

## Case 1: Explaining (the sign of) stock returns

**Deadline: Monday November 9, 2020, 11.00 AMS (upload)**

This assignment should be done in teams of 2 students.

### Requirements

- Put your **group number, names, VU ID numbers**, and **email addresses** of all group members on the front at a separate title page.
- You have to answer each (sub) question separately.
- Reporting style as explained in class. Title, text layout, clarity of tables and figures, no screen dumps, tables no vertical lines, tables and figures in Journal of Finance layout (style + explanatory note, to be stand-alone), etc. Failure to meet the guidelines results in loss of points.
- If you are asked to **test** something, always provide  $H_0/H_A$ , the test-statistic (formula and actual outcome) and your **economic** conclusion
- Only one of you has to upload the final version before the deadline. Do **\*not\*** upload two versions of the paper / files, one by each one of you.
- You **upload** your **report in PDF** format via Canvas. Name of the file: student ID 1 nr \_ student ID 2 nr plus "case1\_EF.pdf", e.g., 2056791\_111222\_case1\_EF.pdf.  
Be aware that you should provide your **annotated / commented DO file** in the appendix of your report for me to run. The do file should be such that I can run it to get precisely the tables and figures you have in your report. Also make sure your do file has a "`cd your-map-name`" command at the start, such that all file names in the rest of the file are relative to this map [see the lab sessions].
- Clarity of descriptions, replicability, clarity of argumentation, interpretation and conclusions do earn you points.

## Description and Data

In this case you will work with the famous Fama-French 5-factor model (Fama and French; 2014 JFE). This asset-pricing model aims to describe stock returns. Whereas the traditional CAPM model relates the excess stock return  $r_t - r_{f,t}$  at time  $t$  only to the excess market return  $r_{m,t} - r_{f,t}$ , Fama and French argue that factors based on value, size, profitability and investment also play a role in modeling stock returns.

In this case, two questions are asked. First, are there any other macro/financial “factors” that explain excess returns, corrected for the five Fama-French (FF) factors? Second, suppose we are not interested in explaining the excess return itself, but its *sign*. Do macro/financial variables have additional explanatory power over the typical FF factors in explaining the sign of excess stock returns?

You are handed three datasets. The Excel file **monthly\_returns\_and\_volume** contains sheets with the monthly log-returns and monthly trading volume of 110 U.S. stocks. **You need to select the stock number that corresponds with your group number from Canvas.** The second Excel file **macro\_fin\_data** contains some U.S. macro/financial variables. Have a look at the README sheet for more information about this data. The final dataset (**FF\_5\_factor\_data.xls**) captures the five monthly Fama-French factors.

In summary, you have the following data:

Variable	Description
<b>Dataset I: stock returns and Volume</b>	
<b>Date</b>	Date of the time series
<b>Stock return</b>	return of stock $i$ ( $i = 1, \dots, 110$ ) (perc)
<b>Volume</b>	Trading volume of stock $i$
<b>Dataset II: Macro/fin data</b>	
<b>IP-growth</b>	Industrial production Index growth (perc)
<b>Inflation</b>	CPI of All Urban Consumers growth (perc)
<b>Credit growth</b>	Total Consumer Credit growth (perc)
<b>M1 growth</b>	M1 Money stock growth (perc)
<b>TB3MS</b>	3-Month Treasury Bill (perc, <b>ann</b> )
<b>GS10</b>	10-year Treasury Constant Maturity Rate (perc, <b>ann</b> )
<b>BAA</b>	Interest rate on BAA bond (perc, <b>ann</b> )
<b>AAA</b>	Interest rate on AAA bond (perc, <b>ann</b> )
<b>Credit spread</b>	BAA – AAA
<b>Term spread</b>	GS10 – TB3MS

### Dataset III: FF factors

<b>Excess market return</b>	Daily market return over the risk-free rate (perc)
<b>Risk-free rate (Rf)</b>	The risk-free rate (perc)
<b>HML</b>	High-Minus-Low factor (perc)
<b>SMB</b>	Small-Minus-Big factor (perc)
<b>RMW</b>	Profitability factor (perc)
<b>CMA</b>	Investment factor (perc)

*Table: Overview of data. NOTE: The monthly FF factors, stock returns and growth rates are all expressed in percentages. The variables TB3MS, GS10, BAA, AAA are expressed in percentages as well, but they are **annualized**!*

## Part I: Explaining returns

**1a)** Create summary statistics (mean, sd-dev, min, max and the  $p$ -value of the test on normality) of all your variables that you use in Part I. Put these into **Table 1** with an adequate caption.

Interpret the summary statistics of your excess stock return and the Term Spread variable. What is the average annualized return of your stock? Are there any outliers? If so, treat these with care. Put your original sum stat table with outliers in the appendix and provide the adjusted sum stat table in your main answer to 1a).

**1b)** Estimate two linear regression models with the excess return as the dependent variable but with different independent variables:

**Model 1:** the 5 Fama-French factors

**Model 2:** extend *Model 1* by including macro and/or financial variables and the logarithm of volume

*Note 1:* you decide by your own which macro/financial variables you include. You might use economic reasoning. It could be that you don't find any significance at all. Nevertheless, include at least **ONE** macro/financial variable!

*Note 2 (STATA):* we have time series here. For convenience, type '**gen time = \_n**', followed by '**tsset time**'. Now STATA knows that you have time series data.

Put the estimation results in **Table 2** using `outreg2`. Also, include at least two Goodness-of-Fit statistics in the table. Finally, write down the exact model equation that corresponds with the two models.

**1c)** Interpret the estimated coefficient related to one of your own selected macro/financial variables, one FF factor and the log-volume variable.

**1d)** Answer the main research question of part I. Perform an adequate **test** to confirm your answer.

**1e)** Create a dummy variable that equals 1 during the Global Financial Crisis (GFC) and 0 elsewhere. It is up to you when you think the GFC **precisely** started and/or ended. (there will be no points subtracted if you deviate a few months).

Estimate **Model III**, which is an extension of Model II in the following two ways:

- 1) account for a different intercept during the GFC
- 2) account for a different impact of the market-factor, your own included macro/financial variables and the log-volume during the crisis period.

Create a new table (**Table 3**) using `outreg2`, where you extend **Table 2** with the new output.

Also, write down the exact model equation corresponding to **Model III**. Interpret the coefficients related to the market factor AND one of your own included macro/financial variables.

Finally, **test** whether the impact of the market factor, your own included macro/finance variable(s) and log-volume indeed jointly changed during the crisis? Use the output of this test **and** Table 3 to answer this question.

## Part II: Explaining the sign of the returns

In this part we will change the dependent variable into a binary variable. You will investigate whether there is any impact of your independent variables on the sign of the excess returns.

**2a)** Create a binary indicator **return\_sign** which is equal to 1 if the excess return equals zero or higher and 0 elsewhere.

Test whether the proportion of positive excess returns equals 0.5. (Hint: just run a very particular regression in STATA)

**2b)** Estimate two Logit (or Probit) models that has exactly the same independent variables as model I and Model II. The dependent variable is now the new binary variable **return\_sign**.

Put the estimation results of both models into **Table 4** using `outreg2`. Provide at least one Goodness-of-Fit measure in the table. Finally, provide the exact model equation for both models.

**2c)** Compute the (average) marginal effects of all parameters of your Logit/Probit **Model II**. Show these into **Table 5**. Interpret the marginal effect of market factor and one of your own added macro/financial variables.

**2d)** Re-estimate your Logit/Probit Model I and II using the first 70% of your data. Then construct the ROC curve for both models for 1) the in-sample data and 2) the remaining 30% of your data. Interpret possible differences between both ROC curves. Which model do you prefer? In your answer, show the ROC curves and your conclusion based on these curves.

**2f)** Suppose the Financial Times would like to know your overall main conclusion, given your results of part I **and** II. Provide this main conclusion using a **maximum of two sentences**.

**BONUS (0.5 pt)**

**2g)** So-far you have estimated contemporaneous relationships. Now suppose that you would like to **forecast** the sign of the excess return **in the next month**, given information up-to-and-including **the current month**. Try to build a model that can do this.

In your answer, you should report:

- 1) An estimation table with your in-sample results using 70% of your data
- 2) The exact model equation corresponding with the estimation table
- 3) An adequate out-of-sample analysis on the last 30% of the data using your model and a proper **benchmark model**. Then provide a conclusion whether we can indeed forecast the sign of the excess returns. Can you relate this conclusion to a famous hypothesis about markets?

**Note: Do not forget to include an annotated DO-file into your final PDF!**