

Asset Pricing

Homework 15

Due at 24:00 pm (KST) on Thursday

Submit one file: written solutions with executable Python code in Jupyter Notebook(.ipynb)

**Subjects**

**[Session11. Latent Factor - Principal Components Analysis]**

1. Investment Science

**Chapter 9.** Data and Statistics

2. Advanced Data Science

**Chapter 15.** Principal Components Analysis

**Assignment 1.**

Summarize this week's study

**Assignment 2.**

Solve the following problems

**Problem 1.**

Gavin Jones figured out a clever way to get 24 samples of monthly returns in just over one year instead of only 12 samples; he takes overlapping samples; that is, the first sample covers Jan. 1 to Feb. 1, and the second sample covers Jan. 15 to Feb. 15, and so forth. He figures that the error in his estimate ofr, the mean monthly return, will be reduced by this method. Analyze Gavin's idea. How does the variance of his estimate compare with that of the usual method of using 12 nonoverlapping monthly returns?

**Problem 2.**

Suppose a stock's rate of return has annual mean and variance of and . To estimate these quantities, we divide 1 year into equal periods and record the return for each period. Let and be the mean and the variance for the rate ofreturn for each period. Specifically, assume that and . If and are the estimates of these, then and . Let and be the standard deviations of these estimates.

(a) Show that is independent of .

(b) Show how depends on . (Assume the returns are normal random variables).

**Problem 3.**

Suppose that instead of projecting on to a line, we project on to a q-dimensional subspace, defined by q orthogonal length-one vectors . We want to show that minimizing the mean squared error of the projection is equivalent to maximizing the sum of the variances of the scores along these q directions.

1. Write for the matrix forms by stacking the . Prove that = .

2. Find the matrix of q-dimensional scores in terms of and . Hint: your answer should reduce to   when .

3. Find the matrix of p-dimensional approximations based on these scores in terms of and . Hint: your answer should reduce to when .

4. Show that the MSE of using the vectors is the sum of two terms, one of which depends only on and not , and the other depends only on the scores along those directions (and not otherwise on what those directions are). Hint: look at the derivation of Eq. 15.5, and use Exercise 41.

5. Explain in what sense minimizing projection residuals is equivalent to maximizing the sum of variances along the different directions.

**Problem 4.**

Using Python,  
(a) Compute the change of daily returns and cumulative returns of 500 stock in the S&P 500(from 2020)(you can choose the number of stocks that you want to plot)

(b) Process the data via PCA by computing the 1st principal component

(c) Replicate the returns of the S&P500 approximately using (b)