Università Commerciale "Luigi Bocconi"

MAFINRISK
Master of Quantitative Finance and Risk Management

XVIII Cycle

Individual Assignment

10060 Econometrics
Professors: Manuela Pedio, Massimo Guidolin

Marco Lingua
ID number: 3193131

Academic Year: 2021 - 2022

1 Introduction

The relationship between the performance of a firm and the remuneration of the management has been largely investigated in financial literature over the last decades. The scope of this brief work is to investigate this topic by using a univariate linear regression model over a small group of 25 American firms randomly selected from the Standard & Poor 500 list. In the following sections, data are analyzed and commented, then, the linear regression model is presented, together with the obtained results and parameters; finally, potential issues are presented and explained. For the purpose of this work, every data are selected via Refintiv Workspace® and reported as millions of American dollars.

1.1 Data type and preliminary analysis

The S&P 500 is a stock market index tracking the performance of 500 large companies listed on stock exchanges in the United States and it is widely considered as the most common market benchmark. The index comprises 505 stocks issued by large-cap companies belonging to 11 different sectors and includes about 80% of the American equity market by capitalization.

Sector	**Number of Firms**
Communication Services	26
Consumer Discretionary	63
Consumer Staples	32
Energy	22
Financial	65
Health Care	65
Industrials	74
Information Technology	73
Materials	28
Real Estate	29
Utilities	28
Total	505

Table 1: S&P 500 composition

As previously said, a small sample of 25 firms has been randomly selected from the S&P population; in particular, for the aim of this work, the independent variable is set as the total revenues from goods and services of the firm, and, for dependent one, the total seniors compensation, both referred to the end of 2019 (for a complete view of the sample see 6). As can be seen in the following table, the result of random sampling, (carried out in according to one of the key assumptions of the linear model, by usign the `sample` function in R on the complete list) has given an unbalanced sample, with greater weights on the financial and IT sectors and zero observations coming from the consumer staples one.

The data collected are cross-sectional: in fact, both the explained and the explanatory variables are referred to a specific observation at a single point in time (i.e. the

Sector	Number of Firms
Communication Services	3
Consumer Discretionary	3
Consumer Staples	-
Energy	1
Financial	7
Health Care	3
Industrials	1
Information Technology	4
Materials	1
Real Estate	1
Utilities	1
Total	25

Table 2: Sample composition

end of financial year) for each individual of the sample (so for each firm selected from the list). The first step consists of analyzing data by doing a normality test, which can be performed through some graphical and numerical methods. Two key parameters, known as Skewness and Kurtosis, are usually computed in order to give mathematical evidence of how variables are distributed. The first of these two indicators explain the position of the majority of data values in the distribution around its mean, while Kurtosis tells the tail shape of a distribution. For univariate distributions can be respectively calculated via these following equations:

$$\text{Skewness[X]} = \text{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

$$\text{Kurtosis}[X] = \text{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

A normal distribution has a skewness nearly close to zero, otherwise, if the value is above zero, then the data are positively skewed towards the right and the histogram presents a right-oriented asymmetry, with the right tail longer than the left one; if skewness is below zero, the asymmetry goes towards the left, with the corresponding tail longer than the opposite. Concerning Kurtosis, It is actually the measure of outliers present in the distribution: standard normal distribution has a typical value of three: values above suggest the presents of outliers and the distributions present a fat-tail, otherwise, if below three, it has light tails and/or lack of outliers. Another graphical method for indagate the normality of a distribution is the Quantile-Quantile plot, in which theoretical quantiles from a standard normal distribution are plotted against the sample quantiles: the more these points dispose on the red dotted line (which means that they are equally distributed) the more the sample is close to a normal distribution. Deviations from this line suggest the non-normality of collected data and far away values are to be considered as outliers. Finally, the presence of outliers can be furtherly investigated through a sample boxplot, where, in order, minimum,

1st quartile, median, 3rd quartile and maximum are plotted as a "threshold": also, in this case, observations which fall outside of the minimum-maximum boundaries have to be considered as outliers. In addition to these graphical methods, a table with values of the principal statistics is provided for a better understanding of the analysis. As previously said, the dependent or explained variable is the total senior executives' compensation (*Remun*): usually, companies rely on different policies of remuneration, in which the managers receive a fixed salary in cash (executive pay), performance-based bonuses (cash, shares, or call options on the company stock) and another kind of benefits (typically non-financial) in return for their work and configured to take into account regulations and law. As can be seen from Figure 1, the data collected are far away from a normal distribution. The histogram evidences a strong positive right asymmetry, with a mean greater than the median, and a right fat-tail that indicates the presence of outliers, which is confirmed by the QQplot and Boxplot. This can be partially explained by the randomness of selected data: even if all individuals belong to the same index, there is high variability in the population itself due both to the presence of more remunerative sectors and by the so-called "tech giants" (Amazon, Apple, Alphabet, Netflix etc.), which, if selected in the sample, would probably be outliers due to their dimensions. In addition, remuneration policies tend to vary in companies involved in strong performing industries: favourable economic conditions and higher growth rates tend to generate a greater pay increase in respect of revenues, while those in hard-hit industries may see flat or declining pay. Selecting random companies, given a specific sector, could maybe give more unifrom results, but the presence of outliers seems to be strongly radicated to the nature of the population.

Mean	Std. dev.	Min.	Max.	Median	Skewness	Kurtosis
41.9457	32.7416	4.6276	142.7060	31.7187	1.7441	2.6042

Table 3: Remun summary statistics

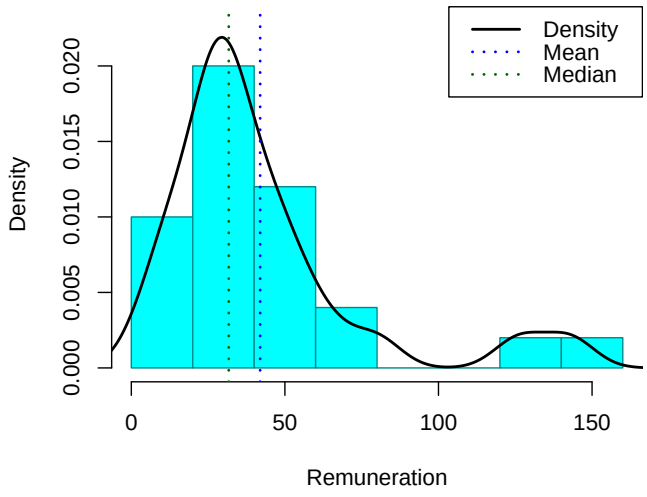


Figure 1: Remun Histogram, Density, Mean and Median

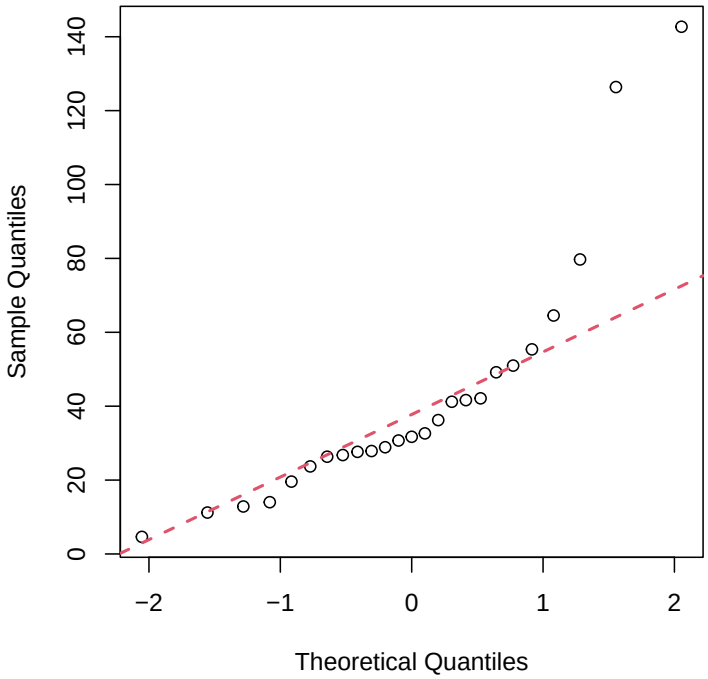


Figure 2: QQ Plot for Remun

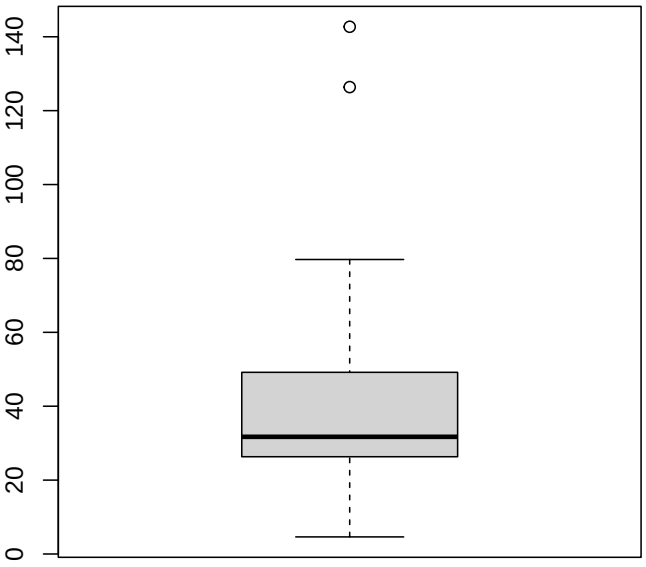


Figure 3: Boxplot for Remun

In this model, the independent explanatory variable is *Revenues*, considered as the profit generated by the sale of goods and services, and it consists of the income generated by the core business of the company Intuitively, also this distribution is characterized by strong levels of Skewness and Kurtosis, even greater than the dependent variable. The variability in the population is still very high even for this variable, despite the fact that we are considering one of the largest market capitalization index: for same reasons previously exposed, it seems quite common that the results of random sampling gives non uniform values, and some observations are way more greater (or smaller) in respect to the sample mean. This fact can be seen also graphically in the followinf figures.

Mean	Std. dev.	Min.	Max.	Median	Skewness	Kurtosis
21,772.34	26,198.28	10.92	108,942.00	12,347.00	1.8435	2.8450

Table 4: Revenues summary statistics

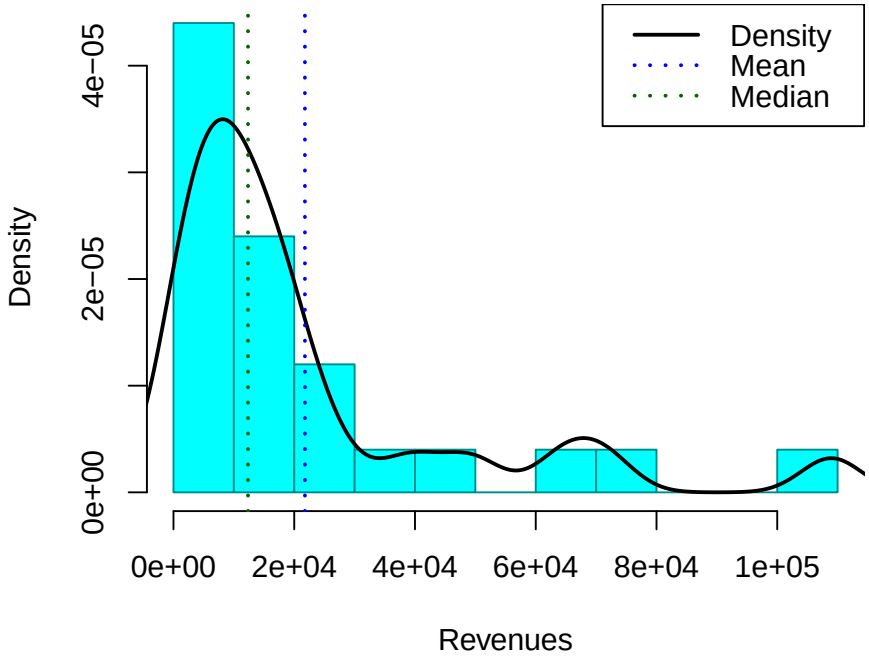


Figure 4: Revenues Histogram, Density, Mean and Median

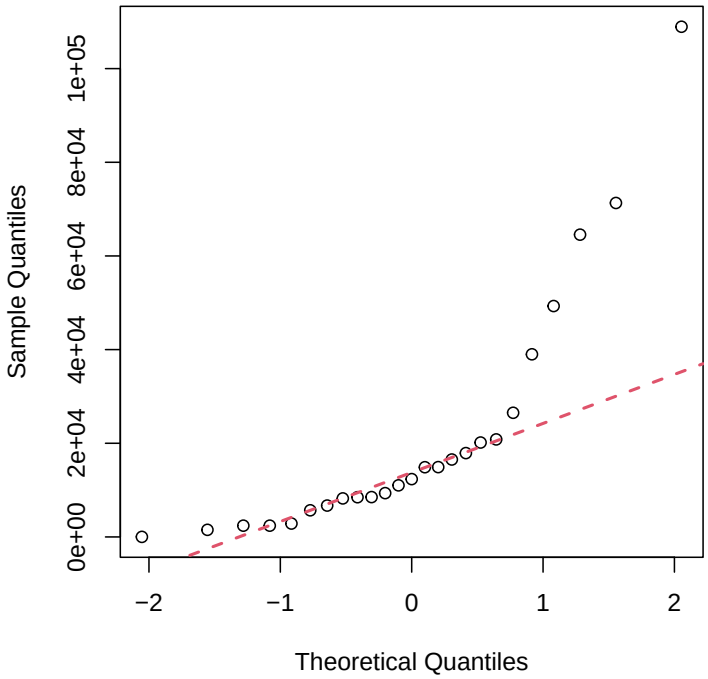


Figure 5: QQ Plot for Revenues

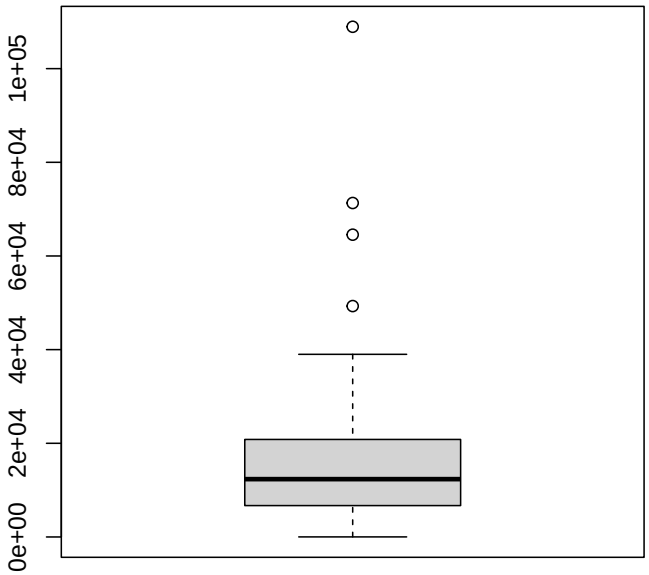


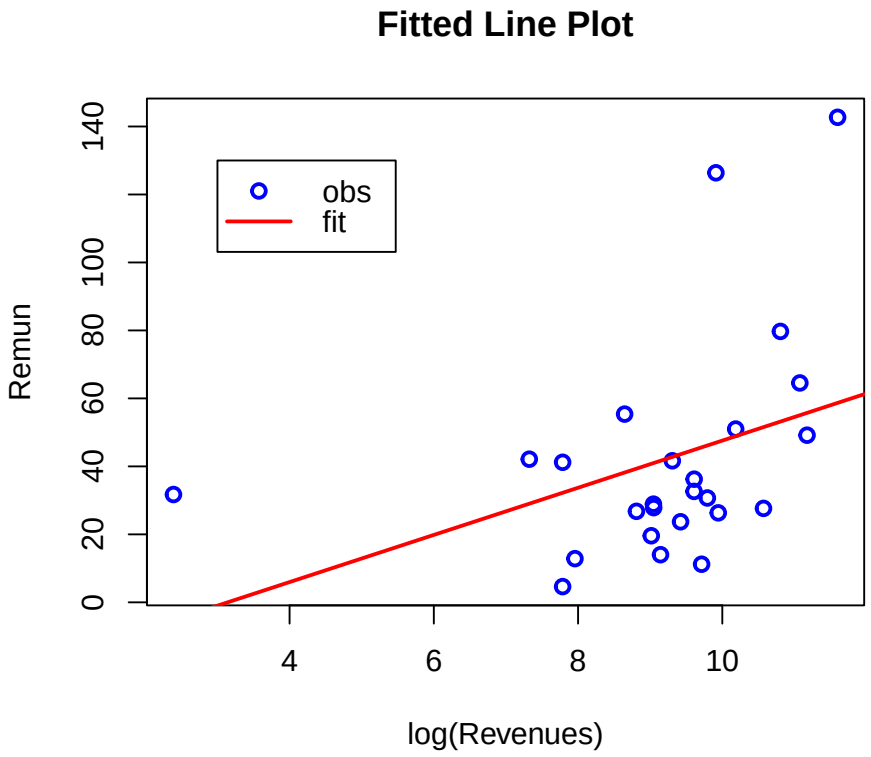
Figure 6: Boxplot for Revenues

1.2 The linear regression model

The proposed model for this work consists of a univariate linear regression, which can be presented as the following equation:

$$Remun_i = \beta_0 + \beta_1 log(Revenues_1) + u_i \tag{1}$$

where β_0 is the intercept, β_1 is the slope (also known as the sensitivity coefficient) and u_i can be considered as the error term, a constant which contains the residual effect of other unobserved phenomena.



Parameter	Est.	S.E.	t val.	p	$R^2 = 0.14338$
β_0	-21.7903	33.0683	0.6590	0.1647	
β_1	6.9379	3.5359	1.9621	0.0620	

Table 5: Regression output

Given that this is a linear-logarithmic model, the relationship between Remun and Revenues is considered non-linear, and the coefficient β_1 says that a 1% change in the independent variable generates a change in the dependent one equal to $\beta_1 0.01$. As reported in Table 5, the intercept β_0 and the regressor β_1 are both statistically non-significative, because the respective t-values are both smaller than the standard critical value (that is equal to 2.0687, for a first type error $\alpha = 0.05$ and a number of degrees of freedom equal to 23) and they fall outside of the rejection region. So, both the null

hypothesis ($\beta_0 = 0$ and $\beta_1 = 0$) can not be rejected at a confidence level of 0.95. This can also be addressed by checking p-values that are both greater than 0.05 (α, the probability of rejecting the null while it is actually true). Then, it is possible to state that, for this regression fitted on this dataset, there is not a significant relationship between the remuneration of executives managers and revenues from sales, for that confidence level. Otherwise, the slope coefficient is indicated to be significative for a first type error of 10%, even if this is quite an uncommon value for hypothesis testing in this particular field. Given that $R^2 = 0.14338$ is the output value computed by the software for this regression, It can be said that the portion of the variance of *Remun* explained by *Revenues* is quite low, suggesting that this particular model, fitted on these data, has not much explanatory power. Since there are more than 85% of the variation of remuneration still to be explained, this can be addressed to the error terms that should include all the other characteristics that are not included in this model, but that must influence the amount of salary paid to the managers.

Another reflection that can be made is about the assumption of homoskedasticity ($Var[u|x] = \sigma^2$), meaning that the conditioned distribution of the errors has constant variance), which, in addition to the other assumptions, is fundamental in determining the variance of the estimated coefficients in OLS regression. Focusing now on what is the purpose of this works -checking a relation between revenues and remuneration- it could be said that non necessarily the variability in remuneration around its mean is constant because any firm chooses different methods of paying its managers: as already said, the total remuneration consists in cash and other benefits, and it could be that for any given firm, the "liquid" (cash) part of the total compensation differ, based on the company's decisions. This seems to be not only related to the sales revenues level, but also to the internal board decision on how to pay managers' salaries. In this case, the variance is no more constant, the error terms are said to be heteroskedastic and the variance of the estimators can not be properly estimated, since their variance estimators are biased even with larger samples. For being sure about the homoskedasticity or not of the errors, it is possible to rapidly check it by launching a Breusch-Pagan test in which the null hypothesis is that the variance of residuals is constant, leading to homoskedasticity. This test resulted in a p-value equals to $0.3809 > 0.05$ so the null cannot be rejected and there is no heteroskedasticity (at 95% confidence level). If this would not be the case, taking also in addition the presence of outliers, the use of a robust regression would be a better strategy, because it aims to minimize the impact of outliers on the coefficient estimates, computing smaller standard errors and leading to a more precise model.

In order to check univariate alternative models to this research, linear-linear and logarithm-logarithm regressions have been tested on the same data sample. Results, which can be found in Appendix, suggest that in this specific case a lin-lin model would perform better in terms of explanatory power and statistical significativity of the coefficients (see Table 7) whether the log-log is affected by a low r-squared and lacks of significativity. However, the use of a different sample in terms of dimensions and firms would probably show different outcomes either in the goodness of the fit, either in statistical relevance for the different regression type which has been tested. However, if we consider the lin-lin model, the intuition that there must be some form of non-linearity between the variables is yet non considered, because in this case one would check for a simple linear relationship.

Despite the fact that this sample is too small and, for a more serious analysis there

would be required more observations because It is not representative of the population at all, probably, even with a larger sample, the explanatory power would increase but not that much: this because in the reality, there are many other aspects that affect the remuneration of the executives' manager. So, for a better investigation, different variables should be taken into account in multivariate linear regression. The relationship between remuneration and the overall performance of the firm is not always strictly positive and in reality, It does not only rely on the firm's revenues and the theoretical foundations of this lies in the so-called *agency theory*. Briefly, executives managers (which can be considered as the *agents*) would not purely pursue the ambition of the firm (i.e. the shareholders) -the *principal*- if their interests are not completely aligned with the one of their company. As stated by the theory, there is a conflict of interests between the managers, which aim to maximize the growth rate of the firm, and the shareholders, which wants to maximize profits and dividends. This can partially explain why the investigated relation is generally weak, and other factors must be also considered.

The main determining factor for executives' remuneration is often considered the size of the firm, based on the company's turnover. Some researchers agree that basing managers bonuses on sales growth is not always the most optimal remuneration policy. In fact, considering an executive which is close to retirement, clearly expects a bonus based on sales or sales growth, and for this precise reason, He may approve excessive spending on an advertisement or other salesforce increases to ensure larger sales revenue, with the concrete risk of reducing profits. In considering the principal-agent bond, it could be appropriate to define and measure performance in terms of shareholder return instead of accounting revenues. In a more appropriate multivariate model, other variables could be considered in order to capture more uncertainty, like:

- ROA, ROE, and others ratios
- Share price, EPS ratio
- Debt/Equity ratio or market capitalization

When utilizing indicators of profitability like ROA and ROE it must be paid attention to the possible multicollinearity of the twos because they are often strictly correlated; as a proxy for market size, it can be considered the D/E ratio or the market cap. Another reflection can be made over the data type utilized in this work, in fact, conducting a time series analysis of executives remuneration and firm revenues would be more suitable for this purpose, because the action of directors (and the consequent remuneration) has usually an effect not in the immediate, but in the medium-long term. In this case, the use of panel data would be more suitable, because it allows investigating a multiperiod behaviour on a sample of individuals.

2 Conclusions

In order to investigate the existence of a relationship between company revenues and executives managers remuneration, a simple univariate linear regression has been fitted to a small sample of listed 25 firms randomly selected from the S&P 500 index. The results show a lack of significativity for the regressor and it can be stated that financial performance can not be considered as the main driver for direct compensation. Finally,

homoskedasticity of error terms has been investigated to check the possibility of using a robust regression model. The firm usually implements remuneration policies based on salaries and other benefits (such as share options) in order to induce managers not to pursue their own interests. Beyond the statistical weakness of the results, It is evident that this model suffers from omitted variables and more factors should be considered for a multivariate regression, such as firm's size indicators and other profitability measures.

3 Appendix

3.1 Sample data

Symbol	Name	Sector	Revenues	Remun
WAT	Waters Corporation	Health Care	2406.60	4.63
EIX	Edison International	Utilities	12347.00	23.70
LOW	Lowe's	Consumer Discretionary	71309.00	49.18
PBCT	People's United Financial	Financials	16529.20	11.21
PSA	Public Storage	Real Estate	2855.11	12.85
CCL	Carnival Corporation	Consumer Discretionary	20825.00	26.33
TROW	T. Rowe Price	Financials	5685.80	55.38
URI	United Rentals	Industrials	9351.00	14.01
L	Loews Corporation	Financials	14909.00	36.23
NOV	Nov	Energy	8479.00	28.89
SBUX	Starbucks	Consumer Discretionary	26508.60	50.99
NFLX	Netflix	Communication Services	20156.45	126.38
NWS	News Corp	Communication Services	8505.00	27.86
ANSS	Ansys	Information Technology	1515.89	42.12
SYK	Stryker Corporation	Health Care	14884.00	32.64
CSCO	Cisco Systems	Information Technology	49301.00	79.71
SPGI	S&P Global	Financials	6699.00	26.78
SHW	Sherwin-Williams	Materials	17900.80	30.71
ALGN	Align Technology	Health Care	2406.80	41.20
AON	Aon	Financials	11013.00	41.65
PGR	Progressive Corporation	Financials	38997.70	27.64
NVDA	Nvidia	Information Technology	10.92	31.72
BLK	BlackRock	Financials	64545.31	64.55
CMCSA	Comcast	Communication Services	108942.00	142.71
APH	Amphenol Corp	Information Technology	8225.40	19.59

Table 6: Sample composition and respectives variables

3.2 On residuals

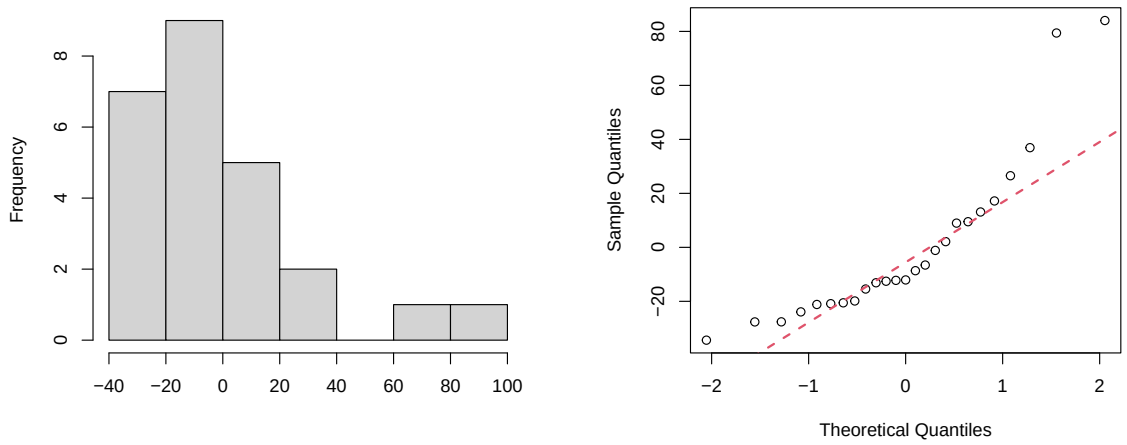
The first assumption which has to be made in fitting a linear model to a set of data is the zero conditional mean of the errors, that means, assumed independence between residuals and the x:

$$E[x|u] = E[x] = 0$$

Since these error terms are unknowns, they can be estimated by doing the difference between the observed value and the predicted value (i.e. the one obtained by the regression):

$$\hat{u}_i = y_i - \hat{y}_i$$

When imposing the first-order condition in estimating the regression parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ the total sum of residuals should be set equal to zero. In this model, checking for this condition gives a sum of residuals of 4.440892e-14: while this sum is very small but actually not zero, this may be addressed to the floating-point numbers limited precision that is rounded to the nearest one representable (so this difference can just be imputed to a computational issue). Even if the result is not mathematically zero, it can be considered as it was. One other step that can be made is to check the normality of residuals, by using the same graphical instruments already presented for *Remun* and *Revenues*.



(a) (b)

Figure 7: Histogram and QQ-plot of residuals

As can be seen, residuals are also non normally distributed, and this can be stated also because the dependent and independent variables are not also normal. As a confirmation method to this, a Shapiro test has been launched on the residuals (it consists of a statistical test where the null hypothesis that residuals are from a normal distribution is verified): this test has returned a p-value of 0.0005108, which allows rejecting the null and to confirm the previous intuition. When errors are not normally distributed, also estimations are not, and seeking for significance with the t-test and p-value are no longer consistent, because hypothesis testing, which is based on the normal-related distribution (such as the Student), can't be applied. In this case, since the parameter estimates are not significant, the non-normal errors do not further invalidate the obtained results. Otherwise, the problem of non normal residuals can be worked out by trasforming in some way the variables, in order to achieve distributions which are more normally distribuited.

3.3 Alternative regressions

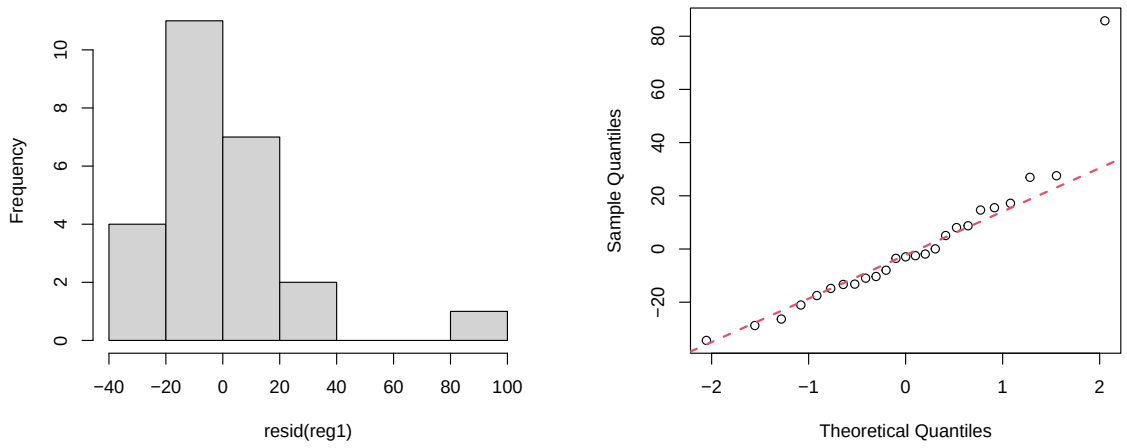
As previously said, lin-lin and log-log models have been fitted on this sample. As a reminder, these alternatives provide a different interpretation of the coefficients. For the first, if x changes of one unit, y changes of β_1 units and for the second one, a one per cent change in x produces a percentual change in y equals to β_1 (which can be interpreted as the elasticity of y in respect to x). The respective summary statistics are reported in the tables below.

Parameter	Est.	S.E.	t val.	p	$R^2 = 0.4516$
β_0	2.366e+01	6.495e+00	3.643	0.001360	
β_1	8.399e-04	1.930e-04	4.352	0.000234	

Table 7: Linear-linear regression output

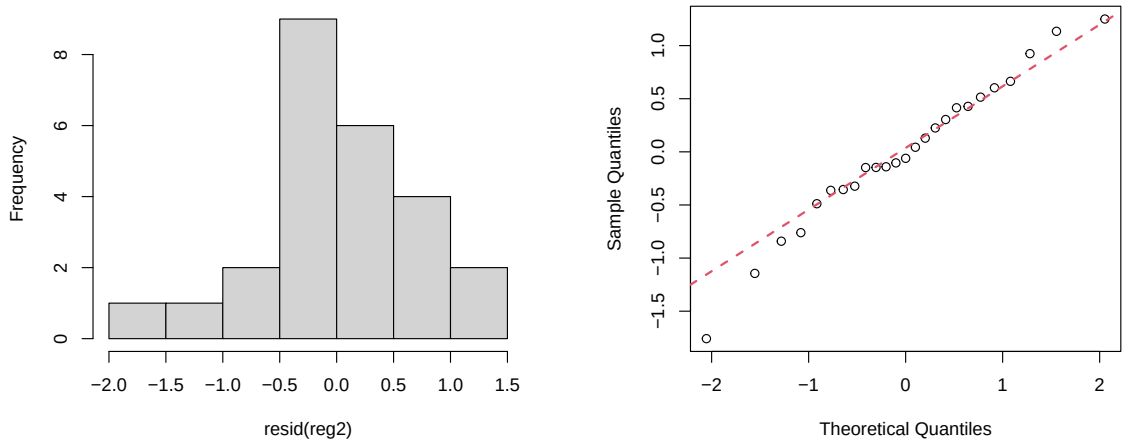
Parameter	Est.	S.E.	t val.	p	$R^2 = 0.1148$
β_0	2.19681	0.76052	2.889	0.00829	
β_1	0.14048	0.08132	1.727	0.09749	

Table 8: Logarithm-logarithm regression output



(a) (b)

Figure 8: Histogram and QQ-plot of residuals for the lin-lin regression



(a) (b)

Figure 9: Histogram and QQ-plot of residuals for the log-log regression

For both the alternative regressions have been conducted a Shapiro test and a BP test in order to check normality e homoskedasticity of residuals: in the log-log model errors are not normally dystribuited and have constant variance, while in the lin-lin are normal and homoskedastic. This can suggest that the first alternative regression would perform better in this analysis, and further investigation are needed in order to conclude that.