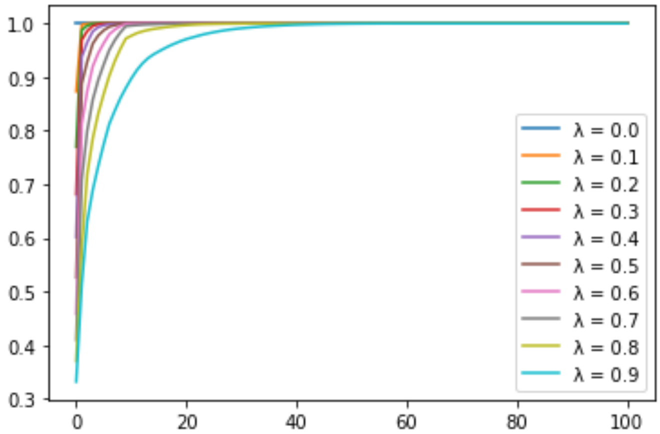
Problem 1

The plot of cumulative variance explained by each eigenvalue gives us insights into the relationship between λ and the covariance matrix used in the PCA analysis. The value of λ in an exponentially weighted covariance matrix determines the weight given to the historical returns.

A smaller value of λ gives more weight to recent returns and less weight to older returns. This results in a covariance matrix that more accurately reflects the current covariance structure of the returns. However, this also means that the covariance matrix is more sensitive to changes in the returns and may not be as stable over time. On the other hand, a larger λ gives more weight to older returns and less weight to recent returns. This results in a covariance matrix that is more stable over time, as it is less sensitive to changes in the returns. However, this also means that the covariance matrix may not accurately reflect the current covariance structure of the returns.



The plot shows that as λ increases, the cumulative variance explained by each eigenvalue approaches a stable value. This suggests that for a large enough λ, the covariance matrix becomes relatively stable and the eigenvalues of the PCA analysis become relatively constant. This means that the most important sources of risk in the returns are relatively constant over time, even though the covariance structure may be changing.

In conclusion, the choice of λ in an exponentially weighted covariance matrix is a trade-off between stability and accuracy. A smaller λ provides more accurate information about the current covariance structure of the returns, while a larger λ provides more stability over time.

Problem 2

The Frobenius norm is a measure of the difference between two matrices. The lower the Frobenius norm, the closer the matrices are to each other. The near_psd method produces the matrix with the lowest Frobenius norm, indicating that it is the closest to the original matrix. Higham's method produces a matrix with a higher Frobenius norm, indicating that it is less accurate than the nearest correlation matrix method.

In terms of runtime, Higham's method is faster than the nearest correlation matrix method. However, as the size of the matrix increases, the runtime of both methods will increase as well.

```
Original matrix Frobenius norm: 450.10494610591377
Near PSD matrix Frobenius norm: 0.6275226557678608
Higham's method Frobenius norm: 0.2352090499515368
Near PSD runtime: 0.13599s
Higham's method runtime: 0.04657s
```

The pros of the near_psd method are that it produces a matrix that is both positive-semidefinite and closest to the original matrix in terms of the Frobenius norm. The cons are that it can be computationally expensive for large matrices.

The pros of Higham's method are that it is computationally efficient and simple to implement. The cons are that it may produce a less accurate result compared to the nearest correlation matrix method.

In general, if accuracy is the most important factor, the nearest correlation matrix method should be used. If computational efficiency is the most important factor, then Higham's method can be used.

Problem 3

	method	direct_error	pca_100_error	pca_75_error	pca_50_error	time
0	1	7.754331e-07	0.144634	0.116594	0.104614	0.835892
1	2	6.388433e-07	0.121601	0.092600	0.070756	0.717177
2	3	7.785416e-07	0.139940	0.116055	0.103626	0.740705
3	4	6.542831e-07	0.124598	0.096417	0.070810	0.643070

we can see that the direct simulation method is the fastest, but also the least accurate with the highest errors. On the other hand, the PCA method with 100% variance explained is the most accurate but also the slowest. As we decrease the percentage of variance explained, the time to run decreases, but the errors also increase. So, there is a trade-off between the time to run and the accuracy of the simulation, and it depends on the specific use case and the required level of accuracy. If accuracy is more important, then using the PCA method with high variance explained may be the best option, even if it takes longer to run. If time is more important and accuracy can be sacrificed to some degree, then using the direct simulation method may be a better option. The PCA method with lower variance explained can be a good compromise between time and accuracy.