Aalto University
School of Business

FORMING AND EVALUATING SOCIALLY RESPONSIBLE INVESTMENT PORTFOLIOS UNDER UNKNOWN INVESTOR RISK PREFERENCES

Measuring the risk-adjusted return of socially responsible investment portfolios built with stochastic dominance criteria

Master's Thesis
Joni Pirskanen
Aalto University School of Business
Information and Service Management
Spring 2020

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Joni Pirskanen

Title of thesis Forming and Evaluating Socially Responsible Investment Portfolios Under Unknown Investor Risk Preferences

Degree Master of Science in Economics and Business Administration

Degree programme Information and Service Management

Thesis advisors Juuso Liesiö, Peng Xu

Year of approval 2020 **Number of pages** 60 **Language** English

Abstract

Socially responsible investing (SRI) has gained growing interest both in practice and academia during recent years. Many investors who implement responsibility themes in their investment strategy gain utility from following their values but also from financial returns. Hence, their goal is not to completely sacrifice the financial performance of the portfolio in order to follow responsibility themes. The risk-return performance of SRI has been studied in many contexts, but no consensus has been reached whether it offers suboptimal returns or possibly some long-lasting advantages.

This thesis studies how portfolios with responsibility constraints have performed compared to an unconstrained benchmark portfolio in Helsinki stock market during 2009-2018. Responsibility constraints are applied by using company-specific ESG (environmental, social and governmental) scores. ESG scores have become a widely accepted method for applying responsibility into portfolio selection.

In practice, investor risk preferences are often unknown when a portfolio is selected. This thesis applies the theoretically appealing method of stochastic dominance to optimize portfolio weights. Stochastic dominance requires only simple assumptions on investor risk preferences and assets' underlying return distributions. Recent development in methods has enabled using stochastic dominance to efficiently diversify a portfolio across a wide set of possible assets. Methodology by Kuosmanen (2004) is applied to select portfolios that are efficient in the sense of second-degree stochastic dominance over OMXH25 index benchmark.

The results imply that an investor is able to exclude up to 40% of companies based on ESG scores and still form a portfolio that dominates the OMXH25 index. If the ESG limit is increased more, the optimization method does not find portfolios that dominate the index by stochastic dominance criteria, implying that increasing the ESG limit higher prevents forming efficient portfolios in the sense of SSD.

Keywords Socially Responsible Investing, Portfolio Selection, Stochastic Dominance

**Aalto-yliopisto
Kauppakorkeakoulu**

**Aalto-yliopisto, P.O. BOX 11000, 00076 AALTO
www.aalto.fi**

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Joni Pirskanen

Työn nimi Vastuullisten sijoitusportfolioiden muodostaminen ja suorituskyvyn arviointi tuntemattomien riskipreferenssien alaisuudessa

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Tieto- ja palvelujohtaminen

Työn ohjaajat Juuso Liesiö, Peng Xu

Hyväksymisvuosi 2020 **Sivumäärä** 60 **Kieli** Englanti

Tiivistelmä

Vastuullinen sijoittaminen on saanut osakseen kasvavaa huomiota sekä käytännön sovelluksissa että tieteellisessä tutkimuksessa viimeisten 15 vuoden aikana. Monet sijoittajat, jotka lisäävät vastuullisuuskriteerejä sijoitusstrategiaansa kokevat saavansa hyötyä sekä omien arvojen mukaisesta strategiasta että taloudellisesta tuotosta. Vastuullisen sijoittamisen tarkoituksena ei siis yleensä ole uhrata tuottoa vastuullisuusteemojen vuoksi. Vastuullisen sijoittamisen riskikorjattua tuottoa on tutkittu eri yhteyksissä, mutta yhtenäistä näkemystä vastuullisen sijoittamisen tehokkuudesta ei ole muodostunut.

Tämä tutkielma tutkii kuinka vastuullisuuskriteerein muodostetut portfoliot suoriutuvat verrattaessa rajoittamattomaan verrokkiportfolioon Helsingin pörssissä 2009-2018. Vastuullisuuskriteerinä portfolion muodostuksessa käytetään ESG-pisteytyksiä, jotka mittaavat yhtiöiden vastuullisuutta ympäristön sekä sosiaalisten ja hallinnollisten tekijöiden suhteen. ESG-pisteytykset ovat yleisesti hyväksytty ja helppo tapa sisällyttää vastuullisuusteema sijoitusstrategiaan.

Kun sijoitusportfoliota valitaan käytännössä, sijoittajien riskipreferenssejä ei usein tunneta etukäteen. Tässä tutkimuksessa portfolioiden muodostuksessa hyödynnetään stokastista dominanssia, joka tarjoaa teoreettisia etuja tavanomaisempiin menetelmiin verrattaessa. Stokastista dominanssia käytettäessä tarvitaan vain hyvin yleisluontoisia oletuksia sijoittajien preferensseistä sekä sijoituskohteiden tuottojakaumista. Stokastista dominanssia on käytetty laajasti kahden riskisen vaihtoehdon vertailuun, mutta sen hyödyntäminen tehokkaasti hajautetun portfolion muodostamiseen on ollut mahdollista vasta jonkin aikaa. Tutkimuksessa käytetty metodi on Kuosmasen vuonna 2004 julkaisema tehokkuustesti, joka optimoi portfolion painotukset toisen asteen stokastiseen dominanssiin perustuen.

Tämän tutkimuksen tulosten perusteella sijoittaja voi stokastista dominanssia hyödyntäen jättää jopa 40 % ESG-mittarilla huonoiten pärjäävää osaketta pois portfolion valinnasta ja muodostaa silti portfolion, joka dominoi OMXH25-indeksiä. ESG-kriteerin lisääminen portfolion valintaan ei toisaalta tarjonnut myöskään ylimääräistä tuottoa. Jos ESG-rajoitetta kasvatetaan 40 prosentista ylöspäin, metodi ei kykene enää löytämään portfoliota, joka dominoisi indeksiä toisen asteen stokastiseen dominanssiin perustuen. Vastuullisuuskriteerin tiukentuessa riittävästi on siis teoriassa uhrattava portfolion tehokkuutta vastuullisuuskriteerin täyttämisen puolesta.

Avainsanat Vastuullinen sijoittaminen, Portfolion valinta, Stokastinen dominanssi

Acknowledgements

I would like to thank my thesis advisors Juuso Liesiö and Peng Xu for the guidance in the beginning of the project and all the support through the whole research process.

Table of Contents

1 Introduction and motivation 1

1.1 Research objectives 2

1.2 Structure of the thesis 3

2 Literature review 4

2.1 SRI investment strategies 4

2.2 SRI investments and return 6

2.3 ESG scores 8

2.4 Introduction to stochastic dominance 9

2.5 Formulating first- and second-degree dominance 11

2.6 Stochastic dominance and efficient diversification 14

2.7 Portfolio selection and performance measurement in SRI studies 15

3 Data and methodology 18

3.1 Research process 18

3.2 Data exploration 19

3.2.1 ESG scores provided by Refinitiv 20

3.3 Data preprocessing 21

3.3.1 Descriptive statistics 22

3.4 Investment strategy 25

3.4.1 Applying ESG scores 26

3.5 Methodology 28

3.5.1 Tools 28

3.5.2 Implementing Kuosmanen model 29

3.5.3 Computational burden 31

4 Results 34

4.1 Finding dominance over OMXH25 34

4.2 Composition of the portfolios 35

4.3 Monthly out-of-sample performance 37

4.4 Cumulative returns 40

4.4.1 Testing for significance 41

5 Discussion 43

6 Conclusion 48

6.1 Research summary 48
6.2 Limitations of the study 49
6.3 Managerial implications 50
6.4 Suggestions for future research 51
References 53
Appendix A: Company-specific return distributions 58
Appendix B: Implementation of LP problem 59
Appendix C: Stocks selected in portfolios 60

List of Tables

Table 1. Return data, daily descriptive statistics. 23
Table 2. Company-specific ESG scores. 24
Table 3. Computational burden of the test statistic $\theta 2n$ 32
Table 4. Number of stocks selected in portfolios. 35
Table 5. Most commonly selected stocks in each portfolio. 36
Table 6. Most weighted stocks. 36
Table 7. Monthy out-of-sample performance. 38
Table 8. Cumulative returns. 41

List of Figures

Figure 1. Illustration of a situation in which the random variable X dominates Y in the sense of FSD. 13
Figure 2. Illustration of a situation in which the random variable X dominates Y in the sense of SSD. 13
Figure 3. Research process. 18
Figure 4. Refinitiv ESG scores. 20
Figure 5. Mean ESG scores distribution. 25
Figure 6. Dividing sample periods. 26
Figure 7. Industry classification. 28
Figure 8. Monthly return distributions of optimized portfolios. 39
Figure 9. Cumulative returns. 41

1 Introduction and motivation

"While each of our individual companies serve its own corporate purpose, we share a fundamental commitment to all of our stakeholders."(US Business Round Table, 2019)

While traditionally the only responsibility of a company has been a financial one, maximizing the shareholder equity, the statement by the US Business Round Table representing the leaders of 181 major US companies reflects the changing attitudes towards the responsibility of companies and business in general. Stemming from global issues such as the climate change, the change of attitudes among both public and policy makers has given a raise to implement the responsibility theme in investing, leading to an ever-growing interest towards socially responsible investing (SRI) in both practice and academia (Renneboog et al., 2008b).

While there is a growing interest towards SRI in academia and a considerable number of researchers have studied the performance of investment portfolios selected with responsibility criteria, there is no consensus on whether SRI could lead to inferior or superior risk-adjusted returns. Majority of studies on SRI are measuring the performance of mutual funds, but only a few have tested the performance of portfolios explicitly built with SRI criteria compared to a similar, unconstrained portfolio. (Auer, 2016)

Responsibility themes can be applied to portfolio selection by using several different strategies. In this thesis, company-specific ESG (Environmental, Social and Corporate Governance) scores are utilized in a negative screening strategy that excludes companies that are performing poorly from a responsibility perspective. ESG scores aggregate the three dimensions of responsibility into one comparable measure, and they are provided by third-party rating agencies. ESG scores are a widely accepted method, and they enable applying a simple responsibility logic to portfolio selection (Auer & Schuhmacher, 2016).

As in mainstream financial literature in general, studies on SRI rely on standard portfolio selection methods and performance measures. However, a question remains whether these methods are able to account for all the risks related to responsible investments. This thesis applies the theoretically appealing method of stochastic dominance (SD) to build several ESG-constrained SRI portfolios and an unconstrained benchmark portfolio. In practical investment portfolio selection applications, investor risk preferences are often unknown prior to the selection. Stochastic dominance allows for choosing between the risky alternatives with only simple assumptions on investors' risk preferences or the underlying return distributions of the choice alternatives. Stochastic dominance offers thus interesting theoretical advantages compared to other portfolio selection methods. (Hadar & Russell, 1971; Hanoch & Levy, 1969; Post, 2003)

Stochastic dominance is a well-known method for selecting between two risky choice alternatives. However, until rather recent development there were no methods for building efficiently diversified portfolios with stochastic dominance criteria. Several methods have been developed during the last 15 years by e.g. Kuosmanen (2004) and Kopa and Post (2014). These methods formulate stochastic dominance into linear programming tests that allow for testing the efficiency of portfolios in the sense of stochastic dominance. The methodology by Kuosmanen will be applied in this thesis to build the investment portfolios.

After building the portfolios by obtaining optimal portfolio weights from the linear programming model, the out-of-sample risk-adjusted returns of the portfolios are measured. Performance of the portfolios is evaluated with several simple and risk-adjusted measures. It will be also examined how well the optimization model is able to find dominance over an index benchmark when the ESG constraint is constantly increased.

1.1 Research objectives

This thesis has a dual objective as it combines SRI and stochastic dominance methods to address the problem of unknown risk preferences. The research objectives is thus divided into two intercepting research questions. The first question examines the risk-return performance of socially responsible portfolios:

1. *How do portfolios with responsibility criteria perform compared to an unconstrained benchmark portfolio under unknown investor risk preferences?*

The second research question concerns the unknown risk preferences. As shown in the literature review, second-degree stochastic dominance (SSD) can be used for building portfolios with very limited information about the preferences. Hence the second question is formulated as follows:

2. *Does the introduction of increasingly strict responsibility criteria make it impossible to construct a portfolio that dominates the OMXH25 index in the sense of SSD?*

1.2 Structure of the thesis

This thesis consists of six chapters. In *Literature review*, a theoretical frame is built by introducing earlier literature. The chapter is divided into two subsections, socially responsible investing and stochastic dominance. In *Data and methodology,* the data used for this thesis is presented. Regarding the methodology, the application of Kuosmanen linear programming model, investment strategy specifications and computational burden of the solution are discussed. *Results* present the empirical findings, while in chapter *Discussion* the results are examined in a critical manner. Implications to theory and practice are presented, and the limitations of the study are discussed. The final chapter concludes the findings of the study and presents recommendations for future research.

2 Literature review

In this chapter, the theory relevant to the topic of the thesis is presented. The chapter has two main themes: socially responsible investing (SRI) and stochastic dominance. Chapter 2.1 gives an overall introduction to SRI, different SRI strategies and some numerical data on the popularity of SRI around the world. Chapter 2.2. discusses different views on the relation between SRI and risk-return efficiency. Chapter 2.3. introduces ESG scores: how they have been used in earlier literature and what kind of problems ESG scores might cause.

Chapters 2.4-2.6 focus on stochastic dominance. First, a general introduction to the theory of stochastic dominance is given, followed by the mathematical notation required for defining first-degree and second-degree dominance. Second, efficient diversification with stochastic dominance methods is introduced. Finally, in 2.7. portfolio selection methods and performance measures in existing SRI studies are discussed to justify the use of stochastic dominance in the empirical part of the thesis.

2.1 SRI investment strategies

While no universal definition of socially responsible investing exists, it can be described as "integration of certain non-financial concerns, such as ethical, social or environmental, into the investment process" (Sandberg et al., 2008, p. 519; von Wallis & Klein, 2015).

Following a rapid expansion during the recent years, the scale of socially responsible investing has reached massive scales. According to Revelli (2017) the expansion of SRI investment strategies into mainstream investment markets occurred after the financial crisis of 2008. In 2018, global responsible investment in 5 major markets (Europe, USA, Japan, Canada and Australia) reached $30.7 trillion following a 34 percent increase in just two years. The proportion of investment assets under some responsible investment strategy account for 46% of all investing assets in Europe and 39% in the US. (GSIA, 2019) The need and demand for SRI has been understood also on a governmental level: the regulatory environment in many western countries has been improved during the last 15 years to stimulate the amount of SRI (Renneboog et al., 2008b).

The Global Sustainable Investment Alliance has introduced sustainable investment definitions that have become the global standard of classification for sustainable investment strategies (GSIA, 2019):

1. **Negative/exclusionary screening:** excluding certain sectors, companies or practices from a portfolio. Typically, companies involved in industries such as alcohol, tobacco, weaponry or gambling are screened (Renneboog et al. 2008a). A negative screen can also be applied using company-specific ESG (Environmental, Social and Corporate Governance) scores: the companies with the worst x% of ESG scores are excluded from the analysis.
2. **Positive/Best-in-class screening:** investment in sectors, companies or projects based on a positive ESG performance compared to industry peers. Allows for choosing the very best companies among different industries.
3. **Norms-based screening:** screening of investments against minimum standards of business practice based on international norms, such as those issued by the OECD, ILO, UN and UNICEF.
4. **ESG Integration**: systematically and explicitly integrating ESG factors into financial analysis.
5. **Sustainability themed investing**: investment in specific sustainability themes or assets specifically related to sustainability (for example clean energy).
6. **Impact/community investing:** targeted investments aimed at solving social or environmental problems, and directing capital to underserved individuals or communities, as well as financing that is provided to businesses with a clear social or environmental purpose.
7. **Corporate engagement and shareholder action**: the use of shareholder power to influence corporate behavior, including through direct corporate engagement (i.e., communicating with senior management and/or boards of companies), filing or co-filing shareholder proposals, and proxy voting that is guided by comprehensive ESG guidelines.

Out of these strategies, negative screening based on excluding controversial industries is the most commonly implemented in portfolio selection, especially in mutual funds (Renneboog et al., 2008b). There is some evidence that negative screening could cause an opportunity cost for investors, since some of the controversial industries have outperformed the market (Trinks & Scholtens, 2017). In this study, a negative screening strategy based on company-specific ESG scores will be implemented.

2.2 SRI investments and return

Investors who follow an investment strategy with SRI elements can be assumed to gain utility both from wealth-maximization and social responsibility: investors try to gain financial returns with investments that are in line with their personal values and beliefs. Rather than donating money to charity, SRI investors seek positive risk-adjusted returns from their investments. Hence, the efficiency of responsible investment strategies has become an important question for both academia and investment professionals. (Renneboog et al., 2008a)

The performance of SRI funds has been studied already widely, yet the question whether investors must pay a price for the social utility remains interesting. A common hypothesis for the studies done in the field is that investors might be willing to pay for aversion of unethical corporate behavior included in their portfolio (Renneboog et al., 2008a; Revelli & Viviani, 2015). It can be simultaneously studied whether SRI funds could reveal some hidden value-relevant information that isn't fully reflected in asset prices (Renneboog et al., 2008a). In a second-hand study of over 2000 studies conducted from various geographical areas and time frames, Friede et al. (2015) find that over 90% of the studies show a nonnegative relation between responsibility factors and corporate financial performance, hinting implementing responsibility factors could provide interesting investment opportunities.

The existing studies on SRI fund performance have provided contradictory results. There are in general three distinct views on the economic viability of SRI investing: the first view is that implementing SRI into investment strategies provides suboptimal returns. The second view is that SRI could improve the long-term performance of investments, while the supporters of the third view argue that no significant costs or positive returns can be associated with SRI strategies. (Auer, 2016)

Supporting the negative view on SRI risk-adjusted performance, Renneboog et al. (2008a, 2008b) find in two widely cited studies that investment funds with SRI strategies slightly underperform benchmark funds with no SRI limitations in several countries. Renneboog et al. discuss two possible reasons behind their findings. First, they argue that SRI imposes a constraint on the investment opportunities, leading to weaker diversification possibilities and reduced profitability of SRI strategies. From a portfolio theory point of view, any constrains set to diversification, selection and exclusion of financial assets reduce the investment opportunities and should indeed lead to inefficiency in portfolio selection (Markowitz, 1952). However, Renneboog et al. note that it remains unclear whether the asset pricing models they are using (CAPM and Fama-French factor model) are able to capture the effect of SRI factors to the risk level of the assets. Second, Renneboog et al. (2008a) discuss that companies meeting high ethical standards might be overpriced in the stock markets, resulting from aversion to the stocks of unethical companies. Supporting this view, Hong and Kacperczyk (2009) find evidence that 'sin' stocks, stocks of companies in controversial fields such as tobacco or gambling, provide higher expected returns. Hong and Kacperczyk argue that these 'sin' stocks are held less by large norm-based investment institutions such as pension funds, leaving room for excess returns.

Despite of the theoretical disadvantage of SRI portfolios, there is compelling empirical evidence from several studies supporting the second, positive view that SRI investing does not provide suboptimal returns. Kempf and Osthoff (2007) find SRI strategies to provide significant return premiums by using a long-short strategy that invests in a long position in best SRI assets and a short position in the worst ones. Derwall et al. (2005) find similar, positive results and provide two possible explanations for why SRI strategies could in theory provide abnormal returns. First, they argue that the stock market might undervalue the environmental information. Derwall et al. (2011) call this the error-in-expectations hypothesis: for SRI strategies to provide superior returns, CSR (corporate social responsibility) information should include some effects to returns that investors are not able to expect. Derwall et al. (2011) predict that if this kind of hypothesis holds, it should diminish over time as investors learn to value CSR correctly. Both Kempf and Osthoff (2007) and Derwall et al. (2005) give another possible explanation for return premiums: they argue that the premium from SRI strategies might capture some risk factors that are missing from the asset pricing models they are using.

There is also empirical evidence that SRI funds are more resistant to financial crises than conventional ones: Nakai et al. (2016) found that in Japan, SRI funds performed significantly better during the 2008 financial crisis compared to conventional funds.

The third view is that SRI strategies do not affect risk-adjusted returns either positively or negatively: in a meta-analysis of 85 first-hand studies on SRI Revelli & Viviani (2015) argue that there should be neither a significant cost nor significant positive returns when investing with an SRI strategy.

Regarding the methodology, existing research on the performance of SRI investments can be divided into two major groups: i) comparing mutual funds and ii) comparing specifically built ESG portfolios to a benchmark. (Auer, 2016; Kempf & Osthoff, 2007). A large part of the studies compares the performance of socially responsible mutual funds to conventional mutual funds (see e.g. Renneboog et al. 2008a, 2008b). However, there's a drawback when studying mutual funds: the performance of mutual funds can differ based on the skills of the portfolio manager or management fees, and hence the performance of socially responsible funds cannot be separately attributed to the impact of responsibility factors (Auer, 2016). In addition to that, mutual funds with high ESG scores have been shown to gradually transform to conventional funds, i.e. the high ESG scores are not persistent when fund managers rebalance the portfolios with other priorities. This implies that not all funds with an SRI label follow pure SRI principles (Wimmer, 2013).

The second strand of literature studies the performance of portfolios that have been explicitly built using a positive or negative responsibility screen (e.g. Auer & Schuhmacher, 2016; Kempf & Osthoff, 2007). This method enables to ignore portfolio management related factors. It also allows to control for the ESG constraint, and the level of ESG can be held constant over time. Despite of the theoretical advantage of the second approach, the first method of comparing complete funds has been dominating in SRI research.

2.3 ESG scores

Many investors have only a vague understanding of what is 'socially responsible' – it is easy to understand excluding certain industries such as tobacco, but more accurate analysis of responsibility is not possible for many individual investors. ESG (Environmental, Social and Governance) scores have become an industry standard to evaluate and quantify the social responsibility of companies. (Auer, 2016) ESG scores try to capture information from several non-financial dimensions of a stock – namely environmental, social and corporate governance dimensions. ESG scores are provided by several rating and information provider agencies that collect data from various sources such as company filings, media and third-party data providers. For an individual investor, ESG scores are an easy and efficient way to analyze the responsibility of potential investments. (Auer, 2016; van Duuren et al., 2016)

While ESG scores are the easiest way to incorporate complete information on companies' total responsibility, the scores have several disadvantages. Auer & Schuhmacher

(2016) note that a drawback of ESG strategies is that they require active trading, causing higher costs than passive funds. It is also important to note that currently, there is no accepted and complete standard methodology to evaluate the responsibility of a company. The criteria used by different ESG rating agencies are based on some global standards, but the agencies are using differing weightings to analyze the criteria. In addition to that, the rating agencies might not be fully transparent with the criteria they are using, causing lack of information for investors. (Escrig-Olmedo et al., 2010)

Concerns of ESG factors being a tool for greenwashing have also been raised. Issues in ESG valuations such as governance conflicts, lack of resources used for evaluation and a dual role of ESG rating agencies should be noted. The raise of the popularity of responsible investing creates financial incentives for companies to manipulate the ESG score and for the providers of SRI funds to use ESG criteria for greenwashing the marketing of investment funds. (Revelli, 2017) Despite of these issues, integrating ESG scores in the portfolio selection screening remains a major responsible investment strategy in practical applications as it allows easy application of responsibility themes into investment strategies.

2.4 Introduction to stochastic dominance

Stochastic dominance (SD) is a well-established analytical tool for decision making under uncertainty, and it has been applied in various research areas – especially in finance and economics (Bawa, 1982). The application of the concept of stochastic dominance in decision theory began in 1962 by Quirk and Saposnik, and the methodology and notation needed for finance extensions were developed in 1969 by Hanoch and Levy. In finance literature, stochastic dominance is used to compare investment alternatives based on their observed rates of return (Post, 2003).

According to Post (2008) the advantage of SD is that it lets the data 'speak for itself', rather than being forced to follow strict assumptions. To be more specific, stochastic dominance is an appealing method for investment choice due to several theoretical advantages: it is nonparametric since it does not require explicit specification of an investor's utility function nor the form of statistical distribution of the choice alternatives. Only general assumptions of the distributions are needed. In other words, one does not need specific information on investors' preferences to define stochastic dominance between risky alternatives. Stochastic dominance takes the complete probability distribution of an investment's returns into account (Kuosmanen, 2004; Post, 2003).

Despite of the theoretical attractiveness of stochastic dominance, portfolio selection research has been dominated by the Portfolio Theory and mean-variance analysis (MVA) that evaluates investments in terms of their means and variances (Markowitz, 1952). The mean-variance analysis and methods built on it are useful due to their simplicity, but they have major limitations: The MVA is only consistent with the expected utility theory if and only if highly restrictive conditions are set for the investor preferences and assets' return distributions (Post, 2003). MVA and stochastic dominance vary fundamentally regarding the treatment of risk: MVA uses variance as the quantified measure for risk. Portfolio selection in MVA is based on the assumption that a risk-averse investor always selects the portfolio with a lower variance in case of equal means (Hanoch & Levy, 1969)

In general, measuring risk based on a single measure such as variance provides uncertain results (Hanoch & Levy, 1969). Variance has several limitations as a measure of risk: it is a symmetric measure, treating both positive and negative deviations from the mean similarly. There is also compelling evidence that the distribution of many assets is in reality not normal, and many assets are showing skewness and fat tails. This means that variance measures the risk of the assets erroneously. While several other risk measures such as value at risk (VaR) have been introduced, selecting the correct measure for risk is challenging (De Giorgi & Post, 2008). Stochastic dominance does not imply any quantified measure for risk; risk is treated as a qualitative criterion that relies on the preference orderings of the risky options (Kuosmanen, 2004).

For a long time, a major drawback of stochastic dominance methods was the lack of applications for efficient diversification. Stochastic dominance methods were only applied to cases where pairwise comparison was possible. When the number of choice alternatives grows, pairwise comparison methods quickly become computationally unfeasible. In other words, stochastic dominance could be used for comparing two single assets or two complete funds, but it could not be used for efficiently diversified investment portfolio selection. (Post, 2003)

Ground-breaking development has been made by Kuosmanen (2004), followed by several other contributions such as Kopa and Post (2014) in order to build applications for testing SD efficiency with full diversification possibilities. These methods will be discussed in detail in chapter 2.6.

2.5 Formulating first- and second-degree dominance

As discussed in the previous chapter, the foundational idea in stochastic dominance is that it requires only simple assumptions on investor's risk preferences. Stochastic dominance can be thus used to define 'dominance' between two risky alternatives without specific knowledge on the investor's utility function.

Stochastic dominance rules rely on von Neumann-Morgenstern expected utility paradigm (EUT). It means that any utility function is assumed to be determined up to a linear transformation, is non-decreasing and all investors are assumed to maximize their expected utility $E[u(X)]$, where X is a random variable capturing the uncertain outcome. In financial applications, the expected utility can be associated with the return of an investment portfolio. (Hanoch & Levy, 1969) Notably, SD rules can be modified to fit non-EUT theories in a straightforward manner, making it less dependent on the constraints of EUT (Kuosmanen, 2004).

Several rules have been defined to find the preference between two risky options: stochastic dominance has been formulated for first, second and third-degree dominance, denoted by FSD, SSD and TSD, respectively. Higher degrees of dominance can be also formulated. (Hadar & Russell, 1971) TSD and higher orders of stochastic dominance are irrelevant for the scope of this thesis, hence only FSD and SSD are introduced.

Next, the mathematical notation for FSD and SSD will be introduced by adapting the notation from Hadar and Russell (1971) and Hanoch and Levy (1969). Consider two risky portfolios. The portfolio returns are modelled with random variables X and Y. Preference between the two risky portfolios can be defined using expected utility: X is preferred over Y if and only if the expected utility from X is higher than from Y, denoted by

$$E[u(X)] \geq E[u(Y)] \text{ for all } u \in U^o. \tag{1}$$

In equation (1), U^o is the set of all strictly increasing utility functions. That means that FSD implies preference by all investors who have a strictly increasing utility function. Whether or not dominance holds between two portfolios can be established based on a comparison of the cumulative distribution functions (CDFs) of the random variables. Specifically, X dominates Y by FSD if and only if

$$F_X(t) \leq F_y(t) \text{ for all } t \epsilon T. \tag{2}$$

According to Hanoch & Levy (1969), the interpretation of FSD is simple: X dominates Y in the sense of FSD if and only if for every value of t, the probability of getting t or less from X is smaller than from Y.

Next, second-degree stochastic dominance (SSD) is defined. SSD requires an assumption of a risk-averse or risk-neutral decision maker. In other words, X is preferred over Y by SSD if the expected utility from X is higher than from Y for all increasing concave utility functions, denoted by

$$E[u(X)] \geq E[u(Y)] \text{ for all } u \in U^{CCV}, \tag{3}$$

where U^{CCV} is the set of all increasing concave utility functions. These assumptions have a well-defined economic interpretation as they model the non-satiation and risk-aversion of an investor (Post, 2003). These assumptions lead to the conclusion that SSD efficiency means any rational and risk-averse investor would choose the SSD efficient portfolio over all other portfolios. SSD is a weaker form of dominance than FSD, hence if X FSD Y, X also dominates Y by SSD. (Hadar and Russell, 1971)

Whether or not second-degree dominance between two portfolios holds can be established by using integrals. X is said to second-degree stochastically dominate Y if and only if

$$\int_{-\infty}^{z} F_X(t) \leq \int_{-\infty}^{z} F_Y(t)\, dt \ \ \forall z \in T. \tag{4}$$

Figure 2 shows a simple case of how SSD can be visualized so that the area between the two distributions to the left from any value is positive. Hence it can be more intuitive to state SSD as the integral of the difference between the two distributions. X SSD dominates Y if and only if

$$\int_{-\infty}^{z} [\, F_X(t) - F_Y(t)]\, dt \ \geq 0 \ \forall z \in T. \tag{5}$$

Figures 1 and 2 illustrate simple cases of FSD and SSD. In figure 1, X dominates Y by FSD: X is always down (or to the right) from Y. In figure 2 there is no first-degree dominance, since the two distributions intersect. However, X second-degree dominates Y since to the left from any point t, the area between X and Y is positive.

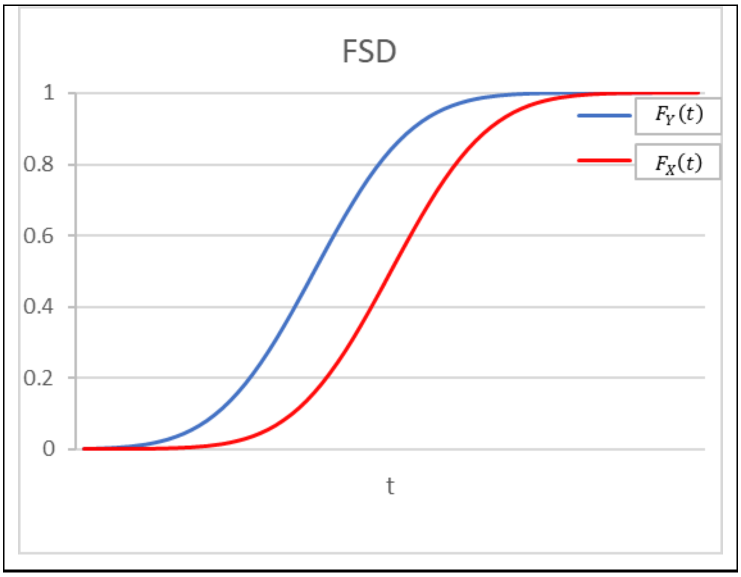


Figure 1. Illustration of a situation in which the random variable X dominates Y in the sense of FSD.

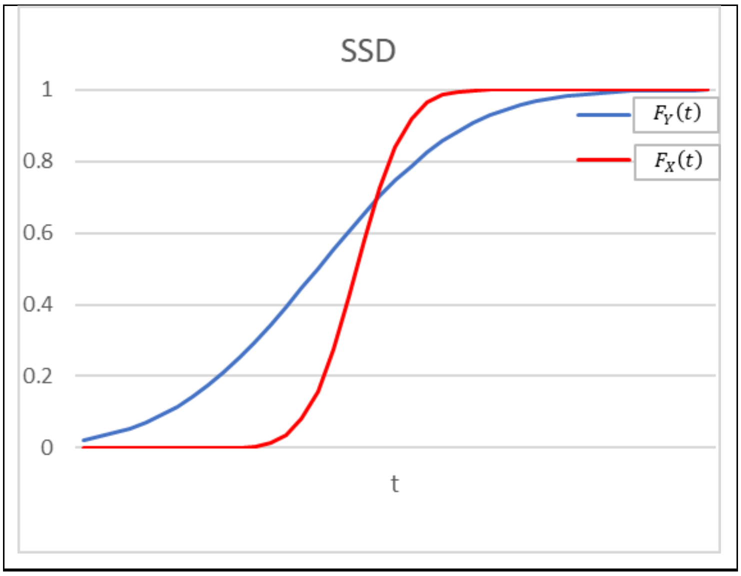


Figure 2. Illustration of a situation in which the random variable X dominates Y in the sense of SSD.

2.6 Stochastic dominance and efficient diversification

Conventional financial applications of stochastic dominance use the asset return data to form empirical distribution functions by ordering the realized returns into an empirical distribution function (EDF). The process of ordering the return data and forming an empirical distribution has a clear downside of losing the information of cross-sectional dependencies between different assets, making it impossible to form an efficiently diversified portfolio. (Kuosmanen, 2004)

First solutions to using stochastic dominance in portfolio choice problem have been proposed by Dentcheva et al. (2003), Post (2003) and Kuosmanen (2004). Out of these early developments, Dentcheva et al. propose a solution of a more general nature, while the solution by Post is simple and computationally lighter, but does not provide a general dominating portfolio (Lizyayev, 2010). Hence, the methodology developed by Kuosmanen is perhaps the most foundational for portfolio selection with stochastic dominance.

The solution by Kuosmanen to the information loss when ordering the data to an EDF is to take a reverse approach, where data is not transformed to EDFs, but rates of return are represented as state-specific vectors. Kuosmanen introduces an efficiency test for FSD and SSD that can be solved using standard linear programming techniques. If the tested portfolio is found inefficient, the linear programming method yields weights for a portfolio that provides dominance by SD criteria over the tested portfolio. In the empirical part of this thesis, the SSD test by Kuosmanen will be implemented.

These developments have been followed since by several proposals by multiple researchers. Kopa & Post (2011; 2014) propose a linear programming method for testing SSD efficiency. Their method yields two types of information: if the tested portfolio is SSD efficient, it identifies a vector for a utility function that rationalizes the efficiency. In case of inefficiency, the method identifies an efficient portfolio that dominates the tested portfolio.

According to Hodder et al. (2015), stochastic dominance methods for portfolio selection by Kuosmanen (2004) and Kopa and Post (2011) perform well out-of-sample when using several performance measures: Sharpe ratio measuring the proportion of excess return to standard deviation, value at risk (VaR) measuring left-tail risk and certainty equivalents. Hodder et al. find stochastic dominance strategies to dominate the value-weighted market portfolio used as a benchmark. They also report stochastic dominance strategies to perform better than traditional, simple portfolio selection strategies that are using Sharpe ratio and information ratio for portfolio construction.

Longarela (2016) develops SSD portfolio selection further by defining a method that allows for building an SSD efficient frontier. This means that the method does not only yield a single optimal portfolio, but an investor is able to select among all the SSD efficient portfolios. While the method has clear advantages, it is not yet as applicable to empirical problems as other portfolio selection methods due to computational limitations (Longarela, 2016). One of the latest developments by Liesiö et al. (2020) introduces SD methodology under incomplete probability information, which is theoretically more sound for future investment decisions than equal state probabilities used in many of the earlier developments.

2.7 Portfolio selection and performance measurement in SRI studies

Many studies that measure SRI portfolio performance use rather simple methods to build the portfolios. Kempf & Osthoff (2007) and Derwall et al. (2005) form simple value-weighted portfolios after applying SRI constraints. Auer & Schuhmacher (2016) apply first an ESG screen and then form simple equal-weighted portfolios. Out of the widely cited studies on SRI portfolio performance, none is found to take the approach of first applying an SRI constraint and then *optimizing* the portfolio weights.

Regarding the measurement of performance, the majority of studies on financial performance of SRI rely on the mainstream financial models such as CAPM-based Jensen's alphas or alternative alphas from more complicated, yet well-known multifactor models by Fama and French (1993) and Carhart (1997) (Derwall et al., 2011). Built on the theoretical assumptions of Markowitz's (1952) Portfolio Theory, the capital asset pricing model (CAPM) proposed by Sharpe (1964) has been a dominating model in financial literature. While the CAPM has not been able to fully withstand empirical tests (Black et al. 1972), it has held its importance in both theoretical and practical applications.

Jensen's alpha measures the risk premium per unit of systematic risk (i.e. market risk). Based on the assumption that CAPM represents the correct equilibrium model, Jensen's alpha can be calculated by

$$\alpha = (R_p - R_f) - ß_p(R_m - R_f), \tag{6}$$

where R_p is the portfolio return, R_f the risk-free rate, R_m the market portfolio return and $ß_p$ the beta of the portfolio. Jensen's alpha measures thus the risk-adjusted return premium compared to the market portfolio.

If CAPM is assumed to hold, Jensen's alpha suits a well-diversified investor who is concerned with the exposure to market risk. Sauer (1997) argues that when an SRI constraint is applied, the investment universe is restricted, and the investor is exposed to unsystematic risk as well. Hence, an SRI investor should be more interested in the total risk of the portfolio rather than only market risk.

Factor models by Fama and French (1993) and Carhart (1997) have withstood empirical tests better than Jensen's alpha. For example, the Fama-French three factor model can be formulated by

$$\alpha = (R_p - R_f) - ß_1(R_m - R_f) - ß_2 SMB - ß_3 HML, \tag{7}$$

where SMB denotes the small-minus-big size premium and HML the high-minus-low value premium. It adds thus two factors to the CAPM-based Jensen's alpha by controlling for size and value factors. A positive alpha would imply a return premium.

Auer and Schumacher (2016) conclude the issues of factor models in SRI research: the number of factors is still under debate, as it is not concluded which factors are relevant for capturing all significant risk factors. It is thus difficult to ensure that these methods correctly account for risks related to responsibility factors.

Methods that measure portfolio alphas require an assumption of normality, which in practice is not fulfilled in typical samples of return data. Non-normality of returns disturbs the functionality of single- and multi-factor models. (Chung et al., 2004) This is yet another major reason why these methods are not fully suitable for comparing SRI and non-SRI investments (Auer, 2016).

There is also an on-going debate on whether active portfolio management is beneficial compared to passive investing. In passive investing, the investor holds a portfolio following the constituents of a market index. While answering to this debate is out of scope for this thesis, the passive market indices have been empirically shown to be SSD dominated, i.e.

there exist portfolios in the dominating set that SSD dominate the index (Kopa & Post, 2014; Kuosmanen, 2004; Post, 2003).

The fact that passive index can be SSD dominated argues for active portfolio management. Furthermore, SRI requires active management anyways since company-specific ESG scores are changing over time and the negative screening requires modifying the feasible region of stocks regularly.

3 Data and methodology

This chapter presents first the research process of the empirical part. Second, the process of exploring, limiting and gathering the data used for the study is introduced. The thesis required four types of data: return data in the form of a total return index (RI), responsibility scores in the form of combined ESG scores, industry classification data and OMXH25 index returns. RI and ESG scores are discussed separately. Third, the investment strategy that is used in the empirical part is defined. Fourth, the tools used in the empirical part and the implementation of Kuosmanen linear programming model are presented in detail.

3.1 Research process

The research process of this thesis is illustrated in Figure 3. *Data exploration* explains the process of finding, limiting and gathering the data needed for this thesis. During this phase, descriptive statistics were collected from return and ESG data. During *data preprocessing*, the obtained datasets were prepared to meet the requirements of the upcoming phases. Before the optimization model was ready to be implemented, *investment strategy* had to be defined to justify the investment logic used with the optimization model. *Model implementation* includes selecting the tools for the empirical part and building the Kuosmanen model used for optimizing the portfolio weights. *Analysing results* covers analyzing and comparing the out-of-sample returns with several simple and risk-adjusted measures.

Figure 3. Research process.

3.2 Data exploration

Data for the empirical part of the thesis was acquired from Refinitiv Datastream database. Datasream is a comprehensive historical financial database accessible through an Excel add-in provided by Refinitiv, one of the leading commercial financial information providers. (Refinitiv.com, 2019). Datastream was accessed in December 2019 for RI and ESG data and in January 2020 for industry classification and OMXH25 index data.

There were some preliminary constraints to data selection: first, Datastream provides return data for a wide set of assets, but ESG data for only a selection of companies. Second, in order to construct a clear benchmark to test for stochastic dominance, a well-defined geographical and temporal limitation is needed. A common practice in stochastic dominance portfolio applications is to test the dominance against an index, which requires a clear restriction into a certain stock market (Kuosmanen, 2004). Third, the dataset needs to be large enough to enable finding significant results. An exploratory approach to data was taken to fulfil these limitations.

After exploring the available data, data selection was limited to Helsinki stock market (OMX Helsinki). Time frame was selected to be 10 years, 1.1.2009-31.12.2018 since Refinitiv only offers robust ESG data from the mid-2000s.

For stock returns, one needs to consider both changes in stock prices and dividends. Datastream provides a total return index that shows the theoretical growth in value of an asset by assuming that dividends are re-invested to buy the same asset at the closing price applicable on the ex-dividend date. Total return index is calculated by

$$RI_t = RI_{t-1} \times \frac{P_t}{P_{t-1}} \tag{8}$$

or

$$RI_t = RI_{t-1} \times \frac{P_t + D_t}{P_{t-1}}, \tag{9}$$

where RI_t is the return index on day t, $RI_{t-1}$the return index on previous day, P_t is the price of the asset on day t, P_{t-1} the price of the asset on previous day and D_t is the dividend payment on ex-date t. Equation (8) is used except for the ex-date of dividend payment, when equation (9) is used. RI uses gross dividends, ignoring tax and re-investment charges. The value of RI is 100 on the base date of each asset, however only percent changes are of interest in this thesis, and the absolute value of the return index does not matter.

Besides return data from individual stocks, OMXH25 index returns were needed as a benchmark input for the optimization model. Index returns were obtained using the same measure, total return index to ensure comparability with stock return data.

3.2.1 ESG scores provided by Refinitiv

ESG scores have been chosen to be used in this study due to easy access to the scores and the wide recognition of their use as an SRI criterion. Despite of the criticism on ESG scores introduced in chapter 2.3., it is reasonable to assume that the largest and most well-known agencies are providing ESG scores robust enough for the purposes of this study.

Refinitiv collects over 400 company-level responsibility measures based on company-reported information. Based on comparability, availability and relevance of the data, Refinitiv has selected 178 separate measures that are included in the overall assessment of the companies. The total ESG score by Refinitiv is based on 10 main responsibility themes that are shown in Figure 4.

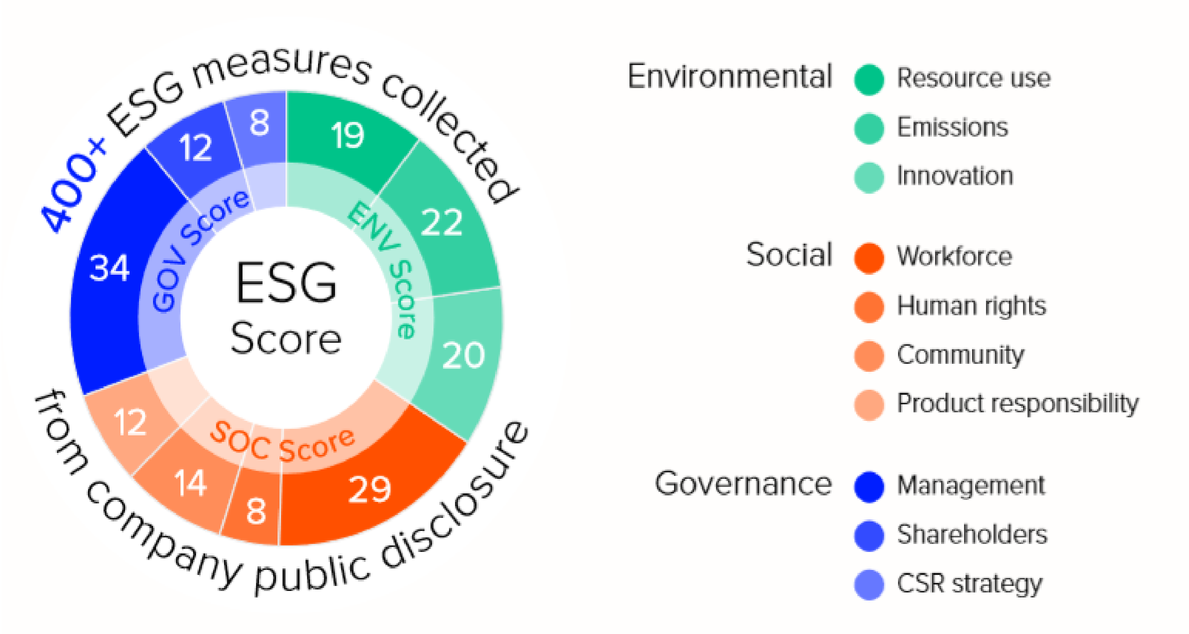


Figure 4. Refinitiv ESG scores. Refinitiv.com (2019)

Refinitiv also provides an ESG Combined score (ESGC) that incorporates controversial media publications about companies to the regular ESG score. When companies are involved in controversial incidents, ESGC score is weighted down. When a company is not involved in any controversies, ESGC score equals the regular ESG score. ESGC gives thus additional information to the company-reported measures. ESGC scores are used instead of the basic ESG scores in the empirical part of this thesis due to the extra information that they provide. However, abbreviation ESG is still used instead of ESGC in the empirical part to slightly simplify the terminology.

The ESGC scores are calculated as a percentile based on company rank. For that reason, using variance or standard deviation in ESG comparison is not very useful. Industry and geography affect the benchmarks where company performance is compared to in order to ensure the relevance of all measures, and hence it makes the most sense to compare ESGC scores within same geographical region and by using industry groupings. The basic formula for calculating a score is

$$Score = \frac{no.of\ firms\ with\ worse\ value + \frac{no.of\ firms\ with\ same\ value}{2}}{no.of\ firms\ with\ a\ value}. \tag{10}$$

Refinitiv uses an industry group classification as a benchmark when calculating the environmental, social and controversy scores since the scores are more relevant between similar companies. The industry group classification has 10 high-level groups that have been divided into 54 subgroups. In this thesis, the industry groups are aggregated according to the procedure introduced in chapter 3.4.1.

3.3 Data preprocessing

The study required four types of data: return data (total return indexes), ESG data (Combined ESG scores), industry data and OMXH25 index returns. All these datasets required some preprocessing before they were ready to be used for the analysis. Also, datasets containing information whether the stocks were ‘included’ or ‘not included’ based on the negative ESG screening were prepared in Excel. Return data, ESG scores, index returns, and ‘included’-data were entered to Jupyter Notebook as csv files and transformed to Python’s Pandas-package DataFrame objects.

In order to measure the portfolio performance, the total return index had to be transformed into percentage changes with a ready-made Pandas function. This led to the first day of return data to drop off. Return data included also some days that were not active trading days such as national holidays. These days were dropped off from the dataset. After these changes, the dataset included daily return data for 2511 days. Lastly, dates were matched among all datasets so that stock return data, OMXH25 index, ESG score dataset and 'inlcuded'-datasets all had 2511 rows of data.

The rolling investment procedure described in the part 3.4 required the stock return data to be divided into in-sample and out-of-sample periods. Both in-sample and out-of-sample datasets were divided into 108 separate periods. Since the number of companies varied during 2009-2018 due to companies entering and exiting the market, companies with no data in a specific period were dropped out. Notably, companies were dropped out only from those periods where they didn't have data instead of completely dropping them out of the analysis to avoid survival bias. This procedure allowed each time period to have a different number of stocks available.

In-sample data and out-of-sample data had to be obtained separately for all the portfolios: unconstrained benchmark portfolio and five ESG constrained portfolios. Hence, there were a total of 12 return data datasets: six in-sample datasets and six out-of-sample datasets, each containing 108 periods of data.

3.3.1 Descriptive statistics

After data preprocessing, the dataset included daily return data for 2511 trading days and 48 companies. In total there were 111,889 observations of daily returns. Table 1 describes the company-specific daily return data as percentages. Mean return varies from -0.047% (Ahtium) to 0.136% (DNA). Smallest standard deviation is 1.30% (Aktia Bank A) and the highest is 3.37% (Outokumpu A). Ahtium has the lowest daily minimum return, -43.4%, while Kone has the highest daily minimum, -6.45%. Ahtium has also the highest daily maximum, 46.5%. OMXH25 index that will be used as a benchmark input in Kuosmanen model has a mean of 0.06% with a standard deviation of 1.36% and minimum and maximum returns -8.38% and 8.35%.

Table 1. Return data, daily descriptive statistics.

Stock	Count	Mean	Stdev	Min	Max
NORDEA BANK	2511	0.066	2.09	-12.50	15.07
NOKIA	2511	0.020	2.60	-17.81	33.94
KONE 'B'	2511	0.095	1.60	-6.45	12.11
NESTE	2511	0.111	2.15	-11.47	23.69
FORTUM	2511	0.047	1.54	-13.22	10.96
SAMPO 'A'	2511	0.074	1.48	-9.40	10.72
KONECRANES	2511	0.074	2.28	-9.85	17.86
UPM-KYMMENE	2511	0.078	2.09	-12.28	13.16
WARTSILA	2511	0.090	2.05	-11.84	13.93
CITYCON	2511	0.046	1.74	-9.23	13.09
ELISA	2511	0.079	1.40	-10.32	7.35
METSO	2511	0.095	2.33	-11.01	19.43
TIETOEVRY	2511	0.082	1.87	-14.98	14.56
DNA	521	0.136	1.66	-8.74	9.94
HUHTAMAKI	2511	0.101	1.78	-14.10	12.88
KEMIRA	2511	0.079	2.02	-13.50	17.22
NOKIAN RENKAAT	2511	0.090	2.26	-11.85	16.25
SANOMA	2511	0.042	2.21	-16.35	21.42
VALMET	1254	0.107	1.85	-8.30	8.78
CARGOTEC 'B'	2511	0.090	2.48	-13.89	14.91
CAVERION	1381	0.037	2.11	-13.10	15.45
CRAMO	2511	0.093	2.43	-20.92	13.42
FINNAIR	2511	0.042	2.18	-15.05	20.43
KESKO	2511	0.073	1.72	-12.83	13.97
ORION B	2511	0.075	1.65	-12.39	15.31
OUTOTEC	2511	0.055	2.86	-36.75	15.96
PONSSE	2511	0.099	1.97	-12.93	12.63
STORA ENSO R	2511	0.064	2.19	-13.54	12.76
TIKKURILA	2201	0.020	1.57	-9.53	9.10
UPONOR	2511	0.046	2.23	-17.69	15.63
YIT	2511	0.061	2.43	-12.01	14.04
OUTOKUMPU 'A'	2511	-0.009	3.37	-24.43	21.88
F-SECURE	2511	0.046	2.24	-15.89	17.89
LEHTO GROUP	673	-0.013	2.35	-27.28	14.92
ORIOLA B	2511	0.055	2.18	-24.53	21.46
METSA BOARD B	2511	0.130	2.95	-24.09	28.00
AKTIA BANK A	2324	0.027	1.30	-8.20	7.49
TOKMANNI	672	0.042	1.77	-14.76	9.44
KESKO A	2511	0.058	1.59	-11.07	9.93
ORIOLA A	2511	0.053	2.14	-22.80	19.21
ORION A	2511	0.076	1.72	-12.12	13.40
METSA BOARD A	2511	0.129	2.92	-23.95	31.49
STORA ENSO A	2511	0.067	2.29	-11.18	17.02
AHTIUM	2423	-0.047	3.23	-43.35	46.50
RAMIRENT	2511	0.068	2.48	-18.63	15.95
RAUTARUUKKI K	2511	0.006	1.98	-9.32	12.84
AMER SPORTS	2511	0.120	2.04	-13.86	18.79
POHJOLA PANKKI	2511	0.054	1.77	-16.65	21.68
OMXH25 index	2511	0.060	1.36	-8.38	8.35

In order to further describe the return data, daily returns were aggregated to monthly returns. In total, there were 120 one-month periods. Appendix A shows the monthly return distributions of individual stocks. Based on a visual inspection, some of the stocks' return distributions show long tails and high kurtosis. If returns are not perfectly normally distributed, it could disturb other, widely used portfolio selection and performance measurement methods such as the Fama-French factor model (Chung et al., 2004). As discussed in the literature review, stochastic dominance methods do not require assumptions of normality or any other forms of the distribution (Kuosmanen, 2004; Post, 2003). The average of monthly returns across all companies during January 2009-December 2018 was 1.2%, while the minimum was -1.4% and maximum 2.8%.

Regarding the ESG scores, following descriptive statistics were calculated. Robust ESG Data was found in total for 48 companies. Out of the 48 companies with ESG data, a vast majority is in the large cap of Helsinki stock market. The number of stocks with ESG data available varies during the 10-year span from 30 to 44 companies due to companies entering and exiting the market. The arithmetic mean of all ESG scores in Helsinki stock market during the studied period 2009-2018 is 52.38. The lowest mean of a single company is 18.68 (Lehto Group) and the highest mean is 75.68 (UPM Kymmene). Table 2 reports the company specific industry category and mean, minimum and maximum ESG scores for each company. Chapter 3.4.1 will introduce the industry categorization more closely. In brief, industry categories include Basic Materials and Utilities, Industrials, Consumer Goods and Services and Finance and Technology. Figure 5 visualizes the distribution of mean ESG scores: scores are concentrated around 50, with a longer left tail.

Table 2. Company-specific ESG scores.

Stock	Category	Mean	Min	Max	Stock	Category	Mean	Min	Max
NORDEA BANK	4	62.8	42.4	82.4	**ORION B**	3	49.0	43.0	55.7
NOKIA	4	54.7	40.8	92.8	**OUTOTEC**	2	68.9	61.9	80.7
KONE 'B'	2	55.2	31.3	67.9	**PONSSE**	2	32.3	32.3	32.3
NESTE	1	67.8	48.3	84.8	**STORA ENSO R**	1	56.7	36.3	84.1
FORTUM	1	58.3	35.8	70.5	**TIKKURILA**	3	51.0	51.0	51.0
SAMPO 'A'	4	46.0	28.2	66.9	**UPONOR**	2	54.9	49.8	63.3
KONECRANES	2	42.5	21.9	71.2	**YIT**	2	48.8	28.4	65.2
UPM-KYMMENE	1	75.7	67.3	88.0	**OUTOKUMPU 'A'**	1	58.0	36.4	74.4
WARTSILA	2	57.8	40.3	65.4	**F-SECURE**	4	48.5	48.5	48.5
CITYCON	4	60.0	60.0	60.0	**LEHTO GROUP**	2	18.7	18.7	18.7
ELISA	1	57.0	50.6	61.2	**ORIOLA B**	3	33.2	23.3	48.1
METSO	2	65.8	40.7	78.3	**METSA BOARD B**	1	55.1	53.4	56.7
TIETOEVRY	4	65.0	44.7	74.0	**AKTIA BANK A**	4	43.2	43.2	43.2
DNA	4	49.1	47.9	50.3	**TOKMANNI**	3	47.3	47.3	47.3
HUHTAMAKI	3	52.2	36.6	63.4	**KESKO A**	3	69.0	61.4	76.8
KEMIRA	1	51.5	35.5	57.6	**ORIOLA A**	3	33.2	23.3	48.1
NOKIAN RENKAAT	3	45.4	33.7	59.8	**ORION A**	3	49.0	43.0	55.7
SANOMA	3	52.8	36.8	65.5	**METSA BOARD A**	1	55.1	53.4	56.7
VALMET	2	60.5	60.1	61.0	**STORA ENSO A**	1	56.7	36.3	84.1
CARGOTEC 'B'	2	50.4	37.3	56.6	**AHTIUM**	2	49.5	37.1	57.5
CAVERION	2	43.1	43.1	43.1	**RAMIRENT**	2	55.4	55.4	55.4
CRAMO	2	45.2	45.2	45.2	**RAUTARUUKKI K**	2	45.7	35.5	73.5
FINNAIR	3	48.7	48.7	48.7	**AMER SPORTS**	3	60.1	29.1	79.6
KESKO	3	69.0	61.4	76.8	**POHJOLA PANKKI**	4	38.5	24.7	44.8

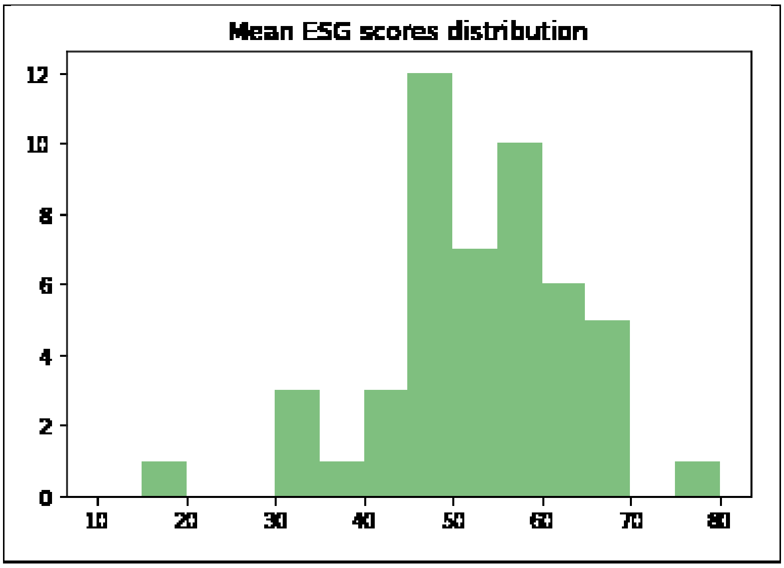


Figure 5. Mean ESG scores distribution.

3.4 Investment strategy

The investment strategy chosen for this study is an adapted momentum strategy combined with a negative ESG screening strategy. Momentum investment strategies exploit the finding that past winners continue to outperform past losers in the near future and aim to invest in the past winners and short sell past losers. Momentum strategies have gained popularity among investment professionals, becoming a widely accepted investment logic (Chan et al., 1996). In this study, the logic is used by defining the 'winner'-portfolio as the portfolio that SSD dominates the OMXH25 index portfolio during in-sample periods. Then these 'winner'-portfolios are held during out-of-sample periods.

In this study, short selling is restricted, since in practice short selling is often difficult to implement due to margin requirements and restrictions set for institutional investors (Sharpe, 1991). Another reason for restricting short selling is to simplify the required SSD test. Unfortunately, Post (2008) shows that limiting short sales reduces momentum gains.

Investment strategies with a rolling period approach have also been used by other researchers studying stochastic dominance in empirical settings (see Hodder et al., 2015; Post, 2008): by dividing the investment horizon into shorter periods, stochastic dominance can be defined in a more robust way.

Adapted from the methodology used by Hodder et al. (2015), in-sample return data from $t_0 - \Delta t, t_0$ is used to optimize the portfolio weights. Moskowitz and Grinblatt (1999) show that momentum effects are strongest with an intermediate length (6-12 months) formation period, thus time period Δt is defined to be one year. Using the weights obtained from the in-sample period, the out-of-sample return of the portfolio is measured for a one-month period t_0, t_1, again following Moskowitz and Grinblatt's recommendation for a rather short holding period. Same length for in-sample and out-of-sample returns were also used by Hodder et al. (2015).

This method allows to find the portfolio that dominates the index benchmark portfolio in the sense of SSD during the period $t_0 - \Delta t, t_0$, and using the obtained portfolio weights to calculate the returns during the out-of-sample period. Next, the in-sample period is moved forward by one month, including the old out-of-sample period in the new in-sample period. Out-of-sample data is simultaneously moved forward by one month.

The dataset contained 120 months, but since the first 12 months are only used as an in-sample period, this rolling procedure led to 108 one-year long in-sample periods and 108 one-month long out-of-sample periods. Because the number of days varies between months, in-sample datasets contained data for 248-254 trading days and out-of-sample datasets contained data for 20-23 trading days. Figure 6. Dividing sample periods. shows the rolling procedure used for defining the in-sample and out-of-sample periods.

number	**In-sample period**	**Out-of-sample period**
1	1.1.2009-31.12.2009	1.1.2010-31.1.2010
2	1.2.2009-31.1.2010	1.2.2010-28.2.2010
3	1.3.2009-28.2.2010	1.3.2010-31.3.2010
…	…	
108	1.12.2017-30.11.2018	1.12.2018-31.12.2018

Figure 6. Dividing sample periods.

3.4.1 Applying ESG scores

A rolling period strategy enables also easy implementation of responsibility in the strategy. As discussed in chapter 2.2, a common issue with mutual fund ESG scores is that they are not persistent over time when portfolios are rebalanced in response to other criteria than ESG (Wimmer, 2013). The rolling period approach of this study allows to continuously adapt the feasible region, always maintaining the desired level of ESG constraint.

This study applies a negative screening strategy. In practice, the assets with worst ESG scores are excluded from selection during each investment period. A similar approach has been used by Auer and Schuhmacher (2016): they construct the portfolio by selecting the portfolio from assets that are above a certain cut-off rate at $x\%$ of ESG scores each month.

More specifically, the ESG limit is set to be 10, 20, 30, 40 and 50 percent. A negative screen that excludes $x\%$ of the worst assets equals a positive screen that includes $(100 - x\%)$ of the best assets. However, applying a positive screen with increasingly high limit would lead to a small number of assets in the feasible region. This would lead to strongly reduced diversification possibilities (Auer, 2016). That's why this study uses a negative screen with a maximum of 50% limit.

It is also likely that when the feasible region of stocks is increasingly limited, the optimization model is not able to find stochastic dominance over an index benchmark anymore. This is the reason for testing the model with so many limits as the study seeks to find if there is a maximum cutoff rate that still enables the optimization model to find optimized weights. This issue will be discussed further in chapter 5.

Comparing ESG scores makes the most sense among industry peers (Auer & Schuhmacher, 2016). Industry data is obtained from Refinitiv Datastream. Datastream provides a categorization that divides companies into 10 industry groups. Since the number of assets is not sufficient for all specific industries, industries are divided into high-level categories by adapting the procedure used by Hyde and Sherif (2005). Companies are divided into four high-level categories:

1. Basic Materials and Utilities
2. Industrials
3. Consumer Goods and Services
4. Finance and Technology.

The details of dividing the industries are shown Figure 7. While these high-level categories are not optimal, they increase the comparability of ESG scores.

Datastream ICB industry code	Industry category
0001 Oil & Gas	Basic Materials and Utilities
1000 Basic Materials	Basic Materials and Utilities
2000 Industrials	Industrials
3000 Consumer Goods	Consumer goods and services
4000 Health Care	Consumer goods and services
5000 Consumer Services	Consumer goods and services
6000 Telecommunications	Finance and technology
7000 Utilities	Basic Materials and Utilities
8000 Financials	Finance and technology
9000 Technology	Finance and technology

Figure 7. Industry classification.

The number of stocks in each category varies during the selected time period 2009-2018: 8 to 10 stocks in Basic Materials and Utilities, 8 to 13 in Industrials, 9 to 13 in Consumer Goods and Services and 4 to 8 in Finance and Technology.

In order to exclude the worst scoring companies, companies had to be ranked. Stocks were ranked inside the four industry categories based on their ESG scores in an ascending order so that the best ESG score inside each industry group gets the rank 1. Next, the ESG limit is applied to these ranks. For example, if the limit is set to 50%, stocks that have a ranking that is among the lowest (best) 50% are set to be feasible and the highest (worst) 50% of ranks are excluded. This procedure was repeated for each investment period to get updated ESG rankings for each of the 108 periods. The procedure was also done separately to get feasible regions for 10%, 20%, 30%, 40% and 50% ESG cut-off rates.

An unconstrained portfolio was constructed using the same rolling procedure, but without any ESG constraints. The optimization algorithm was thus free to optimize weights among all the stocks that were available during any specific period T.

3.5 Methodology

3.5.1 Tools

Return data, ESG scores and OMXH25 returns were originally loaded as Excel tables from Datastream. Some preliminary modifications to the data were done in Excel, however the linear programming model by Kuosmanen (2004) was implemented by using Python programming language. To be more specific, Anaconda Jupyter Notebook was used as the platform: Jupyter Notebook is an open-source, web-based application that allows for building data analysis in an informative and visual way (jupyter.org, 2019) The optimization problem was implemented by using Gurobi Optimizer: Gurobi is a powerful solver for mathematical programming problems, such as linear programming (Gurobi.com, 2019). Gurobi can be used with several coding languages, but Python was selected due to the extensive availability of data analysis packages and simple and straightforward syntax. Jupyter Notebook was also used for descriptive statistics and obtaining and visualizing the results.

3.5.2 Implementing Kuosmanen model

While several alternatives for stochastic dominance portfolio selection have been developed after Kuosmanen's foundational work, Kuosmanen (2004) model was selected to be used in this thesis due to three main reasons. These reasons were wide recognition in literature, simplicity of the method compared to alternatives and proven robustness of out-of-sample performance by other researchers (e.g. Hodder et al. 2015).

Adapting the notation from Kuosmanen and Hodder et al. (2015), in equation (11) N stands for the number of assets available for selection, T for the number of return periods and t for a single return period. As explained in chapter 3.4, each investment period can have a different number of stocks available for selection, and a different number of trading days. The return of the ith stock on day t is denoted by y_{it}. It is important to note that y_{it} include only the stocks that were available for selection after implementing the ESG constraint. A benchmark portfolio return profile that is being dominated is denoted by y_0, and y_{0j} are the daily returns of the benchmark portfolio during a specific time period. The model output, optimized portfolio weights, are denoted by λ, and λ_i is the weight of the ith stock. The SSD test is defined as

$$\theta_2^n(y_0) = \max_{\lambda,W}\left(\sum_{t=1}^{T}\sum_{i=1}^{N} y_{it}\lambda_i - \sum_{t=1}^{T} y_{0t}\right)\Big/T$$

$$s.t. \sum_{i=1}^{N} y_{it}\lambda_i \geq \sum_{j=1}^{T} W_{tj}y_{0j} \qquad for\ t = 1, \dots, T$$

$$W \in \left\{[W_{i,j}]_{T\,x\,T}\middle| 0 \leq W_{i,j} \leq 1;\ \textstyle\sum_{i=1}^{T} W_{i,j} = \sum_{j=1}^{T} W_{i,j} = 1\ \forall_{i,j} = 1, \dots, T\right\} \tag{11}$$

$$0 \leq \lambda_i \leq 1, \forall_i = 1, \dots, N$$

$$\textstyle\sum_{i=1}^{N} \lambda_i = 1.$$

In equation (11) W is a doubly stochastic matrix that allows for ordering the state-specific return vectors in an arbitrary order: W takes the size of T x T, each item $W_{i,j}$ can take values between 0 and 1 with all rows and columns summing up to unity. When the model is optimized, values for W are obtained, but they are arbitrary in the sense that they do not provide any relevant information.

As noted, y_0 is a benchmark that the model is trying to dominate by SSD. A common practice is to use a comparable index, hence OMXH25 index is used as y_0. OMXH25 includes the 25 biggest companies in the Helsinki stock market with market value based weights. Since the majority of the 48 stocks included in this study are in the large cap of OMX Helsinki, the OMXH25 index is the best benchmark index for the purposes of this study. This is enabled by a feature of the Kuosmanen model that the benchmark y_0 does not have to be contained in the market set, i.e. it is not required that y_0 can be replicated from the tested feasible region.

The first constraint implies SSD dominance: the portfolio with optimized weights must give a higher return than any arbitrary ordering of y_0 for all periods $t = 1, \dots, T$. The objective function aims to maximize $\theta_2^n(y_0)$, which is a test statistic for SSD efficiency. $\theta_2^n(y_0) = 0$ is a necessary condition for SSD efficiency of y_0, meaning that when $\theta_2^n(y_0)$ gets positive values, the benchmark portfolio is inefficient, and a dominating portfolio in the sense of SSD can be found. The test statistic can be described as the degree of inefficiency of the benchmark portfolio, or alternatively as the maximum increase of mean return that could be gained by optimizing the portfolio weights.

According to the second constraint, portfolio weights are restricted to take values between 0 and 1, implying that short sales are restricted. The last constraint requires the portfolio weights to sum up to 1.

The actual linear programming algorithm took thus four inputs separately for each of the 108 in-sample periods: number of assets N, number of return days T, in-sample return data of the specific period and index benchmark returns y_0. Model objective was set to be the test statistic θ_2^n, which Gurobi aimed to maximize. The model variables were set to be the portfolio weights λ and the elements of W.

Next, the constraints were coded: For all periods $t = 1, \dots, T$, the return from the obtained optimal portfolio weights had to be higher than those of any arbitrary ordering of benchmark portfolio modeled with $W_{tj}y_{0j}$. Portfolio weights were bounded between 0 and 1 and they needed to sum up to 1. Elements of doubly stochastic matrix were also bounded between 0 and 1, and the sum of each row and column in W had to sum up to 1. Python implementation of the model is shown in appendix B.

While maximizing the test statistic, Gurobi returned the variable values. This enabled extracting the values of portfolio weights that were optimized based on SSD criteria. Kuosmanen (2004) finds it very likely that the model is able to find an optimal solution which means that there is a portfolio that dominates the index. However, it is possible that when the feasible region of stocks is limited with an increasingly high ESG cutoff rate, the model is not able to find dominance in all of the 108 periods. This would mean that the index is efficient compared to that specific feasible region and period. Unfortunately, it also means that the model cannot return weights for an efficient portfolio. Handling this kind of a situation is discussed in the results.

3.5.3 Computational burden

Kuosmanen's linear programming model was implemented with Gurobi optimization six times: for unconstrained portfolio (denoted by ESG0) and five ESG constrained portfolios. ESG constrained portfolios excluded the worst 50%, 40%, 30%, 20% and 10% of companies based on ESG scores, denoted by ESG50, ESG40, ESG30, ESG20 and ESG10, respectively. For each portfolio, optimizing weights was done separately for 108 in-sample periods. This led to 6 x 108= 648 optimizations of portfolio weights. Each optimization took on average 26 seconds, and the total runtime was 281 minutes with a computer having Intel Core M processor with CPU of 0.80 GHz/998 MHz and 8GB RAM.

In terms of computational complexity, in each round of optimization the model had

- $T^2 + N$ variables (elements of W and portfolio weights λ)
- $T^2 + T + N$ inequalities (for elements of W, SSD criteria and λ)
- $2T + 1$ equality constraints (for rows and columns of W and sum of λ)

Varying the length of in-sample dataset (T) and the number of stocks available for selection (N) affects thus the computational burden of the model. Varying the length of T has clearly a larger effect on computational burden than varying the number of assets N.

In Table 3 some options for length of T and N are considered. Table 3 reports in-sample data length T of 1, 6, 12 and 24 months as trading days. For each T, results are reported with $N = 30$ and $N = 45$ stocks (approximately the median and maximum of number of stocks available with different ESG constraints applied).

Table 3. Computational burden of the test statistic θ_2^n.

Months	T	N	Variables	Inequalities	Equality constraints
1	**21**	**30**	471	492	43
1	**21**	**45**	486	507	43
6	**125**	**30**	15655	15780	251
6	**125**	**45**	15670	15795	251
12	**250**	**30**	62530	62780	501
12	**250**	**45**	62545	62795	501
24	**500**	**30**	250030	250530	1001
24	**500**	**45**	250045	250545	1001

Kuosmanen (2004) states that for any given N, the SD efficiency criteria are more easily met when T is increased. However, computational burden should be considered when choosing for T. Increasing T quickly increases the number of variables and constraints. As discussed in the literature review, the optimal length for in-sample period from momentum perspective is medium-length of 6-12 months. Balancing between power and computational burden, the length of in-sample datasets is justified to be 12 months. Also, the length of the whole time period, 10 years, limits the length of T. To conclude, computational burden factors strengthen the justification of using a 12-month in-sample period.

As the result of the optimization, optimal portfolio weights were known for all six portfolios and separately for each of the 108 periods. Next, these weights from in-sample data were used to measure the performance with out-of-sample data. Out-of-sample performance was evaluated using daily data from the holding period, resulting into 108 out-of-sample observations for each portfolio (Jan 2010, Feb 2010… Dec 2018).

4 Results

In this chapter, the results from the empirical part are introduced. First it is examined how well the optimization model found stochastic dominance over the index. The possible case of not finding dominance is solved before measuring performance. Second, the composition of different portfolios is studied to understand how the portfolios were formed under different ESG constraints. Third, the key results from measuring out-of-sample performance of the portfolios are introduced. Fourth, the cumulative returns are plotted to improve the interpretation of the results.

4.1 Finding dominance over OMXH25

After running the optimization model, optimal weights from in-sample periods were known. Gurobi was able to find optimal solutions when SSD dominance was found, i.e. there were portfolio weights that provided second degree stochastic dominance over the OMXH25 index. Gurobi model was able to find optimal solutions in all of the 108 periods for the unconstrained (ESG0) portfolio, ESG10 and ESG20 portfolios. However, as predicted in chapter 3.5.2, the model could not find optimal solutions in all the periods when cutoff rate was continuously increased. For ESG30 and ESG40, Gurobi found optimal solutions for 106/108 periods, and for ESG50 Gurobi found optimal solutions for 96/108 periods.

When optimal solution was not found, portfolio weights were not obtained. Those periods where weights were not obtained could be 'artificially' filled with any chosen method. Since optimized weights for only two periods were missing from ESG30 and ESG40 optimization, those periods were filled by using equal weights

$$\lambda_i = \frac{1}{N} \; for \; all \; i = 1, \ldots, N \quad \text{and} \quad \sum_{i=1}^{N} \lambda_i = 1. \tag{12}$$

When optimizing the ESG50 portfolio, 12 periods were missing optimal weights. It is reasonable to state that if 11% of the periods would be filled with other than optimized weights, the result would be biased, i.e. the out-of-sample performance would not be based on SSD efficiency. An interpretation of the situation for practical applications would be that a cutoff rate of 50 or higher is too strict in the sense that SSD efficiency cannot be always found. ESG50 portfolio was hence decided to be left out from further analysis.

After excluding the portfolio with a 'too strict' ESG constraint, ESG10, ESG20, ESG30 and ESG40 portfolios were left for analysis. Excluding up to 40% of stocks represents the idea of negative screening strategy well.

4.2 Composition of the portfolios

The dataset had data for 48 stocks. However, the number of companies available varied through the 10-year span. Number of stocks needed to be aligned inside each in-sample and out-of-sample period, i.e. the number of stocks had to stay constant during each 12 + 1 month period. This led to the actual number of stocks available during each period to be smaller than what the actual size of the dataset was. The feasible region was also reduced when the cutoff rate of ESG constraint was increased. Table 4 shows the number of stocks that were available for selection for each portfolio. It also shows the minimum, maximum and mean number of stocks that were actually selected in each portfolio. To avoid the extreme case of choosing only 1-2 stocks, the model was constrained to choose at least 4 stocks in the portfolio.

Table 4. Number of stocks selected in portfolios.

	Available, min	Available, max	Selected, min	Selected, max	Selected, mean
ESG 0%	30	33	4	11	6.62
ESG 10%	24	29	4	10	6.56
ESG 20%	19	23	4	10	6.59
ESG 30%	13	18	4	15	6.62
ESG 40%	11	16	4	13	6.69

Table 4 shows that even when the ESG constraint is increased and the feasible region is reduced, the average number of stocks selected stays surprisingly stable. In some extreme cases, Kuosmanen model selected only four stocks in the portfolio regardless of the constraint level. The maximum number of stocks selected in the portfolio varies: at most, 11 stocks were selected in the unconstrained portfolio, 10 stocks in ESG10 and ESG20, and a maximum of 15 and 13 stocks were selected in ESG30 and ESG40 portfolios, respectively. Table 5 reports the two most commonly selected stocks in each portfolio.

Table 5. Most commonly selected stocks in each portfolio.

	Most selected stock	frequency	2. Most selected stock	2. frequency
ESG 0%	Neste	49/108	Elisa	46/108
ESG 10%	Elisa	51/108	Orion A	43/108
ESG 20%	Elisa	50/108	Orion A	46/108
ESG 30%	Orion A	68/108	Amer Sports	57/108
ESG 40%	Kesko A	72/108	Amer Sports	67/108

The most selected stock gradually changed when ESG cutoff rate was increased. Also, some stocks' selection frequency changed radically when ESG cutoff rate was increased. The biggest negative change occurred to Oriola B which was selected 32 times to the unconstrained ESG0 portfolio but 0 times in ESG20, ESG30 and ESG40 portfolios, and Konecranes which was selected 14 times in the unconstrained ESG0 portfolio but zero times to all of the ESG constrained portfolios. Changes to other direction occurred as well: Kesko A was selected only 27 times in the unconstrained portfolio but 72 times in ESG40 portfolio. UPM-Kymmene had a frequency of only 15/108 in the unconstrained portfolio but a frequency of 46/108 in ESG40 portfolio.

In order to measure the weighted 'popularity' of stocks, weights for each stock among all periods were summed. The most weighted stocks in each portfolio are reported in Table 6. *Sum of weights* is a measure for the weighted 'popularity'.

Table 6. Most weighted stocks.

	Most weighted stock	Sum of weights	2. Most weighted stock	2. Sum of weights
ESG 0%	Neste	9.24	Elisa	7.47
ESG 10%	Elisa	8.62	Amer Sports	7.32
ESG 20%	Orion A	8.04	Elisa	7.75
ESG 30%	Orion A	12.06	Amer Sports	10.60
ESG 40%	Amer Sports	11.81	Kesko A	10.34

Regarding the risk and return of the portfolios, it is in a sense irrelevant which stocks were selected in each portfolio. These results are reported to illustrate how varying the ESG cutoff rate clearly had a significant effect on the composition of the portfolios. Whether the changes in the composition of portfolios affected the risk-adjusted returns is examined in the next chapter.

4.3 Monthly out-of-sample performance

Monthly out-of-sample returns from one-month long holding periods were calculated for each portfolio. In order to measure the out-of-sample performance to evaluate the risk-return efficiency of the ESG constrained portfolios, several performance measures were calculated. Performance measures were chosen by following the practices of similar studies by Hodder et al. (2015) and Liesiö et al. (2020). The following measures were applied:

- Mean return (%): the average simple return during the 108 holding periods.

- Standard deviation (%): Standard deviation of the monthly returns.

- Skewness: measures the asymmetry of the monthly returns. A positive skew indicates that the return distribution is right-tailed, and a negative skew indicates it is left-tailed.

- Kurtosis: measures the ‘thickness’ of the tails of the distribution.

- CVaR: Conditional value-at-risk (also known as expected shortfall) measures the average of the losses that are worse than a certain limit. The limit is often set at 5%. Hence it measures the risks related to the worst outcomes.

- Sharpe ratio (Sharpe, 1964): one of the most commonly used measures for risk adjusted returns. Sharpe ratio considers the excess return over the risk-free rate compared to portfolio standard deviation. Sharpe ratio can be formulated by

$$Sharpe\ ratio = \frac{R_p - R_f}{\sigma_p}. \tag{13}$$

  In equation (13), R_p is portfolio return, R_f is the risk-free rate and σ_p standard deviation of portfolio returns. 3-month Euribor rate from 12/2018 is used as the risk-free rate. Due to the exceptional time of negative central bank interest rates the value of R_f is -0.032%.

- Sortino ratio (Sortino, 1994): Sortino ratio is similar to Sharpe ratio, but instead of the total standard deviation of the portfolio, it takes into account only the downside standard deviation, i.e. the standard deviation of negative returns. Sortino ratio is hence built on the statement that positive and negative deviations should not be treated equally, and only negative deviations are considered as risk.

Table 7. Monthy out-of-sample performance.

Measures	OMXH25	ESG 0%	ESG 10%	ESG 20%	ESG 30%	ESG40%
Mean (%)	0.98%	1.10%	1.05%	1.14%	0.88%	1.05%
Std.dev. (%)	4.40%	5.04%	4.94%	5.16%	4.97%	4.68%
Skewness	-0.17	-0.10	-0.31	-0.13	-0.22	0.04
Kurtosis	0.41	0.16	-0.03	0.16	0.17	-0.03
CVaR(5%)	-9.34%	-9.57%	-9.55%	-9.92%	-10.19%	-8.52%
Sharpe ratio	0.30	0.28	0.28	0.28	0.24	0.29
Sortino ratio	0.47	0.47	0.44	0.46	0.39	0.52

Table 7 presents the results from Unconstrained ESG0 portolio compared to ESG constrained portfolios. OMXH25 index is also included in the table for comparison. Overall, Table 7 shows very similar out-of-sample performance between the unconstrained ESG0 portfolio and ESG10, ESG20 and ESG40 portfolios. ESG30 performs weaker than the rest of the portfolios. ESG30 yields a monthly mean return of 0.88%, while the other portfolios yield mean returns of 1.05%-1.14%. Highest mean return is yielded by ESG20. Standard deviation of monthly returns stays rather stable across all the portfolios: the lowest value is 4.68% (ESG40) and the highest 5.16% (ESG20). All of the portfolios display rather low values of skewness and kurtosis.

Based on CVaR on a 5 % level, ESG10 and ESG20 yield very similar risks as the unconstrained portfolio which displays an average return of -9.6% out of the worst 5% of monthly returns. ESG30 is slightly riskier with CVaR of -10.2%. ESG40 performs slightly better, it has a CVaR of -8.5%. Examining Sharpe ratio shows results that could be deducted from mean return and standard deviation: performance of ESG10, ESG20 and ESG40 match closely the performance of the unconstrained portfolio, while ESG30 performs slightly worse. Sortino ratio reveals that when only downside deviations are considered, ESG40 provides the best performance out of all the portfolios. ESG30 yields the worst performance also based on Sortino ratio.

Based on the measures in Table 7, ESG40 is in fact the best performing portfolio as it performs better than the unconstrained portfolio based on CVaR and Sortino ratio and has essentially similar performance based on the rest of the measures. This means that ESG40 provided slightly better risk-adjusted returns, i.e. the same return was obtained with slightly less risk when compared to the unconstrained portfolio. The differences in CVaR and Sortino are small, however. ESG10 and ESG20 display performance that is very close to the unconstrained portfolio. ESG30 is the weakest portfolio based on all of the measures.

Figure 8 visualizes the distribution of monthly returns in each portfolio. As seen from the rather low values of skewness and kurtosis, the portfolios do not show heavy or skewed tails. OMXH25 index is included in the figure for comparison.

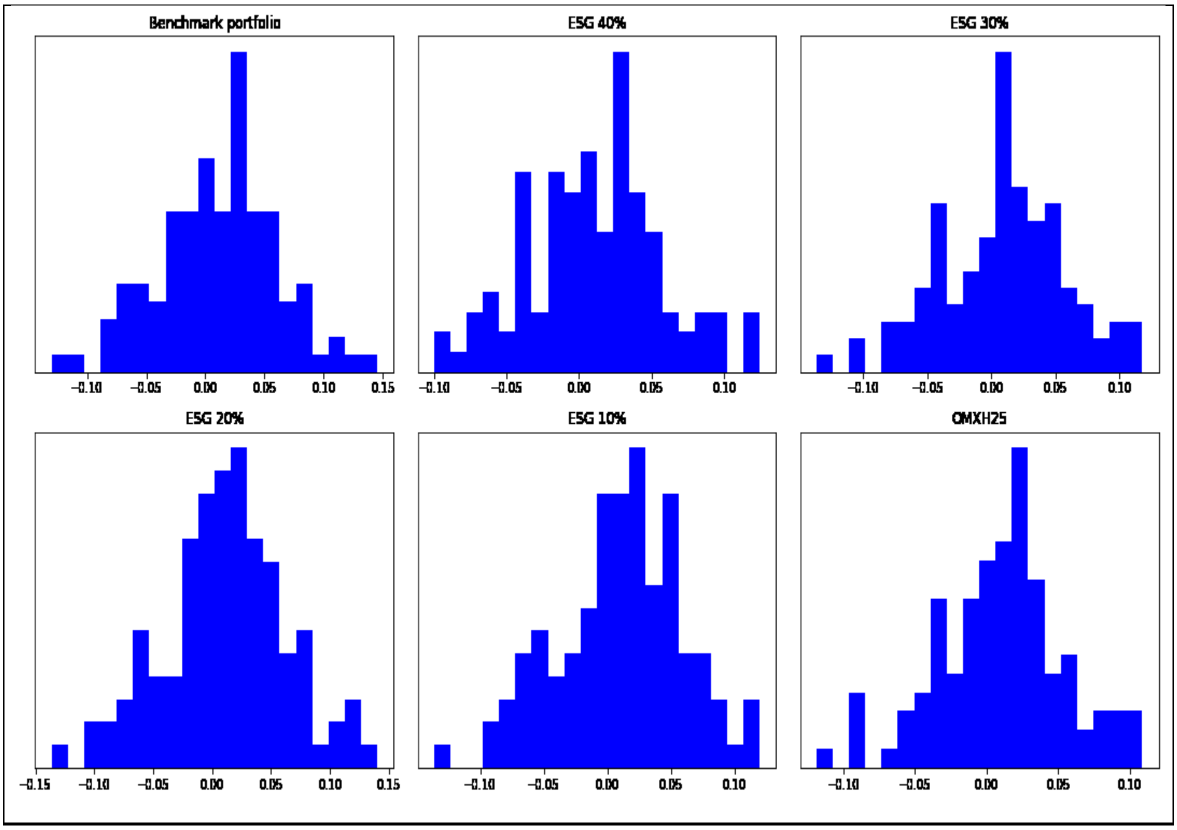


Figure 8. Monthly return distributions of optimized portfolios.

4.4 Cumulative returns

In order to get a better grasp of the performance of the portfolios, cumulative out-of-sample returns were calculated. The cumulative out-of-sample returns from Jan 2010 to Dec 2018 are visualized in Figure 9. Each subplot compares one of the ESG constrained portfolios to the unconstrained portfolio. The first subplot also plots the cumulative returns from OMXH25 index during the same time period.

Cumulative returns show how equal the performances of the unconstrained portfolio and ESG10, ESG20 and ESG40 portfolios were during the whole time period. ESG30, however, appears to provide weaker performance: from 2014 onwards, it runs below all the other portfolios. Table 8 reports the cumulative return as percentage after the last out-of-sample period.

While the focus of the results is on comparing the unconstrained portfolio and ESG portfolios, it is noteworthy that OMXH25 index that was used as a benchmark in the Kuosmanen model yields rather similar cumulative returns as the portfolios with optimized weights. All optimized portfolios except ESG30 resulted into somewhat higher cumulative return than OMXH25 index. Further analysis of the index compared to the optimized portfolios is out of scope of this thesis.

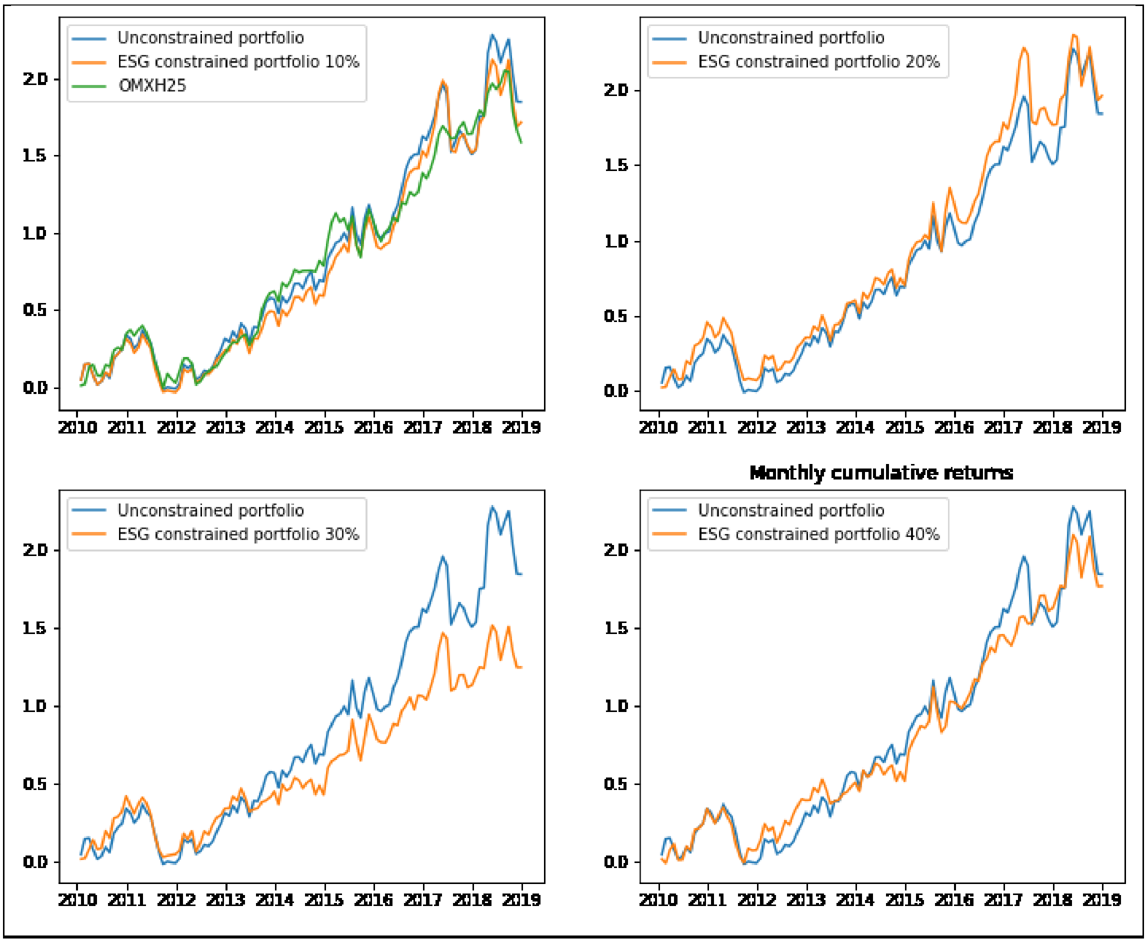


Figure 9. Cumulative returns.

Table 8. Cumulative returns.

	ESG 0%	ESG 10%	ESG 20%	ESG 30%	ESG40%	OMXH25
CUMULATIVE RETURN (%)	185%	171%	196%	125%	179%	159%

4.4.1 Testing for significance

While ESG portfolios displayed in general very similar performance as the unconstrained portfolio, it is still worthwhile to test whether there are any significant differences between portfolio returns. It is in effect only necessary to test if ESG30 returns were significantly lower than returns from the unconstrained portfolio.

A t-test can be applied to test the hypothesis that the two returns are in fact from a distribution with the same mean. If the null hypothesis is rejected, it can be concluded in some significance level that there is a difference in the mean. The two portfolio returns are paired, i.e. each return has a pair from the same day. This calls for a paired t-test.

Two-sided paired t-test between unconstrained portfolio and ESG30 yields a test statistic of 0.81 and a p-value of 0.41. The p-value is so high that the null hypothesis cannot be rejected on any significance level. This means that even though ESG30 yields a weaker cumulative return, the difference between the mean returns of unconstrained portfolio and ESG30 is not significant.

The same test can be applied to test the portfolios against OMXH25 index. While the portfolios have been formed based on SSD dominance in in-sample data, the out-of-sample performance of the portfolios appears to be very close to the performance of the index. The t-test confirms that none of the portfolios differ significantly from the index: the largest negative value of t-stat is between OMXH25 and ESG20, -0.53 with a p-value of clearly insignificant 0.59. Similarly, the paired t-test between OMXH25 and ESG30 yields a p-value of 0.74.

5 Discussion

Perhaps the most important implication from the empirical results is that an investor is able to exclude up 40% of the worst ESG scoring companies and still find portfolios that dominate the passive OMXH25 index on SSD criteria during almost the whole 10-year span from Jan 2009 to Dec 2018. However, finding portfolios that dominate the index during in-sample periods does not automatically transform into risk-return efficiency.

In order to answer the first part of the research question, out-of-sample performance of ESG constrained portfolios were compared to the performance of an unconstrained portfolio with several risk-return measures in the empirical results.

Mean return of the portfolios stayed rather stable despite of limiting the feasible region of stocks with different levels of ESG cutoff rate. Also, increasing the ESG limit did not systematically increase the riskiness of the portfolios: standard deviation and conditional value at risk (CVaR) did not display any increasing trends. Neither did increasing the ESG constraint cause any decreasing effects on the riskiness. Risk-adjusted measures, Sharpe ratio and Sortino ratio, showed stable performance among the analyzed portfolios, confirming that no trends were observed when ESG cutoff rate was increased.

These results imply that as long as the feasible set of stocks enable dominance by SSD over the index, an investor has been able to set a desired ESG constraint without a loss of performance in Helsinki during 2009-2018. Portfolios with ESG constraints, up to a cutoff rate of 40%, offered closely similar performance as an unconstrained portfolio with no ESG constraints.

However, the results show that some caution should be included: none of the ESG portfolios offered risk-adjusted return premiums compared to the unconstrained portfolio – at best, the performance was at par with the unconstrained portfolio. Also, one of the ESG constrained portfolios, ESG30, performed weaker than the rest of the portfolios. While this might have occurred by chance since no significant difference between ESG30 and unconstrained portfolio was found, it shows that one cannot automatically expect the ESG constrained portfolios to perform as well as unconstrained portfolios. There remains a risk that setting an ESG constraint leads to suboptimal returns.

In the literature review, three views on the efficiency of SRI were distinguished: some of the researchers have found SRI factors to provide superior returns. On the contrary, a large part of studies have argued that applying responsibility factors in the investment strategy would cause negative performance effects. The third view has been that investors are able to incorporate SRI factors without a loss of efficiency, but no additional return can be gained.

In this context, the results of this study support the third view. In Helsinki stock market, an investor could've applied an ESG constraint and rather safely yield equal risk-adjusted returns as from a strategy with no responsibility constraints.

Notably, when the ESG cutoff rate was increased higher than 20%, Kuosmanen model was not able to find optimal weights for all periods. With 30 % and 40 % cutoff rates, only two of the 108 were missed. However, when increasing the cutoff rate to over 40%, the model was practically unable to yield portfolio weights since over 11% of the time no optimal weights were found. The reason for that is most likely that the feasible region of stocks available for selection gets too strictly limited, and the model cannot find weights that would provide dominance by SSD over the index benchmark.

In the empirical part it was decided that if only a couple periods were missing portfolio weights, the weights were filled by using equal portfolio weights for that period. However, if the model was left with more 'gaps', it was decided to leave the portfolio out of analysis since the portfolio performance was based increasingly on other logic than dominance by SSD. This decision led to including ESG10-ESG40 portfolios but excluding ESG50 from the analysis.

The benchmark in Kuosmanen (2004) model could have been changed from index to something else to find SSD efficient portfolios easier. When the benchmark is changed from OMXH25 index to an equally weighted portfolio, finding SSD dominance was in fact easier and even with a 50% cutoff rate, the model was able to always find efficient portfolios. However, for the sake of comparability of the different portfolios and in order to follow earlier literature on SSD methods, this study sticks to the index benchmark.

Excluding ESG50 portfolio from the analysis can also be justified based on the definition in chapter 2.2: SRI investors seek to maximize utility from both wealth maximization and from following social responsibility values (Renneboog et al., 2008b). If stochastic dominance methods are used and the ESG constraint is set so strict that it blocks the method from finding SSD efficient portfolios, it could be interpreted that the constraints are preventing the investor from forming efficient portfolios. An investor investing in a strategy that imposes such a strict constraint could be seen to gain more utility from the responsibility constraint than from wealth maximization. This kind of a value-driven investor would be ready to sacrifice financial performance for the sake of responsibility.

When ESG scores are applied, the available investment universe, i.e. the number of stocks in the feasible region is reduced. Understanding the effects of limiting the investment universe require the concepts of market risk and unsystematic risk. As discussed in the literature review, the capital asset pricing model (CAPM) does not withhold empirical tests, but it offers a useful interpretation of dividing risk into market risk and unsystematic or 'diversifiable' risk. A well-diversified investor can be left with only market risk that cannot be diversified away. (Sharpe, 1964) However, if the feasible region of stocks is limited, then an investor is not able to diversify all the unsystematic risk away – the selected portfolio will inevitably include some unsystematic risk. This was one of the main reasons why this study does not resort to measuring alphas, as we're not only interested in the market risk.

The discussion on systematic (market) and unsystematic risk becomes more important when we examine the size of the optimized portfolios. The optimization model chose on average 6,5 stocks in the portfolios. With this few stocks in a portfolio it could be argued that the portfolios include quite a lot of company-specific risk that could be diversified away simply by increasing the number of stocks. Yet as discussed on the literature review, unlike MVA, stochastic dominance methods do not account for any specific risk measures such as variance, but the portfolio is chosen solely based on the SSD criteria. Further research would be required to address the issue of SSD methods choosing 'too small' portfolios.

Applying Kuosmanen model (or practically any other SD portfolio selection method) by using a similar rolling period strategy as in the empirical part of this thesis requires actively rebalancing the portfolio after each rolling period. This would in practice cause transaction costs. However, this is justified when an ESG constraint is introduced, since updating the feasible region of stocks that fulfil the ESG constraint would require rebalancing the portfolio anyways. It is anyways important to note that the results introduced in chapter 4 do not include transaction costs or taxes. This does not reduce the comparability between the portfolios since they've all been built and measured with the same logic. However, one cannot transform the returns obtained from the portfolios to be compared to real-life returns. Comparison to the OMXH25 index is also biased due to omitting transaction costs and taxes from results.

In practice, one could minimize the transaction costs by reducing the frequency of updating the portfolio. The frequency could be e.g. every 6 months, which would match the frequency of Refinitiv's ESG score updates. In this way, an investor would always have up-to-date ESG constraint and reduced transaction costs. The effect of lengthening the holding period has not been tested in this study.

The selection for in-sample and out-of-sample datasets (forming period and holding period) were selected to be 12 and 1 months, respectively. These selections were justified based on existing literature: momentum effects are strong with medium length forming period and a shorter holding period. Also, other researchers studying the out-of-sample performance of stochastic dominance methods have used similar choices. However, these selections are not fixed, and one could use for example a five-year in-sample period similarly to Kuosmanen (2004) in the original study. Increasing the length of T would've required more computational power than what was available for this thesis. One could also experiment with different lengths of out-of-sample periods - different holding periods could suit different investors. Increasing the length of the holding period would lead to reduced frequency of optimizing the weights, but also to reduced transaction costs.

Computational burden limits the use of the model when the scope is expanded further; some other methods such as Kopa and Post (2014) could provide computationally lighter solutions. That being said, increasing the number of stocks N adds to the computational burden of the model way less than increasing the length of T. This means that the model could be used with ease when the number of stocks is increased from 48 to e.g. contain all the 130 stocks in Helsinki stock market. Increasing the length of T would require more computational power than what was available for this study.

Despite of the theoretical advantages and proved performance on out-of-sample data, stochastic dominance methods have not yet been widely applied for investment portfolio selection outside academia. There are several possible reasons for this. First, the earliest efforts for building diversified portfolios with SD were published 2003-2004, and there is still constant development and extensions to the existing models in academia. Hence, no industry standards on SD methods have formed. Second, the models require understanding the theory behind them. While the efficiency test by Kuosmanen, for example, can be solved by using standard linear programming techniques, it is built on an extensive body of theory. The methods can be also argued to be less intuitive than several other, simpler portfolio selection methods. For a larger audience the methods might thus appear as 'black box' methods that yield portfolio weights using unclear logic. Third, there is a lack of ready-made, 'off-the-shelf' commercial tools for applying SD portfolio selection methods.

Building the solutions from scratch requires some technical skills and exposes for errors in the application.

This study has demonstrated that the SD optimization model by Kuosmanen can be successfully and rather easily applied to a portfolio selection problem where the method itself has a secondary role. Since SSD allows for building portfolios that are in theory preferred by all risk-averse investors, this study does not need to focus on investors' varying risk preferences, and it could focus solely on the effects of the responsibility constraints. The underlying distribution of the return data was in a sense irrelevant, as SSD should provide theoretically sound results without assuming normality in the underlying distribution.

6 Conclusion

6.1 Research summary

SRI has gained growing popularity both in academia and practice during recent 15 years. There is a lack of consensus over the performance of SRI: there are results for both abnormal and suboptimal returns compared to benchmarks. This study evaluates the performance of a negative screening strategy in the Finnish stock market. The negative screen was applied by using company-specific ESG scores that aggregate three dimensions of responsibility into one score.

In the empirical part, companies were first divided into industry groups to improve the comparability of ESG scores. Then, companies that had the worst x% of ESG scores were excluded from portfolio selection. Several ESG constrained portfolios were constructed by varying the ESG cutoff rate.

These portfolios were then compared to an unconstrained portfolio to measure performance. The theoretically appealing method of stochastic dominance was used to build both ESG constrained portfolios and the unconstrained portfolio. To be more specific, the linear programming optimization model by Kuosmanen (2004) was applied to yield optimized portfolio weights, simultaneously following the ESG constraints.

Measuring the out-of-sample results from 108 holding periods during 2010-2018 revealed that portfolios with ESG constraints on 10%, 20% and 40% level performed at par with the unconstrained portfolio when measured with several simple and risk-adjusted measures. ESG portfolio with 30% cutoff rate performed slightly, but not significantly worse than the rest of the portfolios.

One of the most important observations was that when the ESG cutoff rate was increased to over 20%, the model was not able to find stochastic dominance over OMXH25 index during all of the holding periods. If only couple periods were missed, the optimal weights could be filled with approximations, but when ESG cutoff rate reached 50%, the model was basically unable to perform. This implies that increasing the ESG cutoff rate too high leads to inefficiency in the sense of second-degree stochastic dominance.

6.2 Limitations of the study

The amount of data available created a limitation to the scope of this study. This concerns especially the availability of ESG data: robust data was only available from around 2005, which lead to limiting the time period to be 10 years, Jan 2009 – Dec 2018. More robust results might have been obtained if it would have been possible to expand the tested time period.

In addition to the time limitation, the number of stocks with Datastream ESG scores in Helsinki stock market was quite low. The number of stocks with ESG scores is likely to increase in the future as there is more and more demand for such information. However, this study was limited to analyze a total of 48 stocks. More comprehensive results could be obtained when the number of stocks is increased. In the current scope, majority of the companies were in the large cap of Helsinki. Expanding the coverage of ESG scores to smaller companies as well would provide possibilities to both investors and future research.

Some studies such as the one by Auer and Schuhmacher (2016) note that using the aggregated ESG scores might cause issues. A score that combines all the three dimensions of responsibility might hide some controversies in one dimension under the aggregated score. For example, a company with issues with environmental responsibility might still have a good ESG score if it excels in social and governance responsibility dimensions. Evaluating all the three dimensions separately and applying a separate constraint on all the dimensions could improve the accuracy of ESG scores as a responsibility measure. Datastream and other ESG score providers do offer this kind of detailed data. However, dividing the ESG score into E, S and G dimensions would've made the logic behind the screening strategy more complicated and more difficult to interpret. This is why this study resorts to aggregated ESG scores.

The industry classification that was implemented in chapter 3.4.1 to improve the comparability of ESG scores between companies is not ideal. The companies were divided into 4 groups based on upper level industry groupings. The comparability of ESG scores could've been improved by increasing the number of industry groups, however, the number of stocks created a constraint for the classification. It is likely that in practical applications the number of stocks under evaluation would significantly increase compared to this study, and thus a more accurate industry classification would be possible to further increase the usability of ESG scores.

While the method used in this study for portfolio selection, Kuosmanen (2004) linear programming model, suits testing for SSD dominance in historical return data and has also proven to perform with out-of-sample data in the study of Hodder et al. (2015), it is not the optimal method for selecting future investments. The reason for that is that it uses historical returns as equally likely states for future returns (Liesiö et al., 2020). In that sense, some of the later developments in SD portfolio selection that were discussed in the literature review such as the one by Longarela (2016) or Liesiö et al. (2020) could be theoretically better for making future investment decisions.

6.3 Managerial implications

The performance of ESG constrained portfolios compared to the unconstrained portfolio imply that portfolio managers, as well as individual investors, should be able to apply a negative screen of ESG scores without loss of performance when compared to a similar unconstrained portfolio. This is in a sense 'good news' for fund providers that are marketing their responsible investing products. Based on the results of this study, it could be justified to state that adding responsibility themes in the investment strategy should not lead to worse performance than in comparable products.

Investors should be aware though that applying ESG screening with a rolling period strategy requires active trading and creates transaction costs and taxes, that were not accounted for in this study. This should not be a major issue compared to other SRI strategies, since basically any investment strategy with responsibility themes augmented in it requires active trading and balancing the portfolio in order to sustain the 'responsibility' of the portfolio. If a portfolio would be only initially selected with sustainability criteria and then held without trading or traded with other criteria than responsibility, the 'responsibility level' would eventually diminish and the portfolio would converge into a conventional one, as discussed in the literature review.

SSD portfolio selection methods could be relevant for fund managers in general: clients usually have different risk preferences, but SSD dominance implies that any risk-averse or risk-neutral client would prefer the dominating portfolio to the dominated ones. In other words, the fund manager can assume that the obtained portfolio is preferred by all the clients that are rational and not risk-seeking (Liesiö et al., 2020).

Regarding SRI, if the level of risk aversion can be overlooked, a fund manager can focus on investors' preferences on the level of responsibility criteria. Fund managers could thus create several different funds with different levels of ESG constraints, for example one that excludes only the worst of the worst and another one that excludes a large part of companies based on ESG scores, investing only in the best-in-class companies. Based on the results of this study, portfolios with different ESG constraints can be formed without a loss of performance, as long as dominance by SSD can be found.

While this study was fully focused on selecting investment portfolios, the same model could be used for selecting any kind of portfolio when the number of risky alternatives is too high for pairwise comparisons. There could be applications in management science and operations research, for example in selecting which research projects should a company invest in when the projects need to fulfil some prior responsibility criteria.

6.4 Suggestions for future research

Different terms and concepts of SRI remain still somewhat vague in academia and likely even more so in practice. This could be one of the reasons why different researchers have found such controversial results on SRI performance. While some efforts have been done to unify SRI research (e.g. Derwall et al., 2011), more work remains to be done to improve the comparability of SRI research. This work could be extended to cover investment practioners: research on how SRI strategies are in fact utilized by different investors could be studied more closely.

Since this study focused only in Helsinki stock market and a rather limited time period, more research should be done on how the results would change in different regions and with different time periods. Also, this study used 10%-point steps in ESG cutoff rates when forming the ESG constrained portfolios. This decision was partly forced due to computational limitations, but in future research, one could try to conduct sensitivity analysis on whether an 'optimal' level of responsibility constraints could be found. 'Optimal' could mean the strictest constraint that would still dominate the benchmark by SSD and yield satisfactory returns.

Aggregated ESG scores were used in this study as the measure of responsibility of companies. More research could be conducted on the effects of environmental, social and governmental factors separately on portfolio performance.

This study resorted in only one stochastic dominance portfolio selection method by Kuosmanen (2004), but other, similar methods could be utilized in similar empirical settings. There has been constant development in SD portfolio selection in academia, and as noted in the limitations of the study, some of the more recent developments could provide theoretically even more sound investment decisions. Comparison of the different SD portfolio selection methods in different empirical settings would also be an interesting topic for future research.

References

Auer, B. R. (2016). Do Socially Responsible Investment Policies Add or Destroy European Stock Portfolio Value? *Journal of Business Ethics*, *135*(2), 381–397.

Auer, B. R., & Schuhmacher, F. (2016). Do socially (ir)responsible investments pay? New evidence from international ESG data. *The Quarterly Review of Economics and Finance*, *59*, 51–62.

Bawa, V. S. (1982). Research Bibliography—Stochastic Dominance: A Research Bibliography. *Management Science*, *28*(6), 698–712.

Bruni, R., Cesarone, F., Scozzari, A., & Tardella, F. (2017). On exact and approximate stochastic dominance strategies for portfolio selection. *European Journal of Operational Research*, *259*(1), 322–329.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, *52*(1), 57–82.

Chan, L. K. C., Jegadeesh, N., & Lakonishok, J. (1996). Momentum Strategies. *The Journal of Finance*, *51*(5), 1681.

Chung, Y. P., Johnson, H., Schill, M. J., Bailey, W., Estrada, J., Ferson, W., … Sarkissian, S. (2004). Asset Pricing When Returns Are Nonnormal: Fama-French Factors vs. Higher-order Systematic Co-Moments*. *The Journal of Business,* 79(2):923-940.

De Giorgi, E., & Post, T. (2008). Second-Order Stochastic Dominance, Reward-Risk Portfolio Selection, and the CAPM. *Journal of Financial & Quantitative Analysis*, *43*(2), 525–546.

Dentcheva, D., Ruszczy, A., & Optim, S. J. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization, 14(2),* 548-566.

Derwall, J., Guenster, N., Bauer, R., & Koedijk, K. (2005). The Eco-Efficiency Premium Puzzle. *Financial Analysts Journal*, *61*(2), 51–63.

Derwall, J., Koedijk, K., & Horst, J. Ter. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking and Finance, 35(8),* 2137-2147.

Environmental, Social and Governance (ESG) Scores from Refinitiv. (2019). Available online: www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf [20.11.2019]

Escrig-Olmedo, E., Muñoz-Torres, M., & Fernandez-Izquierdo, M. (2010). Socially responsible investing: sustainability indices, ESG rating and information provider agencies. *International Journal of Sustainable Economy*, *2*, 442–461.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance and Investment*, *5*(4), 210–233.

Global Sustainable Investment Review 2018. (2019). Available online at http://www.gsi-alliance.org/trends-report-2018/[10.11.2019]

Gurobi. (2019). Online. Available at: www.gurobi.com [25.11.2019]

Hadar, J., & Russell, W. R. (1971). Stochastic dominance and diversification. *Journal of Economic Theory*, *3*(3), 288–305.

Hanoch, G., & Levy, H. (1969). The Efficiency Analysis of Choices Involving Risk. *The Review of Economic Studies*, *36*(3), 335–346.

Hodder, J. E., Jackwerth, J. C., & Kolokolova, O. (2015). Improved Portfolio Choice Using Second-Order Stochastic Dominance. *Review of Finance*, *19*(4), 1623–1647.

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, *93*(1), 15–36.

jupyter.org. (2019). Online. Available at: www.jupyter.org [10.1.2020]

Kempf, A., & Osthoff, P. (2007). The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management*, *13*(5), 908–922.

Kopa, M., & Post, T. (2014). A general test for SSD portfolio efficiency. *OR Spectrum*, *37*(3), 703–734.

Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science*, *50*(10), 1390–1406.

Levy, H. (1992). Stochastic Dominance and Expected Utility: Survey and Analysis. *Management Science*, *38*(4), 555–593.

Liesiö, J., Xu, P., & Kuosmanen, T. (2020). Portfolio Diversification based on Stochastic Dominance under Incomplete Probability Information. *European Journal of Operational Research, forthcoming.*

Lizyayev, A. (2010). *Stochastic dominance in portfolio analysis and asset pricing*. Doctoral thesis, Erasmus University Rotterdam.

Longarela, I. R. (2016). A Characterization of the SSD-Efficient Frontier of Portfolio Weights by Means of a Set of Mixed-Integer Linear Constraints. *Management Science*, *62*(12), 3549–3554.

Markowitz, H. (1952). Portflio Selection. *The Journal of Finance*, *7*(1), 77–91.

Moskowitz, T. J., & Grinblatt, M. (1999). Do Industries Explain Momentum? *The Journal of Finance*, *54*(4), 1249–1290.

Nakai, M., Yamaguchi, K., & Takeuchi, K. (2016). Can SRI funds better resist global financial crisis? Evidence from Japan. *International Review of Financial Analysis, 48,* 12-20.

Post, T. (2003). Empirical Tests for Stochastic Dominance Efficiency. *The Journal of Finance*, *58*(5), 1905–1931.

Post, T. (2008). On the dual test for SSD efficiency. With an application to momentum investment strategies. *European Journal of Operational Research*, *185*(3), 1564–1573.

Quirk, J. P., & Saposnik, R. (1962). Admissibility and Measurable Utility Functions. *The Review of Economic Studies*, *29*(2), 140.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008a). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, *32*(9), 1723–1742.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008b). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance*, *14*(3), 302–322.

Revelli, C. (2017). Socially responsible investing (SRI): From mainstream to margin? *Research in International Business and Finance*, *39*, 711–717.

Revelli, C., & Viviani, J. (2015). Financial performance of socially responsible investing (SRI): what have we learned? A meta-analysis. *Business Ethics: A European Review*, *24*(2), 158–185.

Sandberg, J., Juravle, C., Hedesström, T. M., & Hamilton, I. (2008). The Heterogeneity of Socially Responsible Investment. *Journal of Business Ethics*, *87*(4), 519.

Sauer, D. A. (1997). The impact of social-responsibility screens on investment performance: Evidence from the Domini 400 social index and Domini Equity Mutual Fund. *Review of Financial Economics*, *6*(2), 137–149.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, *19*(3), 425–442.

Sharpe, W. F. (1991). Capital Asset Prices with and without Negative Holdings. *The Journal of Finance*, *46*(2), 489–509.

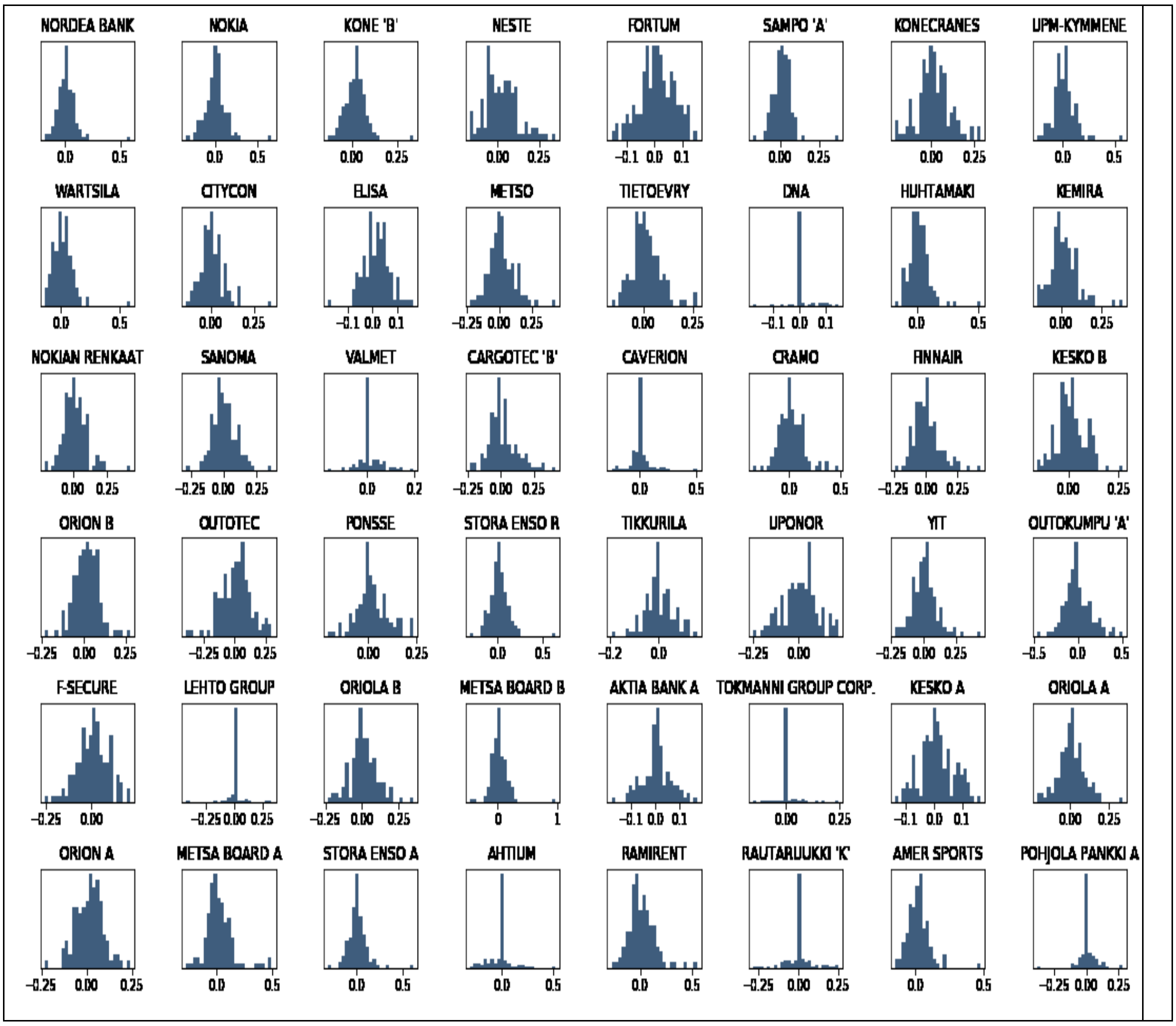
Trinks, P. J., & Scholtens, B. (2017). The Opportunity Cost of Negative Screening in Socially Responsible Investing. *Journal of Business Ethics*, *140*(2), 193–208.

US Business Round Table. (2019). Business Roundtable Redefines the Purpose of a Corporation to Promote 'An Economy That Serves All Americans.' Online. Available at: https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans[12.12.2020]

van Duuren, E., Plantinga, A., & Scholtens, B. (2016). ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. *Journal of Business Ethics*, *138*(3), 525–533.

Wimmer, M. (2013). ESG-persistence in Socially Responsible Mutual Funds. *Journal of Management and Sustainability*, *3*(1).

Appendix A: Company-specific return distributions



Appendix A. Company-specific return distributions.

Appendix B: Implementation of LP problem

```
def optimize(d, T_vars, N_vars, y0):

    # Creating the model with N + T^2 variables:
    m = Model("Test statistic for SSD dominance")
    m.update()
    x = m.addVars(range(N_vars)) #portfolio weights
    w = m.addVars(range(T_vars), range(T_vars)) #doubly stochastic matrix

    # Portfolio weights need to sum up to 1

    for i in range(N_vars):
        weight_sum = sum(x[i] for i in range(N_vars))
    m.addConstr(weight_sum == 1)

    #(b) The weights are between 0 and 1
    m.addConstrs(x[i] <= 1 for i in range(N_vars))
    m.addConstrs(x[i] >= 0 for i in range(N_vars))

    #elements of doubly stchastic matrix are between 0 and 1
    for j in range(T_vars):
        m.addConstrs(w[j,i] <= 1 for i in range(T_vars))
        m.addConstrs(w[j,i] >= 0 for i in range(T_vars))

    #Rows and columns of W sum up to one
    for j in range(T_vars):
        rowsum = sum(w[j,i] for i in range(T_vars))
        m.addConstr(rowsum == 1)

    for j in range(T_vars):
        columnsum = sum(w[i,j] for i in range(T_vars))
        m.addConstr(columnsum == 1)


    # set the constraint against a doubly stochastic matrix
    for j in range(T_vars):
        port_return = quicksum(d.iloc[j,i]*x[i] for i in range(N_vars))
        bench_permutation = quicksum(w[j,i] * y0[i] for i in range(T_vars))
        m.addConstr(port_return >= bench_permutation)

    m.update()

    # set the objective function
    obj = 0
    for j in range(T_vars):
        obj = obj+quicksum(d.iloc[j][i]*x[i] for i in range(N_vars))
    obj = obj - sum(y0[t] for t in range(T_vars))
    obj = obj / T_vars


    m.setObjective(obj, GRB.MAXIMIZE)

    # Solve the problem:
    m.optimize()
    var_values = m.getVars()
    m.update()
    weights = []
    for i in x:
        weights.append(var_values[i].x)

    return(weights)
```

Appendix B. Implementation of LP problem.

Appendix C: Stocks selected in portfolios

	Selected in Unconstrained	Selected in ESG10%	Selected in ESG20%	Selected in ESG30%	Selected in ESG40%
AHTIUM	1	1	2	1	5
AKTIA BANK A	0	0	0	0	0
AMER SPORTS	41	39	42	57	67
CARGOTEC 'B'	21	25	33	32	20
CAVERION	0	0	0	0	0
CITYCON	0	0	0	0	0
CRAMO	0	0	0	0	0
DNA	12	12	12	0	0
ELISA	46	51	50	31	31
F-SECURE	0	0	0	0	0
FINNAIR	0	0	0	0	0
FORTUM	23	26	19	27	12
HUHTAMAKI	22	22	25	8	11
KEMIRA	9	8	9	11	0
KESKO A	27	36	40	51	72
KESKO B	17	17	22	40	54
KONE 'B'	30	31	33	24	25
KONECRANES	14	0	0	0	0
LEHTO GROUP	0	0	0	0	0
METSA BOARD A	8	8	8	0	0
METSA BOARD B	2	1	0	0	0
METSO	5	5	9	24	23
NESTE	49	40	45	43	48
NOKIA	15	15	15	2	1
NOKIAN RENKAAT	30	30	36	16	17
NORDEA BANK	6	1	2	16	26
ORIOLA A	29	30	0	0	0
ORIOLA B	32	30	0	0	0
ORION A	39	43	46	68	39
ORION B	27	26	30	47	25
OUTOKUMPU 'A'	21	11	10	9	7
OUTOTEC	20	27	28	42	43
POHJOLA PANKKI A	9	0	0	0	0
PONSSE	0	0	0	0	0
RAMIRENT	0	0	0	0	0
RAUTARUUKKI 'K'	0	0	0	0	0
SAMPO 'A'	27	19	22	11	11
SANOMA	19	23	24	16	20
STORA ENSO A	26	32	27	20	21
STORA ENSO R	12	11	12	9	10
TIETOEVRY	28	29	35	45	55
TIKKURILA	0	0	0	0	0
TOKMANNI GROUP	0	0	0	0	0
UPM-KYMMENE	15	18	25	33	46
UPONOR	10	12	15	6	5
VALMET	2	3	5	5	5
WARTSILA	21	25	27	22	30
YIT	7	8	8	6	0