Fall 2019

Final Project Description and Requirements: Portfolio Optimization

Project Description

This problem is posted by McKinley Capital LLC as their Portfolio Optimization Horse Race for Summer 2019. The goal of the project is to construct a portfolio of U.S. stocks which tracks the expected return of a given benchmark index and yields the highest information ratio (as defined by the annualized monthly return divided by the standard deviation of the returns). The benchmark index is the Russell 1000 Index from Dec. 2002 to Aug. 2019.

Courtesy to McKinley, a set of more than 20 variables (a.k.a., factors) on a universe of about 1000 US and ADR stocks for 16 years are provided. These monthly factor data cover the period from 12/2002 to 11/2018. The data is contained in a zip file posted on Canvas. The definitions of the factors are given in the Appendix.

Complete the following tasks:

1. Use the monthly returns of the Russell 1000 index from Jan. 2003 to Nov. 2018 contained in "Benchmark Returns.csv" and compute the information ratio and maximum drawdown of the buy-and-hold strategy for this index.

2. Consider the data in t = Dec. 2004. Let R_i denote the monthly return of stock i. Using the monthly return data over the past 24 months (i.e., a 24-month look-back period starting from and including Dec. 2004) in the factor data file, compute the expected returns of R_i (denoted by μ_i), Var[R_i], and Cov(R_i, R_j) (denoted by σ_{ij}) for all stock pairs (i, j) in the universe obtained in Dec. 2004.

3. Fit a 10-factor return prediction model as given by equation (15) in Reference [1] using all stocks contained in universe obtained in part 2. You may use any programming tools to fit this multivariate linear regression model across all securities, either Python or C++. Interpret the fitted results and explain whether the 11 fitted coefficients of the factors (a0, a1, a2, … , a10) obtained for Dec. 2004 are statistically significant.

4. Apply a Python package cvxopt to construct a Markowitz mean-variance optimal portfolio $w \equiv (w_1, w_2, \dots, w_N)$ where N is the number of stocks in the universe obtained in part 2. This portfolio maximizes the portfolio return based on the expected returns $\tilde{\mu}_i$ of all stocks predicted by the fitted model in part 3 (instead of the realized returns recorded in Dec. 2004). (Remark: to get robust predicted returns, it is suggested that one repeats the model fitting task described in part 3 for Nov. 2004, Oct. 2004, …, and Jan. 2004. Then use

the average of the 12 sets of coefficients (a0, a1, a2, … , a10) to get the predicted returns in Dec. 2004.)

The covariance matrix of returns (R_1, R_2, R3, … , R_N), denoted by Σ, is constructed by setting the diagonal elements σ_{ii} to Var(R_i), and off-diagonal elements σ_{ij} to Cov(R_i , R_j) for all stock pairs (for i not equaling j) obtained in part 2 (remark: the covariance matrix Σ needs to be positive-definite. You shall write a module to check whether all the eigenvalues of Σ are positive. If Σ is not positive definite, then you may replace it with $\sqrt{\Sigma * \Sigma^T}$.). The constraint for this optimal portfolio is that the variance as computed by the above covariance matrix needs to be no greater than 0.0064 (which corresponds to a tracking error of no more than 8% with respect to the benchmark index). Namely, the portfolio weight vector w is the solution of the following mean-variance portfolio optimization problem:

$$max_w \sum_{i=1}^{N} w_i * \tilde{\mu}_i$$

Such that: $\sum_{i=1}^{N} w_i = 1$

$w\Sigma w^T <= 0.0064$

$0 <= w_i <= U_i$, for i = 1, 2, … , N,

where U_i is the upper limit of the weight of stock i to hold in the portfolio. Set $U_i = 0.05$ for all i.

Reference [3] explains how to use the cvxopt package to solve for the above mean-variance optimal portfolio w. (Remark: an optional practical constraint is to require the weighted average of the 0-100 ranked ES value of the portfolio being no less than 80.)

5. Use the portfolio weight vector w obtained in Dec. 2004 to compute the portfolio return in Jan. 2005. Repeat part 2 – 4 for t = Jan. 2005, Feb. 2005, Mar. 2005, … , Nov. 2017, Dec. 2017 to obtain a series of 168 monthly returns of the monthly-optimized portfolio. Compute the information ratio and maximum drawdown of this portfolio strategy.

6. Consider a different portfolio strategy which maximizes the CTEF score of the portfolio subject to the same set of constraints as defined in part 4. This is simply done by repeating part 4-5 for t = Jan. 2005, Feb. 2005, Mar. 2005, … , Oct. 2018, Nov. 2018 with the predicted returns of the stocks replaced by the CTEF scores observed in (t-1). Compute the information ratio and maximum drawdown of this portfolio strategy. Remark: this strategy is based on a simply return-prediction model which uses the CTEF score of a stock in month (t-1) to be its return in month t.

7. Implement a mean-variance optimal portfolio strategy based on TWO of your own return-prediction models which utilizes any machine learning/deep learning model to predict the stock returns in month t based on factor values observed in previous months. Namely, repeat part 3-5 with your own return-prediction model replacing the 10-factor model. Compute the information ratio and maximum drawdown of your portfolio strategy based on the 168 monthly returns.

Submission requirements:

1. All source codes and a README.txt file (same requirements on the content of the README.txt as before). Source codes shall be readily run, and tested for expected outputs without any issue. (These include all the C++/header files and Python scripts used in the project).

   The codes and implementation shall demonstrate proficiency in the following aspects:

   - Object-oriented design
   - Clean structure of the task flows and the corresponding functional modules
   - Adequate use of functions
   - Clear input/output interfaces

2. A written report describes the approaches taken, results found and analysis performed which pertain to each of the 7 parts in the project description.
   a. The construction of the input variables (i.e., features) and output variables to the price prediction models need to be clearly explained.
   b. The statistical validity and the accuracy of the price prediction models shall be provided.
   c. Clear justification of no look-forward bias in the models/strategies.
   d. The report shall contain the following sections:
      i. Problem description and data preparation.
      ii. Specification of each of the 4 return-prediction models (10-factor, CTEF, two of your own models). Discussions of the model fitting results and the validity of fitted parameters.
      iii. Performance metrics (using tables and figures) and discussions of each of the 4 portfolios constructed based on the 4 return-prediction models.
      iv. Conclusions.
      v. Appendix: explicit description of the contribution of each group member in “Contribution” section. Under the name of each member, list the following information,
         1. the work done such as performed data cleaning/feature generation, built a Support Vector Machine regression model to predict price, conducted model validation with backtesting, etc.

2. the codes or parts of a code written by the member.

e. Four csv output files generated by the TWO of your own models so one can directly compute the monthly returns of the two respective portfolios by multiplying the two matrices contained in these csv files. Specifically, for each model, submitting the following two csv files:
    i. One csv file has the first row being all the SEDOLs which you use to create portfolios at a monthly frequency in Dec. 2004, Jan. 2005, Feb. 2005, Mar. 2005, … , Nov. 2017. Put all SEDOLs of the union of the stocks appearing in the 168 portfolios as the labels of all columns in the first row of the csv file, then for the subsequent rows, use YYYY-MM as the index of these rows, and put the portfolio weights of the portfolio in YYYY-MM in the corresponding row with each weight entered into the corresponding cell with the correct SEDOL column label.
    ii. The other csv file has the exact row index and column headers. All the cells contains the monthly returns predicted by the model for each of the SEDOLs appearing in the column headers in all YYYY-MM rows.

References

[1] Guerard, J.G., S.T. Rachev, and B.P. Shao (2013). "Efficient global portfolios: big data and investment universes", IBM J. RES. & DEV. Vol. 57, No. 5.

[2] Guerard Jr., J. B., Markowitz, H.M., & Xu, G. (2014). The role of effective corporate decisions in the creation of efficient portfolios, *IBM Journal of Research and Development* 58, No. 4, Paper 11.

[3] Blog: Markowitz Portfolio Optimization in Python. https://plot.ly/ipython-notebooks/markowitz-portfolio-optimization/

Appendix

Explanation of data labels in the factor data file

1.DATE: MM/YYYY

2.CUSIP

3.SYMBOL

4.COMPANY NAME

5.SEDOL: stock identifier

6. FS_ID: FactSet ID

7. EP - Earnings/Price

8. BP - Book/Price

9. CP - Cash Flow/Price

10.SP - Sales/Price

11.DP - Dividend Yield

12.EP1 or FEP1 - 1 year ahead IBES Forecasted EPS to Price/Last year's forecasted earnings per share

13.EP2 or FEP2 - 2 year ahead IBES Forecasted EPS to Price/Last year's forecasted earnings per share

14.RV1 - FEP1 IBES Revisions

15.RV2 - FEP2 IBES Revisions

16.BR1 - IBES Breadth

17.BR2

18.CTEF - Consensus EPS I/B/E/S forecast, revisions and breadth

19.PM1 - price momentum as price(t-1)/price(t-12)

20.PM2 - price momentum as price(t-1)/price(t-7)

21.ES - (Eli Schwartz) Corporate Exports

22.RETURN - Returns

23.REP - Current EP/Average EP of last 5y

24.RBP - Current BP/Average BP of last 5y

25.RCP - Current CP/Average CP of last 5y

26.RSP - Current SP/Average SP of last 5y

27.RDP - Relative Dividend Yield

28.VOL - Monthly stock Volume

29.CRET - Monthly Stock Return

30.STATPERS - Date of IBES Forecast

31.USFIRM - US Firm (=1)

32.CURCODE - Currency Code

33.TOT - Total Number of FY1 Analysts

34.FGR1 - 1 year ahead forecast earnings per share monthly breadth

35.FGR2 - 2 year ahead forecast earnings per share monthly breadth

36.MRV1 - Mckinley definition of revisions in 2005

37.MRV2

38. ROIC: past 12-month return on invested capital

39. RSTDEV: standard deviation of past 12 monthly returns

40. RPM71 (see PM2): reverse price momentum of month-7 price divided by month-1 price.

41. ROA1, ROA3: 1-year, 3-year return on asset

42. ROE1, ROE3, ROE5: 1-year, 3-year, 5-year return on equity

43. 9MFR: return forecast by Mckinley 9-factor model

44. 8MFR: return forecast by Mckinley 8-factor model

45. LIT: legal insider trading index