Computer Project (Quantitative Methods for Finance 2019)

You need to answer all the questions below and bring the answers in the Answer Sheet provided at the end of the document to the test; the test will be on 7 November (at 8:30). The test will ask you questions related to your answers (but not about the MATLAB code). At the exam you CANNOT bring other material. Hence, you need to understand what you do even if you collaborate. You will need to hand in these answers and all supporting documents at the end of the test. Only include originals (no photocopies) of relevant software output in an appendix if asked.

If you need any help, you should ask during the lecture on 24 October. I will not provide any explanation via email to avoid being unfair to others. Make sure you start working on this from now, as you may find it challenging.

Portfolios Analysis

You are going to conduct some analysis along the lines of Fama and French (1992) plus a simple backtest. Download the Fama and French (1992) paper from the URL:

https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.1992.tb04398.x

Skim through it and try to understand the main ideas. The details are not in the scope of this course. Other courses may discuss such details. Also briefly look in Wikipedia:

https://en.wikipedia.org/wiki/Fama–French_three-factor_model

Look also at the index of your book Brooks to find explanation of the methodology. Also check the Fama and MacBeth regression procedure in Brooks as well as the original paper:

http://ecsocman.hse.ru/data/965/126/1231/fama_macbeth_tests_1973.pdf

Try to understand the main ideas. We shall replicate some steps in the following exercises, so do not worry too much.

Download the portfolios: the Y variables (the Portfolios) and the X variables (the Factors)

Go to the URL:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Download data in csv format corresponding to 25 Portfolios Formed on Size and Book-to-Market (5 x 5) (towards the top middle of the page). You must make sure you download this dataset and not another in csv format. You can also directly download it from:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/25_Portfolios_5x5_CSV.zip

Download data in csv format corresponding to Fama/French 3 Factors (top of page). You must make sure you download this dataset and not another in csv format. You can also directly download it from:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/
F-F_Research_Data_Factors_CSV.zip

Note: Ensure that you have downloaded files with monthly frequencies and not daily.

The portfolios are not excess returns. The risk free interest rate is in the 3-factor file. Use this to compute excess returns on the portfolios. If you do not follow the paper, see the equation in

https://en.wikipedia.org/wiki/Fama–French_three-factor_model

Prepare the files

Put the zipfiles in a folder in your computer. For the sake of definiteness, I will suppose that the folder directory is "/media/marketdata/FRENCH/"; you can use another directory. If you use Windows, I think you need to use \ instead of /, but both might work (I only use Linux or Mac). Unzip the files. You will find the files 25_Portfolios_5x5.CSV and F-F_Research_Data_Factors.CSV. Rename them to portfolios.csv and factors.csv, respectively. Open the file portfolios.csv (in excel or similar) and delete the first 15 rows. This file also has Average Equal Weighted Returns and other datasets that follow the top dataset. Delete the remaining data and make sure you are left only with the dataset at the top. Open the file factors.csv and delete the first 3 rows. This file also has annual data that follow the monthly data towards the end of the file. Delete the annual data and make sue you are left only with the dataset at the top. For both files you should be left with a file with headings on the first row and data below. If not, I leave to you to select how many rows to delete in order to achieve this. This makes it easier for MATALB to read the files.

Load the Data

You must to use MATALB for the analysis. Load the data into MATLAB. To do so, use the function: importdata. I will show you the steps to make it easier for you:

```
myDir = '/media/marketdata/FRENCH/';
portfoliosFile = 'portfolios.csv';
factorsFile = 'factors.csv'; delimiter = ',';
portfolios = importdata([myDir, portfoliosFile], delimiter);
factors = importdata([myDir, portfoliosFile], delimiter);
```

I have been told that some people had problems loading the data in Mac (I think). If that so, try to follow the Matlab notes. Also, you may delete the headings (column names) from the files, but make sure that you know which is which.

The data are monthly observations. For both files only use data for the period '196001' to '201812' included. If you do not know how to extract the subset of the data using MATALAB, just open the files and delete the rows outside this range and then load the data in MATLAB.

Make sure you only include this dates, as failing to do so may change the results.

Exercises

Exercise 1 *Compute the following summary statistics for each portfolio excess return (Average value weighted returns, Monthly): mean, variance, skewness and kurtosis. Print the plots for these summary statistics; e.g. plot(mean(portfoliosData)), plot(var(portfoliosData)), where portfoliosData is an* $n \times 25$ *matrix where* n *is the sample size for the period '199001' to '201908' included. Make sure you know to which portfolio each entry in the x axis corresponds to; e.g. the first entry is 'SMALL LoBM'. Attach these figures as an appendix to the Answer Sheet.*

Record the statistics (the actual numbers) for 'SMALL LoBM', 'SMALL HiBM', 'BIG LoBM', 'BIG HiBM' in the answer sheet.

Exercise 2 *(Time series regression) For each portfolio* $p = 1, 2, ..., 25$*, consider the regression*

$$\left(R^{(p)} - r_f\right) = \alpha^{(p)} + \beta_m^{(p)} \left(R_m - r_f\right) + \beta_s^{(p)} SMB + \beta_v^{(p)} HML + \varepsilon^{(p)}$$

where $R^{(p)}$ *is the* $n \times 1$ *vector of returns on the* p^{th} *portfolio in the portfolios dataset.* $R_m - r_f$ *(Mkt-RF) is the market excess return, SMB is small minus big, and HML is high minus low. All vectors are* $n \times 1$*. These columns are labeled in the 3-factor file. Put the model in the usual form*

$$Y = Xb_0 + \varepsilon,$$

where

$$Y = \begin{pmatrix} R_1^{(1)} - r_{f,1} & R_1^{(2)} - r_{f,1} & \cdots & R_1^{(25)} - r_{f,1} \\ R_2^{(1)} - r_{f,1} & R_2^{(2)} - r_{f,2} & \cdots & R_2^{(25)} - r_{f,2} \\ \vdots & \vdots & \ddots & \vdots \\ R_n^{(1)} - r_{f,n} & R_n^{(2)} - r_{f,n} & \cdots & R_n^{(25)} - r_{f,n} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & R_{m,1} - r_{f,1} & SMB_1 & HML_1 \\ 1 & R_{m,2} - r_{f,1} & SMB_2 & HML_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & R_{m,n} - r_{f,1} & SMB_n & HML_n \end{pmatrix}$$

$$b_0 = \begin{pmatrix} \alpha^{(1)} & \alpha^{(2)} & \cdots & \alpha^{(25)} \\ \beta_m^{(1)} & \beta_m^{(2)} & \cdots & \beta_m^{(25)} \\ \beta_s^{(1)} & \beta_s^{(2)} & \cdots & \beta_s^{(25)} \\ \beta_v^{(1)} & \beta_v^{(2)} & \cdots & \beta_v^{(25)} \end{pmatrix}$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1^{(1)} & \varepsilon_1^{(2)} & \cdots & \varepsilon_1^{(25)} \\ \varepsilon_2^{(1)} & \varepsilon_2^{(2)} & \cdots & \varepsilon_2^{(25)} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_n^{(1)} & \varepsilon_n^{(2)} & \cdots & \varepsilon_n^{(25)} \end{pmatrix}$$

Save the OLS estimator in a 4×25 *matrix, say* B. *Plot each column of the matrix* B' *separately. Add these graphs as an appendix to the Answer Sheet.*

Compute the residuals for each individual portfolio regression $p = 1, 2, ..., 25$*:*

$$\hat{\varepsilon}^{(p)} := Y^{(p)} - XB^{(p)}$$

where $Y^{(p)}$ *is the* p^{th} *column in* Y, *and* $B^{(p)}$ *is the* p^{th} *column in* B.

Compute an estimator for $Var\left(B^{(p)}\right)$. *In the Answer Sheet record the values of* $B^{(p)}$ *(the estimated alphas and betas) and their standard errors for the portfolios corresponding to 'SMALL LoBM', 'SMALL HiBM', 'BIG LoBM', 'BIG HiBM'.*

Exercise 3 *(Cross-sectional regression) Suppose that there are three scalars such that for each portfolio* $p = 1, 2, ..., 25$,

$$\mathbb{E}\left(R_t^{(p)} - r_f\right) = \beta_m^{(p)}\lambda_m + \beta_s^{(p)}\lambda_s + \beta_v^{(p)}\lambda_v. \tag{1}$$

The λ*'s are the risk premia associated to the three (pricing) factors: market, size (SMB) and value (HML). To estimate them we can run the following regression. At time* t *(the* t^{th} *row in the dataset)*

$$\left(R_t^{(p)} - r_{f,t}\right) = \alpha_t + \beta_m^{(p)}\lambda_{m,t} + \beta_s^{(p)}\lambda_{s,t} + \beta_v^{(p)}\lambda_{v,t} + Z_t^{(p)},\ p = 1, 2, ..., 25.$$

Hence you are supposing that this equation holds for each $p = 1, 2, ..., 25$ *where* α_t *is zero when we average across* t *(if (1) holds) . The term* $Z_t^{(p)}$ *is a mean zero pricing error. We want to estimate the unknown parameters* $\lambda_{m,t}, \lambda_{s,t}, \lambda_{v,t}$ *and* α_t. *We do not know* $\beta_m, \beta_s, \beta_v$, *but we have estimated them in the previous exercise and saved them in the matrix* B. *We suppose that the estimated betas in* B' *are the matrix of explanatory variables and regress at time* t *the 25 portfolio returns on columns 2 to 4 in* B'*; this is made clear next. To do so, let* R_t *be the* 25×1 *vector of returns at time* t *where the* p^{th} *row corresponds to the* p^{th} *portfolio. At each time* t, *run the regression supposing the true model*

$$Y = Xb_0 + \varepsilon$$

where

$$Y = \begin{pmatrix} R_t^{(1)} - r_{f,1} \\ R_t^{(2)} - r_{f,t} \\ \vdots \\ R_t^{(25)} - r_{f,t} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & \hat{\beta}_m^{(1)} & \hat{\beta}_s^{(1)} & \hat{\beta}_v^{(1)} \\ 1 & \hat{\beta}_m^{(2)} & \hat{\beta}_s^{(2)} & \hat{\beta}_v^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \hat{\beta}_m^{(25)} & \hat{\beta}_s^{(25)} & \hat{\beta}_v^{(25)} \end{pmatrix}$$

$$b_0 = \begin{pmatrix} \alpha_t \\ \lambda_{m,t} \\ \lambda_{s,t} \\ \lambda_{v,t} \end{pmatrix}$$

$$\varepsilon = \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \vdots \\ \varepsilon_t^{(25)} \end{pmatrix}$$

Your goal is to find the estimator $\hat{b} = \begin{pmatrix} \hat{\alpha}_t \\ \hat{\lambda}_{m,t} \\ \hat{\lambda}_{s,t} \\ \hat{\lambda}_{v,t} \end{pmatrix}$ *for* b_0. *Repeat the procedure for each* t *and save* $\hat{b}$ *as the* t *column in the* $4 \times n$ *matrix* $riskPremia$. *You have just computed a time series of risk premia. Plot the time series of risk premia for all three factors separately and the pricing errors* α_t*'s. Compute the following statistics for the three risk premia and the pricing error: mean, variance, skewness and kurtosis. Record these sample statistics in the Answer Sheet.*

Exercise 4 *(Backtest) Use the excess returns on the 25 portfolios to compute the sample covariance matrix on the sample period 200801-201312. Denote by* COV *this sample covariance. Compute the shrunk covariance matrix* $COVS = (1-a) * mean\,(diag\,(COV)) * eye\,(25) + a * COV$ *with* $a = .5$. *Ensure that you understand what "*$mean\,(diag\,(COV)) * eye\,(25)$*" is.*

Compute the portfolio weights that solve the minimum variance portfolio problem, with weights summing to one, using the two different estimators of the covariance matrix and denote them by w *and* ws. *Also consider the equally weighted portfolio, and denote its weights by* wEW. *Carry out a backtest of the three portfolios. To do so, you must use data unseen by the estimators* w *and* ws. *Hence, consider the testing period 201401-201808 to compute the portfolio excess returns for the strategy that uses* w, ws *and* wEW *as portfolio weights. (Conducting a* *backtest on data used for estimation is said to be subject to forward looking bias and must be avoided when possible.) Also, consider the excess returns on the market ($R_{m,t} - r_{f,t}$) as benchmark. Compute the mean, variance, skewness and kurtosis for the three portfolios and the benchmark and report their values in the Answer Sheet. Plot the cumulative (monthly compounded) excess returns of the portfolios and the benchmark and put them in an appendix. The excess returns are in percentage points. Hence, make sure that you divide them by 100 in order to compute the cumulative compounded returns so that for example, 0.15 means 15%. If R_s is the one month simple return at time s, the cumulative compounded return at time t is $C_t = \prod_{s=1}^{t}(1 + R_s)$. The plot is the graph of C_t against t.*

Exercise 5 *(Test) Given the estimated pricing errors $\hat{\alpha}_t$ in the period 199001-201908, test the null hypothesis $\mathcal{H}_0 : \alpha^{(1)} = \alpha^{(2)} = \ldots = \alpha^{(25)} = 0$, where $\alpha^{(k)}$ is the α corresponding to the $k-th$ portfolio. Write down the statistics corresponding to this hypothesis and the critical value of the distribution of the test. Do you reject or not reject the null hypothesis? Comment the results.*

Answer Sheet: Record Your Answers Here Put Graphs in an Appendix

Name______________________ Last Name______________________

Question	Answer			
Q1				
	'SMALL LoBM'	'SMALL HiBM'	'BIG LoBM'	'BIG HiBM'
mean				
var				
skewness				
kurtosis				
Q2				
	'SMALL LoBM'	'SMALL HiBM'	'BIG LoBM'	'BIG HiBM'
$\hat{\alpha}$ value (s.e.)				
$\hat{\beta}_m$value (s.e.)				
$\hat{\beta}_s$ value (s.e.)				
$\hat{\beta}_v$ value (s.e.)				
Q3				
	$\hat{\alpha}$	$\hat{\lambda}_m$	$\hat{\lambda}_s$	$\hat{\lambda}_v$
mean				
var				
skewness				
kurtosis				
Q4				
	w	ws	wEW	benchmark
mean				
var				
skewness				
kurtosis				
Q5				
	statistic	critical value		