

# MMF1921 – Operations Research

## Project 1 (Summer 2025)

**Due Date: 07–May–2025 by 11:59 PM**

---

Please use MATLAB or Python (exclusively) to solve this project. This is a group project, with groups of 2 or 3 students per group. Each group is expected to submit their own original work. You are welcome to form your own group, but no more than 3 students per group is allowed.

### You are given:

- Raw market data consisting of adjusted closing prices for 20 U.S. stocks [1], and factor rates of return corresponding to 8 different factors [2].
- **MATLAB**
  - A MATLAB program template.
  - Five MATLAB function templates to code 4 different factor models and an optimization model.
- **Python**
  - A Python program template.
  - Five Python function templates to code 4 different factor models and an optimization model.
  - An env.yml file is provided that can be installed via the command `conda env create -f env.yml`. The environment is set-up to work on Windows. Please ensure your final code runs error-free in the provided environment.
  - The environment includes the python package `cvxpy` a general purpose package used to solve convex optimization problems
  - BEFORE running the environment set up command, please follow the `cvxpy` [instructions](#) for your OS to ensure your machine is ready for `cvxpy` to be installed.
  - Useful packages such as `Gurobi` (mixed integer optimization solver) are also included in the environment.
  - DO NOT modify the environment unless the TA approves the modification. The TA will not install new packages to run your code unless they are approved.
- (Optional) A LaTeX template for you to write your report. Anyone interested in learning how to use LaTeX is welcome to use this template. However, no penalties or rewards will be given for using LaTeX.

### You should hand in:

- A formal report composed of an introduction, methodology, results and conclusion.
  - The MATLAB or Python program and functions you wrote to solve this project.
- ⇒ Please submit all your files inside a compressed folder, including your report and MATLAB/Python code. The name of the submitted file should be **FirstnameLastname.zip**. Only one submission per team is required. Please submit this file electronically through the Quercus portal (a printed copy is not required).
-

# 1 Introduction

The purpose of this project is to study and compare four different factor models, and to use these models to estimate the parameters required for portfolio optimization. We will build portfolios using mean–variance optimization (MVO) as our investment strategy<sup>1</sup>.

Our investment universe in this project consists of 20 stocks ( $n = 20$ ), with the company tickers shown in Table 1. We are given monthly adjusted<sup>2</sup> closing prices corresponding to these 20 stocks from 31–Dec–2005 to 31–Dec–2016 [1]. We can use the historical prices to compute our observed asset monthly returns.

In addition, we are also given monthly factor returns for eight different factors [2] corresponding to the period 31–Jan–2006 to 31–Dec–2016. This includes the monthly risk-free rate.

**Table 1:** List of assets by ticker

F	CAT	DIS	MCD	KO	PEP	WMT	C	WFC	JPM
AAPL	IBM	PFE	JNJ	XOM	MRO	ED	T	VZ	NEM

**Table 2:** List of factors

Market (‘Mkt_RF’)	Size (‘SMB’)	Value (‘HML’)	Short-term reversal (‘ST_Rev’)
Profitability (‘RMW’)	Investment (‘CMA’)	Momentum (‘Mom’)	Long-term reversal (‘LT_Rev’)

The factor models that we will implement in this project are:

- (a) OLS regression on all eight factors (OLS model)
- (b) Fama–French three-factor model (FF model)
- (c) Least absolute shrinkage and selection operator model (LASSO model)
- (d) Best subset selection model (BSS model)

We will use the factor models to estimate the asset expected returns and covariance matrix, which will use as the inputs for portfolio optimization. Our investment strategy in this project will be mean–variance optimization (MVO). This will allow us to compare the out-of-sample performance of the portfolios built using the different factor models.

## 2 Factor models

Factor models are very popular in both academia and industry, but they are nothing more than linear regression models. The reason we refer to them as factors is because of their financial relevance in explaining systematic return and risk. In other words, we can attribute and measure an asset’s return and risk based on its exposure to some relevant factors.

<sup>1</sup>We will use the term ‘investment strategy’ to refer to our portfolio optimization model.

<sup>2</sup>Your final report should explain why we use *adjusted* closing prices instead of closing prices when using historical data for a backtest.

For this project, we will focus on modelling the asset *excess* returns, where we measure the return in excess of the risk-free rate. Therefore, we must ensure that we subtract the monthly risk-free rate from our monthly asset returns. The monthly risk-free rate is provided with the factor data [2].

All the factors in Table 2 stem from synthetic portfolios of assets with shared properties. Thus, most factors will exhibit some degree of correlation, meaning our factor models will not respect the ideal environment. However, the factor covariances still convey valuable information about the asset risk. Therefore, we must include the factor covariance terms in our calculation of our asset covariance matrix.

In this project we will implement the four factor models described below.

- (a) **OLS model:** We will construct a multi-factor model using all the factors shown in Table 2. The factor model for asset  $i$  should look like this

$$r_i - r_f = \alpha_i + \sum_{k=1}^8 \beta_{ik} f_k + \varepsilon_i$$

where  $r_i$  is the return of asset  $i$ ,  $r_f$  is the risk-free rate,  $\alpha_i$  is the intercept from regression,  $f_k$  is the return of factor  $k$ ,  $\beta_{ik}$  is its corresponding factor loading, and  $\varepsilon_i$  is the stochastic error term of the asset.

The monthly factor returns for all eight factors and the monthly risk-free rate are provided in this project [2].

- (b) **FF model:** The Fama–French three-factor model is a subset of our ‘OLS model’, where we use only the Market, Size, and Value factors from Table 2. The FF model is

$$r_i - r_f = \alpha_i + \beta_{im}(f_m - r_f) + \beta_{is}f_s + \beta_{iv}f_v + \varepsilon_i$$

where  $r_i$  is the return of asset  $i$ ,  $r_f$  is the risk-free rate,  $\alpha_i$  is the intercept from regression,  $(f_m - r_f)$  is the excess market return factor and  $\beta_{im}$  is its corresponding factor loading,  $f_s$  is the size factor and  $\beta_{is}$  is its corresponding factor loading,  $f_v$  is the value factor and  $\beta_{iv}$  is its corresponding factor loading, and  $\varepsilon_i$  is the stochastic error term of the asset.

The data for these factors is given as part of the data set provided. However, we must select only the pertinent columns corresponding to the three Fama–French factors.

- (c) **LASSO model:** We will use the penalized form of LASSO on all eight factors, where the model is the following

$$\min_{\mathbf{B}_i} \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 + \lambda\|\mathbf{B}_i\|_1$$

As part of the research component of this project, we must select an appropriate value for  $\lambda$  and we must justify our choice. Alternatively, we can test different values of  $\lambda$  to study the impact on the resulting factor model. An ideal value of  $\lambda$  will result in a sparse factor model where only two to five coefficients are non-zero. This accounts for both the factor loadings and the intercept. We will let LASSO determine whether we should include the intercept or not.

- (d) **BSS model:** We will use the constrained form of the Best Subset Selection model using all

eight factors as inputs. This version of BSS is the following

$$\begin{aligned} \min_{\mathbf{B}_i} \quad & \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{B}_i\|_0 \leq K \end{aligned}$$

Our basis BSS model should have  $K = 4$ , i.e., we will construct a model with either four factors or three factors plus the intercept. We will let the BSS model determine whether we should include the intercept or not. As part of the research component of this project, you may want to test other values of  $K$  to study how this affects both the in-sample measures of fit and the out-of-sample portfolio performance.

For additional detail on how to setup the BSS model, see Page 8 out of 40 (shown as Page 820 on the document) of the paper by [Bertsimas et al.](#) [3]. A copy of this paper is available on the Quercus portal.

Once the factor models are calibrated, we will use them to estimate our asset expected returns and covariance matrix,  $\boldsymbol{\mu}$  and  $\mathbf{Q}$ . These parameters will serve as the inputs for portfolio optimization.

Note: You are expected to implement these models in MATLAB/Python by writing the code yourself. In other words, you should not use *regress()*, *lasso()*, *sklearn* built-in MATLAB/Python functions. You should calculate the coefficients of the OLS and FF models using linear algebra, as shown during the lectures. *numpy* is recommended for linear algebra operations in Python. Additionally, you should formulate the LASSO and BSS models as optimization models and solve them with *Gurobi*/*quadprog()*/*cvxpy*.

### 3 Portfolio optimization

Our investment strategy to optimize our portfolios will be Mean–Variance Optimization (MVO). We will use the version of MVO that seeks to minimize variance subject to a target expected return. The target return should be the average expected excess return of the market for the corresponding calibration period (i.e., take the geometric mean of the market factor using the historical data pertinent to the calibration period). Note that, by design, our target return will change every time we rebalance our portfolios. In addition, short sales are disallowed.

We will construct four portfolios, one for each factor model. Use the estimated  $\boldsymbol{\mu}$  and  $\mathbf{Q}$  from each factor model to construct your MVO portfolios. This will allow us to evaluate the impact that the different factor models have in the out-of-sample financial performance of the portfolios.

We will simulate a five-year investment horizon, from the start of 2012 until the end of 2016. In addition, we will rebalance<sup>3</sup> our portfolios every year, at the end of each year. This means we have a total of five investment periods. We must use four years of historical returns to calibrate each factor model. We can then use the factor models to estimate  $\boldsymbol{\mu}$  and  $\mathbf{Q}$ . The calibration period should immediately precede the start of the next investment period. For example, if our first investment period starts on Jan–2012, we must use data from Jan–2008 to Dec–2011 to calibrate our factor models. Once an investment period is over, we will re-calibrate our parameters using the most recent 4-year window available. The MATLAB/Python template provided is already setup to select the appropriate windows for in-sample calibration and out-of-sample testing.

---

<sup>3</sup>Portfolio rebalancing means we will update our portfolio weights by buying and selling shares of our stocks.

## 4 Results

We will conduct an in-sample and out-of-sample analysis to evaluate the performance of the factor models and their corresponding portfolios. The in-sample analysis will focus on the measure of fit of the factor models (i.e., we wish to assess how ‘good’ the factor models are at explaining our asset returns). The out-of-sample analysis will focus on evaluating the financial performance of the corresponding portfolios.

### 4.1 In-sample analysis

We should calculate the coefficient of determination (also known as  $R^2$  or ‘ $R$ -squared’) to compare the measures of fit between the four factor models. In particular, we will use the *adjusted*  $R^2$  as our measure of fit to penalize models that use additional factors.

Note: You are expected to calculate the adjusted  $R^2$  values on your own and without relying on a built-in MATLAB/Python function. For simplicity, refer to the [Wikipedia page](#) on the ‘coefficient of determination’ for a simple explanation on how to calculate the adjusted  $R^2$  values.

### 4.2 Out-of-sample analysis

For the out-of-sample analysis, we will measure the portfolio average return and volatility (standard deviation) incurred during the entire experiment. We may also want to calculate the risk-adjusted excess return, also known as the Sharpe ratio. The Sharpe ratio is simply the average return divided by the risk incurred. We should plot the portfolio value through time to visualize the wealth evolution of our portfolios.

We can also use an area plot to show the changes per period in the composition of our portfolios (i.e. to see how our optimal weights changed every time we rebalanced a portfolio). This will allow us to evaluate if our portfolios were concentrated or well-diversified.

## 5 Deliverables

### 5.1 Report (90%)

Prepare a formal report. The report should introduce the purpose of this project, explain your methodology in detail, show a summary of your computational results, and present an analysis of these results. Finally, you should include a discussion and conclusion section.

The report should demonstrate your understanding of the four factor models, as well as your understanding of MVO. Your analysis of the computational experiment should reflect your knowledge of the material we have seen in class. The discussion section should provide any insights you are able to derive from the results, as well as any strengths or weaknesses of the different models.

The report is worth 90% of the total. The distribution is the following

- Formal report structure and presentation: 10%
- Factor model description and implementation: 25%
- Portfolio optimization description and implementation: 10%
- In-sample analysis: 15%
- Out-of-sample analysis: 15%
- Discussion and conclusion: 15%

## 5.2 MATLAB/Python program (10%)

Prepare a MATLAB/Python program and functions to perform the computational experiment. Use the templates provided. You are allowed to modify these templates as much as you see fit. Be sure to properly comment on your code to briefly explain what you are doing. Your code should be easy to read and the TA should be able to run it. The TA will not debug your code and should not have to search for the results within the code.

## References

- [1] Quandl.com. *Wiki – Various End-Of-Day Stock Prices*. <https://www.quandl.com/databases/WIKIP/usage/export>. [Online; accessed 07-Nov-2017]. 2017. (Visited on 11/07/2017).
- [2] K. R. French. *Data Library*. [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). [Online; accessed 01-Feb-2020]. 2020. (Visited on 2020).
- [3] D. Bertsimas, A. King, and R. Mazumder. “Best subset selection via a modern optimization lens”. In: *The annals of statistics* (2016), pp. 813–852.