# Project Week03

## Problem 1

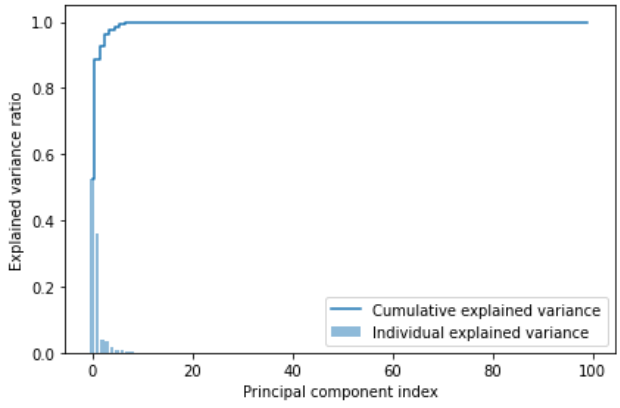
$\lambda = 0.5$

Exponentially weighted covariance matrix:

| | AAPL | MSFT | AMZN | ... | GM | TFC | TJX |
|---|---|---|---|---|---|---|---|
| AAPL | 0.000058 | 0.000088 | 0.000158 | ... | 1.352818e-05 | -0.000032 | -1.588678e-05 |
| MSFT | 0.000088 | 0.000140 | 0.000260 | ... | 1.329909e-05 | -0.000063 | -2.014420e-05 |
| AMZN | 0.000158 | 0.000260 | 0.000626 | ... | -2.969253e-05 | -0.000227 | 2.543016e-05 |
| TSLA | 0.000099 | 0.000161 | 0.000312 | ... | 1.576838e-05 | -0.000088 | -6.350284e-07 |
| GOOGL | 0.000273 | 0.000420 | 0.000936 | ... | -7.913438e-07 | -0.000311 | -1.184947e-05 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| LMT | 0.000024 | 0.000035 | 0.000042 | ... | 1.224781e-05 | -0.000002 | -2.737007e-05 |
| SYK | -0.000040 | -0.000029 | 0.000055 | ... | -3.957780e-05 | -0.000086 | 9.335525e-05 |
| GM | 0.000014 | 0.000013 | -0.000030 | ... | 4.310161e-05 | 0.000043 | -1.745585e-05 |
| TFC | -0.000032 | -0.000063 | -0.000227 | ... | 4.284140e-05 | 0.000150 | -4.091695e-05 |
| TJX | -0.000016 | -0.000020 | 0.000025 | ... | -1.745585e-05 | -0.000041 | 7.176188e-05 |

[100 rows x 100 columns]

PCA cumulative variance explained:



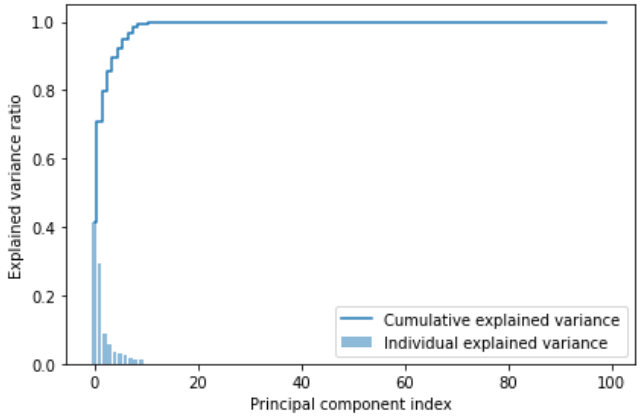

$\lambda = 0.7$

Exponentially weighted covariance matrix:

| | AAPL | MSFT | ... | TFC | TJX |
|---|---|---|---|---|---|
| AAPL | 6.369757e-05 | 0.000097 | ... | 7.275718e-08 | -7.201519e-06 |
| MSFT | 9.666339e-05 | 0.000169 | ... | -2.046466e-05 | -2.866199e-05 |
| AMZN | 1.533963e-04 | 0.000249 | ... | -1.520634e-04 | 8.492705e-07 |
| TSLA | 1.049676e-04 | 0.000177 | ... | -4.175581e-05 | -5.452518e-06 |
| GOOGL | 2.753656e-04 | 0.000449 | ... | -1.940872e-04 | -1.402611e-05 |
| ... | ... | ... | ... | ... | ... |
| LMT | 2.281140e-05 | 0.000040 | ... | -1.934380e-05 | -3.551386e-05 |
| SYK | -2.971833e-05 | -0.000029 | ... | -1.509945e-05 | 7.205655e-05 |
| GM | 3.277675e-05 | 0.000034 | ... | 1.076652e-04 | -4.229112e-07 |
| TFC | 7.275718e-08 | -0.000020 | ... | 2.920111e-04 | 3.102641e-05 |
| TJX | -7.201519e-06 | -0.000029 | ... | 3.102641e-05 | 1.332939e-04 |

[100 rows x 100 columns]

PCA cumulative variance explained:



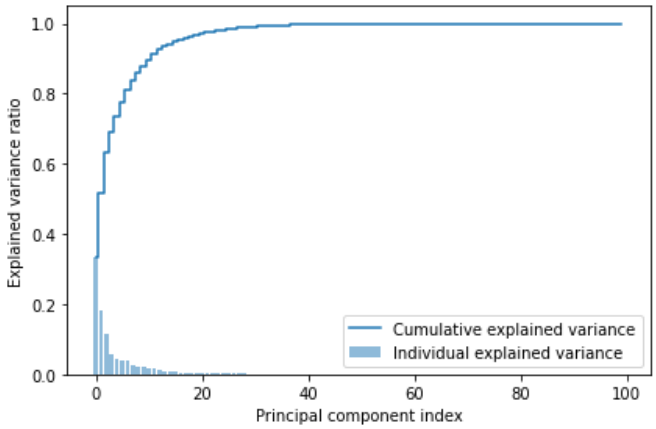

$\lambda = 0.9$

Exponentially weighted covariance matrix:

| | AAPL | MSFT | AMZN | ... | GM | TFC | TJX |
|---|---|---|---|---|---|---|---|
| AAPL | 0.000076 | 0.000111 | 0.000137 | ... | 0.000072 | 0.000089 | 0.000041 |
| MSFT | 0.000111 | 0.000237 | 0.000230 | ... | 0.000064 | 0.000090 | 0.000005 |
| AMZN | 0.000137 | 0.000230 | 0.000372 | ... | 0.000057 | 0.000008 | 0.000008 |
| TSLA | 0.000095 | 0.000177 | 0.000185 | ... | 0.000060 | 0.000027 | 0.000022 |
| GOOGL | 0.000269 | 0.000522 | 0.000581 | ... | 0.000185 | 0.000199 | 0.000122 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| LMT | 0.000014 | 0.000024 | 0.000030 | ... | 0.000011 | -0.000031 | -0.000033 |
| SYK | 0.000005 | -0.000007 | 0.000007 | ... | 0.000049 | 0.000083 | 0.000074 |
| GM | 0.000072 | 0.000064 | 0.000057 | ... | 0.000228 | 0.000238 | 0.000076 |
| TFC | 0.000089 | 0.000090 | 0.000008 | ... | 0.000238 | 0.000661 | 0.000230 |
| TJX | 0.000041 | 0.000005 | 0.000008 | ... | 0.000076 | 0.000230 | 0.000267 |

[100 rows x 100 columns]

PCA cumulative variance explained:



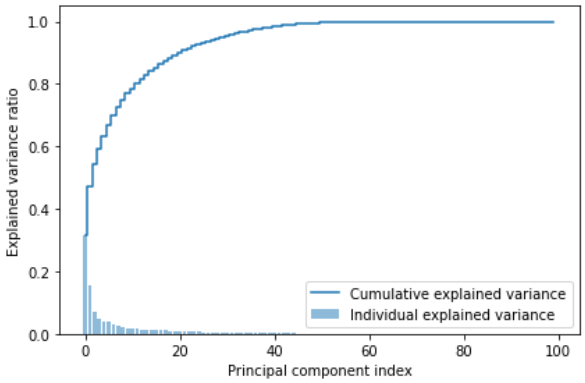

$\lambda = 0.97$

Exponentially weighted covariance matrix:

| | AAPL | MSFT | AMZN | ... | GM | TFC | TJX |
|---|---|---|---|---|---|---|---|
| AAPL | 0.000084 | 1.069457e-04 | 0.000122 | ... | 0.000090 | 0.000125 | 0.000081 |
| MSFT | 0.000107 | 2.687523e-04 | 0.000198 | ... | 0.000078 | 0.000116 | 0.000037 |
| AMZN | 0.000122 | 1.975317e-04 | 0.000291 | ... | 0.000089 | 0.000083 | 0.000033 |
| TSLA | 0.000093 | 1.711137e-04 | 0.000149 | ... | 0.000070 | 0.000035 | 0.000044 |
| GOOGL | 0.000231 | 4.327650e-04 | 0.000409 | ... | 0.000184 | 0.000330 | 0.000185 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| LMT | 0.000011 | 3.951258e-06 | -0.000002 | ... | 0.000010 | -0.000002 | 0.000005 |
| SYK | 0.000021 | 1.140771e-07 | 0.000010 | ... | 0.000056 | 0.000093 | 0.000067 |
| GM | 0.000090 | 7.844996e-05 | 0.000089 | ... | 0.000265 | 0.000220 | 0.000119 |
| TFC | 0.000125 | 1.156588e-04 | 0.000083 | ... | 0.000220 | 0.000748 | 0.000268 |
| TJX | 0.000081 | 3.749775e-05 | 0.000033 | ... | 0.000119 | 0.000268 | 0.000308 |

[100 rows x 100 columns]

PCA cumulative variance explained:



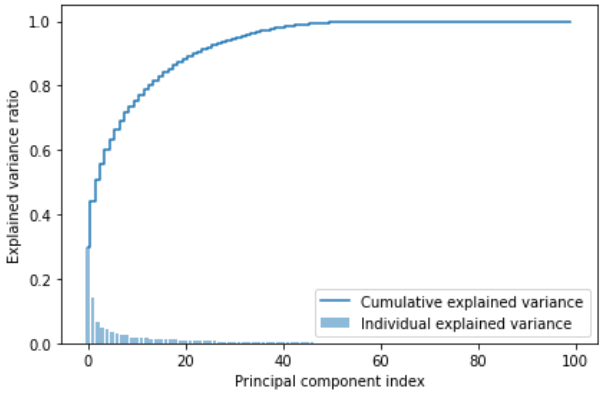
$\lambda = 0.99$

Exponentially weighted covariance matrix:

| | AAPL | MSFT | AMZN | ... | GM | TFC | TJX |
|---|---|---|---|---|---|---|---|
| AAPL | 0.000081 | 9.685917e-05 | 0.000108 | ... | 0.000088 | 0.000123 | 0.000085 |
| MSFT | 0.000097 | 2.612496e-04 | 0.000170 | ... | 0.000078 | 0.000102 | 0.000042 |
| AMZN | 0.000108 | 1.700454e-04 | 0.000264 | ... | 0.000086 | 0.000079 | 0.000030 |
| TSLA | 0.000088 | 1.650178e-04 | 0.000135 | ... | 0.000069 | 0.000029 | 0.000041 |
| GOOGL | 0.000204 | 3.592000e-04 | 0.000335 | ... | 0.000154 | 0.000323 | 0.000181 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| LMT | 0.000010 | 4.024847e-06 | -0.000017 | ... | 0.000013 | 0.000012 | 0.000021 |
| SYK | 0.000023 | -7.142292e-07 | 0.000004 | ... | 0.000050 | 0.000093 | 0.000061 |
| GM | 0.000088 | 7.766248e-05 | 0.000086 | ... | 0.000261 | 0.000194 | 0.000123 |
| TFC | 0.000123 | 1.021598e-04 | 0.000079 | ... | 0.000194 | 0.000733 | 0.000248 |
| TJX | 0.000085 | 4.205904e-05 | 0.000030 | ... | 0.000123 | 0.000248 | 0.000303 |

[100 rows x 100 columns]

PCA cumulative variance explained:




What does this tell us about values of $\lambda$ and the effect it has on the covariance matrix?

The smaller $\lambda$ is, the fewer observations that actually matter, the more weight it puts on the first principle component.

Problem 2

Implement Higham's 2002 nearest psd correlation function:

Compare run time:

```
Converged in 31 iterations.
```

```
n=500
Higham Took: 22.88 seconds
Near_PSD Took: 0.22 seconds
```

```
Distance near_psd() = 94969370.39
Distance higham_nearPSD() = 0.07
```

When N increases:

```
n = 1000
Converged in 60 iterations.
```

```
Higham Took: 228.30 seconds
Near_PSD Took: 0.97 seconds
```

When N increase, the run time difference between Higham and near_psd becomes larger: near_psd only increases 0.75 seconds, but Higham increases over 200 seconds. However the Frobenius Norm of Higham is significantly smaller than that of near_psd.

# Problem 3

Pearson correlation :

```
array([[1.        , 0.64575234, 0.71447293, ..., 0.50278815, 0.56059983,
        0.57718442],
       [0.64575234, 1.        , 0.60820213, ..., 0.21540239, 0.15653013,
        0.29330838],
       [0.71447293, 0.60820213, 1.        , ..., 0.16569986, 0.09000614,
        0.24738206],
       ...,
       [0.50278815, 0.21540239, 0.16569986, ..., 1.        , 0.50353749,
        0.4319246 ],
       [0.56059983, 0.15653013, 0.09000614, ..., 0.50353749, 1.        ,
        0.32431423],
       [0.57718442, 0.29330838, 0.24738206, ..., 0.4319246 , 0.32431423,
        1.        ]])
```

Pearson variance:

```
SPY     0.000077
AAPL    0.000253
MSFT    0.000251
AMZN    0.000256
TSLA    0.002018
           ...
LMT     0.000308
SYK     0.000256
GM      0.000718
TFC     0.000296
TJX     0.000273
```

EW correlation:

```
            SPY      AAPL      MSFT  ...        GM       TFC       TJX
59 SPY   1.000000  0.710633  0.778326  ...  0.501009  0.514452  0.578360
   AAPL  0.710633  1.000000  0.705569  ...  0.258490  0.138101  0.306037
   MSFT  0.778326  0.705569  1.000000  ...  0.179253  0.125740  0.257065
   AMZN  0.660029  0.682885  0.570939  ...  0.085151  0.173785  0.410929
   TSLA  0.560856  0.588838  0.533351  ...  0.270013  0.244344  0.145091
...           ...       ...       ...  ...       ...       ...       ...
   LMT   0.184400  0.007880  0.059033  ...  0.257829  0.273345  0.157703
   SYK   0.609059  0.297169  0.329218  ...  0.495279  0.416505  0.521769
   GM    0.501009  0.258490  0.179253  ...  1.000000  0.563080  0.455820
   TFC   0.514452  0.138101  0.125740  ...  0.563080  1.000000  0.306881
   TJX   0.578360  0.306037  0.257065  ...  0.455820  0.306881  1.000000

[101 rows x 101 columns]
```

EW variance :

```
SPY     0.000086
AAPL    0.000274
MSFT    0.000294
AMZN    0.000237
TSLA    0.002040
          ...
LMT     0.000176
SYK     0.000270
GM      0.000764
TFC     0.000310
TJX     0.000265
```

4 different covariance matrices:

covariance matrix 1 :Pearson correlation + var()

```
array([[7.71482917e-05, 9.02467314e-05, 9.93342377e-05, ...,
        1.18295582e-04, 8.47623427e-05, 8.37341979e-05],
       [9.02467314e-05, 2.53165588e-04, 1.53179376e-04, ...,
        9.18064504e-05, 4.28733187e-05, 7.70818109e-05],
       [9.93342377e-05, 1.53179376e-04, 2.50553256e-04, ...,
        7.02574703e-05, 2.45249979e-05, 6.46760280e-05],
       ...,
       [1.18295582e-04, 9.18064504e-05, 7.02574703e-05, ...,
        7.17531002e-04, 2.32187575e-04, 1.91096784e-04],
       [8.47623427e-05, 4.28733187e-05, 2.45249979e-05, ...,
        2.32187575e-04, 2.96328665e-04, 9.22099840e-05],
       [8.37341979e-05, 7.70818109e-05, 6.46760280e-05, ...,
        1.91096784e-04, 9.22099840e-05, 2.72803673e-04]])
```

covariance matrix 2 :Pearson correlation + EW var()

```
array([[8.57359775e-05, 9.89397075e-05, 1.13464748e-04, ...,
        1.28678477e-04, 9.13663832e-05, 8.70103404e-05],
       [9.89397075e-05, 2.73808020e-04, 1.72609477e-04, ...,
        9.85173980e-05, 4.55903509e-05, 7.90173141e-05],
       [1.13464748e-04, 1.72609477e-04, 2.94162453e-04, ...,
        7.85516103e-05, 2.71717531e-05, 6.90774891e-05],
       ...,
       [1.28678477e-04, 9.85173980e-05, 7.85516103e-05, ...,
        7.63974127e-04, 2.44975725e-04, 1.94366752e-04],
       [9.13663832e-05, 4.55903509e-05, 2.71717531e-05, ...,
        2.44975725e-04, 3.09815997e-04, 9.29378544e-05],
       [8.70103404e-05, 7.90173141e-05, 6.90774891e-05, ...,
        1.94366752e-04, 9.29378544e-05, 2.65063196e-04]])
```

covariance matrix 3 : EW correlation + var()

```
array([[7.71482917e-05, 9.93141168e-05, 1.08211783e-04, ...,
        1.17877044e-04, 7.77848655e-05, 8.39048125e-05],
       [9.93141168e-05, 2.53165588e-04, 1.77701690e-04, ...,
        1.10170722e-04, 3.78254905e-05, 8.04268498e-05],
       [1.08211783e-04, 1.77701690e-04, 2.50553256e-04, ...,
        7.60040800e-05, 3.42618923e-05, 6.72074559e-05],
       ...,
       [1.17877044e-04, 1.10170722e-04, 7.60040800e-05, ...,
        7.17531002e-04, 2.59643561e-04, 2.01668759e-04],
       [7.77848655e-05, 3.78254905e-05, 3.42618923e-05, ...,
        2.59643561e-04, 2.96328665e-04, 8.72534547e-05],
       [8.39048125e-05, 8.04268498e-05, 6.72074559e-05, ...,
        2.01668759e-04, 8.72534547e-05, 2.72803673e-04]])
```

covariance matrix 4 :EW correlation + EW var()

```
array([[8.57359775e-05, 1.08880505e-04, 1.23605143e-04, ...,
        1.28223203e-04, 8.38452737e-05, 8.71876303e-05],
       [1.08880505e-04, 2.73808020e-04, 2.00242334e-04, ...,
        1.18224077e-04, 4.02226242e-05, 8.24463460e-05],
       [1.23605143e-04, 2.00242334e-04, 2.94162453e-04, ...,
        8.49766275e-05, 3.79594601e-05, 7.17811906e-05],
       ...,
       [1.28223203e-04, 1.18224077e-04, 8.49766275e-05, ...,
        7.63974127e-04, 2.73943899e-04, 2.05119630e-04],
       [8.38452737e-05, 4.02226242e-05, 3.79594601e-05, ...,
        2.73943899e-04, 3.09815997e-04, 8.79422002e-05],
       [8.71876303e-05, 8.24463460e-05, 7.17811906e-05, ...,
        2.05119630e-04, 8.79422002e-05, 2.65063196e-04]])
```

PCA simulate:

1. cov1_df: Pearson correlation + var()

```
Simulating with 59 PC Factors: 100%  total variance explained
sim1 took: 0.0657410622 seconds
Norms 1 is 0.00028
Simulating with 11 PC Factors: 74%  total variance explained
sim2 took: 0.0153689384 seconds
Norms 2 is 0.00027
Simulating with 12 PC Factors: 76%  total variance explained
sim3 took: 0.0158350468 seconds
Norms 3 is 0.00021
Simulating with 4 PC Factors: 53%  total variance explained
sim4 took: 0.0079588890 seconds
Norms 4 is 0.00024
Simulating with 3 PC Factors: 48%  total variance explained
sim5 took: 0.0069370270 seconds
Norms 5 is 0.00021
```

2. cov2_df: Pearson correlation + EW var()

```
Simulating with 59 PC Factors: 99%  total variance explained
sim1 took: 0.0667240620 seconds
Norms 1 is 0.00027

Simulating with 12 PC Factors: 76%  total variance explained
sim2 took: 0.0181050301 seconds
Norms 2 is 0.00023

Simulating with 3 PC Factors: 49%  total variance explained
sim3 took: 0.0082480907 seconds
Norms 3 is 0.00023
```

3.cov3_df: EW correlation + var()

```
Simulating with 59 PC Factors: 100%  total variance explained
sim1 took: 0.0664336681 seconds
Norms 1 is 0.00027

Simulating with 10 PC Factors: 75%  total variance explained
sim2 took: 0.0159509182 seconds
Norms 2 is 0.00028

Simulating with 3 PC Factors: 53%  total variance explained
sim3 took: 0.0065071583 seconds
Norms 3 is 0.00028
```

4. cov4_df : EW correlation + EW var()

```
Simulating with 59 PC Factors: 100%  total variance explained
sim1 took: 0.1044352055 seconds
Norms 1 is 0.00030

Simulating with 10 PC Factors: 76%  total variance explained
sim2 took: 0.0151259899 seconds
Norms 2 is 0.00028

Simulating with 3 PC Factors: 53%  total variance explained
sim3 took: 0.0075919628 seconds
Norms 3 is 0.00032
```

The more principal component factors, the more total variance explained. The run times are larger and the Norm generally grow larger, but there are exceptions as well.