

FROM CLASSICAL REGRESSION ANALYSIS
TO QUALITATIVE ANALYSIS:
A SHARE PRICE fsQCA EMPIRICAL APPLICATION

Fábio André Pereira Luís

Dissertation submitted as partial requirement for the conferral of
Master in Finance

Supervisor:

Prof. José Dias Curto, Associate Professor, ISCTE-IUL Business School,
Department of Quantitative Methods for Management and Economics

October 2019

FROM CLASSICAL REGRESSION ANALYSIS
TO QUALITATIVE ANALYSIS:
A SHARE PRICE fsQCA EMPIRICAL APPLICATION

Fábio André Pereira Luís

Dissertation submitted as partial requirement for the conferral of
Master in Finance

Supervisor:

Prof. José Dias Curto, Associate Professor, ISCTE-IUL Business School,
Department of Quantitative Methods for Management and Economics

October 2019

Abstract

This thesis studies the difference between classical regression analysis and qualitative comparative analysis. Several authors argue that any preference for one approach over the other one should not be taken since both should complement themselves and therefore both should be used. This research is composed by a sample of 265 enterprises listed in European stock markets, using financial information of 2016, through the application of a classical regression analysis and a qualitative comparative analysis. More than testing the impact of the *size of the company*, the *leverage level*, the *book value per share*, the *earnings per share*, the *return on asset*, the *cashflow from operations on asset* and the *ownership by a billionaire* on the *share price*, this research aims at comparing classical regression analysis and comparative qualitative analysis through the results obtained from the empirical assessment. The main conclusion shows that qualitative comparative analysis helps to expand the comprehension regarding the conditions needed to achieve the outcome. In fact, the study contributes for the corroboration that regression analysis can be complemented by qualitative comparative analysis. The main limitations of this study are related to the use of a one-year data, which is also relatively outdated, since refers to 2016.

Keywords

Regression Analysis, Qualitative Comparative Analysis, Fuzzy-set, Share Price

JEL Classification System

C02, C31

Resumo

Esta tese estuda a diferença entre a análise de regressão clássica e a análise comparativa qualitativa. Vários autores argumentam que qualquer preferência sobre uma delas não deve ser tida em consideração, uma vez que ambas se devem complementar e têm de ser utilizadas. Para esse propósito, foi utilizada uma amostra constituída por 265 empresas listadas em bolsas de mercado europeias, utilizando informação financeira de 2016, que será utilizada quer na análise de regressão, quer na análise comparativa qualitativa. Mais do que testar o impacto da *dimensão da empresa*, do *nível de endividamento*, do *valor contabilístico das ações*, dos *ganhos por ação*, do *retorno dos ativos*, dos *fluxos de caixa das operações sobre os ativos* e da *estrutura patrimonial no preço das ações*, este estudo pretende comparar as diferentes metodologias utilizadas através dos respetivos resultados. As principais conclusões do estudo revelam que a análise qualitativa comparativa ajuda a compreender as condições necessárias para alcançar o resultado desejado. De facto, esta investigação corrobora estudos anteriores que concluem que a análise de regressão pode ser complementada com a análise comparativa qualitativa. As principais limitações deste trabalho estão relacionadas com o uso de uma base de dados referente a um só ano que, adicionalmente, também está relativamente desatualizada visto que se refere a 2016.

Palavras chave

Análise de Regressão, Análise Comparativa Qualitativa, Conjuntos Incertos, Preço das Ações

Sistema de Classificação JEL

C02, C31

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. José Dias Curto, for the support given to elaborate and write this dissertation, especially because the respective scope includes a type of analysis that is not widely known. In addition, I also thank Prof. Isabel Lourenço, who supported all efforts regarding the data base.

Secondly, I would like to thank my parents for the support given during my second round of my higher education.

Last but not least, I would like to thank my wife for being by my side during these years, prompting me every day that the dissertation was to be done and never letting me give up.

Index

1. Introduction	1
2. Literature review	3
2.1. Classical regression analysis	3
2.1.1. General remarks.....	3
2.1.2. The goodness-of-fit of a model	5
2.1.3. Estimation.....	5
2.1.4. Inference	7
2.2. Limitations of classical regression analysis.....	8
2.3. Complexity Theory	11
2.4. Qualitative comparative analysis.....	13
2.4.1. fsQCA, a kind of QCA	14
2.4.2. Methodology of QCA.....	16
2.4.3. An empirical example.....	17
2.5. Share price conditions: An application.....	19
3. Methodology	22
3.1. Sample characterization.....	22
3.2. Variables description	22
3.2.1. Potential dependent variables / outcome conditions	23
3.2.2. Potential independent variables / antecedent conditions	23
3.3. Hypotheses.....	23
3.4. Empirical application.....	24
3.4.1. Qualitative comparative analysis.....	24
3.4.2. Regression analysis	25
4. Data analysis and empirical results	28
4.1. Qualitative comparative analysis.....	28
4.2. Regression analysis.....	29
4.2.1. Testing the model	30
4.2.2. Testing the assumptions of the model	30
4.2.3. Results interpretation.....	33
5. Conclusions	34
Bibliography.....	38
Annexes	40

Index of annexes

Annex 1: Sample characterization.....	40
Annex 2: Variable definition.....	41
Annex 3: Description of variables.....	42
Annex 4: [RA] Chow test.....	43
Annex 5: [RA] RESET test.....	44
Annex 6: [RA] Matrix of correlations.....	45
Annex 7: [RA] Variance inflation factors.....	46
Annex 8: [RA] Jarque-Bera test.....	47
Annex 9: [RA] Breusch-Godfrey Serial Correlation LM test.....	48
Annex 10: [RA] Breusch-Pagan-Godfrey test.....	49
Annex 11: [RA] White test.....	50
Annex 12: [RA] White test with cross terms.....	51
Annex 13: [RA] Generalised Least Squares method.....	52
Annex 14: [QCA] Definition of the property space.....	53
Annex 15: [QCA] Calibration.....	55
Annex 16: [QCA] Descriptive statistics of antecedent conditions and outcome.....	56
Annex 17: [QCA] Truth table solution.....	57
Annex 18: [RA] Output of the initial model.....	58
Annex 19: [RA] Output of Chow test to the initial model.....	59
Annex 20: [RA] Output of RESET test to the initial model.....	60
Annex 21: [RA] Residuals graph of the initial model.....	61
Annex 22: [RA] Output of the initial model after the exclusion of the residual outlier.....	62
Annex 23: [RA] Output of RESET test to the initial model after the exclusion of the residual outlier.....	63
Annex 24: [RA] Matrix of correlations output.....	64
Annex 25: [RA] Computation of the variance inflation factors.....	65
Annex 26: [RA] Output of Jarque-Bera test to the initial model after the exclusion of the residual outlier.....	66
Annex 27: [RA] Output of Breusch-Godfrey Serial Correlation LM test to the initial model after the exclusion of the residual outlier.....	67
Annex 28: [RA] Output of Breusch-Pagan-Godfrey test to the initial model after the exclusion of the residual outlier.....	68

Annex 29: [RA] Output of White test to the initial model after the exclusion of the residual outlier	69
Annex 30: [RA] Output of White test with cross terms to the initial model after the exclusion of the residual outlier	70
Annex 31: [RA] Output of the GLS estimation	71
Annex 32: [RA] Output of the final model	72
Annex 33: [RA] Output of Breusch-Pagan-Godfrey test to the final model.....	73
Annex 34: [RA] Output of White test to the final model.....	74
Annex 35: [RA] Output of White test with cross terms to the final model.....	75
Annex 36: [RA] Output of Breusch-Godfrey Serial Correlation LM test to the final model ..	76

Index of figures

Figure 1: A symmetrical relationship between x and y	10
Figure 2: An asymmetrical relationship between x and y (<i>sufficient but not necessary</i> condition) ...	10
Figure 3: Comparison between RA and QCA nomenclatures	13
Figure 4: Kind of fuzzy-sets	14

List of abbreviations

BIL: Billionaire

BLUE: Best linear unbiased estimators

BVPS: Book value per share

CFOA: Cash flow from operations on asset

CLM: Central limit theorem

CT: Complexity Theory

EPS: Earning per share

fsQCA: Fuzzy-set qualitative comparative analysis

GLS: Generalised least squares

LE: Leverage level

OLS: Ordinary least squares

QCA: Qualitative comparative analysis

PPS: Price per share

\bar{R}^2 : Adjusted R^2

RA: Classical regression analysis

ROA: Return on asset

SZ: Size of a company measured by the natural logarithm of its total assets

VIF: Variance inflation factors

WLS: Weighted least square

1. Introduction

Fuzzy-set qualitative comparative analysis approach has led some researchers to change their methodology, from a classical regression to a qualitative analysis.

Several authors have empirically found that the dominant logic regression analysis is not enough to respond to the complexity of reality due to its simplicity and disregarding the effect of the relationship among independent variables on dependent one. Additionally, some weaknesses of regression analysis, such as the effect size, the symmetrical effect and the linear relationship, are enough to justify the change from regression analysis to qualitative analysis. In fact, qualitative analysis allows researchers to describe multiple realities and consider complex antecedent conditions into their analysis, since it is more important how independent variables are related with each other than the importance of each individual one. The Complexity Theory takes into account all of these considerations and so it is a theoretical explanation that supports the change from classical regression analysis to qualitative analysis. This study aims at comparing the results obtained from a regression and a qualitative analysis, using a one-year date, with reference date of 2016, that includes accounting information about 265 companies listed in European stock markets. The empirical analysis is focused on the examination of the factors that influence the share price, such as the *size of the company*, the *leverage level*, the *book value per share*, the *earnings per share*, the *return on asset*, the *cashflow from operations on asset* and the *ownership by a billionaire*. Note that for the purpose of this research, it is more important the comparison of both analysis than the assessment of the empirical results.

This research makes two main contributions to the literature: first, a careful review of the literature that focus on relevant theories and papers about the topic; second, provide insights to develop more research in this area in furtherance of scientific quality improvement.

This thesis reveals that the use of a qualitative analysis, in particular the fuzzy-set qualitative comparative analysis, is not enough to explain the share price conditions since regression analysis also provides relevant information about the factors that influence the share price. While a qualitative analysis treats the sample in a qualitative way and the respective conclusions highlight a potential qualitative relationship between antecedent conditions and outcome, a regression analysis specifies and measures the impact of each independent variable on the dependent one. Therefore, qualitative analysis helps to expand the comprehension regarding the conditions needed to achieve the outcome. In fact, this conclusion is aligned with the literature review.

Regarding the limitations of this study, it is important to take into account that the sample is composed by a one-year data. Also, this data is relatively outdated since refers to 2016. In this sense, a sample with a long period of time is required to produce more accurate results through the capture of the volatility of the share prices in stock markets.

In what respects to the structure of this document, following this introduction, the next section presents the review of literature. The third section describes the methodology, which comprises the data description, the hypothesis that this research pertains to examine and the description of the empirical application. After that, the data analysis and empirical results are reflected in the fourth section. Lastly, the conclusions are presented in the respective section, followed by the bibliography and the annexes.

2. Literature review

This section is composed by a brief review of the classical regression analysis and a description of its limitations. After that, the Complexity Theory is presented as a useful way to go beyond the classical regression analysis. Since a qualitative analysis is a methodological tool that respects the Complexity Theory, its description is reflected in own division, which includes a presentation of the fuzzy-set qualitative comparative analysis as a kind of qualitative analysis, the respective methodology and an empirical example.

Lastly, the literature review comprises a brief analysis of the factors that influence the share price because this thesis also aims at investigating which factors have major impact on share price.

2.1. Classical regression analysis

2.1.1. General remarks

The **classical regression analysis** (hereinafter, RA), as an inferential methodology, has been applied in several contexts to establish a relation between cause-effect, under an empirical analysis. This relationship can be achieved through the definition of a regression model, which can be a simple or a multiple one. While the former model allows to assess the relation between two variables (the **dependent** or **explained variable** – y – and the **independent** or **explanatory variable** – x), the latter relates many factors (k dependent variables – x_1, x_2, \dots, x_k) that can influence the dependent variable, which is desirable to predict (Wooldridge, 2014).

Some variables can be treated in a binary way, known by **dummy variables**, in particular the qualitative ones since their information are only restricted to a “presence” or “absence” of a given factor. For instance, the gender of a female worker is a kind of dummy variable. If its value is equal to 1, it means that is a female, otherwise the worker is a man (Wooldridge, 2014). However, some careful is needed because 0 does not always mean the opposite of 1. For example, if a dummy variable is about the married status, 0 can mean single, divorced, widower or non-marital partnership.

Equation (1), which is assumed to hold in the population that researchers intents to study, represents a **simple regression model** and aims at explaining the relationship between education and wage (Wooldridge, 2014).

$$wage = \beta_0 + \beta_1 educ + u \quad (1)$$

While *wage* is the dependent variable and is measured in Euros *per* hour, *educ* is the independent variable and is measured in years of education. Thus, this model pretends to

explain the effect of one more year of education on person's wage. However, others unobserved factors that can influence the wage are included in the term u (called the **error term** or **disturbance**), such as labour force experience, innate ability and work ethic, among others. In its turn, β_1 is a parameter of the model that describes the relationship between the dependent variable (wage) and the factor that is used to determine it (education). In this case, β_1 measures the alteration in hourly wage given another year of education, *ceteris paribus* (i.e., holding all other factors in u fixed). At last, β_0 is another parameter, called the **intercept parameter** or **constant term**, that gives the expected value of wage when the person does not have any year of education (Wooldridge, 2014).

Equation (1) is considered a **simple regression model** because it only relates two variables: *wage* and *educ*. However, if the researcher aims at controlling k factors, such as workforce experience (*exper*) and week spent in job training (*training*), that simultaneously have impact on *wage*, a **multiple regression model**, represented by Equation (2), can be helpful (Wooldridge, 2014).

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u \quad (2)$$

Multiple regression models are more realistic and predicts better the dependent variable since more factors are used (Wooldridge, 2014). Apart from the explanation of wage by year of education, through the interpretation of the parameter β_1 , Equation (2) also explains how the wage is influenced by the years of workforce experience and by the weeks spent in job training, through the parameters β_2 and β_3 , respectively (Wooldridge, 2014).

The models above mentioned are called **linear regression models** because they are linear in the parameters, meaning that the relationship between y and x is linear. However, this type of relationship is not sufficient to explain the dependent variable and then it is not enough for economic or finance applications. Usually, some dependent variables are better explained through non-linear relationships. Instead of a constant change in wage given one additional year of education due to the linear nature of the model, as represented by Equation (1), a log-level model is more reasonable to explain on how wage changes with one more year of education, as represented by Equation (3) (Wooldridge, 2014). This is a type of **non-linear regression model** with $\log(y)$ as dependent variable and x as independent variable, as follows.

$$\log(wage) = \beta_0 + \beta_1 educ + u \quad (3)$$

This kind of model does not explain the variation of wage with education by a constant absolute value but by a constant percentage.

2.1.2. The goodness-of-fit of a model

The goodness-of-fit of a model (i.e. how well the regression predict the real data) is given by the **coefficient of determination** (R^2 ; R -squared), which corresponds to the fraction of the variance in the dependent variable that is predictable from the independent variable(s). This statistical tool typically ranges from 0 to 1, which the latter indicates that the model perfectly fits the data. However, R^2 is sensible to the number of independent variables since it increases as the number of independent variables increase. Otherwise, the **adjusted R^2** (\bar{R}^2) is used because it includes a penalty for adding other independent variables to the regression. This statistical measure increases if, and only if, the independent variable recently added improves the model and decreases when a predictor improves the model less than what is predicted by chance.

Also, the significance of independent variables can be analysed by looking for the **information criteria** since this statistic takes also into account the complexity of the model: for smaller values of the information criteria, the model is more reliable (Wooldridge, 2014).

2.1.3. Estimation

In order to estimate the parameters in a linear regression model (i.e., $\beta_0, \beta_1, \dots, \beta_k$), the researchers are used to the method of **Ordinary Least Squares** (OLS). For that purpose, the researchers select a random sample from a population and, through the OLS method, use the sample to estimate the parameters of that population (Wooldridge, 2014).

Considering the population model represented by the Equation (2), the correspondent estimated OLS equation (i.e., the sample model) is

$$wage_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 exper_i + \hat{\beta}_3 training_i + \hat{u}_i \quad (4)$$

where $\{(wage_i, educ_i, exper_i, training_i): i = 1, \dots, n\}$ denote a random sample of size n . Moreover, $\hat{\beta}_k$ represents the estimators that aim at determining the parameters of population and \hat{u}_i denotes the residual that includes all factors affecting $wage_i$ apart from $educ_i$, $exper_i$ and $training_i$ (Wooldridge, 2014).

However, the estimation only provides trust results as long as the conditions of OLS method are verified, otherwise the results obtained cannot be reliable. These conditions are known as **Gauss-Markov assumptions** and are described below:

Assumption LR.1

Linear in parameters

The parameters $\beta_0, \beta_1, \dots, \beta_k$ are unknown and the error u is an unobserved factor.

Assumption LR.2

Random sampling

A random sample is composed by n observations.

Assumption LR.3

No perfect collinearity

In the sample and as consequence in the population, none of independent variables is constant. Moreover, there are not full linear relationship over the independent variables.

Assumption LR.4

Zero conditional mean

The expected value of the error u is zero given any value of the independent variables.

Assumption LR.5

Homoskedasticity

The variance of the error u is constant given any value of the independent variables. Otherwise, the residuals are heteroskedastic.

Under assumptions LR.1 through LR.4 the estimators are **unbiased**, meaning that the expected value of an estimator is equal to the population value. If all Gauss-Markov assumptions are considered (i.e., assumptions LR.1 through LR.5), then the estimators are the **Best Linear Unbiased Estimators** (BLUE), meaning that the estimator is unbiased and is the one with smallest variance, when compared with all linear and unbiased estimators (i.e., when the expected value of an estimator has the lowest spread from the population value) (Wooldridge, 2014).

In addition, one more assumption is considered for cross-sectional regression applications:

Assumption LR.6

Normality

The error u is independent of the independent variables and is normally distributed with zero mean ($E(u) = 0$) and constant variance ($Var(u) = \sigma^2$): $u \sim Normal(0; \sigma^2)$.

A model that complies with all above-mentioned assumptions is called classical linear model since it is under the **Classical Linear Model assumptions** (LR.1 through LR.6). The respective estimators are strongly efficient when compared with those under the Gauss-Markov assumptions, which means that the estimators have the smallest variance over unbiased estimators.

2.1.4. Inference

In order to determine if the conclusions from the sample can be generalized to the population, researchers should calculate inferential statistics. For that purpose, a testing hypothesis about the parameters in the population regression model should be performed, such as F -test and t -test, described below.

The **F -test** allows researchers to test every hypotheses of the regression function. This means, it tests the global insignificant of the parameters and consequently the regressors relevance.

Considering the multiple regression model that explains the hourly wage represented by the Equation (2), the hypotheses are

$$\begin{cases} H_0: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_1: H_0 \text{ is not true} \end{cases} \quad (5)$$

where the null hypothesis (H_0) refers to the globally insignificance of the regression model and therefore the years of education, the years of workforce experience and the weeks spent in job training (i.e., the independent variables) have no effect on hourly wage (i.e., the dependent variable).

Under the Classical Linear Model Assumptions, the statistic for the F -test can be written as

$$F = \frac{\frac{R^2}{k}}{\frac{1 - R^2}{n - k - 1}} \sim F_{k, n-k-1} \quad (6)$$

where R^2 is the y variation's percentage that is explained by the model, k is the number of independent variables (in this particular case, 3) and n the number of observations. If the null hypothesis is rejected ($p - value < \alpha^1$), the model is globally significant but none conclusion about the relevance of the regressors can be done. For more information, it is needed to do the t -test.

The **t -test** allows researchers to test the hypotheses relative to one parameter of the regression function. Considering the regression model expressed by Equation (2), if researchers are interested to know whether one year of workforce experience or one week spent in job training have the same impact on person's wage (H_0), the hypotheses are stated as follow.

$$\begin{cases} H_0: \beta_2 = \beta_3 \\ H_1: H_0 \text{ is not true} \end{cases} \quad (7)$$

¹ Usually, the critical value used is 5% considering the 95% confidence level.

Under the Classical Linear Model Assumptions, the statistic for the t -test can be written as

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_3}{se(\hat{\beta}_2 - \hat{\beta}_3)} \sim t_{n-k-1} \quad (8)$$

where t is a t -student distribution, β_2 and β_3 are the estimated parameter to assess, $se(\bullet)$ the standard deviation, k is the number of independent variables (in this particular case, 3) and n the number of observations. If the null hypothesis is rejected ($p - value < \alpha$), the impact of an additional year of workforce experience is not equal to the impact of one more week spent in job training on hourly wages.

Moreover, researches can also assess the significance of the independent variable, such as *exper*, expressing the null hypothesis as $H_0: \beta_2 = 0$ (i.e., the years of workforce experience has not impact on hourly wage). In this particular case, if the null hypothesis is rejected, *exper* is relevant and the parameter is statistically different from zero.

2.2. Limitations of classical regression analysis

Some weaknesses of RA have been identified due to its inaccurate application, namely in social sciences (Armstrong, 2012). However, RA has helped several scientists in their researches, estimating relevant models (Woodside, 2014). Nevertheless, some authors stated that caution is needed in the use of RA because how much complex a regression is, more septic the researcher should be (Friedman and Schwartz, 1991). Also, Soyer and Hogarth (2012) verified that some RA outcomes, as t -statistics, F -statistics, $p - values$ and coefficient of determination (R^2) lead scientists to make inadequate decisions. In particular, a high value of R^2 does not necessarily mean that the model is good since, in many cases, does not make good forecasts (Wu *et al.*, 2014). On contrary, a low value of R^2 can lead researchers to make wrong conclusions and to disregard the model when, in fact, the model can be adequate (Woodside, 2013). Even so, a considerable number of researchers have undervalued these weaknesses, claiming that a large sample is enough to mitigate the issues related to the standard statistics (Armstrong, 2012).

In what respects to the fragilities of RA, Woodside (2013) gave three reasons to be careful with RA, in special with the multiple regression analysis, such as:

(i) **Effect size**

The effect size is defined as the individual effect of independent variables on the dependent one, through the significance (or insignificance) statistic of net effects². However, it is possible that a given independent variable does not have individual influence on a dependent one but it can have together with others (Fotiadis, 2018; Mattke, Muller and Maier, 2019). In fact, Wu *et al.* (2014) revealed that effect size cannot strongly explain variations of the dependent variable. Because of this, the effect size makes RA unreliable since the opposite cases can occur, indeed (i.e., independent variable cannot have any net influence on the dependent variable, although the combination among independent variables can have effect on the dependent variable) (Ordanini, Parasuraman and Rubera, 2014).

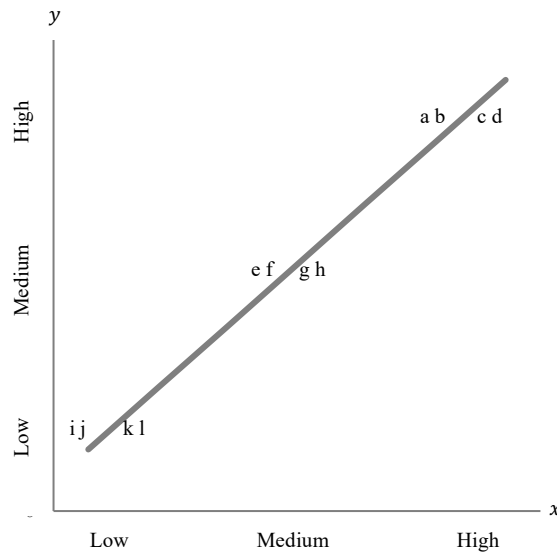
(ii) **Symmetrical effect**

The symmetrical effect³ occurs when high values of y are only achieved with high values of x , which represents a *necessary* and *sufficient* condition (Shering, Korhonen-Kurki and Brockhaus, 2013; Woodside, 2013; González-Velasco, González-Fernández and Fanjul-Suárez, 2017). While a *necessary* condition requires always the presence of a given factor for the occurrence of the outcome (for instance, factors that have to be presented if a media leads to a positive brand attitude), the *sufficient* condition means that whenever a given factor occur, the outcome will also occur (for instance, a specific set of attributes that together lead to a positive brand attitude), although the outcome can be achieve as a result of another factor (Shering, Korhonen-Kurki and Brockhaus, 2013; Mattke, Muller and Maier, 2019; Mello, 2019). The symmetrical effect can be represented as the Figure 1.

² The net effect corresponds to the impact of each potential independent variable on the dependent variable after the segregation of the influence of others independent variables on the dependent one (Woodside, 2013).

³ Symmetric relationship has a correlation above 0.8 (Woodside, 2013).

Figure 1: A symmetrical relationship between x and y

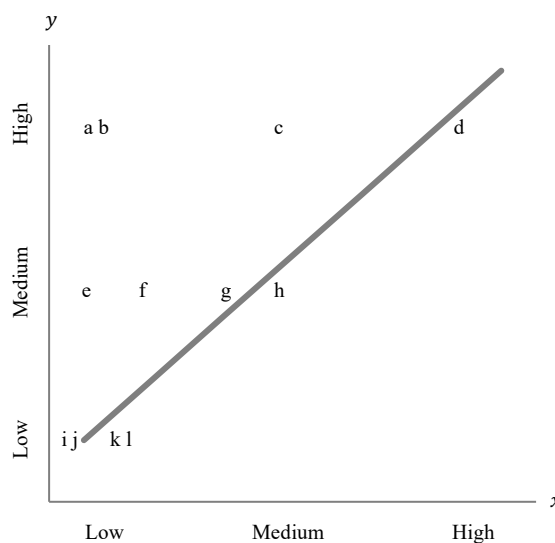


Source: adapted from Wu *et al.* (2014)

However, the empirical evidences show that symmetrical effect does not fully fit the reality. So, the symmetrical effect is considered as a weakness of RA since this kind of analysis assumes either a symmetric relationship between the dependent and independent variable or a net effect of the independent variables on the dependent one. In fact, the reality shows that the asymmetrical effect⁴ is more common than the symmetrical one (Woodside, 2014).

One type of asymmetrical effect is a *sufficient* but *not necessary* relationship, meaning that high values of x are *sufficient* to achieve high values of y but is *not necessary* since high

Figure 2: An asymmetrical relationship between x and y (*sufficient* but *not necessary* condition)



Source: adapted from Wu *et al.* (2014)

⁴ Asymmetric relationship has a correlation between 0.3 and 0.7 (Woodside, 2013).

values of y can be obtained with low values of x or a given set of x (Woodside, 2013; Wu *et al.*, 2014; Mello, 2019). The Figure 2 represents the above-mentioned condition.

Another type of asymmetric relationship can occur when high values of x results not only in high values of y but also in low values of y (*insufficient* but *necessary* condition) (Wu *et al.*, 2014).

In addition, while symmetric tests take into account the cause effect of high (low) values of x on high (low) values of y , asymmetric tests consider any cause effect, either the effect of low (high) values of x on high (low) values of y or the effect of high (low) values of x on high (low) values of y (Woodside, 2014).

(iii) **Linear relationship**

The multiple regression analysis undertakes that the relationship between dependent and independent variables is linear and well explained by the square of correlation coefficient in case of simple regression (R^2). Nevertheless, the reality shows the opposite path (McClelland, 1998).

Other limitation of RA is related to the matrix algebra, as stated by Woodside (2013) and Wu *et al.* (2014). Moreover, these authors concluded that a Boolean algebra can contribute to mitigate some issues mentioned above, through testing the relationships among independent variables as well as solving the symmetrical effect issue. This can be achieved by the **Complexity Theory** (hereinafter, CT), useful to go beyond the dominant logic of RA (Woodside, 2014) and to be applied in accounting, consumer research, finance, management and marketing (Woodside, 2013).

Additionally, in many social science applications, the estimators are not unbiased under assumptions LR.1 through LR.4 since omitted factors in the error term are often correlated with the independent variables, known by endogeneity, and then the error term has not zero mean (Wooldridge, 2014).

Despite these limitations, RA should not be avoided but carefully used. In case of falling out its scope or abilities, RA should preferably be substituted by an adequate tool.

2.3. Complexity Theory

The CT considers that RA, as dominant logic, lacks objectivity in what respects to the use of independent variables and the challenge of hypothesis approaches (Armstrong, Brodie and Parsons, 2001). This theory accepts the nonlinear relationship between variables, since the cause effect of huge changes can produce different results (Woodside, 2014). On this way, the CT evaluates if the relationship among variables depends on the complex antecedent conditions (Wu *et al.*, 2014).

Therefore, several authors consider the reality too complex to disregard the dynamic, stochastic and nonlinear processes, considering the RA as a poor tool to fit the reality. Hence, a **configural analysis** is needed to estimate and to describe multiple realities because the simplicity of RA is not sufficient (Woodside, 2014).

Woodside (2014) and González-Velasco, González-Fernández and Fanjul-Suárez (2017) defined the tenets of the CT to mitigate the lack of rigor in order to formalise it. Thus, the CT is defined under six tenets, as follows:

Tenet T.1

Asymmetry principle: *insufficient* but *necessary* condition

A singular independent variable may be *necessary*, although it is mostly *insufficient* for predicting the value of the dependent variable.

Tenet T.2

Recipe principle

Two or more independent variables are *sufficient* for high values of the dependent variable.

Tenet T.3

Equifinality principle: *sufficient* but *not necessary* condition

A model that is *sufficient* is *not necessary* since another independent variable or a combination of independent variables can achieve the same results.

Tenet T.4

Causal asymmetry principle

A rejection does not mean the opposite situation of acceptance.

Tenet T.5

Relationship between independent variables

The presence of a given independent variable can positively or negatively influence the dependent variable depending on the presence or absence of another independent variable(s).

Tenet T.6

Non-perfect correlation

In a set of independent variables, that is relevant for the occurrence of the dependent variable, not all of them are individually significant for the result. As a result, the correlation is always less than 1.

Tenet T.7

Exemptions to the non-perfect correlation

The CT assumes the possibility of the existence of high values of x that predict high values of y as an exception.

2.4. Qualitative comparative analysis

Along different type of qualitative researches (Bansal, Smith and Vaara, 2018), Qualitative Comparative Analysis (hereinafter, QCA) is a **methodological tool** that mitigates the weaknesses of RA and respects the tenets of CT, being a real alternative to the dominant logic. Despite the name, QCA is not a qualitative method but a mix of qualitative⁵ and quantitative⁶ methodologies (Ragin, 2008; Mello, 2019). In addition, QCA is an **approach** since reflects better the social behaviour, the social thinking and the complexity of the reality. Nevertheless, several scientists and researchers use both approaches (i.e., RA and QCA) or other instruments, in order to get a better performance for their investigations (Shering, Korhonen-Kurki and Brockhaus, 2013). According to Berger and Kuckertz (2016), despite the application of QCA for political science and sociology as an accurate method, QCA has been increasingly used in business and management researches. In particular, QCA is preferentially applied at country level and organizational level analysis.

In addition, QCA is an asymmetric model that indicates all the cases or almost of them with relatively high values of the dependent variable that are caused by relatively high values of independent variable(s) (Wu *et al.*, 2014). In fact, neither a simple nor a multiple regression are *necessary* to achieve high values of y . Rather than the net effects of independent variables on the dependent one foreseen by RA (Ragin, 2008), multiple combinations between independent variables are more relevant for the results. To sum up, QCA assumes that the dependent variable depends on how different independent variables are related, rather than the importance of each individual one (Woodside, 2013; Ordanini, Parasuraman and Rubera, 2014; Mattke, Muller and Maier, 2019).

Compared to RA, some nomenclature needs to be adjusted in QCA, which will be used hereinafter, as follows.

Figure 3: Comparison between RA and QCA nomenclatures

Regression analysis	Qualitative comparative analysis
Dependent variable	Outcome
Independent variable	Antecedent condition
Observation	Case
Correlation	Consistency index
Correlation matrix	Truth table
R-squared (R^2)	Coverage index

⁵ Qualitative means non-numerics and inductive theorizing (Bansal, Smith and Vaara, 2018).

⁶ Quantitative means numerics that can be manipulated (Bansal, Smith and Vaara, 2018).

One advantage of QCA is its application on small/intermediate data size since it provides more accurate results, even if the data is small for a quantitative analysis, such as RA, or big for a qualitative analysis, such as QCA (Shering, Korhonen-Kurki and Brockhaus, 2013; Berger and Kuckertz, 2016).

2.4.1. fsQCA, a kind of QCA

Since QCA treats the conditions in a set way, this methodology is also known by set-theoretical method. One kind of QCA approaches is express the conditions in a binary way, such as dummy variables, called Crisp-set Qualitative Comparative Analysis (Ragin, 2008; Shering, Korhonen-Kurki and Brockhaus, 2013). This QCA classifies the conditions in a gradual scale, called by crisp-set, such as “absence” or “presence”, where the 0 means absence and 1 means presence.

However, it is possible to measure the conditions with more exactness, like “absence”, “more absence”, “more presence” and “presence”, achieving more precision and discrimination. This approach is called **Fuzzy-set Qualitative Comparative Analysis (fsQCA)** and it is considered as an extension of the Crisp-set Qualitative Comparative Analysis since it allows the researchers to grade set memberships in fuzzy-sets that range between 0 and 1, where 0 corresponds to “absence”, 1 to “presence” and somewhere between these values will be “more absence” and “more presence” (Rihoux and Regin, 2007; Ragin, 2008; Shering, Korhonen-Kurki and Brockhaus, 2013; Mello, 2019).

The definition of the limits for a fuzzy-set and the consequent attribution of a scale from 0 to 1 is based on judgment and own knowledge of the researcher and/or based on empirical evidence and statistical data (Ragin, 2008; Shering, Korhonen-Kurki and Brockhaus, 2013).

Figure 4: Kind of fuzzy-sets

Crisp-set	Three-value fuzzy-set	Four-value fuzzy-set	Six-value fuzzy-set	Continuous fuzzy-set
$x_i = 1.0$ fully in	$x_i = 1.0$ fully in	$x_i = 1.00$ fully in	$x_i = 1.0$ fully in	$x_i = 1.0$ fully in
		$x_i = 0.75$ more in than out	$x_i = 0.8$ mostly but not fully in	$0.5 < x_i < 1.0$ more in than out
	$x_i = 0.5$ Neither fully in nor fully out		$x_i = 0.6$ More or less in	$x_i = 0.5$ cross-over
		$x_i = 0.25$ more out than in	$x_i = 0.4$ More or less out	
			$x_i = 0.2$ mostly but not fully out	$0.0 < x_i < 0.5$ more out than in
$x_i = 0.0$ fully out	$x_i = 0.0$ fully out	$x_i = 0.00$ fully out	$x_i = 0.0$ fully out	$x_i = 0.0$ fully out

Ragin (2008) clearly stated different types of fuzzy-sets through the figure above.

In the light of the above-mentioned, fsQCA is more than a qualitative approach since it bridges the qualitative to the quantitative approach. Therefore, the QCA is also considered as a quantitative method due to the numerical information between these qualitative states, in particular regarding the *continuous* fuzzy-set (Ragin, 2008; Mello, 2019).

Many authors, such as González-Velasco, González-Fernández and Fanjul-Suárez (2017) and Fotiadis (2018), generally defined three breakpoints values to scale the fuzzy-sets: 0.95 to represent the full membership since original values cover 95% of data values, 0.50 to represent the cross-over and 0.05 to represent the full non-membership since original values cover 5% of data values.

In addition, fsQCA aims at analysing of casual sufficiency to evaluate which antecedent conditions are *sufficient* to obtain the outcome. On one hand, the sufficiency is verified if the cause is a subset of the outcome, since the membership score of the cause is less or equal than its membership score in the outcome. On the other hand, the necessity is verified if the outcome is a subset of the cause. In this case, the membership score of the outcome is less or equal than the membership score in the cause (González-Velasco, González-Fernández and Fanjul-Suárez 2017; Mello, 2019).

QCA uses the Boolean algebra to represent the operations on fuzzy-sets. The three most common operations are the *negation*, the logical *or* and the logical *and*. As the name suggests, the former is the opposite of the membership score and is represented as \sim or with lowercase⁷. In case of a crisp-set, the *negation* of a score of 1 is 0 and *vice-versa*, while with a fuzzy-set the negation is achieved through the following equation:

$$\sim A = 1 - A \tag{9}$$

In its turn, the logical *or*, represented as $+$ or \cup , refers to the union of two or more sets and corresponds to the maximum value across sets. On contrary, the logical *and* is expressed by $*$ or \cap and refers to the intersection of sets. Thus, the logical *and* corresponds to the minimum value across sets (Ordanini, Parasuraman and Rubera, 2014; Mello, 2019).

⁷ In this case, the presence (full membership) is represented as uppercase.

2.4.2. Methodology of QCA

Ordanini, Parasuraman and Rubera (2014) referred that the application of QCA involves four steps:

(i) **Property space**

The definition of the property space consists in determining all possible combinations of antecedent conditions that lead to the occurrence of the outcome.

(ii) **Set-membership measures**

This step, also known by calibration, consists in transforming the original variables, expressed in a continuous scale, into sets in order to make a range from 0 to 1.

According to Longest and Vaisey (2008), the combinations defined in the property space should also include the negation of each antecedent condition. By this way, all cases have some degree of membership measure in every combinations of antecedent conditions, although each case has a membership measure higher than 0.50 in only one combination, called best-fit case.

(iii) **Consistency in set relations**

The third step consists in assessing the combinations that acts as *sufficient* conditions for the occurrence of the outcome, called by consistent cases.

As reported by González-Velasco, González-Fernández and Fanjul-Suárez (2017), consistency is one of the key concepts related to QCA, is equivalent to correlation coefficient and can assessed through the proportion of consistent cases, computed as follows.

$$Consistency (x_i \leq y_i) = \frac{\sum_i^n [\min(x_i; y_i)]}{\sum_i^n x_i} \quad (10)$$

where x_i represents the antecedent conditions, y_i the outcome condition and n the number of observations.

A condition is considered as *sufficient* when its consistency shall statistically exceed a given threshold. Usually, researchers consider a consistency threshold of 0.80 to treat the condition as *sufficient* (Longest and Vaisey, 2008; Ordanini, Parasuraman and Rubera, 2014; Mattke, Muller and Maier, 2019).

(iv) **Logical reduction**

The last step consists in assessing the *sufficient* conditions and eliminating the unneeded elements since some of them are indifferent to achieve the outcome. Hence, it is used another key concept for QCA, the coverage measure (González-Velasco, González-

Fernández and Fanjul-Suárez, 2017), which is computed in order to evaluate the relevance of the *sufficient* conditions.

$$\text{Coverage } (x_i \leq y_i) = \frac{\sum_i^n [\min(x_i; y_i)]}{\sum_i^n y_i} \quad (11)$$

where x_i represents the antecedent conditions, y_i the outcome condition and n the number of observations.

According to González-Velasco, González-Fernández and Fanjul-Suárez (2017), coverage is equivalent to variance in RA.

2.4.3. An empirical example

Several studies have been performed over the years and a lot of researchers have concluded about the usefulness of QCA.

In their research, Ordanini, Parasuraman and Rubera (2014) studied the impact of innovativeness on new hotel service adoption, in particular which combinations of attributes lead to the adoption of the service, since empirical evidences had revealed inconclusive. Through the comparison between RA and QCA, these researchers accomplished that the net effects are too simpler to represent the reality. So, the study concluded that different combinations of antecedent conditions act as *sufficient* conditions for the adoption of a new service.

These authors used the attributes described below as antecedent conditions for the occurrence of new service adoption, that were measured as a degree of perception and were collected through a questionnaire:

Relative advantage

[Adv]

The new service is perceived as better than other alternatives.

Complexity

[Compl]

Complexity corresponds to the perception of how the new service is hard to understand and then an additional effort is needed to adopt the service (for instance, learning lessons or trainings).

Meaningfulness

[Mean]

The new service is perceived as useful.

Novelty

[*Nov*]

The new service is perceived as incongruent, compared with other alternatives, and as uncertain, regarding the consequence of the adoption.

Coproduction requirements

[*Copr*]

Coproduction requirements reflects the organisational choice made by the service provider in which the customer is involved in the service.

Through the use of QCA, the study concluded that the combinations of attributes⁸ that are *sufficient*⁹ to achieve the new service adoption are:

(i) ***mean * compl * ADV * COPR***

The new service is seen as a good alternative, non-complex and with a high degree of coproduction, although it is not immediately perceived as useful.

(ii) ***NOV * ADV * copr***

The adoption is induced by the perception of the new service as a good alternative and as novel but requiring low level of coproduction.

(iii) ***NOV * MEAN * ADV***

The adoption of the new service can be induced when customers perceive it as being novel, useful and a good alternative.

Taking into account these combinations, the researchers concluded that relative advantage is a *necessary* but *not sufficient* condition for the occurrence of the new service since its presence can induce the adoption, however individually presence does not mean the adoption.

Moreover, the antecedent conditions novelty, non-complexity and meaningfulness are neither *necessary*, nor *sufficient* conditions since are not present in the three combinations. In addition, meaningfulness can be either absent (first combination), irrelevant (second combination) or present (third combination) for the occurrence of the new service adoption and non-complexity and novelty can be either present (first combination in case of non-complexity and second and third combinations in case of novelty) or irrelevant (first combination in case of novelty and second and third combinations in case of non-complexity) for the occurrence of the new service adoption.

⁸ Lowercase and uppercase correspond to the absence and presence of the attributes, respectively.

⁹ These three combinations explain 78% of the adoption of the new service (total coverage measure).

Through the using of RA, researchers concluded that, regarding individual effects, relative advantage and novelty have positive effect on the service adoption ($\beta_{Adv} = 0.50^*$ and $\beta_{Nov} = 0.20^*$), being the former the most important predictor. On contrary, complexity and coproduction show negative effects ($\beta_{Compl} = -0.20^*$ and $\beta_{Copr} = -0.18^*$). Additionally, meaningfulness is not relevant as a predictor since it is not statistically significant¹⁰. In what respects to interaction effects, RA reveals the following models as predictors of the new service adoption:

(i) Model 1

$$Adoption_i = \beta_1 Nov * Mean * Copr_i + \beta_2 Mean * Compl * Adv_i + \beta_3 Mean * Adv * Copr_i + u_i \quad (12)$$

(ii) Model 2

$$Adoption_i = \beta_1 Mean * Compl * Adv * Copr_i + u_i \quad (13)$$

The highest order of significance is for *Mean * Compl * Adv * Copr*, which corresponds to the first combination of attributes that are sufficient for the new service adoption revealed by QCA, described above.

Comparing the results obtained with both approaches, the authors concluded that RA revealed small size effects of the independent variables and did not detect the trade-off effects between them while QCA captured the *sufficient* and *necessary* conditions and the relationships between the antecedent conditions for the occurrence of the new service adoption even if some of them had to be absent.

2.5. Share price conditions: An application

This study is based on an investigation of potential factors that influence the *share price* of listed companies. In fact, several studies have been developed in order to find the variables that can trigger the *share price* of enterprises, most of them related to **accounting information**, although some of results have not been conclusive and have shown contradictory results. Moreover, all papers used the *OLS* regression to figure out the contributions for the *share price* variations.

Menaje (2012), Lestari (2017), Nautiyal and Kavidayal (2018) and Hung, Ha and Binh (2018) assessed the impact of some factors on the *share price* of companies listed in Asian stock

* $p < 0.05$.

¹⁰ $p = 0.24$.

markets, such as Philippian, Indonesia, Indian and Vietnam, respectively. It is worth mentioning that only Menaje (2012) used a one-year data (2009), while the others used a multi-year data to perform their analysis (2012-2014, 1995-2014 and 2006-2016, respectively). In fact, the findings revealed inconsistent results in what respects to the influence of the *earning per share* factor (hereinafter, *EPS*) on the *share price* (Menaje, 2012; Nautiyal and Kavidayal, 2018). While Menaje (2012) concluded about the strong positive correlation with the *share price*, Nautiyal and Kavidayal (2018) found a poor relationship between *EPS* and the *share price*. Also, contradictory outcomes were verified for the influence of the *return on asset* factor (hereinafter, *ROA*) because whilst Hung, Ha and Binh (2018) revealed a positive correlation, Menaje (2012) concluded about a weak negative relationship.

In his turn, Lestari (2017) verified that the *retained earnings to total assets* have a positive impact on the *share price*, not only individually, but collectively too, all together with *sales growth* and *sales to current assets*. Similarly, positive relations with the *share price* were also found for the *economic value added* (Nautiyal and Kavidayal, 2018), the *company size* (measured by the net revenue), the *current ratio* (measured by short-term assets over short-term liabilities) and the *accounts receivable turnover* (measured by net revenue over receivables) (Hung, Ha and Binh, 2018). On contrary, Nautiyal and Kavidayal (2018) showed that *dividend per share* and *dividend payout* have a negative effect on the *share price*. Finally, Hung, Ha and Binh (2018) found that the capital structure, in particular the *leverage level* of the company (hereinafter, *LE*), does not have any impact on the *share price*.

In a European research, Avdalovic and Milenkovic (2017) studied the *share price* conditions of companies listed in Serbian stock market, through a multi-year data (2010-2014). The results revealed that the *book value per share* (hereinafter, *BVPS*) and *ROA* had the major contribution for a positive variation of the *share price*. Additionally, *LE* and *price to book ratio* also provided positive contributions for the *share price* fluctuation, despite of a lower meaningful. On contrary, *EPS* and the *company size* (measured by the assets) had a negative impact on the *share price*.

Another factor that has been analysed in several researches, given its influence on the share price fluctuation, is the **structure of corporate ownership**, although the results still remain ambiguous. In fact, corporate governance has become one of the most discussed matter after the last financial crisis, which led several companies to the bankruptcy due to governance issues. However, the corporate ownership also assumes a huge importance considering a direct effect on corporate power in case of an ownership control. In the light of the above mentioned, it is important to assess the type of corporate ownership since each enterprise has a particular

structure: domestic ownership, foreigner ownership, diversified structure of ownership, qualified ownership, managers who have a stake, among others.

Vintila and Gherghina (2014), Alves, Canadas and Rodrigues (2015) and Jankensgard and Vilhelmsson (2018) performed their resourches in European countries, which assessed the impact of corporate ownership's struture on the *share price* of companies listed in the stock markets of Romania, Portugal and Spain, and Swedeen, respectively. In general, the *share price* volatility increased with the number of relatively large shareholders and the portion of shares held by shareholders with stakes lower than 0.1% (Jankensgard and Vilhelmsson, 2018). However, Alves, Canadas and Rodrigues (2015) concluded that the biggest ownership had a negative impact on the share price. Although Vintila and Gherghina (2014) did not obtain statistical significant results regarding the influence of the large ownership, the results revealed that the second and third largest shareholders, as well as the sum of the three largest shareholders, were positively related to the *share price* volatility. On contrary, ownerships lower than 13.08% had negative influence on the *share price* volatility.

In addition, the positive effect of the first and fifth largest shareholder in an individual basis were verified by Alzeaideen and Al-Rawash (2014) in a study of enterprises listed in the Jordanian stock market. However, ElGhouty and El-Masry (2017) did not find any relationship between the ownership concentration and the stock return. These authores only concluded about a positive impact on the *ex ante* risk.

3. Methodology

The comparison between RA and QCA is performed in the context of a business and management research, through a cross-sectional data that includes financial information about companies of 2016.

3.1. Sample characterization

The sample is composed by **265 European listed companies**, which are from the following countries: Austria, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Poland, Portugal, Russia, Spain, Sweden, Switzerland, Turkey and United of Kingdom.

First of all, the 2016 World's Billionaires list is gathered from the Forbes website. The billionaires who have a stake in European listed companies are selected from this list (in a total of 89 billionaires) and the respective companies are added to the sample. It is worth mentioning that Forbes provides a real time list of the world's billionaires, which is updated every day: while the value of public holdings is updated every five minutes, when the correspondent stock markets are open, the billionaires' wealthiness tied to private companies are updated once a day. If a billionaire holds an ownership on a company that represents more than 20% of his/her net worth, the value is adjusted following the industry or region market index.

Secondly, other European listed companies are added to the sample taking into account the same sector/industry and similar size, but without any relationship with the billionaires from the Forbes' list (in a total of 176).

The description of the sample is attached in Annex 1.

3.2. Variables description

The data is composed by the following 8 variables/conditions: billionaire (*BI*), price per share (*PPS*), book value per share (*BVPS*), earnings per share (*EPS*), leverage level (*LE*), return on asset (*ROA*), size of company (*SZ*) and cashflow from operations on asset (*CFOA*). It is worth mentioning that *BIL* is a dummy variable that is defined as a binary variable equal to 1 if a company is owned by a billionaire and 0 otherwise.

The definition and information¹¹ regarding these variables are described in Annex 2 and Annex 3, respectively.

¹¹ The accounting information was collected from the ISCTE-IUL database.

3.2.1. Potential dependent variables / outcome conditions

In this data, the dependent variable or the outcome (in case of RA or fsQCA, respectively) is *PPS*, which is measured in Euros and refers to the price of a single share of a number of saleable stocks issued by a listed company. Therefore, this research aims at assessing the independent variables that have impact on the share price of a European listed company and the antecedent conditions that leads to a higher score of the share price, through the application of RA and QCA, respectively.

3.2.2. Potential independent variables / antecedent conditions

The potential independent variables / antecedent conditions are the remaining ones that are referred in several researches as factors that can influence *PPS*.

While the variables *BVPS* and *EPS* are measure in Euros, *LE*, *ROA*, *CFOA* and *SZ* are percentages since refers to financial ratios, except the variable *SZ* that corresponds to the natural logarithm of company's assets.

3.3. Hypotheses

The hypotheses intend to achieve the objectives of this thesis and therefore verify which factors defined in previous section have more impact on the share price as well as compare both methodologies used. For that purpose, the hypotheses are supported by the literature review. In this sense, their drafting takes into account the relevant papers on these matters.

Hypothesis 1

None antecedent condition regarding accounting information is *sufficient* or *necessary* for a high score of *PPS* (concluded through QCA).

The first hypothesis to be tested intends to demonstrate that none accounting information factor, such as *BVPS*, *EPS*, *LE*, *ROA*, *SZ* or *CFOA*, contributes for *PPS*, considering that several studies, mostly conducted through RA, are not conclusive regarding the factors that have significant impact on *PPS*. The test is performed though the QCA.

Hypothesis 2

Billionaires that hold a stake in a company do not have positively or negatively influence on the company's *PPS* (concluded through RA), neither produces a high score of *PPS* (concluded through QCA).

The second hypothesis aims at verifying the particular impact of the ownership structure, in particular if a company is owned by a billionaire, on *PPS*. This hypothesis is related to this

specific factor because does not exist enough studies about this subject, neither using RA, nor QCA. Due to this fact, this hypothesis is tested through both methodologies, RA and QCA.

Hypothesis 3

Overall, QCA provides similar results to RA.

The third hypothesis to be tested proposes to corroborate some authors' point of view that claim the complementary between both kind of analysis. The test will be conducted through the comparison of the results obtained from both RA and QCA.

Hypothesis 4

Both RA and QCA do not produce univocal results with previous researches in what respects to the share price conditions.

Several researches about these matters have not produced conclusive results. Additionally, some authors argue that RA and QCA should be used as complement tools of each other, as previously referred. Due to these, the fourth hypothesis intends to report that the results provided by both approaches are ambiguous, when the literature review is considered.

3.4. Empirical application

The empirical strategy adopted in this research is the following. Firstly, the **fsQCA software** is used to estimate the model. For that, a definition of the property space is needed based on the knowledge and judgment. Moreover, the calibration is needed since the data is analysed through sets, in particular fuzzy-sets. Additionally, the Boolean algebra is applied and the subset relationships assessed. The final step is to reduce the *sufficient* combinations by deleting redundant elements.

In order to compare the results from the previous approach, a data analysis through a classical software is required. For that, the **Eviews software** is used.

3.4.1. Qualitative comparative analysis

As referred in the literature review section, the first step in the fsQCA methodology is to define the **property space**. For that purpose, all antecedent conditions in Section 3.2.2. are considered since they are drivers that aim at explaining the outcome condition. After that, the non-best-fit cases are excluded from the property space, following the methodology used by some authors. In the second place, the original measures of the conditions are replaced by the set-membership measures of fsQCA methodology, through a process known by **calibration**. Therefore, the original scale of values is transformed into a fuzzy-set scale. For that purpose, three breakpoints

values are defined according to the literature review: 0.95 to represent the full membership, 0.50 to represent the cross-over and 0.05 to represent the full non-membership.

Third, the consistency in set relations and the logical reduction are assessed in the **truth table**, which is extracted from the fsQCA software. For that purpose, it is only considered the combinations of antecedent conditions that have a consistency value higher than 0.80. After that, the truth table solution is generated in order to disclose the conclusions provided by QCA.

3.4.2. Regression analysis

In order to explain the influence of accounting information and ownership structure on company's *PPS*, an econometric analysis is conducted. Thus, it is possible to obtain a model that explain *PPS* as far as possible. After this, the model is tested.

The most accurate model is found out by adding the potential independent variables to the model and decide about their statistical significance by looking for the *t*-test as well as for the **adjusted R^2 (\bar{R}^2)** and the **information criteria**.

As describe in the Section 2.1.2., when variables are added to the model it is possible to verify if they are statistically significant by looking for the raise of the \bar{R}^2 . Unless this coefficient increases, the variables are not statistically significant to explain the dependent variable. On contrary, smaller values of the information criteria means that the model is more reliable.

Firstly, the potential independent variables are individually added and, if \bar{R}^2 increases and the information criteria decreases, the decision is to keep them in the model. Aiming at improving the model, squares of the independent variables and the combination between *dummy - non-dummy* variables are also included. Notice that *non-dummy - non-dummy* combinations are excluded since they do not have great economic and financial interpretation. Once again, variables are excluded if their introduction led a negative impact on \bar{R}^2 or increases the information criteria.

Secondly, the relevance of the independent variables is assessed. Hence, if the respective parameter is not statistically different from zero (*p - value* > α), the independent variable is not relevant to explain the dependent one and is removed as well.

Finally, if a new independent variable is added to the model and makes irrelevant another independent variable already included, it is necessary to find the combination that offers a greater \bar{R}^2 and a lower information criteria.

After that, the **F-test** and the **t-test** (enunciated in the Section 2.1.4.) are applied to verify the statistical significance of the model's parameters, in particular to assess the global significance of the model and the individual relevance of each independent variable, respectively. Moreover, the stability and specification tests, namely **Chow** and **RESET tests**, are also used to detect the presence of specification errors in the model. These tests are described in Annex 4 and in Annex 5, respectively.

The next stage is to test the Classical Linear Model assumptions, which are already defined in the Section 2.1.3.. While the assumptions LR.1 (linear in parameters), LR.2 (random sampling) and LR.4 (zero conditional mean) are considered as truth, the assumptions LR.3 (no perfect collinearity), LR.5 (homoskedasticity) and LR.6 (normality) required the need of performing some tests.

The no perfect (or strong) collinearity assumption is tested using the **matrix of correlations** and the computation of **variance inflation factors** (VIF), which are provided in Annex 6 and Annex 7, respectively.

Jarque-Bera test, which is described in Annex 8, is used to assess the normality of the errors. As complemented, the skewness and kurtosis of the errors' distribution can also support the conclusion about their normality. This approach is also used and is reported in the same annex. Additionally, it is also performed an autocorrelation test.

Although the autocorrelation in the residuals is more common in time series data, since the variables of a cross section data should be independent from each other and therefore none autocorrelation problem is normally verified, it is also important to apply this test in order to guarantee the inexistence of correlation between residuals. The test used is **Breusch-Godfrey Serial Correlation LM test**. Annex 9 offers further details about it.

Regarding the homoskedasticity assumption, the tests used are **Breusch-Pagan-Godfrey test**, **White test** and **White test with cross terms**, which are provided in Annex 10, Annex 11 and Annex 12, respectively. As it is found heteroskedasticity in residuals, that cannot be solved through the logarithmic transformation¹² of neither the dependent variable nor independent ones, it is necessary to model the heteroskedasticity, using **Weighted Least Squares** (WLS) method, in order to get more efficient estimators since the OLS estimators for β are no longer BLUE. Therefore, the estimators for standard errors become biased and inconsistent and as a result the variances and standard errors for OLS estimated coefficients are incorrectly computed. Consequently, the statistical inference can be misleading because it will be based on

¹² The logarithmic transformation is a way to deal to heteroskedasticity when the root cause is the misspecification of the model.

t and F -tests that are no longer valid. All processes of correction for heteroskedasticity are described in Annex 13.

4. Data analysis and empirical results

The results are presented in two different sections, according to the type of analysis performed (QCA or RA).

4.1. Qualitative comparative analysis

Firstly, the property space is defined with all antecedent conditions (in a total of 7), which are attached in Annex 14, and correspond to all possible combinations or configurations that could generate the outcome. Since the antecedent conditions are treated in QCA in a binary way, as “present” or “absent”, the property space included 128 combinations of antecedent conditions (2^7). However, the non-best-fit cases are excluded and then only 59 combinations are considered for the study.

Secondly, the antecedent conditions and the outcome are calibrated taking into account the 0.95, 0.50 and 0.05 percentiles of each antecedent condition and outcome, which are described in Annex 15. However, none calibration is required for the variable *BIL* since it is a dummy variable and therefore it assumes values of 1 (in case a company that is owned by a billionaire) and 0 otherwise. Annex 16 offers further details about the statistical description of the antecedent conditions and the outcome, both calibrated to fuzzy-set membership scores, which reveals that the variables range from 0 to 1, in fact.

Therefore, the model used to study the share price conditions through a QCA is the following:

$$fsPPS = f(fsBVPS, fsEPS, fsLE, fsROA, fsSZ, fsCFOA, BIL) \quad (14)$$

where *fs* means that the respective variable is calibrated to the fuzzy-set scores.

The truth table solution, which reveals the main conclusions of QCA, is provided in Annex 17. The above-mentioned table shows 5 combinations of the antecedent conditions that are *sufficient* for the occurrence of high scores for *PPS* and now are presented below.

$$\begin{aligned} &fsBVPS \\ &fsEPS * fsCFOA \\ &fsEPS * BIL \\ &\sim fsLE * fsSZ * \sim fsCFOA \\ &fsROA * \sim fsSZ * fsCFOA * BIL \end{aligned} \quad (15)$$

These 5 combinations explain 75.7371% of the share price conditions of European listed companies (solution consistency). Additionally, the complete solution explains 95.8152% of the outcome (solution coverage).

The first combination of the antecedent conditions is characterized by high scores for *BVPS*. In its turn, the second combination requires high scores for both *EPS* and *CFOA*. The third combination is also characterized by high scores for *EPS* but it is also required that, at the same time, the company is held by a billionaire (*BIL*). The fourth combination presents a high score for *SZ*, but low scores for *LE* and *CFOA*. On contrary, the last configuration requires high scores for *CFOA* as well as for *ROA*, but low scores for *SZ*, although in any case the company has to be owned by a billionaire (*BIL*).

The analysis of these 5 *sufficient* combinations allows to conclude that *BVPS* is, in fact, a *sufficient by not necessary* condition for the occurrence of high scores for *PPS* because although its solely presence can produce high scores for *PPS*, its presence in other combinations is not necessary for the production of high scores for *PPS*. This solution explains 88.958% of the outcome.

In addition, *CFOA*, *EPS*, *SZ*, *ROA*, *BIL* and *LE* are neither *sufficient* nor *necessary* to achieve high scores for *PPS*. In particular, while *EPS*, *ROA* and *BIL* can be either present or irrelevant to produce a high score for *PPS*, *LE* can be either absent or irrelevant. In its turn, *CFOA* and *SZ* can be present, absent or irrelevant to achieve high scores of *PPS*.

Lastly, it is worth mentioning that, apart from the first combination, the solution terms of the remaining combinations cannot solely explain very well the high scores for *PPS* since the unique coverage are lower than 3%. However, considering the relationship with other solution terms (raw coverage), the first and second combinations explain higher than 65% of high scores for *PPS* (each one), the third and fourth combinations explain 35% (each one) and the fifth combination only can explain 15%.

4.2. Regression analysis

As referred previously, modelling process in regression analysis starts with all potential independent variables. After that, the assessment about their relevance reveals that the model with the highest \bar{R}^2 and the lowest **information criteria** (72.8156% and 10.87875, respectively) is the following:

$$PPS_i = \beta_0 + \beta_1 BVPS_i + \beta_2 EPS_i + \beta_3 EPS_i^2 + \beta_4 BVPS * BIL_i + \beta_5 CFOA * BIL_i + \beta_6 SZ * BIL_i + u_i \quad (16)$$

This model is considered as the **initial model** and further details are presented in Annex 18.

4.2.1. Testing the model

As revealed in Annex 18, the p – *value* of the **F-test** is lower than 5% significance level. Therefore, the null hypothesis is rejected and then it is possible to conclude about the existence of at least one parameter statically significant and, consequently, at least one relevant regressor. In its turn, the results of the **t-test** applied to each parameter reveal that all independent variables are statistically different from zero since the respective p – *values* are also lower than 5% significance level and then the null hypothesis is rejected (see the output of t -tests in Annex 18).

Regarding to stability and specification tests, **Chow test**¹³ reveals that the impact on *PPS* is different between companies that are owned by billionaires and other companies since the p – *value* is lower than 5% significance level and then the null hypothesis is rejected (see the output of Chow test in Annex 19). Concerning **RESET test**, which the output is attached in Annex 20, the null hypothesis is also rejected (p – *value* is lower than 5% significance level), meaning that the model is not well specified. In other words, the model can have relevant independent variables omitted, incorrect functional form or correlation between independent variables and errors. Several changes in the model are performed in order to get a well-defined model and then with a correct functional form. However, none of them produces positive results in RESET test since the null hypothesis is always rejected (p – *value* lower than 5% significance level). Also, the residuals are analysed through a plot, which is attached in Annex 21, in order to find the root cause for this issue. In fact, there is an outlier in the residuals (its value is 657.42), which is excluded from the data. Therefore, the sample is now composed by 264 observations and, although the model is the same, \bar{R}^2 is higher (86.4058% vis-à-vis 72.8156%) and information criteria is lower (9.948823 vis-à-vis 10.87875), as shown in Annex 22. However, RESET test continues to show that the model is not well specified, unless it is considered 1% significance level. In this case, the null is not rejected because the p – *value* is higher than 1% (Annex 23).

4.2.2. Testing the assumptions of the model

Either the **matrix of correlations** between the independent variables or the computation of **VIFs** reveals the inexistence of perfect collinearity (see the respective outputs in Annex 24 and Annex 25, respectively) since, in general, the former shows correlations lower than 0.8 in absolute value and the latter presents results lower than 10.0. However, the variables *EPS*² and

¹³ The breakpoint used was 177 that was achieved after sorting the dummy variable BIL.

EPS present a slightly strong correlation according the matrix of correlations (Equation 17) as well as EPS^2 has a slight high value of centered VIF (Equation 18).

$$\text{Corr}(EPS^2, EPS) = 0.915565 \quad (17)$$

$$VIF_{EPS^2} = 10.60458 \quad (18)$$

These results are easily justified by the presence of related variables in the model, in particular the inclusion of the squares of variable itself. For this reason, it is expected that *EPS* is strong correlated with EPS^2 since the latter is the square of the former.

Together with the linear in parameters¹⁴, the random sampling¹⁴ and zero conditional mean¹⁴ assumptions, it is possible to conclude that OLS estimators for β are **unbiased** and, assuming the asymptotic properties since the sample is sufficiently big, **consistent**.

In what respects to the normal distribution of the errors, **Jarque-Bera test** concludes that the residuals point for a non normal distribution since the null hypotheses is rejected considering 5% significance level (*p – value* is equal to zero). Although the skewness has been closed to zero (0.808742), the kurtosis is different of the desired value (13.08751 vis-à-vis 3.0) that corroborated the non normal distribution of the errors (Annex 26). However, as the sample is considered as large ($n = 264$), the OLS estimator for β continues to follow asymptotically the normal distribution according to the Central Limit Theorem¹⁵ (CLM). Therefore, the violation of this assumption does not have practical consequences in the OLS estimators for β since it does not influence their potential efficiency and consistency.

Testing for the serial correlation, through **Breusch-Godfrey Serial Correlation LM test**, reveals the inexistence of correlation between errors since the null hypothesis is not rejected considering a 5% significance level (*p – value* = 0.4211; Annex 27). In fact, this conclusion is expected because the serial correlation is not common in cross section but in time series or panel data. However, it is important to test for the serial correlation, together with the remaining tests for the other assumptions, in order to guarantee that the OLS estimators for β are BLUE, indeed.

Finally, **Breusch-Pagan-Godfrey test**, **White test** and **White test with cross terms** (all of them testing for heteroskedasticity) lead to the same conclusion (Annex 28, Annex 29 and Annex 30, respectively). Based on the *p – value* lower than 5% significance level (zero for all

¹⁴ These three assumptions were considered as truth, as explained in the Section 3.4.2..

¹⁵ The Central Limit Theorem foresees that the distribution of the sum of independent random variables, when standardized by its standard deviation, tends to be normal distributed as the size of the sample increases.

of them), the null hypotheses are rejected and consequently the errors are heteroskedastic, meaning that the variance of the errors is not zero. As the BLUE characteristic requires the homoskedasticity assumption, which is not verified through the homoskedasticity tests, the OLS estimators for β are **no longer BLUE**, as referred above. Although the OLS estimators for β continue to be unbiased and consistent, they are no longer asymptotically efficient. This means that it is possible to find other linear estimators for β with smaller variance than the OLS estimators. As mentioned in the previous chapter, the WLS¹⁶ is used to get more efficient estimators. According to the output attaches in Annex 31, the independent variable *EPS* is no longer statistically significant, as the *p* – *value* of the *t* test is higher than 5% significance level, and a new model is found, called as **final model**.

$$PPS_i = \beta_0 + \beta_1 BVPS_i + \beta_2 EPS_i^2 + \beta_3 BVPS * BIL_i + \beta_4 CFOA * BIL_i + \beta_5 SZ * BIL_i + u_i \quad (19)$$

The final model presented explains better the dependent variable than the initial model¹⁷ since \bar{R}^2 is higher (99.6427% vis-à-vis 86.4058%) and information criteria is lower (8.346315 vis-à-vis 9.948823), as shown by Annex 32.

In order to confirm if the errors are no longer heteroskedastic, the respective tests are applied once again. As can be analysed in Annex 33, Annex 34 and Annex 35, although Breusch-Pagan-Godfrey test and White test with cross terms reveal that errors continue to be heteroskedastic (the *p* – *value* of both tests are close to zero and the null hypotheses are rejected), the White test concludes about the homoskedasticity of the errors as the null hypothesis is not rejected based on a 5% significance level (*p* – *value* = 0.0944).

Finally, it is also important to perform the testing for the serial correlation once again (the output is presented in Annex 36), which shows that the errors respect the no autocorrelation assumption because the *p* – *value* of the test is higher than 5% significance level. Therefore, the GLS estimators for β are BLUE.

Comparing the outputs in Annex 22 and Annex 32, it is possible to see that the GLS estimators for β are more efficient than the OLS estimators for β , in fact. Since the GLS estimators for β are BLUE, they can be interpreted without any problem. However, this does not mean the estimators can be compared as the differences are related with changes in signal or in magnitude.

¹⁶ In particular, the Generalised Least Squares (GLS).

¹⁷ Considering the exclusion of the outlier in the residuals.

4.2.3. Results interpretation

Equation 20 presents the significant¹⁸ parameter coefficients for each independent variable that explain the price per share of companies listed in European stock markets.

$$PPS_i = 5.535671 + 1.168503 BVPS_i - 0.030817 EPS_i^2 + 1.083480 BVPS * BIL_i + 140.4552 CFOA * BIL_i - 0.761170 SZ * BIL_i + u_i \quad (20)$$

The output of the final model suggests a R^2 of 0.996490. This indicates that the model explains 99.65% of the company's price per share. Nevertheless, the R^2 is sensitive to the number of explanatory variables, never decreasing with their inclusion. Facing this issue, it is also important to interpret the \bar{R}^2 , which means that when the relationship between the PPS and its independent variables is established, it is possible to explain or eliminate 99.6422% of the PPS 's variance. Thus, the remainder 0.3578% represents part of the PPS 's variation that cannot be explained by the model.

Additionally, the model reveals the following conclusions:

- (i) The marginal effect of EPS on PPS is not constant. When EPS is null (zero), PPS hits a peak, which is also zero. This means that when EPS is zero, PPS is also zero, *ceteris paribus*. Notice that, until the peak, the marginal effects of EPS are positive at decreasing rates and from the peak, the marginal effects are negative at increasing rates, everything else constant;
- (ii) The impact of $BVPS$ on PPS depends on the ownership structure, in particular if a billionaire holds a stake in a European listed company. In this case, an additional Euro in $BVPS$ has a positive impact on PPS of 2.25 EUR , *ceteris paribus*. Otherwise, one more Euro in $BVPS$ has only a positive impact of 1.17 EUR , everything else constant;
- (iii) $CFOA$ only has impact on PPS when a billionaire holds a stake in a European listed company. Therefore, an additional Euro in $CFOA$ leads to an increase of 140.46 EUR in PPS , *ceteris paribus*;
- (iv) The impact of the company size (SZ) on PPS is similar to the impact of $CFOA$, meaning that it depends on the existence of a billionaire that holds a stake on a European listed company. Since the size of the enterprise is measured as the natural logarithmic of its assets, a 100% increase of a European listed company's assets, which is held by a billionaire, leads to a decrease of 0.76 EUR in respective PPS , *ceteris paribus*.

¹⁸ Taking into account a significance level of 5%.

5. Conclusions

This study empirically compares two types of analysis: RA and QCA. For that purpose, it was used a cross-sectional sample composed by 265 European listed companies that includes financial information of 2016, such as *PPS*, *BVPS*, *EPS*, *LE*, *ROA*, *SZ*, *CFOA* and *BIL*, in order to study the factors that influence *PPS*. The outputs obtained from both methodologies are used to draft the main conclusions.

Through QCA, the combinations of antecedent conditions that are *sufficient* to produce high scores of *PPS* are achieved after the calibration process based on statistical data (0.95 represents the full membership, 0.50 represents the cross-over and 0.05 represents the full non-membership) and the exclusion of antecedent conditions that have a consistency value lower than 0.80. Therefore, it is obtained five combinations that explain 75.7371% of the share price conditions in European listed companies (given by the solution consistency), as follows.

$$\begin{aligned}
 & fsBVPS \\
 & fsEPS * fsCFOA \\
 & fsEPS * BIL \\
 & \sim fsLE * fsSZ * \sim fsCFOA \\
 & fsROA * \sim fsSZ * fsCFOA * BIL
 \end{aligned}$$

Except for the first case, all combinations reveal that two or more antecedent conditions can explain high score of the outcome, aligned to the *recipe principle* of the Complexity Theory (Tenet T.2). Moreover, these combinations also show that the relationship between a given set of variables produce different results compared with the combinations of others. For instance, while high score of *PPS* can be achieve with the presence of *CFOA* (second and fifth combinations), high score of *PPS* can also be accomplished with its absence (fourth combination). In fact, this situation reflects the fifth tenet of the Complexity Theory (Tenet T.5).

Through RA, it is possible to achieve a model that explains 99.65% of the company's price per share.

$$PPS_i = \beta_0 + \beta_1 BVPS_i + \beta_2 EPS_i^2 + \beta_3 BVPS * BIL_i + \beta_4 CFOA * BIL_i + \beta_5 SZ * BIL_i + u_i$$

This model had to be estimated through the GLS since the tests for heteroskedasticity revealed that the variance of residuals was not constant and as result the respective estimators for β were no longer BLUE.

With the findings described in the previous section, it is possible to examine the four posed hypotheses, as follows.

Hypothesis 1

None antecedent condition regarding accounting information is *sufficient* or *necessary* for a high score of *PPS* (concluded by QCA).

This hypothesis respects to the assessment of the share price conditions through QCA. This analysis reveals that *CFOA*, *EPS*, *SZ*, *ROA*, *BIL* and *LE* are neither *sufficient* nor *necessary* to achieve high scores for *PPS*. In particular, while *EPS*, *ROA* and *BIL* can be either present or irrelevant to produce a high score for *PPS*, *LE* can be either absent or irrelevant. In its turn, *CFOA* and *SZ* can be present, absent or irrelevant to achieve high scores of *PPS*. However, *BVPS* is, in fact, a *sufficient* condition, despite not *necessary*, for the occurrence of high scores for *PPS*, that corroborates the *equifinality principle* of the Complexity Theory (Tenet T.3). In addition, the respective solution explains 88.958% of *PPS* score. Although these findings do not corroborate the first hypothesis, it figures out that share price is hard to explain since several studies show different conclusions. Taking into account the antecedent conditions used in this research, the studies that are considered in the literature review only show that *EPS* and *ROA* have strong impact on *PPS* and *EPS*.

Hypothesis 2

Billionaires that hold a stake in a company do not have positively or negatively influence on the company's *PPS* (concluded by RA), neither produces a high score of *PPS* (concluded by QCA).

A lot of studies that aim at assessing the impact of the ownership structure on the corporate performance have been conducted. However, none of them recognizes a possible relationship between a company that is owned by a billionaire and the share price. Despite this fact, some studies on this matter indicate that ownership structures have a direct influence in share price, although they are not conclusive about the ones that affect most the share price. In particular, this research reveals that, through RA, the impact on *PPS* is different between the companies that are owned by billionaires and the ones that are not. In fact, the influence of *BVPS*, *SZ* and *CFOA* is enhanced by a presence of a billionaire on the ownership of the company. However, QCA reveals that *BIL* can be either present or irrelevant to produce a high score for *PPS*. In general, this finding corroborates the literature in sense the ownership structure influences the share price, but it does not show a strong and accurate conclusion. Therefore, this hypothesis is verified.

Hypothesis 3

Overall, QCA provides similar results to RA.

Some authors argue that both approaches should be used in order to get a better performance of researches. Therefore, both kind of analysis cannot product contradictory results. The findings indicate that the positive and strong impact of *BVPS* on *PPS* is confirmed by QCA results since it also reveals a relation between these two variables. Moreover, QCA assesses this relationship in a deep manner, revealing that *BVPS* is a *sufficient* condition for the occurrence of high score of *PPS*. Therefore, it can be concluded that QCA complements the results obtained from RA. Although QCA does not confirm the results of RA regarding *EPS*, *BIL*, *SZ* and *CFOA*, it can be considered that this hypothesis is true based on the univocal conclusion about *BVPS*. In fact, as this analysis falls in the scope or abilities of RA, since it is a predominantly quantitative analysis, QCA can be used as a complement and not as a substitute tool.

Hypothesis 4

Both RA and QCA do not produce univocal results with previous researches in what respects to the share price conditions.

It seems to be common that researches, which pertain to explain the factors that influence share price, are not conclusive in the literature. In fact, the findings of this research also show that the factors with major contribution for the share price (such as, *BVPS*) are not referred in the studies that support the literature review. However, as referred in the conclusions about the third hypothesis, the results obtained from QCA complement the results from RA. Therefore, it can be possible to conclude that this hypothesis is true: despite the fact that QCA and RA offer similar results, they are not aligned with the literature, which is inconclusive.

Additionally, this study reveals that the use of QCA is not enough to assess the share price conditions since RA also provides relevant information, in general. While QCA treats the sample in a qualitative way and the respective conclusions highlight a potential qualitative relationship between antecedent conditions and outcome, since antecedent conditions can be *sufficient* and *necessary* for the occurrence of the outcome, RA specifies and measures the impact of each independent variable on the dependent one. Therefore, QCA helps to expand the comprehension regarding the conditions needed for the attainment of the outcome.

Regarding the limitations of this study, it is important to take into account that the sample is composed by a one-year data, which can provide distorted results. In this sense, a sample with

a long period of time is required to produce more accurate results through the capture of the volatility of the prices in stock markets. Moreover, it is also recommended to use updated data, since this study uses data from 2016.

In addition, it is suggested to include in the sample more conditions / independent variables that potentially explain the share price, such as *growth of sales*, *sales to current assets*, *economic value added*, *dividend payout*, *dividend per share*, among others, in order to enhance the results and conclusions. Also, a deep review of the literature helps to support the choice of variables that can be considered in the estimation, since most of them have heterogeneous conclusions.

Bibliography

- Alves, H., Canadas, N., & Rodrigues, A. M. 2015. Determinants of share price and share liquidity: An analysis using SEM model. *Procedia Economics and Finance*, 25: 318-331.
- Alzeaideen, K. A., & Al-Rawash, S. Z. 2014. The effect of ownership on share price volatility of listed companies in Amman Stock Exchange. *Research Journal of Finance and Accounting*, 5(6): 192-201.
- Armstrong, J. S., Brodie, R. J., & Parsons, A. G. 2001. Hypotheses in marketing science: Literature review and publication audit. *Marketing Letters*, 12(2): 171-187.
- Armstrong, J. S. 2012. Illusions in regression analysis. *International Journal of Forecasting*, 28: 689-694.
- Avdalovic, S. M., & Milenkovic, I. 2017. Impact of company performances on the stock price: An empirical analysis on select companies in Serbia. *Economics of Agriculture*, 64: 561-570.
- Bansal, P., Smith, W., & Vaara, E. 2018. New ways of seeing through qualitative research. *Academy of Management*, 61(4): 1189-1195.
- Berger, E., & Kuckertz, A. 2016. *Complexity in entrepreneurship, innovation and technology research – Applications of emergent and neglected methods*. Switzerland: Springer International Publishing.
- ElGhouty, A. S., & El-Masry, A. A. 2017. Effect of ownership structure on firm stock return performance: Evidence from the Egyptian Stock Market. *International Research Journal of Finance and Economics*, 160: 63-83.
- Fotiadis, A. 2018. Modelling wedding marketing strategies: An fsQCA analysis. *Journal of Tourism, Heritage & Services Marketing*, 4: 23-26.
- Friedman, M., & Schwartz, A. J. 1991. Alternative approaches to analysing economic data. *American Economic Review*, 81: 39-49.
- González-Velasco, C., González-Fernández, M., & Fanjul-Suárez, J. 2017. Does innovative effort matter for corporate performance in Spanish companies in a context of financial crisis? A fuzzy-set QCA approach. *Empirical Economics*, 56: 1707-1727.
- Jankensgard, H., & Vilhelmsson, A. 2018. The shareholder base hypothesis of stock return volatility: Empirical evidence. *Financial Management*, 47(1): 55-79.
- Hung, D. N., Ha, H. T., & Binh, D. T. 2018. Impact of accounting information on financial statements to the stock price of the energy enterprises listed on Vietnam's stock market. *International Journal of Energy Economics and Policy*, 8(2): 1-6.
- Lestari, A. 2017. *The effect of fundamental factors on stock price (empirical study share sector manufacturing)*. Unpublished paper.
- Longest, K., & Vaisey, S. 2008. Fuzzy: A program for performing qualitative comparative analyses (QCA) in STATA. *The Stata Journal*, 8(1): 79-104.
- Mattke, J., Muller, L., & Maier, C. 2019. *Paid, owned and earned media: A qualitative comparative analysis revealing attributes influencing consumer's brand attitude in social media*. Paper presented at the 51th Hawaii International Conference on System Sciences, Hawaii.
- McClelland, D. C. 1998. Identifying competencies with behavioral-event interviews. *Psychological Science*, 9(5): 331-339.
- Mello, P.A. 2019. *Qualitative comparative analysis: Research design and application*. Washington, D. C.: Georgetown University Press.
- Menaje, P. 2012. Impact of selected financial variables on share price of publicly listed firms in the Philippines. *American International Journal of Contemporary Research*, 2(9): 98-104.

- Nautiyal, N., & Kavidayal, P. C. 2018. Analysis of insitutional factors affecting share prices: The case of national stock exchange. *Global Business Review*, 19(3): 707-721.
- Ordanini, A., Parasuraman, A., & Rubera, G. 2014. When the recipe is more important than the ingredients: A qualitative comparative analysis (QCA) of service innovation configurations. *Journal of Service Research*, 17(2): 134-149.
- Ragin, C. C. 2008. *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago: University of Chicago Press.
- Shering, J., Korhonen-Kurki, K., & Brockhaus, M. 2013. *Qualitative comparative analysis (QCA): An application to compare national REDD+ policy processes*. Working paper no. 121, CIFOR, Bogor, Indonesia.
- Soyer, E., & Hogarth, R. M. 2012. The illusion of predictability: How regression statistics mislead experts. *International Journal of Forecasting*, 28(3): 695-711.
- Vintila, G., & Gherghina, S. C. 2014. The impact of ownership concentration on firm value: Empirical study of the Bucharest Stock Exchange listed companies. *Procedia Economics and Finance*, 15: 271-279.
- Woodside, A. G. 2013. Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. *Journal of Business Research*, 66: 463-472.
- Woodside, A. G. 2014. Embrace perform model: Complexity theory, contrarian case analysis, and multiple effects. *Journal of Business Research*, 67: 2495-2503.
- Wooldridge, J. M. 2014. *Introduction to econometrics: Europe, Middle East and Africa edition*. Andover: Cengage Learning.
- Wu, P., Yeh, S., Huan, T., & Woodside, A. G. 2014. Applying complexity theory to deepen service dominant logic: Configural analysis of customer experience-and-outcome assessments of professional services for personal transformations. *Journal of Business Research*, 67: 1647-1670.

Annexes

Annex 1

Sample characterization

Distribution of companies by countries

Countries	#
Germany	26
Austria	3
Denmark	3
Spain	17
Finland	6
France	46
Netherlands	9
Italy	38
Norway	5
Poland	11
Portugal	9
Russia	33
Sweden	16
Switzerland	3
Turkey	24
United Kingdom	16
TOTAL	265

Distribution of companies by dummy variable billionaire (*BIL*)

Shareholders	#
Billionaires	89
Non-billionaires	176
TOTAL	265

Annex 2

Variable definition

Variable	Definition	Formula	Unit
Billionaire [BIL]	A person with a net worth of at least one billion.	Dummy variable, defined as a binary variable equal to 1 if a company is owned by a billionaire and 0 otherwise	-
Price per share [PPS]	The price of a single share of a number of saleable stocks issued by a listed company.	-	EUR
Book value per share [BVPS]	The value of shares based on the common share outstanding.	$BVPS = \frac{\text{Common shareholders' equity}}{\text{Common shares outstanding}}$	EUR
Earnings per share [EPS]	The fraction of the profit allocated to each common share outstanding.	$EPS = \frac{\text{Net income}}{\text{Common shares outstanding}}$	EUR
Leverage level [LE]	The amount of debt that a company uses to finance assets.	$LE = \frac{\text{Total debt}}{\text{Total assets}}$	%
Return on asset [ROA]	Indicator of how profitable a company is relative to its total assets.	$ROA = \frac{\text{Net income}}{\text{Total assets}}$	%
Size of company [SZ]	In this case, the size of a company is measured as the natural logarithm of assets.	$SZ = \ln(\text{total assets})$	%
Cashflow from operations on asset [CFOA]	An efficiency ratio that charges the cashflows from the ongoing regular business activities to assets.	$CFOA = \frac{\text{Cashflow from operations}}{\text{Total assets}}$	%

Annex 3

Description of variables

Regressor	Mean	Median	Standard deviation	Minimum	Maximum
PPS	43.44	14.75	105.50	0.00	950.25
BVPS	19.27	5.33	65.14	-10.89	848.32
EPS	2.75	0.74	9.24	-8.15	107.25
LE	0.43	0.47	0.31	0.00	1.83
ROA	0.03	0.02	0.08	-0.82	0.30
SZ	10.81	14.30	6.54	0.28	19.29
CFOA	0.06	0.05	0.07	-0.10	0.35
BIL		Dummy variable		0.00	1.00

Annex 4

[RA] Chow test

In case of cross-sectional data, the Chow test allows the researcher to test the differences in regression functions across groups.

If dummy variables and/or cross regressors with dummy variables (such as dummy - dummy, non-dummy - dummy interactions) are included in the model, implying the existence of two different groups, it is important to know if these groups are effectively relevant. In other words, this test is often used to determine if the independent variables have different impacts on different subgroups of the population.

The hypotheses for the Chow test are:

$$\begin{cases} H_0: \text{there is no difference between two groups} \\ H_1: \text{there is difference between two groups} \end{cases} \quad (21)$$

The respective statistic is:

$$F = \frac{\frac{SSR - (SSR_1 + SSR_2)}{k + 1}}{\frac{SSR_1 + SSR_2}{n - 2(k + 1)}} \sim F_{k, n-2(k+1)} \quad (22)$$

where SSR is the sum of squared of residuals, which measures the difference between the data and the values predicted by the model, k is the number of independent variables and n the observation numbers.

Given the p - value, the decision rule is:

- (i) p - value $< \alpha$: the null hypothesis is rejected;
- (ii) p - value $> \alpha$: the null hypothesis is not rejected.

If the null hypothesis is not rejected, there is no difference between two groups tested. On contrary, if the null hypothesis is rejected, there is difference between these two groups.

Annex 5

[RA] RESET test

The RESET test allows researchers to test the omission of relevant explanatory variables, incorrect function form and correlation between explanatory variables and the errors of the model. If one of them occurs, it is enough to conclude that the *OLS* estimators for β are biased and inconsistent.

The RESET test assess if any non-linear function of the variables added to the model represented by the Equation 23, such as functions with squared and cubic terms (represented by \hat{y}^2 and \hat{y}^3 in the Equation 24), are statistically different from zero.

$$y = X\beta + u \tag{23}$$

$$y = X\beta + \delta_1\hat{y}^2 + \delta_2\hat{y}^3 + error \tag{24}$$

Therefore, the hypotheses for the RESET test are:

$$\begin{cases} H_0: \delta_1 = \delta_2 = 0 \\ H_1: \exists \delta_j \neq 0 \end{cases} \quad j = 1, 2 \tag{25}$$

The respective statistic is:

$$F = \frac{\frac{R_{UR}^2 - R_R^2}{2}}{\frac{1 - R_{UR}^2}{n - k - 3}} \sim F_{2, n-k-3} \tag{26}$$

where R_{UR}^2 and R_R^2 represent R^2 of the unrestricted model (related to the Equation 24) and of the restricted model (related to the Equation 23), respectively, k is the number of independent variables and n the observation numbers.

Given the *p – value*, the decision rule is:

- (i) *p – value* < α : the null hypothesis is rejected;
- (ii) *p – value* > α : the null hypothesis is not rejected.

If the null hypothesis is not rejected, the model is well specified. However, if the null hypothesis is rejected, the estimates for δ_1 and δ_2 are statistically significant, meaning that there are non-linear functions of independent variables omitted in the model represented by the Equation 23. Therefore, it needs to change the functional forms, for instance, to linear-log, log-linear, log-log or quadratic forms.

Annex 6

[RA] Matrix of correlations

The matrix of correlations shows the correlations between the independent variables. Correlation means how strongly pairs of variables are related, whether causal or not, and it is measured through the correlation coefficient that ranges between -1 and 1 . The matrix of correlations is a set of correlation coefficients between all independent variables.

If a correlation coefficient is higher than 0.8 in absolute terms [$Corr(x_i, x_j) > |0.8|$], there is a strong collinearity. However, this correlation coefficient is not enough to violate the no perfect collinearity assumption. On contrary, the assumption is violated in case of perfect collinearity [$Corr(x_i, x_j) = |1|$] or when almost there.

Annex 7

[RA] Variance inflation factors

The variance inflation factors, also known by VIFs, correspond to the terms in the sampling variance affected by correlation between independent variables and therefore allow researchers to assess the collinearity among independent variables.

VIFs are computed as follow.

$$VIF_j = \frac{1}{1 - R_j^2} \quad , \quad j = 2, 3, \dots, k \quad (27)$$

where R_j^2 is the R^2 of the model with the dependent variable j .

Strong symptoms of collinearity are considered as above 10, from which the no perfect collinearity assumption is considered violated.

Annex 8

[RA] Jarque-Bera test

The Jarque-Bera test allows researchers to test the error's normality. It compares the skewness and kurtosis of the error's distribution with a normal distribution. Here, the skewness should tend to zero and the kurtosis should tend to three to be considered a normal distribution.

This test only is valid for big samples. When it has small samples, the *Chi-squared* approximation is very sensible. This means that it can consider no error's normality when they are really normal distributed.

Given the linear regression function expressed by the Equation 23, the hypotheses for the Jarque-Bera test are:

$$\begin{cases} H_0: u|X \sim Normal \\ H_1: u|X \not\sim Normal \end{cases} \quad (28)$$

The respective statistic is:

$$JB = n \left\{ \frac{\widehat{S}(X)^2}{6} + \frac{[\widehat{K}(X) - 3]^2}{24} \right\} \sim \chi^2_{(2)} \quad (29)$$

Given the *p – value*, the decision rule is:

- (i) *p – value* < α : the null hypothesis is rejected;
- (ii) *p – value* > α : the null hypothesis is not rejected.

If the null hypothesis is not rejected, the errors are normal distributed. In this case, the error's kurtosis and skewness should tend to three and zero, respectively. However, if the null hypothesis is rejected, the errors are not normal distributed. In special case of a big sample, the errors could not be normal distributed, if the null hypothesis is rejected, but the *OLS* estimators for β are normal distributed, under CLT. Nevertheless, the *OLS* estimators for β still continue to be unbiased and consistent.

Annex 9

[RA] Breusch-Godfrey Serial Correlation LM test

Another assumption that is common to test is the no autocorrelation of the errors. This is as important as the homoskedasticity because it leads to efficient estimators, even keeping their unbiased and consistence. That means, if this assumption is violated, the OLS estimators for β are no longer BLUE.

The null hypothesis of the LM test is that there is no serial correlation.

The test statistic is computed by an auxiliary regression as follows. First, suppose you have estimated the regression:

$$y_i = X_i\beta + u_i$$

where β are the estimated coefficients and u are the errors.

The test statistic for lag order p is based on the auxiliary regression for the residuals:

$$u_i = y_i - X\hat{\beta}$$

$$e_i = X_i\gamma + \left(\sum_{j=1}^p \alpha e_{i-j} \right) + v_i$$

This is a regression of the residuals on the original regressors X and lagged residuals up to order p . The F -statistic is an omitted variable test for the joint significance of all lagged residuals.

The $\text{Obs} \cdot R^2$ statistic is the Breusch-Godfrey LM test statistic. This LM statistic is computed as the number of observations times the (uncentered) R^2 from the test regression. Under quite general conditions, the LM test statistic is asymptotically distributed as a $\chi^2(p)$.

Annex 10

[RA] Breusch-Pagan-Godfrey test

The Breusch-Pagan-Godfrey test allows researchers to verify if the square of the error term is related to one or more independent variables, i.e. the expected value of the error term square might be some function, at least, of one out of the explanatory variables.

Given the linear regression function from the Equation 30, the regression of the error term square estimator can be written as the Equation 31.

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i \quad (30)$$

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + error \quad (31)$$

Therefore, the hypotheses for the Breusch-Pagan-Godfrey test are:

$$\begin{cases} H_0: \delta_1 = \dots = \delta_k = 0 \\ H_1: \exists \delta_j \neq 0 \end{cases} \quad j = 1, 2, \dots, k - 1 \quad (32)$$

The respective statistic is:

$$LM = n * R_{\hat{u}^2}^2 \sim \chi_{(k-1)}^2 \quad (33)$$

Given the $p - value$, the decision rule is:

- (i) $p - value < \alpha$: the null hypothesis is rejected;
- (ii) $p - value > \alpha$: the null hypothesis is not rejected.

If the null hypothesis is rejected, there is heteroskedasticity and hence the error conditional variance on the independent variable is not constant.

Annex 11

[RA] White test

The White test is similar to the Breusch-Pagan-Godfrey test, but considers the square and the cube of the all independent variables from the regression of the error term square estimator.

Given the linear regression function expressed by the Equation 30, the regression of the error term square estimator can be written as follow.

$$\hat{u}^2 = \theta + \delta_1 x_1 + \dots + \delta_k x_k + \gamma_1 x_1^2 + \dots + \gamma_k x_k^2 + \varphi_1 x_1^3 + \dots + \varphi_k x_k^3 + error \quad (34)$$

Therefore, the hypotheses for the White test are:

$$\begin{cases} H_0: \delta_1 = \dots = \delta_k = \gamma_1 = \dots = \gamma_k = \varphi_1 = \dots = \varphi_k = 0 \\ H_1: \text{at least one different from 0} \end{cases} \quad (35)$$

The respective statistic is also represented by the Equation 33.

Given the $p - value$, the decision rule is:

- (i) $p - value < \alpha$: the null hypothesis is rejected;
- (ii) $p - value > \alpha$: the null hypothesis is not rejected.

As the previous test, if the null hypothesis is rejected, there is heteroskedasticity. Consequently, when the null hypothesis is no rejected, there is homoskedasticity.

Annex 12

[RA] White test with cross terms

This test is an extension of the White test because it includes the crossing between the explanatory variables.

Given the linear regression function expressed by the Equation 30, the regression of the error term square estimator with $k = 3$ can be written as follow.

$$\hat{u}^2 = \theta + \delta_1 x_1 + \delta_2 x_2 + \gamma_1 x_1^2 + \gamma_2 x_2^2 + \varphi_1 x_1^3 + \varphi_2 x_2^3 + \omega_1 x_1 x_2 + error \quad (36)$$

Therefore, the hypotheses for the White test with cross terms are:

$$\begin{cases} H_0: \delta_1 = \dots = \delta_k = \gamma_1 = \dots = \gamma_k = \varphi_1 = \dots = \varphi_k = 0 \\ H_1: \text{at least one different from 0} \end{cases} \quad (37)$$

The respective statistic is also represented by the Equation 33.

Given the $p - value$, the decision rule is:

- (i) $p - value < \alpha$: the null hypothesis is rejected;
- (ii) $p - value > \alpha$: the null hypothesis is not rejected.

As the previous test, if the null hypothesis is rejected, there is heteroskedasticity. Consequently, when the null hypothesis is no rejected, there is homoskedasticity.

Annex 13

[RA] Generalised Least Squares method

According to the correction of the heteroskedascity problems, the GLS estimator uses a function/weight to correct the coefficients and standard errors that were given wrongly by OLS. This function is called h and use the data to estimate the unknown parameters in this model. This results in an estimate of each h_i , denoted as \hat{h}_i . Using \hat{h}_i instead of h_i in the GLS transformation yields an estimator called the feasible GLS estimator. This is commonly called estimated GLS.

Assuming that,

$$\text{Var}(u/\mathbf{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k) v,$$

Where v has mean equal to unity, conditional on $\mathbf{x} = (x_1, x_2, \dots, x_k)$ and δ_j are unknown variables. The exponential is used to correct heteroskedascity where the variances must be positive to perform WLS. Since the value for δ_j is not realistic even it is known. Thus, to estimate this the equation will be transformed into a linear form that, with a slight modification, can be estimated by OLS.

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k) v,$$

If it is assumed that v is actual independent of \mathbf{x} , this can be written as

$$\log(u^2) = a_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + e,$$

Where e has a zero mean and is independent of \mathbf{x} . An important note is that the intercept is different from δ_0 , but this is not limitative to implement the GLS. Since the dependent variable is log of the squared errors, and satisfies Gauss-Markov assumptions, the unbiased estimators of the δ_j can be obtained by OLS.

Then, a feasible GLS procedure to correct heteroskedascity issue from OLS is:

- (i) Run the regression of y on x_1, x_2, \dots, x_k and obtain the residuals, \hat{u} ;
- (ii) Create the $\log(\hat{u}^2)$ by first squaring the OLS residuals and then taking the natural log;
- (iii) Run the regression $\log(\hat{u}^2)$ on x_1, x_2, \dots, x_k and obtain the fitted values, \hat{g} ;
- (iv) Exponentiation of the fitted values from the previous regression: $\hat{h} = \exp(\hat{g})$;
- (v) Estimate the equation $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$ by WLS, using weights $1/\text{sqrt}(\hat{h})$.

Annex 14

[QCA] Definition of the property space

Best-fit configurations	Cases		
	Absolute frequency	Relative frequency	Cumulative frequency
~BIL*~fsBVPS*~fsEPS*~fsLE*~fsROA*~fsSZ*~fsCFOA	32	13.91%	13.91%
BIL*~fsBVPS*~fsEPS*~fsLE*~fsROA*~fsSZ*~fsCFOA	17	7.39%	21.30%
~BIL*fsBVPS*fsEPS*fsLE*fsROA*fsSZ*fsCFOA	16	6.96%	28.26%
BIL*fsBVPS*fsEPS*~fsLE*fsROA*fsSZ*fsCFOA	15	6.52%	34.78%
~BIL*fsBVPS*fsEPS*~fsLE*~fsROA*~fsSZ*~fsCFOA	14	6.09%	40.87%
~BIL*fsBVPS*fsEPS*~fsLE*fsROA*fsSZ*fsCFOA	11	4.78%	45.65%
BIL*fsBVPS*fsEPS*fsLE*fsROA*fsSZ*fsCFOA	8	3.48%	49.13%
~BIL*~fsBVPS*~fsEPS*fsLE*~fsROA*fsSZ*~fsCFOA	7	3.04%	52.17%
~BIL*fsBVPS*~fsEPS*fsLE*~fsROA*fsSZ*~fsCFOA	5	2.17%	54.35%
~BIL*fsBVPS*fsEPS*~fsLE*fsROA*~fsSZ*fsCFOA	5	2.17%	56.52%
~BIL*fsBVPS*~fsEPS*~fsLE*~fsROA*~fsSZ*~fsCFOA	4	1.74%	58.26%
BIL*fsBVPS*fsEPS*~fsLE*~fsROA*~fsSZ*~fsCFOA	4	1.74%	60.00%
~BIL*~fsBVPS*~fsEPS*~fsLE*fsROA*~fsSZ*~fsCFOA	4	1.74%	61.74%
~BIL*fsBVPS*fsEPS*fsLE*~fsROA*fsSZ*~fsCFOA	4	1.74%	63.48%
BIL*fsBVPS*fsEPS*~fsLE*fsROA*~fsSZ*fsCFOA	4	1.74%	65.22%
~BIL*~fsBVPS*~fsEPS*fsLE*fsROA*fsSZ*fsCFOA	4	1.74%	66.96%
BIL*~fsBVPS*~fsEPS*fsLE*fsROA*fsSZ*fsCFOA	4	1.74%	68.70%
~BIL*~fsBVPS*fsEPS*fsLE*fsROA*fsSZ*fsCFOA	4	1.74%	70.43%
BIL*~fsBVPS*fsEPS*~fsLE*~fsROA*~fsSZ*~fsCFOA	3	1.30%	71.74%
~BIL*~fsBVPS*~fsEPS*fsLE*~fsROA*~fsSZ*~fsCFOA	3	1.30%	73.04%
~BIL*fsBVPS*fsEPS*~fsLE*fsROA*fsSZ*~fsCFOA	3	1.30%	74.35%
~BIL*~fsBVPS*~fsEPS*fsLE*~fsROA*~fsSZ*fsCFOA	3	1.30%	75.65%
~BIL*~fsBVPS*~fsEPS*~fsLE*fsROA*~fsSZ*fsCFOA	3	1.30%	76.96%
~BIL*fsBVPS*~fsEPS*~fsLE*~fsROA*fsSZ*fsCFOA	3	1.30%	78.26%
~BIL*~fsBVPS*fsEPS*~fsLE*~fsROA*~fsSZ*~fsCFOA	2	0.87%	79.13%
~BIL*~fsBVPS*fsEPS*~fsLE*fsROA*~fsSZ*~fsCFOA	2	0.87%	80.00%
~BIL*fsBVPS*fsEPS*fsLE*fsROA*~fsSZ*~fsCFOA	2	0.87%	80.87%
BIL*fsBVPS*~fsEPS*fsLE*~fsROA*fsSZ*~fsCFOA	2	0.87%	81.74%
~BIL*~fsBVPS*~fsEPS*fsLE*fsROA*fsSZ*~fsCFOA	2	0.87%	82.61%
~BIL*fsBVPS*fsEPS*fsLE*fsROA*fsSZ*~fsCFOA	2	0.87%	83.48%
BIL*fsBVPS*fsEPS*fsLE*fsROA*fsSZ*~fsCFOA	2	0.87%	84.35%
~BIL*~fsBVPS*~fsEPS*~fsLE*~fsROA*~fsSZ*fsCFOA	2	0.87%	85.22%
BIL*~fsBVPS*~fsEPS*~fsLE*fsROA*~fsSZ*fsCFOA	2	0.87%	86.09%
~BIL*~fsBVPS*fsEPS*~fsLE*fsROA*~fsSZ*fsCFOA	2	0.87%	86.96%
~BIL*~fsBVPS*~fsEPS*fsLE*~fsROA*fsSZ*fsCFOA	2	0.87%	87.83%
BIL*~fsBVPS*~fsEPS*fsLE*~fsROA*fsSZ*fsCFOA	2	0.87%	88.70%
~BIL*~fsBVPS*~fsEPS*~fsLE*fsROA*fsSZ*fsCFOA	2	0.87%	89.57%
BIL*~fsBVPS*~fsEPS*~fsLE*fsROA*fsSZ*fsCFOA	2	0.87%	90.43%
~BIL*fsBVPS*~fsEPS*~fsLE*fsROA*fsSZ*fsCFOA	2	0.87%	91.30%
~BIL*fsBVPS*~fsEPS*fsLE*~fsROA*~fsSZ*~fsCFOA	1	0.43%	91.74%
BIL*~fsBVPS*~fsEPS*~fsLE*fsROA*~fsSZ*~fsCFOA	1	0.43%	92.17%
BIL*~fsBVPS*fsEPS*~fsLE*fsROA*~fsSZ*~fsCFOA	1	0.43%	92.61%
BIL*fsBVPS*fsEPS*~fsLE*fsROA*~fsSZ*~fsCFOA	1	0.43%	93.04%
~BIL*fsBVPS*fsEPS*~fsLE*~fsROA*fsSZ*~fsCFOA	1	0.43%	93.48%
~BIL*~fsBVPS*~fsEPS*~fsLE*fsROA*fsSZ*~fsCFOA	1	0.43%	93.91%
~BIL*fsBVPS*~fsEPS*~fsLE*fsROA*fsSZ*~fsCFOA	1	0.43%	94.35%
BIL*~fsBVPS*~fsEPS*fsLE*fsROA*fsSZ*~fsCFOA	1	0.43%	94.78%
~BIL*fsBVPS*~fsEPS*fsLE*fsROA*fsSZ*~fsCFOA	1	0.43%	95.22%
BIL*~fsBVPS*~fsEPS*~fsLE*~fsROA*~fsSZ*fsCFOA	1	0.43%	95.65%
~BIL*~fsBVPS*~fsEPS*fsLE*fsROA*~fsSZ*fsCFOA	1	0.43%	96.09%

Best-fit configurations	Cases		
	Absolute frequency	Relative frequency	Cumulative frequency
BIL*~fsBVPS*fsEPS*fsLE*fsROA*~fsSZ*fsCFOA	1	0.43%	96.52%
BIL*fsBVPS*fsEPS*fsLE*fsROA*~fsSZ*fsCFOA	1	0.43%	96.96%
~BIL*~fsBVPS*~fsEPS*~fsLE*~fsROA*fsSZ*fsCFOA	1	0.43%	97.39%
~BIL*fsBVPS*~fsEPS*fsLE*~fsROA*fsSZ*fsCFOA	1	0.43%	97.83%
BIL*fsBVPS*~fsEPS*fsLE*~fsROA*fsSZ*fsCFOA	1	0.43%	98.26%
~BIL*fsBVPS*fsEPS*fsLE*~fsROA*fsSZ*fsCFOA	1	0.43%	98.70%
BIL*fsBVPS*~fsEPS*~fsLE*fsROA*fsSZ*fsCFOA	1	0.43%	99.13%
BIL*~fsBVPS*fsEPS*~fsLE*fsROA*fsSZ*fsCFOA	1	0.43%	99.57%
BIL*~fsBVPS*fsEPS*fsLE*fsROA*fsSZ*fsCFOA	1	0.43%	100.00%
TOTAL	230	100.00%	-

Annex 15

[QCA] Calibration

5%, 50% and 95% percentiles of the variables

Regressor	5%	50%	95%
PPS	0.21	14.75	162.14
BVPS	0.08	5.33	55.08
EPS	-0.61	0.74	10.08
LE	0.01	0.47	0.86
ROA	-0.02	0.02	0.16
SZ	0.30	14.30	17.46
CFOA	0.00	0.05	0.20

Calibration process

```
compute: fsPPS = calibrate(PPS,162.14,14.75,0.21)
```

```
compute: fsBVPS = calibrate(BVPS,55.08,5.33,0.08)
```

```
compute: fsEPS = calibrate(EPS,10.08,0.74,-0.61)
```

```
compute: fsLE = calibrate(LE,0.86,0.47,0.01)
```

```
compute: fsROA = calibrate(ROA,0.16,0.02,-0.02)
```

```
compute: fsSZ = calibrate(SZ,17.46,14.30,0.30)
```

```
compute: fsCFOA = calibrate(CFOA,0.20,0.05,0.00)
```

Annex 16

[QCA] Descriptive statistics of antecedent conditions and outcome

Variable	Mean	Std. Dev.	Minimum	Maximum	N Cases	Missing
BIL	0.3358491	0.4722864	0	1	265	0
fsBVPS	0.4409811	0.31516	0	1	265	0
fsEPS	0.4508679	0.2740947	0	1	265	0
fsLE	0.4108302	0.3224565	0.05	1	265	0
fsROA	0.4576981	0.2691503	0	1	265	0
fsSZ	0.4906415	0.3387424	0.05	0.99	265	0
fsCFOA	0.4271321	0.312835	0	1	265	0
fsPPS	0.4066792	0.3029869	0.05	1	265	0

Annex 17

[QCA] Truth table solution

```
*****
*TRUTH TABLE ANALYSIS*
*****
```

```
File: /Users/fapl/Documents/Cursos académicos/MSc. Finance/2016:2017/
Thesis/03 Data/01 Base dados 2016/03 Modelling fsQCA/02 Calibration.csv
Model: fsPPS = f(fsBVPS, fsEPS, fsLE, fsROA, fsSZ, fsCFOA, BIL)
Algorithm: Quine-McCluskey
```

```
--- TRUTH TABLE SOLUTION ---
frequency cutoff: 1
consistency cutoff: 0.802952
Assumptions:
```

	raw coverage	unique coverage	consistency
	-----	-----	-----
fsBVPS	0.88958	0.179085	0.820383
fsEPS*fsCFOA	0.650737	0.0197644	0.869128
fsEPS*BIL	0.37413	0.0236617	0.860987
~fsLE*fsSZ*~fsCFOA	0.342767	0.000556707	0.79974
fsROA*~fsSZ*fsCFOA*BIL	0.145681	0.00139183	0.839572
solution coverage:	0.958152		
solution consistency:	0.757371		

Annex 18

[RA] Output of the initial model

Dependent Variable: PPS
 Method: Least Squares
 Date: 10/16/19 Time: 00:25
 Sample: 1 265
 Included observations: 265

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	0.702322	0.086091	8.157920	0.0000
EPS	8.843228	1.077460	8.207475	0.0000
EPS^2	-0.118844	0.013531	-8.783313	0.0000
BIL*BVPS	1.624796	0.175775	9.243633	0.0000
BIL*CFOA	598.2118	105.4664	5.672061	0.0000
BIL*SZ	-4.592660	0.952341	-4.822494	0.0000
C	5.300183	4.143484	1.279161	0.2020
R-squared	0.734334	Mean dependent var		43.44046
Adjusted R-squared	0.728156	S.D. dependent var		105.4956
S.E. of regression	55.00401	Akaike info criterion		10.87875
Sum squared resid	780563.9	Schwarz criterion		10.97331
Log likelihood	-1434.434	Hannan-Quinn criter.		10.91674
F-statistic	118.8575	Durbin-Watson stat		1.938506
Prob(F-statistic)	0.000000			

Annex 19

[RA] Output of Chow test to the initial model

Chow Breakpoint Test: 177

Null Hypothesis: No breaks at specified breakpoints

Varying regressors: BVPS EPS EPS^2 C

Equation Sample: 1 265

F-statistic	3.622776	Prob. F(4,254)	0.0068
Log likelihood ratio	14.70313	Prob. Chi-Square(4)	0.0054
Wald Statistic	14.49110	Prob. Chi-Square(4)	0.0059

Annex 20

[RA] Output of RESET test to the initial model

Ramsey RESET Test

Equation: EQ19_01

Specification: PPS BVPS EPS EPS^2 BIL*BVPS BIL*CFOA BIL*SZ C

Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	3.973225	257	0.0001
F-statistic	15.78651	(1, 257)	0.0001
Likelihood ratio	15.79755	1	0.0001

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	45172.26	1	45172.26
Restricted SSR	780563.9	258	3025.441
Unrestricted SSR	735391.6	257	2861.446
Unrestricted SSR	735391.6	257	2861.446

LR test summary:

	Value	df
Restricted LogL	-1434.434	258
Unrestricted LogL	-1426.536	257

Unrestricted Test Equation:

Dependent Variable: PPS

Method: Least Squares

Date: 10/16/19 Time: 19:26

Sample: 1 265

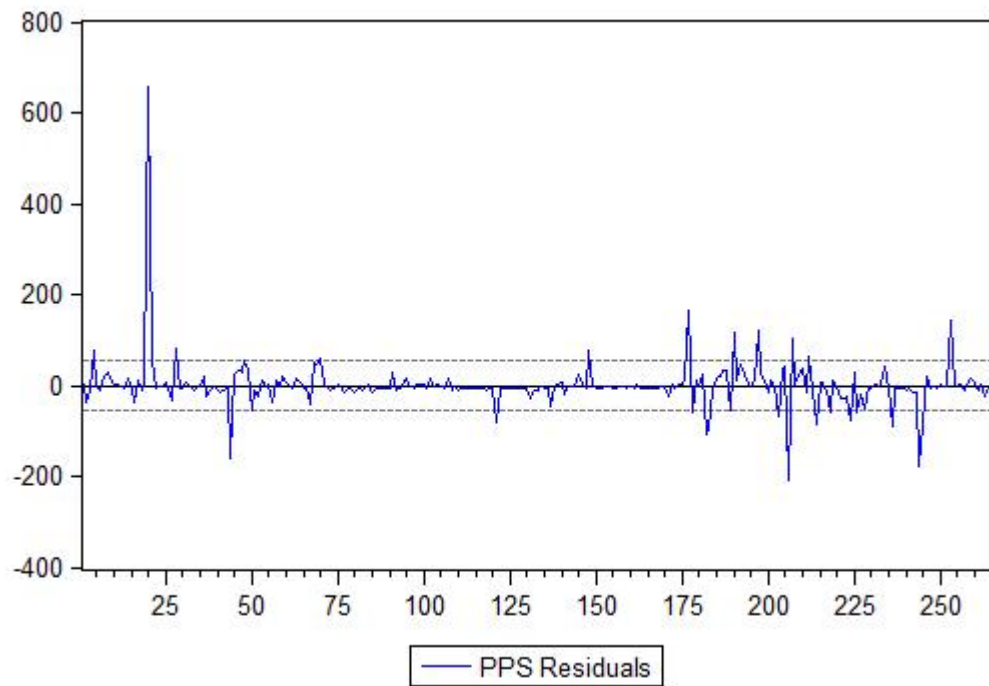
Included observations: 265

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	-0.040116	0.204760	-0.195919	0.8448
EPS	10.73221	1.150663	9.326983	0.0000
EPS^2	-0.138582	0.014065	-9.852742	0.0000
BIL*BVPS	0.742448	0.280247	2.649257	0.0086
BIL*CFOA	412.3599	112.7308	3.657918	0.0003
BIL*SZ	-2.713822	1.039905	-2.609683	0.0096
C	9.533281	4.168085	2.287209	0.0230
FITTED^2	0.001146	0.000288	3.973225	0.0001

R-squared	0.749709	Mean dependent var	43.44046
Adjusted R-squared	0.742891	S.D. dependent var	105.4956
S.E. of regression	53.49249	Akaike info criterion	10.82668
Sum squared resid	735391.6	Schwarz criterion	10.93475
Log likelihood	-1426.536	Hannan-Quinn criter.	10.87010
F-statistic	109.9719	Durbin-Watson stat	1.907759
Prob(F-statistic)	0.000000		

Annex 21

[RA] Residuals graph of the initial model



Annex 22

[RA] Output of the initial model after the exclusion of the residual outlier

Dependent Variable: PPS
 Method: Least Squares
 Date: 10/19/19 Time: 00:50
 Sample: 1 19 21 265
 Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	0.847057	0.054562	15.52472	0.0000
EPS	4.872345	0.705522	6.906018	0.0000
EPS^2	-0.087823	0.008640	-10.16418	0.0000
BIL*BVPS	1.638291	0.110411	14.83811	0.0000
BIL*CFOA	639.5194	66.27880	9.648927	0.0000
BIL*SZ	-4.363394	0.598303	-7.292954	0.0000
C	5.805022	2.602761	2.230332	0.0266
R-squared	0.867160	Mean dependent var		40.44246
Adjusted R-squared	0.864058	S.D. dependent var		93.70576
S.E. of regression	34.54955	Akaike info criterion		9.948823
Sum squared resid	306773.5	Schwarz criterion		10.04364
Log likelihood	-1306.245	Hannan-Quinn criter.		9.986923
F-statistic	279.6093	Durbin-Watson stat		2.129171
Prob(F-statistic)	0.000000			

Annex 23

[RA] Output of RESET test to the initial model after the exclusion of the residual outlier

Ramsey RESET Test
 Equation: UNTITLED
 Specification: PPS BVPS EPS EPS^2 BIL*BVPS BIL*CFOA BIL*SZ C
 Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	2.423502	256	0.0161
F-statistic	5.873362	(1, 256)	0.0161
Likelihood ratio	5.988468	1	0.0144

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	6880.394	1	6880.394
Restricted SSR	306773.5	257	1193.671
Unrestricted SSR	299893.1	256	1171.457
Unrestricted SSR	299893.1	256	1171.457

LR test summary:

	Value	df
Restricted LogL	-1306.245	257
Unrestricted LogL	-1303.250	256

Unrestricted Test Equation:
 Dependent Variable: PPS
 Method: Least Squares
 Date: 10/19/19 Time: 14:41
 Sample: 1 19 21 265
 Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	0.543745	0.136328	3.988521	0.0001
EPS	6.167575	0.879846	7.009831	0.0000
EPS^2	-0.101240	0.010194	-9.931261	0.0000
BIL*BVPS	1.280625	0.183696	6.971430	0.0000
BIL*CFOA	563.2611	72.80967	7.736076	0.0000
BIL*SZ	-3.609230	0.669434	-5.391463	0.0000
C	7.268899	2.648236	2.744808	0.0065
FITTED^2	0.000444	0.000183	2.423502	0.0161

R-squared	0.870139	Mean dependent var	40.44246
Adjusted R-squared	0.866588	S.D. dependent var	93.70576
S.E. of regression	34.22656	Akaike info criterion	9.933715
Sum squared resid	299893.1	Schwarz criterion	10.04208
Log likelihood	-1303.250	Hannan-Quinn criter.	9.977259
F-statistic	245.0488	Durbin-Watson stat	2.096878
Prob(F-statistic)	0.000000		

Annex 24

[RA] Matrix of correlations output

	BVPS	EPS	EPS^2	BIL*BVPS	BIL*CFOA	BIL*SZ
BVPS	-					
EPS	0.795009	-				
EPS^2	0.715266	0.915564	-			
BIL*BVPS	0.548930	0.720786	0.813079	-		
BIL*CFOA	0.070552	0.132015	0.075554	0.264358	-	
BIL*SZ	0.098065	0.136877	0.087571	0.341068	0.808717	-

Annex 25

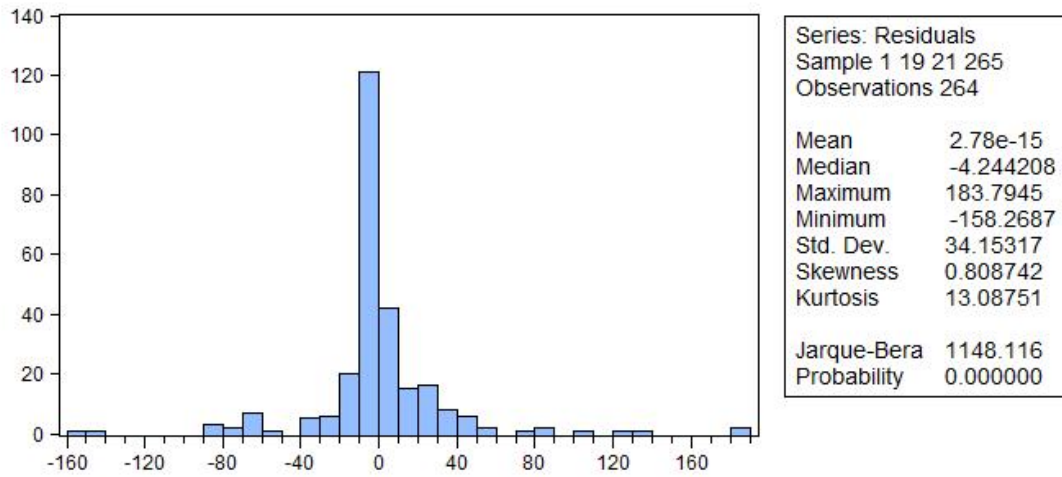
[RA] Computation of the variance inflation factors

Variance Inflation Factors
 Date: 10/18/19 Time: 22:44
 Sample: 1 19 21 265
 Included observations: 264

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
BVPS	0.002977	3.032880	2.791455
EPS	0.497761	10.04827	9.258252
EPS^2	7.47E-05	10.74214	10.60458
BIL*BVPS	0.012191	4.081324	3.916372
BIL*CFOA	4392.880	3.595141	2.927567
BIL*SZ	0.357966	4.376244	3.285338
C	6.774364	1.498262	NA

Annex 26

[RA] Output of Jarque-Bera test to the initial model after the exclusion of the residual outlier



Annex 27

[RA] Output of Breusch-Godfrey Serial Correlation LM test to the initial model after the exclusion of the residual outlier

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.840869	Prob. F(2,255)	0.4325
Obs*R-squared	1.729687	Prob. Chi-Square(2)	0.4211

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 10/19/19 Time: 14:42

Sample: 1 19 21 265

Included observations: 264

Presample and interior missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	-0.000956	0.054666	-0.017494	0.9861
EPS	-0.120403	0.712260	-0.169044	0.8659
EPS^2	0.002221	0.008825	0.251702	0.8015
BIL*BVPS	-0.007918	0.111289	-0.071149	0.9433
BIL*CFOA	9.149517	66.75588	0.137059	0.8911
BIL*SZ	-0.044313	0.599689	-0.073893	0.9412
C	0.123723	2.606506	0.047467	0.9622
RESID(-1)	-0.074453	0.064752	-1.149831	0.2513
RESID(-2)	0.030147	0.063587	0.474112	0.6358
R-squared	0.006552	Mean dependent var		2.78E-15
Adjusted R-squared	-0.024615	S.D. dependent var		34.15317
S.E. of regression	34.57096	Akaike info criterion		9.957401
Sum squared resid	304763.6	Schwarz criterion		10.07931
Log likelihood	-1305.377	Hannan-Quinn criter.		10.00639
F-statistic	0.210217	Durbin-Watson stat		1.990028
Prob(F-statistic)	0.988959			

Annex 28

[RA] Output of Breusch-Pagan-Godfrey test to the initial model after the exclusion of the residual outlier

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	33.67041	Prob. F(6,257)	0.0000
Obs*R-squared	116.1902	Prob. Chi-Square(6)	0.0000
Scaled explained SS	665.4800	Prob. Chi-Square(6)	0.0000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/19/19 Time: 00:19

Sample: 1 19 21 265

Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-129.6963	230.8125	-0.561912	0.5747
BVPS	-11.39145	4.838536	-2.354318	0.0193
EPS	592.9518	62.56556	9.477288	0.0000
EPS^2	-7.911181	0.766230	-10.32482	0.0000
BIL*BVPS	61.73969	9.791231	6.305611	0.0000
BIL*CFOA	34653.01	5877.596	5.895780	0.0000
BIL*SZ	-201.4421	53.05740	-3.796682	0.0002
R-squared	0.440115	Mean dependent var		1162.021
Adjusted R-squared	0.427043	S.D. dependent var		4047.683
S.E. of regression	3063.849	Akaike info criterion		18.91889
Sum squared resid	2.41E+09	Schwarz criterion		19.01371
Log likelihood	-2490.293	Hannan-Quinn criter.		18.95699
F-statistic	33.67041	Durbin-Watson stat		2.009034
Prob(F-statistic)	0.000000			

Annex 29

[RA] Output of White test to the initial model after the exclusion of the residual outlier

Heteroskedasticity Test: White

F-statistic	73.06259	Prob. F(6,257)	0.0000
Obs*R-squared	166.4297	Prob. Chi-Square(6)	0.0000
Scaled explained SS	953.2268	Prob. Chi-Square(6)	0.0000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/19/19 Time: 00:19

Sample: 1 19 21 265

Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	181.8290	176.0992	1.032537	0.3028
BVPS^2	-0.058933	0.005481	-10.75131	0.0000
EPS^2	30.56584	2.049668	14.91258	0.0000
(EPS^2)^2	-0.005505	0.000366	-15.04945	0.0000
(BIL*BVPS)^2	1.196715	0.082455	14.51352	0.0000
(BIL*CFOA)^2	219342.7	17904.23	12.25089	0.0000
(BIL*SZ)^2	-10.49602	1.928778	-5.441796	0.0000
R-squared	0.630416	Mean dependent var	1162.021	
Adjusted R-squared	0.621787	S.D. dependent var	4047.683	
S.E. of regression	2489.285	Akaike info criterion	18.50354	
Sum squared resid	1.59E+09	Schwarz criterion	18.59835	
Log likelihood	-2435.467	Hannan-Quinn criter.	18.54164	
F-statistic	73.06259	Durbin-Watson stat	2.254713	
Prob(F-statistic)	0.000000			

Annex 30

[RA] Output of White test with cross terms to the initial model after the exclusion of the residual outlier

Heteroskedasticity Test: White

F-statistic	85.45138	Prob. F(23,240)	0.0000
Obs*R-squared	235.2703	Prob. Chi-Square(23)	0.0000
Scaled explained SS	1347.511	Prob. Chi-Square(23)	0.0000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/19/19 Time: 14:43

Sample: 1 19 21 265

Included observations: 264

Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-129.6117	149.2200	-0.868595	0.3859
BVPS	50.85233	17.65518	2.880306	0.0043
BVPS^2	-0.787819	0.335680	-2.346935	0.0197
BVPS*EPS	0.359983	3.058799	0.117688	0.9064
BVPS*(EPS^2)	0.231657	0.108433	2.136404	0.0337
BVPS*(BIL*BVPS)	3.484021	0.763131	4.565432	0.0000
BVPS*(BIL*CFOA)	-3446.488	642.5335	-5.363904	0.0000
BVPS*(BIL*SZ)	-12.63968	5.065288	-2.495354	0.0133
EPS	-85.28544	63.46711	-1.343774	0.1803
EPS^2	19.46381	9.072054	2.145469	0.0329
EPS*(EPS^2)	-0.499825	0.498790	-1.002075	0.3173
EPS*(BIL*BVPS)	-13.84500	6.947583	-1.992779	0.0474
EPS*(BIL*CFOA)	22216.02	3882.916	5.721477	0.0000
EPS*(BIL*SZ)	-52.32276	27.47652	-1.904272	0.0581
(EPS^2)^2	-0.004195	0.005151	-0.814254	0.4163
(EPS^2)*(BIL*BVPS)	-0.131475	0.064707	-2.031836	0.0433
(EPS^2)*(BIL*CFOA)	-378.1843	125.7930	-3.006403	0.0029
(EPS^2)*(BIL*SZ)	3.974009	0.874852	4.542491	0.0000
BIL*BVPS	239.6843	36.81237	6.510971	0.0000
BIL*CFOA	32800.67	26022.25	1.260485	0.2087
(BIL*CFOA)^2	150782.2	51751.05	2.913607	0.0039
(BIL*CFOA)*(BIL*SZ)	-3629.021	1874.188	-1.936316	0.0540
BIL*SZ	-344.1654	198.5132	-1.733716	0.0843
(BIL*SZ)^2	28.67467	13.31485	2.153585	0.0323
R-squared	0.891175	Mean dependent var	1162.021	
Adjusted R-squared	0.880746	S.D. dependent var	4047.683	
S.E. of regression	1397.792	Akaike info criterion	17.40968	
Sum squared resid	4.69E+08	Schwarz criterion	17.73477	
Log likelihood	-2274.078	Hannan-Quinn criter.	17.54031	
F-statistic	85.45138	Durbin-Watson stat	2.147795	
Prob(F-statistic)	0.000000			

Annex 31

[RA] Output of the GLS estimation

Dependent Variable: LOG(RESID^2)
 Method: Least Squares
 Date: 10/19/19 Time: 16:44
 Sample: 1 19 21 265
 Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	-0.001599	0.003728	-0.429079	0.6682
EPS	0.268062	0.048201	5.561291	0.0000
EPS^2	-0.003306	0.000590	-5.600333	0.0000
BIL*BVPS	0.010390	0.007543	1.377435	0.1696
BIL*CFOA	10.84860	4.528189	2.395792	0.0173
BIL*SZ	0.030336	0.040876	0.742146	0.4587
C	3.385210	0.177821	19.03713	0.0000
R-squared	0.254530	Mean dependent var	4.249185	
Adjusted R-squared	0.237126	S.D. dependent var	2.702501	
S.E. of regression	2.360436	Akaike info criterion	4.581727	
Sum squared resid	1431.916	Schwarz criterion	4.676544	
Log likelihood	-597.7880	Hannan-Quinn criter.	4.619827	
F-statistic	14.62485	Durbin-Watson stat	1.643478	
Prob(F-statistic)	0.000000			

Dependent Variable: PPS
 Method: Least Squares
 Date: 10/19/19 Time: 16:46
 Sample: 1 19 21 265
 Included observations: 264
 Weighting series: 1/@SQRT(EXP(LOG(RESID^2)F))
 Weight type: Inverse standard deviation (EViews default scaling)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	1.139022	0.066147	17.21967	0.0000
EPS	0.350146	0.305238	1.147125	0.2524
EPS^2	-0.033881	0.005462	-6.202543	0.0000
BIL*BVPS	1.109820	0.089285	12.43008	0.0000
BIL*CFOA	135.1618	49.30662	2.741250	0.0066
BIL*SZ	-0.782742	0.302418	-2.588280	0.0102
C	5.942067	1.047178	5.674360	0.0000

Weighted Statistics

R-squared	0.996508	Mean dependent var	35.89919
Adjusted R-squared	0.996427	S.D. dependent var	269.8839
S.E. of regression	15.50465	Akaike info criterion	8.346315
Sum squared resid	61781.32	Schwarz criterion	8.441132
Log likelihood	-1094.714	Hannan-Quinn criter.	8.384415
F-statistic	12224.56	Durbin-Watson stat	1.897980
Prob(F-statistic)	0.000000	Weighted mean dep.	72.78236

Unweighted Statistics

R-squared	0.789950	Mean dependent var	40.44246
Adjusted R-squared	0.785047	S.D. dependent var	93.70576
S.E. of regression	43.44487	Sum squared resid	485076.4
Durbin-Watson stat	1.998919		

Annex 32

[RA] Output of the final model

Dependent Variable: PPS
 Method: Least Squares
 Date: 10/19/19 Time: 16:47
 Sample: 1 19 21 265
 Included observations: 264
 Weighting series: 1/@SQRT(EXP(LOG(RESID^2)F))
 Weight type: Inverse standard deviation (EViews default scaling)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	1.168503	0.060987	19.15977	0.0000
EPS^2	-0.030817	0.004768	-6.463392	0.0000
BIL*BVPS	1.083480	0.086335	12.54975	0.0000
BIL*CFOA	140.4552	49.12026	2.859415	0.0046
BIL*SZ	-0.761170	0.302017	-2.520285	0.0123
C	5.535671	0.986033	5.614081	0.0000

Weighted Statistics

R-squared	0.996490	Mean dependent var	35.89919
Adjusted R-squared	0.996422	S.D. dependent var	269.8839
S.E. of regression	15.51414	Akaike info criterion	8.343846
Sum squared resid	62097.66	Schwarz criterion	8.425118
Log likelihood	-1095.388	Hannan-Quinn criter.	8.376503
F-statistic	14651.27	Durbin-Watson stat	1.905291
Prob(F-statistic)	0.000000	Weighted mean dep.	72.78236

Unweighted Statistics

R-squared	0.784193	Mean dependent var	40.44246
Adjusted R-squared	0.780010	S.D. dependent var	93.70576
S.E. of regression	43.95086	Sum squared resid	498372.8
Durbin-Watson stat	1.991915		

Annex 33

[RA] Output of Breusch-Pagan-Godfrey test to the final model

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	4.962176	Prob. F(5,258)	0.0002
Obs*R-squared	23.16061	Prob. Chi-Square(5)	0.0003
Scaled explained SS	85.57269	Prob. Chi-Square(5)	0.0000

Test Equation:

Dependent Variable: WGT_RESID^2

Method: Least Squares

Date: 10/19/19 Time: 16:48

Sample: 1 19 21 265

Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	173.5281	52.36366	3.313903	0.0011
BVPS*WGT	11.44724	2.674774	4.279703	0.0000
EPS^2*WGT	-0.466683	0.198020	-2.356743	0.0192
BIL*BVPS*WGT	-1.520374	3.548414	-0.428466	0.6687
BIL*CFOA*WGT	-1037.952	2079.465	-0.499144	0.6181
BIL*SZ*WGT	-16.83937	12.61058	-1.335337	0.1829
R-squared	0.087730	Mean dependent var		235.2184
Adjusted R-squared	0.070050	S.D. dependent var		655.5219
S.E. of regression	632.1455	Akaike info criterion		15.75858
Sum squared resid	1.03E+08	Schwarz criterion		15.83985
Log likelihood	-2074.133	Hannan-Quinn criter.		15.79124
F-statistic	4.962176	Durbin-Watson stat		1.931518
Prob(F-statistic)	0.000235			

Annex 34

[RA] Output of White test to the final model

Heteroskedasticity Test: White

F-statistic	1.828878	Prob. F(6,257)	0.0938
Obs*R-squared	10.81057	Prob. Chi-Square(6)	0.0944
Scaled explained SS	39.94235	Prob. Chi-Square(6)	0.0000

Test Equation:

Dependent Variable: WGT_RESID^2

Method: Least Squares

Date: 10/19/19 Time: 16:49

Sample: 1 19 21 265

Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	230.7018	54.55609	4.228709	0.0000
WGT^2	5.506233	26.91048	0.204613	0.8380
BVPS^2*WGT^2	0.051384	0.017693	2.904143	0.0040
(EPS^2)^2*WGT^2	-8.38E-05	5.79E-05	-1.446926	0.1491
(BIL*BVPS)^2*WGT^2	-0.016575	0.022999	-0.720681	0.4718
(BIL*CFOA)^2*WGT^2	-39937.15	49149.24	-0.812569	0.4172
(BIL*SZ)^2*WGT^2	0.206626	1.286995	0.160549	0.8726
R-squared	0.040949	Mean dependent var		235.2184
Adjusted R-squared	0.018559	S.D. dependent var		655.5219
S.E. of regression	649.4105	Akaike info criterion		15.81616
Sum squared resid	1.08E+08	Schwarz criterion		15.91098
Log likelihood	-2080.734	Hannan-Quinn criter.		15.85426
F-statistic	1.828878	Durbin-Watson stat		1.857792
Prob(F-statistic)	0.093821			

Annex 35

[RA] Output of White test with cross terms to the final model

Heteroskedasticity Test: White

F-statistic	2.768673	Prob. F(16,247)	0.0004
Obs*R-squared	40.14735	Prob. Chi-Square(16)	0.0007
Scaled explained SS	148.3345	Prob. Chi-Square(16)	0.0000

Test Equation:

Dependent Variable: WGT_RESID^2

Method: Least Squares

Date: 10/19/19 Time: 16:49

Sample: 1 19 21 265

Included observations: 264

Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	506.8731	97.00964	5.224976	0.0000
WGT^2	-293.9065	72.17620	-4.072069	0.0001
BVPS^2*WGT^2	-0.119403	0.107296	-1.112835	0.2669
BVPS*WGT^2	13.09073	3.371161	3.883152	0.0001
BVPS*(EPS^2)*WGT^2	0.041707	0.033855	1.231949	0.2191
BVPS*(BIL*BVPS)*WGT^2	1.143610	1.314825	0.869781	0.3853
BVPS*(BIL*CFOA)*WGT^2	-1968.984	1480.796	-1.329679	0.1849
BVPS*(BIL*SZ)*WGT^2	3.336643	6.545088	0.509794	0.6107
(EPS^2)^2*WGT^2	-0.002639	0.001255	-2.102204	0.0365
(EPS^2)*WGT^2	6.656706	1.548079	4.299977	0.0000
(EPS^2)*(BIL*BVPS)*WGT^2	-0.038501	0.062835	-0.612729	0.5406
(BIL*BVPS)*WGT^2	5.643204	48.52869	0.116286	0.9075
(BIL*CFOA)^2*WGT^2	-111743.8	57438.51	-1.945450	0.0529
(BIL*CFOA)*WGT^2	7049.302	14336.72	0.491696	0.6234
(BIL*CFOA)*(BIL*SZ)*WGT^2	24.73658	1088.530	0.022725	0.9819
(BIL*SZ)^2*WGT^2	0.174844	6.005919	0.029112	0.9768
(BIL*SZ)*WGT^2	-38.84652	79.67646	-0.487553	0.6263
R-squared	0.152073	Mean dependent var	235.2184	
Adjusted R-squared	0.097147	S.D. dependent var	655.5219	
S.E. of regression	622.8676	Akaike info criterion	15.76877	
Sum squared resid	95827126	Schwarz criterion	15.99904	
Log likelihood	-2064.478	Hannan-Quinn criter.	15.86130	
F-statistic	2.768673	Durbin-Watson stat	1.950074	
Prob(F-statistic)	0.000412			

Annex 36

[RA] Output of Breusch-Godfrey Serial Correlation LM test to the final model

Breusch-Godfrey Serial Correlation LM Test:

Obs*R-squared	0.000000	Prob. Chi-Square(2)	1.0000
---------------	----------	---------------------	--------

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 10/19/19 Time: 16:49

Sample: 1 19 21 265

Included observations: 264

Presample and interior missing value lagged residuals set to zero.

Weight series: 1/@SQRT(EXP(RESIDF))

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BVPS	-0.147802	0.055772	-2.650126	0.0085
EPS^2	0.019765	0.006800	2.906565	0.0040
BIL*BVPS	-0.229605	0.147786	-1.553634	0.1215
BIL*CFOA	-28.36855	57.51175	-0.493265	0.6222
BIL*SZ	0.062401	0.306174	0.203808	0.8387
C	1.506001	0.594378	2.533743	0.0119
RESID(-1)	-0.041372	0.027005	-1.532035	0.1267
RESID(-2)	0.025723	0.027354	0.940364	0.3479

Weighted Statistics

R-squared	-0.136317	Mean dependent var	1.036895
Adjusted R-squared	-0.167389	S.D. dependent var	15.33080
S.E. of regression	16.60227	Akaike info criterion	8.486790
Sum squared resid	70562.65	Schwarz criterion	8.595152
Log likelihood	-1112.256	Hannan-Quinn criter.	8.530334
Durbin-Watson stat	1.874313	Weighted mean dep.	-3.52E-14

Unweighted Statistics

R-squared	-0.104619	Mean dependent var	6.013120
Adjusted R-squared	-0.134823	S.D. dependent var	43.11216
S.E. of regression	45.92657	Sum squared resid	539967.9
Durbin-Watson stat	0.244934		

FROM CLASSICAL REGRESSION ANALYSIS TO QUALITATIVE ANALYSIS:
A SHARE PRICE fsQCA EMPIRICAL APPLICATION

Fábio André Pereira Luís