COURSE: **LABORATORY OF DATA ANALYTICS FOR INVESTMENT**

STUDENT: **PAOLO BASSI**

IS SIZE ANOMALY STILL VALID IN THE ERA
OF
GROWING ECONOMIES AND SECTORS?

THESIS

Size anomaly is a phenomenon that has been observed both by academics and in the market in the past, and at the same time other professors criticize its legitimacy because it could only be a non-repeatable event in a systematic way. Thus, this small work wants to take a step further: if the size anomaly is real, is the same for all companies regardless of the market sector of belonging or the latter is the relevant factor for this anomaly to occur? And if this is the case, is there any real benefit to this type of diversification across sectors in terms of performances?

This general analysis without entering depth into the macroeconomic aspect, arises from the curiosity to understand how the different sectors are decisive in determining the actual investment returns. Each sector has unique features both positive, competitive, and flexible market, and negative, presence of incumbent and barriers to entry.

Finally, what I would expect to see is that the size anomaly is more evident in the tertiary or service sectors, while I would presume a less obvious phenomenon in the primary or industrial sectors.

ACADEMIC LITERATURE

To begin with, the starting point for this document comes from some important academic papers:

⇒ *The Relationship Between Return and Market Value of Common Stocks by R.W. Banz (1981)*

"The evidence presented in this study suggests that the CAPM is misspecified. On average, small NYSE firms have had significantly larger risk adjusted returns than large NYSE firms over a forty-year period. This size effect is not linear in the market proportion but is most pronounced for the smallest firms in the sample. The effect is also not very stable through time.
"There is no theoretical foundation for such an effect. We do not even know whether the factor is size itself or whether size is just a proxy for one or more true but unknown factors correlated with size.
"To summarize, the size effect exists but it is not at all clear why it exists. Until we find an answer, it should be interpreted with caution."

⇒ *The Cross-Section of Expected Stock Returns by Eugene Fama and Kenneth R. French (1992)*

"There is a negative relation between average return and size.
"Our main result is that two easily measured variables, size and book-to-market equity, seem to describe the cross-section of average stock returns."

⇒ *Multifactor Explanations of Asset Pricing Anomalies by Eugene Fama and Kenneth R. French (1996)*

"Fama and French (1993) find that the three-factor risk-return relation is a good model for the returns on portfolios formed on size and book-to-market-equity."

To sum up:
Various empirical tests show that stock returns cannot be fully described by a single factor, as assumed by the CAPM.

The most important mathematical models developed subsequently are the well known: the Three Factor Model (1993) by Fama and French, which considers, in addition to the traditional market factor of the CAPM, also the value and size factors.

The idea that stock returns may manifest some level of predictability, and may therefore not be totally random, gained some credibility beginning in 1993, when E. Fama and K. French show that equity portfolios overweighted on value or small-cap stocks tend to outperform the market portfolio over the long term. In their three-factor model, Fama and French assert that the excess return on a stock - $E(R_i)-R_f$ - is explained by the co-movement of its returns with three factors:

1. The excess return of a market portfolio $E(R_M)-R_f$, as in the CAPM (also known as market factor);
2. The difference in the returns of a portfolio of small-cap stocks and the returns of a portfolio of large-cap stocks (SMB), also known as the size effect;
3. The difference in returns between a portfolio of value stocks - i.e. stocks with a high book-to-market (Book-to-market is a multiple used to find the value of a company by comparing the book value of its net worth with its market capitalization value). - and a portfolio of growth stocks - i.e. stocks with low book-to-market (HML, from "High Minus Low"), also known as the "value" effect.

For relevance to the experiment we want to carry out, we will only consider the size factor in detail.

This type of anomaly is generally attributable to so-called "distress risk"; in fact, companies with lower capitalization suffer from less liquidity and greater economic problems than larger companies, and offer lower returns precisely when investors' marginal utility is high, i.e. during recessions.

PROCEDURE

Let's try to find the answers to these questions with the help of two methods:

1) **Approach A** analyzes the capitalization factor and the sectors by using two different types of equally weighted full-invested portfolio strategies:

   a) One that diversifies across market capitalizations but invests in one sector only.
   b) One that spans across sectors, while investing in one segment of market cap a time.

2) **Approach B** considers the industries in their entireness and their link with factor size from Fama & French factor model.

---

1) Approach A

Therefore, we start from the first method, and we initially divide companies according to:

- **Market capitalization** (i.e., small and big cap) exploiting the scheme states in Fama & French 3 factor model in which size is obtained by considering the median capitalization of companies and then, if the cap is greater than this threshold, the firm is considered as a big cap, otherwise as a small cap.

- **Sectors**, term which refers to a classification used to describe companies that belong to the same segment of the economy, namely:
   1. Energy
   2. Industrials
   3. Real Estate
   4. Financials
   5. Information Technology
   6. Consumer Discretionary
   7. Health Care
   8. Materials
   9. Utilities
   10. Communication Services
   11. Consumer Staples

Pre-process:

Considering the data provided (close prices, capitalization and industries), I first handled missing values (companies with missing data even just in one dataset have not been considered), calculated the interquartile range and removed the values from the above range in order to obtained clearer results.

After controlling for correspondence between available companies' data both in close prices matrix and in capitalization matrix, I computed the median capitalization to divide firms according to size as explained above.

The next step has regarded the computation of the annualized average returns for each firm and the partition of these ones in small and big datasets.

Afterward, I extracted the sectors which matched the companies and concatenated them with firm's returns and market capitalization to build a dictionary with two keys: the first is about the industries, and the second concerns the capitalization (big or small); in addition to this there are the values which regard the returns for each company.
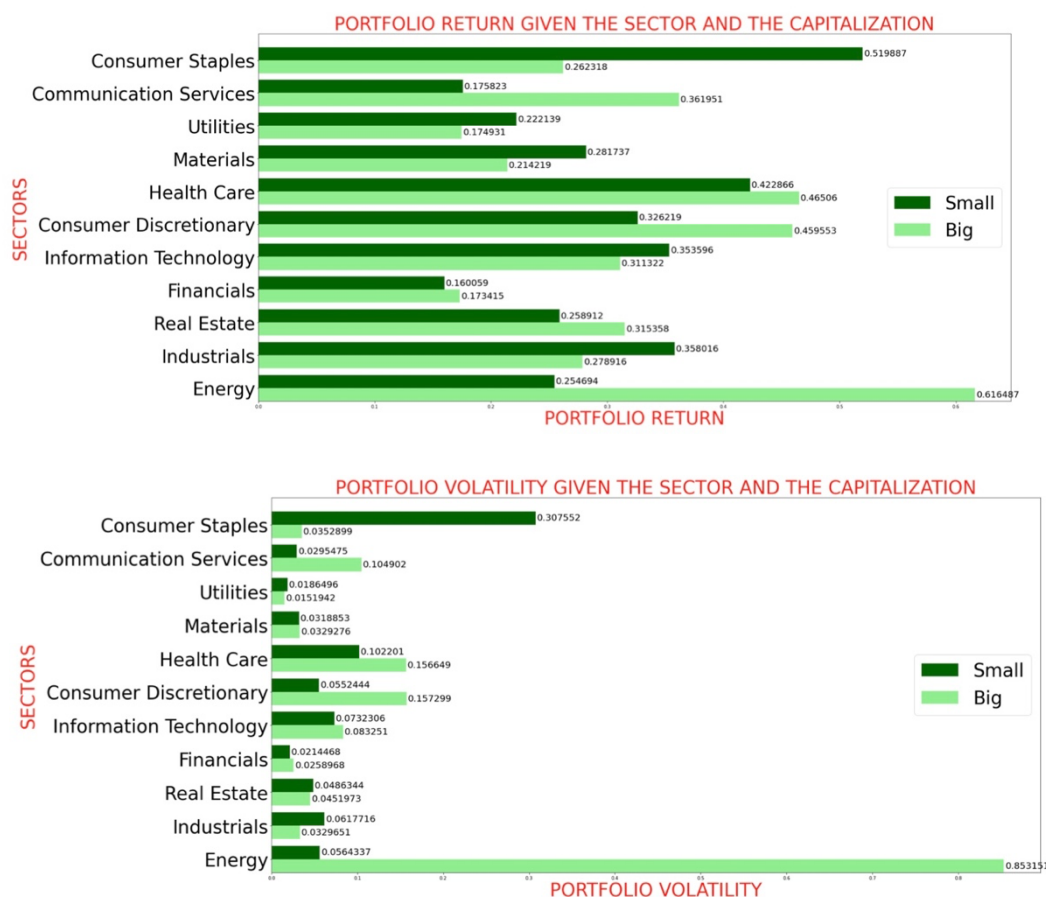
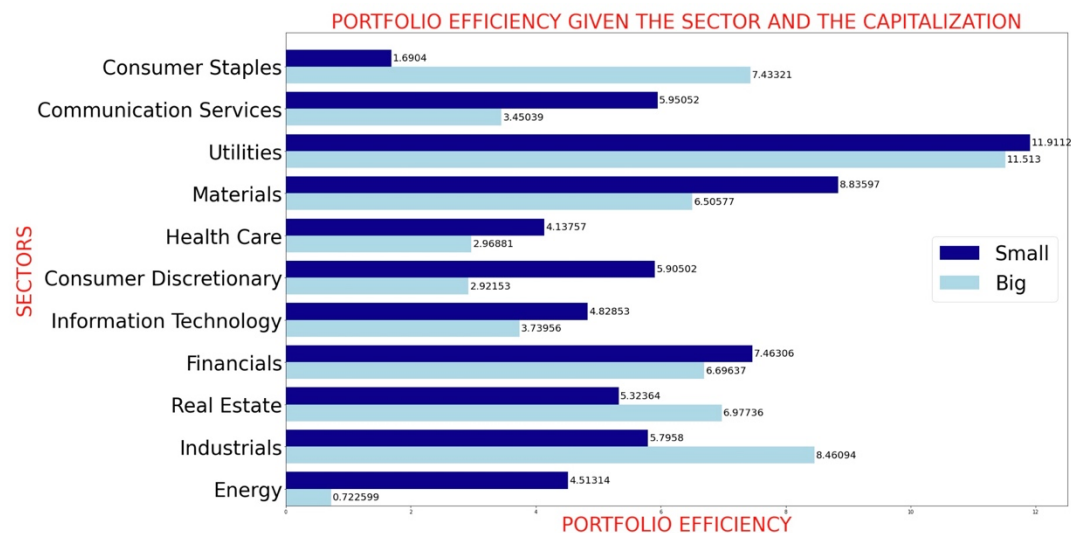Finally, all the needed data to build the different portfolio strategies previously describe were ready.

a) **First strategy**:

Hence, the first strategy implemented wants to compute the expected return, the volatility, and the efficiency for different portfolio with the respect to market capitalization but considering just one sector at a time.
As said before, I assumed equal-weighted, full invested portfolios for the sake of simplicity.
Below are shown the results.

PORTFOLIO EFFICIENCY GIVEN THE SECTOR AND THE CAPITALIZATION

Sector	Small	Big
Consumer Staples	1.6904	7.43321
Communication Services	5.95052	3.45039
Utilities	11.9112	11.513
Materials	8.83597	6.50577
Health Care	4.13757	2.96881
Consumer Discretionary	5.90502	2.92153
Information Technology	4.82853	3.73956
Financials	7.46306	6.69637
Real Estate	5.32364	6.97736
Industrials	5.7958	8.46094
Energy	4.51314	0.722599

Starting from the first graph and looking at the results the first conclusions we can draw are the following:

Only 5 portfolios out of 11 reflect the hypothesis of Fama & French regarding size effect. In particular, some big portfolios overperform by far the smaller ones in some sectors. Therefore, we might already infer that size anomaly depends also on the industry considered and not just the size of companies.

On the other hand, the chart regarding portfolio volatility shows us that just 4 out of 11 higher volatile portfolio stems from small cap firms. This is somehow strange because we could expect more volatility associated with smaller size since big companies are usually stable.

Putting all together, we can see that there is a general correlated trend in return and volatility: the more profitable portfolios are also the more volatile ones, i.e., if the lucrative portfolio in one sector matches with the smaller (or the larger), it is likely that the latter is also the most volatile one. Nothing new so far because we are considering risky assets therefore higher returns are associated with higher volatility on average.

Finally, considering the efficiency of single portfolios we can see how, instead, 8 out of 11 small companies overperform the corresponding large companies. It seems plausible given that Real Estate and Industrials sectors are evergreen segments of the economy in which physical assets have a great deal of weight in companies' financials, and thus also in dimension both in effective dimension and in market capitalization.
This last point seems to be evidence of Fama & French conclusion regarding size anomaly.
In fact, independently of the sector, we have a generally higher performance from smaller portfolios.

b) Second strategy:

The second strategy, instead, considers two types of portfolios regarding size but this time computed across sectors.

The first graph shows the expected return, the second the volatility while the last one the efficiency of the two portfolios.

At the first glance the outcome appears somehow reversed, because we can clearly see how the big companies overperform the small ones though at the expense of a very higher volatility. However, considering the efficiency of the given portfolios we can acknowledge again how actually companies with lower capitalization have better results compared to the larger firms even if considered across different sectors.

In terms of diversification benefit we can also notice that having a portfolio made by small capitalization company belonging to different industries results in a lower volatility, in proportion, than the one gotten from a portfolio constructed by companies for each sector.

At the same time, the return is higher by far if we had this kind of portfolio compared to a strategy that considers only the cap and the one-sector 'factor'.

So far, we can draw the following conclusions:

- Size anomaly is real but may differ regarding sectors due to different levels of demand for particular products, materials or services.
- The diversification benefit exists, even if it depends on the single investor's risk appetite as well.

2) Approach B

This approach instead wants to consider industries in their completeness, and this is useful because a further explanation of the previous results could depend on the different types of sectors in which the portfolios have been formed, consequently let's first take a more specific look at each industry while considering obtained outcomes; the following definitions have been found on '*https://www.fidelity.com*'.

Consumer Staples: The Consumer Staples sector consists of companies that provide goods and services that people use on a daily basis, like food, clothing, or other personal products.

Communication Services: The Communication Services sector includes companies that facilitate communication or provide entertainment content and other information through various types of media.

Utilities: The Utilities sector includes companies that engage in the production and delivery of electric power, natural gas, water, and other utility services, such as steam and cooled air.

Materials: The Materials sector consists of companies that are engaged in the manufacturing or processing of chemicals and plastics, or they may harvest forests or extract metals and minerals

Health Care: The Health Care sector includes companies engaged in the production and delivery of medicine and health care–related goods and services.

Consumer Discretionary: Companies in the Consumer Discretionary sector manufacture goods or provide services that people want but don't necessarily need, such as high-definition televisions, new cars, and family vacations.

Information Technology: The Information Technology sector is comprised of companies that offer goods and services, including hardware, software, semiconductors, and consulting services.

Financials: Companies in the Financials sector are engaged in businesses such as banking and brokerage, mortgage finance, and insurance.

Real Estate: The Real Estate sector is primarily made up of companies that own commercial real estate properties. A large portion of the companies are structured as Real Estate Investment Trusts (REITs).

Industrials: The Industrials sector includes companies that manufacture and distribute capital goods in support of industries such as aerospace and defense, construction and engineering, electrical equipment and heavy machinery.

Energy:
The Energy sector consists of companies involved in the exploration, production, or management of energy resources such as oil and gas, as well as companies that service these industries.


First, I downloaded data for each sector from '*https://www.fidelity.com*', then I computed annualized average returns for all the industries and put the results in a single data frame.
Below is shown the graphical outcome.
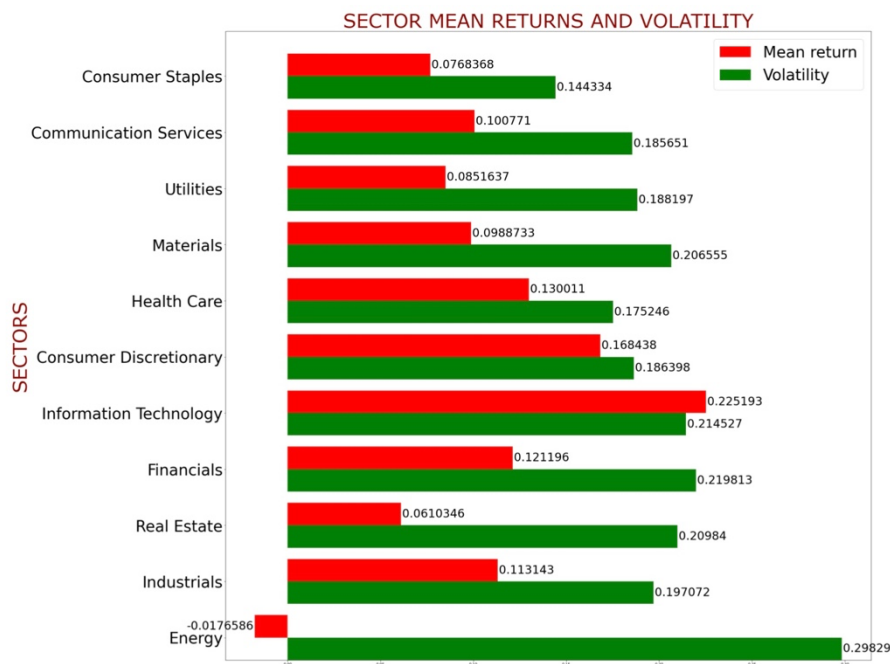
What immediately catches the eye is the still anomalous result of the Energy sector, which here even shows a negative value in terms of return, though with a very high volatility, which likely depends again on oil volatility primary.

Here we have the larger variations and performance's gaps between small and big portfolios. Given that this huge volatility cannot be found in a generally stable economy, we can easily infer that the phenomenon is closely related to change in the price of commodities, in particular oil which has been through a roller coaster in terms of price fluctuation in the past decade, with a volatility of 23.567154.
Below is shown this evidence up to late 2018 (same date range of the datasets).

Natural Gas $ price

Oil $ price
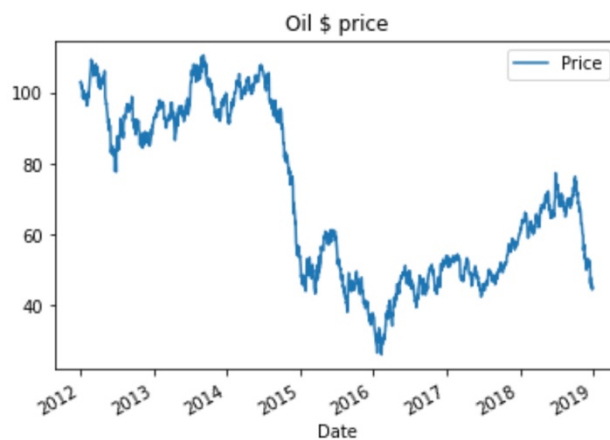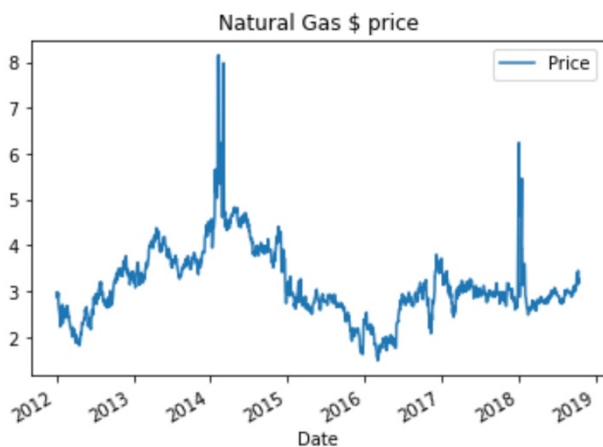
Another seeming result is that volatility is largely higher than the returns in almost every sector. Therefore, we can state that diversify capital according to companies' size (and sectors) instead of just investing in one industry at a time regardless of the dimension give an evident better result.

We want to see if there exist some dependencies between sectors, so we compute correlation using a heatmap.
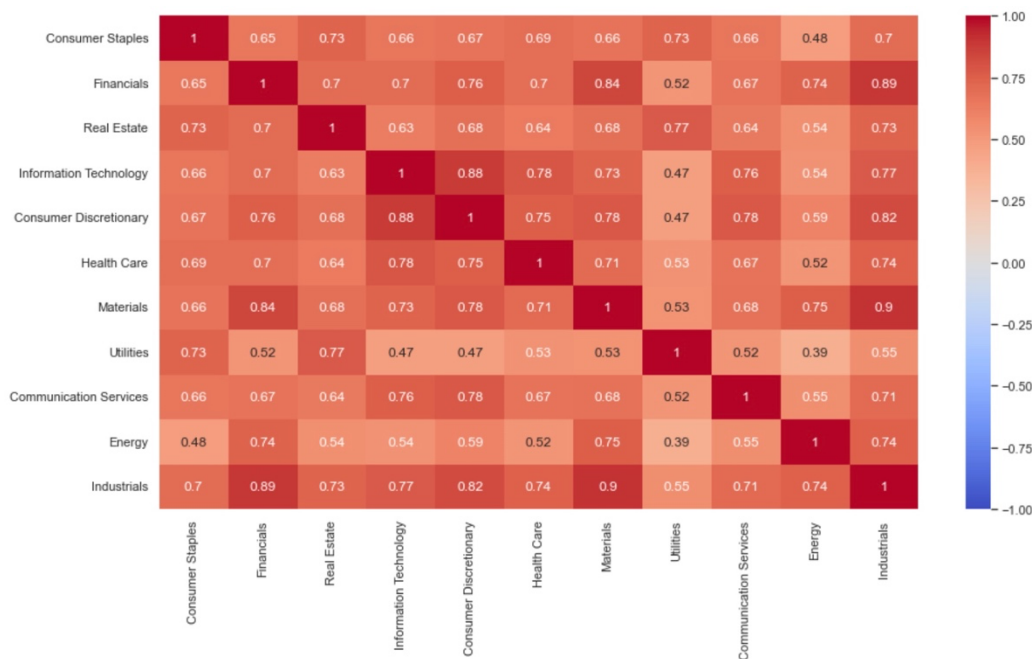
Last step considers Fama-French 3-factor model using the data 'F-F_Research_Data_Factors_daily' downloaded from '*https://mba.tuck.dartmouth.edu*' in order to compute a multivariate regression of sector returns on market return, SMB, and HML factor.
I followed two different ways here, achieving two different results.

a) Compute the regression considering all the values in the data frame below.

Consumer Staples	Financials	Real Estate	Information Technology	Consumer Discretionary	Health Care	Materials	Utilities	Communication Services	Energy	Industrials	Mkt-RF	SMB	HML
-0.003350	0.004268	-0.025132	-0.004439	-0.001623	-0.008539	-0.000356	-0.039556	0.007173	-0.001303	-0.002094	-0.20	0.08	0.36
-0.006387	-0.005579	-0.004158	-0.001137	-0.003734	-0.011286	-0.002139	-0.009631	0.000133	0.002614	-0.000470	-0.46	-0.37	0.11
0.011502	0.006311	0.002923	0.014848	0.011925	0.015620	0.007856	0.019785	0.009423	-0.003475	0.004673	1.05	-0.31	-0.72
0.005351	0.000348	-0.001665	0.004398	0.000000	0.001183	-0.001414	-0.021375	-0.000107	-0.006443	0.000698	0.03	-0.09	-0.33
0.003387	0.010104	0.011259	0.016265	0.011111	0.005196	0.017039	-0.004032	0.001702	0.015248	0.009414	1.00	0.15	-0.06
...	...	...	...	...	...	...	...	...	...	...	...	...	...
-0.002803	0.013575	-0.012788	-0.009630	0.004456	-0.012094	0.010177	-0.011196	-0.001060	0.038516	0.003184	-0.10	1.27	2.39
-0.010775	-0.015023	-0.005505	-0.031521	-0.018943	-0.019740	-0.011170	-0.012527	-0.026111	0.002747	-0.013256	-2.18	-0.42	1.75
0.008700	0.000735	0.007489	-0.002343	-0.000978	0.005776	-0.004651	0.013418	-0.004468	0.000000	0.000757	0.03	-0.29	0.80
-0.017546	-0.014139	-0.017453	-0.005116	-0.018104	-0.009623	-0.014909	-0.009148	-0.001122	-0.013014	-0.018907	-1.02	0.30	-0.77

	Consumer Staples	Financials	Real Estate	Information Technology	Consumer Discretionary	Health Care	Materials	Utilities	Communication Services	Energy	Industrials
Mkt-RF	0.006607	0.010440	0.008400	0.011899	0.009888	0.008767	0.009856	0.006652	0.008823	0.011048	0.009885
SMB	-0.003270	-0.000095	-0.000513	-0.000974	0.001444	-0.000080	0.001988	-0.003897	0.000439	0.002900	0.001075
HML	0.000508	0.007366	0.001988	-0.003720	-0.001293	-0.002244	0.003841	0.001397	-0.000653	0.009731	0.003761

SMB - positive correlation means the portfolio is weighted towards small cap companies.
HML - positive correlation means that returns are attributable to value premium.

b) First compute mean annualized values for returns and factors, then compute the multivariate regression.

```
Mkt-RF    14.263952
SMB        0.815427
HML       -4.848931
```

```
Consumer Staples         0.062461
Financials               0.122658
Real Estate              0.054058
Information Technology    0.225325
Consumer Discretionary   0.169866
Health Care              0.111537
Materials                0.092482
Utilities                0.060711
Communication Services   0.118743
Energy                  -0.020317
Industrials              0.111313
```

	Consumer Staples	Financials	Real Estate	Information Technology	Consumer Discretionary	Health Care	Materials	Utilities	Communication Services	Energy	Industrials
Mkt-RF	0.003914	0.007686	0.003387	0.014119	0.010644	0.006989	0.005795	0.003804	0.007441	-0.001273	0.006975
SMB	0.000224	0.000439	0.000194	0.000807	0.000608	0.000400	0.000331	0.000217	0.000425	-0.000073	0.000399
HML	-0.001330	-0.002613	-0.001152	-0.004800	-0.003618	-0.002376	-0.001970	-0.001293	-0.002529	0.000433	-0.002371

As shown here above, the results are different.
In the last approach the size factor is more relevant than the first one, where just 5 out of 11 sectors show a positive coefficient; on the other hand almost every sector has a positive SMB coefficient by using the second method.

CONCLUSIONS

We can distinguish two types of results:

- The first concerning the calculation of the expected return in investing in low-capitalized companies both in block and spacing in the different sectors.

Here in general the initial expectations regarding the tertiary and primary sectors were met: with the exception of some anomalies, in the services sector we have been able to notice a better performance of the portfolios constituted by low capitalization companies.

- The second using a multivariate regression to understand the relationship between expected returns and Fama & French formula coefficients.

Here the first method shows how the coefficient SMB is in a negative majority, and the sectors in which it is positive, to indicate a balance of the portfolio in favor of small companies, is more evident in the primary sectors.
The second method instead shows a general trend oriented to portfolios of small companies, regardless of the primary sector or tertiary of belonging.

Ultimately, by adding all the results and trying to make a sort of average of these, we can assert that the small capitalization companies outperform the large capitalization ones; although we have been able to observe how in some sectors the phenomenon is more evident than others, without a more in-depth sectoral and macroeconomic analysis, the initial thesis regarding differences between primary and tertiary sectors cannot be demonstrated with absolute certainty.