University of Tartu

Faculty of Science and Technology

Institute of Computer Science

Oluwagbemi Omobolanle Kadri

GENDER-BASED SEGREGATION IN COMPANY BOARDS AND WELL-BEING

Master's Thesis (30 ECTS)
Software Engineering Curriculum

Supervisors:

Rajesh Sharma PhD
Peep Kungas PhD

Tartu 2020

Abstract:

Gender-based segregation in company boards and well-being

Segregation is an act of division from a supreme body to smaller groups because of the characteristics of the body. In order terms among humans, we can refer to it as an unwarranted detachment or separation resulting in traits a person possesses; for example, gender, occupation, race, resident, income, religion, age, etc. Analyses based on segregation impact in our society have increased over the years. It has spawned enormous controversial discussions in our modern-day world; and elicited several researchers' interests to identify the origin of segregation.

In this thesis, we investigated if gender and age segregation exist in Estonian companies' boards and its relationship with the labour market. In addition, we examine if it leads to high credit risks and a negative correlation to the well-being of Estonian society.

The key measurement factors for comparison and drawing conclusions are the unemployment rate measured as the labour market, financial key performance indicators measured as credit risk, a well-deprivation index measured as well-being and segregation indexes from 'SCube' data model measured as segregation. 'SCube' originated from a model created by researchers at the University of Pisa, it uses a data science framework to deal with the problem of social and occupational segregation. Analysis from the 'SCube' data-set will be measured with segregation indexes ranging from 0 to 1 in accordance to this range high level of segregation means high value of the segregation index meaning a value close to 1.

The Estonian statistics ready-made data-set is used in conjunction with the data-set from 'SCube' model to examine and draw conclusions of the occupational segregation problem discussed in this work. In addition, statistical techniques correlation and causal inference are used to determine the relationship and causal effects between segregation and the various factors.

CERCS: P170 Computer science, numerical analysis, systems, control.

Keywords: Segregation, well-being, gender-based segregation, labor, company's board, minority group.

Abstraktne:
Sooline segregatsioon ettevõtete juhatustes ja heaolu

Eraldamine on jagunemine kõrgeimast organist väiksematesse rühmadesse, kuna see on organ. Inimeste seas võib seda nimetada põhjendamatuks eraldamiseks või eraldamiseks, mille tulemuseks on isiku omandiomadused, näiteks soo, ametikoht, rass, elanik, sissetulekud, religioon, vanus jne. Analüüse, mis põhinevad segregatsiooni mõjul meie ühiskonnas, on aastate jooksul suurenenud. See on tekitanud meie tänapäeva maailmas tohutuid vastuolulisi arutelusid ja õhutanud mitmete teadlaste huve eraldatuse päritolu kindlakstegemiseks.

Selles väitekirjas uurisime, kas Eesti ettevõtete juhatustes eksisteerib sooline ja vanuseeraldus ning selle suhted tööturuga. Lisaks uurime, kas see toob kaasa kõrged krediidiriskid ja negatiivse korrelatsiooni Eesti ühiskonna heaoluga.

Võrdluse ja järelduste tegemise põhitegurid on tööpuuduse määr, mida mõõdetakse tööturul, krediidiriskina mõõdetavad finantstulemuste põhinäitajad, heaoluindeksid, mida mõõdetakse heaolu ja segregatsiooni indeksid SCube'i andmemudelist, mida mõõdetakse segregatsioonina. Pisa ülikooli teadlaste loodud mudelist lähtuv SCUBE kasutab andmeteaduse raamistikku, et tegeleda sotsiaalse ja ametialase eraldatuse probleemiga. SCube-andmekogumi analüüsi mõõdetakse eraldamisindeksitega, mis ulatuvad 0-st 1ni, vastavalt sellele vahemikuse kõrgele eraldamistasemele tähendab eraldamisindeksi kõrget väärtust, mis tähendab lähedast väärtust.

Eesti statistikat kasutatakse koos SCube'i mudeli andmete kogumiga, et uurida ja teha järeldusi tööalase segregatsiooni probleemist, mida selles töös arutatakse. Lisaks kasutatakse statistilisi järeldusi ja põhjuslikke järeldusi segregatsiooni ja erinevate tegurite vahelise seose ja põhjusliku mõju kindlaksmääramiseks.

CERCS: P170 Arvutiteadus, numbriline analüüs, süsteemid, kontroll.
Keywords: palgaline ebavõrdsus, firma kasv, empiiriline analüüs, ennustav analüüs

Contents

1 Introduction

'Segregation', a practice that separates things or people into groups based on their characteristics. Segregation among people include unfair treatment of the segregated group; it occurs in different forms (race, gender, age, religious beliefs, language, cultural traits and more) and can take place, for example, at workplaces, neighbourhood; or other units in a community [7]. Segregation by a person's gender is gender-based segregation or gender segregation; a similar definition for age segregation.

Several articles claim that segregation is the principal reason for gender inequality and gender pay gap in the labour market [13]. What is the labour market? Labour market trades with demand and supply of labour; a market where employer demands and the person employed supplies; employers compete to hire the most effective, whereas employees compete to deliver excellent satisfying jobs.

Most findings related to segregation issue among board members narrow down their results to how it affects the company and not the country. However, in this work, the scope broadens beyond a specific company to the entire country. For one of the research hypothesis of this study, we will analyze even distribution among gender and age in boards of companies; how it leads to improved credit risk management or low credit risks for companies in Estonia; the credit risk management measure used in this work is the total corporate financial performance of companies in different counties within the country.

The corporate financial performance in this study uses financial KPI as a measurement. We can define credit risk as to the possibility of loss [1] that a bank borrower or counterparty defaults on its loan repayments or commitments and agreed terms. It's a risk that the lender won't receive the amount lent to the borrower or its interest [23]. One aspect leading to sustenance and improvement of financial companies or banks in today's competitive market; is their strength to use strategic and competitive ways to maintain risks; for example, credit risks [2]. We can define credit risk management as the practice of preventing or mitigating credit risk; it identifies, analyzes and implements various risk factors [3] or risk control strategies or measures to forestall loan losses at any point in time.

In the aspect of credit risk concerning this work, few studies inspect issues borrowers have that cause loan payment default; for instance, issues like employees unhappiness working in an organization; segregation or discrimination among boards of directors within the company and its effect on the company's performance. Adequate knowledge of those problems and reduce in them might help to predict a high level of credit risk or mitigate and stop credit risk. Managing credit risks is a challenging task for financial companies and banks, but can segregation play a role in increasing such risk?.

[1] https://www.investopedia.com/terms/c/creditrisk.asp

[2] https://medium.com/@vietnamcreditmedia/what-is-credit-risk-management-9844f649b13c

[3] https://www.wallstreetmojo.com/credit-risk-management/

It's impossible to speak about segregation or discrimination amongst people and not mention its relationship with well-being; both concepts revolve around humans psychological state; in researches concerning discrimination, a common practice is detecting its effects with well-being. Oxford Dictionary describes well-being as being in a healthy, happy or comfortable state; their definition defines well-being using happiness. However, the concept is far broader than that and has caused several debates since the 3rd century BC [4].

This thesis will answer three research questions: (i) Is there a relationship between gender and age segregation among boards of directors in Estonia and the labour market? (ii) Is there a relationship between gender and age segregation among boards of directors in Estonia and credit risk? (iii) Is there a relationship between gender and age segregation among boards of directors in Estonia and the country's well-being?.

This work aims to determine the correlation and causal relationship of gender- and age-based segregation in companies boards with the labour market; credit risk and well-being in Estonia; if well-being has a strong positive relationship with segregation. Few researchers at the University of Tartu made initial findings on this subject; results from their research case study showed a negative correlation between unemployed young men from Lääne-Virumaa, Estonia and the county's unemployment rate for the year 2008 to 2015. But further studies like the one from this thesis is required to verify this hypothesis. The University of Pisa researchers have also developed a segregation-aware framework known as 'SCube' to measure segregation in the boards of companies; the data from its models are obtainable daily, and they use metrics measured as segregation indexes ranging from 0 to 1; high values of the index close to 1 means a high level of discrimination. The metrics or quantitative measures from SCube data uses indexes to measure segregation level and compares its relationship with indicators derived from the labour market, companies financial performances and well-being. Using segregation metrics to predict unemployment might raise the decision-making standard for unemployment, and eradicate quarterly delays in measuring policy changes effects on unemployment.

The data used in this research are SCube data: it originated from SCube Framework developed by researchers at the University of Pisa; Statistics Estonia data: derived from a statistics database of an Estonian government agency that governs data within the country. There are seven chapters or sections in this thesis divided into Chapter 1 the current section explaining the introductory description of the topic; Chapter 2 illustrates detailed works related to this study and their contributions to this research; recent trends similar to this topic. Chapter 3 explains the data used in this thesis; it explains how and where they fit in the study; it tells information of collected, calculated and measured data. Chapter 4 informs about testing the research hypothesis; it also confirms and gives more details about the hypothesis questions. Chapter 5 outlines the problems of this study and explains the method used to tackle the research problem. Chapter 6 shows, reports and explains the study results; it tests the research and confirms its approach. Chapter 7 is the concluding part of this work; it summarizes the research topic and findings.

[4] https://www.wellbeingpeople.com/2018/07/20/what-does-wellbeing-actually-mean/

2 Related work

This section will compare other works and use its comparison to describe this exploration. It will summarize, feature differences and similarities of diverse researches; discuss recent global trends relating to the former and this study and review how subjects from related studies helped this work. Various discussions and topics to consider are (i) Emerging Trends (ii) Gender diversity in companies' boards. (ii) Gender Pay Gap in Estonia. (iii) Age diversity and discrimination. (iv) Well-being measure. Table 2.6 includes a summary of all studies discussed in this section.

2.1 Emerging Trends

There is a rising trend that advocates for more women in companies' board; this movement promotes a gender-based diverse team of board members. Recent studies of gender diversity in companies' boards often use the number of women holding a position in corporate board seats as a measurement for the board's gender diversity [5]; because on a global scale, men occupy more board seats than women [9]. Many countries have adopted this trend; for example, in 2003, Norway was first to introduce gender quota law, requiring that public companies' boards should include a minimum of 40% female directors [8]. In March 2015, Germany also set 30% gender quota for the boardroom [10]; which made room for women on the board seats of Europe's top companies by 2016. In 2017 California, US signed a bill that imposed monetary fines on companies without at least one woman as a board member; they also motivated larger boards to appoint three or more female board members [20]. In 1994, Switzerland adopted a Gender Equality Act (GEA) to prevent gender-based discrimination in the work environment; Estonia adopted the same act in 2004[5]. PwC's 2018 director's survey[6] for the USA revealed that the significant consideration of gender diversified boards by investors compelled several corporations to make adjustments. Regarding this trend; in this exploration, we will analyze the relationship between segregation in a gender-diverse board and its strong positive correlation with well-being in Estonia.

Several analyses in board diversity focus more on gender dispersion issue than on age heterogeneity. They also concentrate on gender or age diversity and its effects on a firm's performance more than segregation effects or correlation with the labour market or well-being. An article from Professor Dr Uschi Backes-Gellner and Stephan Veen (2009) [6] stated that age diversity influences positive productivity of innovative companies because of task creativity requirements; this is, however, the opposite for companies with routine tasks. A blog post from A.R. Mazzotta in 2018[7], claimed that age diversity in an organization reduces the turnover of

[5]https://news.err.ee/98113/gender-equality-in-estonia

[6]https://www.pwc.com/us/en/governance-insights-center/annual-corporate-directors-survey/assets/pwc-annual-corporate-directors-survey-2018.pdf

[7]https://www.armazzotta.com/blog/2018/07/10/3-reasons-why-age-diversity-in-the-workplace-is-important/

employees from age 55 or older. Likewise, in the earlier paragraph of this section, we see articles concerning gender diversity in companies boards and their performance. We seldom see studies specific to gender and age segregation amongst the board of directors and its effect on well-being or the labour market. Therefore, we put forward a suggestion that this topic is one of the earliest research for such a problem; but its country of observation is Estonia.

Another intriguing movement in academic fields is the well-being measurement; one might ask these questions: how do we define well-being?; how can we measure a society's welfare?. Or in the aspect of this exploration, one might inquire: can gender and age segregation amongst board members affect well-being?. A known misinterpretation for measuring well-being is using GDP or national pay as measurement, although statisticians never professed this as a measure; there is a general misunderstanding to use it as one. For example, Aristotle, an archaic Greek philosopher and polymath in one of his prominent; and extant work titled "Nicomachean Ethics" described human well-being as happiness; he acknowledged its influence on physical factors in our lives and our position in the society. Likewise, in the 20th century, there were several advocates for an alternative measure to economic progress aside from national income: from Arthur Cecil Pigou, founder of contemporary welfare economics; to Robert Kennedy, an American politician and lawyer; Les Bury an Australian treasurer, Simon Smith Kuznets an American economist and statistician, and few more [15]. The updated shift in policy circles has changed well-being measures to go beyond using GDP; it has fostered subjective and objective well-being measures, creating a more individualistic approach to measuring economic wellness. The various advocation in welfare economics birthed: the holistic approach used in the 20th century, the capabilities approach and preference-based approach used in the 21st century for well-being measurement. These approaches produced several measurement indicators used in measuring economic wellness. Nevertheless, in this exploration, the writer adopted indicators from the holistic and capabilities approach; to produce a new well-being measurement index; which will be discussed further in section 5 of this study.

2.2 Gender diversity in companies' boards

In accordance to MSCI-ESG 2015 Research [19] and its world index enterprise 2012 to 2015 data, companies with higher gender diversity comprising at least a woman director; and one female CEO generated 2.7% more return on equity per year. But the study did not examine the relationship between teams' diverse directors and their corporate performance. Corporations with more than their country's average board diverseness; produced 24% governance-related issues. Likewise, the findings claim to have no evidence that more female managers on seat result in solid risk management. With its three main approaches; it predicted on a global scale; 30% of women filling boards seats in either 2020, 2021 or 2027. The 30% estimate originated from an unmet goal set by the UK and USA; to increase the number of women on management seat by 2015, in this aspect it states the USA is slow-moving compared to other countries.

Like this exploration, they measured women as a minority of companies' board. However, In contrast, it posits no measurement for age group affected by its discoveries; no estimate for discrimination in a gender diversified board and none for the relationship between segregation from gender diversity on boards with corporate performance. Besides, the data adopted in their study is specific to MSCI-ESG corporations; excluding the country used in this exploration and a few other countries.

Katherine Klein's 2017 [18] peer-review analyzed several meta-analyses and over 100 studies from companies' in thirty-five countries and five continents. They revealed that a gender diversified board does not increase or decline companies' performance; such diversified boards have limited or no relationship with company achievement, and it may not give the company a cognitive variety or advantage. It stated that researches suggesting otherwise were from consulting firms or financial institutions; which might not be rigorous, peer-reviewed and academic as theirs. A meta-analysis from Post and Byron 2015 [24] conducted for the same research, showed a small correlation between gender diversity and accounting returns; and a moderate relationship with corporate social responsibility (CSR). The study also stated they are not sure if the correlation meant causation since it not reviewed. Compared to this exploration, Klein's study used gender diversity as a measurement tool; and motivated the author of this exploration to make unbiased realizations and findings. In contrast, this exploration uses gender diversity as a subordinate measurement tool with segregation as the measurement tool; this is because a company can be gender diversified, and they may or may not have segregation. This study estimates the relationship between segregation in companies boards with their corporate financial performance and unlike [18], this study is specific to Estonia.

2.3 Gender Pay Gap in Estonia:

A gender pay gap is the difference in average income women receive compared to men. An article from Sten & Tairi Rõõm 2010 [4] discussed the various factors that influence the gender pay gap in Estonia: the hours both genders worked, the enterprise's ownership type either domestic owned or foreign-owned; their level of education, family factors, work experience; glass-ceiling, segregation, their membership in trade unions, ethnicity, hours worked, the expansion in real estate; the 2007-2008 financial crisis where men were paid higher than the women for working more hours. Sten & Tairi Rõõm article observed that amongst all factors; segregation affects gender pay gap the most. It also explained that if there is gender-diversity of 50% in all occupations in Estonia and there is no segregation, the average gender pay gap will reduce by 32%. Similar to their work this study; will measure the impact of income inequality on the company and society, but it will look at it from the aspect of segregation in companies' boards. This study will use the Gini segregation index from SCube dataset for measuring income inequality in a company's board; and its various segregation indexes as a measurement indicator for segregation in companies board.

In 2010, one of Estonia's broadcasting agencies named ERR; wrote an article[8] of their interview with Mari-Liis Sepper (a gender equality commissioner) about gender inequality in Estonia; it stated that in Europe the directors' board seat occupied by women was 11%. Whereas, in Estonia, women filled boards' seats at 6%. Mari-Liis said Estonia had the highest gender pay gap in the EU; with a 30% larger rate contrasted to other countries in the EU; and high gender-segregation alongside occupational segregation issue. A recent update from Statistics Estonia website [9] claims that from 2019 till the time of this exploration gender pay gap has decreased in Estonia by 0.9%. An analyst at Statistics Estonia website Karina Valma stated that from 2013 to 2019 Estonia gender pay gap has reduced in total by 7.7%. ERR and Statistics Estonia article gave the author of this exploration more insight on income inequality in Estonia; and its changes over the years. Unlike this exploration, Statistics Estonia article analyzed its data using employees; and ERR is a 2010 news update concerning the situation.

2.4 Age diversity and discrimination

In 2018 PwC [10] released an annual survey of public companies' corporate directors at the United States. It stated that 21% of directors consider that age diversity is an important issue; whereas, 46% instead consider gender diversity. Their reports claim that younger executives less than age 50; promotes the company with their innovative skills. An empirical study from Maria Jesus Munoz-Torres, Idoya Ferrero-Ferrero and Maria Angeles Fernandez-Izquierdo [14] claim that age diversity improves companies because all age group in the board has their unique abilities. Their article also suggests that corporate governance guidelines should encourage a board with age or generational diversity; to boost companies' performance in areas of creating diverse views and making deliberated decisions.

In chapter 6 of a 2018 article about modern viewpoints on ageism by Justyna Stypińska and Pirjo Nikander; they explained the conceptual distinction between ageism and age discrimination within the labour market and its effects on older employees. Their studies claim that the issue became a subject of scholarly interest from the first 21st century. It stated that ageism against older workers continues to be prevalent in Europe today. They suggested that an increase in unemployed older workers; might help to reduce ageism in the labour market because it might lead to a change in perceptions of treating older workers; for example, their relationship between work and forced retirements [27]. Nevertheless, the article did not observe ageism on equal terms; it promotes ageism on older workers than on younger employees. Similar to these articles; this study will look at age diversity issue and its relationship with a financial aspect of corporate performance; it will observe their effect on the labour market, well-being and credit risk management. In contrast, it will use segregation measurement in a generation diversified

[8] https://news.err.ee/98113/gender-equality-in-estonia

[9] https://www.stat.ee/news-release-2020-087

[10] https://www.pwc.com/us/en/services/governance-insights-center/library/younger-directors-bring-boardroom-age-diversity.html

board to determine its results.

2.5 Well-being measure:

In 1972, the fourth King of Bhutan named Jigme Wangchuck used the holistic approach to generate (GNH) [12] "Gross National Happiness" Index; ever since then, this idea has influenced the economic and social policy of the world at large including Bhutan itself. His concept of the GNH Index; was to give importance to non-economic aspects of well-being and improve Bhutan's well-being. GNH Index bases on the Alkire-Foster method [30] that uses vigorous multi-dimensional schemes. In 2011, the GNH Index generated the General Assembly resolution adopted by the United Nations; it targeted at encouraging continual happiness and well-being using the holistic approach. In the same year, the Centre of Bhutan Studies released an updated version of GNH Index; it comprises 33 indicators and nine domains which includes common areas related to social and economic interaction. The GNH Index uses a person's "achievements" described as happiness to measure the well-being of a nation; it definition of happiness differs from the usual happiness self-survey that requests an individual to rate how happy or satisfied they are with life. GNH indicators focus on how policies can increase achievements among people who have less of it. [29]

Another historical approach in well-being measure is a theoretical framework called capabilities approach; founded by Indian economist-philosopher Amartya Kumar Sen in 1980 and developed by American philosopher Martha Nussbaum; alongside other scholars in humanities social sciences. Capability approach proposes that well-being depends on; a person's freedom to achieve their functionings meaning: what they value in life, or what they can do [1]. It involves two normative claims: the first is everyone should have the principal and moral freedom to achieve well-being; the second is freedom should depend on people's capabilities [11].

Theoretical strands from this approach led to the creation of several indices one of the few are: Human Development Index (HDI); Gender-related Development Index (GDI); Gender Inequality Index (GDI), Human Poverty Index (HPI) and more [12]. Although, opinions vary on well-being measurement; the theoretical, conceptual ideas that influence well-being research are subjective and objective measure. Holistic and capabilities approach uses both theories, unlike some perspectives that use solely subjective or objective, for example, the use of GDP and national income to measure well-being are objective; the preference-based and the happiness survey test is subjective. This study will adopt some ideas and methodological concept, from the holistic and capabilities approach for well-being measurement. Although there have been several criticisms on both propositions; they remain one of the most proposed means to measure well-being. [13]

[11] https://plato.stanford.edu/entries/capability-approach/

[12] https://en.wikipedia.org/wiki/Capability_approach#Capabilities-based_indices

[13] https://theconversation.com/how-do-we-measure-well-being-70967

2.6 Summary:

Table 2.6 summarizes the related works discussed in this section.

Paper	Objective	Data set	Size of data set	Country	Year of publication	Method	Techniques	Criticism
1	Do women occupy boardroom seats than men? [5]	Time frame (1995-1996 and 2002)	Not included	Tennessee, United States	2004	Qualitative and Quantitative	Not included	Research is only specific to one state in the USA.
2	Men occupy boardroom seats than women [9]	Not included	Not included	Global	April 2002	Not included	Not included	Research only contains works from previous studies
3	Norway's Board Measures and the Female Labour Market [8]	Norwegian Registry data (1986–2014)	None	Norway	2018	Qualitative	Regression	None

Paper	Objective	Dataset	Size of data set	Country	Year of publication	Method	Techniques	Criticism
4	Gender quota effects on boardroom of Germany largest corporations. [10]	Muessing database (2000-2015)	2763 private firms	Germamy	2018	Descriptive	Fixed effects regression	None
5	California and board gender quotas on firm, director and labour market performance [20]	Data entry (12 months till end of September 2018)	2,562 firms	California, United States	2018	Qualitative	OLS regressions	None
6	Gender equality issue in Estonia[14]	None	Not included	Estonia	Oct, 18 2010	None	None	None
7	Changes in the gender composition of boardroom [15]	Directors Survey PwC's (2018)	Survey on 714 directors	United States	May 18, 2017	Descriptive	None	None

Paper	Objective	Dataset	Size of data set	Country	Year of publication	Method	Techniques	Criticism
8	Age diversity impact on a company's performance [6]	LIAB (1993-2003)	18,000 companies with around 2 million employees	Germany	10, Jun 2009	Empirical	(Standard deviation), (Variation coefficient), (Fixed effects estimation)	None
9	Gender pay gap in Estonia [4]	Eurostat database (2000-2008)	Not included	Estonia	2011	Quantitative	Mincer-type regression	None
10	Gender pay gap difference slowly reducing in Estonia [a]	None	Not included	Estonia	Jul, 22 2020	None	None	None
11	Age difference in the board room [b]	PwC's Directors Survey (2018)	Not included	United States	None	Qualitative	None	None
12	Age diversity in the workplace and its importance [c]	None	Not included	Not included	10, July 2018	None	None	None

[a] https://www.stat.ee/en/uudised/news-release-2020-087

[b] https://www.pwc.com/us/en/services/governance-insights-center/library/younger-directors-bring-boardroom-age-diversity.html

[c] https://www.armazzotta.com/blog/2018/07/10/3-reasons-why-age-diversity-in-the-workplace-is-important/

Paper	Objective	Dataset	Size of data set	Country	Year of publication	Method	Techniques	Criticism
13	Board diversity and company's performance [14]	Time frame (2009)	2,152 individual observations from 205 boards	Countries within Europe	2015	Empirical	(OLS) regression	None
14	Age segregation in the labour market [27]	Not Included	Not included	Global (35 countries)	May 2018	Quantitative	None	Promotes ageism for older people than younger people
15	GNH and Bhutan's Happiness Index [29]	None	Not included	Bhutan	2012	None	None	None
16	Capability approach for well-being measures [1]	None	Not included	None	07 Sep 2003	None	Capabilities approach	Not included
17	Gender pay gap a reason for segregation [13]	Time frame (1989-1990, 1993)	2,747,051 samples	United States	1997	Quantitative	Standardization and Regression	None

Paper	Objective	Dataset	Size of data set	Country	Year of publication	Method	Techniques	Criticism
18	About the capability approach used to generate well-being indexes [30]	Time frame (2006, 2008, 2010)	8700 for initial samples, 7142 respondents for final survey	Bhutan	May 2012	Qualitative and Quantitative	Alkire-Foster methodology	None
19	Description of the capability approach [a]	None	Not included	None	Not Included	None	None	None
20	Four approaches used in measuring well being [15]	None	Not included	Global	March 2011	Descriptive	None	None
21	About well-being measurement and the capabilities approach in public policy [25]	None	Not included	None	None	None	Sen's capability approach	None

[a] https://en.wikipedia.org/wiki/Capability_approach#Capabilities-based_indices

22	Gender variety as a global trend in corporate boards [19]	Global director universe (Dec 15, 2009 to Aug 15, 2015)	42,344 board members from 4,218 companies	Global	Nov 2015	Descriptive	None	Data is more specific to financial companies.
23	Gender diversified boards and company's performance [18]	Meta-analyses data	Not included	Not included	May 18, 2017	Descriptive	None	It does not specify the analysis of board diversity issue like segregation.
24	Gender diversified boards with women and a firm's financial performance [24]	Meta-analysis data from 140 studies	More than 90,000 firms samples	5 continents and 35 countries	5 Nov, 2014	Descriptive	None	Board diversity was the principal focus and not the effects of segregation on diversified boards.
25	Bhutan's GNH Analysis [12]	Not included	Not included	Bhutan	2015	Qualitative	Not included	None

Table 1: Summary of related works

In this thesis, to give estimates concerning this work and analyze findings; there is a need to use the right data. The next chapter includes information about handling and processing of data.

3 Data description

This section will give comprehensive details about the data applied in this thesis with a description of its origin, structure, formatting procedure and techniques used to obtain results of findings; as well as an explanation about information concerning raw data processing, data features, data functionalities and importance.

3.1 Data Information

The data used in this thesis originated from two different sources with the names:

i SCube Framework data [7]: developed by researchers at the University of Pisa; a multi-dimensional segregation data cube from SCube system. SCube system[16] is a data mining framework from relational data and attributed graphs; it is a segregation-aware tool used for segregation discovery. SCube data contains information about segregation indexes for boards of directors from 87,330 companies' in Estonia within the year 1998 to 2016. It is used in this work to measure segregation level in companies boards for counties and regions in Estonia; it is available on SoBigData research infrastructure [17]. For clarification purpose, the author of the thesis prefers to refer to it as 'Segregation Data'.

ii Statistics Estonia data [18]: derived from the statistics database of an Estonian government agency that governs data within the country; in this work, the data values help to generate indicators for measuring labour market, corporate performance and well-being of counties within Estonia. Also, The (NUTS) of regions classification is used to retrieve data by region. For clarification purpose, the thesis author prefers to refer to all the data used to formulate labour market indicators as to the 'Labour Market Data'. The one for credit risk management as 'Corporate financial performance Data' and for well-being as 'Well-being Deprivation' data.

3.1.1 Segregation Data:

The raw segregation data contains 21 columns and 87,330 rows; its first row comprises the names or title of the columns. According to the requirements of this study; the author will use ten columns ('Age', 'County', 'Sex', 'Dissimilarity', 'Entropy', 'Gini', 'Isolation', 'Interaction', 'Atkinson', 'timeUnit') from the segregation data to perform analysis and observations. To describe gender diversity level of the boardroom in this raw data; we will use two terms called the majority and minority group; a majority group is the amount of gender in the boardroom with the most population; whereas, a minority group is the one with least population. The Age column in Figure.1 contains the age range of the data, categorized into '16-36', '37-45',

[16] https://docs.google.com/document/d/19A5xyYIQdkzZ_6qpNNI7C4_W9CXeA4vhlWD14DI3JDQ/edit
[17] http://www.sobigdata.eu/access/virtual
[18] https://andmed.stat.ee/en/stat

‘46-54’, ‘55-65’, ‘66-99’. In the County column in Figure.1, there are 15 levels stated as 15 counties (Harjumaa, Pärnumaa, Ida-Virumaa, Tartumaa, Lääne-Virumaa, Viljandimaa, Läänemaa, Saaremaa, Raplamaa, Põlvamaa, Valgamaa, Järvamaa, Võrumaa, Jõgevamaa, Hiiumaa) within Estonia. Properties of the sex column in Figure.1 include gender values (male and female) of the minority group in boardrooms.

residence	age	sex	legalType	shareCapital	status	workersHigh	RegistrationDistrict	County	City	Activities	M
EE	37-45	M	OsaÃ¼hing	2570-3327	1	5-Jan	EEK	Harju	Tallinn	5.783522128	872
EE	37-45	M	OsaÃ¼hing	2570-3327	1	5-Jan	EEK	Harju	Tallinn	5.862897128	967
EE	37-45	M	OsaÃ¼hing	3328-4970124	1	5-Jan	EUR	Harju	Tallinn	8.633900154	633
EE	37-45	M	OsaÃ¼hing	2570-3327	1	5-Jan	EEK	Harju	Tallinn	5.993522128	797
EE	37-45	M	OsaÃ¼hing	2570-3327	1	5-Jan	EEK	Harju	Tallinn	5.762897128	1196
EE	37-45	M	OsaÃ¼hing	2570-3327	1	5-Jan	EEK	Harju	Tallinn	5.698452683	1313

Figure 1: An example image of the first 12 columns of the raw segregation data

The Dissimilarity column in Figure.2 is a segregation indicator used to store distribution values of a minority group in boardrooms; it ranges from the lowest 0 to the highest value 1, with a higher value equating to a higher level of segregation. Entropy column in Figure.2 contains measures of the diversity of all groups in the boardroom; it reaches a minimum of 0 (low segregation) when all the groups respect the global entropy (full integration); and the maximum 1 (high segregation) when the boardroom contains only one group (complete segregation). Gini column as in Figure.2 represents an income distribution inequality measure among board members; similar to the initial indexes; its estimate is from 0 to 1. Atkinson column in Figure.2 is a more sensitive measurement for income inequality; board members with higher segregation level differ in analysis from the one with lower segregation level; it also ranges from 0 to 1. Isolation column in Figure.2, another type of segregation indicator to measure the probability of isolation for the minority group; meaning the exposure level of a minority group to its group; it also ranges from 0 to 1. Whereas, Interaction column in Figure.2 is the opposite meaning their level of exposure to the majority group; used to measure the possibility of interaction among the board of directors; it also ranges from 0 to 1 in contrast to initial indicators, it shows a higher level of segregation when equal to 0 and vice versa. And the ’TimeUnit’ column in Figure.2 stores values of the time frame for all columns and rows within the year 1998 to 2016.

T	Dissimilarity	Entropy	Gini	Isolation	Interaction	Atkinson	partitionGraphA	timeUnit	Regions
1361	0.923171899	0.844869204	0.988057292	0.844452554	0.155547446	0.942611825	filterEdgeGCFor'	1998	Northern Estonia
1361	0.923171899	0.844869204	0.988057292	0.844452554	0.155547446	0.942611825	filterEdgeGCFor'	1998	Northern Estonia
1861	0.923171899	0.844869204	0.988057292	0.844452554	0.155547446	0.942611825	filterEdgeGCFor'	1998	Northern Estonia
1861	0.923171899	0.844869204	0.988057292	0.844452554	0.155547446	0.942611825	filterEdgeGCFor'	1998	Northern Estonia

Figure 2: An example image of the remaining 9 columns of the raw segregation data

3.1.2 Corporate Financial Performance (CFP) Data:

(CFP) data comprises of nine different finance data from statistics Estonia for enterprises with more than 20 or more employees within Estonia. Apart from the income statements data they

are labelled as code EM13[19] - (for time unit 1998-1999), EM12[20] - (for time unit 2000-2004), EM012 [21] - (for time unit 2005-2016) and are formatted and calculated with their specific designated indicators to generate some financial key performance indicators for measuring credit risk. The income statements data in short income data instead has code EM05[22] - (for time unit 1998-1999), EM04[23] - (for time unit 2000-2004) and EM004[24] - (for time unit 2005-2016). For simplification purposes in the process of calculating the financial (KPI), we chose to categorize the several collected enterprise data into (cash spent, current assets, current-liabilities, income, inventories, prepayments, total assets, total liabilities).

(CFP) data includes five financial performance indicators (Current Ratio, Quick Ratio, Burn Rate, Runway, Return on Equity). The cash-spent data from statistics Estonia uses the indicator (Cash, bank and marketable securities) to calculate (Burn Rate) and (Runway) available for all counties in Estonia. The cash-spent data is composed of two columns that include information about the total cash amount spent by companies for the beginning and end of the year; its units are in thousands with currency converted to euros from Estonian kroons. The current liabilities data contains the total current liabilities of companies for each county in Estonia for the beginning and end of the year; it has similar information as cash-spent data, on the contrary. It uses the indicator (Current liabilities total) in place of the cash-spent indicator, and it helps to calculate (Quick Ratio) and (Current Ratio). The current assets data is similar to current liabilities data but with the indicator (Current assets total). The inventories and prepayment data are almost the same as current liabilities data; they use different indicators defined as inventories (Inventories total), prepayment (Prepayment to suppliers) and they help to calculate only the (Quick Ratio). The inventories data contains companies total cost of inventories for counties in Estonia for the beginning and end of the year, and the prepayment data contains the total cost of prepayment made to suppliers. The income data includes information on several enterprises income statement; it has a similar structure as previous data; with its financial indicator (Financial Income) used to calculate (Return on Equity). The total assets and liabilities data is similar to the income data. They do not measure income statement; they have different indicators for measurement named as (assets total) and (liabilities total); they contain information about total assets and liabilities of companies for each county in Estonia.

[19] https://andmed.stat.ee/en/stat/majandus__ettevetete-majandusnaitajad__ettevetete-vara-kohustused__aastastatistika/EM13

[20] https://andmed.stat.ee/en/stat/majandus__ettevetete-majandusnaitajad__ettevetete-vara-kohustused__aastastatistika/EM12

[21] https://andmed.stat.ee/en/stat/majandus__ettevetete-majandusnaitajad__ettevetete-vara-kohustused__aastastatistika/EM012

[22] https://andmed.stat.ee/en/stat/majandus__ettevetete-majandusnaitajad__ettevetete-tulud-kulud-kasum__aastastatistika/EM05

[23] https://andmed.stat.ee/en/stat/majandus__ettevetete-majandusnaitajad__ettevetete-tulud-kulud-kasum__aastastatistika/EM04

[24] https://andmed.stat.ee/en/stat/majandus__ettevetete-majandusnaitajad__ettevetete-tulud-kulud-kasum__aastastatistika/EM004

3.1.3 Labor Market Data:

The labour market data comprises of four different data from Statistics Estonia; merged into segregation data to solve this thesis first research question. The four raw data used as labour market indicators were collected to have a time frame of 1998 to 2016, similar to the segregation data. The first data derived from Estonia Statistics with code (TT240)[25] is the annual average employment rate which includes information about yearly employment rate about the entire country with the indicators (Males' employment rate, %) and (Females' employment rate, %); it uses percentage unit as measurement. The second data with code (TT453) [26] contains information about inactive people in the labour force it comprises of data about the entire country; each county in Estonia and persons from age 16 until pension; it has the unit thousands for the overall population of inactive people in the country. The third data with code (TT43) [27] contains information about the total unemployed person (male and female) in the country with their duration of unemployment. The unemployment duration has a period of (less than six months, 6 to 11 months, 12 months or more and 24 months or more, total duration), and its measurement unit is in thousands. The fourth data with code (TT50) [28] is about the rate of unemployment by regions (Northern Estonia, Western Estonia, Central Estonia, Southern Estonia, Northeastern Estonia); its measurement unit is in percentage.

3.1.4 Well-being Data:

Well-being data is composed of the labour market data, (CFP) data and other additional data from statistics Estonia like health data, poverty rate data and wages data. The well-being data is related to the objective well-being approach because, during data collection, we could only obtain data for this approach. Its time frame is from 2004 to 2016 because the health data is available within this period.

Health data from Estonia statistics with code (TH79)[29] stores health information of people older than 16. It has three indicators identified as (very good or good), (neither good nor bad), (bad or very bad). These indicators have a total percentage of values for citizens health status by regions. The poverty data with code (LES20) [30] has an (Absolute poverty rate %) indicator that stores information about the overall poverty rate by county and regions; this thesis uses only information about region poverty rate. The wages data with code (PA5321) [31] contains average monthly net wages of income earners by county; it is calculated in Euros and uses the (Average monthly net wages (salaries), euros) indicator for measurement.

[25] https://andmed.stat.ee/en/stat/sotsiaalelu__tooturg__heivatud__aastastatistika/TT240

[26] https://andmed.stat.ee/en/stat/sotsiaalelu__tooturg__mitteaktiivsed__aastastatistika/TT453

[27] https://andmed.stat.ee/en/stat/sotsiaalelu__tooturg__tootud__aastastatistika/TT43

[28] https://andmed.stat.ee/en/stat/sotsiaalelu__tooturg__tootud__aastastatistika/TT50

[29] https://andmed.stat.ee/en/stat/sotsiaalelu__tervishoid__tervislik-seisund/TH79

[30] https://andmed.stat.ee/en/stat/sotsiaalelu__sotsiaalne-terjutus-laekeni-indikaatorid__vaesus-ja-ebaverdsus/LES20

[31] https://andmed.stat.ee/en/stat/majandus__palk-ja-toojeukulu__palk__aastastatistika/PA5321

3.2 Data preparation

3.2.1 Segregation Data Imputation:

Segregation raw data contains 31% missing values; one way to handle this missing values problem is with likewise deletion method, but in this thesis, we will neglect this method because of the possibility of losing important information. To manage the issue, we used a nonparametric imputation method (missForest) [32] it is one of the robust packages of R for implementing (random-forest) algorithm. We used the tree-based algorithm (random-forest) because it is one of the best-supervised learning methods; it supports non-linearity and is easy to interpret; enables high stability with predictive models [33].

A nonparametric computation method is a distribution-free statistic type. With nonparametric computation, the analyzed population data does not have to meet specific assumptions or distribution types. It does not require precise assumptions about the operative form of a function f- (a random function) [34]. Instead, it estimates f in a realistic way close to its data points. Another important reason for using (missForest) library is the fact that it produces an OOB- Out of the bag error estimate to explain the accuracy of the imputation. The error estimate shows two types of errors; the first one is NRMSE (normalized mean squared error) used to report error received from the imputation of continuous values. PFC (proportion of falsely classified) describes error obtained from the imputation of categorical values. Table 2 contains the error estimate for the imputed data; we will use the imputed data for segregation analysis.

Error-type	Error-rate	Error-rate (%)
NRMSE	0.00002537149296535	0.00254%
PFC	0.0211589949362259	2.12%

Table 2: Data Imputation Error Estimate OOB (out of bag)

Also, from the perspective of data collection, we chose data imputation to increase the probability of having more female data. We used the indexes (Dissimilarity, Entropy, Atkinson, Interaction, Gini, Isolation) from segregation data as a measurement of segregation because one of the main requirement of the thesis is to verify initial findings made by few researchers at the University of Tartu on this subject. Results from their research case study showed a negative correlation between unemployed young men from Lääne-Virumaa, Estonia and the county's unemployment rate for the year 2008 to 2015. But further studies like the one from this thesis is required to verify this hypothesis.

[32] https://cran.r-project.org/web/packages/missForest/missForest.pdf

[33] https://www.analyticsvidhya.com/blog/2016/04/tree-based-algorithms-complete-tutorial-scratch-in-python/

[34] https://www.analyticsvidhya.com/blog/2016/03/tutorial-powerful-packages-imputing-missing-values/

3.2.2 Segregation Data Formatting:

The segregation data needs to be formatted because we will merge it with the labour market, (CFP) and well-being data, and it needs to fit into their requirement. We will use the combined data to find the results of correlation and causal relationship for all research questions. To obtain findings based on region, we will categorize the segregation data with (NUTS). The NUTS (Nomenclature des Unités Territoriales Statistiques) categorization is a statistical classification of regional units in European Union used to gather, produce and disseminate regional statistics [35]. It is important to note that the unformatted segregation data contains segregation indexes values of several companies for each county in a year; we need to aggregate this data to have an average of segregation index by county. We used equation (1) and (2) to solve this problem, equation (1) is without variables and equation (2) is the opposite. Let (S) be the mean of segregation indexes by county, (t) be the total count of companies with segregation indexes, (si) is a segregation index of each company in a county.

$$S = \frac{\text{(sum of segregation indexes of companies)}}{\text{(total count of companies)}} \tag{1}$$

$$S = \frac{\sum_{i=1}^{t} si}{t} \tag{2}$$

3.2.3 Generate Labor-Market Data:

To generate the labour-market data; we should edit and format the statistics data before merging it with segregation data because the statistics Estonia labour force datasets have units in percentage and thousands. For this work, we prefer to use a percentage unit for all measurement of datasets indicators to provide clarity during data observation; as a result, we generated a simple formula for this conversion. The conversion formula(3) is (R) - rate or indicator value converted to a percentage; (Px) - dimension value in thousands divided by (Py) - average population of the country for the given year multiplied by 100. The formula is valid for inactive persons and unemployed persons by duration data because their unit is in thousands.

$$R = (\sum_{t=1}^{Py} \frac{Px}{Py}) * 100 \tag{3}$$

3.2.4 Corporate Financial Performance (CFP) Indicators:

To generate the (CFP) data we have to merge the (CFP) indicators with the segregation data; In general, there are several indicators used to measure companies finances. In this thesis, we used five financial key performance indicators for short (KPI) for this measurement. We decided to use corporate performance measures to handle credit risk management because we looked at a

[35] https://vana.stat.ee/296050

scenario of a company high in debt or liabilities, and noticed that there is a high probability of such company to default in the loan payment. And if a company is unable to refund the cash or services owed this is a risk called credit risk for the borrower that issued money to such a company. So how do we manage such risk? We suggest to find the root cause of the issue or deal with the problem that triggers such risk. What if the root cause of the problem is segregation? Does a high level of segregation among boards of members play a role in reducing company financial performance? Can low company financial performance lead to default in debt payment leading to credit risk for the borrower of the company? Hence, we also ask does segregation have a relationship with companies performance that leads to credit risk? We assume that if company performance is low; this might increase the chances of credit risk for the borrower, and if company financial performance is high; it might reduce the probability of credit risk. The (CFP) data will combine these various finance statistics indicator data to measure the chances of credit risk occurrence.

From the (CFP) data, we generated five financial key performance indicators (Current ratio, Quick ratio, Burn rate, Runway and Return on Equity) for credit risk measurement. In general, there are more than 20 financial key performance indicators to measure corporate financial performance; however, for this work, we collected and used these indicators based on the availability of the financial data and our intuition the (KPI) that fits into the project requirement.

Current Ratio [36] measures the ability for companies in specific counties to pay their short-term obligations due within a year. Equation (5) and (4) interprets Current Ratio as

$$CR = \frac{\text{(Total current assets for all companies in a county at year-end)}}{\text{(Total current liabilities for all companies in a county at year-end)}} \tag{4}$$

$$CR = \frac{\sum_{c=1}^{Ca} CA}{\sum_{c=1}^{Cl} CL} \tag{5}$$

Where (CR) is the Current Ratio; (CA) is Current Assets and (CL) as Current Liabilities. We chose the end of the year as this information will contain the latest update of their finance. Quick ratio [37] tells if a company short-term assets are enough to cover its near-future liabilities. It gives a more accurate survey of a business financial health than the Current Ratio that ignores liquid assets such as inventories. Equation (10) defines Quick Ratio (QR), and it includes equation (6) (CAt) which equals (current assets for all companies in a county at year-end), equation (9) (INt) meaning (inventories for all companies in a county at year-end), equation (8) (PRt) as (prepayments for all companies in a county at year-end) and equation (7) (CLt) as (current liabilities for all companies in a county at year-end).

$$CA_t = \sum_{c=1}^{ca} CA \tag{6}$$

[36] https://www.investopedia.com/terms/c/currentratio.asp

[37] https://www.scoro.com/blog/financial-kpis-for-financial-kpi-dashboard/

$$CL_t = \sum_{c=1}^{cl} CL \tag{7}$$

$$PR_t = \sum_{p=1}^{pr} PR \tag{8}$$

$$IN_t = \sum_{i=1}^{in} IN \tag{9}$$

$$QR = \frac{(CA_t) - (IN_t) - (PR_t)}{(CL_t)} \tag{10}$$

Burn rate [38] describes the rate at which a new firm spends its enterprise capital to finance cost before generating positive cash flow from operations; the spending can be weekly, monthly or annual basis. In this work, Burn rate [39] tells if the organization's annual operating costs are sustainable for a longer-term. Equation (13) represents Burn rate (BR), it includes equation (12) CBt for (Cash spent for companies in a county at the start of the year), equation (11) CEt as (Cash spent for companies in a county at the end of the year) and Mt the denominator variable in equation (13) is explained as (total number of months in a year).

$$CE_t = \sum_{c=1}^{ce} CE \tag{11}$$

$$CB_t = \sum_{i=1}^{cb} CB \tag{12}$$

$$BR = \frac{(CB_t) - (CE_t)}{M_t} \tag{13}$$

The runway indicator measures how much time a company has before it goes out of cash. Equation (15) shows Runway and (14) interprets it as follow:

$$RW = \frac{\text{(cash amount spent by companies in a county year-end)}}{\text{(Burn Rate)}} \tag{14}$$

$$RW = \frac{CE_t}{BR} \tag{15}$$

Return on Equity [40] indicator tells how a business is capable of using shareholder's investments to generate high profits. It measures the profitability of a firm before taxes per shareholder equity unit[41]. In equation (16) Return on Equity (RE) is interpreted as

$$RE = \frac{\text{(total financial income of companies in a county)}}{\text{(total average shareholder equity per county)}} * 100 \tag{16}$$

[38] https://www.investopedia.com/terms/b/burnrate.asp

[39] https://www.scoro.com/blog/financial-kpis-for-financial-kpi-dashboard/

[40] www.investopedia.com/terms/r/returnonequity.asp

[41] https://www.scoro.com/blog/financial-kpis-for-financial-kpi-dashboard/

To derive Equation (16) we use FIt the total financial income of companies in a county. Financial income also (book income) is a report of the corporation performance for its shareholders; it tells the financial interest of the firm before taxes.

$$FI_t = \sum_{c=1}^{fi} FI \tag{17}$$

And S$avgt$ in equation (18) as the total average shareholder equity per county. The average shareholder equity is a credible averaging concept used to level out estimation results of measurement about return on equity. And it is estimated as

$$Savgt = \frac{\text{(shareholder equity at year-start + shareholder equity at year-end)}}{2} \tag{18}$$

In equation (19) we set SBt as (shareholder equity for the start of the year) and SEt in equation (20) as (shareholder equity for the end of the year).

$$SB_t = (AB_t) - (LB_t) \tag{19}$$

$$SE_t = (AE_t) - (LE_t) \tag{20}$$

Equation (19) SBt accepts variable ABt described in equation (21) and LBt in equation (23); ABt from equation (21) returns total assets of companies in a county at the beginning of the year. And LBt from equation (23) shows companies total liabilities at the beginning of the year for one county. Also, in equation (20) SEt accepts AEt shown in equation (22) and LEt in equation (24). AEt from equation (22) returns total assets of companies in a county at the end of the year. LEt in equation (24) as companies total liabilities at the end of the year for one county.

$$AB_t = \sum_{i=1}^{ab} AB \tag{21}$$

$$AE_t = \sum_{i=1}^{ae} AE \tag{22}$$

$$LB_t = \sum_{i=1}^{ab} LB \tag{23}$$

$$LE_t = \sum_{i=1}^{ae} LE \tag{24}$$

Hence we derived the equation (25) for average shareholder equity S$avgt$ used as the denominator for Return on Equity (RE). Equation shows that RE (26) accepts the results of (S$avgt$) and (NIt).

$$S_avgt = (SB_t + SE_t)/2 \tag{25}$$

$$RE = \frac{NI_t}{S_a v g t} \tag{26}$$

3.2.5 Wellbeing Deprivation Data:

There have been several options as to the measurement of well-being; for numerous years, researchers have debated the right way to measure well-being. In this work, we will not suggest the right or wrong way for well-being measurement, instead will use available welfare data from Estonian statistics to cover many aspects of measuring well-being. We collected 11 indicators from statistics Estonia data to measure well-being. The labour market indicators (employment gap, employment rate and unemployment by duration), (CFP) indicators (Current ratio, Burn rate, Runway, Quick ratio, ROE (Return on Equity)), Health indicator, Poverty rate indicator and average wages per county indicator. In this section, for clarity, we will refer to these indicators as well-being indicators. To use this data for measurement we will merge it with segregation data.

To generate the (well-being deprivation) data we used the Alkire-Foster (AF) [1] methodology because of its simplicity and clarity; it is easy to compute; it is also intuitive; it conveys information on multiple deprivations and combines them. It is applicable and expandable; it is valid for both ordinal and cardinal data. We used two main basic techniques from the Alkire-Foster (AF) method [2]; they are (dimensional cutoffs and union identification) [3]. Dimensional cutoffs determine the deprivation in a dimension or well-being data indicator. Union identification validates the deprived in the sense that if there is at least one deprived, the data indicator is lacking in well-being.

Here is how we used the methodology in this work, we liken each step to a matrix and each data indicator as a dimension in a matrix. Let (I) be well-being indicators, Let (I1) be employment rate indicator, (I2) be current ratio indicator, (I3) be the health indicator with (Very good or good) health values. (X) as the data to check for deprivation, (c) as deprivation or dimensional cutoffs for each dimension. So we suggest that the employment rate has a cutoff of 60% if the value is more than 60% there is deprivation; the same for Current ratio it values should be between the range of 1.2 to 2.0, and if the health indicator value is below 90% there is a deprivation.

$$\text{X} = \begin{Bmatrix} I_1 & I_2 & I_3 \\ 70.1 & 1.36 & 58.1 \\ 58.5 & 2.06 & 92 \\ 66 & 1.69 & 49.7 \end{Bmatrix}$$

$$\text{c} = \begin{pmatrix} 60 & (1.2-2) & 90 \end{pmatrix}$$

We replace the matrix with (1) if the dimension has deprivation and (0) if it does not. Hence, we generate the deprivation matrix (D).

$$\mathrm{D} = \begin{Bmatrix} I_{d1} & I_{d2} & I_{d3} \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{Bmatrix}$$

Let (Dc) be the total deprivation count and (Idc) be the vector of the deprivation count. And (Dm) be the sum of dimensions or data indicators, let (Idm) be its vector.

$$\mathrm{Dc} = \begin{Bmatrix} I_{dc} \\ 2 \\ 1 \\ 2 \end{Bmatrix} \mathrm{Dm} = \begin{Bmatrix} I_{dm} \\ 3 \\ 3 \\ 3 \end{Bmatrix}$$

We go on further to calculate the intensity of the deprivation (Di). The result of (Di) matrix will return well-being deprivation vector (did) that tell to what extent there is a deprivation in the data or dimension. We used this vector as a measure for well-being.

$$\mathrm{Di} = \begin{Bmatrix} di_d \\ 2/3 \\ 1/3 \\ 2/3 \end{Bmatrix} = \begin{Bmatrix} di_d \\ 0.66 \\ 0.33 \\ 0.66 \end{Bmatrix}$$

Deprivation cutoffs: For every well-being indicator, we used specific deprivation cutoffs to determine its deprivation. Here is a list of the dimensional cutoffs based on their categories:

Indicators	Deprivation cut-off (c)	Deprivation (d)
Employment gap	5%	greater than 5%
Employment rate	60%	greater than 60%
Unemployment by duration	4.5%	greater than 4.5%

Table 3: Dimensional cutoffs for labor market indicators

Table 3 displays the cutoffs (c) list for labour market indicators and (d) as the requirement passed to have deprivation; we can see that the employment gap [42] between male and females is set to 5%, if the value is more than 5% there is a deprivation. The employment gap, employment rate and unemployment by duration follow the same procedure. The employment gap and employment rate cutoff can be debatable because it is chosen based on intuition. The reason behind choosing such value as cutoff is because we wanted a one that is not too small; that is close to perfect or perfect as this might be unrealistic, and it is not too high. We chose 4.5% as unemployment by duration cutoff because according to Federal Reserve claims; natural unemployment rate falls between 3.5% and 4.5% [43].

[42] https://www.ilo.org/infostories/en-GB/Stories/Employment/barriers-womenintro

[43] https://www.thebalance.com/natural-rate-of-unemployment-definition-and-trends-3305950

Indicators	Deprivation cut-off (c)	Deprivation (d)
Current Ratio	1.2 to 2	outside the range of (1.2 to 2)
Quick Ratio	1	less than 1
Burn Rate	0	greater than 0
Runway	0.5	less than 0.5
Return on Equity	15% to 20%	outside the range of (15% to 20%)

Table 4: Dimensional cutoffs for (CFP) indicators

Table 4 displays the cutoffs (c) list for (CFP) data indicators; Current ratio cutoff is within the range of 1.2 to 2 because this is good current ratio for any business [44]. It is the same procedure for Quick ratio [45]; a number less than (1) might mean that a company does not have sufficient liquid assets to cover its current liabilities. And the Return on Equity as well, 15 to 20% is in general considered good value to have for a company. The burn rate [46] cutoff is (0) because a negative value means that the company is spending more money than they earn[47]. Runway cutoff is 0.5, which translates to 6 months and (1) translates to 12 months; because Runway [48] estimation is for annual expenditure.

Indicators	Deprivation cut-off (c)	Deprivation (d)
Health (Very good or good)	90%	greater than 90%
Health (Neither good nor bad)	70%	greater than 70%
Health (Bad or very bad)	3%	greater than 3%
Absolute poverty	3%	greater than 3%

Table 5: Dimensional cutoffs for (Health and Poverty) indicators

Table 5 displays the cutoffs (c) list for (Health and Poverty) indicators; the health data from Statistics Estonia contains three indicators for health measurement; these indicators (Very good or good), (Neither good or bad), (Bad or very bad) measure the overall health rate by county. We chose 90% as a cutoff value for the Health- (Very good or good) based on the highest score of healthiest people in the world, according to 2019 ranking, Spain had a score of 92% [49]; this score inspired our cutoff value; the cutoff value for the remaining indicators are also from on intuition.

[44]https://www.freshbooks.com/hub/accounting/good-liquidity-ratio
[45]https://investinganswers.com/dictionary/q/quick-ratio
[46]https://www.liveplan.com/blog/metrics-in-a-minute-cash-burn-rate/
[47]https://founderscpa.com/calculating-burn-rate-runway-startups/
[48]https://scalefactor.com/ask-the-experts/what-is-cash-runway/
[49]https://worldpopulationreview.com/country-rankings/healthiest-countries

Indicators	Deprivation cut-off (c)	Deprivation (d)
Wages (2004)	(2480 kroons) as 158.50 Euros	less than 158.50 Euros
Wages (2005)	(2690 kroons) as 171.92 Euros	less than 171.92 Euros
Wages (2006)	(3000 kroons) as 191.73 Euros	less than 191.73 Euros
Wages (2007)	(3600 kroons) as 230.08 Euros	less than 230.08 Euros
Wages (2008)	(4350 kroons) as 278.01 Euros	less than 278.01 Euros
Wages (2009)	(4350 kroons) as 278.01 Euros	less than 278.01 Euros
Wages (2010)	(4350 kroons) as 278.01 Euros	less than 278.01 Euros
Wages (2011)	278,02 Euros	less than 278,02 Euros
Wages (2012)	290 Euros	less than 290 Euros
Wages (2013)	320 Euros	less than 320 Euros
Wages (2014)	355 Euros	less than 355 Euros
Wages (2015)	390 Euros	less than 390 Euros
Wages (2016)	430 Euros	less than 430 Euros

Table 6: Dimensional cutoffs for (Wages) indicators

Table (6) displays the cutoffs (c) list for (Wages) indicators; it uses information about standard wages cut off from the Estonian Tax and Customs Board. The cut off value is a monthly wage of full-time work from the year 2004 to 2016; it is according to the time frame of the well-being data.

3.3 Data Segregation:

In this thesis, to measure data relationship with segregation, and check for even distribution among board members, each data is separated based on gender (Females and Males), age group (‘16-36’, ‘37-45’, ‘46-54’, ‘55-65’, ‘66-99’) and either their county or regions. The counties are (Harjumaa, Pärnumaa, Ida-Virumaa, Tartumaa, Lääne-Virumaa, Viljandimaa, Läänemaa, Saaremaa, Raplamaa, Põlvamaa, Valgamaa, Järvamaa, Võrumaa, Jõgevamaa); regions are (Northern Estonia, Western Estonia, Central Estonia, Southern Estonia, Northeastern Estonia).

After segregating the data, we apply statistical methods that estimate the relationships. In the next chapter, we will describe the testing phase of each segregated data for results received from such implementation process.

4 Hypothesis Testing

A hypothesis is a conditional assumption less than an approved theory, where its logical or practical outcomes need to be tested and proven as true [50]. Hypothesis testing in statistics is a way to test the results of a hypothesis for meaningful results [51]. The purpose of hypothesis testing in this work is to determine if there is enough statistical evidence in support of the research questions. Even with such evidence, this does not prove 100% that the research hypothesis is correct. Instead, it gives evidence that there is a slight probability results from the hypothesis testing occurred by chance. This section will focus on the procedure and techniques used to test such a condition.

The three research questions are:

i Is there a relationship between gender and age segregation among boards of directors in Estonia and the labour market?

ii Is there a relationship between gender and age segregation among boards of directors in Estonia and credit risk management?

iii Is there a relationship between gender and age segregation among boards of directors in Estonia and the country's well-being?

To test for the relationship in all research questions, we will use correlation and causation statistical measure to determine if there is a correlation or causal relation among each research question. These statistical measures require the use of two types of statistical hypothesis for testing relationships: the null hypothesis (that states that there is no difference between the characteristics of a population) and the alternative hypothesis (that states a difference). The next paragraph shows an example of both hypothesis type. In the example, the term (indicator) can be either labour market indicators (employment rate, inactive persons, unemployment by duration, unemployment by regions); financial key performance indicators (Current ratio, Quick ratio, Runway, Burn rate, Return on equity) for credit risk and well-being deprivation indicators.

For example, given the research question: Is there a relationship between segregation and (indicator)?

i Null hypothesis: There is no (correlation or causation) between (segregation index) and (indicator).

ii Alternative hypothesis: There is (correlation or causation) between (segregation index) and (indicator).

Before accepting a hypothesis, it needs to have statistical and practical significance. During hypothesis testing, we make use of the $null$ and $alternative$ hypothesis to test for statistical

[50] https://www.merriam-webster.com/dictionary/hypothesis

[51] https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/

significance; the condition is either accepted or rejected based on its statistical significance. In this work, we will determine statistical significance for a causal relationship with F-test and use T-test for correlation relationship.

4.1 Causation statistical significance (Wald test):

For testing the hypothesis of causal relationship we used the (lmtest) package [16] from R; (lmtest) package has a function (grangertest) that calculates the causal relationship for each research questions and handles significance tests. (grangertest) uses a statistical concept known a granger causality for measuring relationship. It returns restricted and unrestricted models [52], residual degree of freedom, (degree of freedom), the wald test and p–value [53]. The degree of freedom [54] is the number of values in the final estimation that can change. The Wald test from (grangertest) function compares the unrestricted model regarded as (Model 1) which includes granger causality terms, and the restricted model regarded as (Model 2) with no granger causality terms to determine their significance. The comparison gives a p–value that informs if the null hypothesis is statistically significant.

```
Granger causality test

Model 1: finance_data[["QuickRatio"]] ~ Lags(finance_data[["QuickRatio"]], 1:2) + Lags(finance_data
[["Interaction"]], 1:2)
Model 2: finance_data[["QuickRatio"]] ~ Lags(finance_data[["QuickRatio"]], 1:2)
  Res.Df Df      F Pr(>F)
1  87323
2  87325 -2 0.4835 0.6166
```

Figure 3: Sample of Granger causality test result using females (CFP) data

4.2 Correlation statistical significance (t-test):

A t-test is a type of inferential statistics used in this work to determine the statistical significance of the hypothesis; in general, it checks if two population means are reliably different from each other. It uses the t-statistics, t-distribution, degrees of freedom and p-value for this purpose. The two main methods for calculating correlation relationships in this work are Pearson and Spearman correlation; during the calculation of these methods, one can test for their hypothesis significance as well. For testing the significance of (Pearson and Spearman) association, we used t-test because it is a common standard and a statistical procedure used for a similar case in many studies. Also, there is a package in R (a statistical software computing environment) that computes t-test and correlation association between paired samples. We will make use of this package for the significance test of each hypothesis. This R stats package uses (cor.test) function to test for such association. Significance tests result from the (cor.test) function returns t-statistic, degrees of freedom for t-distribution; confidence interval level from (Fisher Z

[52] https://stats.stackexchange.com/questions/131261/granger-causality-interpretation-using-r

[53] https://www.rdocumentation.org/packages/lmtest/versions/0.9-38/topics/grangertest

[54] https://en.wikipedia.org/wiki/Degrees$_o f_f reedom_(statistics)$

transform), and p-value for Pearson correlation. And for Spearman correlation, it returns only t-statistic and p-value. Below is an example of a result obtained for both cases:

```
t = 31.184, df = 68996, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1105259 0.1252417
sample estimates:
      cor
0.1178902
```

Figure 4: Sample of Pearson correlation t-test result using females (CFP) data

```
data:  finance_data_F[["CurrentRatio"]] and finance_data_F[["Atkinson"]]
S = 1.1443e+12, p-value < 2.2e-16
```

Figure 5: Sample of Spearman correlation t-test result using females (CFP) data

From the example above, we can see that the t-test result from Pearson correlation or association contains (t, cor, df, confidence interval, and p-value) values, and t-test result from Spearman correlation or association returns values for (S, $p\text{-}value$). Spearman correlation t-test result did not give the degree of freedom and confidence interval because it does not use t-distribution to make assumptions about the data distribution. The t and S value [55] is the t-test statistics [56] used to determine if the $null$ hypothesis should be accepted or rejected. The df value in full degrees of freedom [57] is the probability distributions of the test statistics for this hypothesis tests it uses the t-distribution [58] to estimate the statistical significance of the $null$ hypothesis. The result value (cor) is the correlation value or Pearson correlation coefficient (r) it gives information about correlation association strength. The confidence interval value is a range to estimate if the value of the correlation value (r) falls between the minimum and maximum confidence interval value. In this work, If (r) is not within the interval range, we reject the null hypothesis. The 95% confidence interval information explains the level of certainty of the correlation coefficient. The p–value helps to determine the probability that a hypothesis result is statistically significant; it is expressed as a value from 0 to 1.

By study standards, the p–value when compared with significance level denoted as (α) can determine when to reject or accept the null hypothesis. The significance level is usually chosen before data collection and often set or required to have a value of (0.05) or 5% sometimes much lower or higher depending on the field of study. In this section, we questioned the traditional

[55] https://blog.minitab.com/blog/adventures-in-statistics-2/understanding-t-tests-t-values-and-t-distributions
[56] https://en.wikipedia.org/wiki/T-statistic
[57] https://statisticsbyjim.com/hypothesis-testing/degrees-freedom-statistics/
[58] https://www.investopedia.com/terms/t/tdistribution.asp

approach of using α = 0.05 as the significance level because we wanted to choose an optimal value that is less error-prone. We decided to look deeper into the criteria of choosing a significance level and noticed that it is possible to have a value other than the standard 5% as long as it is optimal for the research work and data.

4.3 Choosing an optimal significance level for Correlation:

The significance level denoted as (α) is a specific value set to accept or reject a $null$ hypothesis. If the significance level is lesser than a p–value the $null$ hypothesis is rejected, and the $alternative$ hypothesis is accepted. Whereas, if the significance level is higher than p–value, then the $null$ hypothesis is accepted, and the $alternative$ hypothesis not accepted [17]. When contemplating to choose an optimal significance level, one should take into account several considerations like the statistical power of the hypothesis test, the sample size, the effect size, the consequences or risk of having a Type I and Type II error. Type I error expressed as α is the error made when we refuse a correct null hypothesis and take the $alternative$ hypothesis. Type II error denoted as β is the error made when we reject a correct $alternative$ hypothesis and choose the $null$ hypothesis. The decision made on significance level $alpha$ or p–value can increase or decrease the likelihood of having a Type II error $power$ (1 - β). Reducing the chance of having a Type I error $power$ (1 - α) will increase the probability of Type II error. [22].

It is preferable to know the risk of having a Type I and Type II error for each hypothesis test; that is to measure the cost or disadvantage of having a Type I error in comparison with a Type II error. Based on the null hypothesis of this work, there is no advantage and disadvantage in having a Type I error. Instead, there is a disadvantage in having a Type II error as this might result to an increase in credit risk; adverse effect on the well-being of the society; rise in the rate of unemployment, and one can risk declining the labour market progress. In this work, Type II error costs more than the Type I error for the null hypothesis, and we cannot afford to make such error. To reduce the chances of having a Type II error, we will use the pwr package [11] from R. This package accepts four quantities: which are the sample size; the effect size (correlation value); significance level also (the chance of having a Type I error) and statistical power of the hypothesis test meaning (the probability of having a Type II error). The pwr function takes a compulsory three inputs the input not set can be auto-generated. We used (pwr.r.test) and (pwr.t.test) function from pwr package and provided three inputs; the function calculates the fourth input as its result [59]. In this situation, we provided the first three inputs the sample size (row number for segregated data), the effect size (correlation value from the correlation test) and statistical power (set to 60% because we cannot afford to make a type II error). And in return, the optimal significance level is given as a result of the pwr function.

In the next chapter, there is a description of when the significance level is applied and a more detailed explanation of the correlation and causation methods used for measurement.

[59] https://www.statmethods.net/stats/power.html

5 Methods

This section discusses the techniques used in this research work. It explains how we tested the research questions and the methods used for evaluation. In this thesis, we worked on experimental research that uses a quantitative method. It originated from a study of social segregation among a network of companies by researchers at the University of Pisa; where they created a segregation-aware framework SCube for measuring segregation and proposed it as a measurement tool for segregation. Researchers at the University of Pisa and Tartu did the exploratory research of this work and analysed segregation based on gender and age among boards of directors in Italian companies. The researchers at the University of Tartu tested the framework from the occupational segregation study at the University of Pisa on a case study of unemployed young-men in Laane-Viru county in Estonia for 2008 to 2015; results showed a negative correlation of unemployment rate with isolation index.

This work is required to test the segregation framework as an analyst and verify hypothesis from the earlier study in Laane-Viru county with more case studies in Estonia. It should examine the issue from a broader outlook other than the unemployment rate by using different factors that affect the labour market; credit-risk and well-being. Results from this study can lead to the creation of techniques that might help predict daily or weekly changes in the Estonian labour market, or find one of the root causes of credit risk for financial companies in Estonia. And help to study the effects of a successful business in society. We used the relationship between segregation in company boards and these various factors to generate estimated results by using correlation and causation technique.

5.0.1 Correlation Method:

Correlation (r) is a statistical method used to measure the relationships between two or more variables; its result is a coefficient often within the range of -1 and +1. For instance, to determine the association between two variables, where one is the dependent variable (y), and the other is the independent variable (x). We estimate a correlation coefficient that measures the strength of the relationship; where the dependent variable is the effect or result of the events and is affected by changes from the independent variable; and the independent variable causes the event to happen. In this work, for correlation measurement, the dependent variable is either the labour-market or credit risk or well-being indicators, and the independent variable is segregation indexes. There are two different methods of performing correlation analysis the parametric also Pearson correlation (r) and non-parametric that is Spearman and Kendall correlation.

5.0.2 Correlation Tests:

We adopted either the Pearson or Spearman correlation method for correlation tests because they are used often for detecting associations between variables. We applied the Spearman

method for whenever the transformed and non-transformed data fails to meet Pearson correlation assumptions. In the case of non-linear data, a transformation of data means converting it to linear data. If the transformed data or non-transformed data fails to meet the requirement for the Pearson correlation method, then we use the Spearman method. For data transformation, we used the logarithmic transformation to convert the highly skewed dataset to linear data. In this study, the data transformation case was valid for only female-segregated well-being data.

Pearson correlation will measure the linear relationship among two variables with the occurrence of the p-value to test for statistical significance; show their relationship strength and direction using the Pearson coefficient r that is between the range of -1 and +1. The coefficient size measures the correlation strength with 1 being a perfect positive correlation and 0 as no relationship. The relationship strength increases when the coefficient r is further away from 0. One of the steps for analysing the data used to calculate Pearson correlation; involves reviewing the data to make sure it is suitable for analysis to generate a valid result. It is only proper to use Pearson correlation if the data passes the assumptions expected for Pearson's correlation.

Table 7 has a summary of Pearson Correlation assumptions. The first assumption requires that the dependent and independent variables should be continuous; if this requirement fails, we use the Spearman correlation method. The variables in this study passed this need because they have units in percentage or within the range 0 to 1. Assumption 2 demands both variables to have a linear relationship; if the data fails this need, then we transform the dependent variable with the logarithmic model and reapply the Pearson method if the data defaults again then we use Spearman method. The lm function in (R) package tested this need, the coefficient p-value and adjusted R-Squared value from its result measures the data linearity. If the coefficient p-value, is less than the standard significance level (0.05), and the adjusted R-squared value is within the range of 15% to 75%; then we assume the data is linear.

In table 7, assumption 3 requires that there should be minimum outliner or no significant outliner in the bivariate data; the outliner means data points that do not follow the pattern of other data points. We set a minimum of three outliners for each data; meaning less than four outliners in a data. This requirement is the same for Assumption 4, but instead of checking for outliners, it looks for influence values. Influence values or points are types of outliers that influences the statistical measure or correlation coefficient of the data. Several studies have suggested that outliers can be an error point or are incorrect values that can affect the regression model. Some researches countered this saying; for example, based on a book from John Wiley Sons [21] there should be enough non-statistical prove about removing outliers in a data. It claimed that the outlier could be an essential value in the data; one that controls the main properties of the model. Therefore, we suggest not more than 3 outliners in the data; then we apply the Spearman correlation method. Assumption 3 and 4 uses the car package in (R) to find outliers.

Assumption 3 uses the function $outlinerTest$ and returns a p-value and cut-off value; the data contains an outlier if the p-value is less than the cut-off value. Whereas assumption 4

No	Assumptions	Description	(R) Packages	(R) Function	Condition
1	Continuous variables	Variables should be continuous	None	None	None
2	Linear relationship	Variables should have linear relationship	Built-in package	$lm()$	if (coefficient p-value) < (0.05) and 0.15 > (adjusted-square value) < 0.75)
3	Significant outliers	Minimum (less than 4) or no significant outliers	car	$outlierTest()$	p-value < (received cut-off value)
4	Influence values	Minimum (less than 4) or no influence values	car	$cooks.distance()$	(variable values) > (0.5)
5	Normality	Data should have normal distribution	Built-in packages	when dataset equals (n) then $shapiro.test()$ uses (n > 3 and n $\leq$ 5000) and $ad.test()$ uses (n > 5000)	(variable values) > (0.05)
6	Homoscedasticity	Data should have homoscedasticity or equal variances	car	$ncvTest()$ for variance test	(variance values) < (0.05)
7	Heteroskedasticity	Data should not heteroskedasticity	car	$bptest()$ for spread location	(spread location value) > (0.05)

Table 7: Summary of Pearson Correlation Assumptions

uses the function of $cooks.distance$ in (R), it returns a data result vector of values for the data variable. (0.5) is used to filter the vector and return a count of all influence values in a given variable.

Assumption 5 from table 7 requires that the data has a normal distribution; if the data fails this assumption, then we implement the Spearman method. It uses $shapiro.test$ and $ad.test$ functions from (R) built-in packages to calculate normality. $shapirot$ function utilises (Shapiro-Wilk) method on data that is greater than (3) and less than (5000), and $ad.test$ function accepts a dataset that is greater than (5000). If the result values from both functions are greater than (0.05), then the data has a normal distribution. In assumption 6 and 7, residuals should have homoscedasticity and not heteroscedasticity. Homoscedasticity also heteroskedasticity is the equal variance of the independent variable for every value of the dependent variable. It is a systematic change in the spread of the residuals over the range of measured values. When the homoscedasticity assumption fails, an issue arises known as heteroscedasticity. Heteroscedasticity is the high invariance in residuals of the data. Assumption 6 and 7 both use the functions $ncvTest$ and $bpTest$ from car package in (R). We assume there is homoscedasticity if the variance or spread location values from the result is less than (0.05). Otherwise, there is heteroscedasticity.

The (R) functions that assumptions 3, 4, 6, 7 uses for estimation accept residual input of linear model that is produced by lm.

5.0.3 Correlation Coefficient Interpretation:

It is often useful to determine how significant or how strong is the relationship between two variables. The correlation coefficient or r is a statistical measure used to measure the strength of such relationships [28]. The correlation coefficient returns several values from -1 to 0 to 1; that explains the degree of their relationship. Table 8 describes and interprets the coefficient result and their association strength. In general, a coefficient value r that returns 0 reveals that there is no correlation between the variables. However, coefficient value r greater than 0 indicates a positive relationship; this implies that an increase in the independent variable matches an increase in the dependent variable. Then r less than 0 shows a negative relationship meaning that an increase in the independent variable shows a decrease in the values of the dependent variable [60]. The terms (Weak), (Moderate), (Strong) and (Very Strong) explains the degree of the positive and negative relationship between the variables; it explains in simple words to what extent the variables are correlated. In this thesis, we used correlation analysis guidelines in table 8 because it is a standard rule for evaluating the strength between our variables.

[60] https://www.sheffield.ac.uk/polopoly_fs/1.536458!/file/MASH_Correlation_R.pdf

Correlation Coefficient (r)	Association or Relationship
r = 0	No association
r < 0.3	Weak Positive
r < -0.3	Weak Negative
0.3 to 0.5	Moderate Positive
-0.5 to -0.3	Moderate Negative
0.5 to 0.9	Strong Positive
-0.9 to -0.5	Strong Negative
-1 to -0.9	Very Strong Negative
0.9 to 1	Very Strong Positive

Table 8: Strength of correlation coefficients [28]

5.0.4 Causation Method:

In statistics, a common idea of the relationship between two events states that correlation does not mean causation. So a different event might be the cause of correlation association between two variables; although the two events are related; but one does not cause the other. So in this study, we estimated both cases meaning correlation and causation or causal-effect relationship. In our scenario, we assumed that segregation has a causal effect or causes changes in either the labour market or credit risk or well-being and vice versa. Causation is a statistical inference technique that shows the causal relationship between two variables; it explains if a variable causes another. A book from Spirtes et al. (2000) [26] explained causation as an event that causes another event to happen.

In this study, we used dependent or independent variables as either (segregation, the labour-market or credit risk or well-being) to measure causation. And a built-in function in (R) $grangertest$ to decide granger cause and effect relationship between our variables. $grangertest$ function uses a stationary process for measurement, so we avoid using non-stationary data.

The next chapter has a comprehensive display of the results from this method. It includes visual and textual information of the causation and correlation results.

6 Results

For this section, we answered based on statistical findings research questions of this thesis. And we displayed results for correlation and causation relations among segregation with labour-market indicators, credit risk indicators and well-being deprivation index. These results are also separated based on data segregation.

6.1 Correlation Results:

If the outcome of the correlation result between segregation and each indicator is significant, we assume that a correlation exists and use the correlation coefficient to describe their relationship.

6.1.1 Segregation and the Labor Market:

Q1: Is there a relationship (Correlation) between gender and age segregation among boards of directors in Estonia and the labour market?

Results for age-segregated data: There is 70.66% association between segregation and labour market indicators for counties age-segregated data in Estonia. The indicator unemployment by duration (6 months to 11), (total) by persons and the employment rate for each county had the highest correlation level. A strong positive relationship is the most prominent association between unemployment by duration (6 months to 11) indicator and segregation. Also (Atkinson, Dissimilarity, Interaction, Isolation) are its most occurring segregation indexes; the two most correlated counties for this indicator are (Harju) and (Tartu). A strong positive relationship is the most pre-eminent association between unemployment by duration (total) and segregation; segregation indexes (Atkinson, Dissimilarity, Entropy) are prominent in this relationship; the most correlated county for this association is (Tartu).

There is also a high number of strong negative correlation between the employment rate and segregation. (Atkinson, Dissimilarity) are the highest segregation indexes in this association. (Harju) and (Tartu) county has the most association for counties with this indicator. The indicator (inactive persons in the labour force) has a percentage of 9.6% correlation and the most correlated association for this indicator is a (strong positive) relationship with segregation, and it has (Harju) as the highest correlated county for its association with segregation. The indicator unemployment by duration (24 months or more) has a percentage of 9.91% correlation and the most correlated association for this indicator is a (strong positive) relationship with segregation, and it has (Harju) as the highest correlated county; (Gini) as its most occurring segregation index. For all age-segregated data (16-36) is the most correlation group.

There is also 60.61% association between segregation and labour market indicators for regions age-segregated data in Estonia. It uses the measurement unemployment by regions. Figure 6 contains a combined information about this result.

Summary	Labor market by age	unemployedByRegion
Segregation Index with highest Correlation	Dissimilarity	Dissimilarity
Age group with highest Correlation	16-36	16-36
Regions with highest Correlation	Southern Estonia	Southern Estonia
Percentage of Correlation	73.15	73.15
Association type with highest Correlation	Str Positive	Str Positive

Figure 6: Age segregated data and labour market indicators for regions

Results for female and age-segregated data For simplicity, we used female and age-segregated data as data segregated by gender (female) and age. There is 54.76% association between segregation and labour market indicators for counties female and age-segregated data in Estonia. The indicator unemployment by duration (24 months or more), (6 to 11 months), inactive persons in the labour force and the employment rate for each county had the highest correlation level.

The indicator unemployment by duration (24 months or more) has a percentage of 8.04% correlation and its most correlated association for this indicator is a (strong positive) relationship with segregation. It has (Harju) as the highest correlated county; segregation indexes (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are the most occurring segregation indexes for this association. The indicator (inactive persons in the labour force) has a percentage of 8.04% correlation and the most correlated association for this indicator is a (strong positive) relationship with segregation, it has (Harju) as the highest correlated county for its association with segregation. (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are the highest segregation indexes in this association.

There is also a high number of strong negative correlation between the employment rate and segregation. (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are the highest segregation indexes in this association. (Harju) has more association compared to other counties in this association.

The indicator unemployment by duration (6 months to 11) has a percentage of 8.04% correlation. A strong positive relationship is the most prominent association between this indicator and segregation. Also (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are its most occurring segregation indexes; the most correlated county for this indicator is (Harju). The indicator unemployment by duration (total) has a percentage of 7.89% correlation. A (strong positive) relationship is the most pre-eminent association between this indicator and segregation; segregation indexes (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are prominent in this relationship; the most correlated county for this association is (Harju).

Female and age-segregated data for regions in Estonia has 56.94% association between segregation and labour market. It uses unemployment by regions indicator for measurement with (Northern Estonia) as the highest correlated region its association with segregation. The

most visible association types are (strong positive, very strong positive, weak positive); (Dissimilarity, Entropy, Gini, Interaction, Isolation) are its most occurring segregation indexes. For all female and age-segregated data (16-36) is the most correlation group.

Results for female-segregated data: The association between segregation and labour market indicators for the female-segregated data has 89.06% of correlation. With this segregated group, the highest correlated indicators are the rate of inactive persons and unemployment duration by county; they have (strong) positive and (very strong) positive relationship with segregation index (Atkinson, Dissimilarity, entropy, Gini, Isolation).

There is 30.95% association between segregation and labour market indicators for counties female-segregated data in Estonia. The indicator unemployment by duration (less than 6 months) has the lowest correlation level compared to other indicators in this association.

The indicator unemployment by duration (24 months or more) has a percentage of 4.46% correlation and the most correlated association for this indicator is a (strong positive, very strong positive) relationship with segregation. It has (Harju, Ida-Viru, Laane-Viru, Parnu, Tartu) as the highest correlated counties; segregation indexes (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are the most occurring segregation indexes for this association. The indicator (inactive persons in the labour force) has a percentage of 4.46% correlation, and the most correlated association types for this indicator are (Moderate Positive, Very Strong Negative). It has (Harju, Ida-Viru, Laane-Viru, Parnu, Tartu) as the highest correlated counties for its association with segregation. (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are the highest segregation indexes in this association.

There is also a high number of strong negative correlation between the employment rate and segregation. (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are the highest segregation indexes in this association. (Harju, Ida-Viru, Laane-Viru, Parnu, Tartu) has more association compared to other counties in this association.

The indicator unemployment by duration (6 months to 11) has a percentage of 4.46% correlation. A (very strong positive) relationship is the most prominent association between this indicator and segregation. Also (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are its most occurring segregation indexes; the most correlated counties for this indicator are (Harju, Ida-Viru, Laane-Viru, Parnu, Tartu). The indicator unemployment by duration (total) has a percentage of 4.46% correlation. (strong positive, very strong) relationships are the most pre-eminent associations between this indicator and segregation. The segregation indexes (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are prominent in this relationship; the most correlated counties for this association is (Harju, Ida-Viru, Laane-Viru, Parnu, Tartu).

Female-segregated data for regions in Estonia has 25% association between segregation and labour market. It uses unemployment by regions indicator for measurement with (Northeastern Estonia, Northern Estonia, Southern Estonia) as its highest correlated regions. The most visible

association type is (very strong positive); (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are its most occurring segregation indexes. For all female-segregated data (16-36) is the most correlation group.

Results for male-age segregated data: For simplicity, we used male and age-segregated data as data segregated by gender (male) and age. The association between segregation and labour market indicators for county male-segregated data is 49.89%; The employment rate for each county had the highest correlation level. (Atkinson, Dissimilarity, Entropy, Gini, Interaction, Isolation) are the highest segregation indexes in this association. Figure 7 and 8 contains a combined information about this result. Whereas, in figure 9 has results of association between male-age segregated data and labour market indicators for regions.

Summary	Labor market by males & age	employmentRate	inactivePercent	unEmployedDuration (24 months or more)
Segregation Index with highest Correlation	Atkinson, Dissimilarity, Interaction, Isolation	Atkinson, Dissimilarity	Atkinson, Dissimilarity	Gini
Age group with highest Correlation	16-36	16-36	16-36	16-36
County with highest Correlation	Harju	Tartu	Harju	Harju
Percentage of Correlation	49.89	7.2	6.76	7.03
Association type with highest Correlation	Str Positive	Str Negative	Str Positive	Str Positive

Figure 7: Male-age segregated data and labour market indicators for county (Part 1)

unEmployedDuration (Less than 6 months)	unEmployedDuration (6 to 11 months)	unEmployedDuration (12 months or more)	unEmployedDuration (Total)
Entropy, Interaction, Isolation	Atkinson, Dissimilarity, Interaction, Isolation	Gini	Atkinson, Dissimilarity
16-36	16-36	16-36	16-36
Tartu	Harju, Tartu	Harju	Tartu
6.93	7.52	6.93	7.52
Weak Negative	Str Positive	Str Positive	Str Positive

Figure 8: Male-age segregated data and labour market indicators for county (Part 2)

Summary	Labor market by males & age	unemployedByRegion
Segregation Index with highest Correlation	Atkinson, Dissimilarity	Atkinson, Dissimilarity
Age group with highest Correlation	16-36	16-36
Regions with highest Correlation	Northern Estonia, Southern Estonia	Northern Estonia, Southern Estonia
Percentage of Correlation	60.61	60.61
Association type with highest Correlation	Str Positive	Str Positive

Figure 9: Male-age segregated data and labour market indicators for region

Results for male-segregated data: The association between segregation and labour market indicators for county male-segregated data is 28.63%; for this segregated group, the highest correlation level is (strong positive) relationship between the indicator and segregation. The indicator with the highest correlation is unemployment by duration (total). The most occurred

segregation indexes in this association are (Atkinson, Dissimilarity, Entropy, Interaction, Isolation). (Dissimilarity) is the highest segregation index in this association. The age group with the most correlation is (16-36) and the counties with most correlation are (Harju, Jorgeva, Laane, Laane-Viru, Tartu, Viljandi).

The association between segregation and labour market indicators for regions male-segregated data is 19.7%. It uses only unemployment by regions as measurement. The unempolyment by regions indicator has a strong positive association with segregation for the regions (Northeastern Estonia, Northern Estonia, Southern Estonia, Western Estonia).

6.1.2 Segregation and Credit risk:

Q2: Is there a relationship (Correlation) between gender and age segregation among boards of directors in Estonia and credit risk management?

Results for age-segregated data: The association between segregation and credit risk or (CFP) indicators for the age gender is 68.28%. With this segregated group, there is a weak negative relationship between Runway financial indicator and segregation. (Dissimilarity, interaction, isolation) are the segregation indexes with highest correlation. There is also predominantly a strong negative correlation between segregation and the County's Return on Equity indicator. Figure 10 contains more information about this result.

Summary	Credit risk for age	CurrentRatio	QuickRatio	BurnRate	Runway	ReturnOnEquity
Segregation Index with highest Correlation	Gini	Gini	Gini	Gini	Interaction, Isolati...	Dissimilarity, Interaction, Isolation
Age group with highest Correlation	16-36	16-36	16-36	16-36	16-36	16-36
County with highest Correlation	Harju	Harju	Harju	Harju	Harju	Harju, Tartu
Percentage of Correlation	68.28	14.41	13.76	12.58	14.41	13.12
Association type with highest Correlation	Str Negative	Moderate Negati...	Str Negative	Weak Negative	Weak Negative	Str Negative

Figure 10: Age segregated data and credit risk indicators for county

Results for female and age-segregated data: The association between segregation and credit risk or (CFP) indicators for the female and age segregation is 48.12%. For this segregated group, the (Runway, Current ratio, Burn rate and Return on Equity) financial indicator has the highest level of correlation. Also, there is a (Moderate Negative) relationship between Quick ratio and segregation. (Gini) index is the highest segregation index for this association. (Harju) county has more correlation compared to other counties in this association.

Results for female-segregated data: The association between segregation and credit risk or (CFP) indicators for the female segregation is 23.96%. For this segregated group, the (Runway, Current ratio, Burn rate and Return on Equity) financial indicator has the highest level of correlation. Also, there is a (weak negative) relationship between Quick ratio and segregation. (Gini) index is the highest segregation index for this association. (Harju and Laane-Viru) county has more correlation compared to other counties in this association.

Results for male and age-segregated data: The association between segregation and credit risk or (CFP) indicators for the male gender is 47.73% level of correlation. Current ratio

indicator has the highest level of correlation for this segregated data. There is a predominantly (Strong Negative) relationship in this association. (Gini) index has more correlation compared to others in this association. Figure 11 contains more information about this result.

Summary	Credit risk for males and age	CurrentRatio	QuickRatio	BurnRate	Runway	ReturnOnEquity
Percentage of Correlation	47.73	10.23	9.7	8.41	10.15	9.24
Age group with highest Correlation	16-36	16-36	16-36	16-36	16-36	16-36
County with highest Correlation	Harju	Harju	Harju	Harju, Tartu	Harju	Harju, Tartu
Segregation Index with highest Correlation	Gini	Gini	Gini	Gini	Interaction, Isolation	Interaction, Isolation
Association type with highest Correlation	Str Negative	Str Negative	Str Negative	Weak Negative	Weak Negative	Str Negative

Figure 11: Males age-segregated data and credit risk indicators for county

Results for male-segregated data: The association between segregation and credit risk or (CFP) indicators for the male gender is 24.55% level of correlation. Runway indicator has the highest level of correlation for this segregated data. There is a predominantly (Weak Negative) relationship in this association. Segregation indexes (Interaction, Isolation) have more correlation compared to others. Rapla, Tartu county has the most association in relation to other counties.

6.1.3 Segregation and Well-being:

Q3: Is there a relationship (Correlation) between gender and age segregation among boards of directors in Estonia and the country's well-being?

For all types of segregated groups (age, males and females) they correlated by 100%. The correlation between segregation (Gini) and well-being deprivation index for the age-segregated has (very strong) positive, (strong) positive and (weak) negative relationship. For the male-segregated data, there is a tie of (weak) negative relationship and (very strong) positive relationship between segregation (Atkinson, dissimilarity) and well-being deprivation index. For the female-segregated data, there is a predominantly (very strong) positive relationship and strong positive relationship between segregation (Atkinson, dissimilarity, isolation, interaction, Gini, entropy) and well-being deprivation index. In table 9 and 10 there is a summary of correlation results by county and regions.

Segregated group	Result	Labour-market association	Credit-risk association	Well-being association
Females	Sample size	670	2865	448
	Correlation (%)	89.06%	23.96%	100%
Females and age	Sample size	670	2865	448
	Correlation (%)	54.76%	48.12%	100%
Age	Sample size	675	2865	896
	Correlation (%)	70.66%	68.28%	100%
Males	Sample size	670	2865	448
	Correlation (%)	28.63%	24.55%	100%
Males and age	Sample size	670	2865	448
	Correlation (%)	49.89%	47.73%	100%

Table 9: Summary of Correlation results by County

Segregated group	Result	Labour-market association	Well-being association
Females	Sample size	114	114
	Correlation (%)	25%	100%
Females and age	Sample size	114	448
	Correlation (%)	56.94%	100%
Age	Sample size	114	114
	Correlation (%)	73.15%	100%
Males	Sample size	114	114
	Correlation (%)	19.17%	100%
Males and age	Sample size	114	114
	Correlation (%)	60.61%	100%

Table 10: Summary of Correlation results by Regions

6.2 Causation Results:

Q1: Is there a relationship (Causation) between gender and age segregation among boards of directors in Estonia and the labour market?

6.2.1 Segregation and Labor Market:

The association result between segregation and labour market indicators for the age-segregated data for all age groups (16-36, 37-45, 46-54, 55-65, 66-99) claim that segregation (Interaction) granger cause increase in county's employment rate and increase in inactive persons at the workforce. Whereas, 68.75% of results for all age groups claim that the county's employment rate and increase in inactive persons in the labour force can granger cause an increase in segregation. For the male-segregated data alone result showed that unemployment by regions granger cause segregation (Interaction) at the rate of 33.33%.

6.2.2 Segregation and Credit risk:

Q2: Is there a relationship (Causation) between gender and age segregation among boards of directors in Estonia and credit risk? The association between segregation and credit risk or (CFP) indicators showed that for the male-segregated data (Current ratio) granger cause segregation at the rate of 33.33%.

6.2.3 Segregation and Well-being:

Q3: Is there a relationship (Causation) between gender and age segregation among boards of directors in Estonia and the country's well-being? The association between segregation (interaction) and well-being deprivation index showed that only the female-segregated data showed granger causes the county's (that is Harju, Ida-Viru) well-being deprivation at the rate of 100%. It also showed that segregation (interaction) granger causes regions' well-being deprivation at the rate of 50%.

7 Conclusion

In this thesis, we analyzed the relationship between segregation among boards of directors in Estonia and the changes in the labour-market; credit risk and well-being. The societal impact of segregation on companies in Estonia and well-being of residents in Estonia. We covered topics related to the best composition of a company board. The influence of gender distribution in the company's success.

The presence of a good company and how it influences people's well-being. In this story, we understood the relationship between the distribution of gender and age in boards of companies and credit risk management. We found out if the differences lead to an advantage, and created a well-being deprivation index to study the effects of the presence of a successful company in the society. And solve the three main research questions with statistical methods used in estimating the relationship between segregation and factors of consideration.

In the findings, we realized that when there is a high level of segregation among boards of directors people within the age group (16-36) had more relationship with segregation compared to other age groups; which means that they tend to be affected more. Females had the highest level of relation with segregation and had more well-being deprivation when segregation among boards of directors increases. Also, the labour market indicator had the highest relationship with segregation.

We noticed that correlation and causal relationship existed between segregation and all measurement factors. Although the relationship was of different types; this gives a glimpse of hope that segregation among boards of directors might indeed play a huge role in influencing the labour market, credit-risk and well-being.

References

[1] S. Alkire. The capability approach as a development paradigm. *Material for*, 2003.

[2] S. Alkire and J. Foster. Counting and multidimensional poverty measurement. *Journal of public economics*, 95(7-8):476–487, 2011.

[3] S. Alkire and J. Foster. Understandings and misunderstandings of multidimensional poverty measurement. *The Journal of Economic Inequality*, 9(2):289–314, 2011.

[4] S. Anspal, T. Rõõm, S. Anspal, L. Kraut, and T. Rõõm. Gender pay gap in estonia: empirical analysis. *Report for the Estonian ministry of social affairs. Tallinn: Ministry of Social Affairs*, 2011.

[5] D. E. Arfken, S. L. Bellar, and M. M. Helms. The ultimate glass ceiling revisited: The presence of women on corporate boards. *Journal of Business ethics*, 50(2):177–186, 2004.

[6] U. Backes-Gellner and S. Veen. The impact of aging and age diversity on company performance. *Available at SSRN 1346895*, 2009.

[7] A. Baroni. Segregation aware data mining. 2017.

[8] M. Bertrand, S. E. Black, S. Jensen, and A. Lleras-Muney. Breaking the glass ceiling? the effect of board quotas on female labour market outcomes in norway. *The Review of Economic Studies*, 86(1):191–239, 2019.

[9] Z. Burgess and P. Tharenou. Women board directors: Characteristics of the few. *Journal of business ethics*, 37(1):39–49, 2002.

[10] N. Burrow, A. Fedorets, and A. Gibert. The effects of a gender quota on the board of german largest corporations. *Berlin: German Institute for Economic Research*, 2018.

[11] S. Champely, C. Ekstrom, P. Dalgaard, J. Gill, S. Weibelzahl, A. Anandkumar, C. Ford, R. Volcic, H. De Rosario, and M. H. De Rosario. Package 'pwr'. *R package version*, 1(2), 2018.

[12] H. Chen. An analysis of bhutan's gross national happiness. *Seven Pillars Institute Moral Cents*, 4(2):66–74, 2015.

[13] D. A. Cotter, J. DeFiore, J. M. Hermsen, B. M. Kowalewski, and R. Vanneman. All women benefit: The macro-level effect of occupational integration on gender earnings equality. *American sociological review*, pages 714–734, 1997.

[14] I. Ferrero-Ferrero, M. Á. Fernández-Izquierdo, and M. J. Muñoz-Torres. Age diversity: An empirical study in the board of directors. *Cybernetics and Systems*, 46(3-4):249–270, 2015.

[15] J. Hawkins. 7. the four approaches to measuring wellbeing. *Measuring and promoting wellbeing: how important is*, page 191, 2014.

[16] T. Hothorn, A. Zeileis, R. W. Farebrother, C. Cummins, G. Millo, D. Mitchell, and M. A. Zeileis. Package 'lmtest'. *Testing linear regression models. https://cran. r-project. org/web/packages/lmtest/lmtest. pdf. Accessed*, 6, 2015.

[17] J. H. Kim and I. Choi. Choosing the level of significance: A decision-theoretic approach. *Abacus*, 2019.

[18] K. Klein. Does gender diversity on boards really boost company performance? *Social Impact, Wharton–University of Pennsylvania, available at: http://knowledge. wharton. upenn. edu/article/will-gender-diversity-boards-really-boost-company-performance*, 2017.

[19] L.-E. Lee, R. Marshall, D. Rallis, and M. Moscardi. Women on boards. *Global Trends*, 2015.

[20] F. v. Meyerinck, A. Niessen-Ruenzi, M. Schmid, and S. Davidoff Solomon. As california goes, so goes the nation? the impact of board gender quotas on firm performance and the director labor market. *SSRN Electronic Journal*, 2018.

[21] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to linear regression analysis*, volume 821. John Wiley & Sons, 2012.

[22] J. F. Mudge, L. F. Baker, C. B. Edge, and J. E. Houlahan. Setting an optimal α that minimizes errors in null hypothesis significance tests. *PloS one*, 7(2):e32734, 2012.

[23] B. C. on Banking Supervision and B. for International Settlements. *Principles for the management of credit risk*. Bank for International Settlements, 2000.

[24] C. Post and K. Byron. Women on boards and firm financial performance: A meta-analysis. *Academy of management Journal*, 58(5):1546–1571, 2015.

[25] S. Ringen. Well-being, measurement, and preferences. *Acta Sociologica*, 38(1):3–15, 1995.

[26] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman. *Causation, prediction, and search*. MIT press, 2000.

[27] J. Stypińska and P. Nikander. Ageism and age discrimination in the labour market: A macrostructural perspective. In *Contemporary perspectives on ageism*, pages 91–108. Springer, Cham, 2018.

[28] R. Taylor. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1):35–39, 1990.

[29] K. Ura, S. Alkire, and T. Zangmo. Bhutan: Gross national happiness and the gnh index. 2012.

[30] K. Ura, S. Alkire, T. Zangmo, and K. Wangdi. *An extensive analysis of GNH index*. 2012.

Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Oluwagbemi Kadri**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   Gender-based segregation in company boards and well-being,

   supervised by Rajesh Sharma.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Oluwagbemi Kadri
11/11/2020