VRIJE UNIVERSITEIT AMSTERDAM

EMPIRICAL FINANCE

Case 2

Authors: **Group 56**
Soheel Gholamy: 2668182
Mark de Kwaasteniet: 2649271

S.gholamy@student.vu.nl
M.G.de.Kwaasteniet@student.vu.nl

Supervisors:
Anne Opschoor
Eva Mynott

Amsterdam, October 4, 2021

1 Question 1a

The data set contains 36 companies with a time series of 408 months per company, resulting in a total of 14688 observations. All the macro economic and Fama French factors are presented in % with an exception for trading volume of the stock of each company. The Skewness and Kurtosis test for every variable in table 1 indicates that the probability that each of them, separately, is normally distributed, is lower than 1%. For further analyses, we have generated the Excess Stock Return using the following formula,

$$ExRet_{i,t} = Ret_{i,t} - RF_t \tag{1}$$

where $ExRet$, Ret and R_f denote the Excess Stock Return, The Stock Return and the Risk Free Rate, respectively. Furthermore, the i and the t denote the id code of the stock and the time in months and years.

Table 1

Summary Statistics

Variable Names	Mean	Std. Deviation	Min	Max	P-Value
Excess Return	.924	8.782	-72.404	127.366	0
Ln_Volume	12.633	1.631	5.136	17.275	0
IDCODE	18.5	10.389	1.000	36	0
MktRF	.69	4.376	-23.240	12.47	0
SMB	.054	2.947	-14.860	18.05	0
HML	.169	2.889	-11.050	12.6	0
RMW	.347	2.468	-18.480	13.38	0
CMA	.241	2.002	-6.860	9.56	0
RF	.26	.206	0.000	.79	0
Unemp	5.857	1.518	3.500	10	0

This table contains the summary statistics of the variables in the data set of case 2 - group 56. The mean, standard deviation, minimum value and maximum values are represented in column 2,3,4 and 5 respectively (N = 14688, i = 36 and t = 408). P-value denotes the probability that the Chi-Squared distributed variables is greater than the test statistic. The p-value is calculated with the Jarque Bera test, which jointly tests the skewness and kurtosis. It represents the probability of accepting the null-hypothesis of the test, which is a joint hypothesis of the skewness being equal to zero and kurtosis being equal to three.

The Excess Stock Return has a monthly average of 0.924%, which means that with all the combined companies the monthly mean of the Excess Stock Return is 0.924%. The monthly volatility of the Excess Stock Returns is 8.782%, to determine the annualized volatility of the Excess Stock Returns, we multiply the monthly volatility by the square root of the number of months in a year, shown in equation 2.

$$Annualized - Volatility = 8.782\% \times \sqrt{12} = 30.42\% \tag{2}$$

When we look at the mean and the standard deviation of the $Unemployment$ variable, we don't notice a deviant pattern. The overall average is 5.857%, which seems realistic when compared to the current unemployment rate in the US of 5.2%. Furthermore, the unemployment statistics are positive, which is also realistic, since the unemployment rate cannot be negative. The only oddities can be found with the maximum and minimum values of Unemployment, since they are perfectly rounded numbers, which seems unlikely. This makes us wonder whether these number might have been cut at the point of 3.5- and 10

2 Question 1b

Model 2: Excess Stock Return on log Volume and FF and Macro Economic Factors.

$$\begin{aligned} ExRet_{i,t} = \alpha + \beta_1 MkrtRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \\ \beta_6 Unemp_t + \beta_7 Inflation_t + \beta_8 Ln\_Volume_{i,t} + \epsilon_{i,t} \end{aligned} \tag{3}$$

where i and t denote the IDCODE of the stock and the time in Year/Month, respectively. Furthermore, Ln is used to indicate that a variable has been log transformed.

Table 2

FF and Economic Factors affecting the Excess Stock Return

VARIABLES	Model 1	Model 2 (Including Economic Factors)
MktRF	1.083***	1.081***
	(0.0165)	(0.0166)
SMB	0.0156	0.0163
	(0.0239)	(0.0239)
HML	0.237***	0.232***
	(0.0306)	(0.0308)
RMW	0.374***	0.378***
	(0.0311)	(0.0313)
CMA	0.227***	0.228***
	(0.0449)	(0.0451)
Unemp		0.0465
		(0.0417)
Inflation		0.332
		(0.255)
ln_volume		-0.0568
		(0.0392)
Constant	-0.0484	0.327
	(0.0664)	(0.562)
Observations	14,688	14,688
R-squared	0.246	0.247
R-adjusted	0.246	0.246
F-stat	960.3	601

This table presents the regression results of model 1 and 2. Model 1 is the regression of the Excess Stock Return on the Market Risk Premium (MktRF), the Small-Big Factor (SMB), the High-Low Factor (HML), the Profitability Factor (RMW) and the Investment Factor (CMA). Model 2 is an extension of model 1, including the Macro Economic Factors; Unemployment Rate and Inflation Rate, and the natural log of Trading Volume of the stock. The standard errors are reported in the parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted by ***, ** and *, respectively. (N = 14688)

3 Question 1c

Interpretation of the Log Volume Variable

β_8 : If the Trading Volume of a stock increases by 1%, the Excess Stock Return of this stock will decrease by 0.0006 percentage points, corrected for the other factors.
The sign of the Log Trading Volume Coefficient is negative, meaning that an increase in Trading Volume will have a negative effect on the Excess Stock Return of a company.

The possible differences are noticeable in both the change in Fama French coefficients of both models and the F-statistic. The change in the coefficients is minimal in Model 2, meaning that the added Macro Economic Factors and the Log Volume aren't that highly correlated with the dependent variable. However the F-statistic is much higher in Model 1, indicating that the added variables in Model 2 are little correlated with the dependent variable resulting in an increase of the mean squared errors.

4 Question 1d

The estimated OLS standard errors are possibly inappropriate because there could be a strong firm-specific component in the error term. If there is somewhat a correlation between the standard errors over firms than the estimated standard errors are wrong, meaning that we should cluster the standard errors over firms. In case we don't, the coefficient will remain the same, but the inference of the coefficient will be incorrect since we cannot compute the t-statistic correctly. To implement the clustered standard errors, we run the same regression, but cluster over the firms using the 'IDCODE' in the data set.

When we look at Table 2 - Augmented, we see that the coefficients in Model 1 with- and without clustered standard errors remained the same. The standard errors of all the coefficients have increased in Model 1 with the clustered standard errors, meaning that this could influence the statistical significance of the coefficients. However, this is not the case in Model 1 with clustered standard errors, where all the coefficients still remain significant at a 1% significance level. Note, that our inference should be interpreted with caution, since we conclude the significance of the coefficients using a total number of clusters that is too low. In our case, we cluster over 36 firms, while instead we want the number of clusters to be equal or greater than 50 for true asymptotic inference.

Table 2 - Augmented

Factors affecting the Excess Stock Return

VARIABLES	Model 1	Model 2 (Including Economic Factors)	Model 1 (Clustered Std Errors)
MktRF	1.083***	1.081***	1.083***
	(0.0165)	(0.0166)	(0.0540)
SMB	0.0156	0.0163	0.0156
	(0.0239)	(0.0239)	(0.0595)
HML	0.237***	0.232***	0.237***
	(0.0306)	(0.0308)	(0.0839)
RMW	0.374***	0.378***	0.374***
	(0.0311)	(0.0313)	(0.0572)
CMA	0.227***	0.228***	0.227***
	(0.0449)	(0.0451)	(0.0579)
Unemp		0.0465	
		(0.0417)	
Inflation		0.332	
		(0.255)	
ln_volume		-0.0568	
		(0.0392)	
Constant	-0.0484	0.327	-0.0484
	(0.0664)	(0.562)	(0.0683)
Observations	14,688	14,688	14,688
R-squared	0.246	0.247	0.246
R-adjusted	0.246	0.246	0.246
F-stat	960.3	601	191.9

This table presents the regression results of model 1, 2 and model 1 including Clustered Standard Errors. Model 1 is the regression of the Excess Stock Return on the Market Risk Premium (MktRF), the Small-Big Factor (SMB), the High-Low Factor (HML), the Profitability Factor (RMW) and the Investment Factor (CMA). Model 2 is an extension of model 1, including the Macro Economic Factors; Unemployment Rate and Inflation Rate, and the natural log of Trading Volume of the stock. The fourth column contains the regression results of Model 1 including IDCODE Clustered Standard Errors. The standard errors are reported in the parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted by ***, ** and *, respectively. (N = 14688)

5 Question 1e

Model 2, Industry FE: Excess Stock Return on log Volume and FF and Macro Economic Factors.

$$ExRet_{i,t} = \alpha + \beta_1 MkrtRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 Unemp_t + \beta_7 Inflation_t + \beta_8 lnVolume_{i,t} + \sum_{j=2}^{10} \delta_j Industry_i^j + \epsilon_{i,t} \tag{4}$$

Interpretation of the Excess Market Return Variable

β_1 : For a stock in a given industry, if the Excess Market Return goes up by 1 percentage point then the Excess Stock Return will increase by 1.081 percentage points, corrected for the other factors.

To determine whether or not we really need the industry fixed effects, we look at the F-statistic, calculated in equation 5, which tests the hypotheses in hypothesis set 1. If we look at the $F_{statistic}$ in equation 6, we can see that it is not larger than the F_{crit}. Thus, we cannot reject the null-hypothesis in hypothesis set 1, meaning that the coefficients of the included Industry Dummy variables are all equal to zero. We conclude, based on the F-statistic, that we don't really need the industry fixed effects in our model.

Hypothesis Set 1

- H_0: $\delta_2 = \delta_3 = ... = \delta_{10} = 0$
- H_A: At least one of the included Industry Dummy coefficients is not equal to zero.

$$F_{statistic} = \frac{RSS_0 - RSS_1}{RSS_1} \frac{N - k_1}{k_1 - k_0} \sim F(k_1 - k_0, N - k_1) \tag{5}$$

where the RSS_0, RSS_1, k_0 and k_1 denote the Residual Sum of Squares and the number of parameters in model 2 without- and model 2 with industry fixed effects, respectively.

$$F_{statistic} = \frac{853234.39 - 853078.67}{853078.67} \frac{4688 - 18}{18 - 9} = 0.3 \sim F_{5\%(9,14670)} \tag{6}$$

$$F_{crit} \approx 2.70 \sim F_{5\%(9,14670)} \tag{7}$$

6 Question 1f

In order to construct a Time Fixed Effects model, we include Monthly Time Dummies into the model equation. Since all the FF and macro economic factors have the same value for all stocks at time point t (collinearity), these will be omitted in a time fixed effects model.

Model 2, Time FE: Excess Stock Return on Log Volume.

$$ExRet_{i,t} = \gamma_t + \beta_1 Ln\_Volume_{i,t} + \epsilon_{i,t} \tag{8}$$

where i and t denote the IDCODE of the stock and the time in Year/Month, respectively. Furthermore, Ln is used to indicate that a variable has been log transformed.

Interpretation of the Log Volume Variable

β_1 : For stocks at a given time (Year/Month), if the trading volume increases by 1% then the Excess Stock Return will decrease by 0.00133 percentage points, corrected for the other factors.

7 Question 1g

Similar to Question 1f, we again omit the factors that have collinearity.

Model 2, Time & Firm FE: Excess Stock Return on Log Volume.

$$ExRet_{i,t} = \alpha_i + \gamma_t + \beta_1 Ln\_Volume_{i,t} + \epsilon_{i,t} \tag{9}$$

where i and t denote the IDCODE of the stock and the time in Year/Month, respectively. Furthermore, Ln is used to indicate that a variable has been log transformed.

Interpretation of the Log Volume Variable

β_1 : For a given stock at a given time (Month/Year), if the trading volume increases by 1% then the Excess Stock Return will decrease by 0.00354 percentage points, corrected for the other factors.

8 Question 1h

To determine which of the four models is preferred, we use the F-statistics in equation 5, we've used before to calculate the necessity of the included fixed effects. In question 1e, we have already concluded that we don't need Industry Fixed Effects in model 2. We use the same testing approach to test whether we need Time Fixed Effects or Firm and Time Fixed Effects in model 2. Equation 10 and 11 denote the F-statistics for Time Fixed Effects and Firm and Time Fixed Effects, respectively. Equation 10 denotes the F-statistic of adding Time Fixed Effects to Model 1, indicating that the null-hypothesis of this F-test, denoted in Hypothesis Set 2, can be rejected using the critical value in equation 7. This means that at least one of coefficients of the added Time Fixed Effects is significantly different from zero, therefore, we should add these Time Fixed Effects in the model.
Looking at equation 11, we cannot reject the null-hypothesis, denoted in Hypothesis Set 3, using the critical value in equation 7. This means that the coefficients of the added Firm Fixed Effects onto the Time Fixed Effects model are equal to zero and that it is not necessary to include Firm Fixed Effects. Therefore, we conclude that model 2 with only Time Fixed Effects (Question 1f) is the best model out of the total four. This model explains more than the original model 2 with no fixed effects and is also more parsimonious than model 2 with firm and time fixed effects.

$$F_{TimeFixed} = \frac{1131489.26 - 819097.75}{819097.75} \frac{14688 - 409}{409 - 2} \approx 13.38 \sim F_{5\%(407,14279)} \tag{10}$$

$$F_{FirmFixed} = \frac{819097.75 - 817946.44}{817946.44} \frac{14688 - 444}{444 - 409} \approx 0.57 \sim F_{5\%(35,14244)} \tag{11}$$

Hypothesis Set 2

- H_0: All the coefficients of the included Time Fixed Effects are equal to zero.
- H_A: At least one of the included Time Fixed Effects Coefficients is not equal to zero.

Hypothesis Set 3

- H_0: All the coefficients of the included Firm Fixed Effects are equal to zero.
- H_A: At least one of the included Firm Fixed Effects Coefficients is not equal to zero.

Table 3

Factors affecting the Excess Stock Return (including multiple fixed effects)

VARIABLES	Model 2	Model 2 (Industry Fixed Effects)	Model 2 (Time Fixed Effects)	Model 2 (Time and Firm Fixed Effects)
ln_volume	-0.0568	-0.0644	-0.133**	-0.354***
	(0.0392)	(0.0416)	(0.0567)	(0.110)
MktRF	1.081***	1.081***		
	(0.0166)	(0.0166)		
SMB	0.0163	0.0165		
	(0.0239)	(0.0239)		
HML	0.232***	0.232***		
	(0.0308)	(0.0308)		
RMW	0.378***	0.377***		
	(0.0313)	(0.0313)		
CMA	0.228***	0.228***		
	(0.0451)	(0.0451)		
Unemp	0.0465	0.0467		
	(0.0417)	(0.0417)		
Inflation	0.332	0.323		
	(0.255)	(0.255)		
Constant	0.327	0.499	0.376	2.720
	(0.562)	(0.627)	(1.398)	(1.722)
Industry_FE	NO	YES	NO	NO
Time_FE	NO	NO	YES	YES
Firm_FE	NO	NO	NO	YES
Observations	14,688	14,688	14,688	14,688
#Firms				36
#Year_Months			408	408
#Industries		10		
R-squared	0.247	0.247	0.277	0.277
F-stat	601	282.8	13.40	13.40

This table presents the regression results of model 2 with Industry, Time and Firm and Time Fixed Effects. Model 2 is the regression of the Excess Stock Return on the Market Risk Premium (MktRF), the Small-Big Factor (SMB), the High-Low Factor (HML), the Profitability Factor (RMW) and the Investment Factor (CMA), Unemployment Rate, Inflation Rate and the natural log of Trading Volume of the stock. Column 2 contains model 2 with Industry Fixed Effects, Column 3 contains model 2 with Time Fixed Effects (omitted collinearity variables) and Column 4 contains model 2 with Time and Firm Fixed Effects (omitted collinearity variables). The standard errors are reported in the parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted by ***, ** and *, respectively. (N = 14688)

9 Question 1i

Model 3: Excess Stock Return on Log Volume, Introduction- and Affection Dummies of 2001 Tax Law.

$$ExRet_{i,t} = \alpha + \beta_1 lnVolume_{i,t} + \beta_2 AfterLaw_t + \beta_3 Affected_i + \beta_4 AfterLaw_t \times Affected_i + \epsilon_{i,t} \quad (12)$$

where the variables AfterLaw and Affected denote generated dummy variables for stock observations at or after January 2001 and stocks that are affected by the Tax Law (Financial and Health Care industry), respectively.

β_2: Time effect in Excess Stock Returns

For a stock that is not in the Financial or Health Care Industry at or after January 2001, the Excess Stock Return will be 0.0414 percentage points lower, compared to a stock that is not in the Financial or Health Care Industry at a time point before January 2001, corrected for the other factors.

β_3: Difference in Excess Stock Returns between Finance and Health Care and other industries, already before the Tax Law introduction

For a stock that is in the Financial or Health Care Industry at a time point before January 2001, the Excess Stock Return will be 0.374 percentage points higher, compared to a stock that is not in the Financial or Health Care Industry at a time point before January 2001, corrected for the other factors.

β_4: The causal effect of the Tax Law introduction on the Excess Stock Return

This is the effect of the tax law on the excess stock return, meaning that the Excess Stock Return decreases with 0.480 percentage points more, than it would have without the Tax Law Introduction.

To determine whether there is an actual effect of the Tax Law that is introduced in January 2001, we test whether the β_4 is statistically significant or not. Equation 13 denotes the t-test that has been performed, with the results in equation 14. Since the t_{stat} is not lower than the t_{crit}, we cannot reject the null-hypothesis in Hypothesis Set 4. Therefore, we conclude that there is no significant (at 5%) causal effect of the Tax Law introduction in January 2001 on the Excess Stock Returns on the two affected industries.

Hypothesis Set 4

- H_0: $\beta_4 = 0$
- H_A: $\beta_4 \neq 0$

$$t_{stat} = \frac{\hat{\beta}_4 - \beta_4}{S.E.} \sim t(0, 1, N - k) \quad (13)$$

$$t_{stat} = \frac{-0.480}{0.351} = -1.368 \quad (14)$$

$$t_{crit,5\%} = -1.96 \quad (15)$$

Table 4

Difference in Difference analysis

VARIABLES	Model 3
after_law	-0.0414
	(0.209)
affect_industry	0.374
	(0.262)
c.after_law#c.affect_industry	-0.480
	(0.351)
ln_volume	-0.144**
	(0.0596)
Constant	2.741***
	(0.693)
Observations	14,688
R-squared	0.001
F-stat	4.592

This table presents the regression results of model 3, which is the Excess Stock Return on the Natural Log of the Trading Volume of the stock, the dummy for the Tax Law Introduction (AfterLaw), the Tax Law Affection (Affected) and the cross product of both the Introduction and Affection Dummy. The AfterLaw dummy equals one when the time date t is at or after January 2001, it is zero otherwise. The Affected dummy equals one when the industry of the stock is equal to the Financial or the Health Care Industry, it is zero otherwise. The cross product of both of the dummies equals one if the stock is both at or after January 2001 and belongs to the affected industries. The standard errors are reported in the parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted by ***, ** and *, respectively. (N = 14688)

```
//Prepare Workspace
clear all
ssc install outreg2
ssc install asdoc

//Change directory
cd "/Users/markdekwaasteniet/Documents/Master Finance/Empirical Finance/Case 2"

//Data structurizing
import excel "/Users/markdekwaasteniet/Documents/Master Finance/Empirical Finance/Case 2/data_case_II_group_56.xls", sheet("Data") firstrow

// Exercise 1a

// Create the variable: excess_ret
gen excess_ret = RETURN-RF
label var excess_ret "Return over the Risk Free rate"

// Creating the log volume variable using the natural log
gen ln_volume=ln(VOLUME)


//Test the normality of all variables except Date
sktest excess_ret VOLUME IDCODE MktRF SMB HML RMW CMA RF Unemp Inflation

mat sktest = r(Utest)

//Subtract the last column of the matrix including the p-values
matrix Pvalues = sktest[.,4...]

//Summarize all the variables except Date
//Sut the variables into a table and convert to matrix
tabstat excess_ret ln_volume IDCODE MktRF SMB HML RMW CMA RF Unemp Inflation, statistic(mean sd min max) columns(statistics) save
mat statistics = r(StatTotal)'

//Merge the two matrices together
mat table1 = statistics,Pvalues\Pvalues,statistics
mat table1 = table1[1..10,1...]

//Output the matrix
asdoc wmat, mat(table1) title(Summary Statistics Table 1) dec(3) replace
//Rename Myfile.doc name to Table 1.doc

//Set Panel Data
gen year_month = ym(StataYear,StataMonth) //Create own time variable that is easier to read
xtset IDCODE year_month, monthly

// Exercise 1b
// Model 1:
```

```
reg excess_ret MktRF SMB HML RMW CMA
outreg2 using "Table 2.tex", replace word addstat(R-adjusted, e(r2_a), F-stat, e(F))
ctitle(Model 1) title(FF and Economic Factors affecting the Excess Stock Return)

// Model 2:
reg excess_ret MktRF SMB HML RMW CMA Unemp Inflation ln_volume
outreg2 using "Table 2.tex", append word addstat(R-adjusted, e(r2_a), F-stat, e(F))
ctitle(Model 2 - Including Economic Factors) title(FF and Economic Factors affecting the
Excess Stock Return)


// Exercise 1d
// Use clustered standard errors.
reg excess_ret MktRF SMB HML RMW CMA, vce(cluster IDCODE)
outreg2 using "Table 2.tex", append word addstat(R-adjusted, e(r2_a), F-stat, e(F))
ctitle(Model 1 - Clustered Std Errors) title(Factors affecting the Excess Stock Return)


// Exercise 1e
reg excess_ret MktRF SMB HML RMW CMA Unemp Inflation ln_volume
outreg2 using "Table 3.tex", replace word addstat(F-stat, e(F)) ctitle(Model 2) title(Factors
affecting the Excess Stock Return (including industry effects))


//Create Industry fixed effects model using industry dummies:
reg excess_ret MktRF SMB HML RMW CMA Unemp Inflation ln_volume i.Industry
outreg2 using "Table 3.tex", append word addstat(F-stat, e(F)) ctitle(Model 2 - Industry Fixed
Effects) title(Factors affecting the Excess Stock Return (including industry effects))


// Exercise 1f
// Per month Time FE model:
reg excess_ret ln_volume i.year_month
outreg2 using "Table 3.tex", append word addstat(F-stat, e(F)) ctitle(Model 2 - Time Fixed
Effects) title(Factors affecting the Excess Stock Return (including industry effects))


// Exercise 1g
xi i.year_month, pre(Y)
xtreg excess_ret ln_volume Yyear_month*, fe
outreg2 using "Table 3.tex", append word addstat(F-stat, e(F)) ctitle(Model 2 - Time and
Firm Fixed Effects) title(Factors affecting the Excess Stock Return (including industry
effects))


// Exercise 1h
//Ftest for time fixed effects
reg excess_ret ln_volume i.time
testparm i.time
```

```
//Ftest for firm fixed effects on top of time fixed effects.
reg excess_ret ln_volume i.time i.IDCODE
testparm i.IDCODE

// Exercise 1i

// Create two new dummy variables. 1 is the before event. 1 is for the affected firms.
// January 2001 is equal to time = 181
gen after_law = 0
replace after_law = 1 if time >= 181

gen affect_industry = 0
replace affect_industry = 1 if (Industry == 2 | Industry == 6)

// then we also let the dummy variables interact with each other in the model.
reg excess_ret after_law affect_industry c.after_law#c.affect_industry ln_volume
outreg2 using "Table 4.tex", replace word addstat(F-stat, e(F)) ctitle(Model 3 )
title(Difference in Difference analysis)
```