MIT Sloan School of Management

Analytics of Finance 15.450
Hui Chen Spring 2023

Problem Set 1

(Due: 1:00 PM, Thursday, February 23)

Instructions:

- Please submit your homework on Canvas. List the names of all of your team members and their IDs in your writeup. Each team only needs to submit once.
- Submit the following files: (1) a PDF file for the writeup of your answers to all the questions; (2) for each question that involves coding, a separate zip file with the code and data for that question.

1. **Interview questions**
   (a) Provide an example of a consistent but biased estimator, and an example of an inconsistent but unbiased estimator.
   (b) True or false: If A and B are positively correlated, and B and C are positively correlated, then A and C are definitely positively correlated. Please explain.
   (c) Return of stock X has mean 20% and volatility 50%. Market return has mean 8% and volatility 20%. The correlation between stock X and market is 50%. If we run regression of stock X return on market return (with intercept), what is the regression coefficient on market return (i.e. beta)?

2. Do Question 2.8 and 2.10 from ISL Chapter 2 in Python or R.

3. **Predicting market returns:** As a portfolio manager, you are trying to develop a market timing strategy by exploiting the predictability of excess returns for the market index. We will study the following two models:
   - **IID model**: We assume monthly market excess returns r_t^e are IID normal,
   $$r_t^e \sim \mathcal{N}(\mu, \sigma^2). \qquad (\dagger)$$
   - **AR model**: Here we assume monthly market excess returns follow
   $$r_{t+1}^e = a_0 + a_1 r_t^e + \varepsilon_{t+1} \qquad (*)$$
   where ε_t is IID normal with $\varepsilon_t \sim \mathcal{N}(0, \sigma_e^2)$.

(a) Download the data for monthly market excess returns and one-month Treasury bill rates from Ken French's website (https://goo.gl/xtDApU). The data span the period from 07/1926 to 12/2022. Make a time series plot and a histogram of the market excess returns.

(b) Estimate the two models (†) and (∗) using the full sample (07/1926 to 12/2022). In the case of the IID model, you can use maximum likelihood or method of moments. For the AR model, you can use OLS.

(c) Given the parameter estimates from above, predict the monthly returns for each month from 01/2014 to 12/2022 for the two models. Plot the time series of your forecasts on top of the realized excess returns. Compute the in-sample MSE for each model. Which model performs better in terms of MSE? Is this a good way to compare the two models? Please explain.

(d) Suppose we are currently at the beginning of January 2014 and pretend that only the data prior to 2014 are available to us. What do you predict the mean excess return to be in the month of 01/2014 based on the IID model? What about the AR model?

(e) Use the same procedure as in 3(d), produce your one-month ahead predictions of market excess returns for the period of 01/2014 to 12/2022 for the two models. Again, plot the two time series of your forecasts on top of the realized excess returns and compute the MSEs. (Hint: Each month, you will use all the data available until that point as the traininig set to re-estimate the models, while the next month will be your test set. As you observe more data, your training sample is also growing.)

(f) Based on the out-of-sample MSE, which of the two models does a better job predicting returns?