MF796 Assignment 5

Shi Bo

U56082126

February 22, 2021

1: Covariance Matrix Decomposition

(Q1)

First, I used the SPY data from yahoo finance but the first 100 components of which the sequence follows the alphabet (From A to Z but actually to C!) Then, I tried to use ‘ffill’ method to fill NaN data out. There were some missing data since not all the companies can be traded at the beginning of 2016-02-22.

```
In [131]: print('The number of NaN after cleaning: \n', data.isna().sum())
The number of NaN after cleaning:
 A       0
AAL     0
AAP     0
AAPL    0
ABBV    0
       ..
LNT     0
MMM     0
MO      0
SCHW    0
T       0
Length: 101, dtype: int64
```

(Q2)

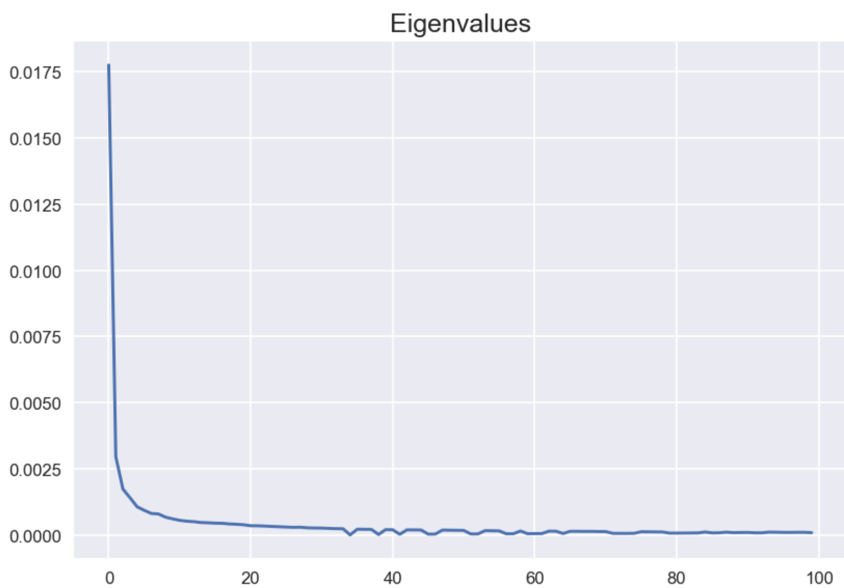
```
A sequence of daily log returns for each asset for each date:
                   A       AAL       AAP  ...        MO      SCHW         T
Date                                      ...
2016-02-23 -0.022873 -0.011327 -0.002166  ...  0.005052 -0.016849 -0.003261
2016-02-24  0.008306  0.006910  0.019264  ... -0.002604 -0.018825  0.009751
2016-02-25  0.003994  0.017070 -0.002263  ...  0.016487  0.017166  0.007252
2016-02-26 -0.001064 -0.011918 -0.000200  ... -0.013557  0.055312 -0.006443
2016-02-29 -0.006405  0.003176 -0.010788  ...  0.000487 -0.016235 -0.004860
...              ...       ...       ...  ...       ...       ...       ...
2021-02-11  0.014919 -0.024420 -0.046797  ... -0.005526  0.007358  0.004892
2021-02-12  0.008002  0.016346 -0.010740  ...  0.002076  0.039613  0.003827
2021-02-16 -0.000234  0.031350  0.017515  ...  0.000230  0.046340  0.005885
2021-02-17  0.016818  0.009495  0.039775  ...  0.004367  0.005726  0.020500
2021-02-18 -0.018930 -0.015687  0.034797  ...  0.000688  0.010225 -0.011565

[1258 rows x 100 columns]
```

(Q3)

Covariance matrix of daily returns:

	A	AAL	AAP	...	MO	SCHW	T
A	0.000274	0.000194	0.000113	...	0.000084	0.000189	0.000095
AAL	0.000194	0.001223	0.000220	...	0.000162	0.000359	0.000181
AAP	0.000113	0.000220	0.000506	...	0.000102	0.000175	0.000121
AAPL	0.000161	0.000182	0.000125	...	0.000096	0.000165	0.000100
ABBV	0.000135	0.000126	0.000110	...	0.000074	0.000135	0.000093
...	...	...	...	...	...	...	...
LNT	0.000081	0.000094	0.000100	...	0.000095	0.000056	0.000098
MMM	0.000131	0.000218	0.000117	...	0.000106	0.000175	0.000111
MO	0.000084	0.000162	0.000102	...	0.000241	0.000103	0.000111
SCHW	0.000189	0.000359	0.000175	...	0.000103	0.000481	0.000148
T	0.000095	0.000181	0.000121	...	0.000111	0.000148	0.000212

[100 rows x 100 columns]



After eigenvalue decomposition of the covariance matrix, there were no positive eigenvalues here. If so, there must be some data that was wrong.

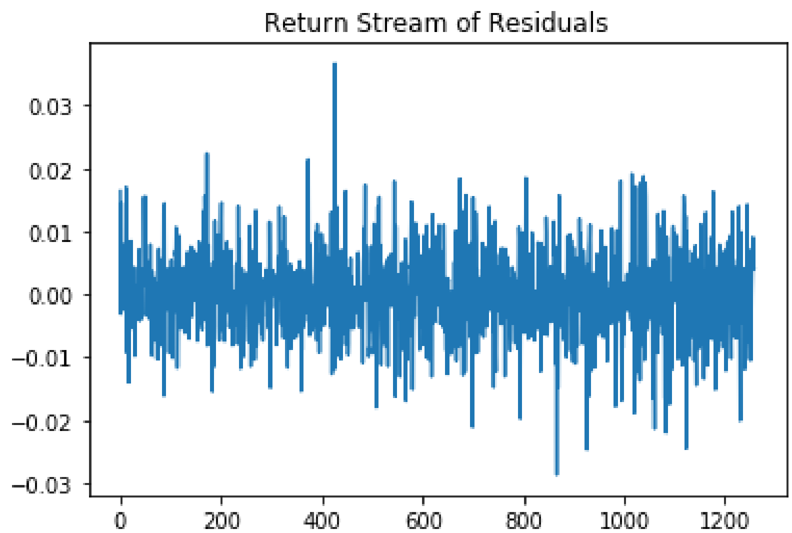
(Q4)

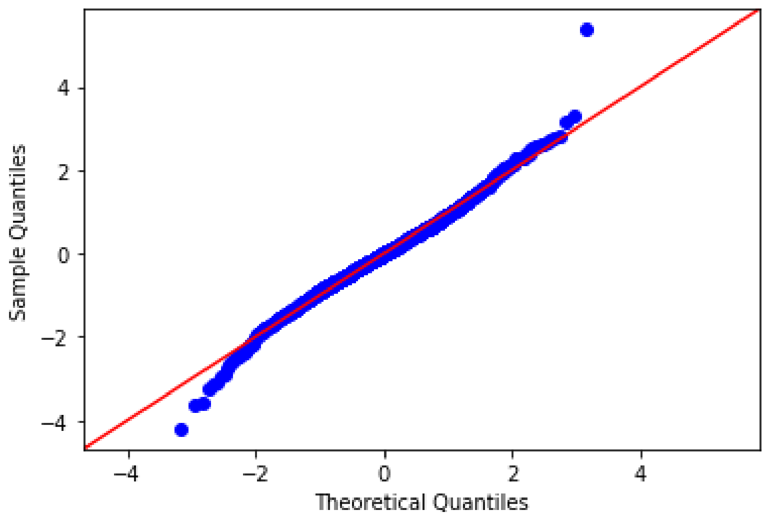
The number of eigenvalues that are required to account for 50% of the variance is 2.

The number of eigenvalues that are required to account for 90% of the variance is 47.

It makes sense according to principal component analysis (PCA).

(Q5)





From the above figure, we could see that the residual return that remove the impact of top 90% eigenvalues presents a smaller volatility compared with the original return and its mean is almost zero which implies that it could be i.i.d and makes the model stable.

2: Portfolio Construction

(Q1)

$$L(\omega,\lambda) = \langle R,\omega\rangle - a\langle\omega, C\omega\rangle - \sum_{k=1}^{K}\lambda_k\left(\left\langle \vec{g}_k,\omega\right\rangle - c_k\right) \qquad (1)$$

We can rewrite it in matrix form by defining G to be the matrix whose rows are the vectors $\vec{g}_k$: $\vec{g}_1$ is the first row of G, $\vec{g}_2$ is the second, and so on. Also, let $\lambda = (\lambda_1 \,.\,.\,.\,, \lambda_k)$ be the vector of Lagrange multipliers, and c = (c1, . . . , cK) the vector of constraint values. Then we can rewrite (1) in matrix form:

$$L(\omega,\lambda) = \langle R,\omega\rangle - a\langle\omega, C\omega\rangle - \langle\lambda, G\omega - c\rangle \qquad (2)$$

1. Calculate the gradients $\nabla_\omega L$ and $\nabla_\lambda L$. (Hint: $\langle\lambda, G\omega\rangle = \langle G^T\lambda,\omega\rangle$).

$\nabla_\omega L = R - 2\mathrm{a}C\omega - G^T\lambda$

$\nabla_\lambda L = G\omega - c$

2. Use $\nabla_\omega L = 0$ to solve for w in terms of λ and substitute the result into the equation $\nabla_\lambda L = 0$. This gives equation for λ, which for me came out to be

$$GC^{-1}G^T\lambda = GC^{-1}R - 2\mathrm{ac} \qquad (3)$$

where C is the covariance matrix and c is the vector of constraint values. Please don't confuse them.

let $\nabla_\omega L = 0$, $\omega = \dfrac{1}{2a}C^{-1}(R - G^T\lambda)$

substitute into $\nabla_\lambda L = 0$. $\lambda = \left(GC^{-1}G^T\right)^{-1}\left(GC^{-1}R - 2\mathrm{ac}\right)$

The G matrix here is :

```
In [136]: G
Out[136]:
matrix([[1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
         1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
         1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
         1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
         1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
         1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
         1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
         1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
         0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
         0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
         0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
         0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
         0., 0., 0., 0.]])
```

$$GC^{-1}G^T = \begin{matrix} 4.96367394e-05 & -1.36920518e-06 \\ -1.36920518e-06 & 6.79241207e-06 \end{matrix}$$

(Q2)

It might not be acceptable to most mutual funds since some elements of w is negative and it means short position of stocks.

Hence, if stock is not short-able, we may set another constraint that all the element of w is non-negative.