FIN42330 Python for Financial Data Science

Group Project Assignment Details

(Please read all instructions very carefully)

Due Date: Friday, 19th December 2025 at 5pm

Submit your Python source code (preferably as a Jupyter notebook) and write-up of your results/answers (succinctly described and as a PDF file) via the dedicated submission folder under Assignments in the Assessment section on Brightspace. Source code can be separated into multiple files and function should be appropriately named. Please ensure that all tables, figures, and code are well annotated with appropriate information to help the reader discern what you are doing.

The submitted report should be attached to the appropriate submission form, which should closely adhere to the template provided in Brightspace and clearly provide all information required therein, including the assignment number, the group number, the name and student number of each group members and details on their individual contribution.[1]

1. Given the monthly prices of two asset classes: S&P 500 stock market index (SP500) and the Bloomberg Barclays U.S. Aggregate Bond Index (LBUSTRUU) for the period from December 1979 to December 2023, calculate the simple rate of returns on these two datasets.

   Using the monthly risk-free rate of return data provided for the period January 1980 to December 2023, calculate the stock and bond simple excess returns.

   The dataset on the stock index, bond index, and riskfree rate is provided in the folder 'Data_on_assets'.

   Write a Python function to compute the following statistics for the stock and bond simple excess returns:

   (a) annualized mean
   (b) annualized volatility
   (c) annualized Sharpe ratio
   (d) skewness
   (e) kurtosis

   Report the summary statistics in a table and explain the results.

[1] Please note that this is a requirement and, if not fulfilled, I will not accept the submission, applying any relevant late submission penalty until a proper submission is made.

2. Divide the total sample of monthly excess returns on the bond and stock indexes calculated in question 1 above into two periods: an in-sample from January 1980-December 2000 and an out-of-sample period from January 2001-December 2023. Using a rolling window estimation approach, generate a time-series of monthly out-of-sample constant expected (mean) excess return forecasts (call it the *benchmark forecast*) for each of the two asset classes. Write a function to do this computation.

3. Select the monthly data for five variables that you believe are plausible predictors of the excess returns on stocks and bonds (see, for example, Rapach, Ringgenberg, and Zhou, 2016 for stocks, Lin, Wu, and Zhou, 2017, and the references therein for inspiration) from the 'Predictors2023' dataset provided. Based on each of your five predictors and using the same rolling window estimation approach, generate monthly out-of-sample excess return forecasts for each of the asset classes using the following three predictive models (call these *model forecasts*):

   (a) an ordinary least squares (OLS) predictive regression model for each of the five predictors

   (b) a combination forecasts of excess returns that is a simple average of the forecasts based on the five predictors from the OLS model in (a) above.

   The above leads to a total of 6 model forecasts each for stock and bond excess returns.

   Compute the mean squared forecast error (MSFE) for the benchmark forecast and the ratio of MSFE for each of the model forecasts relative to the MSFE of the benchmark forecast MSFE for each asset. For example, the MSFE of the benchmark forecast for stock excess returns is given by

$$\text{MSFE} = \frac{1}{T}\sum_{t=1}^{T}\left(R_t - \hat{R}_t\right)^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2, \tag{1}$$

   where $\hat{R}_t$ is the benchmark forecast of stock excess return for period t, R_t is the actual (realized) stock excess return in period t, $R_t - \hat{R}_t = e_t$ is the forecast error, and T is the number of observation in the out-of-sample period.

   Compare the performance of the model forecasts relative to the benchmark forecast using the Diebold and Mariano (1995) test for equal predictive ability (you should write your own function to perform this test). The test assesses we assess whether the difference between the MSFEs of the benchmark forecast and the model forecast is statistically significant. You should clearly state the null hypothesis being tested and also provide a discussion of your table of results.

   In addition, create a figure showing the time-series of the benchmark and combination forecasts for each of the two asset classes.

4. Generate the out-of-sample forecasts of the (2-by-2) sample variance-covariance matrix for a portfolio of the two asset classes using the same rolling window estimation approach.

5. Using the benchmark excess return forecasts on the two assets and the sample variance-covariance matrix forecasts, construct out-of-sample optimal portfolio weights for a mean-variance investor. Compute the annualized summary statistics (mean, volatility, and Sharpe ratio) for the optimal portfolio's excess return. The estimate of the weights of the optimal portfolio at time t is

$$\hat{\boldsymbol{w}}_t = \frac{1}{\lambda} \times \hat{\boldsymbol{\Sigma}}_t^{-1} \hat{\boldsymbol{\mu}}_t \tag{2}$$

where $\hat{\boldsymbol{\mu}}_t$ is a 2×1 vector of benchmark excess return forecasts on the two asset assets at time t, $\hat{\boldsymbol{\Sigma}}$ is a 2×2 matrix, and λ is investor's risk aversion parameter, which you should set to a value of 3.

The out-of-sample portfolio excess return at time t is given by

$$R_{p,t+1} = \hat{\boldsymbol{w}}_t' \boldsymbol{r}_{t+1}, \tag{3}$$

where $\hat{\boldsymbol{w}}_t$ is the 2×1 vector of optimal portfolio weight at time t calculated from equation (2), $\boldsymbol{r}_{t+1}$ is a 2×1 vector of actual (realized) excess return on the two assets at time $t + 1$, and $'$ is the transpose operator.

Repeat the exercise with the 6 model forecasts in place of the benchmark excess return forecasts and compute the annualized summary statistics for portfolios generated.

Report the summary statistics in a table and explain the results. Does any statistical evidence of out-of-sample forecasting performance translate into economic gains/significance? In addition, create a figure showing the time-series of portfolio weights and cumulative excess returns for the optimal portfolio based on the benchmark forecast and the combination forecast.

References

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & economic statistics 13, 253–263.

Lin, H., Wu, C., Zhou, G., 2017. Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. Management Science .

Rapach, D. E., Ringgenberg, M. C., Zhou, G., 2016. Short interest and aggregate stock returns. Journal of Financial Economics 121, 46–65.