Fintech545 Homework 2

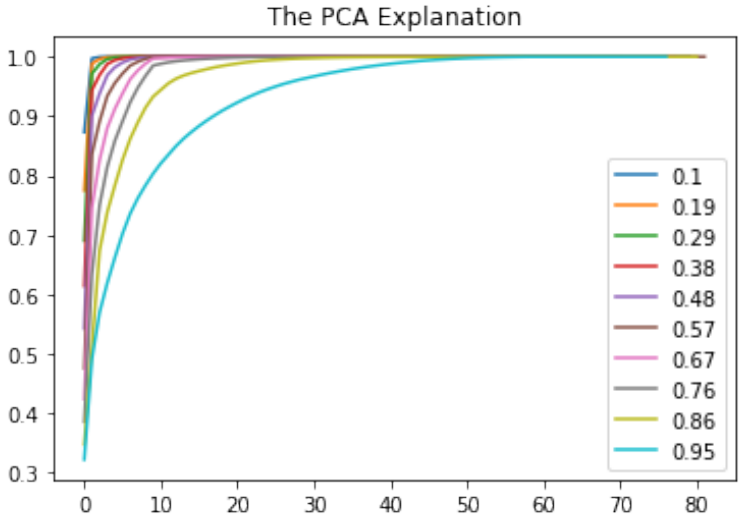
qw112 Qiyu Wu

Question 1.Exponentially Weighted Covariance Matrix

For the question, there are steps that have been done to make the weighted covariance matrix, assume the weight λ has been known:

- Read csv file data and adjust index and column name for the dataset;
- Generate weights for each time point (60 time points in total);
- Applied exponentially weighted covariance formula to generate 101×101 covariance matrix;

After finishing all the above steps, we can get one covariance matrix for each specific λ. Then, we can investigate the relationship between the explanationary power of PCA and the λ value.

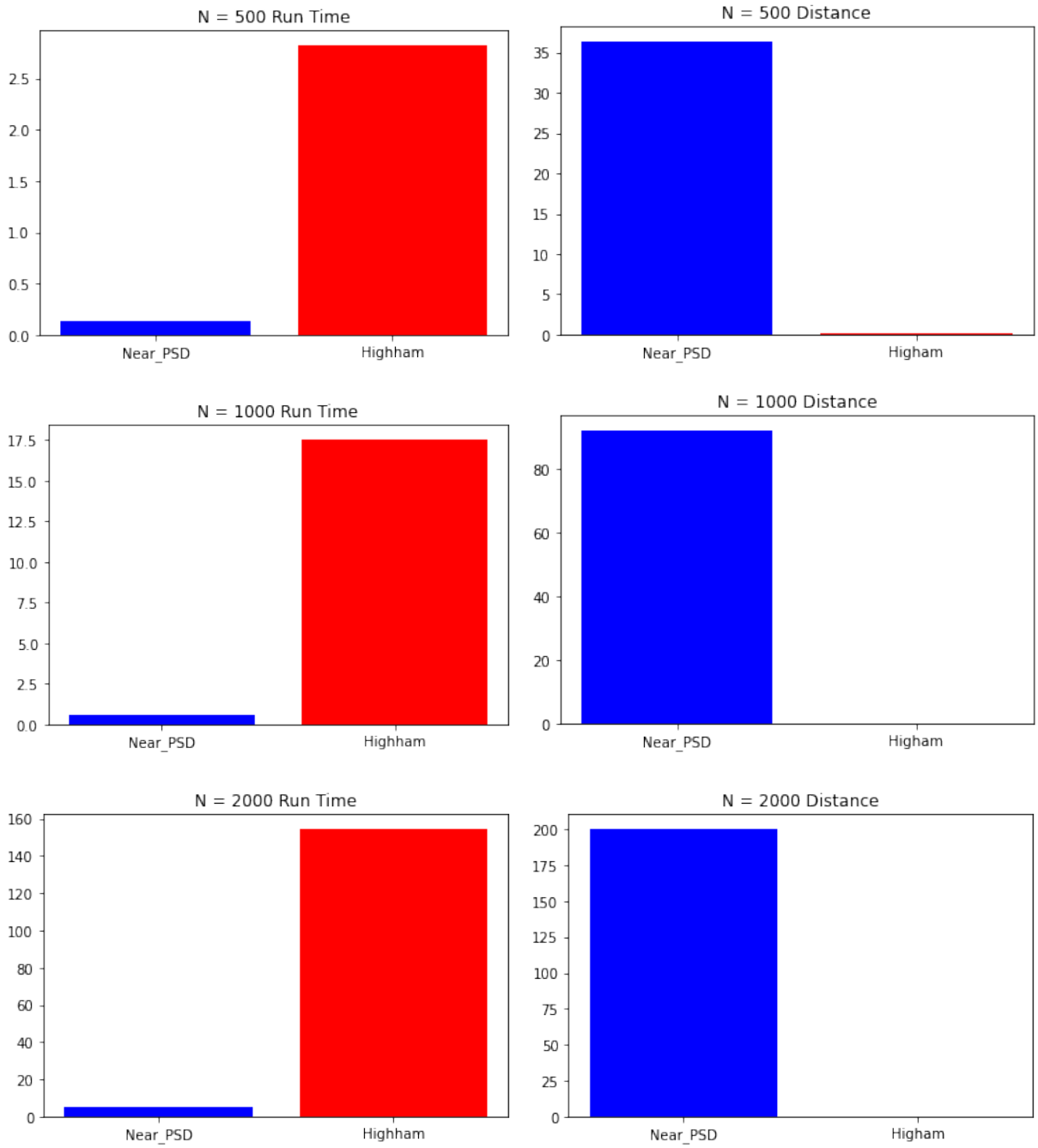


Conclusion:

- **Findings:** For the same n value, when λ approaches to 0, the less explanationary power will be.
- **Reasons:** Exponentially Weighed means that weights put on the time near to the current time, which also means the smaller time lag. So, when λ become smaller, it means that people put more consideration on the older time. In this way, We need more time point to generate high explanationary power in the PCA method.

Question 2. Methods to Deal with Non-Positive Semidefinite Matrix

Within the question, we try to use near_psd() and Higham method to fix a non-PSD Matrix. According to the new eigenvalue based on the fixed matrix, the covariance matrix is now PSD. Then, we would like to know more to compare the



Conclusion:

- From the general point of view, out of the accuracy needs, we need use the Higham's method to adjust non-psd matrix. It has a small distance difference under all cases.
- To improve the run time efficiency, the Near_PSD would be a better choice. It is consistently fast even when N is larger.

Question 3. Principle Component Analysis

Construct 4 Covariance Matrix

Here are Combinations about my construction:

- Pearson Correlation & Pearson Variance
- Exponentially Weighted Correlation & Exponentially Weighted Variance
- Pearson Correlation & Exponentially Weighted Variance
- Exponentially Weighed Correlation & Pearson variance

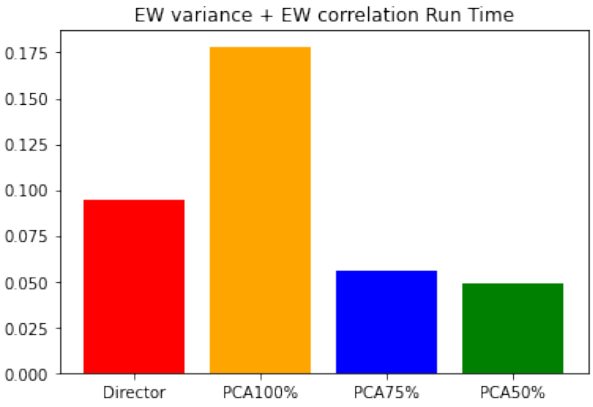
According to the 4 Covariance Matrix, We perform analysis about the 4 covariance matrix from 4 perspectives: Direct Simulation, and 100%, 75%, 50% PCA.

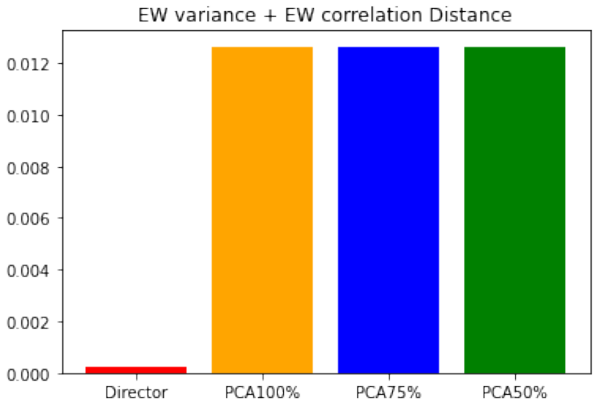
Visutalization Analysis

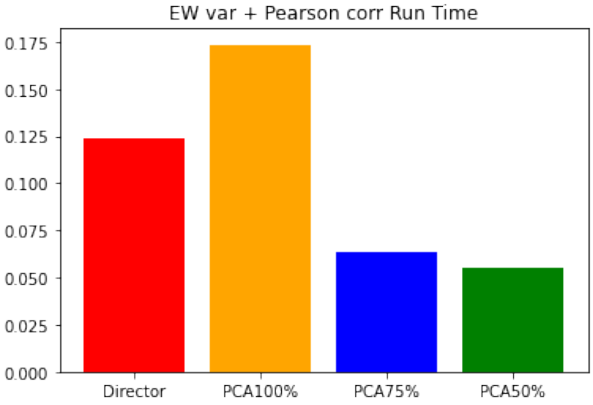
Here are visualization about their run time and distane between real covariance matrix and covariance matrix gain from simulation. Since the difference of distance between different PCA explanation requirements, we made a table for it.

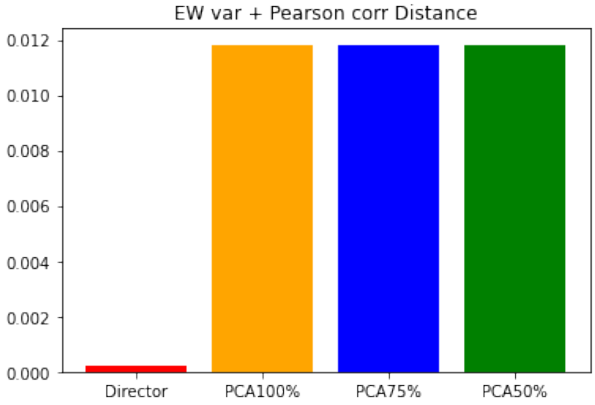
	Direct Simulation	PCA 100%	PCA 75%	PCA 50%
EW var + EW corr	0.00233200	0.01185840	0.01185803	0.01185922
Pearson var + corr	0.00029433	0.01187555	0.01187420	0.01187649
EW var + Pearson corr	0.00111533	0.01187173	0.01187169	0.01187397
Pearson var + EW corr	0.01078994	0.01196630	0.01196632	0.01196639

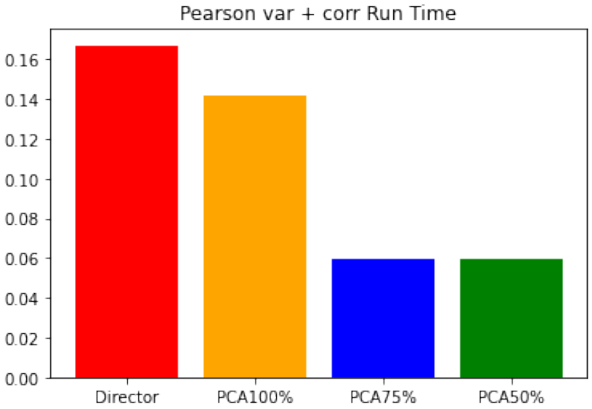
Figure 1: Distance Table of Covariance Matrix

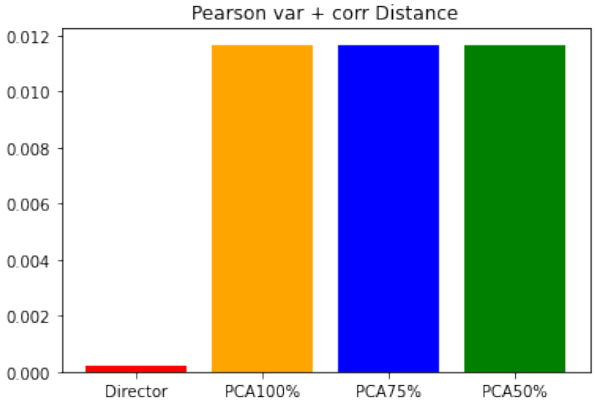

EW variance + EW correlation Run Time
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Director
PCA100%
PCA75%
PCA50%

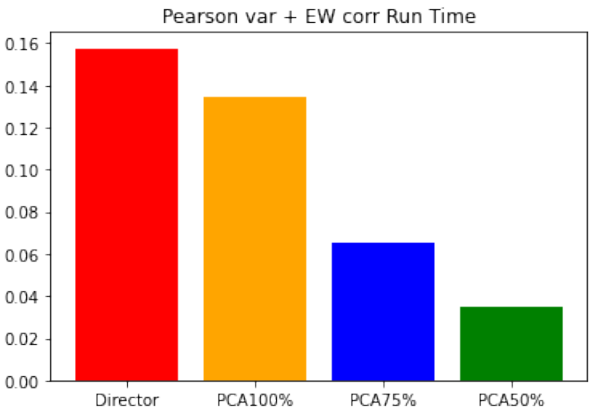

EW variance + EW correlation Distance
0.012
0.010
0.008
0.006
0.004
0.002
0.000
Director
PCA100%
PCA75%
PCA50%

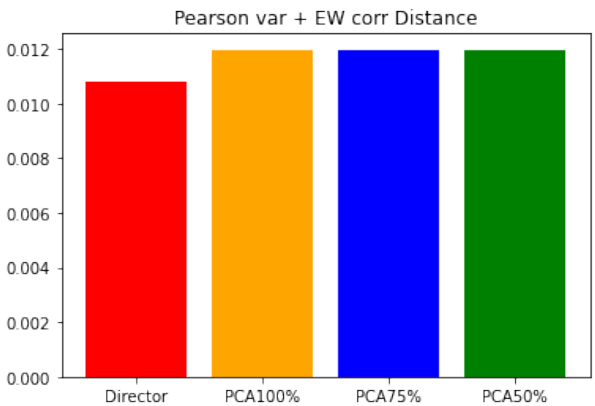

EW var + Pearson corr Run Time
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Director
PCA100%
PCA75%
PCA50%


EW var + Pearson corr Distance
0.012
0.010
0.008
0.006
0.004
0.002
0.000
Director
PCA100%
PCA75%
PCA50%


Pearson var + corr Run Time
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Director
PCA100%
PCA75%
PCA50%


Pearson var + corr Distance
0.012
0.010
0.008
0.006
0.004
0.002
0.000
Director
PCA100%
PCA75%
PCA50%


Pearson var + EW corr Run Time
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Director
PCA100%
PCA75%
PCA50%


Pearson var + EW corr Distance
0.012
0.010
0.008
0.006
0.004
0.002
0.000
Director
PCA100%
PCA75%
PCA50%

Conclusion:

- When PCA become smaller, it is time friendly. When we have the larger dataset, seting a relative low requirement on the percentage PCA explanation will give us an overview about the simulation results in time.
- The Difference about distance between PCA percentage explanation is relatively small, when the percentage is above 50%. In order to make the most efficiency simulation, to use the PCA 50% is acceptable.
- From covariance combination side, the Pearson variance combination sets have the best time efficiency. When we do not need to weight more for the time with small lag, Pearson would be better choice for the efficiency.
- Moreover, compared with other covariance calculation combinations, the Pearson Variance and Exponentially Weighted Correlation can have the highest distance between simulated X and real sigma.