Introduction to Econometrics: Assignment 2

Yanis Ferhat (11288574) Félix Poirier (11298801) Joel Chrispin (11288382)

Thursday 13th October, 2022

1 Research question and exploration

1.1 Our research question

Our research question is the following: can the Fama-French 3 factor model explain Canadian technology stock returns over the past decade? In recent history, a popular research subject in portfolio theory and asset pricing has been the use of factor models to explain stock returns. This has led to many known models such as the Capital Asset Pricing Model (CAPM), Arbitrage Price Theory (APT) and the Fama-French multi-factor models. In this work, we are interested in the famous Fama-French 3 factor model, which states that portfolio or stock returns can be mostly explained by 3 variables: overall market returns, the performance of value stocks (value), and the performance of small-cap stocks (size). Since the 3 factor model is widely known to be significant in explaining returns, we expect our regression to be quite strong. We also anticipate overall market returns to clearly be the most significant explanatory variable, since the industry already considers these returns to have strong a guiding influence on stock and portfolio returns.

The stock returns we wish to explain will be pulled from the Canadian technology sector and will be represented by returns on the XIT ETF. This will be the dependent variable of our model, and will be called **tech**. The market variable, named **market**, will be represented by returns on the XIC ETF. The value factor variable, termed **value**, will be represented by the difference between returns on the XCV ETF and the XCG ETF, approximating the difference in returns on value stocks and growth stocks. The size variable, which is the difference between returns on small stocks and large stocks, will be approximated by the difference between returns on the XCS ETF and the FIDCNLCC index. In our econometric model, this explanatory variable is titled Size. The timeframe for all the returns is from November 2nd, 2012 to September 30th, 2022. Furthermore, weekly returns were used for each variable. The data was extracted manually from a Bloomberg terminal at HEC's National Bank Trading Room.

1.2 Data source

Bloomberg Terminal, 2022, "XIT" "XIC" "XCV" "XCG" "XCS" "FIDNLCC" (accessed September 8th, 2022)

1.3 Summary statistics

Table 1: Summary statistics for *tech*, *market*, *value* and *size*

	Tech	Market	Value	Size
count	517.000000	517.000000	517.000000	517.000000
mean	-0.003133	-0.000783	-0.000799	-0.000167
std	0.031860	0.019804	0.022181	0.027590
min	-0.151192	-0.090930	-0.095603	-0.134211
25%	-0.021413	-0.011619	-0.012017	-0.014185
50%	-0.004975	-0.002268	-0.001781	-0.002370
75%	0.013590	0.007087	0.007886	0.012890
max	0.140315	0.165685	0.211631	0.264327

The average weekly return on the XIT ETF during the specified time frame, represented by the **tech** variable, is around -0.31%. The variable's standard deviation, defined as the average distance between an observation and the mean, is approximately 0.0319, or 3.19%. Finally, the variable's range of values spans from about -15.12% to 14.03% in weekly returns. The **market** variable, characterized as weekly returns on the XIC ETF, has a mean of approximately 0.078%. Roughly, the variable's values range from -9.09% to 16.57% with a standard deviation of 0.0198, or 1.98%. The Value variable, depicting the difference in returns on the XCV and XCG ETFs, has a rounded average value of -0.08%. This indicates that in the extracted sample, on average, weekly returns on the XCV ETF are lower than weekly returns on the XCG ETF. The variable's standard deviation is approximately 2.22% and its values range from about -9.56% to 21.16%. **size**, the variable illustrating the difference between returns on the XCS ETF and the FIDCNLCC index, has a mean value of about -0.02%, showing that, on average, the sample contains better weekly index performances. Finally, the variable has a 2.76% standard deviation and a dispersion of values going from -13.42% to 26.43%, approximately.

2 The econometric model

This work is centered around the following multivariate econometric model:

$$tech_i = \beta_0 + \beta_1 market_i + \beta_2 value_i + \beta_3 size_i + u_i$$

Using the OLS function from Python's *statsmodels* package, we were able to estimate the above theoretical model. Here are the results (replace the table with Latex OLS:

Table 2: OLS Results for the Effect of the Fama-French 3 factor model (**market**, **value** and **size**) on the XIT ETF's weekly returns (**tech**)

Dep. Variable:	Tech	**R-squared:**	0.499
Model:	OLS	**Adj. R-squared:**	0.496
Method:	Least Squares	**F-statistic:**	170.1
Date:	Thu, 06 Oct 2022	**Prob (F-statistic):**	1.56e-76
Time:	14:53:06	**Log-Likelihood:**	1227.2
No. Observations:	517	**AIC:**	-2446.
Df Residuals:	513	**BIC:**	-2429.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-0.0022	0.001	-2.247	0.025	-0.004	-0.000
Market	2.8998	0.166	17.423	0.000	2.573	3.227
Value	-1.6893	0.131	-12.932	0.000	-1.946	-1.433
Size	-0.1698	0.072	-2.352	0.019	-0.312	-0.028

Omnibus:	12.445	**Durbin-Watson:**	2.194
Prob(Omnibus):	0.002	**Jarque-Bera (JB):**	23.633
Skew:	-0.007	**Prob(JB):**	7.38e-06
Kurtosis:	4.047	**Cond. No.**	204.

Table 3: OLS estimation of coefficients

coefficients	**estimations**
$\hat{\beta}_0$	-0.0022
$\hat{\beta}_1$	2.8998
$\hat{\beta}_2$	-1.6893
$\hat{\beta}_3$	-0.1698

Here is the theoretical econometric model's estimated OLS equation:

$$\widehat{tech_i} = -0.0022 + 2.8998 market_i - 1.6893 value_i - 0.1698 size_i$$

2.1 Interpretation of the estimated coefficients

The model's estimated intercept, denoted as $\hat{\beta}_0$, has a value of -0.0022. This illustrates that when **market**, **value** and **size** are simultaneously attributed null values, **tech**, or the weekly return on the XIT ETF, tends to have a value of -0.22%. Such a scenario is almost never observed in practice, rendering this interpretation nearly uninformative.

The estimated coefficient for **market** ($\hat{\beta}_1$) is 2.8998. This indicates that when the weekly return on the XIC ETF increases by 0.01, or 1 percentage point, and this increase is not correlated with changes in the other explanatory variables, weekly returns on the XIT ETF tend to be higher by 0.028998, or about 2.90%. If the four assumptions supporting multivariate OLS regression are respected, then this relationship is of causal nature. In such a situation, it could be stated that better overall market performance causes better individual portfolio or stock performance.

value's estimated coefficient ($\hat{\beta}_1$) is -1.6893. This implies that when the difference between weekly returns on the XCV and XCG ETFs increases by 0.01, or 1 percentage point, and this increase is not correlated with variations in the values of the other explanatory variables, weekly returns on the

XIT ETF tend to be lower by 0.016893, or about 1.69%. If the four assumptions supporting multivariate OLS regression are respected, this relationship can be defined as causal, meaning increased value stock performance leads to worse individual portfolio or stock performance.

Finally, the estimated coefficient for the **size** variable ($\hat{\beta_3}$) is -0.1698, arguing that when the difference between weekly returns on the XCS ETF and the FIDCNLCC index increases by 0.01, or 1 percentage point, and this increase is not correlated with fluctuations in the values of the other explanatory variables, weekly returns on the XIT ETF tend to be lower by 0.001698, or about 0.17%. If the four assumptions supporting multivariate OLS regression are respected, this relationship can be defined as causal, meaning increased value stock performance leads to worse returns for stocks in the technology sector.

2.2 OLS properties

For the above estimated coefficients to be without bias, the four following assumptions, also known as the estimators' OLS properties, must be respected:

T-statistics, p-values and confidence intervals

To evaluate each coefficient's significance, the following two-tailed generic hypothesis test will be assessed at the 5% significance level.

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$
$$\forall j = 0, 1, 2, 3$$

The t-statistic is a measurement of an estimation's precision. It follows that the higher an estimator's t-statistic, the more precise the estimation is and the more likely its associated explanatory variable has a statistically significant impact on the dependent variable. Considering the chosen 5% significance level and the model's 513 degrees of freedom, each t-statistic will be compared to a 1.960 critical value (in absolute value). While the significance level defines the maximum probability at which an error would be made by rejecting the null hypothesis when it is in fact true in the unobserved population, the p-value measures the actual probability of committing such an error. Therefore, a p-value below a specified significance level, which in this case is of 5%, is a strong indicator of a coefficient's statistical significance. Thus, both the t-statistic and the p-value can be used to assess a coefficient's statistical significance.

The t-statistics for the **intercept** β_0, **market** β_1, **value** β_2 and **size** β_3 are, in absolute values, 2.247, 17.423, 12.932 and 2.352, respectively. These values are all higher than the 1.960 critical value stated earlier. In consequence, at a 5% significance level, the null hypothesis may be rejected in favor of the alternative hypothesis within all four hypothesis tests. We can therefore conclude that, at the 5% significance level, the model's true intercept is non null and the three explanatory variables have a statistically significant impact on Tech, or, more precisely, on returns on the XIT ETF.

The p-values for the **intercept** β_0, **market** β_1, **value** β_2 and **size** β_3 are 0.025, 0, 0 and 0.019 respectively. In other words, in all four hypothesis tests, the actual probabilities of making an error by rejecting the null hypotheses when they hold in reality are 2.5%, 0%, 0% and 1.9%, respectively. Since these probabilities all lie below 5%, we can again conclude that, at the 5% significance level, the model's true intercept is non null and the three explanatory variables have a statistically significant impact on Tech, the model's dependent variable.

The software used to obtain the OLS estimators calculated confidence intervals for each coefficient with a 95% confidence level. This means that, for each population parameter, the probability that its value lies within the bounds of its estimator's confidence interval is 95%. Therefore, at the 95% confidence level, we estimate that the model's true **intercept** (β_0) lies between -0.004 and 0, that the model's true **market** (β_1) lies between 2.573 and 3.227, that the model's true **value** (β_2) lies between -1.946 and -1.433, and that the model's true **size** (β_3) lies between -0.312 and -0.028. Note that all mentioned confidence intervals are inclusive.

3 F-statistic and R^2

The F-statistic of our model, as presented in the regression results above, it is 170.1. Here, the F-test is testing for the following hypotheses:

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$
$$\forall j = 0, 1, 2, 3$$

Here, the unconstrained model ($H1$) is the initial model with 3 variables and one constant. The constrained model ($H0$) is removing the 3 variables from the model, leaving only the constant β_0 as the explanatory variable. The F-statistic therefore tests whether the explanatory variables of our model are jointly significant. In other words, we are trying to see whether our model is more powerful with 3 variables than with none. At the 1% significance level, the appropriate critical value to test the null hypothesis is 3.78. The F-statistic being much larger (170.1), we can therefore reject H_0 and claim that our 3 variables, **market** (β_1), **value** (β_2) and **size** (β_3), are jointly significant and help explain technology stock returns better than the constant β_0 alone.

The R^2 of our regression is 0.499. This means that our model is able to explain about 49.9% of the variations in the returns of Canadian tech stocks. We can conclude that although our model shows some significance in explaining the dependent variable, as shown by the T and F statistics, the results are still disappointing. Indeed, there are more than half of the variations of the XIT ETF weekly returns over the past 10 years that can't be explained by the Fama-French 3 factor model. Therefore, although we can affirm that the market, value and size factors are significant in explaining tech stock returns, we can't conclude that our model does a great job of explaining the majority of the variations of tech returns. As we expected in our hypothesis at the beginning, the market is the strongest explanatory variable, the 3 variables are statistically significant. However, we expected to observe a better fit across the residuals of the dependent variable.

The reason we can extrapolate our results to the population is because the OLS estimators are unbiased, since they are the BLUE estimators, according to the Gauss-Markov theorem. Indeed, if the estimators didn't respect the 5 hypotheses of the theorem. our estimators would be biased and we wouldn't be able to conclude that the relationship between the explanatory variables and the dependent variable is causal.

4 Resources

Here's a link the GitHub repository containing the data, the images and the code :

`https://github.com/felixpoirier1/Assignment2`

Here's a direct link to the code :

`https://github.com/felixpoirier1/Assignment2/blob/main/assignment2.ipynb`