BOARD DIVERSITY AND THE FINANCIAL PERFORMANCE OF FIRMS

A quantitative examination

Signe Bredsgaard Byrith - 20184734

October 15th, 2024

AALBORG
UNIVERSITY

Aalborg University Business School

4th semester, cand.oecon

Supervisor: Pernille Gjerløv-Juel

ABSTRACT:

The purpose of this study is to examine to what extent diversity in the board of directors correlates with the financial performance of firms. The composition of the board of directors in the modern corporation has become increasingly discussed in combination with it becoming a highly politicized topic. Considering the ongoing debate in relation to good cooperate governance it is fair to say that the literature is characterized by mixed results.

This study is based on a panel data structure containing information about the boards in Danish firms and a series of firm specific characteristics in a time frame spanning from 2017 to 2019. The data is subject to a filtering process to ensure the quality of the data. By utilizing a total of 12 ordinary least squares (OLS) regression models, the study examines the association between the Shannon index, the percentage of women and ethnic minorities in the board of directors and a series of dummy variables indicating an increasing number of either women or ethnic minorities, and the financial performance of firms. To create the theoretical framework for the quantitative examination the resource dependence theory, the human capital theory and the critical mass and token theory is utilized.

The result from the quantitative examination indicates that there are a negative significant association, between the diversity and the firm's financial performance and an in-significant association between an increase in the diversity and the firm's financial performance.

The conclusion of the study is that it seems to be impossible to provide a clear-cut answer to the problem statement.

Title:

Board diversity and the financial performance of firms – a quantitative examination

Tema:

Speciale

Gruppe:

8

Periode:

Efterårssemester 2024

Forfattere:

Signe Bredsgaard Byrith

Studienummer: 20184736

sbyrit18@student.aau.dk

Vejleder:

Pernille Gjerløv-Juel

pgj@business.aau.dk

Normalsider: 52.33

Sider: 132

Afleveringsdato: 15. oktober 2024

1 CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

2 BOARD DIVERSITY AND THE FIRM'S FINANCIAL PERFORMANCE

2.1 INTRODUCTION TO THE PROBLEM

The composition of the modern corporation has become an increasingly more discussed topic in both the political and public sphere. This is especially true when it comes to the question of diversity in the board of directors, which lately has become a highly politicized topic (Hecklen et al., 2022) (Terjesen & Sløk-Madsen, 2023) (Harder, 2023). The advocates of diversity argue that some of the advantages would be a more efficient run board and a broader focus in connection with monitoring. Introducing diversity in the form of (gender, age, educational background, ethnicity, nationality) to the board of directors is anticipated to have a positive impact on innovation, broaden the perspective in problem-solving situations, improve creativity and discourage homogeneous ways of thinking (Ciavarella, 2018). However, this may not come without challenges for the firm. An amplified board diversity is also associated with increased levels of conflict among the members of the board, organization problems and a prolonged decision process due to communication difficulties (Ciavarella, 2018). In other words, the benefits come with a cost.

Even though several studies have proposed diversity as being a core component of good governance among the board of directors, the composition of these boards have remained largely unchanged in the terms of set-up (Gabaldon et al., 2017a)(Gabaldon et al., 2017b)(Rose, 2007). Studies involving mandated quotas for women directors furthermore suggest that the homogeneous nature of top management will not change in a foreseeable future, if no external action is made (Kogut et al., 2014). Consequently, diversity has become a topic of conversation regarding legislative intervention in many countries. The debate mainly focuses on whether it would be best if the companies regulate themselves on a voluntary basis or whether there is a need for government intervention (Kogut et al., 2014).

Norway is one of the frontrunner in implementing legislative measures regarding female membership of the board of directors. In 2005, they introduced a law that entailed that all public-limited companies were required to have a 40% female and male representation in their board rooms. A two-year transition period was allowed and the companies that did not

uphold the standard of the law would be forced to be dissolved by compulsion. When it comes to self-regulation initiatives, the European Commission launched in 2010 a new gender equality strategy, which introduced a new initiative to improve gender equality in the corporate boardroom (European Commission, 2011). The initiative was triggered by the lack of progress in the area in previous years (Deloitte, 2022). In 2012, the Commission launched directives aimed to address the lack of diversity in the boardrooms of listed companies (Deloitte, 2022). The purpose of the proposal was to boost the representation of women in corporate boards across the European Union. It established that a minimum of 40% of the underrepresented gender among non-executives' directors of public listed companies, should become the target, and it mandated that companies with a low level of diversity should implement a transparent and neutral selection process to achieve the objective. The resolution was however denied by the Member States. (European Commission, 2022). The proponents of government intervention argue that it is necessary to create legislative action, for example with the use of quotas, as self-regulation appears to be ineffective (Mensi-Klarbach et al., 2021). Critics, however, claim that self-regulative schemes are more likely to promote a cultural shift in the board of directors rather than a forced change. Furthermore, this approach is claimed to help change the prevailing gender dynamics that currently benefit men in boardroom (Mensi-Klarbach et al. 2021). This is further substantiated by current studies which finds that mandatory quotas carry the risk that the individuals, who by virtue of the law are appointed to the boards, will be regarded as "the token women". Board members included with the help of an quota scheme is therefore in danger of being faced with scrutiny by the existing board members as their presence in the board could be seen as being determined not by merit but legislation. (DuPlessis et al., 2014). Moreover, it has the possibility of fostering a culture of defiance by the affected firms, where they will be more inclined to only follow the letter of the law rather than the spirit (Boyd, 1996).

In the light of the debate regarding diversity of the board of directors, its relation to good practice in corporate governance, and other perspectives, there is a lack of consensus in the literature, which is characterized by inconsistent outcomes. Some of the studies have found a positive relation (Campbell and Minguez-Vera, 2008) (Cook and Glass, 2015) meanwhile others have found either a negative- or non-significant relation (Smith et al., 2006), (Rose et al., 2013). Given the mixed results, there is a need to gain a better understanding of the link between board diversity and firm performance. If there is no significant link between diversity and firm performance, then the issues would be a predominantly public policy

matter. However, if there exists a link between firm performance and board diversity, then it would have economic ramifications for the firm, and as such a matter the board should consider (Carter et al. 2010).

2.1.1 Problem statement

With the above introduction the following problem can be introduced for the project:

How does diversity in the board of directors correlate with the financial performance of firms

2.1.2 Problem area

In the following section, the formulation of the problem will be elaborated on, which aims to narrow down the scope of the problem. Subsequently, the reasoning behind the problem formulations is explained

In the following study, diversity is will be defined as demographic, in terms of women and ethnic minorities. In this context women will be defined as individuals with a koen-index on 0 in the Bef-dataset from "Danmarks Statistiks Forskerservice" database. Moreover, in this context ethnic minorities are defined as individuals with a "ET_TYPE" on either 2 or 3 in the Bef-dataset from the same database. Further restrictions are presented in section 3.

The problem statement will be answered based on a quantitative examination. In the study will there be a focus on the theory presented in the literature review. Based on the theory presented in the literature review a series of hypotheses will be created regarding gender and ethnic minorities and how the board diversity defined as such may have an association with the financial performance of the firms. The hypotheses are subsequently investigated in a quantitative examination based on registry data from "Danmarks Statistiks Forskerservice". The result of the quantitative examination will be considered in a theoretical context, using the before mentioned theory and there will be a discussion on its ability to answer the problem statement.

2.1.3 Structure of the study

Based on the above sections the structure of the study is outlined in the following section. This aims to present the reader with a brief outline of the content of the different chapters of the study.

Chapter 3:

In Chapter 3 the literature review regarding firm performance and board diversity is presented. The chapter is divided into different sections, with each their specific focus within the literature. The first section accounts the relationship between corporate governance and the board of directors. The second section discusses the definition of diversity and the findings in the established studies of board diversity and firm performance. In the third section, some of the different theories that have been utilized in established studies are presented. The fourth section presents some of the methodological considerations in the established studies

Chapter 4:

In chapter 4 there is a presentation of the theories that has been utilized to create the hypotheses that are being examined in the study. This includes the resource dependence theory, the critical mass theory and the human capital theory.

Chapter 5:

In chapter 5 there is a presentation of the hypotheses that are being examined in the quantitative examination. This includes the rationale, which is based on the theoretical background afforded in Chapter 3 and the empirical findings from the existing literature presented in Chapter 2

Chapter 6:

In chapter 6 there is an outline of the econometric methodology. This includes the argumentation for the model choice, the filtering of the cohort and a presentation of the endogeneity problem and its implications.

Chapter 7:

In chapter 7 there is an outline of the research design. This includes the structure of the regression models and a review of the independent, explanatory and control variables.

Chapter 8:

In chapter 8 there is an outline of the descriptive statistics of the variables included in the study. This includes the descriptive statistics of the variables, the transformation of the variables and examination of multicollinearity.

Chapter 9:

In chapter 9, the quantitative examination is performed, including a robustness test.

Chapter 10:

In Chapter 10 the discussion on the generalization of the data, the results of the examination and the causality and endogeneity problem.

Chapter 11:

In Chapter 11 it is concluded how diversity in the board of directors correlates with the financial performance of the firms

3 LITERATURE REVIEW

In the following section the connection between firm performance, corporate governance and the structure of the board is examined with an emphasis on the impact of board diversity. Acknowledging that board diversity has the potential to strengthen board decision-making, this literature review examines how board diversity is defined and its potential impact on firm performance. In the beginning of the review, the connection between corporate governance and the structure of the board is examined. This is followed by a section that outlines different approaches to defining diversity and a summarization of the key findings from the literature. Next, it assesses the most common methodologies used within the field of the study, including the selection of panel- or cross-sectional data, and the choice of proxy for firm performance, such as account- or marked-based measurements. Finally, the theoretical framework that underlines the research regarding board diversity and firm performance is outlined. The literature review aims to highlight the existing research within the field of the study and explain the rationale behind the methodological choices made in this the paper, thereby furthering the discussion on the impact of diversity in the board of directors and firm performance.

3.1 CORPORATE GOVERNANCE AND THE BOARD OF DIRECTORS

In recent years there has been a newfound focus on corporate governance especially when it comes to the issue of the composition of the board of directors and diversity. This has become one of the significant challenges that the modern firm is facing (Carter et al., 2003) (Adams, 2015). The concept of corporate governance refers to the system in which the firms are being controlled and directed (Adams, 2015) (Cadbury, 1992). The board of directors play a crucial role in the corporate governance for the firm. This is due to the fact that economic theory proposes that they act as an important internal control mechanism and make the decision of the overall company strategy (Baysiner and Butler, 1985) (Rose, 2007) (Campbell and Mineguez-Vera, 2007). The connection between the board of directors and firm performance is often explained using the agency framework. Within this theoretical framework, the board of directors acts as a governance mechanism that seeks to control and monitor managers who exhibit opportunistic tendencies and thus do not create value for the shareholders (Rose, 2007) (Carter et al., 2003). For this reason, it is argued that the handling of the agency problem should be a top priority of the board of directors to ensure good governance (Shrader et al., 1997) (Rose et al. 2007) (Campbell and Mineguez-Vera, 2007).

Moreover, given that it is the board of directors that has the responsibility for the governance of the firm it is also argued that the individual members have an influence in the financial performance of the firm (Merendino and Melville, 2019)

The structure of the corporate board is determined by the legislation of the specific country; however, they can be divided into two distinct groups: one-tier structure and two-tier structure. Countries as the United Kingdom and the United States of America employ the one-tier structure. This structure is characterized by having one single board of directors who has the management responsibility of the company. This type of board structure consists of both executive and non-executive members which are elected by the shareholders of the firm. This means that the shareholders have a great influence of whom there has the managerial power in the firm (Adams, 2015) (Jungmann, 2006). Within the one-tier board structure both the executive and non-executive members are entrusted to perform the same tasks. Thus there is no division between the duties of monitoring and strategic planning (Jungmann, 2006). This is in great contrast to the two-tier structure boards. The two-tier structure in used in several European countries such as Germany and Denmark. The structure is characterized by having two separate boards, which have respectively a managerial- and a supervisory role. This type of board structure also consists of both executive and non-executive members; however, the executive members are seated in the managerial board and the non-executive members are seated in the supervisory board. Within this structure it is the responsibility of the members of the supervisory board to monitor and control the actions of the members of the managerial board. This is done to ensure that the interest of the shareholders is upheld and prevent possible conflicts of interest. However, in practice the separation is only theoretical, given that the members of the supervisory board are elected by the members of the management board (Jungmann, 2006).

3.2 DIVERSITY IN BOARDS

In the existing studies of diversity, a distinction is often made between non-observable- (cognitive) and observable (demographic) diversity (Bernile et al., 2016) (Maznevski, 1994) (Milliken and Martins, 1996). When referencing demographic diversity, it involves characteristics that is directly observable, and not possible to change. This includes age, nationality, ethnic background and gender. On the other hand, cognitive diversity is related to more underlying attributes and characteristics that are possible to change. This includes educational background, occupational background, industry experience and organizational

membership (Milliken and Martins, 1996). In the research regarding the link between diversity in the board of directors and firm performance, the focus of the papers is predominantly on how diversity would have an impact on the firm's return on investment and financial performance. However, as mentioned in the above section, there is a lack of consensus regarding the results of the papers. Some researchers find a positive link between diversity and firm performance (Nygaard, 2011) (Tyrefors and Jansson, 2017) while others find a negative performance or no significant link (Adams and Ferreira, 2009) (Rose, 2007) (Comi et al., 2017) (Carter et al., 2010)

Carter et al. (2010) studied the relationship between ethnic- and gender diversity in the board of directors and the firm's financial performance in the United States of America. In the article the researchers focused on assessing if a more diverse board of directors would lead to an improvement in the firm's financial performance. The authors hypothesize that both gender- and ethnic diversity have the potential to have an impact on the firm's financial performance, yet, from a theoretical standpoint, the impact could either be positive or negative. The authors investigate it by using a fixed effect regression. Based on this, they find that gender and ethnic diversity in the board of directors has a positive and significant impact on firm performance, however this relationship depends on which type of measurement that is used as a proxy for firm performance. When using ROA as a proxy they find a positive relationship, however when they use Tobin's Q, they find a non-significant relationship (Carter et al.,2010).

In Erhardt et al. (2003), the authors investigate the relationship between financial performance and the demographic diversity of the firm's board, where the demographic diversity is defined as gender and ethnicity. In the article the author hypothesized that demographic diversity in the board of directors would have a positive impact on organizational performance. They argue that diversity tends to foster an environment that produces more innovation, greater creativity and better decisions. further argue that, given that these results have been found in both group settings and on an individual plan, they may also apply in the settings of a board of directors. Therefore, a heterogeneous board of directors would perform better on these parameters than a homogeneous board of directors. To study this, they make a regression analysis and utilize hierarchical regression. Here the authors find that there is a positive correlation between firm performance in the form of ROA and ROI and the diversity of the boards (Erhardt et al., 2003).

In the paper by Carter et al. (2003) the authors examine the connection between the overall value of fortune 1000 firms and board diversity, where diversity is defined as the percentage of Hispanics, Asians, African Americans and women on the cooperate boards. In the paper they hypnotize that there is no significant connection between the diversity in the board of directors and firm performance measured using the proxy, Tobin's Q. They argue that this is due to the fact that no specific theoretical framework predicts the nature of the relationship. To investigate their hypothesis, they use comparisons of means and regression analysis. In the paper the authors find a positive and statistically significant connection between an increase in the overall diversity of the board of directors and firm valuation (Carter et al., 2003).

In a paper by Rose (2007) the author investigates the relationship between female- and foreign board members and the financial performance of firms in Denmark. In the paper the researcher investigates whether a female board representation has a positive impact on firm's performance measures using Tobin's Q as a proxy. The author argues that economic theory does not sufficiently capture all the facets that apply when trying to predict whether an individual's gender or other diversity measurement may have an impact on a firm's performance. Here it is necessary to also include theories grounded in sociology to hypothesize what would happen. In the paper, Rose hypothesizes that an increased representation of women in the board of directors has a positive impact on the financial performance of a firm. To investigate the hypothesis, the author uses a cross-sectional regression. In the end, Rose finds, that there is a non-significant relationship between women in the board of directors and firm performance. Moreover, the author finds, neither gender nor the fact that an individual is a foreigner impact firm performance (Rose, 2007).

I Gyapong et al. (2016) the authors examine the relationship the gender and ethnic board diversity and firm performance of firms in South Africa. In the paper, the authors investigate whether female and ethnic minority representation in the board of directors has an impact on the firm's performance measured in Tobin's Q, and whether the numerical number of women or ethnic minorities who are included in the board is important for the potential effect on the firm's performance. The researchers argue that to reap the full benefit of board diversity, the number of minorities in the board of directors must reach a specific threshold. If this threshold is not met, the minorities will not be able to utilize their full potential. To investigate this, the authors used panel data and a fixed effect regression model. In the end, Gyapong et al. (2016) find a positive and significant impact of both gender and ethnic

minorities on firm performance. Moreover, they find that the impact of gender diversity increases when there was three or more women present in the board. However, this effect was not found for ethnic minorities where the impact was positive of minority diversity, but the impact was less, if three or more minorities present in the board (Gyapong et al., 2016)

In Liu et al. (2013) the authors examines if gender diversity has an impact on the firm's financial performance in China. In the article the authors investigate the effect of the representation of women in the board of directors in China and if this has an effect on the financial performance of the firms. It is also investigated whether not just the presence of women but also the number of women in the board of directors has an significant impact on the firm's financial performance. In the paper, the authors contend that the number of women in the board of directors may be critical in order to have a potential impact on the firm's financial performance. According to this view, the representation of women needs to reach a specific threshold before it is possible to utilize the potential benefit of the presence of women in the board of directors. In the paper the firm's financial performance is measured using the proxies ROA and ROS. To investigate this the authors, use panel data and utilize a fixed effect regression. Lastly, the authors find that there is a positive and significant impact of gender diversity on firms' financial performance in China. Furthermore, they find that the presence of three or more women in the board of directors has a greater impact on the firm's financial performance then, if there less than two (Liu et al., 2013)

3.3 METHODOLOGY

As it has been reported in the earlier section, there are inconsistent results in the literature surrounding the relationship between firm performance and diversity. In an editorial made by Adams et al. (2015), they assign this tendency to the different methodological choices made through one's investigation, and to the endogeneity of the subject:

"The mixed finding in the literature on the relationship between diversity and firm performance can be attributed to differences across studies measures of performance, methodologies, time horizons, omitted variable biases and other contextual issues… A further reason why it may be hard to come up with clear-cut results may be the endogenous nature of corporate governance in general" (Adams et al., 2015).

In previous studies, scholars have utilized various methods to investigate the relationship between diversity and firm performance. When selecting the most appropriate method it is essential that the method applied is in line with the nature of the data being considered. From

the review of the existing literature, it appears that within the field of study, it is debatable whether one should use cross-sectional- or panel data. On one hand, there is cross-sectional data which consists of information of a series of objects or subject at a specific period. Proponents of using cross-sectional data argue that this type of data is favorable as it avoids issues such as residual serial correlation, which is possible when using data with a time-dimension (Erhardt et al., 2003). Furthermore, the data itself is often more accessible and less troublesome to analyze. This is also one of the reasons why it is often used in studies where the intent is to examine a single point in time. However, one of the greater limitations of this type of data, is its inability to capture any kind of time-depended element in the data (Wooldridge, 2016). As an alternative to cross-sectional data, there is panel data, which also is known as longitudinal data, and consists of both a cross-sectional- and time series-dimension. This means it differs from the cross-sectional dataset by including information from a unit over a period. Proponents of using this type of data argue that it makes it possible to take individual heterogeneity into account. This is, however, only possible if the correct panel date techniques are utilized. Moreover, there are potential risk associated with the data, given the potential for attrition bias and the need for time consistency (Hsiao, 2020) (Gyapong et al., 2016) (Wooldridge, 2016). This indicates that the choice of data largely depends on the nature of the examination and the specific research question.

Within the financial performance literature there is also a debate regarding which proxy for performance is the most appropriate to use when performing financial performance examinations. The debate is mainly concentrated on whether a market-based proxy for financial performance, such as Tobin's Q, or an accounting-based proxy for financial performance, such as ROA[1] or ROI[2], is the best proxy for examining financial performance. On one hand, there is stock market-based measurements which represents a firm's future value on the stock market and is therefore often described as being a long-term financial performance measurement. Given that it is a measurement for the valuation of a firm, it is also often described as a proxy for the ability of a firm to create shareholder return (Rose, 2007) (Gentry and Shen, 2010). The main criticism of this type of measurement is that one of the underlying assumptions of using proxies based on stocks, is market efficiency, which in some circles of economics is heavily criticized. Moreover, some economist argues that stock prices does not provide the full insight into the value of a firm, given that it is affected by the

[1] Return on assets (Supriyadi &Terbuka, 2021),

[2] Return on Investment (Supriyadi &Terbuka, 2021),

information that directors choose to reveal to investors (Gentry and Shen, 2010) (Bromiley, 1990) (Richard et al., 2009). On the other hand, there is accounting-based measurement which is said to represent a firm's past and is therefore often described as being a short-term financial performance measurement. The measurement is calculated using an actuatable source which means that the data reflects what has transpired in the firm, and it is therefore, limited how much the data can tell about the future expectations regarding the firm. Moreover, it is argued that there is a certain sensibility in the data due to the fact that they may be highly influenced by the asset's valuation principle chosen by management. However, this method is still much used in financial performance paper due to the fact that the measurement provided an indication of the firm aggregate profitability (Rose, 2007) (Gentry and Shen, 2010) (Richard et al., 2009). Ultimately, it is essential to consider that within the financial performance literature both types of proxies are commonly acknowledged as acceptable indicators of financial performance. Therefore, both are acceptable to use in financial performance examinations (Gentry and Shen, 2010)

3.4 THEORY

A large part of the literature on diversity and firm performance does not formalize a specific theoretical framework that should be used when attempting to predict the relationship between them. Often when investigating the nature of their relationship, an interdisciplinary approach is used and recommended (Terjesen et al. 2009). This is further supported by Carter et al. (2010) who says that the following:

"No single theory directly predicts the nature of the relationship between board diversity and financial performance but several theories from various fields provide insight into the issue" (Carter et al., 2010)

In Carter et al (2010) they use theories from economic, organizational theory and social psychology to create their hypotheses, and investigate the influence of diversity on firm performance. In the article, they use the resource dependency theory, the human capital theory, the agency theory and social psychological theory. They argue that the resource dependence theory and agency theory can be used to argument how board diversity can be beneficial from an organizational and management point of view. In resource dependency it is argued that a diverse board has the ability to access a broader range of external dependencies, which is necessary for the survival of the firm, and in agency theory, a diverse board has the ability to act as a better monitor of the firm's management due to increased

independence. Moreover, they argue that the human capital theory can be used as a tool to address the role of an individual's soft- and hard skills. This skill has the possibility to have either a positive- or negative influence on a firm's organization, and a different set of skills can be needed in regard to the specific organization, and time in the organization. Lastly, they discuss that the social psychological theory can be used to describe the inner group dynamic of the board (Carter et al, 2010).

In their research about the impact of board diversity on firm performance Bagh et al. (2023) suggest that the relationship between the two is grounded in the management theories; upper echelons theory, management network theory and resource-based theory. Upper echelons theory proposes that the characteristics of the senior management has great impact on the performance of the firm as it is the senior management oversees the organization blueprint. There is in other words, a clear link between the characteristics of the board and the organizational success. In the management network theory, the management can reduce transactions cost for the firm by using their social and political ties. Therefore, the connections of the different members of the board of directors can be used to improve the financial performance of the firm. Lastly resource-based theory suggests that the performance of a firm can be improved by applying an effective utilization of the firms' unique human resources. Therefore, a targeted investment in the firms' human resources can lead to an improvement in their competitive advantage (Bagh et al., 2023)

Low et al. (2015) suggest that a pluralistic approach is necessary when trying to explain the relationship between firm performance and board diversity. They mention that theories within the management theory such as the agency theory, the stakeholder theory, legitimacy theory and resource dependency theory, can be used to describe the relationship. The stakeholder theory describes the relationship between the firm and its outside environment. Within the framework, the board does not only represent the interest of the firm but also those of their outside environment, such as costumers. Therefore, the actions of the board of directors will be judged by them and will trigger actions in return. In line with the theory, the composition of the board of directors should reflect that of others' expectations, as the firm in return will receive benefits. The legitimacy theory suggests that organizations have a need to work within the confines of the social norms in the given society. When doing that, the firm will uphold its legitimacy in the eyes of the stakeholders and thereby keep their support. They do that complying with the laws in the society, but also by implementing symbolic policies that mirror the wishes of the society. Thereby they can uphold the social contract between the

firm and the stakeholders. The authors also argue that behavioral theories can be used. Here they cite stewardship theory, tokenism combined with the critical mass theory and heterogeneous/homogeneous group dynamic theory. The stewardship theory proposes that managers are not motivated by their own goals, but act as a steward for the firm, and therefore are motivated by the goals of the firm. Tokenism is often used in combination with the critical mass theory. It is related to the dynamics in group settings when one, a "token", does not possess the same auxiliary characteristics as most of the group. Giving that the token is often a sole person, they will have little or no influence in the group. The critical mass theory says that the number of "tokens" need to surpass a specific level before they can influence the group. Lastly, they argue that heterogeneous and homogeneous groups function differently in relation to innovation and communication. Heterogeneous groups tend to improve innovation in a firm, due to them having a larger body of knowledge and the ability to have a more comprehensive perspective in the decision situation. However, the same reasons are also said to be the cause of conflict in the group thereby reducing the effectiveness of the communication and making the decision process longer and more rigid. On the other hand, homogenous tends to hamper innovation due to group think and conformity, which in turn result in a less group conflict and quick decision making (Low et al., 2015).

4 THEORY

In the following section theory on the relationship between diversity and firm performance will be presented. The section aims to provide a description of the specific theories employed in the paper. As mentioned in the literature review *"No single theory directly predicts the nature of the relation* (Carter et al., 2010)*,* which highlights the complexity of having no dominate theoretical framework within the literature. In the literature, serval studies recommend a pluralistic theoretical perspective, thus this is adopted (Carter et al., 2010) (Low et al., 2015). Firstly, the resource dependence theory will be introduced. The focus of this theory is on how firms is depended on linkage between the firm and external actors and how diversity can impact these relationships. Thereafter, the critical mass theory will be presented, which focusses on the impact of diversity and the need to reach a certain threshold before it impacts firm performance. Lastly, the human capital theory will be described, which focusses on how the knowledge and skills that diverse members of the board of director can impact the firm performance. These theories are selected as they are accessed to represent a broad theoretical perspective.

4.1 RESOURCE DEPENDENCE THEORY

The resource dependence theory originates from corporate governance literature and was introduced by Pfeffer and Salancik in 1978. Within the confines of the theoretical framework, the authors argue that to ensure good governance, it is necessary that the board of directors of firms to create links between the corporation and its external environment. This is due to the company's existing in an open and competitive system, and to ensure survival it is essential for them to create a symbiotic linkage between them and other companies (Terjesen et al., 2010). They argue that external linkage has four main advantages; 1) the accumulation of information and industry know-how, 2) the creation of networks with entities that are deemed important for the company, 3) the creation of endorsements from parties that are important within the company's external environment and 4) establish authenticity for the corporation in its environment (Pheffer and Salancik, 1978) (Hillman et al., 2000).

In the resource dependence theory, the role of the board of directors contains many dimensions. When appointed to a board of directors, the board members' background will play an essential role in how they benefit the firm. Each board member provides resources

such as their induvial stock of skills, information, their reach to critical stakeholders (public officials, social networks, buyers, suppliers, ect.) and legitimacy. The scope in which each member will benefit the firm is based on, if the skills and network they provide, yield better information, less environmental dependency, more valuable resources or establishes legitimacy for the firm in their environment (Hillman et al., 2000). Therefore, diverse types of directors will benefit the firm in different types of ways and individuals with diverse backgrounds and unique human capital[3] are equipped to handle different types of environmental dependencies (Carter et al., 2003) (Hillman et al., 2000).

4.2 THE CRITICAL MASS AND TOKEN THEORY

The sociologist Rosabeth M. Kanter is widely acclaimed as the founder of the critical mass theory, which proposes that in a group setting, the influence of a minority group must reach a certain threshold of representation or "critical mass" before their influence becomes significant. Kanter's research primarily focused on the dynamics between men and women in organizational contexts, making the theory especially relevant in gender-based studies (Childs and Krook, 2008).

Kanter is seen as the founder of the critical mass theory. The theory proposes that in a group setting the minority group need to reach a certain "critical mass" before their impact is significant. The focus of Kanter's research is mainly the women-men group dynamics, and is therefore often used in gender-based studies (Childs and Krook, 2008)

Kanter (1977) makes a distinction between four different kinds of groups; uniform-, skewed-, tilted- and balanced groups. Within the model framework, the uniform groups refer to a group, where the members have one specific homogeneous characteristic. One might be able to differentiate the members into smaller sub-groups, but the group as such is still considered uniform with respect to a significant external characteristic, for example gender, ethnicity or race. Typically, the ratios for this type of group are 100:0 (Kanter, 1977). The skewed group refer to a group, where there is one numerically dominant group of members (majority) opposed to a numerically smaller group of members (minority). The group dynamics of the skewed group means, that the majority group will be the dominant group when it comes to the culture of the group and the dominate group will have the general control of the group. The minority group will often be referred as the "tokens". This means that they will be seen

[3] The human capital theory will be addressed in the subsequent section

more as a representative of their specific sub-group than actual members of the overall group. Within the group dynamics they will have little to no power. The reason for this is that the "tokens" often will have none or few individuals to create a potential alliance with, and thereby be a contrast to the dominate group. The ratios for the skewed group are according to Kanter (1977) 15:85. The tilted group refer to a group, where the ratio and group dynamics are slightly different from the skewed group. The distribution is far less extreme compared to the skewed group, and the minority group has the potential to create alliances with other group members as the minority groups members are becoming individuals instead of "tokens". The ratios for the tilted group are according to Kanter (1977) 65:35. Lastly, the balanced group referrer to a group, where the ratio is 50:50. This means that potential sub-groups are created instead of a majority- and a minority group. However, this may also create the potential for type specific identification in the sub-groups. Therefore, creating an environment where new skewed- and tilted groups appear withing the balanced group, which have the same implications of the before mentioned characteristics. In the balanced group the dynamic will be far more based on personal and structural factors (Kanter, 1977).

The dynamics of the skewed group were a particular focus of Kanter (1977). In the skewed group the "tokens" not only diverge from the majority group, but they are specially identified by not being part of this group. This means, that they carry the set of assumptions that are generally applicable for the specific groups even if it is not specifically applicable for the "token" on a personal level. Kanter (1977) argues that the dynamics of the skewed group can be further intensified when the two following conditions are met: 1) when the way the "token" diverge from the majority is physically observable, such as race and sex, and 2) when the "token" is in an environment where they are both new and rare (Kanter, 1977).

"Tokens" are also often faced with a bigger performance pressure than the members of the majority, due to their increased visibility. Kanter (1977) argues that the "token" will typically respond to the increased pressure in two ways: 1) they will overperform, which will further amplify their visibility and can lead to further isolation by the majority, and 2) they will try to assimilate with the majority group by lowering their visibility and thereby their performance. Furthermore, when faced with being in a skewed group the majority group's behavior will also result in a non-optimal performance. This is further emphasized by the fact that uniform groups tend to outperform skewed groups (Kanter, 1977).

4.3 HUMAN CAPITAL THEORY

With the publication of the article "Human Capital" in 1964, Becker is often referred to as the father of the human capital theory. The theory proposes that an individual has a unique set of abilities, such as experiences, education and other life skills, that set them apart from others. The set of unique skills are called human capital, and in a governance context it can be used to enhance a firm's competitive advantage (Borjas, 2020) (Carter et al., 2010) (Barney, 1991). Within the theoretical framework, human capital resources of the individuals in an organization refers to the training, the experience, their judgement, their intelligence, their interpersonal relationships and induvial insights, all of which makes them a unique asset for the firm (Barney, 1991).

Becker theorized that human capital theory could be divided into two distinguished categories: specific- and general human capital (Gibbons and Waldman, 2004). On one hand there is specific human capital which is knowledge or skills that are either firm or industry specific, and therefore properties that cannot be assigned directly onto other sectors or work. This covers highly specialized human capital which is learned through targeted education or industry experience. On the other hand, there is general human capital which is knowledge or skills that can be used cross firms and industries. This covers general skills which would be applied no matter which industry the individual is working in (Gibbons and Waldman, 2004) (Borjas, 2020).

The human capital theory is often used in combination with the resource dependence theory. From a resource dependence theory standpoint, each member of the board of directors provides the board with a specific set of essential skill, human capital, to be able to guide and support the management of the firm in accordance to its challenges (Volonté and Gantenbein, 2016) (Hillman et al., 2000). Thus, the stock of human capital of each board member has a significant impact on how they act in different contexts and therefore the performance of the firm. As individuals obtain different types of human capital by way of experience and education, this often becomes one of the foremost reasons they are elected to a board of directors (Volonté and Gantenbein, 2016) (Terjesen et al., 2009).

5 HYPOTHESES

In the following section, a series of hypotheses will be set up. They will be created based on the theories and empirical findings presented in the previous section. The hypothesis will later be used to examine a link between diversity and firm performance.

5.1 HYPOTHESIS 1

The resource dependence theory indicates that the firms can benefit from legitimacy, advice and counsel and access to resources, that comes from board diversity. It states that the board of directors is the primary linkage between the firm and its environment, therefore making the composition of the board and the human capital of the members essential for the firm's financial performance. The human capital theory posits that the abilities of individuals can influence the performance of the firm, yet the influence of various forms of human capital will not impact performance equally. Both theories suggest the presence of a diverse board of directors can offer a variety of advantages to the firm, due to their varied backgrounds and distinctive human capital. The theories indicate that there may be a link between diversity among the board of directors and financial performance of the firm. However, the theories themselves do not predict that there is a specific link between them. Both theories suggest that the kind of diversity that the individual contributes with is important for how the potential effect will impact the firm, yet this effect could also be insignificant. Based on both resource dependence theory and the human capital theory, this could be due to sector specification. Furthermore, the human capital theory indicates that if the ability of the specific individual is not leveraged in an optimal way by the organization, then the benefits of board diversity will not be fully realized and as a result the presence of diversity might be insignificant.

From a critical mass theory perspective, the relationship between board diversity and the firm's financial performance is contingent on reaching a significant threshold before the firm can realize the benefit of board diversity. The theory indicates that there is a significant link between the board diversity and the firm's financial performance. However, the theory also indicates that when the specific threshold is reached, and a balanced group has been created, new tilted and skewed sub-groups within the balanced groups may appear. This occurrence can lead to group imbalance within the dynamic groups such as problems with communication and conflict. The internal imbalance has the potential to undermine the

positive effects of diversity, that it is expected to bring, thereby implying that diversity can be a source of negative impact on the firm's financial performance, if it is managed ineffectively.

From the empirical studies presented in the literature review, it is evident that the examination of representation of female and ethnic minorities in the board of directors, in terms of impact on financial performance of firms, has produced mixed result. In the study by Rose (2007), no significant links were found between female board representation and the firm's financial performance as measured by Tobin's Q. This indicates that in that specific context, the presence of gender diversity in the board of directors neither increases nor decreases the firm's financial performance. In contrast, Erhardt et al. (2003) found a significant positive link between both female and ethnic minority board representation and the firm's financial performance, where the financial performance was measure as ROA and ROI. This implies that in this specific context the demographic diversity has a positive impact on financial performance. Similar results are found in Carter et al. (2003) where the authors find a positive and significant connection between female and minority diversity in the board of directors and the firm's financial performance as measured by Tobin's Q. However, the before mentioned findings are not consistent across the use of all proxies for firm performance, as seen in Carter et al (2010). Here the authors find that there is there is a positive link between female and ethnic minority representation in the board of directors, and the firm's financial performance as measured by ROA. Yet, when measured by Tobin's Q they find an insignificant relationship. The inconsistent results when examining the relationship between female and ethnic minority representation in the board of directors and the firm's financial performance, suggests that the methodological decisions are important when examining the relationship.

The above reasoning gives rise to the following hypothesis:

Hypothesis 1a: There is a positive association between gender diversity in the board of directors and firms' financial performance

Hypothesis 1b: There is a positive association between ethnic minority diversity in the board of directors and firms' financial performance

5.2 HYPOTHESES 2 AND 3

The critical mass theory indicates that for firms to fully be able to realize the benefit of board diversity, they must take account of the quantity of individuals with a diverse background. When a woman or an ethnic minority gain the status of being a "token" or "solo" member of the board, the members of the dominant group might interpret their presence as being symbolic. In this scenario the "token" member will be faced with being the representation of their specific minority and will be placed in position within the board that is related to their specific stereotype. Moreover, the "token" will have little or no influence on the working of the board, and might be a source of disturbance of the board. However, to fully realize the benefit of board diversity, the number of diverse directors added to the board is important. Based on the theory, when more individuals with the same characteristics as the "token" are added to the board of directors, the dominant group dominance decreases. The "token" will no longer be alone, and in turn the benefit of diversity would increase then the number of individuals with a diverse background increase.

From the empirical studies presented in the literature review, both Gyapong et al. (2016) and Lui et al. (2013) argues that the number of women or ethnic minorities added to the board of directors could be important, and the influence of diversity is stronger when more diverse individuals are included in the board. The evidence from their results also suggests that the increasing number of women and ethnic minorities in the board of directors strengthen the relationship between board diversity and firm performance, and makes the relationship more significant (Gyapong et al., 2016) (Lui et al., 2013).

The above reasoning gives rise to the following hypothesis:

Hypothesis 2: There is an increasing positive association between gender diversity in the board of directors and the firm's financial performance

Hypothesis 3: There is an increasing positive association between ethnic minority diversity in the board of directors and the firm's financial performance

6 ECONOMETRIC METHOD

In the following section, the econometric method utilized in the study will be presented. Firstly, the data will be presented, in addition with the nature of the dataset and the restrictions that have been applied. Subsequently, the cohort will be defined including an elaboration of the methodological consideration made in the selection process. Following, the data structure will be detailed, and the selection of the model used for the quantitative examination. Finally, the endogeneity issue will be addressed, and an outline will be made on how to mitigate the problem.

6.1 THE DATA

The data utilized in this study was obtained through access to "Danmarks Statistiks Forskerservice" facilitated by Aalborg University Business School (AAUBS). This access is limited to specific datasets. [4]This means that there may exist data on the database which could be relevant to the quantitative examination but alas not available due to the access restrictions. However, it has been assessed that the data is appropriate to address the problem statement of the study. The data in the database is registry data, which is usually used for administrative purposes and contains information about firms and individuals across the Danish society. The data makes it possible to get information from across the entire life cycle for individuals and firms, making it possible to study both business entities and their boards. However, as the data collection was not collected for this specific examination, it is not possible to account for all factors that might have been important for the examination. Due to the fact that the information on the database contains identifying information about firms and individuals, there are limitations in the data depicted in this study[5]. Any information that has been deemed as being identifiable by Danish Statistics restrictions cannot be published in this study. The focus of the study is on the period from 2017 to 2019, and the data used in the study is limited to annual data. The reason for this is that the data available from the database is reported as annual data. The time frame was selected to provide a retrospective perspective on how the exploratory- and control variables influence the independent variable in 2019. In the study, there is an implicit expectation diversity takes time to materialize on the firm's

[4] Overview of the data used in this examination is displayed in appendix A
[5] There is a requirement for anonymization of the data in the study, which these certain numbers and figures cannot be displayed

financial performance. Therefore, a lag is included to account for the postponed influence of board diversity on the firm's financial performance. In order to capture this postponed influence, the study expands the time frame to include at least 1 lag, with 2 lags ultimately being employed, to further improve the robustness of the quantitative examination and to mitigate the potential endogeneity. Furthermore, the study is geographically limited to Denmark due to the fact, that the origin of the data is Danish.

6.2 THE COHORT

The cohort of the study has been selected to include firms which are limited liability firms. The rationale behind this decision was that the established literature mainly examines a cohort consisting of publicly listed companies, but due to data availability restrictions, it was not feasible to do the same. Therefore, the limited liability firms were selected as an appropriate approximation of the public listed firms. The cohort has been created with the criterium that the firms do not change industries during the selected period. This is based on the perspective that there can be substantial barriers of entry for firms that transit from one industry to another. Consequently, this decision has been made to minimize potential data disturbances from firms that endure these forms of transitions.

Additionally, the cohort was filtered to include firms that have at least three seats in the board of directors. This decision was made to ensure a variation in the board composition, and thus increasing the likelihood of diversity effects. Firms where the CPR number, gender or ethnicity was unknown of their board members were also excluded from the study. The decision was made as this information is necessary to identify the members of the board and therefore the composition of the board. It is acknowledged that this may result in the unintentional filtering of firms with foreign board members, as they may not have a CPR number. Although these criteria help to ensure a balanced dataset, they also present the possibility of introducing selection bias into the study. This has the potential to limit the generalizability of the findings and reduce the studies ability to fully examine the relationship between board diversity and firms' financial performance. However, to ensure a balanced dataset and its quality for the cohort, the criteria were assessed as being necessary.

6.3 DATA STRUCTURE

In order to examine the link between diversity in the board of directors and the financial performance of firms, a regression analysis is prepared using panel data. This data structure has been chosen as it allows the inclusion of lagged versions of the variables in the

regression. This approach utilizes the benefits of the panel data structure, making it possible to examine how lagged versions of the variables influence the independent variable. However, in the examination, linear regression will be utilized. Bhagat and Black (2001) argue that in order to capture the time effect of board diversity a large data set is required, since the composition of the board of directors often gradually over time (Bhagat & Black, 2001) However, the available data from the database surrounding the Danish boards is limited to 9 years combined, which may not be a sufficient amount of data to sufficiently utilize the panel data regression efficiently. Moreover, the cross-section regression methodology is an accepted methodology within the literature (Rose, 2007) (Erhardt et al., 1997). Hence, while utilizing the panel data structure, the study will focus on the cross-sectional regression methodology to perform the quantitative examination.

6.4 CHOICE OF MODEL

In the process of selecting a suitable model to examine the relationship between firm performance and diversity in the board of directors, different models have been considered. As the dependent variable, firm's financial performance, is continuous, the Ordinary Least Squares (OLS), [6]method is assets as being appropriate for the quantitative examination.

The choice is further underlined by the fact that the OLS method is commonly used in studies examining this relationship (Rose, 2007) (Low et al., 2015). Additionally, the regression approach is chosen since it allows to control for several variables simultaneously, which is seen as being beneficial due to the nature of the dependent variable. It is expected that the firm's financial performance is influenced by various factors, and therefore the capability to control for their various influences is crucial for attaining reliable estimates (Wooldridge, 2016). Hence, the OLS regression approach is assessed to be suitable to address the objective of the study, examining the relationship between firm board diversity and firm's financial performance.

6.5 ENDOGENEITY

Endogeneity is of great concern to this study. Hermalin and Weisbach (2016) mentions that endogeneity is a pressing issue in corporate governance literature, since nearly all variables of interest are endogenous. This means, that it is difficult to precisely assess if the change in performance is truly a result of the variables included in empirical analysis, or if it is a result

[6] The model is presented in appendix H

of endogeneity. One of the main issues that they present is the bidirectional relationship between board composition and firms' financial performance, and the fact that past board members decision has an influence on the future performance of the firm, and that the future performance of the firm impacts the selection of future members of the board (Hermalin and Weisbach, 2016). It can therefore be argued that there exists an endogenous relation between the firm's financial performance and demographic diversity of the board of directors. The study is moreover faced with the risk of revers causality, as demographic diversity might have a positive impact on firm performance, but it is also plausible that better performing firms may also just be better at recruiting diverse members to their board of directors (Liu et al., 2014). In the literature, there is different methods which have been employed to account for endogeneity. One of the most common methods used in econometrics is the use of instrument variables. This method estimates the regression using the two-stage least squares ("2SLS") method (Liu et al., 2014) (Adams and Ferreira, 2009). However, one of the challenges when using this method is choosing an appropriate instrumental variable. The requirement for the chosen instrument is that it needs to correlate with the independent variable but uncorrelated with the dependent variable (Adams and Ferreira, 2009). Given the difficulties of finding an appropriate instrumental variable this method has not been chosen. Instead, a one year-lagged version of the explanatory variable is utilized as a method to decrease the endogeneity (Labelle et al., 2015) (Liu et al., 2014).

7 RESEARCH DESIGN

In the following section the research design of the quantitative examination is presented. Firstly, the structure of the regression models will be presented, [7] and subsequently the review of the dependent, the explanatory and control variables.

7.1 HYPOTHESIS 1A

Hypothesis 1a is that there is a positive correlation between gender diversity in the board of directors and firms' financial performance. As it has been presented in section 6, the examination is made in a retrospective perspective meaning, that it is assumed that previous values of the explanatory- and control variables influences the firm's financial performance in 2019.

Equation 1: Model 1aa

$$ROA_t = \beta_0 + \beta_1 SIK_{t-2} + \beta_2 SF_{t-1} + \beta_3 BS_{t-1} + \beta_4 FA_{t-1} + \beta_5 REV_{t-1} + \beta_6 AE_{t-1} + \beta_7 ET_{t-1} + \beta_8 AS_{t-1} + \beta_9 GO_{t-1} + \beta_{10} IN_{t-1} + \varepsilon_t$$

Equation 2: Model 1ab

$$ROA_t = \beta_0 + \beta_1 PK_{t-2} + \beta_2 SF_{t-1} + \beta_3 BS_{t-1} + \beta_4 FA_{t-1} + \beta_5 REV_{t-1} + \beta_6 AE_{t-1} + \beta_7 ET_{t-1} + \beta_8 AS_{t-1} + \beta_9 GO_{t-1} + \beta_{10} IN_{t-1} + \varepsilon_t$$

7.2 HYPOTHESIS 1B

Hypothesis 1a is that there is a positive correlation between ethnic minority diversity in the board of directors and firms' financial performance. As it has been presented in section 6 The examination is made in a retrospective perspective meaning, that it is assumed that previous values of the explanatory- and control variables influences the firm's financial performance in 2019.

Equation 3: Model 1ba

$$ROA_t = \beta_0 + \beta_1 SIE_{t-2} + \beta_2 SF_{t-1} + \beta_3 BS_{t-1} + \beta_4 FA_{t-1} + \beta_5 REV_{t-1} + \beta_6 AE_{t-1} + \beta_7 ET_{t-1} + \beta_8 AS_{t-1} + \beta_9 GO_{t-1} + \beta_{10} IN_{t-1} + \varepsilon_t$$

[7] The variables and their representation is presented in appendix B

Equation 4: Model 1bb

$$ROA_t = \beta_0 + \beta_1 PE_{t-2} + \beta_2 SF_{t-1} + \beta_3 BS_{t-1} + \beta_4 FA_{t-1} + \beta_5 REV_{t-1} + \beta_6 AE_{t-1} + \beta_7 ET_{t-1} + \beta_8 AS_{t-1} + \beta_9 GO_{t-1} + \beta_{10} IN_{t-1} + \varepsilon_t$$

7.3 HYPOTHESIS 2

Hypothesis 2 is that There is an increasing positive association between gender diversity in the board of directors and the firm's financial performance. As it has been presented in section??? The examination is made in a retrospective perspective meaning, that it is assumed that previous values of the explanatory- and control variables influences the firm's financial performance in 2019.

Equation 5: Model 2

$$ROA_t = \beta_0 + \beta_1 Dwomen1_{t-2} + \beta_2 Dwomen2_{t-2} + \beta_3 Dwomen3_{t-2} + \beta_4 SF_{t-1} + \beta_5 BS_{t-1} + \beta_6 FA_{t-1} + \beta_7 REV_{t-1} + \beta_8 AE_{t-1} + \beta_9 ET_{t-1} + \beta_{10} AS_{t-1} + \beta_{11} GO_{t-1} + \beta_{12} IN_{t-1} + \varepsilon_t$$

7.4 HYPOTHESIS 3

Hypothesis 3 is that There is an increasing positive association between ethnic minority diversity in the board of directors and the firm's financial performance. As it has been presented in section??? The examination is made in a retrospective perspective meaning, that it is assumed that previous values of the explanatory- and control variables influences the firm's financial performance in 2019.

Equation 6: Model 3

$$ROA_t = \beta_0 + \beta_1 Detnicity1_{t-2} + \beta_2 Detnicity2_{t-2} + \beta_3 Detnicity3_{t-2} + \beta_4 SF_{t-1} + \beta_5 BS_{t-1} + \beta_6 FA_{t-1} + \beta_7 REV_{t-1} + \beta_8 AE_{t-1} + \beta_9 ET_{t-1} + \beta_{10} AS_{t-1} + \beta_{11} GO_{t-1} + \beta_{12} IN_{t-1} + \varepsilon_t$$

To ensure the robustness of the results of the above regression models, a second series of OLS regressions will be performed using the return of equity as the dependent variable.

7.5 DEPENDENT VARIABLE [8]

The dependent variable for the study is the financial performance of the firms. As it was presented in the literature review, there is a continued debate within the financial performance literature regarding which type of measurement for firm performance is the most

[8] How the dependent variable has been calculated is presented in appendix

appropriate to use. Both accounting- and marked-based measurements are possible to use in these types of studies and each comes with their own benefits and limitations. However, they are both generally accepted within the literature. In this study, return on assets (ROA) has been chosen as the measurement for firms' financial performance. ROA, an accounting-based measurement, gives an indication of the firm's ability to produce profit from its assets (Erhardt et al., 2003). The choice of ROA as the proxy for firm performance is motivated by the widespread use and acceptance within the research literature (Carter et al., 2010) (Erhardt et al., 2003) (Shrader et al, 1997), and the accessibility of the data in the database. Tobin's Q was considered as a proxy for firm performance. However, due to the lack of appropriate and accessible date in the database, it was not possible to utilize this specific measure. This led to the decision to not use Tobin's Q in favor of ROA. Moreover, as a robustness test the return of equity ROE is utilized.

7.6 EXPLANATORY VARIABLES [9]

The variables of interest in the study are demographic diversity, defined by gender and ethnic diversity. It is, therefore, necessary for the independent variables included in the study to properly address these dimensions of diversity. In prior studies, various methods have been utilized to assess diversity within the board of directors. In this study the Shannon index for gender and ethnic minorities, as well as the percentage of board seats occupied by women and ethnic minorities respectively are selected as the explanatory variables for gender and ethnic minority diversity. The choice of the explanatory variable is motivated by the widespread used and acceptance of this within the literature (Campbell and Mineguez-Vera, 2007) (Erhard et al. 2003) (Liu et al., 2013), and the accessibility of the data in the database. The Shannon index is commonly used across academic disciplines and provides an indication of the diversity within a population with respect to specific categorical characteristics. The metric accounts for both the positive and negative impact of diversity and considers the overrepresentation of any single group in the population. This approach is supported by the methods used in Campbell and Mineguez-Vera (2007) and Østergaard et al. (2011). The percentage of the board seats occupied by women and ethnic minorities represents another commonly used diversity metric within the literature. It allows for the observation of changes in the ratio of diversity (Erhardt et al. 2003). This approach is also supported by the methods used in Erhardt et al. (2003), Campbell and Mineguez-Vera (2007) and Liu et al. (2013).

[9] How the explanatory variables have been calculated is presented in appendix C

Other studies have used different metrics such as the total numbers of either females or ethnic minorities in the board of directors to address demographic diversity (Carter et al., 2010), dummy variables to indicate the presence of at least one woman in the board of directors (Campbell and Mineguez-Vera, 2007) and the Blau index (Campbell and Mineguez, 2007). However, this study makes use of the Shannon index, and the percentage of seats occupied by women and ethnic minorities, as these metrics have been considered both reliable and accessible for the analysis.

In addition to the measurement of diversity, the hypothesis also explores the impact of different levels of diversity on the firm's financial performance. This aspect of the study is grounded in the critical mass theory, which states that "one is a token, two is a presence, and three is a voice (Kristie, 2011). To test the critical mass theory, the method suggested by Liu et al. (2013) and Gyapong et al. (2016) will be implemented. For gender diversity, three dummy variables have been created to test the critical mass theory. The first dummy variable indicates that there is 1 woman in the board of directors. The second dummy variable indicates that there are 2 women in the board of directors, and the third dummy variable indicates that there are 3 or more women in the board of directors. In a similar manner, to test the theory in relation to ethnic minority diversity, three dummy variables in a similar design have been created. The first dummy variable indicates that there is 1 ethnic minority in the board of directors. The second dummy variable indicates that there are 2 ethnic minorities in the board of directors, and the third dummy indicates that there are 3 or more ethnic minorities in the board of directors.

7.7 CONTROL VARIABLES[10]

In the study a series of variables are included to account for factors that might influence the firm's financial performance. The selected control variables are based on both economic theory and the established literature to ensure robustness in the quantitative examination.

7.7.1 Firm size

From both an empirical and theoretical perspective, the firm's size is believed to have an impact on the firm's financial performance. One argument suggests a positive relationship based on the assumption that larger firms have a greater probability to benefit from economies of scale which leads to greater production efficiency. Furthermore, larger firms typically have a more direct access to external capital, enabling them to raise profits more

[10] How the control variables have been calculated is presented in appendix C

effectively (Labelle et al., 2015). From another perspective, it can be argued that firm size may impact the firm's financial performance negatively. Larger firms tend to be predisposed to agency problem, such as conflict of interest and information asymmetry, which can weaken profitability and efficiency (Labelle et al., 2015). Moreover, firm size is commonly used in diversity and firm performance studies (Smith et al., 2006) (Campbell and Mineguez-Vera, 2007). Although there are mixed results regarding firm size having either a negative or positive impact on firm performance, the variable is included in the study as a control.

7.7.2 Board size

In the literature, there is a debate whether board size has either a positive or negative impact on the firm's financial performance. On one hand, a positive relationship can be argued. A larger number of members may lead to a greater pool of expertise and knowledge, which have the potential to improve decision-making. Moreover, as discussed in section 4, forming external linkage are important for the performance and survival of the firm. A larger board, with a greater number of members, has an increased likelihood of establishing these crucial connections with its environment (Labelle et al., 2015). On the other hand, a negative relationship can also be argued. As it was presented in the literature review, more heterogeneous groups may be subject to increased conflict and decreased cohesion. Furthermore, the complexity of the group dynamics can result in a prolonged decision process, which may be especially problematic in high-pressure situations (Labelle et al., 2015). The board size is commonly used in diversity and firm performance studies (Carter et al., 2010) (Lui et al., 2013) (Gyapong et al., 2016). Although there are mixed results regarding the board size having either a negative or positive impact on the firm's financial performance, the variable is included in the study as a control

7.7.3 Firm age

From a theoretical perspective, firm age is considered to have an impact on the firm's financial performance. The literature posits that younger firms have a greater chance of failure than older firms, while older firms also tend to outperform younger firms (Mallinguh et al., 2020) (Cressy, 2006). This can be attributed to the concept of "survival of the fittest" which means that successful firms survive and stay in the market, while unsuccessful ones will exit the market (Johnson et al., 2013). In addition, younger firms often face the "liability of newness" which refers to the difficulties firms face in their early stages. These includes unstable costumer connections, underdeveloped problem-solving practices, lack of social linkage to their external environment and challenges of learning new roles. In contrast, older

firms, having already established tires to its environment and customers, have a greater likelihood of survival and outperform younger firms (Yang & Aldrich, 2017) (Coleman, 2004). On the other hand, Schumpeter argued that in the long run older firms will be outcompeted by newer firms through creative destruction. Creative destruction refers to the way process and product innovation replaces existing goods and methods, leading to improvements in productivity (Caballero, 2010). The introduction of innovation to the economy may not only improve efficiency but also make space for new market leaders. Subsequently, creative destruction may tear the competitive advantage of older firms, and thereby allowing for newer and better performing firms (Caballero, 2010). Furthermore, the theory suggests that firm age may become a liability for older firms, as they may be slower to adapt to new innovations. Firms that are not able to adapt to the new market conditions are more likely to fail. Lastly, firm age is commonly used in diversity and firm performance studies (Lui et al., 2016) (Gyapong et al., 2016) (Smith et al., 2006). Thus, the variable is included in the study as a control.

7.7.4 Revenue

From a theoretical perspective, it can be argued that the revenue of the firm could have an impact on firm performance. According to a resource-based view, a firm can improve its competitive advantage and performance by managing their resources optimally. Higher revenue makes it possible for firms to invest in the key resources that are essential for their competitive edge. By combining the resources effectively, the firms can obtain the optimal resource mix to achieve competitive advantages. However, obtaining the optimal mix of resources is difficult, and miss-management of revenue and over-investing in non-core areas may occur. This improper allocation may cause a rise in cost and non-performance enhancing activities, which impact the firm's financial performance negatively (Wernerfelt, 1984). From an agency theory perspective, the interest between the managers and the shareholders of the firm do not always align. In firms with higher revenue, managers have more resources available for value-adding operations, that align with the interest of the shareholders. However, if the managers decisions are driven by self-interest, rather of those of the shareholders, the firm's financial performance could be impacted negatively regardless of high revenue (Carter et al., 2003). Hence, it has been assessed that the variables should be included in the study as a control

7.7.5 Industry

From an empirical and theoretical perspective, it can be argued that it would be beneficial to control industry effects, as the relationship between firm performance and board diversity may differ across industries. The human capital theory proposes that the abilities of the employees have an impact on firm performance. In the resource dependence theory, a great importance is held on external linkage. Each industry encounters different challenges from its environment, which in turn has the potential to impact firm performance differently. In the critical mass theory, diversity may have an impact on firm performance only when a specific threshold is reached. Different industries might be faced with different standards of when the critical mass is reached. Moreover, empirical studies have found that female and ethnic minority representation tends to be industry specific. Farrell and Herch (2005) found that women have a greater likelihood of being present in the board of directors in sectors with a greater emphasis on labor and service (Farrell and Hersch, 2005). On the other hand, others have found a greater concentration of women and ethnic minorities in the financial sectors (Carter et al., 2003). Moreover, in the established literature, industry is commonly used in the diversity and firm performance studies (Rose et al., 2007) (Carter et al., 2003) (Erhardt et al., 2003). Hence, it has been assets that the variable should be included in the study as a control.

7.7.6 Average educational level in the firm

From a theoretical perspective, it can be argued that it would be beneficial to control the average education level of the employees in the firm. According to the human capital theory, education is an important driver of employees' skills, knowledge and overall capabilities, and contributes to the firm's performance and competitive advantage. Studies has found that an individual's educational level is often seen as an indicator of their ability to obtain knowledge, making them more attractive from a firm perspective, given that it implies that potential candidates would have it easier to obtain on-the-job-training. This is an important trait since informal on-the-job-training has a greater likelihood to yield a competitive advantage for the firm (Hitt et al., 2001). Moreover, Biasi et al. (2021) argues that education has the potential of improving innovation and promoting successful innovations (Biasi et al., 2021). Hence, it has been assets to include the average educational level in the firm as a control in the study.

7.7.7 Average employee tenure in the firm

From a theoretical perspective it can be argued that it can be beneficial to control employee tenure. According to the human capital theory the abilities of the employees have an impact

on firm performance and there is a significant difference between firm specific- and general human capital. Over the term of their employment, employees accumulate both firm specific- and general human capital, through on-the-job training and experience. Employee tenure indicates a large amount of firm-specific human capital and high-level productivity potential. Moreover, individuals with long tenures are often associated with having a vast tactic knowledge (Gagliardi et al., 2023) (Dosi & Grazzi, 2010). Dosi and Grazzi describes tactic knowledge as "*the inability by actor(s) implicated, or even by sophisticated observers, to explicitly articulate the sequences of procedures by which "things are done", problems are solved, behavioral patterns are formed, ect"* (Dosi & Grazzi, 2010) (Gagliardi et al., 2023). Tactic knowledge includes a personal element which makes it unable to be transferred from one person to another. Therefore, firms risk losing tactic knowledge when employees changes job. When employees adopt the norms, culture, procedures and goals of the organization the their potential within the organization are impacted positively and improves their ability to perform their core tasks, which eventually impact the firm positively (Gagliardi et al., 2023)(Steffens et al., 2014). However, the positive impact of tenure on employee performance follows the law of diminishing returns. In the beginning of an employee's career at a specific firm, the rate at which they acquire tenure-related resources tends to be higher, than employees who are at a more advanced stage of their career (Gagliardi et al., 2023). Hence, it has been assets to include the average employee tenure in the firm as a control in the study.

7.7.8 Average salary in the firm

From a theoretical perspective, it can be argued that it would be beneficial to control the average salary in the firm. Labor economics proposes a positive association between the productivity of the firm and employee wages. In this context, the compensation of employees is determined by the value of marginal product of labor. A profit-maximizing firm will continue to hire until the monetary gain from hiring one additional employee is equal to the cost of that employee. When the marginal product is equal to the wage, the firm is producing a level of output, where marginal cost is equal to the price. Given that the productivity of the firm is linked with the employees' wages, a rise in the average wage within the firm indicates an increase in the productivity of the firm, and thus performance (Borjas, 2020). Hence, it has been assets to include the average salary in the firm as a control in this study.

8 DESCRIPTIVE STATISTICS

In the following section the descriptive statistics of the untransformed variables included in the study are presented. Subsequently, the variables will be transformed in line with the insight gained by the analysis. Lastly a multicollinearity test is performed to evaluate the potential correlation between the variables.

Table 1 presents the untransformed descriptive statistics of the variables included in the study. The return of assets (ROA) has a mean of 10.61% indicating that, on average, the firms generate profits from their assets. However, with a median 6.32%, it is apparent that half of the firms report a ROA that is inferior to this level, which indicates a significant skewness. In addition, the high level of kurtosis and skewness indicate the presence of considerable outliers which is further underlined by a standard deviation on 343.02. Similarly, the return on equity (ROE) has a mean of 99.55% and a median of 17.23%, suggesting a significant skewness. This is further substantiated by the skewness and kurtosis of 103.78 and 11207.51, and standard deviation on 9030.83, which points to the presence of considerable outliers.

The Shannon index for gender diversity reports a mean of 0,35 and a median of 5,56, indicating that, on average the gender diversity falls short of a 50/50 representation. Moreover, the skewness of -0,21 and kurtosis suggest that the distribution of the variable is approximately balanced. In contrast, the Shannon index for ethnic minority diversity reveals a mean of 0.03 and a median of 0 implying that half of the firms lack ethnic minority diversity altogether. The skewness of 4.01 and kurtosis of 14.41 suggests that many of the firms in the sample have a relatively small Shannon index for ethnic minority diversity.

Regarding the percentage of women in the board of directors of the firms in the sample, the mean is 21.10% and has a median of 25%, suggesting that half of the firm report less than 25% female representation. This is further underlined by the skewness of 0.65 and kurtosis of -0.35, indication a right skewed distribution, meaning that most of the observations are left of the mean. The average percentage of ethnic minorities in the board of directors, has a mean of 1.89% with a median on 0, indication that half of the firms have no ethnic minorities represented in their board of directors. The skewness of 6.50 and kurtosis of 53.34 further confirms the disproportionate distribution, since this suggests a right skewed distribution, meaning that most of the observation are left of the mean.

Table 1: The untransformed variables

	vars	n	mean	median	sd	skew	kurtosis
ROA	1	12,048	10.61	6.35	343.02	91.53	9,199.51
ROE	2	12,048	99.55	17.23	9,030.83	103.78	11,207.51
Shannon_index_koen	3	12,048	0.35	0.56	0.31	-0.21	-1.89
Shannon_index_etnicity	4	12,048	0.03	0	0.13	4.01	14.41
Dwomen1	5	12,048	0.39	0	0.49	0.45	-1.80
Dwomen2	6	12,048	0.15	0	0.36	1.97	1.87
Dwomen3	7	12,048	0.03	0	0.17	5.50	28.22
Detnicity1	8	12,048	0.05	0	0.22	4.15	15.23
Detnicity2	9	12,048	0.01	0	0.07	13.72	186.21
Detnicity3	10	12,048	0.003	0	0.06	16.85	281.81
Percentage_women	11	12,048	21.10	25	21.47	0.65	-0.35
Percentage_etnicity	12	12,048	1.89	0	8.92	6.50	53.34
Firm_age	13	12,048	20.59	17	14.62	1.55	4.02
Average_salary	14	12,048	374,322.50	351,614.80	211,501.60	14.17	459.02
Average_Education	15	12,048	3.43	3.38	0.90	0.36	1.67
Revenue	16	12,048	81,407.15	17,526.48	669,943.20	72.34	6,602.13
BoardSize	17	12,048	3.83	3	1.26	2.53	10.35
FirmSize	18	12,048	85,704.77	10,869.41	1,066,923.00	61.70	4,760.94
Tenure	19	12,048	6.20	5.17	4.79	2.00	6.83

For the dummy variables representing gender and ethnic minority in the board of directors, table 1, shows a pattern, as the number of seats the dummy represents increases, the mean and standard deviation decreases. From the skewness and the kurtosis of the variables it can be observed that only the variables Dwomen1 and Dwomen2 appear normally distributed with the values 0.45 and -1.80, and 1.97 and 1.87 respectively. In contrast, the skewness and kurtosis for Dwomen3 are 5.50 and 28.22, while for Detnicity1 it is 4.15 and 15.23, Detnicity2 it is 13.72 and 186.21 and lastly, for Detnicity3 it is 16.85 and 281.81. The values indicate a great right-skewness, indicating that a great deal of the values for the firms included in the sample are left of the mean. Since the dummy variables are binary, the left-skewness suggests a high prevalence of 0 values, indicating that for many of the diversity threshold is not met. Moreover, the increase in the skewness suggests that the stricter the criteria of the dummy variables, the fewer firms achieve the level of diversity.

From table 1, a series of insights can be made for the control variables included in the sample. Firm age has a mean of 20.59 years, and a median of 17 years. The average salary of the employees has a mean of 374,322.50 kr., and a median of 351,614.80 kr. The variable has a skewness of 14.17 and kurtosis of 459.02, which indicates a right-skewness of the variables. This suggests that a great number of firms have an average wage for their employees that are left of the mean, and a limited number of firms that have significantly higher salaries, creating the right-tailed distribution. This also indicates the presence of outliers.

Regarding the average education in the firms, the mean is 3.43 and the median is 3.38, suggesting an approximately normal distribution, which is further underlined by its skewness and kurtosis of 0.36 and 1.67. This suggests, at lack of outliers in the variable. The average tenure of the employees has a mean of 6.2 and a median of 5.17. based on the skewness and kurtosis of 2 and 6.83, it implies that the variables are approximately normal distribution, however with a higher kurtosis.

For the number of board seats in the firms, the mean is 3.83 and the median is 3, which indicates that half of the boards have 3 seats available[11]. This indicates that half of the firms in the sample have a relatively small board. The skewness and kurtosis of the variable are 2.53 and 10.35, which suggests a right-skewed distribution, and outlier with some boards having a relatively large number of board seats. In terms of the firm size[12], the mean is 85,704.77 DKK[13]. and the median is 10,869.41 DKK[14].. The skewness and kurtosis of the variable are 61.70 and 4760, which indicates a highly left-skewness of the distribution with the potential of outliers. Lastly, the mean of the revenue is 81,407.15 [15]DKK and the median is 17,526.48[16] DKK. the skewness and kurtosis suggest a highly right skewness, indicating the potential of outliers.

8.1 DATA TRANSFORMATION AND OUTLIERS

As illustrated in table 2 and the above presentation of the untransformed data, there is a probability of outliers being present in the dataset, therefore making it necessary for it to be processed before further quantitative examination. To take this into account, a series of steps are undertaken. Firstly, the variables are examined for outliers using histograms. This analysis found that ROA, ROE, Average salary of the firm and revenue all exhibit outliers that could not be addressed using the naturel logarithm. For ROA and ROE, the limitation arose due to the presence of negative values. For the average salary of the firm and revenue, a prior attempt at using the naturel logarithm did not achieve a satisfying result. Therefore, the Box-and-Whisker plot [17]method was utilized to identify outliers for the variables, and subsequently filtered for in the dataset. This step had a significant impact on the kurtosis and skewness of the variables. After the application of the Box-and-Whisker plot, the variables

[11] This can be deduced based on the restrictions mentioned in section 6
[12] This is the untransformed firm size before the natural logarithm is taken of it as it is described in appendix C
[13] Measured in 1.000 DKK
[14] Measured in 1.000 DKK
[15] Measured in 1.000 DKK
[16] Measured in 1.000 DKK
[17] The Box-and-Whisker plot method is presented in appendix G

were visualized using the histogram again, and the distribution were approximately normally distributed[18].

Table 2: The transformed variables

	vars	n	mean	median	sd	skew	kurtosis
ROA	1	7,579	7.55	6.48	7.23	0.59	0.19
ROE	2	7,579	18.05	15.95	16.69	0.37	0.14
Shannon_index_koen	3	7,579	0.37	0.56	0.31	-0.35	-1.82
Shannon_index_etnicity	4	7,579	0.02	0	0.11	5.00	23.33
Dwomen1	5	7,579	0.42	0	0.49	0.31	-1.90
Dwomen2	6	7,579	0.15	0	0.36	1.91	1.65
Dwomen3	7	7,579	0.02	0	0.15	6.14	35.76
Detnicity1	8	7,579	0.04	0	0.19	5.01	23.10
Detnicity2	9	7,579	0.001	0	0.04	26.19	683.82
Detnicity3	10	7,579	0.001	0	0.03	35.49	1,257.83
percentage_women	11	7,579	22.43	25	21.35	0.52	-0.50
percentage_etnicity	12	7,579	1.11	0	6.15	7.13	69.23
log_Firm_age	13	7,579	2.77	2.83	0.79	-0.90	1.36
Average_education	14	7,579	3.35	3.32	0.83	0.26	1.52
log_Average_salary	15	7,579	12.70	12.73	0.33	-0.96	1.67
log_Revenue	16	7,579	9.56	9.66	1.20	-0.77	1.15
log_BoardSize	17	7,579	1.28	1.10	0.24	1.37	1.89
log_FirmSize	18	7,579	9.24	9.20	1.19	0.44	0.97
log_Tenure	19	7,579	1.62	1.71	0.79	-0.73	1.28

In the second step, the natural logarithm was applied to variables to transform them to the state in line with the existing literature and to achieve an approximately normal distribution[19]. The transformation was applied to the board size, the firm size, the revenue, the firm age, the number of employees, employee tenure [20]and the average salary in the firm. In the third and last step, the Cooks distance [21] were calculated to identify potential outliers' dataset that would negatively impact on the output when plotting the regression. This is done by plotting the regressions for the quantitative examination using the independent variable that is going to be used for both the primary examination and the robustness test. Here several observations were identified as potential outlier by due to their Cooks distance were above

[18] Due to the requirement for anonymization of the data, these cannot be displayed
[19] The naturel logarithm is taken of the board size, firm size, revenue, firm age and the number of employees
[20] Due to the fact the variable follows the law of diminishing returns
[21] The Cooks Distance method is presented in Appendix G

the cutoff value[22], and subsequently filtered for[23]. In the end the filtering process leads to the data presented in table 2 which will be used for the quantitative examination.

8.2 MULTICOLLINEARITY

In the following section, multicollinearity in the dataset is investigated. Multicollinearity is present in the dataset when there is perfect linear relationship between the exploratory variables, which violates one of the OLS regression assumptions[24]. Moreover, the presence of multicollinearity can also have implications for variance of the estimated coefficient, where they will enlarge the standard errors, and thereby influence the further quantitative examination (Studenmund, 2014). To investigate the possibility of multicollinearity in the dataset a Pearsons correlation matrix [25]is utilized. The subsequent Pearson correlation matrix is presented in appendix F[26]

The matrix indicates that there is no significant high correlation [27]between the explanatory variables. An inspection of the independent variables, more specifically gender and ethnic minority dummies, suggests that the dummy trap [28]is avoided, since none of the dummies for gender are highly correlated with each other, and the same applies for the ethnic minority dummies. However, as it is noted in appendix E, one of the assumptions of the Pearson correlation matrix is that all the variables included are normally distributed, which is violated by including the dummy variables. Therefore, although the Pearson matrix offers insights, it is not sufficient to rule out multicollinearity entirely. Consequently, further tests will be carried out in the quantitative examination of the regression models, such as variance inflation factors (VIF) (Wooldridge, 2016) (Studenmund, 2014).

[22] The method on how to calculate the cutoff value is presented in appendix G
[23] Due to the requirement for anonymization of the data, the visualization of the Cooks Distances cannot be displayed
[24] The OLS regression assumptions are presented in appendix D
[25] The theory of the Pearsons correlation matrix and its assumptions is presented in appendix E
[26] The Pearsons correlation matrix is presented in the appendix F
[27] An arbitrary rule of thumb is that a correlation over 0.80 is considered high, when consultation the literature (Studenmund, 2014)
[28] The dummy entails occurs when too many dummy variables are included in the regression, where they describe all available outcomes of the variable (Wooldridge, 2016)

9 QUANTITATIVE EXAMINATION

In the following section, the estimated results of the quantitative examination will be presented. Initially, as well as, the decision rule for the support of the hypotheses and subsequently, the estimated results and analysis will be presented.

9.1 DECISION RULE

With a view to examine the hypotheses which have been set up in the previous section, a decision rule is set up. The decision rule will be based on the estimated $\hat{\beta}$ being positive and that the p-value being significant. Here the minimum p-value for it being significant will be at a 5% significant level (Studenmund, 2014).

9.2 RESULTS

In the following section, the quantitative study will be performed. It will be examined if the established hypotheses are supported by the quantitative results. Before the data is utilized in the regression models, all the explanatory- and control variables are standardized. The reason for this is, that it reduces possible collinearity in the dataset, and it reduces the magnitude of the variables included in the regression, so no variable is interpret as contributing to the independent variable more than others (Wooldridge, 2016). The results of the quantitative examinations are presented in a series of tables including information regarding the specific regression model. Before they are presented in the below sections a series of diagnostics to investigate whether they uphold the Ordinary least squares (OLS) assumptions[29] [30] have been performed. Furthermore, the control variable for industry effect is depicted as one collective variable. The reason for this is that it is a categorical control variable and therefore not one of the explanatory variables. This means that the significance of the variable in the OLS regression for the models indicates the highest level of significance for any of the categorical variables.

In the following results of the OLS regression tables, it has been chosen to use the robust standard errors in favor of the original estimate. This is due to the fact that in the diagnosis of all the regression models, heteroskedasticity have been detected. This is also the case for the robustness test of the examination.

[29] The Ordinary least squares (OLS) assumptions are presented in appendix D
[30] The diagnostics of whether they uphold the OLS assumptions are presented in appendix I

9.2.1 Hypothesis 1

Hypothesis 1 is that there is a positive association between board diversity, in terms of gender and ethnic minorities, and the firm's financial performance. This will be examined utilizing the regression models presented in section 7.

In the below Tabel 3, Tabel 4, Tabel 5 And Tabel 6 The results of the model 1aa, model 1ab, model 1ba and model 1bb respectively, are presented using the OLS regression[31]

Table 3: OLS regression for model 1aa

	Estimate	Sdt. Error	T value	P value	
Intercept	8.4623474	1.1355555	7.4522	<0.0000000000000002	***
SIK_{t-2}	-0.1756040	0.0821483	-2.1376	0.0325775	*
$Log(SF_{t-1})$	-1.9295112	0.1028940	-18.7519	<0.0000000000000002	***
$Log(BS_{t-1})$	-0.2800460	0.0770146	-3.6363	0.0002785	***
$Log(FA_{t-1})$	-0.2142311	0.0877425	-2.4416	0.0146456	*
$Log(REV_{t-1})$	2.0272045	0.1010348	20.0644	<0.0000000000000002	***
AE_{t-1}	0.1772914	0.0788946	2.2472	0.02246564	*
$Log(ET_{t-1})$	-0.0469852	0.0896279	-0.5242	0.6001455	
$Log(AS_{t-1})$	0.9481400	0.0826447	11.4725	<0.0000000000000002	***
IN_{t-1}					.
Observations	7579				
R^2	0.1121				
Adjusted R^2	0.1105				
Residual Std. Error	6.817				
F Statistics	68.21 *** (df = 14; 7561)				

[32]

[31] The implication of the Ordinary least squares (OLS) model is presented in appendix H

[32] For all the regression the * indicates a 5% significance level, ** indicates a 1% significance level and *** indicates a 0.1% significance level

Table 4: OLS regression for model 1ab

	Estimate	Sdt. Error	T value	P value	
Intercept	8.4559210	1.1433234	7.3959	<0.0000000000000002	***
PK_{t-2}	-0.1990526	0.0798113	-2.4940	0.0126510	*
$Log(SF_{t-1})$	-1.9291390	0.1029459	-18.7394	<0.0000000000000002	***
$Log(BS_{t-1})$	-0.2867006	0.0770430	-3.7213	0.0001996	***
$Log(FA_{t-1})$	-0.2111087	0.0878537	-2.4030	0.0162871	*
$Log(REV_{t-1})$	2.0171161	0.1015254	19.8681	<0.0000000000000002	***
AE_{t-1}	-0.1753364	0.0788875	2.2226	0.0262714	*
$Log(ET_{t-1})$	-0.0441696	0.895907	-0.4930	0.6220158	
$Log(AS_{t-1})$	0.9449072	0.0827027	11.4253	<0.0000000000000002	***
IN_{t-1}					.

Observations	7564
R^2	0.1123
Adjusted R^2	0.1106
Residual Std. Error	6.817
F Statistics	68.31 *** (df = 14; 7561)

Table 5: OLS regression for model 1ba

	Estimate	Std. error	T value	P value	
Intercept	8.568278	1.132153	7.5681	<0.0000000000000002	***
SIE_{t-2}	-0.114126	0.058035	-1.9665	0.0492753	*
$Log(SF_{t-1})$	-1.924405	0.102880	-18.7054	<0.0000000000000002	***
$Log(BS_{t-1})$	-0.278701	0.077019	-3.6186	0.0002981	***
$Log(FA_{t-1})$	-0.219968	0.087882	-2.5030	0.0123356	*
$Log(REV_{t-1})$	2.054490	0.100408	20.4614	<0.0000000000000002	***
AE_{t-1}	0.188131	0.078629	2.3926	0.0167519	*
$Log(ET_{t-1})$	-0.071538	0.089216	-0.8019	0.4226599	
$Log(AS_{t-1})$	0.968801	0.082397	11.7577	<0.0000000000000002	***
IN_{t-1}					.
Observation			7579		
R^2			0.1118		
Adjusted R^2			0.1102		
Residual Std. Error			6.818		
F Statistics			68.01 *** (df = 14;7561)		

Table 6: OLS regression for model 1bb

	Estimate	Std. error	T value	P value	
Intercept	8.570650	1.132040	7.5710	<0.00000000000004143	***
PE_{t-2}	-0.106943	0.051223	-2.0878	0.0368513	. *
$Log(SF_{t-1})$	-1.924234	0.102865	-18.7064	<0.0000000000000002	***
$Log(BS_{t-1})$	-0.281227	0.076997	-3.6524	0.0002615	***
$Log(FA_{t-1})$	-0.219836	0.87878	-2.5016	0.0123842	*
$Log(REV_{t-1})$	2.054277	0.100401	20.4608	<0.0000000000000002	***
AE_{t-1}	0.186828	0.078635	2.3759	0.0175316	*
$Log(ET_{t-1})$	-0.071434	0.089213	-0.8007	0.4233226	
$Log(AS_{t-1})$	0.967443	0.082367	11.7455	<0.0000000000000002	***
IN_{t-1}					.
Observation	7579				
R^2	0.1118				
Adjusted R^2	0.1102				
Residual Std. Error	6.818				
F Statistics	67.99*** (df = 14;7561)				

The results presented in table 3 indicate that there is a negative association between the Shannon index for gender (SIK_{t-2}) and the return on assets (ROA). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis. The findings suggests that a 1- unit increase in the Shannon index for gender is associated with a reduction of 0.1756040 in the return of assets for the firms in the sample, holding all other factors constant. This implies that a higher Shannon index for women in the board of directors may have a negative influence on the firm's financial performance.

Similarly, the results presented in table 4 indicate that there is a negative association between the percentage of women in the board of directors (PK_{t-2}) and the return on assets (ROA). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis The findings suggests that a 1-unit

increase in the percentage of women is associated with a reduction of 0.1990526 in the return of assets for the firms in the sample, holding all other factors constant. This implies that a higher percentage of women representation in the board of directors may have a negative influence on the firm's financial performance.

The results presented in table 5 indicate that there is a negative association between the Shannon index for ethnic minorities (SIE_{t-2}) and the return on assets (ROA). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis. The findings suggests that a 1-unit increase the Shannon index for ethnic minorities is associated with a reduction of 0.106943 in the return of assets for the firms in the sample, holding all other factors constant. This implies that a higher Shannon index for ethnic minorities in the board of directors may have a negative influence on the firm's financial performance.

Finally, the results presented in table 6 indicate that there is a negative association between the Percentage of ethnic minorities in the board of directors (PE_{t-2}) and the return on assets (ROA). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis. The findings suggest that a 1-unit increase in the Shannon index for ethnic minorities is associated with a reduction of 0.106943 in the return of assets for the firms in the sample, holding all other factors constant. This implies that a higher percentage of ethnic minorities representation in the board of directors may have a negative influence on the firm's financial performance.

9.2.2 Hypothesis 2

Hypothesis 2 is that there is an increasing positive association between gender diversity in the board of directors and the firm's financial performance. This will be examined utilizing the regression model presented in section 7

In the below Tabel 7, the results of the model 2 is presented using the OLS regression[33]

[33] The implication of the Ordinary least squares (OLS) model is presented in appendix H

Table 7: OLS regression for model 2

	Estimate	Std. error	T value	P value	
Intercept	8.7000943	1.1475612	7.5814	<0.0000000000000002	***
$Dwomen1_{t-2}$	-0.3085484	0.1795884	-1.7181	0.0858218	.
$Dwomen2_{t-2}$	-0.6079275	0.2402380	-2.5305	0.0114094	*
$Dwomen3_{t-2}$	0.0085837	0.4014058	0.0214	0.9829399	
$Log(SF_{t-1})$	-1.19334140	0.1029021	-18.7889	<0.00000000000000022	***
$Log(BS_{t-1})$	--0.2695474	0.0813379	-3.3139	0.0009243	***
$Log(FA_{t-1})$	-0.2113080	0.0878773	-2.4046	0.0162150	***
$Log(REV_{t-1})$	2.0250319	0.1013025	19.9899	<0.00000000000000022	***
AE_{t-1}	0.1744018	0.0789107	-2.2101	0.0271268	*
$Log(ET_{t-1})$	-0.0438647	0.0896439	-0.4893	0.6246284	
$Log(AS_{t-1})$	0.9475470	0.0813379	11.4509	<0.00000000000000022	***
IN_{t-1}					.
Observation			7579		
R^2			0.1124		
Adjusted R^2			0.1105		
Residual Std. Error			6.817		
F Statistics			59.84 *** (df = 16;7559)		

The results presented in table 7 indicate that there is a non-significant association between the presence of 1 woman in the board of directors ($Dwomen1_{t-2}$) and the return on assets (ROA). Furthermore, the results indicate that there is a significant negative association between the presence of 2 women in the board of directors ($Dwomen2_{t-2}$) and the return in assets. Additionally, the results from table 7 indicate that there is a non-significant association between the presence of 3 or more women in the board of directors ($Dwomen3_{t-2}$) and the return of the assets. This is evident from the variables $Dwomen1_{t-2}$ and $Dwomen3_{t-2}$ having p-values that are above the 5% significance level, and $Dwomen2_{t-2}$ having a p-value that are below the threshold. Given that the coefficient of $Dwomen2_{t-2}$ is negative, this suggests that the association is negative, however, it also indicates that it is not possible to confirm the null hypothesis. The findings suggest that a 1-unit increase in either $Dwomen1_{t-2}$ or $Dwomen3_{t-2}$ is not associated with a change in the return in assets, holding all other factors constant. However, a 1-unit increase in the

$Dwomen2_{t-2}$ is associated with a reduction of 0.6079275 in the return of assets, holding all other factors constant.

The above examination indicates that the critical mass theory does not apply to women in this specific context, since the quantitative findings are not consistent with its theoretical predictions.

9.2.3 Hypothesis 3

Hypothesis 3 is that there is an increasing positive association between ethnic minority diversity in the board of directors and the firm's financial performance. This will be examined utilizing the regression model presented in section 7

In the below Tabel 8, the results of the model 3 is presented using the OLS regression[34]

Table 8: OLS regression for model 3

	Estimate	Std. error	T value	P value	
Intercept	8.0591074	5.577978	3.3325	0.0008649	***
$Detnicity1_{t-2}$	-0.629275	0.426169	-1.4766	0.1398283	
$Detnicity2_{t-2}$	-0.958504	2.414410	-0.3980	0.6913838	
$Detnicity3_{t-2}$	-0.484049	2.792156	-0.1734	0.8623727	
$Log(SF_{t-1})$	-1.924232	0.115717	-16.6287	<0.00000000000000022	***
$Log(BS_{t-1})$	-0.275666	0.082382	-3.3462	0.0008233	***
$Log(FA_{t-1})$	-0.219933	0.094635	-2.3240	0.0201513	*
$Log(REV_{t-1})$	2.054239	0.116200	17.6785	<0.00000000000000022	***
AE_{t-1}	0.187797	0.087088	-2.1564	0.0310826	*
$Log(ET_{t-1})$	-0.071947	0.098891	-0.7275	0.4669152	
$Log(AS_{t-1})$	0.969080	0.092697	10.4542	<0.00000000000000022	***
IN_{t-1}					
Observation			7579		
R^2			0.1119		
Adjusted R^2			0.11		
Residual Std. Error			6.819		
F Statistics			59.51*** (df = 16;7559)		

The results presented in table 8 indicate that there is a non-significant association between the presence of 1 ethnic minority in the board of directors ($Detnicity1_{t-2}$) and the return on

[34] The implication of the Ordinary least squares (OLS) model is presented in appendix H

assets (ROA). Furthermore, the results indicates that there is a non-significant association between the presence of 2 ethnic minorities in the board of directors ($Detnicity2_{t-2}$) and the return on assets. Additionally, the results from table 8 indicates that there is a non-significant association between the presence of 3 or more ethnic minorities in the board of directors ($Detnicity3_{t-2}$) and the return on assets. This is evident from the variables, $Detnicity1_{t-2}$, $Detnicity2_{t-2}$ and $Detnicity3_{t-2}$ having p-values that are above the 5% significance level. As a result, it is not possible to confirm the null hypothesis. The findings suggests that a 1-unit increase in either $Detnicity1_{t-2}$, $Detnicity2_{t-2}$ or $Detnicity3_{t-2}$ is not associated with a change in the return in assets, holding all other factors constant.

The above examination indicates that the critical mass theory does not apply to ethnic minorities in this specific context, since the quantitative findings are not consistent with its theoretical predictions.

9.3 ROBUSTNESS TEST [35]

In the following section a robustness test for the hypothesis is performed. This is done by estimating a regression model using the same variables, except the independent variable will be the return on equity (ROE) as presented in section 7.

9.3.1 Hypothesis 1

Hypothesis 1 is that there is a positive association between board diversity, in terms of gender and ethnic minorities, and the firm's financial performance. This will be examined utilizing the regression models presented in 7 except for a change in the independent variable.

In the below table 9, table 10, table 11 and table 12 the results of the model 1aa, model 1ab, model 1ba and model 1bb respectively, are presented using the OLS regression[36]

[35] The diagnostics of whether they uphold the OLS assumptions are presented in appendix J
[36] The implication of the Ordinary least squares (OLS) model is presented in Appendix H

Table 9: Robustness test for model 1aa

	Estimate	Sdt. Error	T value	P value	
Intercept	21.02807	2.09190	10.0522	<0.0000000000000002	***
SIK_{t-2}	-0.83503	0.18682	-4.4697	0.000007947	***
$Log(SF_{t-1})$	-4.78004	0.23305	-20.5110	<0.00000000000000022	***
$Log(BS_{t-1})$	-0.52262	0.18038	-2.8974	0.003774	**
$Log(FA_{t-1})$	-0.54672	0.20047	-2.7271	0.006404	**
$Log(REV_{t-1})$	5.90050	0.23029	25.511	<0.00000000000000022	***
AE_{t-1}	0.42598	0.17115	2.4889	0.012837	*
$Log(ET_{t-1})$	-0.91167	0.20253	-4.5013	0.000006855	***
$Log(AS_{t-1})$	1.63819	0.18943	8.6479	<0.00000000000000022	***
IN_{t-1}					**
Observations	7579				
R^2	0.1421				
Adjusted R^2	0.1405				
Residual Std. Error	15.47				
F Statistics	89.48 *** (df = 14; 7561)				

Table 10: Robustness OLS regression for model 1ab

	Estimate	Sdt. Error	T value	P value	
Intercept	21.10292	2.12913	9.9115	<0.0000000000000002	***
PK_{t-2}	-0.78902	0.17939	-4.3985	0.000011051	***
$Log(SF_{t-1})$	-4.77753	0.23355	-20.4560	<0.00000000000000022	***
$Log(BS_{t-1})$	-0.55020	0.18023	-3.0528	0.002275	**
$Log(FA_{t-1})$	-0.53897	0.20086	-2.6833	0.007305	**
$Log(REV_{t-1})$	5.88533	0.23112	25.4644	<0.00000000000000022	***
AE_{t-1}	0.42612	0.17117	2.4895	0.012814	*
$Log(ET_{t-1})$	-0.91747	0.20259	-4.5287	0.000006025	***
$Log(AS_{t-1})$	1.64003	0.18943	8.6577	<0.00000000000000022	***
IN_{t-1}					**
Observations	7579				
R^2	0.1418				
Adjusted R^2	0.1403				
Residual Std. Error	15.48				
F Statistics	89.26 *** (df = 14; 7561)				

Table 11: Robustness regression for model 1ba

	Estimate	Std. error	T value	P value	
Intercept	21.57878	2.16401	9.9717	<0.0000000000000002	***
SIE_{t-2}	-0.28538	0.13017	-2.1924	0.028382	*
$Log(SF_{t-1})$	-4.76432	0.23357	-20.3977	<0.00000000000000022	***
$Log(BS_{t-1})$	-0.52268	0.18029	-2.8991	0.003753	**
$Log(FA_{t-1})$	-0.57426	0.20133	-2.8524	0.004351	**
$Log(REV_{t-1})$	6.03943	0.22913	26.3586	<0.00000000000000022	***
AE_{t-1}	0.47558	0.17045	2.7902	0.005281	**
$Log(ET_{t-1})$	-1.02137	0.20223	-5.0506	0.0000004508	***
$Log(AS_{t-1})$	1.73126	0.18851	9.1859	<0.00000000000000022	***
IN_{t-1}					**
Observation			7579		
R^2			0.1401		
Adjusted R^2			0.1385		
Residual Std. Error			15.49		
F Statistics			88.02 *** (df = 14;7561)		

Table 12: Robustness OLS regression for model 1bb

	Estimate	Std. error	T value	P value	
Intercept	21.58303	2.16274	9.9795	<0.00000000000000022	***
PE_{t-2}	-0.27715	0.11559	-2.3976	0.016526	*
$Log(SF_{t-1})$	-4.76354	0.23355	-20.3962	<0.00000000000000022	***
$Log(BS_{t-1})$	-0.52897	0.18025	-2.9346	0.003350	**
$Log(FA_{t-1})$	-0.57391	0.20133	-2.8506	0.004376	**
$Log(REV_{t-1})$	6.03851	0.22910	26.3570	<0.00000000000000022	***
AE_{t-1}	0.47228	0.17048	2.7704	0.005613	**
$Log(ET_{t-1})$	-1.02139	0.20222	-5.0508	0.0000004504	***
$Log(AS_{t-1})$	1.72830	0.18845	9.1712	<0.00000000000000022	***
IN_{t-1}					**
Observation	7579				
R^2	0.1401				
Adjusted R^2	0.1385				
Residual Std. Error	15.49				
F Statistics	88.01*** (df = 14;7561)				

The results presented in 9 indicate that there is a negative association between the Shannon index for gender (SIK_{t-2}) and the return on equity (ROE). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis. The findings suggest that a 1- unit increase in the Shannon index for gender is associated with a reduction of 0.83503 in the return of equity for the firms in the sample, holding all other factors constant. This implies that a higher Shannon index for women in the board of directors may have a negative influence on the firm's financial performance.

Similarly, the results presented in table 10 indicate that there is a negative association between the percentage of women in the board of directors (PK_{t-2}) and the return on equity (ROE). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis The findings suggests that a 1-unit increase in the percentage of women is associated with a reduction of 0.78902 in the

return of equity for the firms in the sample, holding all other factors constant. This implies that a higher percentage of women representation in the board of directors may have a negative influence on the firm's financial performance

The results presented in table 11 indicate that there is a negative association between the Shannon index for ethnic minorities (SIE_{t-2}) and the return on equity (ROE). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis. The findings suggest that a 1-unit increase the Shannon index for ethnic minorities is associated with a reduction of 0.28538 in the return of equity for the firms in the sample, holding all other factors constant. This implies that a higher Shannon index for ethnic minorities in the board of directors may have a negative influence on the firm's financial performance

Finally, the results presented in table 12 indicate that there is a negative association between the Percentage of ethnic minorities in the board of directors (PE_{t-2}) and the return on equity (ROE). This is evident from the variable's p-value, which is below the 5% significance level, thereby indicating statistical significance. However, given that the coefficient for the variable is negative, it is not possible to confirm the null hypothesis. The findings suggest that a 1-unit increase in the Shannon index for ethnic minorities is associated with a reduction of 0.27715 in the return of equity for the firms in the sample, holding all other factors constant. This implies that a higher percentage of ethnic minorities representation in the board of directors may have a negative influence on the firm's financial performance.

These results from the above examinations are the same as in the main regression.

9.3.2 Hypothesis 2

Hypothesis 2 is that there is an increasing positive association between gender diversity in the board of directors and the firm's financial performance. This will be examined utilizing the regression model presented in 7 except for a change in the independent variable.

In the below Tabel 13 the results of the model 2 is presented using the OLS regression[37]

[37] The implication of the Ordinary least squares (OLS) model is presented in appendix H

Table 13: Robustness OLS regression for model 2

	Estimate	Std. error	T value	P value	
Intercept	22.07561	2.08905	10.5673	<0.00000000000000022	***
$Dwomen1_{t-2}$	-1.65769	0.40985	-4.0446	0.000052923	***
$Dwomen2_{t-2}$	--2.03542	0.53936	-3.7737	0.000162	***
$Dwomen3_{t-2}$	-0.92937	0.86958	-1.0688	0.285215	
$Log(SF_{t-1})$	-4.78328	0.23352	-20.4838	<0.00000000000000022	***
$Log(BS_{t-1})$	-0.48337	0.19110	-2.5818	0.009847	**
$Log(FA_{t-1})$	-0.54259	0.20071	-2.7033	0.006881	**
$Log(REV_{t-1})$	5.90202	0.23095	25.5558	<0.00000000000000022	***
AE_{t-1}	0.42383	0.17115	2.4763	0.013298	*
$Log(ET_{t-1})$	-0.91017	0.20265	-4.4914	0.000009847	***
$Log(AS_{t-1})$	1.63923	0.18958	8.6468	<0.00000000000000022	***
IN_{t-1}					
Observation			7579		
R^2			0.1423		
Adjusted R^2			0.1405		
Residual Std. Error			15.47		
F Statistics			78.39*** (df = 16;7559)		

The results presented table 13 indicate that there is a negative significant association between the presence of 1 woman in the board of directors ($Dwomen1_{t-2}$) and the return on equity (ROE). Furthermore, the results indicate that there is a significant negative association between the presence of 2 women in the board of directors ($Dwomen2_{t-2}$) and the return on equity Additionally, the results from table 13 indicate that there is a non-significant association between the presence of 3 or more women in the board of directors ($Dwomen2_{t-2}$) and the return on equity. This is evident from the variable $Dwomen3_{t-2}$ having a p-value that are above the 5% significance level, $Dwomen1_{t-2}$ and $Dwomen2_{t-2}$ having p-values that are below the threshold. Given that the coefficient of$Dwomen1_{t-2}$ and $Dwomen2_{t-2}$ are negative, it suggests that the association is negative, however, it also indicate that it is not possible to confirm the hypothesis. The findings suggest that a 1-unit increase in $Dwomen3_{t-2}$ is not associated with a change in the return on equity, holding all other factors constant. However, a 1-unit increase in $Dwomen1_{t-2}$ and $Dwomen2_{t-2}$ are associated with a reduction of 1.65769 and 2.03542 respectively in the return of equity,

holding all other factors constant .These results are on the other hand, not the same as in the main regression, but concludes the same.

9.3.3 Hypothesis 3

Hypothesis 3 is that there is an increasing positive association between ethnic minority diversity in the board of directors and the firm's financial performance. This will be examined utilizing the regression model presented in section 7 except for a change in the independent variable.

In the below Tabel 14, the results of the model 3 is presented using the OLS regression[38]

Table 14: Robustness OLS regression for model 3

	Estimate	Std. error	T value	P value	
Intercept	21.636051	5.857605	3.6937	0.0002226	***
$Detnicity1_{t-2}$	-1.601928	0.968328	-1.6543	0.0981031	.
$Detnicity2_{t-2}$	0.021011	5.485951	0.0038	0.9969443	
$Detnicity3_{t-2}$	-1.532675	6.344253	-0.2416	0.8091084	
$Log(SF_{t-1})$	-4.763195	0.262930	-18.1158	<0.00000000000000022	***
$Log(BS_{t-1})$	-0.516879	0.187186	-2.7613	0.0057710	**
$Log(FA_{t-1})$	-0.576132	0.215027	-2.6793	0.0073925	**
$Log(REV_{t-1})$	6.038539	0.264026	22.8710	<0.00000000000000022	***
AE_{t-1}	0.474004	0.197878	-2.3954	0.0166249	*
$Log(ET_{t-1})$	-1.020639	0.224697	-4.5423	0.0000056507002859228	***
$Log(AS_{t-1})$	1.731843	0.210624	8.2224	0.0000000000000002328	***
IN_{t-1}					
Observation			7579		
R^2			0.1402		
Adjusted R^2			0.1383		
Residual Std. Error			15.49		
F Statistics			77.01*** (df = 16;7559)		

[38] The implication of the Ordinary least squares (OLS) model is presented in appendix H

The results presented in table 14 indicate that there is a non-significant association between the presence of 1 ethnic minority in the board of directors ($Detnicity1_{t-2}$) and the return on equity (ROE). Furthermore, the results indicates that there is a non-significant association between the presence of 2 ethnic minorities in the board of directors ($Detnicity2_{t-2}$) and the return on equity. Additionally, the results from table 14 indicates the that there is a non-significant association between the presence of 3 or more ethnic minorities in the board of directors ($Detnicity3_{t-2}$) and the return on equity. This is evident from the variables, $Detnicity1_{t-2}$, $Detnicity2_{t-2}$ and $Detnicity3_{t-2}$ having p-values that are above the 5% significance level. As a result, it is not possible to confirm the null hypothesis. The findings suggests that a 1-unit increase in either $Detnicity1_{t-2}$, $Detnicity2_{t-2}$ or $Detnicity3_{t-2}$ is not associated with a chang in the return in assets, holding all other factors constant. These results from the above examination conclude the same as in the main regression.

10 DISCUSSION

In the following section the discussion of the results of the quantitative examination. Firstly, the representativeness of the data that has been utilized will be discussed, followed by an examination of the causality of the results.

10.1 GENERALIZATION OF THE DATA

As previously presented in section???, the data selected is limited to a specific research project. In order to create the cohort, which has been examined in the quantitative examination, a series of filtering processes were performed. Consequently, the filtering process may have an impact on the representativeness of the data.

First of all, the study is limited to firms that were registered in Denmark from 2017 to 2019. In addition, where the firms that are included in the cohort have been further restricted to, that they have to, firms who have been operating in all of the consecutive years. This has been done to ensure that it was possible to include a 2 lagged version of the diversity measurements. Whereas this method can safeguard temporal consistency in the cohort, it also has the possibility to limit the generalization ability of the findings. By only including firms that are registered in the Danish CVR-system and to this specific time period, the findings of the quantitative examination are not directly transferrable to firms in other countries in this specific time frame. Put another way, in other countries the link between board diversity and firm performance could differ due to the cultural context, variation in the economic structure or legislation.

The cohort has been further restricted to only firms where the board members are registered with a CPR-number. Firms where the data was not available for all members of the board have been omitted from the cohort. This choice was made to ensure the identifiability of the board members in the study, which was essential in regard to ensure gender and ethnicity of the individual and therefor creation of the diversity measurements. However, this may have excluded firms who have board members that are located in other countries, and therefore not registered in the Danish CPR-system.

The cohort has been further restricted to include only individuals where the gender was identifiable in the register data. This was chosen to ensure a precise gender classification and circumvent situations where gender had to be assessed. The cohort has, moreover, been restricted so that only individuals where the ethnicity was identifiable in the register date. This was also done to ensure the accuracy of the ethnicity classification and avoid situations where the ethnicity must be inferred. This choice was made to ensure the reliability of the board diversity measures in the dataset. However, this may also lead to a limitation of the scope of the sample, as firms with missing ethnicity and gender data were not included in the final cohort.

Likewise, the cohort has been limited to include only firms that are limited liability firms. This choice was made due to the fact, that in the established literature, the focus is mainly publicly listed firms, which are included in their cohort (Rose, 2007) (Campbell & Mínguez-Vera, 2008). However, due to restriction on data availability it was not feasible to replicate the studies in the established literature. As result the limited liability firms were selected as an appropriate approximation of the public listed firms. This approximation could have an impact on the comparability of the results of the quantitative examination, as the focus in these studies are on the public listed firms. An additional reason is to the fact that it is assumed that public listed firms operate under other market pressures than limited liability firms and are subjected to other government regulations.

Additionally, the cohort has been limited to include firms where at least 3 board seats are available in the board of directors. The underlying idea behind imposing these criteria are the assumption that there is a requirement of 3 individuals in the board of directors, before the diversity effects can be observed. The criteria was intended to guarantee that diversity effects, such as group decision making and group interaction develop within the structure of the board.

Lastly, the cohort was limited to firms that did not change industries within the specific period. This is based on the perspective that there can be substantial barriers of entry for firms that transit from one industry to another. Consequently, this decision has been made to minimize potential data disturbances from firms that undergo these forms of transitions. However, this has the potential to filter firms that are operating industries that are in constant change, and firms that are undergoing structural changes.

In summary, it can be argued that the subsequent filtering was necessary to guarantee the reliability of the data and temporal consistency, but they do impose a series of limitations on the generalization ability of the data. Moreover, by omitting firms based on the before mentioned method there is risk of selection bias, which limits the representation of the data and the ability to make generalizing conclusions about the results.

10.2 RESULTS OF THE EXAMINATION

The results of the quantitative examination indicates that there could be a negative association between a higher level of diversity, measured as the Shannon index, and the percentage of women and ethnic minorities in the board of directors, and the firm's financial performance. From a resource dependence theory perspective, a diverse board of directors can theoretically improve the firm's financial performance through external linkage to its environment which is a source of good cooperate governance. However, the negative association between board diversity in terms of gender and ethnic minority diversity may imply that the firms are unable to leverage the potential benefits. The negative association may also indicate that the skills and network of the diverse members of the board are misaligned with the firm's environment, which prevent them from addressing the strategic needs of the firm effectively.

The negative association between the percentage of ethnic minorities and women in the board of directors and the return of assets can, from a human capital theory perspective, be explained by the trouble of integrating skills into a firm. The skills of the diverse board may not be aligned with the skills that are needed in the specific industry. This may be even more evident in highly specialized industries where industry experience is essential to understand the inner working mechanisms of the firm. This could explain the negative association between diversity and the return of assets.

According to the critical mass and token theory and the results of the quantitative examination, it is not possible to apply women or ethnic minorities into this specific context. However, in the quantitative examination of the women it was found that the presence of 2 women had a negative association with the return on assets. This could, from a critical mass and token theory perspective, be due to the fact, that the presence of 2 women in the board of directors may not be enough for them to not be seen as the "token". Kanter argues, that the tokenism theory indicates the barriers that are related to being the token, is often intensified in environments where their differences, such as gender, are detectible and in environments

where they are underrepresented. Therefore, the possible negative association between 2 women in the board of directors and the return on assets may be amplified due to their gender-based deviation from the majority. Moreover, uniform groups tend to outperform non-uniform groups, where the inclusion of 2 women in the board of directors may disrupt the otherwise uniform group structure of the board, and thereby influence the performance negatively. From the quantitative examination it appears that the inclusion of 3 or more women in the board of directors have no significant impact on the return on assets. This may be explained by the emergence of a more balanced group, when 3 or more women are in the board of directors. In the balanced group there is a possibility of the forming of new sub-groups, which leads to the potential making of new "tokens" within the sub-groups. These may lead to the creation of dysfunction in the smaller group dynamic, which can influence the efficiency of the board's decision making. This may ultimately limit the potential influence of the new tokens of the sub-groups on the firm's financial performance, create a source of dysfunctional group dynamics and hinder the work in the board of directors.

10.3 CAUSALITY AND ENDOGENEITY

From the quantitative examination alone, is it not possible to determine if a causal relationship exists between board diversity and firms' financial performance. In order to determine if there is a causal relationship between two variables it is necessary to hold every other factor there might impact the variables constant (Wooldridge, 2016). As it has been determined by the quantitative examination, the R^2 for the regressions implies can only explain a relatively small part of the variation of the dependent variable. This means that there are facets of the relationship which cannot be explained by the explanatory variables included in the regression model. This also implies that there are factors which impact the dependent variables that cannot be held constant. Accordingly, a causal relationship between board diversity and firms financial performance, cannot be recognized solely founded on the quantitative examination.

From a theoretical perspective, the human capital theory and the resource dependence theory suggests that there is an association between board diversity and firm performance. The resource dependence theory suggests that the board of directors is the primary linkage between the firm and its environment. The composition of the board and the human capital that the members of the board possess is important for the firm's financial performance. The human capital theory suggests that the abilities of individuals can influence the performance

of the firm, yet the influence of various forms of human capital will not impact performance equally. Moreover, the critical mass theory suggests that there is a positive association between an increasing diversity in the board of directors, due to the changing of the group dynamic of the board.

From the empirical literature in section 3 it is evident that there are mixed results to support whether there a link exists between diversity and firm performance. Erhard et al. (2003) found that there is a positive link between board diversity and firm performance, arguing that diversity is fostering a better environment for innovation, creativity and for decision making. The positive link was also found by Carter et al. (2010). On the other hand, Comi et al. (2017) found a negative link between women in the board of directors and firm performance. While others has found there is no significant link between firm performance and board diversity (Rose, 2007).

Moreover, in section 3 And section 6 It is mentioned that one of the pression issues within this part of the literature, is the endogeneity problem. One of the raised concerns is the endogeneity between diversity in the board of directors and firm performance. This may be, due to the fact that firm performance can both impact the incentive of diverse individuals joining the specific board and the motivations of the firms of hiring diverse directors (Liu et al., 2016). Thus, serval studies have attempted to counter the revers causality problem. In the literature, authors has found different results when trying to address the endogeneity problem using various methods. For example, Liu et al. (2016) employ a series of different econometric techniques to address the endogeneity problem, including instrumental variables, to isolate the influence of diversity in the board directors. Other studies utilize a lagged version of the diversity variable to address the fact that board characteristics may need time to impact the firm's financial performance (Labelle et al., 2015). Despite different methodologies to address the endogeneity problem, the empirical results continue to be mixed. Some studies find that when controlling endogeneity, the impact is neutral or negative, when performed under specific circumstances (Bennouri et al., 2018). Other studies, however, find that when taking endogeneity into account they find a positive impact of diversity's influence on firm performance (Adams & Ferreira, 2009). Moreover, Adams and Ferreira (2009) argue that the published results are often not robust when taking endogeneity into account. The mixed results suggest that there is a complex relationship between firm performance and diversity, and not addressing it could lead to spurious correlation between board diversity and firm performance.

From the results of the quantitative examination, it can be inferred that board diversity in terms of gender and ethnic minorities both have an negative association with the firm's financial performance. Moreover, the results suggested that the critical mass theory does not apply either women or ethnic minorities in this specific context. Both results were reinforced by the results of the robustness test.

Regardless of the finding, it is difficult to irrefutably determine a causal relationship. Whilst it is possible to address endogeneity using different method such, it is improbable that whey can address all variables and factors that may influence the firm's financial performance. Consequently, although the quantitative examination may suggest that there is a negative association between gender and ethnic minority diversity and the firm's financial performance, the endogeneity problem and the mixed empirical results in the literature, implies that the relationship between firm performance and board diversity is complex by nature. Thus, it is assessed that the results of the quantitative examination do not suggest causality. Thus, the results of the quantitative examination do not suggest that there is a causal relationship between board diversity and the firm's financial performance.

11 CONCLUSION

The overall purpose of the study has been to examine to what extent diversity in boards correlates with the financial performance of firms. Within the literature there is a lack of consensus regarding if there is a positive, negative or non-significant association between diversity in the board of directors and the firm's financial performance. Moreover, difference in the definition of diversity further complicates the association. This view is said to be attributed to the difference in methodological considerations, a lack of formalization of a primary theoretical framework and the difference in the definition of diversity further complicates a clear direction of the results.

From the empirical findings in the established literature, it is suggested that there is in fact a positive association between the firm's financial performance and board diversity, in terms of gender and ethnic minority diversity.

Regarding the resource dependence theory and the human capital theory it is suggested that there may be a positive association between the board diversity and the firm's financial performance. Meanwhile the critical mass and token theory also suggest that there may be a increasing positive association.

Based on the quantitative examination several empirical conclusions can be made. The results indicate that there is a negative significant association between the diversity measurement in the form of the Shannon index and the percentage of women or ethnic minorities, and the firm's financial performance. In addition, the analysis indicates a non-significant association between an increase in diversity, in terms of gender and ethnic minorities, and the firm's financial performance.

However, several reservations must be made in relation to the results. The cohort may e.g. be subject to selection bias due to the data filtering process. The filtering process may therefore limit the examinations' ability to produce generalizable results. In addition, the inherent endogeneity- and causality problem and the inability to completely account for this in the data, emphasizes the complexity of the relationship between firm's financial performance and board diversity. This, in combination with the inability to make a direct comparison of the results with the literature due to approximation of firm structure, further substantiates the difficulties of making a definitive definition of their relationship. On this basis it appears to be impossible to either confirm or reject my problem statement.

12 LITERATURE

Adam Petrie (n.d.), 'VIF'. Found the 29. September 2024. URL: https://www.rdocumentation.org/packages/regclass/versions/1.6/topics/VIF

Adams, M. (2015). Board diversity: More than a gender issue?. *Deakin Law Review*, *20*(1), 123-152.

Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of financial economics*, *94*(2), 291-309.

Adams, R. B., De Haan, J., Terjesen, S., & Van Ees, H. (2015). Board diversity: Moving the field forward. *Corporate Governance-An International Review*, *23*(2), 77-82.

Bennouri, M., Chtioui, T., Nagati, H., & Nekhili, M. (2018). Female board directorship and firm performance: What really matters?. *Journal of Banking & Finance*, *88*, 267-291.

Bagh, T., Khan, M. A., Meyer, N., & Riaz, H. (2023). Impact of boardroom diversity on corporate financial performance. *Humanities and Social Sciences Communications*, *10*(1), 1-13.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management*, *17*(1), 99-120.

Baysinger, B. D., & Butler, H. N. (2019). Corporate governance and the board of directors: Performance effects of changes in board composition. In *Corporate Governance* (pp. 215-238). Gower.

Bernile, G., Bhagwat, V., & Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of financial economics*, *127*(3), 588-612.

Bhagat, S., & Black, B. (2001). The non-correlation between board independence and long-term firm performance. *J. CorP. l.*, *27*, 231.

Biasi, B., Deming, D. J., & Moser, P. (2021). *Education and innovation* (No. w28544). National Bureau of Economic Research.

Borjas, G. J. (2020). *Labor Economics* (8th ed.) McGraw-Hill Education

Boyd, C. (1996). Ethics and corporate governance: The issues raised by the Cadbury report in the United Kingdom. Journal of Business Ethics, 15(2), 167–182.

Bromiley, P. (1990). On the use of finance theory in strategic management. *Advances in strategic management*, *6*, 71-98.

Caballero, R. J. (2010). Creative destruction. In *Economic growth* (pp. 24-29). London: Palgrave Macmillan UK.

Cadbury, A. (1992). Cadbury report: The financial aspects of corporate governance. *Tech reprt, HMG, London*.

Campbell, K., & Mínguez-Vera, A. (2008). Gender diversity in the boardroom and firm financial performance. *Journal of business ethics*, *83*, 435-451.

Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate governance, board diversity, and firm value. *Financial review*, *38*(1), 33-53.

Carter, D. A., D'Souza, F., Simkins, B. J., & Simpson, W. G. (2010). The gender and ethnic diversity of US boards and board committees and firm financial performance. *Corporate Governance: An International Review*, *18*(5), 396-414.

Ciavarella, A. (2018). An analysis of the relationship between board diversity and firms performance in Europe.

Childs, S. & Krook, M. L. (2008) Critical Mass Theory and Women's Political Representation. *Political Studies, 56(3),* 725-736

Coleman, S. (2004). The liability of newness and small firm access to debt capital: Is there a link?. *Journal of Entrepreneurial Finance, JEF*, *9*(2), 38-59.

Comi, S., Grasseni, M., Origo, F., & Pagani, L. (2017). Where women make the difference. The effects of corporate board gender quotas on firms' performance across Europe. *University of Milan Bicocca Department of Economics, Management and Statistics Working Paper*, (367).

Cook, A., & Glass, C. (2015). Do minority leaders affect corporate practice? Analyzing the effect of leadership composition on governance and product development. *Strategic Organization*, *13*(2), 117-140.

Cressy, R. (2006). Why do most firms die young?. *Small business economics*, *26*, 103-116.

Deloitte (2022). Women in the boardroom – a global perspective 7th edition, 4-5

Dosi, G., & Grazzi, M. (2010). On the nature of technologies: knowledge, procedures, artifacts and production inputs. *Cambridge Journal of Economics*, *34*(1), 173-184.

Du Plessis, J., O'Sullivan, J., & Rentschler, R. (2014). Multiple layers of gender diversity on corporate boards: To force or not to force. *Deakin L. Rev.*, *19*, 1.

Duke (n.d.) 'Inflation adjustment'. Found 30. September. URL: Inflation adjustment of data for regression and forecasting (duke.edu)

Erhardt, N. L., Werbel, J. D., & Shrader, C. B. (2003). Board of director diversity and firm financial performance. *Corporate governance: An international review*, *11*(2), 102-111.

European Commission (2011). Strategy for equality between women and men 2010-2015, 3-40

European Commision (2022). Commission welcomes political agreement on Gender Balance on Corporate Boards: Found on the 9. September. URL: https://ec.europa.eu/commission/presscorner/detail/da/ip_22_3478

Farrell, K. A., & Hersch, P. L. (2005). Additions to corporate boards: The effect of gender. *Journal of Corporate finance*, *11*(1-2), 85-106.

Gabaldon, P., Mensi-Klarbach, H., & Seierstad, C. (2017a). Gender diversity in the boardroom: The multiple versions of quota laws in Europe. *Gender Diversity in the Boardroom: Volume 1: The Use of Different Quota Regulations*, 233-254.

Gabaldon, P., Mensi-Klarbach, H., & Seierstad, C. (2017b). Gender diversity in the boardroom: The multiple approaches beyond quota regulations. *Gender Diversity in the Boardroom: Volume 2: Multiple Approaches Beyond Quotas*, 261-284.

Gagliardi, N., Grinza, E., & Rycx, F. (2023). Workers' tenure and firm productivity: New evidence from matched employer-employee panel data. *Industrial Relations: A Journal of Economy and Society*, *62*(1), 3-33.

Gentry, R. J., & Shen, W. (2010). The relationship between accounting and market measures of firm financial performance: How strong is it?. *Journal of managerial issues*, 514-530.

Gibbons, R., & Waldman, M. (2004). Task-specific human capital. *American Economic Review*, *94*(2), 203-207.

Gyapong, E., Monem, R. M., & Hu, F. (2016). Do women and ethnic minority directors influence firm value? Evidence from post-apartheid South Africa. *Journal of Business Finance & Accounting*, *43*(3-4), 370-413.

Harder, M. (2023) *DI: Ligestilling opnås bedst i fælleskab og gennem frivillighed, ikke med kvoter oppefra.* Information. https://www.information.dk/debat/2023/02/di-ligestilling-opnaas-bedst-faellesskab-gennem-frivillighed-kvoter-oppefra

Harrison, D. A., & Sin, H. (2006). What is diversity and how should it be measured. *Handbook of workplace diversity,* 191-216

Hecklen, A., Scheel, A. F., & Munch, J. *Ekspert: Ministers forslag om flere kvinder i bestyrrelser får ingen betydning. (*2022, February 21). DR. https://www.dr.dk/nyheder/penge/ekspert-ministers-forslag-om-flere-kvinder-i-bestyrelser-faar-ingen-betydning

Hermalin, B, E., &Weisbach, M, S. (2003). Boards of Directors as an Endogenously Determined Institution: A Survey of the Economic Literature. *Economic Policy Review, 9(1),* 8-26

Hillman, A. J., Cannella, A. A., & Paetzold, R. L. (2000). The resource dependence role of corporate directors: Strategic adaptation of board composition in response to environmental change. *Journal of Management studies*, *37*(2), 235-256.

Hitt, M. A., Bierman, L., Shimizu, K., & Kochhar, R. (2001). Direct and Moderating Effects of Human Capital on Strategy and Performance in Professional Service Firms: A Resource-Based Perspective. *The Academy of Management Journal*, *44*(1), 13–28. https://doi.org/10.2307/3069334

Hsiao, C. (2022). *Analysis of panel data* (No. 64). Cambridge university press.

Jungmann, C. (2006). The effectiveness of corporate governance in one-tier and two-tier board systems–Evidence from the UK and Germany–.

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American journal of Sociology*, *82*(5), 965-990.

Kogut, B., Colomer, J., & Belinky, M. (2014). Structural equality at the top of the corporation: Mandated quotas for women directors. *Strategic Management Journal*, *35*(6), 891-902.

Kristie, J. (2011). The power of three. *Dir. Boards*, *35*(5), 22-32.

Labelle, R., Francoeur, C., & Lakhal, F. (2015). To regulate or not to regulate? Early evidence on the means used around the world to promote gender diversity in the boardroom. *Gender, Work & Organization*, *22*(4), 339-363.

Lincoln, A. (2023). Board Diversity and Inclusivity. In *Encyclopedia of Sustainable Management* (pp. 341-346). Cham: Springer International Publishing.

Liu, Y., Wei, Z., & Xie, F. (2014). Do women directors improve firm performance in China?. *Journal of corporate finance*, *28*, 169-184.

Johnson, D. D., Price, M. E., & Van Vugt, M. (2013). Darwin's invisible hand: Market competition, evolution and the firm. *Journal of economic behavior & organization*, *90*, S128-S140.

Low, D. C., Roberts, H., & Whiting, R. H. (2015). Board gender diversity and firm performance: Empirical evidence from Hong Kong, South Korea, Malaysia and Singapore. *Pacific-Basin Finance Journal*, *35*, 381-401.

Lærd Statistics (n.d.), 'Pearson Product-Moment Correlation'. Found the 10. September 2024 URL: Pearson Product-Moment Correlation - When you should run this test, the range of values the coefficient can take and how to measure strength of association. (laerd.com)

Majumdar, R. (2013). On the alternative proxies for estimating firm growth in empirical corporate finance literature: Evidence from Indian manufacturing sector.

Mallinguh, E., Wasike, C., & Zoltan, Z. (2020). The business sector, firm age, and performance: The mediating role of foreign ownership and financial leverage. *International Journal of Financial Studies*, *8*(4), 79.

Maznevski, M. L. (1994). Understanding our differences: performance in decision-making groups with diverse members. *Human relations*, *47*(5), 531-552.

Mensi-Klarbach, H., Leixnering, S., & Schiffinger, M. (2021). The carrot or the stick: Self-regulation for gender-diverse boards via codes of good governance. *Journal of Business Ethics*, *170*, 577-593.

Merendino, A., & Melville, R. (2019). The board of directors and firm performance: empirical evidence from listed companies. *Corporate Governance: The international journal of business in society*, *19*(3), 508-551.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of management review*, *21*(2), 402-433.

Mohamad Asri, M. N., Hashim, N. H., Mat Desa, W. N. S., & Ismail, D. (2018). Pearson Product Moment Correlation (PPMC) and Principal Component Analysis (PCA) for objective comparison and source determination of unbranded black ballpoint pen inks. *Australian Journal of Forensic Sciences*, *50*(4), 323-340.

NCL (n.d) 'Box and Whisker Plots'. Found 25. September 2024 URL: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/data-presentation/box-and-whisker-plots.html

Nygaard, K. (2011). Forced board changes: Evidence from Norway. *NHH Dept. of Economics Discussion Paper*, (5).

Penn State: Eberly College of Science (n.d.) '9.5 – Identifying Influential Data Points'. Found 25. September 2024 URL: https://online.stat.psu.edu/stat462/node/173/

Pfeffer, J., & Salancik, G. R. (1978). The external control. *New York*.

Raza, H. (2021a) 'Simple linear regression (cont) Multiple linear regression Lecture: 2'

Raza, H (2021a) 'Multiple Regression Analysis with Qualitative data: Binary (or Dummy) Lecture: 8'

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring organizational performance: Towards methodological best practice. *Journal of management*, *35*(3), 718-804.

Romer, C. D. (2008). Business cycles. *The concise encyclopedia of economics*, *26*, 2010.

Rose, C. (2007). Does female board representation influence firm performance? The Danish evidence. *Corporate governance: An international review*, *15*(2), 404-413.

Rossi, M. (2016). The impact of age on firm performance: A literature review. *Corporate Ownership & Control*, *13*(2), 217-223.

RStudio (n.d.) 'Cook's Distance'. Found 25. September 2024. URL: https://rpubs.com/DragonflyStats/Cooks-Distance

Shrader, C. B., Blackburn, V. B., & Iles, P. (1997). Women in management and firm financial performance: An exploratory study. *Journal of managerial issues*, 355-372.

Smith, N., Smith, V., & Verner, M. (2006). Do women in top management affect firm performance? A panel study of 2,500 Danish firms. *International Journal of productivity and Performance management*, *55*(7), 569-593.

Steffens, N. K., Shemla, M., Wegge, J., & Diestel, S. (2014). Organizational tenure and employee performance: A multilevel analysis. *Group & Organization Management*, *39*(6), 664-690.

Studenmund, A. H. (2014). *Using econometrics a practical guide*. Pearson education limited.

Supriyadi, T., & Terbuka, U. (2021). Effect of return on assets (ROA), return on equity (ROE), and net profit margin (NPM) on the company's value in manufacturing companies listed on the exchange Indonesia securities year 2016-2019. *International Journal of Economics, Business and Management Research*, *5*(04), 2021.

Taouab, O., & Issor, Z. (2019). Firm performance: Definition and measurement models. *European Scientific Journal*, *15*(1), 93-106.

Terjesen, S., Sealy, R., & Singh, V. (2009). Women directors on corporate boards: A review and research agenda. *Corporate governance: an international review*, *17*(3), 320-337.

Terjesen, S. & Sløk-Madsen, S, K. (2023). *Professorer: Kønskvoter tjener ingen – heller ikke kvinderne.* Politikken. https://politiken.dk/debat/debatindlaeg/art9202892/K%C3%B8nskvoter-tjener-ingen-%E2%80%93-heller-ikke-kvinderne?srsltid=AfmBOoo4Fm-WUSLPFhmXhDgskr3bJ6LIdAcJzJJicI9zWuOeB4zTJglN

Tyrefors, B., & Jansson, J. (2017). *Gender quotas in the board room and firm performance: evidence from a credible threat in Sweden*. SSRN.

Volonté, C., & Gantenbein, P. (2016). Directors' human capital, firm strategy, and firm performance. *Journal of Management & Governance*, *20*, 115-145.

Vora, G. (2019). Age of Firms: Irrelevance Proposition. *Modern Economy*, *10*(5), 1446-1478.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic management journal*, *5*(2), 171-180.

Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach* (Sixth). Cengage Learning

Yang, T., & Aldrich, H. E. (2017). "The liability of newness" revisited: Theoretical restatement and empirical testing in emergent organizations. *Social science research*, *63*, 36-53.

Østergaard, Christian R., Bram Timmermans, and Kari Kristinsson. "Does a different view create something new? The effect of employee diversity on innovation." *Research policy* 40.3 (2011): 500-509.

APPENDIX A

Table 15: Overview over the data included in the study

Name of the dataset	Number of variables	Number of observations
Udda2017	14	498738
Udda2018	14	4953626
Udda2019	17	5151891
Persbest2017	6	361965
Persbest2018	6	370985
Persbest2019	6	408819
Idan2017	18	4779886
Idan2018	18	4828869
Idan2019	18	4838658
Firm2017	29	811969
Firm2018	29	844144
Firm2019	29	869776
Fire2017	10	222716
Fire2018	10	225082
Fire2019	10	227931
Bef2017	17	4985738
Bef2018	17	4953626
Bef2019	20	5139949

APPENDIX B

In the following appendix the overview of the variables are presented as the symbol used in the study and its representation.

Table 16: Overview over the variables included in the study as a symbol

Symbol	representation
ROA_t	The return of assets in 2019
SIK_{t-2}	Shannon index for gender in 2017
PK_{t-2}	The percentage of women in the board of directors in 2017
SIE_{t-2}	The Shannon index for ethnic minorities in 2017
PE_{t-2}	The percentage of ethnic minorities in the board of directors in 2017
$Dwomen1_{t-2}$	Dummy variable indicating the presence of 1 woman in the board of directors
$Dwomen2_{t-2}$	Dummy variable indicating the presence of 2 women in the board of directors
$Dwomen3_{t-2}$	Dummy variable indicating the presence of 3 or more women in the board of directors
$Detnicity1_{t-2}$	Dummy variable indicating the presence of 1 ethnic minority in the board of directors
$Detnicity2_{t-2}$	Dummy variable indicating the presence of 2 ethnic minorities in the board of directors
$Detnicity3_{t-2}$	Dummy variable indicating the presence of 3 or more ethnic minorities in the board of directors
SF_{t-1}	The Firm Size in 2018
BS_{t-1}	The size of the firm's board in 2018
FA_{t-1}	Firm age in 2018
REV_{t-1}	The firm's revenue in 2018

AE_{t-1}	The average educational level of the firm's employees in 2018
ET_{t-1}	The average of the employee's tenure of the firm in 2018
AS_{t-1}	The average of the employee's salary of the firm in 2018
IN_{t-1}	The specific industry in 2018

APPENDIX C

In the following appendix the method used to compute the variables included in the study is presented. In addition, the datasets used to compute them will be presented.

Dependent variables:

Return on assets:

In the literature (Supriyadi &Terbuka, 2021), the return on asset (ROA) is calculated as follows:

Equation 7: Return on assets

$$ROA = \frac{net\ income}{total\ asset} * 100$$

In the study, the independent variable ROA has been calculated in line with this methodology. To calculate the dependent variable ROA the *FIRE*-dataset from all 6 years is utilized. The datasets contain 10 different variables, and it is chosen to utilize the following variables: *"cvrnr", "aare"* and *"at"*. In addition, an extra variable is created to be able to take into account the specific year the dataset is from. The variables from the *FIRE*-dataset are chosen for the following reason. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr[39]-system. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "aare" variable represents the net income of the specific firm. The "at" variable represents the total assets of the firm in ultimo, and the value is measured in 1.000 DKK. Additionally, to consider the inflation of the specific year, the financial data, the "aare" and "at" were adjusted to 2010 prices by dividing the financial measure by the CPI of the specific year, in regard to index 2010 CPI and multiply it by 100 (Duke, n.d.).

[39] CVR stands for "Det Centrale Virksomhedsregister" and is the Danish Stats register for information regarding firms in Denmark

Return on equity:

In the literature (Supriyadi &Terbuka, 2021), the return on equity (ROE) is calculated as follows:

Equation 8: return on equity

$$ROE = \frac{net\ income}{total\ equity} * 100\%$$

In this study, the independent variable ROE has been calculated in line with this methodology. To calculate the dependent variable ROE the FIRE-dataset from all 3 years is utilized. The dataset contain 10 different variables, and it is chosen to utilize the following variables: "*cvnr", "aare" and "gul".* In addition, and extra variable is created to be able to include the specific year the dataset is from. The variables from the *FIRE*-dataset are chosen for the following reason. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr[40]-system. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "aare" variable represents the net income of the specific firm. The "egul" variable represents the total equity of the firm in ultimo, and the value is measured in 1.000 DKK. Additionally, to consider the inflation of the specific year, the financial data, the "aare" and "egul" were adjusted to 2010 prices by dividing the financial measure by the CPI of the specific year, in regard to index 2010 CPI and multiply it by 100 (Duke, n.d.).

Explanatory variables:

Shannon index women:

From a theoretical standpoint the Shannon index ranges from 0 to positive infinity. The value of 0 will occur if everyone within the population is part of the same category, and the greater the population is distributed among more categories, the index value will rise and become larger. The Shannon index is bound by the number of categories included in its calculations. When the variable is binary, the index value for total equality would be 0.69. This, however, changes when more categories are made possible (Harrison and Sin, 2006).

The variable is calculated by using the following:

[40] CVR stands for "Det Centrale Virksomhedsregister" and is the Danish Stats register for information regarding firms in Denmark

Equation 9: Shannon index

Shannon index = - $\sum p_k * \ln(p_k)$,

where p_k is the proportion of population in the k category, and ln (p_k) is the natural logarithm of p_k(Harrison and Sin, 2006)

To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. This makes it possible to calculate the proportion of the individuals that are a board member and female. Due to the fact we are making a binary Shannin index, the diversity measurement are focused on the gender-based diversity. Therefore, the proportion of individuals that are a board member and male is also calculated.

When two proportions have been calculated the Shannon index for gender is created using equation 9

Shannon index ethnic minority:

From a theoretical standpoint the Shannon index ranges from 0 to positive infinity. The value of 0 will occur if everyone within the population is part of the same category, and the greater the population is distributed among more categories, the index value will rise and become larger. The Shannon index is bound by the number of categories included in its calculations. When the variable is binary, the index value for total equality would be 0.69. This, however, changes when more categories are made possible (Harrison and Sin, 2006).

The variable is calculated by using the following:

Equation 10: Shannon index

Shannon index = $- \sum p_k * \ln(p_k)$,

where p_k is the proportion of population in the k category, and $\ln(p_k)$ is the natural logarithm of p_k(Harrison and Sin, 2006)

To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not

0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. This makes it possible to calculate the proportion of the individuals that are a board member and ethnic minority. Due to the fact we are making a binary Shannin index, the diversity measurement is focused on ethnic minority-based diversity. Therefore, the proportion of individuals that have a Danish origin is also calculated. When two proportions have been calculated the Shannon index for ethnic minority is created using equation 10.

Percentage of women in the board of directors:

The percentage of female used as a measurement for demographic diversity in the board of directors is calculated by taking the total numbers of either females in the board of directors and divided by the total number of members of the board. To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is due to the fact that it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position

in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. Then a counting variable is created to count the total number of women in the board of directors, and the then it is divided with then total number of seats available in the board of directors. Then the result is multiplied by 100, and the variable has been created.

Percentage of ethnic minorities in the board of directors:

The percentage of female used as a measurement for demographic diversity in the board of directors is calculated by taking the total numbers of either females in the board of directors and divided by the total number of members of the board. To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no

cvrnr-number is included. The “type” variable represents the individual’s position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The “koen” variable represents a person’s gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The “IE_TYPE” variable represents a person’s ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the “pnr” as the merging variable. Then a counting variable is created to count the total number of ethnic minorities in the board of directors, and then it is divided with the total number of seats available in the board of directors. Lastly, the result is multiplied by 100, and the variable has been created.

DWoman1:

The dummy variable Dwoman1 indicates that there are 1 woman in the board of directors. To calculate the explanatory variable the “Persbest”-dataset and the “Bef”-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: “cvr”, “pnr” and “type” from the “Persbest”-dataset, and the variables “pnr”, “koen”, “IE_TYPE” from the “Bef”-dataset. The “pnr” variable represents a person’s personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The “cvrnr” variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The “type” variable represents the individual’s position

in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. In the newly merged dataset, a new variable is created that can count the total number of women in the board of directors in the specific firms based on the "cvnr". When this variable is created, subsequently the dummy variable is created. The dummy variable is operationalized so that if the value in the new variable, is equal to 1, then the dummy identifies it with the identifier 1. This is how the variable is created.

Dwomen2

The dummy variable Dwomen2 indicates that there are 2 women in the board of directors. To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so thet if there is a missing value they are filtered for. This is because it is not possible to identify the

firm if no cvrnr-number is included. The “type” variable represents the individual’s position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The “koen” variable represents a person’s gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The “IE_TYPE” variable represents a person’s ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the “pnr” as the merging variable. In the newly merged dataset, a new variable is created that can count the total number of women in the board of directors in the specific firms based on the “cvnr”. When this variable is created, subsequently the dummy variable is created. The dummy variable is operationalized so, that if the value in the new variable, is equal to 2, then the dummy identifies it with the identifier 1. This is how the variable is created

Dwomen3

The dummy variable Dwomen3 indicates that there are 3 or more women in the board of directors. To calculate the explanatory variable the “Persbest”-dataset and the “Bef”-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: “cvr”, “pnr” and “type” from the “Persbest”-dataset, and the variables “pnr”, “koen”, “IE_TYPE” from the “Bef”-dataset. The “pnr” variable represents a person’s personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The “cvrnr” variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The

variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. In the newly merged dataset, a new variable is created that can count the total number of women in the board of directors in the specific firms based on the "cvnr". When this variable is created, subsequently the dummy variable is created. The dummy variable is operationalized so that if the value in the new variable is equal to 3 or higher, then the dummy identifies it with the identifier 1. This is how the variable is created

Detnicity1:

The dummy variable Detnicity1 indicates that there are 1 ethnic minority in the board of directors. To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be

registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. In the newly merged dataset, a new variable is created that can count the total number of ethnic minorities in the board of directors in the specific firms based on"cvnr". When the variable is created, subsequently the dummy variable is created. the dummy variable is operationalized so that if the value of the new variable is equal to 1, then the dummy identifies it with the identifier 1. This is how the variable is created.

Detnicity2:

The dummy variable Detnicity2 indicates that there are 2 ethnic minority in the board of directors. To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is

registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. In the newly merged dataset, a new variable is created that can count the total number of ethnic minorities in the board of directors in the specific firms based on"cvnr". When the variable is created, subsequently the dummy variable is created. the dummy variable is operationalized so that if the value of the new variable is equal to 2, then the dummy identifies it with the identifier 1. This is how the variable is created.

Detnicity3:

The dummy variable Detnicity3 indicates that there are 3 or more ethnic minority in the board of directors. To calculate the explanatory variable the "Persbest"-dataset and the "Bef"-data from all 3 years are utilized. The datasets contain 6 variables and 17 to 20 variables respectively. To create the explanatory variable, it has been decided to utilize the following variables: "cvr", "pnr" and "type" from the "Persbest"-dataset, and the variables "pnr", "koen", "IE_TYPE" from the "Bef"-dataset. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr"

variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Then a counting variable is created to count the total number of available seats in the board of director for the specific firm. The "koen" variable represents a person's gender. If the value of the variable is equal 1 the person is identified as a male, and if the variable is equal to 0, the person is identified as a female. The "IE_TYPE" variable represents a person's ethnicity. If the value of the variable is equal to 2 the person is an immigrant who were born abroad. If the value of the variable is equal to 3 the person is a descendant of an immigrant and were born in Denmark. If the value of the variable is equal to 1 the person is a person with Danish origin. The ethnic minority variable is operationalized so that if the person has a value that is equal to 2 or 3, then the person is identified as an ethnic minority. When the two datasets have been operationalized, the datasets are merged using the "pnr" as the merging variable. In the newly merged dataset, a new variable is created that can count the total number of ethnic minorities in the board of directors in the specific firms based on"cvnr". When the variable is created, subsequently the dummy variable is created. the dummy variable is operationalized so that if the value of the new variable is equal to 3 or more, then the dummy identifies it with the identifier 1. This is how the variable is created.

Control variables:

Firm size:

In the literature, firm size is calculated as follows:

Equation 11: Firm Size

$$Firm\ size = \log(at)$$

In the study, the control variable firm size has been calculated with this methodology. To calculate the control variable firm size, the FIRE-dataset from all 3 years is utilized. The datasets contain 10 different variables, and it is chosen to utilize the following variables: *"cvrnr"* and *"at"*. In addition, an extra variable is created to be able to take into account the

specific year of the dataset. The variables from the FIRE-dataset are chosen due to the following reason. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system under. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "at" variable represents the total assets of the firm in ultimo, and the value is measured in 1.000 DKK.

Revenue:

In the study, the control variable revenue has been calculated with this methodology. To calculate the control variable revenue, the *"FIRE"*-dataset from all 3 year is utilized. The datasets contain 10 different variables, and it is chosen to utilize the following variables. "cvrnr" and "oms". In addition, an extra variable is created to be able to take into account the specific year of the dataset. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system under. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "oms" variable is the revenue of the firm, and the value is measured in 1.000 DKK.

Tenure:

In the study, the control variable tenure has been calculated with this methodology: To calculate the control variable tenure the *"IDAN"*-dataset from all 3 year is utilized. The datasets contain 18 different variables, and it is chosen to utilize the following variables: "pnr", "ansaar", "type_2008", "cvrnr" and "ansdage_2008". The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "ansaar" is the number of years the person have been employed the in the specific firm. The "TYPE_2008" variable represents the type of job. The object is to find the average tenure of employees, therefore, individuals with the "TYPE_2008" value "H" are included as this includes individuals where the employment is a full-time position. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals who have registered residence in Denmark is given. No pnr is the same and are utilized to identify citizens. The "ANSDAGE_2008" variable represents the number of days the individual has in the job in the specific year. The variable is operationalized so that if there are missing values or the

value is 0 they are filtered for. This is because it has been assessed that if the individuals have a missing value, it is not possible to determine whether they will contribute to the daily operations of the firm. In addition, if the individual has the value 0, this means that they have been 0 days in the job in the specific year and it is therefore unlikely that they contribute to the daily operations of the firm. To calculate the control variable, the people with no pnr and the firms with no cvnr are filtered away. The period in ansaar is limited to the period between 1980 and the specific year of the dataset. To ensure the quality of the data, the specific period is filtered, to only include the specific period. Afterward the number of years each person in the dataset has worked for the specific firm is calculated, subtracting the ansaar from the specific year of the dataset. Subsequently, the average tenure of the employees in the firms are calculated.

Industry:

In the literature, it has been suggested that a categorical variable can be made to identify the specific industry. In the study, the control variable for industry effects is a categorical variable, which thereafter will be turned into dummy's that represent the specific industry. To calculate the control variable, the *Firm*-dataset from all 3 years is utilized. The datasets contain 29 variables. To create the categorical variable, it has been decided to utilize the following variables: *"cvrnr"* and *"GF_GRO10_DB07".* The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "*GF_GRO10_DB07"* variable represents the Danish industry's classification system that are used to identify which industry a specific firm is linked to. The industry classifications represented in "*GF_GRO10_DB07"* is presented in table 17

Table 17:Industry classification values

Classification value	Industry
1	Agriculture, forestry and fishing
2	Industry, extraction of raw materials and business activities
3	Construction and civil engineering
4	Trade and transport etc..
5	Information and communication
6	Financing and insurance
7	Real estate transactions and rental
8	Business services
9	Public administration, education and health
10	Culture, leisure and other services
11	Unknown activity

If any of the firms in the study had the classification value 11, they were filtered out of the population.

Firm age:

Firm age, in this study, is defined as the time that has passed since the firm originally was registered as a legal entity. To calculate the control variable, firm age, the *FIRM*-dataset from all 3 years is utilized. The datasets contain 29. To calculate firm age, it has been decided to make use of the following variables: *"cvrnr"* and *"JUR_FRA_DATO"*. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The *"JUR_FRA_DATO"* variable represents the date a specific cvr-number has been created in the cvr-system. Under the Danish Corporate law, it is the firm's incorporation date that appears as starting data in the cvr-system. The *"JUR_FRA_DATO"* displays the precise date the firm was started, however, the objective is obtain the specific year the firm was established. To achieve this the R-library "lubridata" is utilized to isolate that from the *"JUR_FRA_DATA"* variable. Once the year the firm was established has been identified, it is subtracted from the year of the dataset and the firm age has been found.

Average educational level of employees:

The average educational level of employees is in this study defined as the average educational level of all employees there are in a full-time position. To calculate the control variable, the *Udda*-dataset and the *Idan*-dataset from all 3 years have been utilized. The *Udda*-datasets contain variables ranging from 14 to 17 depending on the specific dataset, and the *Idan*-datasets contain variables 18 variables. To calculate the variable, average educational level of employees, it has been chosen to make use of the following variables from the *Udda*-dataset: *"pnr"* and *"hfaudd"*. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals who have registered residence in Denmark is given. No pnr is the same and are utilized to identify citizens. The "hfaudd" variables represents an individual's highest completed education at a specific time. The variable contains a wide spectrum of educational levels, it has been chosen to operationalize it after the DISCED-15 classification, which contains 15 different subgroups of educational levels. It has been chosen to combine some of the subgroups given that they have been assessed to represent a similar educational level. The converted values have been recalibrated to be within a range of 1 to 8, where a value of 8 represents the highest educational level and 1 represents the lowest educational level. During the review of the "hfaudd" variable, some deviations from the DISCED-15 classification criteria was noted. Thus, all value variations in the variable were examined, and the deviations were considered in the code. The information table 18 represents how the operationalization of the educational level in respect to the DISCED-15 classification:

Table 18: Educational level and converted values

Educational level	**Converted value**
05: Pre-school education	1
10: Primary school	1
15: Preparotary education	1
20: Upper secondary education	2
25: Danish language education	2
29: Vocational basic course	2
30: Vocation education	3
35: Qualifying education programs	3
39: Employment-related education and training, AMU	3
40: Short higher education programs, KVU	4
50: Medium-length higher education programs, MVU	5
60: Bachelor degrees	6
70: Long higher education programs, LVU	7
80: Ph.d and doctoral degrees	8
90: Unknown	Deleted from the data

From the above table 18 It can be derived that individuals can have an unknow educational level, and individuals with this level has been filtered from the dataset. It has been chosen to utilize the following variables from the *Idan*-dataset: *"cvrnr", "pnr"* and *"TYPE_2008"*. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals who has registered residence in Denmark is given. No pnr is the same and are utilized to identify citizens. The "TYPE_2008" variable represents the type of job. The object is to find the average education level of employees, therefore, individuals with the "TYPE_2008" value "H" are included as this includes individuals where the employment is a full-time position.

The converted information from the *Udda*-dataset has then been merged into the *Idan*-dataset. Then the mean of the educational level for the employees for a specific firm is calculated based on the individuals affiliated with the firm and the mean of their educational level. In the new operationalized variable, the lowest level of education is 1 and the highest level of education is 8.

Average salary of the employees:

The average salary of the employees is in this study defined as the average salary of all employees who are in a full-time position. To calculate the control variable, the *Idan*-dataset from all 3 years is utilized. The *Idan*-datasets contain variables in a range from 15 to 18 depending on the specific dataset. To calculate the average salary of the employees it has been chosen to utilize the following variables: *"ANSDAGE_2008", "pnr", "TYPE_2008", "cvrnr"* and *"BREDT_LOEN_BELOEB".* The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No pnr is the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operated in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The "ANSDAGE_2008" variable represents the number of days the individual has in the job in the specific year. The variable is operationalized so that if there are missing values or the value is 0 they are filtered for. This is because it has been assessed that if the individuals have a missing value, it is not possible to determine whether they will contribute to the daily operations of the firm. In addition, if the individual has the value 0, this means that they have been 0 days in the job in the specific year and it is therefore unlikely that they contribute to the daily operations of the firm. The *"BREDT_LOEN_BELOEB"* variable represents income subjected to labor market contribution and income that are not subjected to labor market contribution included ATP and benefits. The variable is operationalized so that if there are missing values they are filtered for. The "TYPE_2008" variable represents the type of job. The object is to find the average salary of employees, therefore, individuals with the "TYPE_2008" value "H" are included as this includes individuals where the employment is their full-time position.

The operatized variables from the *Idan*-dataset have then been utilized to calculate the average salary of the employees. It is calculated based on the individuals affiliated with the firm, and the mean of all the employee's salary.

Board size:

In the literature, board size has been calculated as the total individuals seated in the board of directors of a specific firm. To calculate the control variable, board size, the "*Persbest"*-dataset from all 3 years is utilized. The datasets contain 3 different variables, where it is chosen to utilize the following variables: *"pnr", "cvrnr"* and *"type"*. The "pnr" variable represents a person's personal identification number. Pnr is a unique number that all individuals with registered residence in Denmark is given. No two pnr are the same and are used to identify citizens. The variable is operationalized so that if there is a missing value they are filtered for. This is because it is not possible to identify the individual if no pnr-number is included. The "cvrnr" variable represents the cvr-number which is a distinctive number by which a legal entity is registered in the cvr-system. It is required for all businesses operating in Denmark to be registered in the cvr-system, so that they can be identified by the public authorities. The variable has been operationalized so that if there is a missing value they are filtered for. This is due to the fact that it is not possible to identify the firm if no cvrnr-number is included. The "type" variable represents the individual's position in upper-management. The objective is to identify if the individual is in the board of directors or not. The variable is, therefore, operationalized so that if the individual is part of the board of directors, then the value of the variable is given 1 and if not 0. Afterwards a counting function is created to count the number of individuals who are identified as being part of the board of directors. This is done based on the specific firm, using the newly operationalized "type" variable and the "cvrnr" variable.

APPENDIX D

In following appendix, the assumptions for the Ordinary least squares model (OLS) are presented. The method that has been deemed the most appropriate to examine the link between firm performance and diversity in this study is multiple regression model. To estimate this model type OLS is used (Wooldridge, 2016). To ensure unbiased estimation of the regression model, the following assumptions must be met:

Assumption MLR 1: Linear in Parameters

In the population model, the relationship between the dependent- and the independent variable, is linear (Raza, 2021a) (Wooldridge, 2016), which can be expressed as follows:

Equation 12: MLR1

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + u$$

Assumption MLR 2: Random Sampling

It is assumed that the sample is stochastic, and consisting of n observations $\{(x_i, y_i): 1 = 1, 2, \ldots, 2\}$, which follows the population model presented in MLR 1 (Raza, 2021a) (Wooldridge, 2016)

Assumption MLR 3: Multicellularity

It is assumed that none of the explanatory variables in the stochastic sample are the same value, and none of the explanatory variables has a perfect linear relationship with each other (Raza, 2021a) (Wooldridge, 2016)

Assumption MLR 4: Zero conditional mean

The expected value of the error term, u, is zero, conditional on any value of the explanatory variables (Raza, 2021a) (Wooldridge, 2016), which is expressed as follows:

Equation 13: MLR4

$$E(u|x) = 0$$

Assumption MLR 5: Homoskedasticity

The variance of the error term, u, is constant given any value of the explanatory variable (Raza, 2021a) (Wooldridge, 2016), which is expressed as follows:

Equation 14: MLR5

$$Var(u|x) = \sigma^2$$

Assumption MLR 6: Constant variance

The error term of the population model, u, is independent of the explanatory variables, its mean is 0, it has a constant variance, and is normally distributed (Raza, 2021a) (Wooldridge, 2016), which is expressed as follows:

Equation 15: MLR6

$$u = N(0, \sigma^2)$$

Gauss-Markov Theorm:

The Gauss-Markov Theorm states, that if the assumptions **MLR 1** through **MLR 5** are met, the models' estimates are said to be the BLUE, best linear unbiased estimates. If the assumption MLR1 through MLR5 are met, the models are said to be a classical linear model. (Raza, 2021a) (Wooldridge, 2016).

APPENDIX E

In the following appendix the concept of correlation and correlation matrixes is presented.

The correlation between x_i and y_i can be denoted by the following equation:

Equation 16: Correlation

$$Corr(X,Y) = \frac{Cov(XY)}{sd(X) * sd(Y)} = \frac{\sigma_{XY}}{\sigma_X * \sigma_Y}$$

In the above equation???, x_i and y_i denotes two variables and the output of the equation is the correlation coefficient. If $Corr(X,Y) = 0$, then the relationship between x_i and y_i is non-linear, and are denoted as being uncorrelated random variables. However, if $Corr(X,Y) = 0$, it does not mean that there is no relationship between x_i and y_i, but that there is no linear relationship between them. If $Corr(X,Y) \neq 0$, then the relationship between x_i and y_i is denoted as being correlated with each other. If $Corr(X,Y) = 1$, then the relationship between x_i and y_i is denoted as being perfectly correlated (Wooldridge, 2016).

To display the correlation between variables, a correlation matrix is often employed. This is done to give an assessment of the relationship between the variables.

The correlation matrix used in this examination is the Pearson Product Moment Correlation matrix. Peason Product Moment Correlation correlation matrix is used to estimate the linear relationship between two variables. The estimate is calculated by measuring how close two variables are compared with a linear line that has been fitted to represent their relationship. The correlation coefficient can range between -1 and 1, where $Corr(X,Y) =$ -1, indicates a perfect negative linear relationship and $Corr(X,Y) = 1$ indicates a perfect linear relationship. In the literature it is proposed that a correlation coefficient in the range between 0.8 and 1, and -0.8 and -1, implies a strongly correlated relationship. Correlation coefficient in a range between 0.5 and 0.8, and -0.5 and -0.8, implies a medium correlated relationship. And lastly a correlation coefficient in the range 0 and 0.5, and -0.5 and 0, implies a weakly correlated relationship (Mohamad Asri et al., 2018) (Wooldridge, 2016) .

To use Pearson Product Moment Correlation, it is assumed that the variables are normally distributed and continuous. For some of the variables included in this study, the assumption of normality does not hold. This is especially true for the dummies included as independent variables and some of the discrete variables. Other alternatives to the Pearson Product

Moment Correlation matrix has been considered, however the specific type of correlation matrix has been assessed as being the most useable(Lærd.n.d.)

APPENDIX F

The Pearson correlation matrix between the independent variables, the explanatory variables and the control variables is presented in figure 1. * indicate a 5% significance level, ** indicate a 1% significance level and *** indicate a 0.1% significance level. Industry is not included in the correlation matrix because the variable is categorical.

Figure 1: Pearson correlation matrix

Variables	ROA	ROE	Shannon index women	Shannon index ethnicity	Dwomen1	Dwomen2	Dwomen3	Dethnicity1	Dethnicity2	Dethnicity3	Percentage women	Percentage ethnicity	Firm age	Average educational level	Average salary	Revenue	Board Size	Firme Size	Tenure
ROA	1000																		
ROE	0.337***	1000																	
Shannon index women	-0.073***	0.017***	1000																
Shannon index ethnicity	-0.022	-0.024*	0.820	1000															
Dwomen1	-0.028*	-0.065***	-0.065***	-0.028*	1000														
Dwomen2	-0.008***	-0.063***	-0.063***	-0.028*	0.000	1000													
Dwomen3	-0.024*	-0.063***	-0.063***	-0.028*	0.000	0.000	1000												
Dethnicity1	-0.022	-0.024*	0.017	0.072	0.006	0.021	-0.046	1000											
Dethnicity2	-0.004	0.002	0.035	0.38	-0.01	-0.0042	-0.004	0.006	1000										
Dethnicity3	-0.020	-0.021	0.011	0.015	0.014	0.00	-0.0045	-0.006	-0.001	1000									
Percentage women	-0.082***	-0.076***	0.072***	0.029*	0.018	0.029***	0.029***	0.025***	0.004	0.013	1000								
Percentage ethnicity	-0.020*	-0.020*	0.021	0.888***	0.018	0.013	0.006	0.325***	0.258***	0.449***	0.027	1000							
Firm age	-0.070***	-0.076***	0.077***	-0.005	0.009	0.272***	0.095***	-0.007	0.014	0.009	0.028***	-0.004	1000						
Average educational level	0.175***	0.086***	0.389***	0.011	-0.056***	-0.042***	0.001	0.016	0.010	-0.027*	0.389***	-0.00	0.045*	1000					
Average salary	0.083***	0.047***	0.047***	0.015	-0.088***	-0.063***	-0.034**	0.023*	0.005	-0.003	0.089***	-0.003	0.067***	0.693***	1000				
Revenue	0.138***	0.38***	0.215***	0.005	-0.068***	-0.018***	0.232***	0.011	0.008	-0.013**	0.255***	0.036	0.068***	0.031*	0.137***	1000			
Board Size	-0.022	-0.005	0.265***	0.030**	-0.055***	0.007***	0.025*	0.043**	0.008	-0.004	0.298***	0.023*	0.079***	0.00	0.054***	0.027***	1000		
Firme Size	-0.010***	-0.064***	0.023***	0.02*	-0.027***	-0.006***	-0.004	0.025*	0.007*	0.018*	0.128***	-0.022	0.077***	0.005	0.007	0.004**	0.043**	1000	
Tenure	-0.006	-0.010***	0.100**	-0.008	0.053***	0.036***	0.029***	0.007	0.035	0.036**	0.128**	-0.022	-0.003	0.020	0.023*	0.021	-0.072	-0.027*	1000

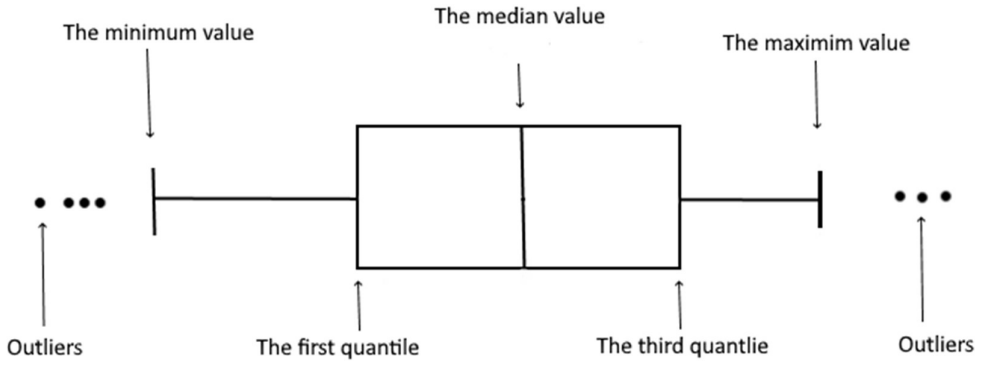
APPENDIX G

In the following appendix the concept of Box-and-Whisker plot and Cooks Distance is presented:

Box-and-Whisker plot:

The Box-and-Whisker plot is a visualization tool to describe the distribution of data by means of quantiles and can be used to detect potential outliers (NCL, n.d.) (Wooldridge, 2016). The Bow-and-Whisker plot is visualized in table 2

Figure 2: Box-and-Whisker plot (the authors own illustration)



The tool displays the minimum value, the first quantile (25th percentile), the median (50th percentile), the third quantile (75th percentile) and the maximum value (NCL, n.d.) (Wooldridge, 2016). The plot is visualized in the form of a box, where the box visualizes the interquartile range (IQR), which is the range between first- and the third quantile and can be described with the following equation:

Equation 17: IQR

$$\mathrm{IQR} = \mathrm{Q}_3 - \mathrm{Q}_1$$

The lines that are stretching parallel with the box are called “whiskers” and display the spread of the data that are outside the first- and third-quantile, ending at the minimum- and maximum value. The outliers, which is represented in the form of the dots outside of the box-and-whiskers, are data that falls far away from the other data and have the potential to interfere with the interpretation of the data (NCL, n.d.) (Wooldridge, 2016). To detect

potential outliers in the data, using the box-and-whisker plot, the following criteria is commonly used:

Equation 18: Upper outliers

$$Upper\ outliers => Q_3 + (1.5 * IQR)$$

Equation 19: Lower outliers

$$Lower\ outliers =< Q_1 + (1.5 * IQR)$$

Where the data point that are placed higher than equation??? And lower than equation???? Are consider outliers in the data (NCL, n.d.) (Wooldridge, 2016).

Cooks Distance:

Cooks Distance is a measure that can be used to detect outliers when performing OLS analysis and can display the magnitude of the influence coming from each data point on the fitted response values. It is also used as a method to calculate the impact of deleting a particular observation from the OLS regression, where outliers have the potential to interfere with the outcome and interpretation of the regression (Pen State, n.d.) (Rstudio, n.d.). Cooks Distance is calculated as the following:

Equation 20: Cook's Distance

$$D_i = \frac{\sum_{j=i}^{n}(\hat{Y}_j - \hat{Y}_{j(i)})^2}{p\ MSE}$$

Where the elements of equation???? Is as follows:

$\hat{Y}_j$ refers to the estimated prediction from the regression for observation j (Penn State, n.d.)

$\hat{Y}_{j(i)}$ refers to the observation j's prediction from a refitted regression model where observation j has been removed (Penn State, n.d.)

p refers to the total number of fitted parameters in the regression model (Penn State, n.d.)

MSE refers to the mean squared error for the regression model (Penn State, n.d.)

To interpret the observations Cooks Distance, and when the distance for the observation is assessed to being influential, different text suggests different methods to calculate the cutoff value. In this study the following method has been chosen:

Equation 21: Cutoff for Cook's Distance

$$\text{cutoff} = 4/\text{N}$$

where N is the number of observations in the data sample (Rstudio, n.d.) (Penn State, n.d.). [41]

[41] This method has been chosen since other methods included the number of explanatory variables. Given that this number will vary dependent on which model it is the distance is calculated based on, this method is chosen to keep the calculation consistent throughout.

APPENDIX H

In the following appendix the notation for the multiple linear regression model is presented. A multiple linear regression model is a variation of the linear regression model, which makes it possible to control several factors that may influence the dependent variable simultaneously. Just as with the linear regression model, the multiple regression model utilizes the ordinary least squares method to estimate its parameters (Wooldridge, 2016). The multiple regression model can be written as follows:

Equation 22: Multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + u$$

Where y is the dependent variable, β_0 is the intercept, β_1 is the parameter associated with x_1, β_2 is the parameter associated with x_2 and u is the error term.

The method used to estimate the parameters in multiple regression is the ordinary least squares method. The estimated OLS regression is as follows:

Equation 23: Estimated OLS regression

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_k$$

Where $\hat{\beta}_0$ is the estimate of the intercept β_0, $\hat{\beta}_1$ is the estimate of the parameter β_1, $\hat{\beta}_2$ is the estimate of the parameter β_2.In the ordinary least squares method the estimates are calculated by minimizing the sum of squared residuals (Wooldridge, 2016). This is done for the n observations on y, x_1 and x_2 $\{(x_{i1}, x_{i2}, y_i): i = 1,2, \dots, n\}$, where the estimates for $\hat{\beta}_0, \hat{\beta}_1 \; and \; \hat{\beta}_2$ are calculated at the same time to create the following:

Equation?????

Equation 24: Minimization of the sum of squared residuals

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2$$

The above equation???? Is then minimized as small as possible.

In the R studio software, the multiple regression model is written as follows:

$$lm([dependent\ variable] \sim [explanatory\ variables], date = [data\ source])$$

APPENDIX I

In the following section the diagnostic of the main regression is performed. This is done in accordance with the Ordinary least squares (OLS) regression assumption presented in appendix D

Figure 3: Graphic representation of regression diagnostic of OLS model 1aa

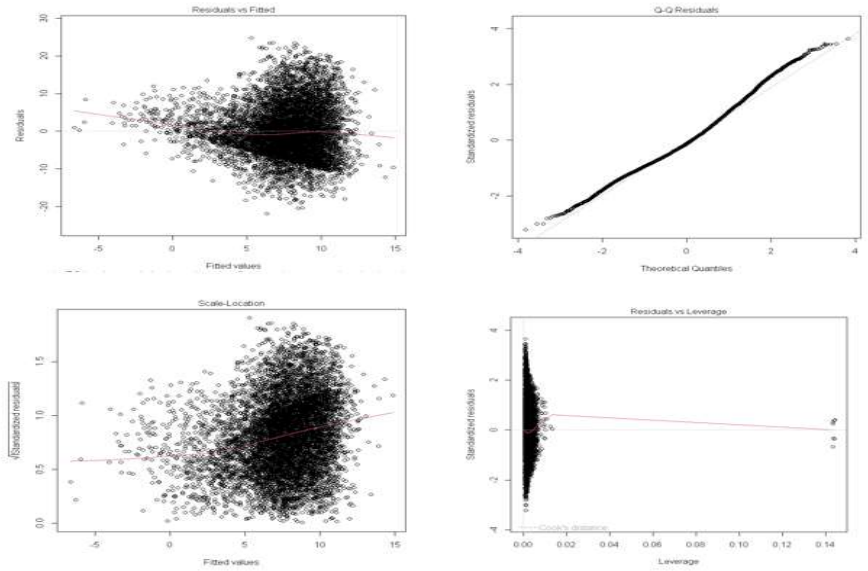


Figure 4: Graphic representation of regression diagnostic of OLS model 1ab

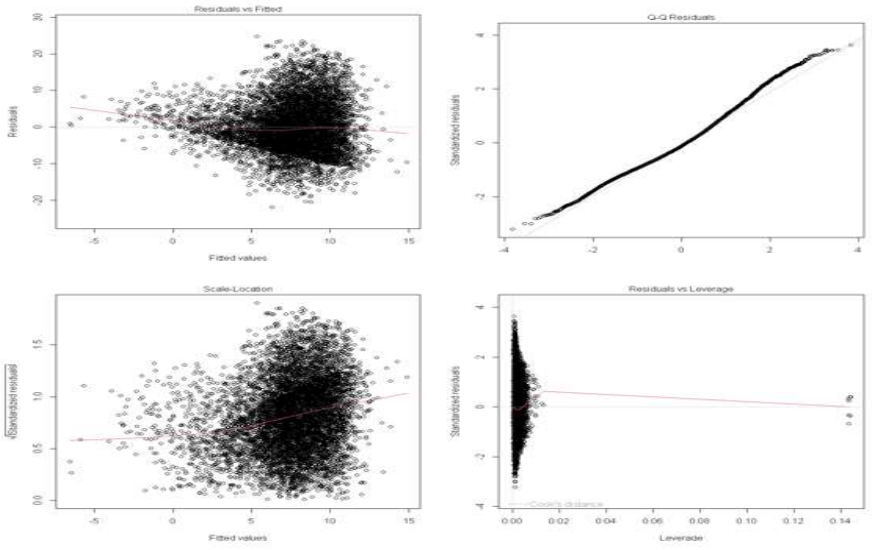


Figure 5: Graphic representation of regression diagnostic of OLS model 1ba

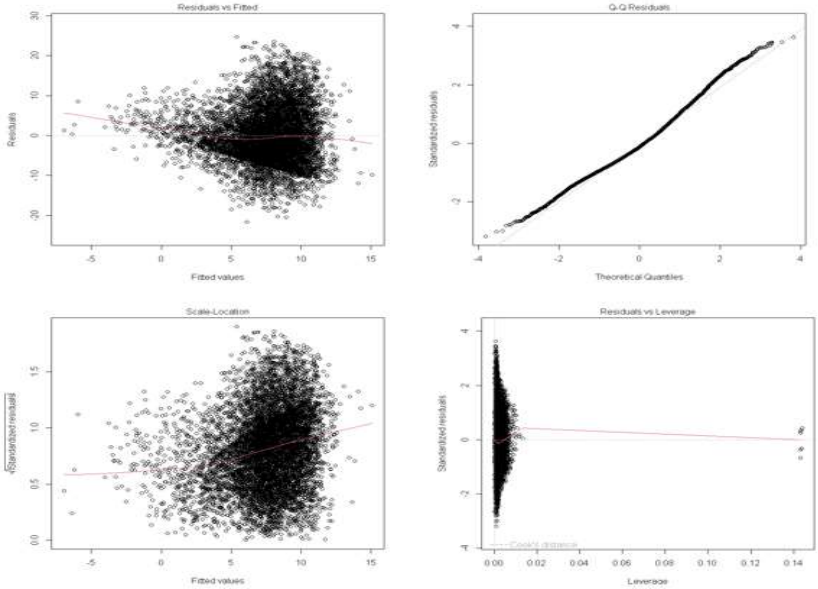


Figure 6: Graphic representation of regression diagnostic of OLS model 1bb

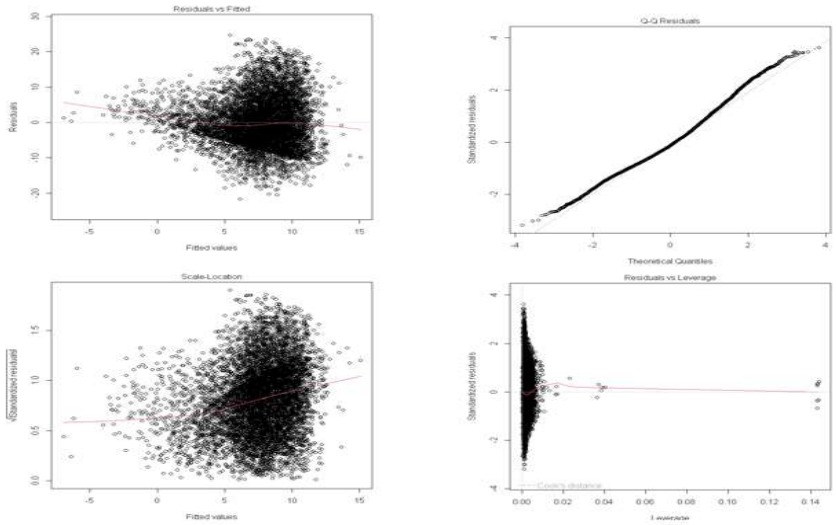


Figure 7: Graphic representation of regression diagnostic of OLS model 2

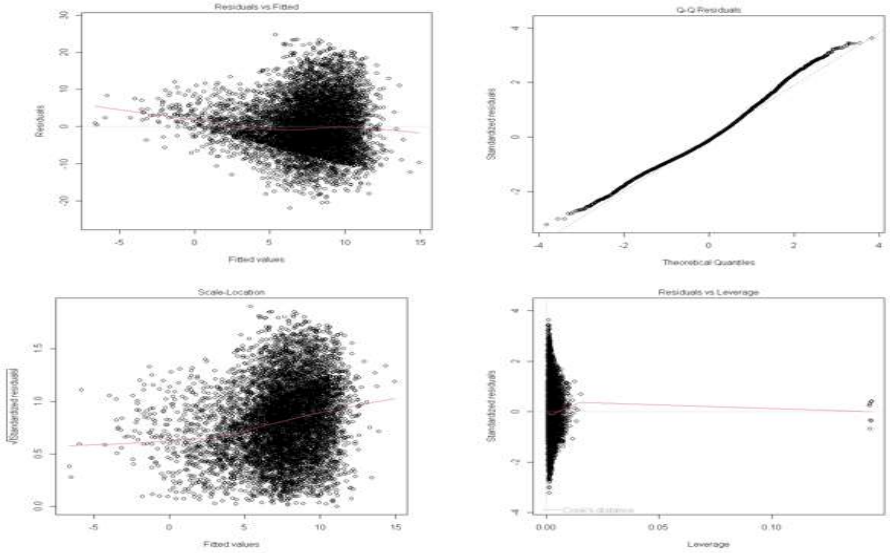


Figure 8: Graphic representation of regression diagnostic of OLS model 3

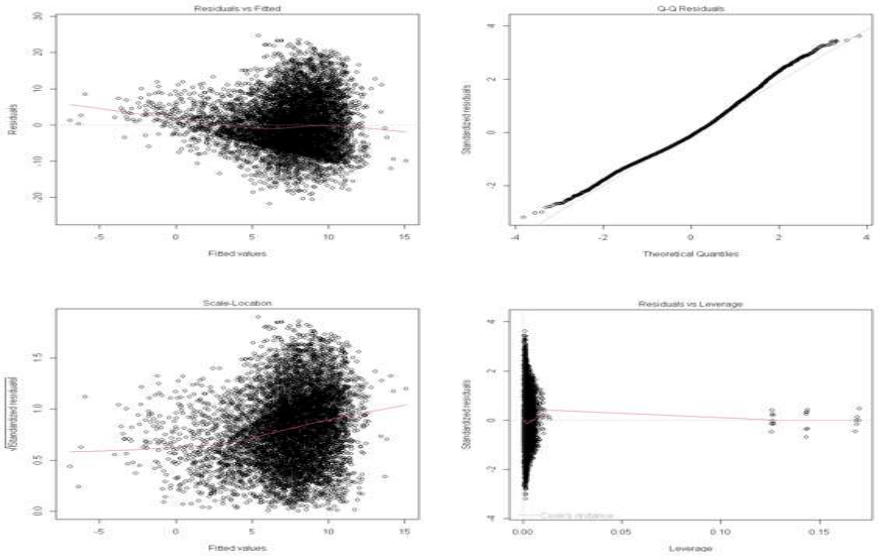


Figure 3 to 8 depict the plots of the regression models. In the graph called *"Residuals vs. Fitted"* the relationship between the residuals and the fitted values are depicted. The graph can be used to illustrate whether the parameters included in the regression exhibit a linear pattern. It applies that, the flatter the horizontal red line, the greater the indication are that the parameters included in the regression have a linear relationship. From the "*Residuals vs. Fitted"* graph it can be inferred that the red horizontal line is approximately flat for the

different models. This suggest a linear relationship between the residuals and the estimated values across the models, and the MLR1 assumption is not violated (Raze, 2021b).

It is not feasible to test the MLR2 assumption directly and it is not possible to test it graphically. However, due to the sample size on 7579 it is assessed that it is stochastic.

The MLR3 proposes no perfect multicollinearity between the variables included in the regression models. This is investigated using the *Variance Inflation Factor* (VIF). The measurement is utilized to illustrate how much of the variance for a regression coefficient, $\hat{\beta}_k$, due to multicollinearity (Wooldridge, 2016). A rule of thumb for the VIF values is that if the VIF is larger than 5 or 10, it indicates multicollinearity (Studenmund, 2014) (Adam Petrie, n.d.).

Table 19: VIF for model 1aa

Variable	VIF
SIK_{t-2}	1,093827
SF_{t-1}	2,179658
BS_{t-1}	1,103069
FA_{t-1}	1,459516
REV_{t-1}	2,232017
AE_{t-1}	1,238446
ET_{t-1}	1,608084
AS_{t-1}	1,411266
IN_{t-1}	1,895709

Table 20: VIF for model 1ab

Variable	VIF
PK_{t-2}	1,120792
SF_{t-1}	2,179622
BS_{t-1}	1,103733
FA_{t-1}	1,460580
REV_{t-1}	2,245728
AE_{t-1}	1,238943
ET_{t-1}	1,608422
AS_{t-1}	1,412484
IN_{t-1}	1,894979

Table 21: VIF for Model 1ba

Variable	VIF
SIE_{t-2}	1,0004643
SF_{t-1}	2,180713
BS_{t-1}	1,103619
FA_{t-1}	1,458303
REV_{t-1}	2,198417
AE_{t-1}	1,234970
ET_{t-1}	1,592558
AS_{t-1}	1,399443
IN_{t-1}	1,887384

Table 22: VIF for Model 1bb

Variable	VIF
PE_{t-2}	1,003646
SF_{t-1}	2,180982
BS_{t-1}	1,102997
FA_{t-1}	1,458307
REV_{t-1}	2,198750
AE_{t-1}	1,234931
ET_{t-1}	1,592605
AS_{t-1}	1,399144
IN_{t-1}	1,885762

Table 23: VIF for Model 2

Variable	VIF
$Dwomen1_{t-2}$	1,253552
$Dwomen2_{t-2}$	1,268271
$Dwomen3_{t-2}$	1,118900
SF_{t-1}	2,181389
BS_{t-1}	1,202052
FA_{t-1}	1,461783
REV_{t-1}	2,242150
AE_{t-1}	1,238940
ET_{t-1}	1,608772
AS_{t-1}	1,413260
IN_{t-1}	1,900344

Table 24: VIF for Model 3

Variable	VIF
$Detnicity1_{t-2}$	1,006100
$Detnicity2_{t-2}$	1,001849
$Detnicity3_{t-2}$	1,005160
SF_{t-1}	2,181370
BS_{t-1}	1,105601
FA_{t-1}	1,458935
REV_{t-1}	2,199599
AE_{t-1}	1,235505
ET_{t-1}	1,593099
AS_{t-1}	1,399800
IN_{t-1}	1,889714

Tabel table 19 to 24 shows, the VIF values from the variables are under the value of 5 and 10. This indicates no perfect multicollinearity in the regression models.

The MLR4 states that the calculated mean of the error term, u, is 0. This assumption can be examined by calculating the mean of the error term of the regression models.

Table 25: Mean of residuals for model 1aa – model 3

Model	Mean of the residuals
Model 1aa	$0.5245056e^{-15}$
Model 1ab	$0{,}9801812e^{-16}$
Model 1ba	$0{,}5947885e^{-17}$
Model 1bb	$0{,}4182516e^{-15}$
Model 2	$0{,}343604e^{-16}$
Model 3	$0{,}2675844e^{-17}$

Table 25 shows that the mean of the residuals for the models are approximately 0. From this it is assessed that the mean of the regression models is 0.

The MLR5 assumption states that the regression model should exhibit homoscedasticity. This can be investigated using several methods. The figures 3 to 8 that was presented in the above sections, depicts a graph called "*Scale-Location"* , which shows how the standardized residuals are spread along the ranges of the predictors. The graph can give an indication into whether the assumption of homoscedasticity is fulfilled or not. If the standardized residual seems to be spread randomly, then the residuals are said to have a constant variance. On the other hand, if the spread is seen following a trend, then the residuals are said to have a non-constant variance (Raza, 2021b). From the graph, a trend may appear, which indicates the presence of heterogeneity. To further investigate this, a Breusch-Pagan (BP) test is performed. In this test, the null hypothesis is that homoskedasticity is present in the regression model.

Table 26: Breusch-Pagan test for model 1aa – model 3

Model	BP	df	P-value
Model 1aa	360,83	14	$<0{,}2e^{-15}$
Model 1ab	267,9	14	$<0{,}2e^{-15}$
Model 1ba	284,19	14	$<0{,}2e^{-15}$
Model 1bb	281,92	14	$<0{,}2e^{-15}$
Model 2	272,48	16	$<0{,}2e^{-15}$
Model 3	285,28	16	$<0{,}2e^{-15}$

Table???? shows the results of the BP test. From the tests p-values, it can be inferred that the null-hypothesis can be rejected and indicate that heteroskedasticity is precent in the regression models. This indicates that the t-statistics and F-statistics may not be reliable (Wooldridge, 2016). However, this can be accommodated by utilizing robust standard errors and utilize these in the regression analysis.

The robust standard errors for the regression models are calculated and shown in the regression models presented in the quantitative examination, section 9.2.1 – 9.2.3.

The MLR6 assumption states that the error term of the population is normally distributed. This can be investigated utilizing a graphical presentation of the error term such as histogram.

Figure 9: Histogram for the error term for model 1aa

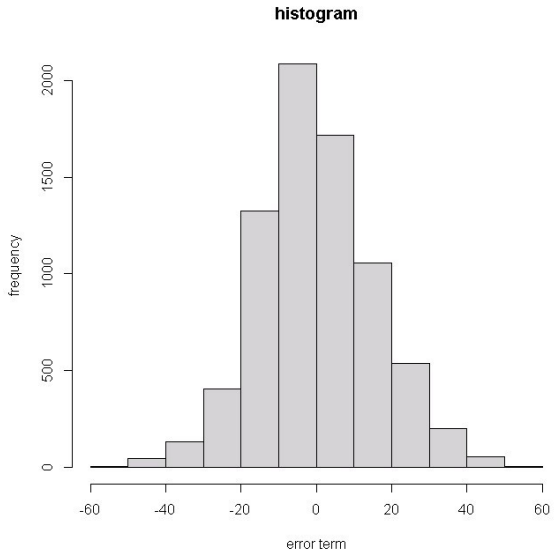


Figure 10: Histogram for the error term for model 1ab

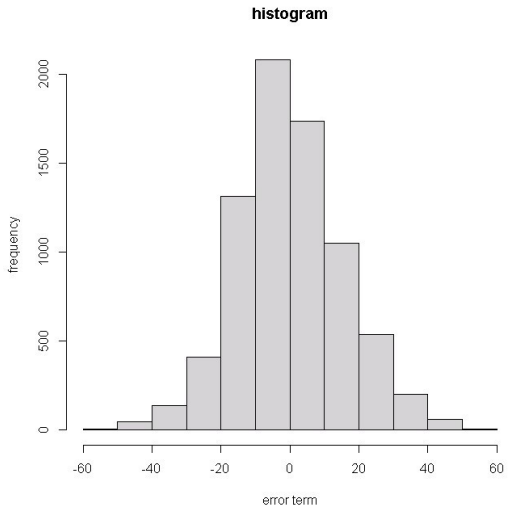


Figure 11: Histogram for the error term for model 1ba

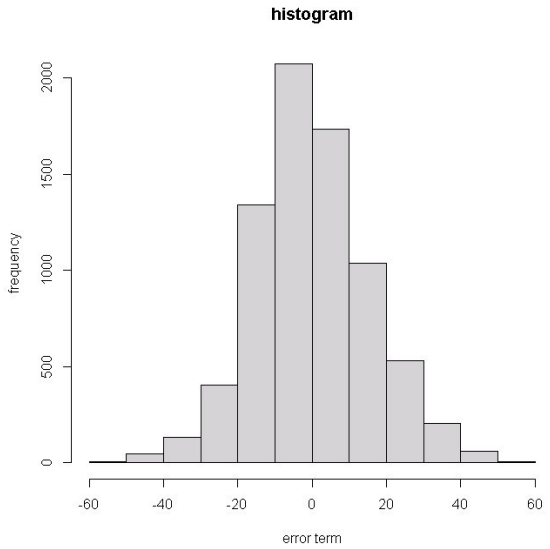


Figure 12: Histogram for the error term for model 1bb

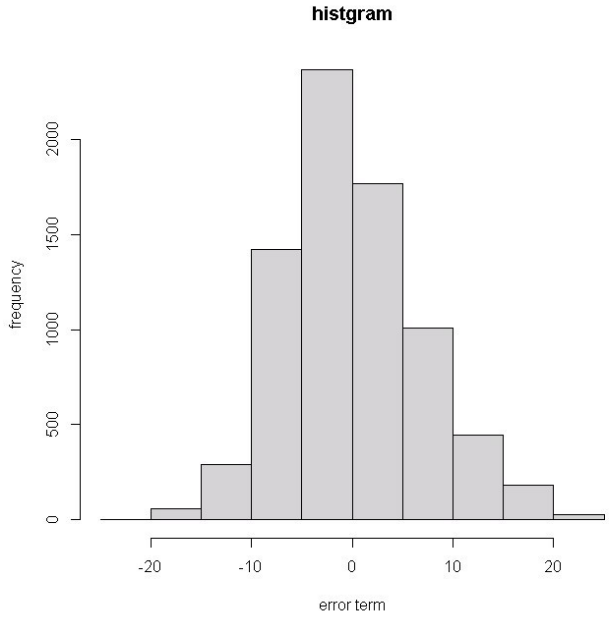


Figure 13: Histogram for the error term for model 2

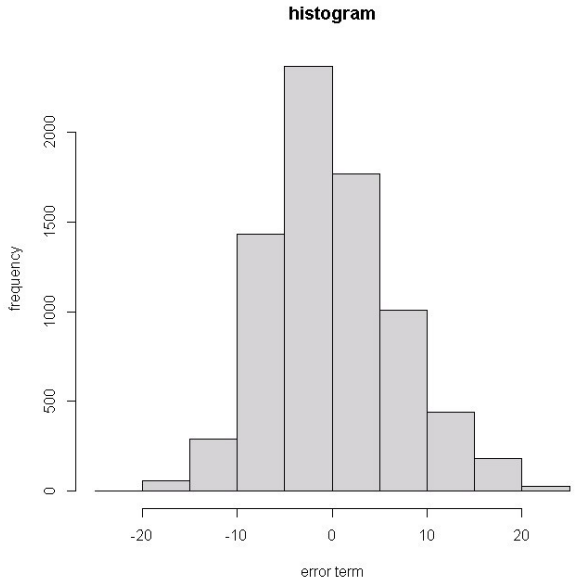
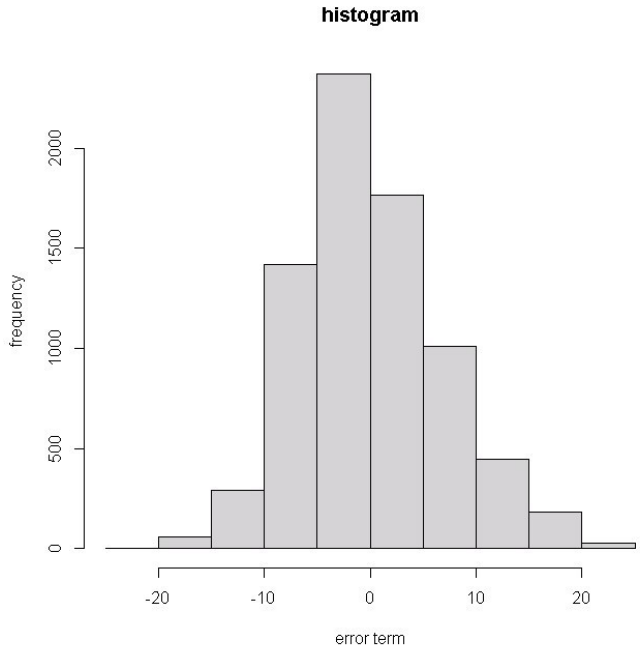


Figure 14: Histogram for the error term for model 3



From the above figure 9 to figure 14, a histogram of the error term, u, is presented. From the figures it can be inferred that the error term follows an approximately normal distribution.

Given that the regression models do not appear to breach the MLR1 to MLR6 assumptions, they can be categorized as classical linear models.

APPENDIX J

In the following section the diagnostic of the robustness regression is performed. This is done in accordance with the Ordinary least squares (OLS) regression assumption presented in appendix D

Figure 15: Graphic representation of regression diagnostic of OLS robustness model 1aa

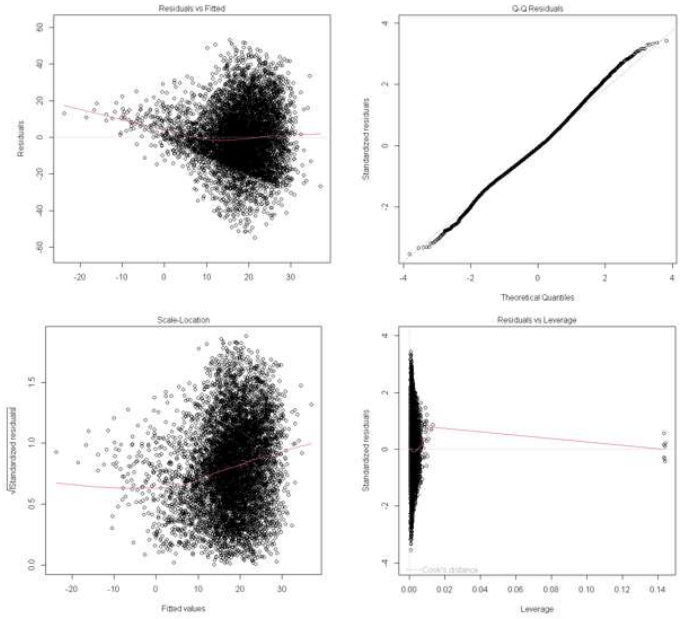


Figure 16: Graphic representation of regression diagnostic of OLS robustness model 1ab

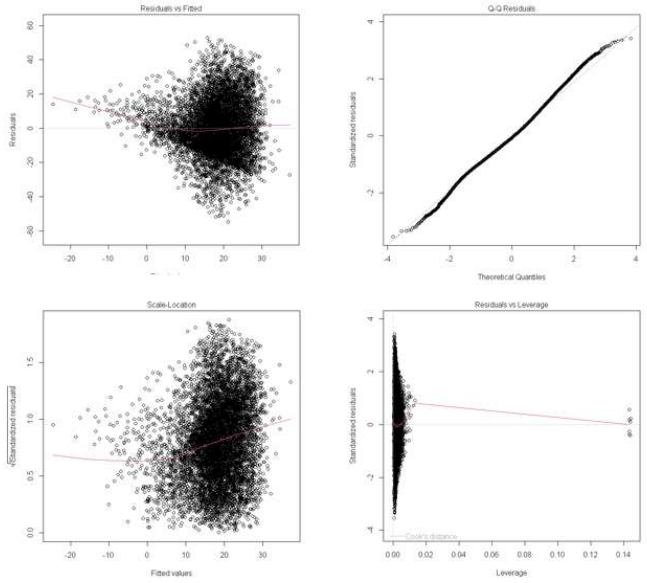


Figure 17: Graphic representation of regression diagnostic of OLS robustness model 1ba

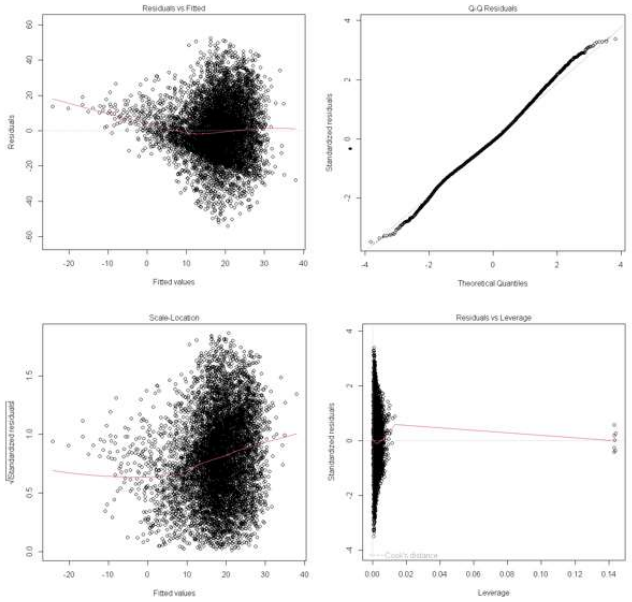


Figure 18: Graphic representation of regression diagnostic of OLS robustness model 1bb

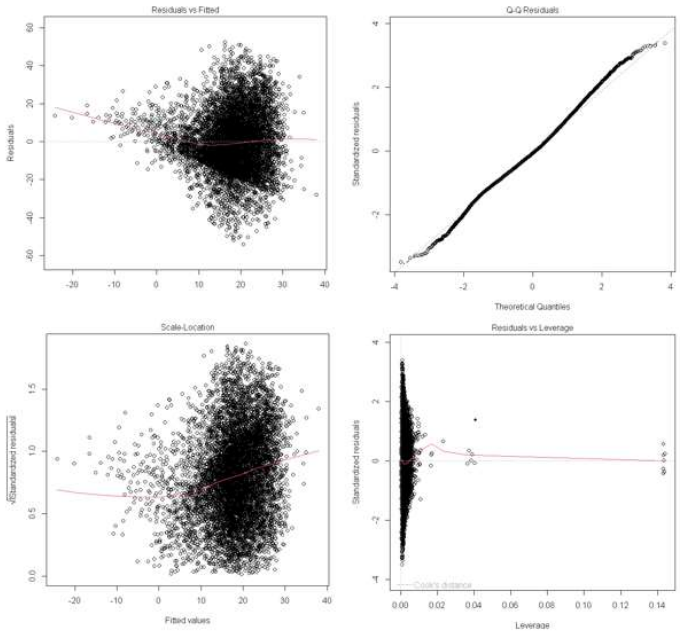


Figure 19: Graphic representation of regression diagnostic of OLS robustness model 2

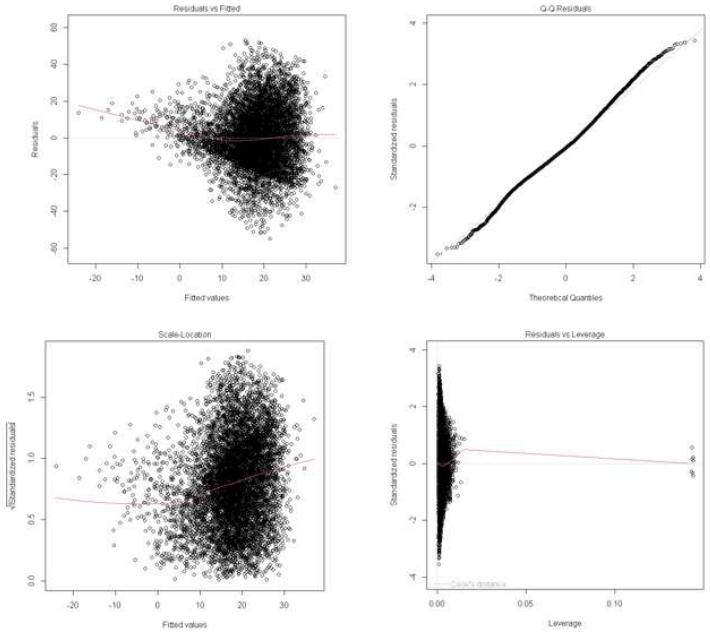


Figure 20: Graphic representation of regression diagnostic of OLS robustness model 3

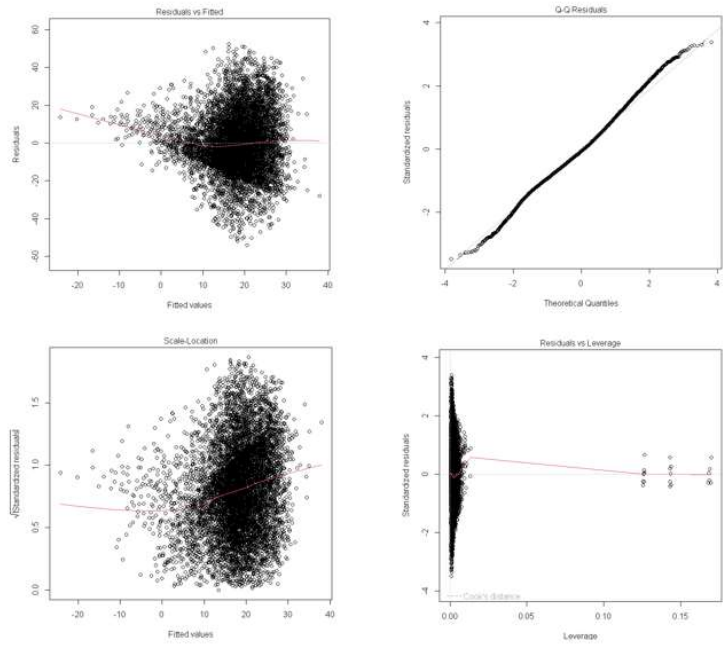


Figure 15 to 20 depict the plots of the regression models. In the graph called *"Residuals vs. Fitted"* the relationship between the residuals and the fitted values are depicted. The graph can be used to illustrate whether the parameters included in the regression exhibit a linear pattern. It applies that, the flatter the horizontal red line, the greater the indication are that the

parameters included in the regression have a linear relationship. From the "*Residuals vs. Fitted"* graph it can be inferred that the red horizontal line is approximately flat for the different models. This suggest a linear relationship between the residuals and the estimated values across the models, and the MLR1 assumption is not violated (Raze, 2021b).

It is not feasible to test the MLR2 assumption directly and it is not possible to test it graphically. However, due to the sample size on 7579 it is assessed that it is stochastic.

The MLR3 proposes no perfect multicollinearity between the variables included in the regression models. This is investigated using the *Variance Inflation Factor* (VIF). The measurement is utilized to illustrate how much of the variance for a regression coefficient, $\hat{\beta}_k$, due to multicollinearity (Wooldridge, 2016). A rule of thumb for the VIF values are, that if the VIF is larger than 5 or 10, it indicates multicollinearity (Adam Petrie, n.d.).

Table 27: VIF for robustness model 1aa

Variable	VIF
SIK_{t-2}	1,093827
SF_{t-1}	2,179658
BS_{t-1}	1,103069
FA_{t-1}	1,459516
REV_{t-1}	2,232017
AE_{t-1}	1,238446
ET_{t-1}	1,608084
AS_{t-1}	1,411266
IN_{t-1}	1,895709

Table 28: VIF for robustness model 1ab

Variable	VIF
PK_{t-2}	1,120792
SF_{t-1}	2,179622
BS_{t-1}	1,103733
FA_{t-1}	1,460580
REV_{t-1}	2,245728
AE_{t-1}	1,238943
ET_{t-1}	1,608422
AS_{t-1}	1,412484
IN_{t-1}	1,894979

Table 29: VIF for robustness model 1ba

Variable	VIF
SIE_{t-2}	1,0004643
SF_{t-1}	2,180713
BS_{t-1}	1,103619
FA_{t-1}	1,458303
REV_{t-1}	2,198417
AE_{t-1}	1,234970
ET_{t-1}	1,592558
AS_{t-1}	1,399443
IN_{t-1}	1,887384

Table 30: VIF for robustness model 1bb

Variable	VIF
PE_{t-2}	1,003646
SF_{t-1}	2,180982
BS_{t-1}	1,102997
FA_{t-1}	1,458307
REV_{t-1}	2,198750
AE_{t-1}	1,234931
ET_{t-1}	1,592605
AS_{t-1}	1,399144
IN_{t-1}	1,885762

Table 31: VIF for robustness model 2

Variable	VIF
$Dwomen1_{t-2}$	1,253552
$Dwomen2_{t-2}$	1,268271
$Dwomen3_{t-2}$	1,118900
SF_{t-1}	2,181389
BS_{t-1}	1,202052
FA_{t-1}	1,461783
REV_{t-1}	2,242150
AE_{t-1}	1,238940
ET_{t-1}	1,608772
AS_{t-1}	1,413260
IN_{t-1}	1,900344

Table 32: VIF for robustness model 3

Variable	VIF
$Detnicity1_{t-2}$	1,006100
$Detnicity2_{t-2}$	1,001849
$Detnicity3_{t-2}$	1,005160
SF_{t-1}	2,181370
BS_{t-1}	1,105601
FA_{t-1}	1,458935
REV_{t-1}	2,199599
AE_{t-1}	1,235505
ET_{t-1}	1,593099
AS_{t-1}	1,399800
IN_{t-1}	1,889714

Tabel 27 to 32 shows, the VIF values from the variables are under the value of 5 and 10. This indicates no perfect multicollinearity in the regression models.

The MLR4 states that the calculated mean of the error term, u, is 0. This assumption can be examined by calculating the mean of the error term of the regression models.

Table 33: Mean of residuals for robustness model 1aa – model 3

Model	Mean of the residuals
Model 1aa	$-0.44685e^{-15}$
Model 1ab	$0{,}8902227e^{-16}$
Model 1ba	$0{,}8274478e^{-16}$
Model 1bb	$0{,}2683657e^{-15}$
Model 2	$-0{,}5528586e^{-16}$
Model 3	$-0{,}3783712e^{-15}$

Table 33 shows that the mean of the residuals for the models are approximately 0. From this it is assessed that the mean of the regression models is 0.

The MLR5 assumption states that the regression model should exhibit homoscedasticity. This can be investigated using several methods. The figures 15 to 20 that was presented in the above sections, depicts a graph called "*Scale-Location"* , which shows how the standardized residuals are spread along the ranges of the predictors. The graph can give an indication into whether the assumption of homoscedasticity is fulfilled or not. If the standardized residual seems to be spread randomly, then the residuals are said to have a constant variance. On the other hand, if the spread is seen following a trend, then the residuals are said to have a non-constant variance (Raza, 2021b). From the graph, a trend may appear, which indicates the presence of heterogeneity. To further investigate this, a Breusch-Pagan (BP) test is performed. In this test, the null hypothesis is that homoskedasticity is present in the regression model.

Table 34: Results of the Breusch-Pagan test for robustness model 1aa - model 3

Model	BP	df	P-value
Model 1aa	214,18	14	$<0,2e^{-15}$
Model 1ab	223,47	14	$<0,2e^{-15}$
Model 1ba	232,97	14	$<0,2e^{-15}$
Model 1bb	230,68	14	$<0,2e^{-15}$
Model 2	235,89	16	$<0,2e^{-15}$
Model 3	234,31	16	$<0,2e^{-15}$

Table 34 shows the results of the BP test. From the tests p-values, it can be inferred that the null-hypothesis can be rejected, and indicate that heteroskedasticity is precent in the regression models. This indicates that the t-statistics and F-statistics may not be reliable (Wooldridge, 2016). However, this can be accommodated by utilizing robust standard errors and utilize these in the regression analysis.

The robust standard errors for the regression models are calculated and shown in the regression models presented in the quantitative examination, section 9.3.1 – 9.3.3.

The MLR6 assumption states that the error term of the population is normally distributed. This can be investigated utilizing a graphical presentation of the error term such as histogram.

Figure 21: Histogram for the error term for robustness model 1aa

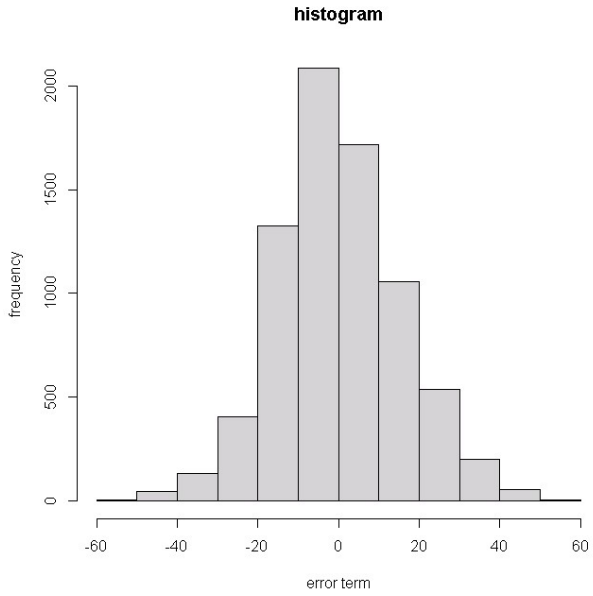


Figure 22: Histogram for the error term for robustness model 1ab

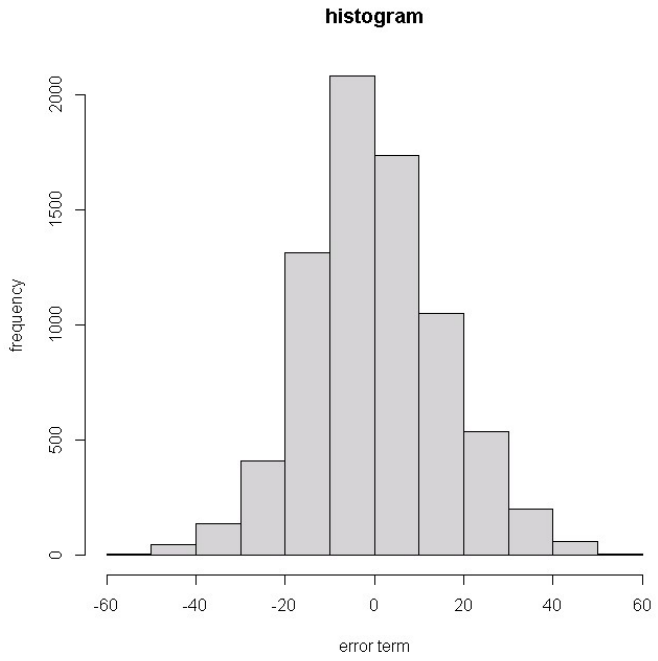


Figure 23: Histogram for the error term for robustness model 1ba

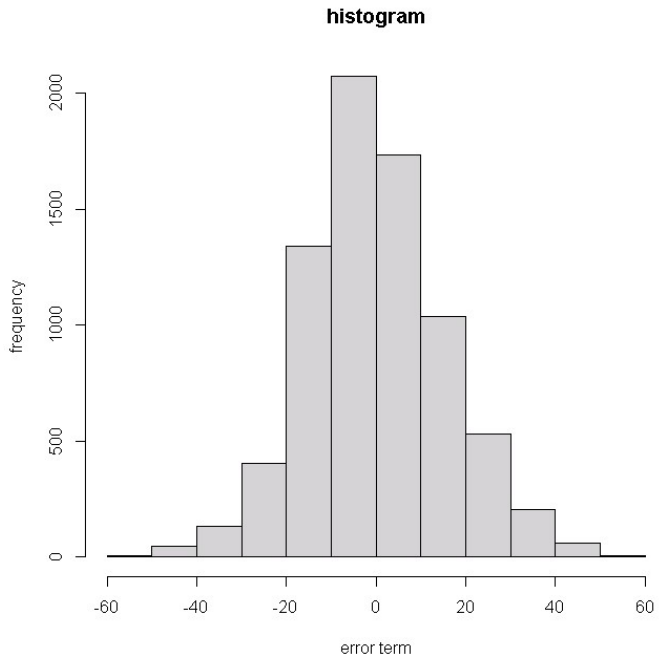


Figure 24: Histogram for the error term for robustness model 1bb

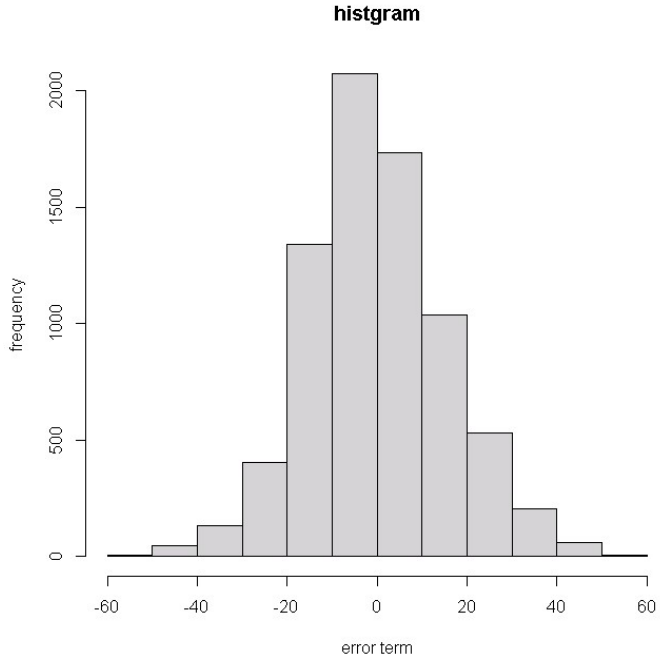


Figure 25: Histogram for the error term for robustness model 2

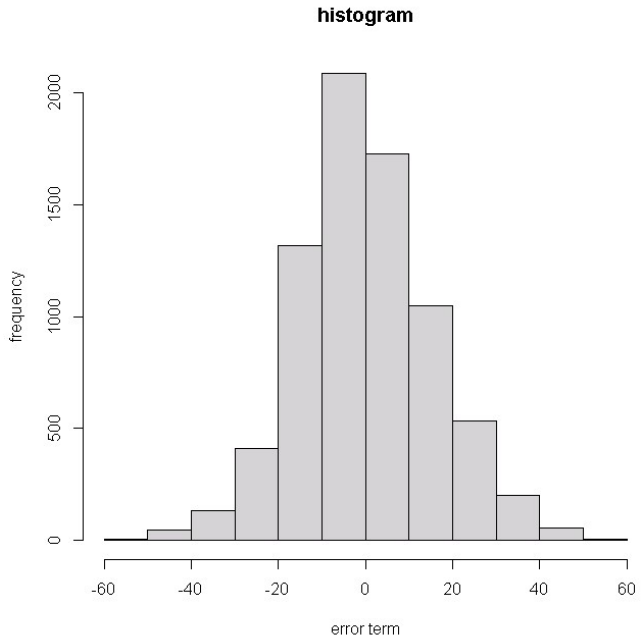
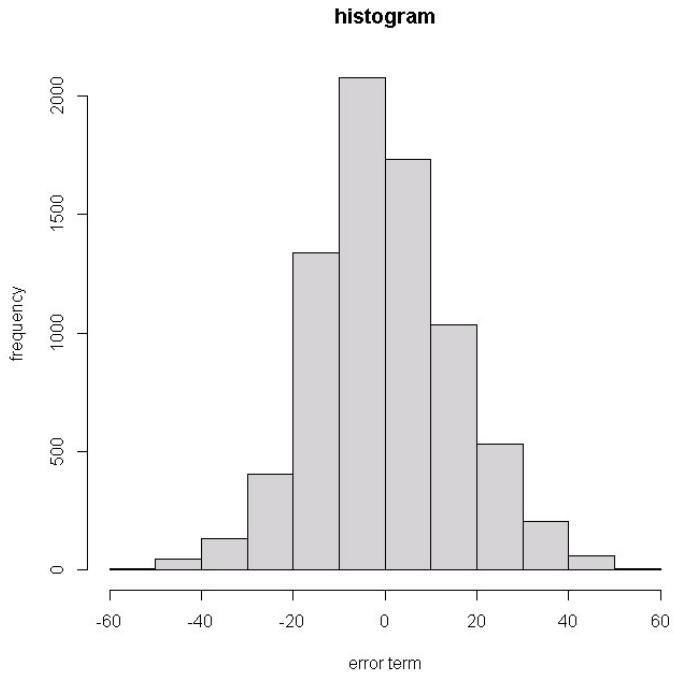


Figure 26: Histogram for the error term for robustness model 3



From the above figure 21 to figure 26, a histogram of the error term, u, is presented. From the figures it can be inferred that the error term follows an approximately normal distribution.

Given that the regression models do not appear to violate the MLR1 to MLR6 assumptions, they can be categorized as classical linear models.