal to 1.73, a standard deviation equal to 0.68 and a Sharpe ratio equal to 2.53.

```
Mean:  1.731903837912697
Std:  0.6800601207546281
Sharpe ratio:  2.5319876660345675
Weights:
[ 2.66266842e-01  1.03933979e+00  4.97699484e-02 -1.34793651e-01
 -3.59713937e-02  6.38486255e-01  1.12702086e+00  3.42115875e-01
  3.10912846e-01 -1.86594462e+00 -3.37183923e-01 -1.99414329e-01
 -6.45431926e-01  1.05985006e+00 -1.63072831e-01 -6.20781826e-02
 -7.29671792e-01 -4.21736020e-01  2.04602065e-01  3.76806122e-02
  6.85879916e-01  4.10474948e-01 -2.89844200e-01  9.62359036e-02
  1.72996703e-01 -4.00955511e-01  9.92280913e-02 -8.16219262e-03
 -2.45030679e-01 -1.56946097e-03]
```

Figure 14: Results obtained using predictive factors

Therefore, if we compare those results to 1f, there is a significant improvement as the mean was equal to 1.35 and the Sharpe ratio was equal to 1.96.

This was expected. Indeed, those results have been obtained thanks to additional information, P/E ratios. Those idiosyncratic factors enable to diminish the specific risk (diversified away).
In fact, the term $(b^i + c^i F^i_{s(t)})/250$ is obtained thanks to a regression that is efficient (based on annual returns, same frequency, good predictive power) and that eventually leads to a good approximation of daily returns. Therefore, working on ϵ^i_t seems to be similar to working on a relative error. The standard deviation can be seen as the spread of a random variable, and a relative error can also be seen this way. Therefore, it makes sense to work on the matrix sigma with residual returns and it genuinely improves the strategy.
Obviously, it is a good sign to do better than the initial market portfolio and this can be understood as not all individuals invest optimally nor have extra factors to make their strategy more efficient.

Yet, there is something I have not understood which the choice made for the mean returns: why do we take the vector of predictive mean returns based on the P/E ratio of 2014 ? I tried to take other years and sometimes I got results that did not make sense (mean=15 for example)