# SUSTAINABLE INVESTING

## EFFECTS ON PORTFOLIO RETURNS AND RISK IN A NORDIC SETTING

**Max Junestrand**

**Benjamin Wahlstedt**

Bachelor Thesis
Stockholm School of Economics
2022

**Sustainable investing: Effects on portfolio returns and risk in a Nordic setting**

Abstract:

ESG investing has gone from niche to mainstream, with the total AUM assigned to ESG-tilted portfolios growing by the day. The objective of this paper is to understand how the return and risk performance of Nordic mutual and exchange-traded funds differ based on sustainability ratings, as there is no consensus in the risk/reward payoff in ESG investing. Previous research indicates that notable differences between so called green and brown portfolios can be seen especially in times of high economic volatility, but the literature is inconclusive. The results in this paper suggest that a higher sustainability score corresponds to lower downside risk for Nordic mutual funds and ETFs, but a higher sustainability score also corresponds to a lower return profile overall. The results are less significant when separating equity portfolios and bond portfolios, however, and the largest differences can be seen when comparing the most sustainable portfolios with others.

Keywords:

ESG investing, Sustainable finance, Corporate Social Responsibility, SRI, Sustainability Rating

Authors:

Max Junestrand (24996)
Benjamin Wahlstedt (24674)

Tutor:

Olga Obizhaeva, Assistant Professor, Department of Finance, SSE

Examiner:

Adrien D'Avernas, Assistant Professor, Department of Finance, SSE

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics
©Max Junestrand and Benjamin Wahlstedt, 2022

# 1 Research question and motivation

## 1.1 Motivation

"There is one and only one social responsibility of business — to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game, which is to say, engages in open and free competition without deception or fraud."

These are the famous words of Milton Friedman, as originally expressed in his 1962 book Capitalism and Freedom, which was then exposed to a larger audience in a 1970 New York Times Magazine column. In the past decades, the notion of a business' responsibility has broadened. With increasing global concerns regarding the environment, social well-being, and fairness, the value of what a company creates outside of its P&L is becoming increasingly important for consumers, investors, and regulators. Sustainable investing has hence grown from a niche segment into a mainstream phenomenon and well-regarded investment strategy, according to industry leaders such as Bernow, Klempner and Magnin [1]. Principles for Responsible Investing (PRI) [2] which has established guidelines for investors and companies, reports that more than 5,000 institutions have signed the principles. The total AUM managed by PRI signatories are growing at an accelerating pace, reaching over $17 trillion in the US alone in 2020, as seen in figure 1.

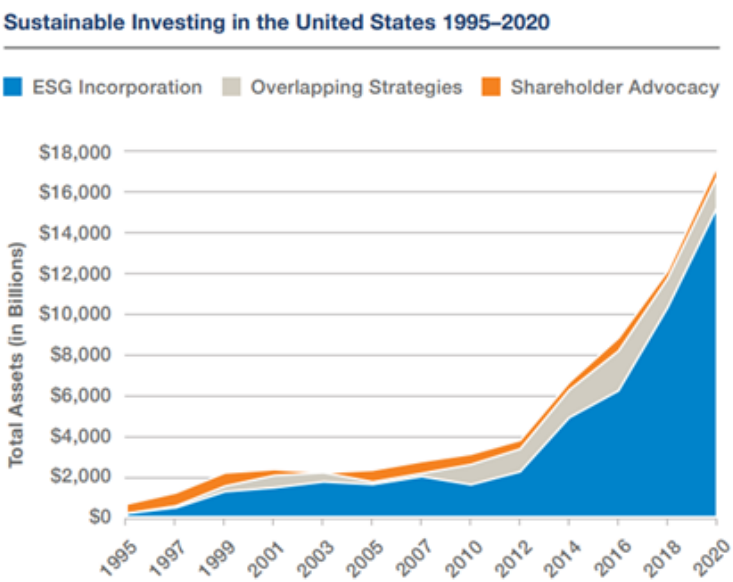


Figure 1: Report on U.S. Sustainable, Responsible and Impact Investing Trends. Source: U.S. Sustainable Investment Forum (2018)

Endorsing the PRIs means these institutions have declared to include corporate social responsibility as an essential part in their due diligence process and as a key factor in their investment decisions. The trend is not solely limited to investors who label themselves sustainable however, as investors who have no outspoken ESG rules are also increasingly pushing for corporate social responsibility and ethics according to Duuren, Plantinga and Scholtens [3].

Sustainable investing is thus on the agenda, and the importance of understanding the performance of such sustainable portfolios in comparison to traditional ones is more important than ever. The relevance of understanding ESG investing only increase when you consider that recent investor surveys indicate that significant parts of the investor community believes incorporating a sustainability view in your investment focus always affects the financial returns or risk profile of the portfolio negatively. In a recent market study containing answers from over 2,000 issuers and investors, HSBC, one of the largest banks in the world, found that when participants were prompted with the statement "I believe that practicing ESG, responsible or sustainable investing" and asked to finish the sentence 17% believe that it always involves accepting lower returns or higher risk, as seen in figure 2.

Figure 2: Sustainable financing and investing survey 2020. Source: HSBC

To recap:

1. More money is funneled into ESG funds
2. A significant share of investors believe that incorporating sustainable or responsible investing practices means accepting lower returns or higher risks

These two facts begs the question of "does sustainable investing outperform traditional funds in terms of financial returns or risk?". This question has previously been investigated from a few angles, the results of which is elaborated in section 2, but in light of recent economic instability due to the invasion of Ukraine

and the following increased market volatility – and clear evidence that without significantly changing the status quo the planet will heat to devastating levels according to UNEP [4] - there is a clear opportunity to contribute to the field. There has furthermore been no significant work done on Nordic funds in particular, which could prove to have different dynamics given the relatively high attention given to ESG topics compared to other markets, as suggested by Dahlberg and Wiklund [5].

### 1.2 Research question

Main research topic: *Is there a difference between Nordic sustainable funds and traditional funds in terms of performance and/or risk?*

The study will be limited in the following ways:

- Considers mutual and exchange-traded funds on the Nordic market.
- Considers annual data 2000-2021, weekly data 2019/01/01-2022/06/30.
- Total net asset value (NAV) returns and downside deviation are selected as performance and risk indicators respectively.
- Morningstar Sustainability Rating (MSR) is used to rank portfolios' level of sustainability.

This study extends previous research by introducing a scale in sustainability scores, as opposed to a binary comparison of "sustainable" versus "non-sustainable" or "green" versus "brown" funds. The study also extends previous research temporally into the current year.

## 2 Literature Review

The literature on sustainable and responsible investments points towards a non-negative relationship between ESG metrics and financial performance, and more importantly a majority of previous research suggest a positive relation. Friede, Busch and Bassen [6] conducted a meta study of previous research, combining the findings of close to 2200 individual papers and found 90% of studies report either a positive return or risk profile of ESG funds when compared to traditional funds over an extensive time period.

Recent studies have supported the hypothesis that sustainable firms and financial instruments benefit from lower risk and are more resilient during financially turbulent times. Hoepner, Oikonomou, Sautner, Starks and Zhou [7] investigated

the financial crisis of 2008-09 and found that U.S. firms with high ESG scores outperformed firms with lower scores during the period. Ilhan, Sautner and Vilkov [8] found that firms that are more carbon-intensive, thereby scoring lower on ESG metrics, have higher tail risk than carbon neutral or positive companies. Additionally, a recent white paper by Morgan Stanley [9] investigating the performance of nearly 11,000 US mutual funds from 2004 to 2018 found that sustainable funds demonstrate lower downside risk whilst not sacrificing any financial return compared to traditional funds. The lower downside risk was especially significant during economic turbulence and overall market downturn.

During extreme economic volatility and market crashes, there is some evidence that sustainable stocks provide lower return volatility – and even significantly higher returns. Albuquerque, Koskinen, Yang and Zhang [10] investigated stock performance during the first market crash of COVID-19 in Q1 2020, and when controlling for other factors such as policy responses, they found that sustainable firms, as quantified by their environmental and social (ES) rating, earned an additional daily return of 0.4% between February 24 until March 17 relative to low ES rated companies, for a cumulative difference of 7.2%. The study also finds no evidence for increased returns or downside risk during times of normal economic activity prior to the COVID-19 market crash. This is in line with the findings of a paper by Broadstock, Chan, Cheng and Wang [11] which studied similar stocks on the Chinese stock exchange and reached the conclusion that ESG performance mitigates financial risk during financial crisis and that the role of ESG plays less of a role in 'normal' times, pointing towards its increased importance during economic crisis. This is in line with the findings of Hoepner et al. [7], Lins, Cervaes and Tamayo [12], and Morgan Stanley [9] and strengthens the view that sustainable funds have advantages against traditional funds in times of economic stress.

With increasing attempts and engagement in green washing found by Delmas and Burbano [13], there could be some divergence between behaviour and sustainability ratings. These issues are to some degree identified and acted on by score providers such as Morningstar [14]. Additionally, there is some prior research on the divergence in the sustainability scores between different providers. Chatterji, Durand, Levine and Touboul [15], Dorfleitner, Halbriten and Nguyen [16] and Conway [17] all suggest that ESG scores differ significantly between providers. Conway finds an especially large divergence if evaluating the E, S, and G separately, with a smaller divergence for combined scores between score providers.

Previous research has also discussed the reason behind sustainability making a difference in stock returns and/or risk. Albuquerque et al.[10] suggests that the difference could be an effect of product differentiation, making customers more loyal to the company brand or product. The paper also suggests that the sustainable portfolio returns could be relatively higher in crises due to an investor preference

mechanism making the selloff in the underlying stocks less severe. The investor preference theory is also supported by Bollen's [18] findings of lower volatility in fund flows to and from socially responsible funds.

# 3 Research Design

### 3.0.1 Research method

The goal of the research is to understand how the return and risk-performance of Nordic high-ESG-scoring mutual and exchange-traded funds differ to their traditional counterparts. To distingush between sustainable and traditional funds, Morningstar Sustainability Rating is used [19]. Returns are quantified using mutual fund/ETF Net Asset Value as retrieved from Refinitiv Eikon Datastream. The year-end NAV between 2001 to 2021 is sampled to test the results in comparison to previous research. Weekly data between Jan 1, 2019 and June 30, 2022 is sampled to extend the data set used in previous research. Risk is denoted by downside deviation, see formula 1. The metric has previously been used for similar analysis [9].

For each week analyzed, the total return and downside deviation is described using the median and interquartile range in subsets. The data is evaluated on the probability that returns and downside deviation normally distributed using the Wilcoxon signed rank test. QQ-plots are used in order to validate the choice of the non-parametric Wilcoxon test to evaluate the null hypothesis that the median of two group distributions are equal.

### 3.0.2 ESG rating

A key part in investigating sustainable contra traditional financial instruments is the methodology used in classifying the sustainability of the underlying assets.

This paper uses of one of the world's most well-regarded ESG ratings, the Morningstar Sustainability Rating, see figure 3. The rating system evaluates mutual funds and exchange-traded funds (ETFs) based on how well the fund are managing ESG risks and opportunities relative to their Morningstar Category peers. The rating is distributed through the Morningstar platform but the work is carried out by independent ESG research, ratings, and analysis firm Sustainalytics.

Morningstar Sustainability Rating

| Distribution | Score | Descriptive Rank | Rating Icon |
|---|---|---|---|
| Highest 10% | 5 | High | |
| Next 22.5% | 4 | Above Average | |
| Next 35% | 3 | Average | |
| Next 22.5% | 2 | Below Average | |
| Lowest 10% | 1 | Low | |

Figure 3: Morningstar Sustainability Ratings. Source: Morningstar

The ratings are determined using a bottom-up approach by analysing the underlying holdings within a portfolio. Importantly, Morningstar's Sustainability Rating is not a metric of how sustainable a portfolio is, but rather measures the degree to which a company's monetary performance is exposed to risks driven by ESG factors. For this paper, it is assumed that this value is tightly correlated with the overall notion of a company's sustainability. If a company is highly exposed to ESG risks, such as higher emissions pricing [20], it is assumed less sustainable than a company with low exposure to ESG risks. Morningstar's ratings are relative to a global comparable portfolio. An overview of the work Sustainalytics go through to derive the sustainability rating is shown in figure 4.

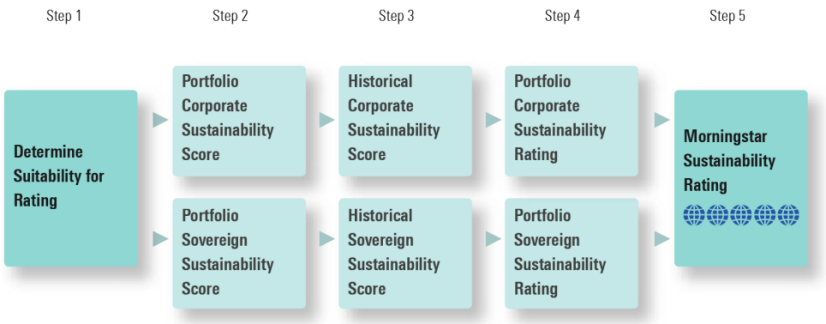


Figure 4: Morningstar Sustainability Rating Method. Source: Morningstar

# 4 Data

Eikon Datastream was used to extract Net Asset Value (NAV) data for the Nordic Fund universe. NAV for all Mutual Funds and ETFs that were tagged in Datastream as having a Nordic exposure was extracted on a yearly basis from 2000-12-31 to 2021-12-31, and on a weekly basis from 2019-01-01 to 2022-06-28. The two groups of data were cleaned separately to encompass Mutual Funds and ETFs with available NAV data for all measurement dates for the respective time period. This cleaning was conducted to avoid measurement errors due to a varying fund composition over the measurement period. After cleaning the data, 244 portfolios remain for the yearly subset and 1,307 portfolios for the weekly subset. In order to be able to separate portfolios by portfolio type for robustness testing, the Eikon Datastream portfolio classification tag is also extracted. MSR data was extracted from Morningstar Direct using portfolios' ISIN codes and paired with the respective NAV data. The most recent MSR is used to classify the fund irrespective of observed time period. Mutual Funds without a MSR is referred to as "Not rated". Summary statistics for weekly and yearly data separated by sustainability rating is found in tables 1 and 2.

Risk is denoted by the downside deviation, to single out the negative portion of risk. The downside deviation takes into account the required return and only considers returns below this required return. The downside deviation is calculated using formula 1, where $n$ is number of return observations, $r_t$ is a specific observed return and $r^*$ is the target return.

$$\text{Downside deviation} = \sqrt{\frac{\sum_{t=1}^{n}[min(r_t - r^*, 0)]^2}{n-1}} \tag{1}$$

To derive the required return from the portfolios, the beta of each individual portfolio is calculated by a regression with the MSCI Nordic index. A risk-free rate of 2 percent with a market risk premium of 6 percent is used for CAPM calculations, . This implies a market return of 8 percent for a portfolio with a beta of 1, in line with historical returns on the Nordic stock exchanges.

Table 1: Weekly data summary statistics

| Morningstar sustainability rating | 5 | 4 | 3 | 2 | 1 | Not rated |
|---|---|---|---|---|---|---|
| n | 136 | 158 | 339 | 202 | 36 | 436 |
| Mean average weekly return | 0.06% | 0.13% | 0.18% | 0.21% | 0.27% | 0.02% |
| Median average weekly return | 0.04% | 0.08% | 0.23% | 0.24% | 0.28% | -0.02% |
| Q1 average weekly return | -0.05% | -0.03% | 0.06% | 0.18% | 0.26% | -0.06% |
| Q3 average weekly return | 0.17% | 0.29% | 0.28% | 0.28% | 0.30% | 0.01% |
| Average st. dev. | 1.70% | 1.97% | 2.65% | 2.94% | 3.57% | 1.55% |
| Median st. dev. | 1.34% | 1.51% | 3.04% | 3.23% | 3.49% | 1.33% |
| Q1 st. dev. | 0.78% | 0.76% | 2.65% | 2.81% | 3.28% | 0.74% |
| Q3 st. dev. | 2.61% | 3.22% | 3.29% | 3.50% | 4.16% | 1.49% |
| Average downside deviation | 1.91% | 2.24% | 3.08% | 3.43% | 4.18% | 1.61% |
| Median downside deviation | 1.45% | 1.73% | 3.57% | 3.76% | 3.99% | 1.47% |
| Q1 downside deviation | 0.80% | 0.76% | 2.95% | 3.25% | 3.84% | 0.82% |
| Q3 downside deviation | 2.91% | 3.62% | 3.88% | 4.15% | 4.92% | 1.77% |

Table 2: Annual data summary statistics

| Morningstar sustainability rating | 5 | 4 | 3 | 2 | 1 | Not rated |
|---|---|---|---|---|---|---|
| n | 35 | 37 | 55 | 41 | 11 | 65 |
| Mean average return | 3.37% | 7.62% | 8.27% | 10.52% | 12.51% | 2.04% |
| Median average return | 0.66% | 7.80% | 8.73% | 11.79% | 13.18% | -0.31% |
| Q1 average return | -0.25% | 1.08% | 4.84% | 8.45% | 11.52% | -0.71% |
| Q3 average return | 7.44% | 13.54% | 10.45% | 13.63% | 14.90% | 1.89% |
| Average st. dev | 12.42% | 19.34% | 22.87% | 27.15% | 28.53% | 12.16% |
| Median st.dev | 6.30% | 24.17% | 24.96% | 29.78% | 31.66% | 8.22% |
| Q1 st. dev. | 5.02% | 5.78% | 22.52% | 25.73% | 25.18% | 5.63% |
| Q3 st. dev. | 22.39% | 27.25% | 27.59% | 32.61% | 34.50% | 9.10% |
| Average downside deviation | 14.99% | 22.02% | 26.15% | 28.03% | 26.45% | 10.48% |
| Median downside deviation | 7.25% | 27.99% | 28.67% | 29.68% | 29.52% | 9.03% |
| Q1 downside deviation | 5.88% | 7.29% | 26.81% | 26.95% | 24.56% | 6.28% |
| Q3 downside deviation | 26.70% | 30.70% | 31.13% | 32.79% | 31.58% | 10.91% |

Table 3: Weekly data asset class distribution

| | | Morningstar sustainability Rating | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 5 | 4 | 3 | 2 | 1 | Not rated | Sum |
| Eikon Fund Type | Equity | 37 | 73 | 257 | 170 | 34 | 54 | 625 |
| | Bond | 51 | 61 | 70 | 31 | 0 | 315 | 528 |
| | Real Estate | 3 | 9 | 2 | 1 | 0 | 3 | 18 |
| | Mixed assets | 32 | 6 | 6 | 0 | 1 | 10 | 55 |
| | Alternatives | 7 | 0 | 2 | 0 | 0 | 10 | 19 |
| | Money market | 6 | 9 | 2 | 0 | 1 | 44 | 62 |
| | Sum | 136 | 158 | 339 | 202 | 36 | 436 | 1307 |

Table 4: Annual data asset class split

| | | Morningstar sustainability Rating | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 5 | 4 | 3 | 2 | 1 | Not rated | Sum |
| Eikon Fund Type | Equity | 19 | 18 | 31 | 28 | 6 | 48 | 150 |
| | Bond | 15 | 19 | 23 | 11 | 5 | 16 | 89 |
| | Real Estate | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Mixed assets | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Alternatives | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| | Money market | 1 | 0 | 0 | 2 | 0 | 1 | 4 |
| | Sum | 35 | 37 | 55 | 41 | 11 | 65 | 244 |

Overall, the most common portfolio types in the dataset as described by Eikon are "Equity" and "Bond". A differing distribution of MSR scores can be observed between the two types. Equity portfolios are sustainability rated by Morningstar to a higher degree than bond portfolios for the weekly dataset, and bond portfolios' ratings are skewed toward higher ratings compared to equity portfolios overall for both the weekly and annual datasets. This paper makes no attempt at explaining this observation. The expected returns within categories are non-normally distributed due to the two different portfolio types' differing return profiles.

# 5 Hypothesis testing

## 5.1 Returns

There are two components to this paper's analysis of portfolio performance. A returns analysis on a yearly basis is conducted between 2000 and 2021 to compare findings with previous research such as [9], and draw new insights into how the Nordic market has performed. The second analysis is of the downside deviation for the portfolio dataset.

### 5.1.1 Portfolio Returns

For each year analyzed in the study, total returns are described using the median, as it is more robust against outliers than the mean. Since this paper, compared to previous research [9], [10] [11] has a more granular sustainability scoring with sustainability scores of 1-5 as opposed to a binary classification of sustainable or non-sustainable, the difference in performance between the funds is determined using the both non-parametric Wilcoxon statistical test [21] between all the different funds, and the Friedman Test [22] to test the equality of medians across all samples. These tests are similar to paired T-Test and ANOVA for repeated measures, but instead of testing the equality of means between samples, the Wilcoxon and Friedman Test test the equality of medians. These tests are chosen due to the non-normal distribution of the data.

The median total return varies by sustainability rating of the portfolios across the years, as seen in figure 5 where the returns of the different portfolio groups are depicted by year. The graph suggests that a lower sustainability rating corresponds with a higher volatility of returns.

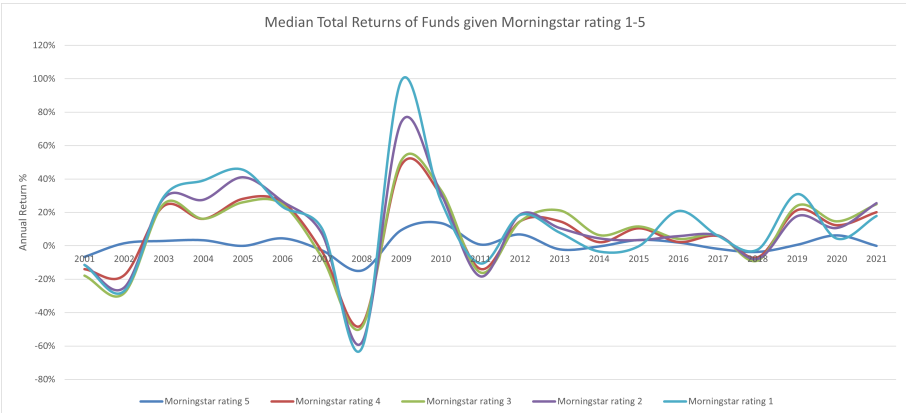


Figure 5: Median total returns of funds between 2001 and 2021

In figure 6 below the percentage-point difference between the five groups of funds in terms of medians is shown.

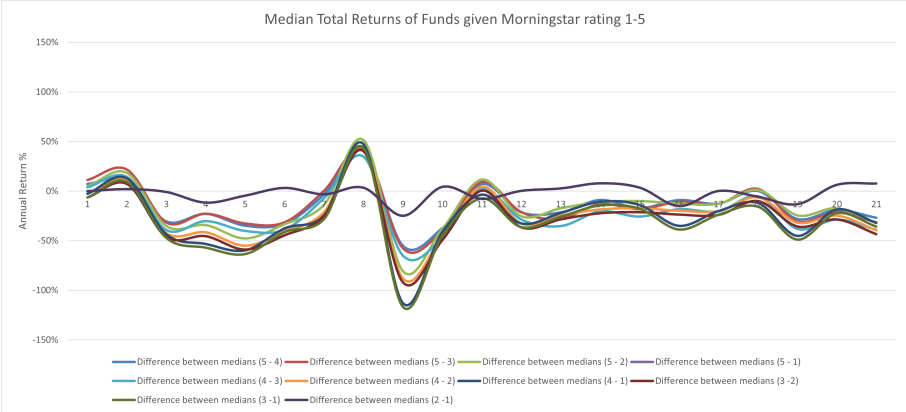


Figure 6: Difference in median total returns between funds between 2001 and 2021

The results from the statistical tests which indicate whether or not the differences can be considered statistically significant is presented. Note that for the group classified with a sustainability rating of 1, the dataset only includes 11 observations which reduces the statistical strength of hypothesis testing. With the Friedman Test, the following hypothesis is tested:

- Null Hypotheses $H_0$ : Median treatment effects of the population are all the same
- Alternative Hypotheses $H_1$ : There is a difference in treatment effects.

Table 5: Friedman Test results

| **Year** | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| **p value** | 5.58E-07 | 1.46E-10 | 9.90E-13 | 9.42E-10 | 2.41E-12 | 1.26E-11 |

| **Year** | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|
| **p value** | 1.01E-07 | 9.74E-14 | 1.33E-09 | 2.05E-11 | 6.72E-07 | 3.86E-11 |

| **Year** | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|
| **p value** | 3.87E-12 | 3.02E-15 | 3.78E-13 | 7.27E-06 | 3.72E-11 | 1.26E-06 |

| **Year** | 2019 | 2020 | 2021 |
|---|---|---|---|
| **p value** | 1.46E-10 | 1.20E-12 | 1.87E-11 |

The equality of median returns for every year and sustainability group is rejected with a statistical significance of 99.99%+ for every single year. This is not unexpected from observing figure 5.

Prior research considers funds on a binary scale, sustainable or non-sustainable. This paper considers portfolios in five groups over 20 years, and statistical analysis is conducted on the difference between each pair for each given year. This yields one test hypothesis per pair, per year. Total yearly returns from funds categorized by ratings 1 through 5 are denoted "1" ... "5" for simplicity.

- For each year between 2001 and 2021 and fund group [5 ... 2], compare it with another fund group [4 ... 1], and form a Null Hypotheses $H_0$ : Median(i) - Median(j) = 0, where $i$ and $j$ denotes fund groups such that $i > j$ and an alternative Hypotheses $H_1$ : Median(i) - Median(j) $\neq$ 0.

Parts of figure 6 is presented in combination with a table of the differences and statistical significance levels below. The full table can be found in the appendix, table 22.

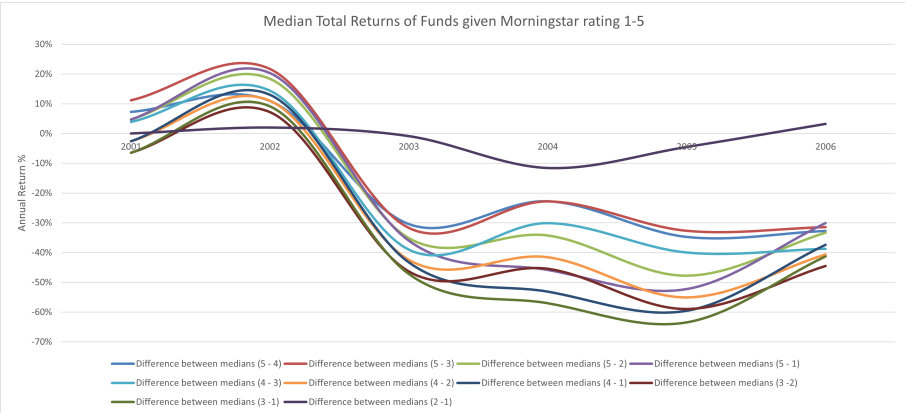


Figure 7: Difference in total returns for each pair of fund categorization between 2001 and 2006

Table 6: Return analysis 2001-2006. Significance levels: * 95%, ** 99%, and *** 99.5%.

| Year | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | 7.3% | 19.2% | -20.9% | -12.7% | -28.1% | -21.6% |
| p-value | 0.047 | 0.002 | 0.001 | 0.007 | 0.005 | 0.001 |
| Significance | * | *** | *** | ** | ** | *** |
| Diff medians (5 - 3) | 11.2% | 29.9% | -22.1% | -12.8% | -26.1% | -20.3% |
| p-value | 0.009 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 |
| Significance | ** | *** | *** | *** | *** | *** |
| Diff medians (5 - 2) | 4.7% | 26.5% | -25.5% | -24.2% | -41.1% | -22.1% |
| p-value | 0.238 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Significance | | *** | *** | *** | *** | *** |
| Diff medians (5 - 1) | 4.7% | 28.5% | -26.4% | -35.8% | -45.6% | -18.9% |
| p-value | 0.350 | 0.051 | 0.021 | 0.003 | 0.000 | 0.027 |
| Significance | | | * | *** | *** | * |
| Diff medians (4 - 3) | 3.9% | 10.7% | -1.3% | -0.1% | 2.1% | 1.3% |
| p-value | 0.262 | 0.061 | 0.324 | 0.380 | 0.145 | 0.397 |
| Significance | | | | | | |
| Diff medians (4 - 2) | -2.6% | 7.4% | -4.7% | -11.5% | -13.0% | -0.5% |
| p-value | 0.148 | 0.034 | 0.037 | 0.000 | 0.000 | 0.386 |
| Significance | | * | * | *** | *** | |
| Diff medians (4 - 1) | -2.6% | 9.4% | -5.6% | -23.0% | -17.5% | 2.7% |
| p-value | 0.103 | 0.449 | 0.416 | 0.016 | 0.051 | 0.350 |
| Significance | | | | * | | |
| Diff medians (3 - 2) | -6.5% | -3.4% | -3.4% | -11.4% | -15.0% | -1.8% |
| p-value | 0.075 | 0.319 | 0.237 | 0.001 | 0.010 | 0.407 |
| Significance | | | | *** | ** | |
| Diff medians (3 - 1) | -6.5% | -1.4% | -4.3% | -23.0% | -19.6% | 1.4% |
| p-value | 0.009 | 0.260 | 0.350 | 0.016 | 0.0416 | 0.120 |
| Significance | ** | | | * | * | |
| Diff medians (2 - 1) | 0.0% | 2.0% | -0.9% | -11.6% | -4.5% | 3.2% |
| p-value | 0.120 | 0.483 | 0.382 | 0.042 | 0.449 | 0.160 |
| Significance | | | | * | | |

There are statistically significant differences between the total yearly returns of some of the fund groups. In particular there is a difference between median returns in funds classified as MSR 5 and the other fund groupings in all but three cases at at least a 95% significance level. These tests suggest that between the years 2001 and 2006 most of the differences in return profile is statistically significant between portfolios rated as 5 and the other, as well as between groups 4 and 2.

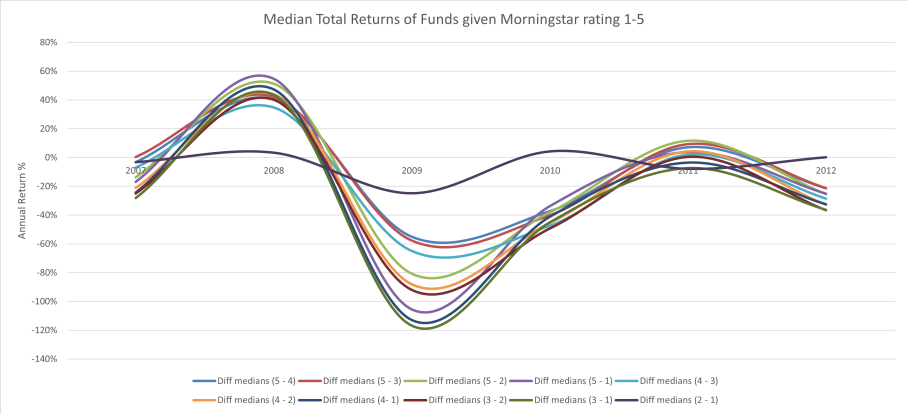


Figure 8: Difference in total returns for each pair of fund categorization between 2007 and 2012

Table 7: Return analysis 2007-2012. Significance levels: * 95%, ** 99%, and *** 99.5%.

| Year | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | 0.7% | 32.3% | -38.8% | -16.8% | 14.6% | -7.8% |
| p-value | 0.053 | 0.007 | 0.002 | 0.000 | 0.078 | 0.002 |
| Significance | | ** | *** | *** | | *** |
| Diff medians (5 - 3) | 4.3% | 33.9% | -41.5% | -19.7% | 16.8% | -7.7% |
| p-value | 0.005 | 0.000 | 0.001 | 0.000 | 0.017 | 0.005 |
| Significance | ** | *** | *** | *** | * | ** |
| Diff medians (5 - 2) | -9.9% | 43.1% | -64.6% | -17.7% | 19.1% | -11.9% |
| p-value | 0.013 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Significance | * | *** | *** | *** | *** | *** |
| Diff medians (5 - 1) | -13.1% | 46.5% | -89.4% | -13.4% | 11.3% | -11.6% |
| p-value | 0.120 | 0.012 | 0.005 | 0.004 | 0.382 | 0.005 |
| Significance | | * | *** | *** | | *** |
| Diff medians (4 - 3) | 3.5% | 1.6% | -2.7% | -2.9% | 2.1% | 0.1% |
| p-value | 0.174 | 0.170 | 0.297 | 0.369 | 0.247 | 0.319 |
| Significance | | | | | | |
| Diff medians (4 - 2) | -10.6% | 10.8% | -25.8% | -0.9% | 4.4% | -4.1% |
| p-value | 0.004 | 0.000 | 0.001 | 0.178 | 0.013 | 0.002 |
| Significance | *** | *** | *** | | * | *** |
| Diff medians (4 - 1) | -13.8% | 14.2% | -50.6% | 3.4% | -3.4% | -3.9% |
| p-value | 0.088 | 0.051 | 0.009 | 0.483 | 0.207 | 0.260 |
| Significance | | | ** | | | |
| Diff medians (3 - 2) | -14.2% | 9.2% | -23.1% | 2.0% | 2.3% | -4.1% |
| p-value | 0.000 | 0.010 | 0.001486 | 0.257 | 0.041 | 0.000 |
| Significance | *** | * | *** | | * | *** |
| Diff medians (3 - 1) | -17.4% | 12.6% | -47.9% | 6.3% | -5.5% | -3.9% |
| p-value | 0.009 | 0.207 | 0.009 | 0.449 | 0.027 | 0.004 |
| Significance | ** | | ** | | * | *** |
| Diff medians (2 - 1) | -3.2% | 3.4% | -24.8% | 4.3% | -7.8% | 0.2% |
| p-value | 0.260 | 0.350 | 0.051 | 0.183 | 0.088 | 0.382 |
| Significance | | | | | | |

The years between 2007 and 2012 covers the financial crisis. This time period has been covered in more detail by the likes of Lins et al. [12] and the period was earlier than the wider attention given to sustainable investing. According to the Swedish Riksbank [23] the financial implications didn't reach the Nordics until the US investment bank Lehman Brothers went bankrupt in Q3 of 2008. As a result of the financial crisis, previous research [12] [10] [11] points towards more significant

differences in returns based on sustainability. The results that for every difference except for 4 and 3, and 2 and 1 returns differ on a 99%+ significance level for 2009 is thus in line with previous research. During 2008 the differences between median returns for more sustainable funds and less sustainable funds were positive across comparisons and significant in six out of ten cases and in 2009 the opposite pattern can be observed.

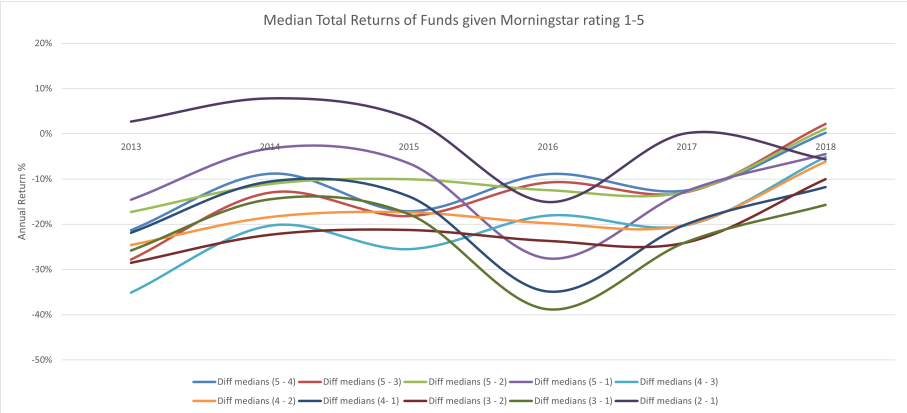


Figure 9: Difference in total returns for each pair of fund categorization between 20013 and 20018

Table 8: Return analysis 2013-2018. Significance levels: * 95%, ** 99%, and *** 99.5%.

| Year | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | -16.8% | -2.5% | -6.9% | -0.3% | -7.7% | 3.2% |
| p-value | 0.024 | 0.105 | 0.002 | 0.330 | 0.001 | 0.006 |
| Significance | * | | *** | | *** | ** |
| Diff medians (5 - 3) | -23.3% | -6.7% | -8.0% | -2.2% | -8.1% | 5.2% |
| p-value | 0.002 | 0.000 | 0.007 | 0.105 | 0.000 | 0.001 |
| Significance | *** | *** | ** | | *** | *** |
| Diff medians (5 - 2) | -12.8% | -4.8% | 0.1% | -3.8% | -8.0% | 4.1% |
| p-value | 0.010 | 0.000 | 0.061 | 0.008 | 0.001 | 0.001 |
| Significance | * | *** | | ** | *** | *** |
| Diff medians (5 - 1) | -10.0% | 3.0% | 3.6% | -18.9% | -7.8% | -1.5% |
| p-value | 0.483 | 0.289 | 0.350 | 0.003 | 0.051 | 0.088 |
| Significance | | | | *** | | |
| Diff medians (4 - 3) | -6.5% | -4.2% | -1.1% | -1.9% | -0.4% | 1.9% |
| p-value | 0.198 | 0.010 | 0.392 | 0.202 | 0.433 | 0.468 |
| Significance | | * | | | | |
| Diff medians (4 - 2) | 4.0% | -2.3% | 7.1% | -3.6% | -0.3% | 0.9% |
| p-value | 0.421 | 0.007 | 0.182 | 0.022 | 0.480 | 0.034 |
| Significance | | ** | | * | | * |
| Diff medians (4 - 1) | 6.7% | 5.6% | 10.6% | -18.7% | -0.2% | -4.7% |
| p-value | 0.088 | 0.289 | 0.042 | 0.034 | 0.416 | 0.007 |
| Significance | | | * | * | | ** |
| Diff medians (3 - 2) | 10.5% | 1.9% | 8.2% | -1.7% | 0.1% | -1.0% |
| p-value | 0.058 | 0.392 | 0.018 | 0.021 | 0.262 | 0.279 |
| Significance | | | * | * | | |
| Diff medians (3 - 1) | 13.2% | 9.8% | 11.7% | -16.8% | 0.2% | -6.7% |
| p-value | 0.003 | 0.034 | 0.021 | 0.001 | 0.490 | 0.009 |
| Significance | *** | * | * | *** | | ** |
| Diff medians (2 - 1) | 2.7% | 7.9% | 3.5% | -15.1% | 0.1% | -5.6% |
| p-value | 0.120 | 0.042 | 0.207 | 0.005 | 0.233 | 0.051 |
| Significance | | * | | *** | | |

For 2013-2018, the tests suggest a statistically significant difference in median returns for several comparisons to group 5.

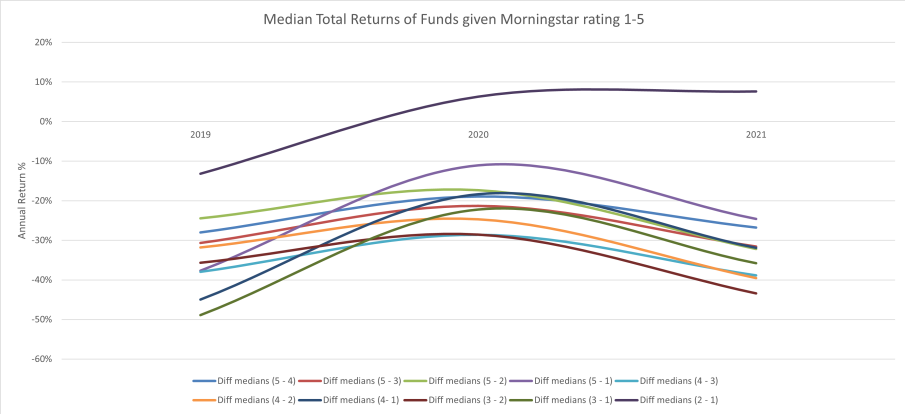


Figure 10: Difference in total returns for each pair of fund categorization between 2019 and 2021

Table 9: Return analysis 2019-2021. Significance levels: * 95%, ** 99%, and *** 99.5%.

| Year | 2019 | 2020 | 2021 |
|---|---|---|---|
| Diff medians (5 - 4) | -20.7% | -6.0% | -20.1% |
| p-value | 0.004 | 0.006 | 0.013 |
| Significance | *** | ** | * |
| Diff medians (5 - 3) | -23.4% | -8.3% | -24.9% |
| p-value | 0.000 | 0.007 | 0.000 |
| Significance | *** | ** | *** |
| Diff medians (5 - 2) | -17.2% | -4.4% | -25.6% |
| p-value | 0.000 | 0.067 | 0.000 |
| Significance | *** | | *** |
| Diff medians (5 - 1) | -30.4% | 1.9% | -17.9% |
| p-value | 0.012 | 0.260 | 0.051 |
| Significance | * | | |
| Diff medians (4 - 3) | -2.7% | -2.3% | -4.8% |
| p-value | 0.357 | 0.282 | 0.094 |
| Significance | | | |
| Diff medians (4 - 2) | 3.5% | 1.6% | -5.4% |
| p-value | 0.340 | 0.186 | 0.021 |
| Significance | | | * |
| Diff medians (4 - 1) | -9.7% | 7.9% | 2.2% |
| p-value | 0.183 | 0.027 | 0.160 |
| Significance | | * | |
| Diff medians (3 - 2) | 6.2% | 3.9% | -0.6% |
| p-value | 0.225 | 0.104 | 0.442 |
| Significance | | | |
| Diff medians (3 - 1) | -7.0% | 10.3% | 7.0% |
| p-value | 0.139 | 0.042 | 0.382 |
| Significance | | * | |
| Diff medians (2 - 1) | -13.2% | 6.3% | 7.6% |
| p-value | 0.003 | 0.103 | 0.183 |
| Significance | *** | | |

In the most recent years, there are statistically significant differences in median returns between the most sustainable portfolios and groups 2, 3 and 4.

Taking all of the above into consideration, this paper suggests that the sustainability profile of portfolios could have an effect on returns. In a majority of the portfolio comparisons, especially between funds of rating 5 compared to the rest, this paper finds significant results.

### 5.1.2 Equity portfolio returns analysis

The equity and bond portfolios' returns are evaluated in isolation below, to test robustness to a different bond/equity portfolio distribution within sustainability score groups. The Eikon Datastream tags for portfolio type is used.

Below the differences in median annual returns from equity-classified portfolios for the respective sustainability rating is analyzed. The Friedman test rejects the null hypothesis of equal medians with considerable statistical significance for each year. A full overview per year is in the appendix for the interested reader. Tables 10, 11, 12, and 13 show the breakdown of the difference between each fund pair with associated Wilcoxon test results.

Significance levels mapping: * 95%, ** 99%, *** 99.5%.

For the individual equity portfolio tests, the following hypotheses are created. This is similar to the previous section, but now only considers equity portfolio data.

- For each year between 2001 and 2021 and fund group with asset class Equity [5 ... 2], compare returns with another group [4 ... 1], and form a Null Hypotheses $H_0$ : Median(i) - Median(j) = 0, where $i$ and $j$ denotes fund groups such that $i > j$ and an alternative Hypotheses $H_1$ : Median(i) - Median(j) $\neq 0$.

Table 10: Equity portfolios return analysis 2001-2006. Significance levels: * 95%, ** 99%, and *** 99.5%.

| Year | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | 7.91% | 21.88% | -25.17% | -15.05% | -31.49% | -25.89% |
| p-value | 0.057 | 0.008 | 0.001 | 0.003 | 0.001 | 0.001 |
| Statistical significance | | ** | *** | *** | *** | *** |
| Diff medians (5 - 3) | 10.46% | 32.43% | -22.62% | -12.89% | -27.61% | -22.82% |
| p-value | 0.010 | 0.000 | 0.000 | 0.003 | 0.001 | 0.000 |
| Statistical significance | * | *** | *** | *** | *** | *** |
| Diff medians (5 - 2) | 4.49% | 27.78% | -24.20% | -21.68% | -41.81% | -21.38% |
| p-value | 0.213 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |
| Statistical significance | | *** | *** | *** | *** | *** |
| Diff medians (5 - 1) | -0.04% | 9.58% | -24.38% | -33.68% | -41.23% | -23.53% |
| p-value | 0.422 | 0.156 | 0.000 | 0.000 | 0.000 | 0.000 |
| Statistical significance | | | *** | *** | *** | *** |
| Diff medians (4 - 3) | 2.55% | 10.55% | 2.55% | 2.16% | 3.88% | 3.07% |
| p-value | 0.461 | 0.030 | 0.189 | 0.478 | 0.213 | 0.282 |
| Statistical significance | | * | | | | |
| Diff medians (4 - 2) | -3.41% | 5.90% | 0.97% | -6.63% | -10.32% | 4.51% |
| p-value | 0.189 | 0.094 | 0.410 | 0.036 | 0.040 | 0.189 |
| Statistical significance | | | | * | * | |
| Diff medians (4 - 1) | -7.95% | -12.30% | 0.79% | -18.63% | -9.75% | 2.36% |
| p-value | 0.281 | 0.422 | 0.422 | 0.031 | 0.281 | 0.219 |
| Statistical significance | | | | * | | |
| Diff medians (3 - 2) | -5.96% | -4.65% | -1.58% | -8.79% | -14.20% | 1.44% |
| p-value | 0.005 | 0.004 | 0.498 | 0.001 | 0.008 | 0.154 |
| Statistical significance | *** | *** | | *** | ** | |
| Diff medians (3 - 1) | -10.50% | -22.85% | -1.77% | -20.79% | -13.63% | -0.71% |
| p-value | 0.031 | 0.031 | 0.219 | 0.031 | 0.000 | 0.500 |
| Statistical significance | * | * | | * | *** | |
| Diff medians (2 - 1) | -4.53% | -18.20% | -0.18% | -12.00% | 0.57% | -2.15% |
| p-value | 0.156 | 0.078 | 0.500 | 0.078 | 0.344 | 0.281 |
| Statistical significance | | | | | | |

In table 10, results for 2001-2006 is shown. The test gives most significant results in comparisons with the portfolios with a sustainability rating of 5. In 2002, investors earned superior returns from holding the sustainability 5 score portfolios, with a lower return the following years.

Table 11: Equity portfolios return analysis 2007-2012. Significance levels: * 95%, ** 99%, *** 99.5%.

| Year | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | 3.80% | 31.13% | -38.47% | -17.78% | 15.59% | -11.03% |
| p-value | 0.044 | 0.009 | 0.004 | 0.001 | 0.080 | 0.018 |
| Statistical significance | * | ** | *** | *** | | * |
| Diff medians (5 - 3) | 5.03% | 31.99% | -31.55% | -21.25% | 17.04% | -8.43% |
| p-value | 0.023 | 0.002 | 0.001 | 0.000 | 0.009 | 0.007 |
| Statistical significance | * | *** | *** | *** | ** | ** |
| Diff medians (5 - 2) | -10.39% | 39.06% | -51.72% | -17.26% | 19.55% | -11.59% |
| p-value | 0.025 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 |
| Statistical significance | * | *** | *** | *** | *** | *** |
| Diff medians (5 - 1) | -15.86% | 37.77% | -49.45% | -12.68% | 12.76% | -10.18% |
| p-value | 0.219 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Statistical significance | | *** | *** | *** | *** | *** |
| Diff medians (4 - 3) | 1.22% | 0.85% | 6.91% | -3.47% | 1.45% | 2.60% |
| p-value | 0.478 | 0.074 | 0.240 | 0.080 | 0.177 | 0.282 |
| Statistical significance | | | | | | |
| Diff medians (4 - 2) | -14.20% | 7.92% | -13.25% | 0.51% | 3.96% | -0.56% |
| p-value | 0.000 | 0.010 | 0.016 | 0.253 | 0.024 | 0.057 |
| Statistical significance | *** | * | * | | * | |
| Diff medians (4 - 1) | -19.67% | 6.63% | -10.98% | 5.09% | -2.83% | 0.85% |
| p-value | 0.156 | 0.422 | 0.500 | 0.344 | 0.500 | 0.500 |
| Statistical significance | | | | | | |
| Diff medians (3 - 2) | -15.42% | 7.07% | -20.16% | 3.98% | 2.51% | -3.16% |
| p-value | 0.001 | 0.005 | 0.001 | 0.038 | 0.049 | 0.001 |
| Statistical significance | *** | *** | *** | * | * | *** |
| Diff medians (3 - 1) | -20.89% | 5.78% | -17.89% | 8.57% | -4.28% | -1.75% |
| p-value | 0.078 | 0.344 | 0.078 | 0.219 | 0.000 | 0.000 |
| Statistical significance | | | | | *** | *** |
| Diff medians (2 - 1) | -5.47% | -1.29% | 2.27% | 4.58% | -6.78% | 1.41% |
| p-value | 0.078 | 0.047 | 0.344 | 0.500 | 0.031 | 0.109 |
| Statistical significance | | * | | | * | |

In table 11 the same analysis is conducted for 2007-2012. Again, the test shows a more significant difference between the portfolios with a sustainability score of 5 and other portfolio groups. Portfolio group 5 showed statistically significant higher median returns for 2008 and 2011, and lower median returns the rest of the years. The higher returns from more sustainable portfolios can be observed in the comparison between group 3 and 2 in 2008, and again in comparing group 3 with

group 1 in 2011.

Table 12: Equity portfolios return analysis 2013-2018. Significance levels: * 95%, ** 99%, *** 99.5%.

| Year | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | -18.75% | -2.30% | -13.18% | -1.53% | -10.33% | 4.38% |
| p-value | 0.004 | 0.068 | 0.003 | 0.189 | 0.001 | 0.087 |
| Statistical significance | *** | | *** | | *** | |
| Diff medians (5 - 3) | -19.80% | -8.35% | -8.76% | -1.72% | -9.84% | 5.50% |
| p-value | 0.002 | 0.000 | 0.007 | 0.064 | 0.000 | 0.025 |
| Statistical significance | *** | *** | ** | | *** | * |
| Diff medians (5 - 2) | -16.51% | -8.23% | -10.03% | -4.27% | -9.04% | 5.56% |
| p-value | 0.004 | 0.000 | 0.014 | 0.027 | 0.000 | 0.000 |
| Statistical significance | *** | *** | * | * | *** | *** |
| Diff medians (5 - 1) | -6.95% | 0.71% | 6.48% | -15.09% | -9.51% | -2.21% |
| p-value | 0.078 | 0.156 | 0.422 | 0.078 | 0.000 | 0.281 |
| Statistical significance | | | | | *** | |
| Diff medians (4 - 3) | -1.05% | -6.04% | 4.41% | -0.19% | 0.49% | 1.12% |
| p-value | 0.101 | 0.022 | 0.461 | 0.166 | 0.496 | 0.344 |
| Statistical significance | | * | | | | |
| Diff medians (4 - 2) | 2.24% | -5.93% | 3.14% | -2.74% | 1.29% | 1.18% |
| p-value | 0.478 | 0.006 | 0.177 | 0.048 | 0.427 | 0.177 |
| Statistical significance | | ** | | * | | |
| Diff medians (4 - 1) | 11.79% | 3.02% | 19.66% | -13.56% | 0.83% | -6.60% |
| p-value | 0.156 | 0.422 | 0.000 | 0.078 | 0.078 | 0.109 |
| Statistical significance | | | *** | | | |
| Diff medians (3 - 2) | 3.29% | 0.11% | -1.27% | -2.54% | 0.80% | 0.06% |
| p-value | 0.024 | 0.287 | 0.149 | 0.054 | 0.133 | 0.377 |
| Statistical significance | * | | | | | |
| Diff medians (3 - 1) | 12.85% | 9.06% | 15.25% | -13.37% | 0.34% | -7.72% |
| p-value | 0.000 | 0.109 | 0.047 | 0.031 | 0.219 | 0.000 |
| Statistical significance | *** | | * | * | | *** |
| Diff medians (2 - 1) | 9.55% | 8.94% | 16.52% | -10.83% | -0.46% | -7.78% |
| p-value | 0.078 | 0.219 | 0.109 | 0.344 | 0.219 | 0.219 |
| Statistical significance | | | | | | |

For the time period 2013-2019 shown in table 12 , group 5 had lower median returns compared to lower sustainability scoring portfolios in 2013-2015 and 2017.

Table 13: Equity portfolios return analysis 2019-2021. Significance levels: * 95%, ** 99%, *** 99.5%.

| Year | 2019 | 2020 | 2021 |
|---|---|---|---|
| Diff medians (5 - 4) | -20.67% | -11.63% | -22.07% |
| p-value | 0.003 | 0.006 | 0.004 |
| Statistical significance | *** | ** | *** |
| Diff medians (5 - 3) | -25.03% | -9.20% | -25.37% |
| p-value | 0.000 | 0.005 | 0.000 |
| Statistical significance | *** | ** | *** |
| Diff medians (5 - 2) | -19.14% | -9.02% | -24.97% |
| p-value | 0.000 | 0.004 | 0.000 |
| Statistical significance | *** | *** | *** |
| Diff medians (5 - 1) | -29.96% | 2.69% | -17.98% |
| p-value | 0.000 | 0.422 | 0.000 |
| Statistical significance | *** | | *** |
| Diff medians (4 - 3) | -4.36% | 2.43% | -3.30% |
| p-value | 0.177 | 0.427 | 0.062 |
| Statistical significance | | | |
| Diff medians (4 - 2) | 1.53% | 2.61% | -2.90% |
| p-value | 0.166 | 0.427 | 0.253 |
| Statistical significance | | | |
| Diff medians (4 - 1) | -9.29% | 14.32% | 4.09% |
| p-value | 0.078 | 0.078 | 0.109 |
| Statistical significance | | | |
| Diff medians (3 - 2) | 5.89% | 0.18% | 0.40% |
| p-value | 0.038 | 0.133 | 0.073 |
| Statistical significance | * | | |
| Diff medians (3 - 1) | -4.93% | 11.90% | 7.39% |
| p-value | 0.281 | 0.000 | 0.500 |
| Statistical significance | | *** | |
| Diff medians (2 - 1) | -10.82% | 11.72% | 6.99% |
| p-value | 0.078 | 0.109 | 0.500 |
| Statistical significance | | | |

The returns analysis for equity portfolios for 2019-2021 is shown in table 13 . Group 5 has underperformed group 4,3 and 2 on median returns with a 99%+ statistical significance over the whole period.

### 5.1.3 Bond portfolio returns analysis

The same analysis is repeated for the bond portfolios in the dataset, using the same method. The same method for constructing the test hypotheses was used as in the previous section.

- For each year between 2001 and 2021 and fund group with asset class Bond [5 ... 2], compare returns with another group [4 ... 1], and form a Null Hypotheses $H_0$ : Median(i) - Median(j) = 0, where $i$ and $j$ denotes fund groups such that $i > j$ and an alternative Hypotheses $H_1$ : Median(i) - Median(j) $\neq 0$.

Table 14: Bond portfolios return analysis 2001-2006. Significance levels: * 95%, ** 99%, *** 99.5%.

| Year | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | 6.33% | 13.57% | -13.17% | -8.77% | -17.87% | -10.78% |
| p-value | 0.483 | 0.252 | 0.098 | 0.143 | 0.171 | 0.080 |
| Statistical significance | | | | | | |
| Diff medians (5 - 3) | 6.89% | 15.29% | -21.20% | -11.72% | -27.30% | -15.38% |
| p-value | 0.372 | 0.483 | 0.119 | 0.186 | 0.108 | 0.131 |
| Statistical significance | | | | | | |
| Diff medians (5 - 2) | 3.81% | 17.52% | -26.71% | -26.26% | -42.65% | -13.17% |
| p-value | 0.260 | 0.034 | 0.004 | 0.005 | 0.001 | 0.062 |
| Statistical significance | | * | *** | *** | *** | |
| Diff medians (5 - 1) | 6.97% | 32.86% | -28.64% | -37.19% | -47.78% | -14.51% |
| p-value | 0.500 | 0.063 | 0.219 | 0.063 | 0.000 | 0.313 |
| Statistical significance | | | | | *** | |
| Diff medians (4 - 3) | 0.56% | 1.71% | -8.03% | -2.95% | -9.44% | -4.61% |
| p-value | 0.424 | 0.081 | 0.488 | 0.472 | 0.318 | 0.488 |
| Statistical significance | | | | | | |
| Diff medians (4 - 2) | -2.52% | 3.95% | -13.54% | -17.50% | -24.78% | -2.40% |
| p-value | 0.449 | 0.074 | 0.027 | 0.042 | 0.016 | 0.350 |
| Statistical significance | | | * | * | * | |
| Diff medians (4 - 1) | 0.64% | 19.29% | -15.47% | -28.42% | -29.91% | -3.74% |
| p-value | 0.500 | 0.219 | 0.313 | 0.094 | 0.094 | 0.500 |
| Statistical significance | | | | | | |
| Diff medians (3 - 2) | -3.08% | 2.23% | -5.51% | -14.55% | -15.34% | 2.21% |
| p-value | 0.449 | 0.074 | 0.042 | 0.139 | 0.139 | 0.319 |
| Statistical significance | | | * | | | |
| Diff medians (3 - 1) | 0.08% | 17.58% | -7.44% | -25.47% | -20.48% | 0.87% |
| p-value | 0.219 | 0.219 | 0.406 | 0.219 | 0.219 | 0.219 |
| Statistical significance | | | | | | |
| Diff medians (2 - 1) | 3.16% | 15.34% | -1.93% | -10.92% | -5.13% | -1.34% |
| p-value | 0.500 | 0.094 | 0.500 | 0.156 | 0.406 | 0.406 |
| Statistical significance | | | | | | |

In table 14, the conclusion is that there is no major returns difference between bonds sorted by sustainability rating. Portfolio group 2 might be an outlier for 2003-2006, as comparisons with that group specifically yields statistically significant results for the period.

Table 15: Bond portfolios return analysis 2007-2012. Significance levels: * 95%, ** 99%, *** 99.5%.

| Year | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | -0.14% | 20.17% | -24.27% | -8.02% | 15.32% | -4.93% |
| p-value | 0.394 | 0.119 | 0.072 | 0.143 | 0.186 | 0.058 |
| Statistical significance | | | | | | |
| Diff medians (5 - 3) | 3.06% | 33.44% | -34.20% | -11.82% | 13.96% | -1.68% |
| p-value | 0.012 | 0.330 | 0.351 | 0.483 | 0.351 | 0.372 |
| Statistical significance | * | | | | | |
| Diff medians (5 - 2) | -0.58% | 44.37% | -29.27% | -7.92% | 19.51% | -11.27% |
| p-value | 0.449 | 0.034 | 0.103 | 0.449 | 0.233 | 0.016 |
| Statistical significance | | * | | | | * |
| Diff medians (5 - 1) | -9.27% | 49.56% | -96.85% | -12.38% | 12.75% | -14.67% |
| p-value | 0.063 | 0.000 | 0.000 | 0.063 | 0.500 | 0.063 |
| Statistical significance | | *** | *** | | | |
| Diff medians (4 - 3) | 3.20% | 13.28% | -9.93% | -3.80% | -1.36% | 3.24% |
| p-value | 0.055 | 0.170 | 0.304 | 0.202 | 0.496 | 0.019 |
| Statistical significance | | | | | | * |
| Diff medians (4 - 2) | -0.44% | 24.20% | -5.00% | 0.10% | 4.18% | -6.34% |
| p-value | 0.382 | 0.160 | 0.183 | 0.483 | 0.382 | 0.088 |
| Statistical significance | | | | | | |
| Diff medians (4 - 1) | -9.13% | 29.39% | -72.58% | -4.37% | -2.57% | -9.74% |
| p-value | 0.219 | 0.094 | 0.063 | 0.219 | 0.500 | 0.156 |
| Statistical significance | | | | | | |
| Diff medians (3 - 2) | -3.64% | 10.93% | 4.93% | 3.90% | 5.55% | -9.58% |
| p-value | 0.207 | 0.139 | 0.103 | 0.207 | 0.319 | 0.042 |
| Statistical significance | | | | | | * |
| Diff medians (3 - 1) | -12.33% | 16.12% | -62.65% | -0.56% | -1.21% | -12.99% |
| p-value | 0.000 | 0.219 | 0.063 | 0.063 | 0.156 | 0.156 |
| Statistical significance | *** | | | | | |
| Diff medians (2 - 1) | -8.68% | 5.19% | -67.58% | -4.46% | -6.76% | -3.40% |
| p-value | 0.000 | 0.219 | 0.063 | 0.094 | 0.406 | 0.219 |
| Statistical significance | *** | | | | | |

In table 15, results are inconclusive overall. Group 1's median return for 2008-2009 differs materially from the other groups, but the low number of portfolios in group 1 means the null hypothesis can not be rejected for more than the comparison with group 5.

Table 16: Bond portfolios return analysis 2013-2019. Significance levels: * 95%, ** 99%, *** 99.5%.

| Year | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | -14.27% | -2.80% | -3.42% | -2.06% | -6.78% | 1.43% |
| p-value | 0.098 | 0.330 | 0.201 | 0.290 | 0.217 | 0.046 |
| Statistical significance | | | | | | * |
| Diff medians (5 - 3) | -23.53% | -5.82% | -6.23% | -1.20% | -4.21% | 2.59% |
| p-value | 0.108 | 0.016 | 0.131 | 0.416 | 0.351 | 0.131 |
| Statistical significance | | * | | | | |
| Diff medians (5 - 2) | -7.32% | -0.93% | 1.79% | -1.02% | -2.92% | 0.87% |
| p-value | 0.103 | 0.103 | 0.382 | 0.416 | 0.382 | 0.183 |
| Statistical significance | | | | | | |
| Diff medians (5 - 1) | -10.66% | 6.83% | -2.53% | -22.12% | -2.92% | -1.20% |
| p-value | 0.500 | 0.313 | 0.500 | 0.000 | 0.219 | 0.500 |
| Statistical significance | | | | *** | | |
| Diff medians (4 - 3) | -9.26% | -3.02% | -2.81% | 0.87% | 2.57% | 1.16% |
| p-value | 0.347 | 0.009 | 0.263 | 0.488 | 0.250 | 0.087 |
| Statistical significance | | ** | | | | |
| Diff medians (4 - 2) | 6.95% | 1.86% | 5.22% | 1.04% | 3.86% | -0.56% |
| p-value | 0.483 | 0.016 | 0.289 | 0.517 | 0.382 | 0.449 |
| Statistical significance | | * | | | | |
| Diff medians (4 - 1) | 3.61% | 9.62% | 0.90% | -20.05% | 3.85% | -2.62% |
| p-value | 0.219 | 0.156 | 0.500 | 0.094 | 0.313 | 0.156 |
| Statistical significance | | | | | | |
| Diff medians (3 - 2) | 16.20% | 4.88% | 8.02% | 0.17% | 1.29% | -1.72% |
| p-value | 0.139 | 0.350 | 0.207 | 0.319 | 0.233 | 0.103 |
| Statistical significance | | | | | | |
| Diff medians (3 - 1) | 12.86% | 12.64% | 3.70% | -20.92% | 1.29% | -3.79% |
| p-value | 0.156 | 0.063 | 0.156 | 0.000 | 0.313 | 0.000 |
| Statistical significance | | | | *** | | *** |
| Diff medians (2 - 1) | -3.34% | 7.76% | -4.32% | -21.09% | 0.00% | -2.06% |
| p-value | 0.313 | 0.063 | 0.406 | 0.000 | 0.313 | 0.094 |
| Statistical significance | | | | *** | | |

It is difficult to draw any general conclusions from table 16.

Table 17: Bond portfolios return analysis 2019-2021. Significance levels: * 95%, ** 99%, *** 99.5%.

| Year | 2019 | 2020 | 2021 |
|---|---|---|---|
| Diff medians (5 - 4) | -13.13% | -3.49% | -15.38% |
| p-value | 0.108 | 0.028 | 0.157 |
| Statistical significance | | * | |
| Diff medians (5 - 3) | -16.90% | -4.28% | -12.12% |
| p-value | 0.186 | 0.064 | 0.252 |
| Statistical significance | | | |
| Diff medians (5 - 2) | -13.98% | 5.02% | -7.93% |
| p-value | 0.062 | 0.416 | 0.160 |
| Statistical significance | | | |
| Diff medians (5 - 1) | -22.30% | 0.75% | -20.05% |
| p-value | 0.063 | 0.500 | 0.094 |
| Statistical significance | | | |
| Diff medians (4 - 3) | -3.77% | -0.78% | 3.26% |
| p-value | 0.440 | 0.472 | 0.170 |
| Statistical significance | | | |
| Diff medians (4 - 2) | -0.85% | 8.51% | 7.45% |
| p-value | 0.483 | 0.074 | 0.289 |
| Statistical significance | | | |
| Diff medians (4 - 1) | -9.17% | 4.24% | -4.67% |
| p-value | 0.313 | 0.406 | 0.406 |
| Statistical significance | | | |
| Diff medians (3 - 2) | 2.92% | 9.30% | 4.20% |
| p-value | 0.183 | 0.160 | 0.103 |
| Statistical significance | | | |
| Diff medians (3 - 1) | -5.40% | 5.03% | -7.92% |
| p-value | 0.313 | 0.406 | 0.313 |
| Statistical significance | | | |
| Diff medians (2 - 1) | -8.32% | -4.27% | -12.12% |
| p-value | 0.219 | 0.406 | 0.094 |
| Statistical significance | | | |

For the last three years, there is no statistically significant differences in bond portfolios depending on sustainability rating, with one exception. This is despite the fact that sustainability ratings are likely most accurate for this period given the temporal proximity.

## 5.2 Risk

In this section, the risk aspect of ESG investing will be evaluated. As described in the methodology section, downside deviation is the chosen measure of risk. The downside deviation is calculated using formula 1. Like in the previous section on returns, Wilcoxon non-parametric median testing is conducted due to the non-normality in the data as shown in QQ plots in the Normality analysis of returns section in the Appendix.

Initially, annual downside deviation is considered. By testing the difference between each group with every other group the following hypotheses were generated.

- Null Hypotheses $H_{0...9}$ : Median(i) - Median(j) = 0 $\{\forall i, j \in [5, 1] : i > j\}$ Where $i$ and $j$ represent annual downside deviation of a given group of funds.
- Alternative Hypotheses $H_{10...19}$ : Median(i) - Median(j) $\neq 0$.

Similarly to the previous section on returns, groupings of funds are denoted according to their respective sustainability ratings as $1 \ldots 5$.

Table 18: Wilcoxon annual downside deviation test results. Significance levels: * 95%, ** 99%, *** 99.5%.

| Groups | Diff medians downside deviation | p-value | Significance level |
|---|---|---|---|
| 5 - 4 | -16.51% | 0.003 | *** |
| 5 - 3 | -17.05% | 0.000 | *** |
| 5 - 2 | -17.95% | 0.000 | *** |
| 5 - 1 | -18.20% | 0.062 | |
| 4 - 3 | -0.55% | 0.082 | |
| 4 - 2 | -1.45% | 0.003 | *** |
| 4 - 1 | -1.69% | 0.382 | |
| 3 - 2 | -0.90% | 0.106 | |
| 3 - 1 | -1.15% | 0.350 | |
| 2 - 1 | -0.24% | 0.207 | |

The null hypothesis can not be rejected for any comparison to the portfolios in group 1 as seen in table 18, this is likely due to the limited annual data sample for funds of sustainability rating 1. The test results in a statistically significant reduction in downside deviation for the most sustainable portfolios compared to portfolios 4, 3, and 2 however. The results imply the rejection of $H_0$, $H_1$, $H_2$ and $H_5$.

For the next analysis, the weekly downside deviation between 2019 and 2022 is considered, as outlined in section 4 on Data. The number of observations is higher

which gives stronger results. Similar to the previous analysis, the hypotheses can be formulated as follows:

- Null Hypotheses $H_{0...9}$ : Median(i) - Median(j) = 0 $\{\forall i, j \in [5, 1] : i > j\}$ Where $i$ and $j$ represent weekly downside deviation of a given group of funds.
- Alternative Hypotheses $H_{10...19}$ : Median(i) - Median(j) $\neq 0$.

Table 19: Wilcoxon weekly downside deviation test results. Significance levels: * 95%, ** 99%, *** 99.5%.

| **Groups** | **Diff medians downside deviation** | **p-value** | **Significance level** |
|---|---|---|---|
| 5 - 4 | -0.28% | 0.010 | ** |
| 5 - 3 | -2.12% | 0.000 | *** |
| 5 - 2 | -2.31% | 0.000 | *** |
| 5 - 1 | -2.54% | 0.000 | *** |
| 4 - 3 | -1.84% | 0.011 | * |
| 4 - 2 | -2.03% | 0.000 | *** |
| 4 - 1 | -2.26% | 0.000 | *** |
| 3 - 2 | -0.19% | 0.015 | * |
| 3 - 1 | -0.42% | 0.000 | *** |
| 2 - 1 | -0.23% | 0.009 | ** |

The larger number of portfolios in the weekly time series makes the results more significant overall, as seen in table 19. The results indicate statistically significant results for all observations with at least a 95% confidence level, thus rejecting hypotheses $H_{0...9}$. Notably, the implication of this is that there is a statistically significant lower risk measured by downside deviation for all more sustainable portfolios relative to less sustainable portfolios.

In table 20 below, the same tests as previously are run on the weekly data for the equity and bond portfolios separately to test the robustness of the results. There are no bond portfolios in the Nordic data set with a sustainability rating of 1, which makes comparison with that group of portfolios impossible. Hypotheses formed are:

- Null Hypotheses $H_{0...9}$ : Median(i) - Median(j) = 0 $\{\forall i, j \in [5, 1] : i > j\}$ Where $i$ and $j$ represents the weekly downside deviation of a given group of portfolios in the "Bond" category.
- Alternative Hypotheses $H_{10...19}$ : Median(i) - Median(j) $\neq 0$.

Table 20: Wilcoxon test for weekly downside deviation, bond portfolios. Significance levels: * 95%, ** 99%, *** 99.5%.

| **Groups** | **Diff medians downside deviation** | **p-value** | **Significance level** |
|---|---|---|---|
| 5 - 4 | -0.03% | 0.365 | |
| 5 - 3 | -0.01% | 0.024 | * |
| 5 - 2 | -0.48% | 0.086 | |
| 5 - 1 | n.m. | N/A | N/A |
| 4 - 3 | 0.03% | 0.115 | |
| 4 - 2 | -0.44% | 0.185 | |
| 4 - 1 | n.m. | N/A | N/A |
| 3 - 2 | -0.47% | 0.005 | *** |
| 3 - 1 | n.m. | N/A | N/A |
| 2 - 1 | n.m. | N/A | N/A |

The statistical power diminishes significantly when separating the sustainability groups by portfolio class as seen in table 20. For bond portfolios, only the null hypotheses $H_1$ and $H_7$ can be rejected at a respective 95% and 99% confidence level. The difference in median downside deviation is also significantly lower at less than 0.5 percentage points for each respective comparison.

An analysis of equity portfolios is conducted in a similar manner to the bond analysis.

- Null Hypotheses $H_{0...9}$ : Median(i) - Median(j) = 0 $\{\forall i, j \in [5, 1] : i > j\}$ Where $i$ and $j$ represents the weekly downside deviation of a given group of portfolios in the "Equity" category.
- Alternative Hypotheses $H_{10...19}$ : Median(i) - Median(j) $\neq 0$.

Table 21: Wilcoxon test weekly downside deviation, equity portfolios. Significance levels: * 95%, ** 99%, *** 99.5%.

| Groups | Diff medians downside deviation | p-value | Significance level |
|---|---|---|---|
| 5 - 4 | -0.13% | 0.082 | |
| 5 - 3 | -0.20% | 0.001 | *** |
| 5 - 2 | -0.31% | 0.001 | *** |
| 5 - 1 | -0.49% | 0.000 | *** |
| 4 - 3 | -0.07% | 0.025 | * |
| 4 - 2 | -0.19% | 0.035 | * |
| 4 - 1 | -0.36% | 0.000 | *** |
| 3 - 2 | -0.12% | 0.158 | |
| 3 - 1 | -0.29% | 0.004 | *** |
| 2 - 1 | -0.17% | 0.003 | *** |

As opposed to the bond portfolio analysis, hypothesis testing suggests a statistically significant lower downside deviation for the more sustainable equity portfolios compared to less sustainable equity portfolios for all but two tests. Findings support the previous literature in suggesting that increased sustainability can be linked to lower risk in investments [9] [10] [11] [12].

# 6 Conclusion

For the Nordic mutual fund and ETF universe, this paper suggests that portfolios with higher sustainability scores as measured by MSR carry lower median returns. The effect is especially significant and consistent across time in comparisons with sustainability group 5. Separating Equity and Bond portfolios shows that the effect is derived from a clear effect in equity portfolios, with limited statistically significant results for the bond portfolios across the annual time period. Hypothesis testing suggests that the delta in median returns for more sustainable portfolios is positive for years 2002, 2008, 2011, and negative for the other years tested. While the full year 2002 was spanned by a NBER recession, the NBER recession of 2008-2010 did not start until December 2008 [24]. The results of previous studies, that more sustainable funds achieve higher returns in bear markets, is therefore only partially aligned with this thesis' findings.

This paper suggests that portfolios scoring higher in sustainability metrics carry lower downside deviation risk. Results are more significant when comparing equity portfolios separately, and the higher number of portfolios in the weekly dataset means the statistical strength of hypothesis testing increases for these tests. For

the annual downside deviation testing, only the highest MSR score group median downside deviations can be assumed lower than the lowest sustainability scoring groups.

There are large differences in equity/bond composition within the MSR groups. Controlling for the differences in asset class distribution within the sustainability groups by separating them diminished the strength of findings significantly. There is an underweight of lower sustainability scores for the Nordic portfolio data sample and although Wilcoxon and Friedman tests are agnostic to differences in sample sizes, the low number of MSR 1 scoring portfolios does affect the statistical strength of hypothesis testing in comparisons to this group.

This paper does not take into account the possibility that portfolios could have moved between MSR scores over time, a possibility as scores are reevaluated regularly. Moreover, it goes beyond the scope for this study to account for attempts of greenwashing, the sustainability ratings published by Morningstar are regarded as accurate. A more extensive study could compare the results using a different sustainability score provider.

# References


[1] Bryce Klempner Sara Bernow and Clarisse Magnin. *From why' to 'why not': Sustainable investing as the new normal*. URL: https://www.mckinsey.com/industries/private-equity-and-principal-investors/our-insights/from-why-to-why-not-sustainable-investing-as-the-new-normal.

[2] Principles for Responsible Investment. *Principles for Responsible Investment*. URL: https://www.unpri.org/.

[3] Emiel Duuren, Auke Plantinga, and Bert Scholtens. "ESG Integration and the Investment Management Process: Fundamental Investing Reinvented". In: *Journal of Business Ethics* 138.3 (Oct. 2016), pp. 525–533. DOI: 10.1007/s10551-015-2610-8. URL: https://ideas.repec.org/a/kap/jbuset/v138y2016i3d10.1007_s10551-015-2610-8.html.

[4] UN Environment. *Emissions gap report 2022*. URL: https://www.unep.org/resources/emissions-gap-report-2022.

[5] Linnea Dahlberg and Frida Wiklund. "ESG Investing In Nordic Countries : An analysis of the Shareholder view of creating value". In: 2018.

[6] Gunnar Friede, Timo Busch, and Alexander Bassen. "ESG and financial performance: aggregated evidence from more than 2000 empirical studies". In: *Journal of Sustainable Finance & Investment* 5.4 (2015), pp. 210–233. DOI: 10.1080/20430795.2015.1118917. eprint: https://doi.org/10.1080/20430795.2015.1118917. URL: https://doi.org/10.1080/20430795.2015.1118917.

[7] Andreas Hoepner et al. "ESG Shareholder Engagement and Downside Risk". In: (Jan. 2020). DOI: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2874252.

[8] Emirhan Ilhan, Zacharias Sautner, and Grigory Vilkov. "Carbon Tail Risk". In: *Review of Financial Studies* 34.3 (2021), pp. 1540–1571. URL: https://EconPapers.repec.org/RePEc:oup:rfinst:v:34:y:2021:i:3:p:1540-1571..

[9] Institute for Sustainable Investing Morgan Stanley. "Sustainable Reality: Analyzing Risk and Returns of Sustainable Funds". In: (2019). URL: https://www.morganstanley.com/content/dam/msdotcom/ideas/sustainable-investing-offers-financial-performance-lowered-risk/Sustainable_Reality_Analyzing_Risk_and_Returns_of_Sustainable_Funds.pdf.

[10] Rui Albuquerque et al. "Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash". In: *The Review of Corporate Finance Studies* 9.3 (July 2020), pp. 593–621. ISSN: 2046-9128. DOI: 10.1093/rcfs/cfaa011. eprint: https://academic.oup.com/rcfs/article-pdf/9/3/593/33880102/cfaa011.pdf. URL: https://doi.org/10.1093/rcfs/cfaa011.

[11] David C. Broadstock et al. "The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China". In: *Finance Research Letters* 38.C (2021). DOI: 10.1016/j.frl.2020.101716. URL: https://ideas.repec.org/a/eee/finlet/v38y2021ics1544612320309983.html.

[12] K V Lins, H Servaes, and A Tamayo. "Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis". In: *Journal of Finance* 72.4 (Aug. 2017). This is the peer reviewed version of the following article: Lins KV, Servaes H & Tamayo A (2017) Social capital, trust, and firm performance: the value of corporate social responsibility during the financial crisis, Journal of Finance, Vol. 72 (4), pp 1785-1824 which has been published in final form at https://doi.org/10.1111/jofi.12505. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving. © 2017 The American Finance Association, pp. 1785–1824. URL: https://lbsresearch.london.edu/id/eprint/579/.

[13] Magali A. Delmas and Vanessa Cuerel Burbano. "The Drivers of Greenwashing". In: *California Management Review* 54.1 (2011), pp. 64–87. DOI: 10.1525/cmr.2011.54.1.64. eprint: https://doi.org/10.1525/cmr.2011.54.1.64. URL: https://doi.org/10.1525/cmr.2011.54.1.64.

[14] Financial Times. "Financial Times, Morningstar cuts 1,200 funds from 'sustainable' list". In: (). URL: https://www.ft.com/content/9cf8c788-6cad-4737-bc2a-2e85ac4aef7d.

[15] Aaron K. Chatterji et al. "Do ratings of firms converge? Implications for managers, investors and strategy researchers". In: *Strategic Management Journal* 37.8 (2016), pp. 1597–1614. DOI: https://doi.org/10.1002/smj.2407. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.2407. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.2407.

[16] Gregor Dorfleitner, Gerhard Halbritter, and Mai Nguyen. "Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches". In: *Journal of Asset Management* 16 (2015), pp. 450–466.

[17] Elaine Conway. "To Agree or Disagree? An Analysis of CSR Ratings Firms". In: *Social and Environmental Accountability Journal* 39 (2019), pp. 152–177.

[18] Nicolas P. B. Bollen. "Mutual Fund Attributes and Investor Behavior". In: *The Journal of Financial and Quantitative Analysis* 42.3 (2007), pp. 683–708. ISSN: 00221090, 17566916. URL: http://www.jstor.org/stable/27647316 (visited on 11/28/2022).

[19] Morningstar. *Morningstar Sustainble Investing Handbook*. URL: https://s21.q4cdn.com/198919461/files/doc_downloads/press_kits/2016/Morningstar-Sustainable-Investing-Handbook.pdf.

[20] WorldBank. *Putting a Price on Carbon with an ETS*. URL: https://www.worldbank.org/content/dam/Worldbank/document/Climate/background-note_ets.pdf.

[21] Frank Wilcoxon. "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83. ISSN: 00994987. URL: http://www.jstor.org/stable/3001968 (visited on 11/19/2022).

[22] Milton Friedman. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance". In: *Journal of the American Statistical Association* 32.200 (1937), pp. 675–701. DOI: 10.1080/01621459.1937.10503522. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1937.10503522. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1937.10503522.

[23] Riksbanken. *The financial crisis 2007-2009*. URL: https://www.riksbank.se/en-gb/financial-stability/the-riksbanks-financial-stability-tasks/preparedness-to-manage-a-financial-crisis/the-financial-crisis-2007-2009/.

[24] NBER. *US Business Cycle Expansions and Contractions*. URL: https://www.nber.org/research/business-cycle-dating.

# 7 Appendix

## 7.1 Normality analysis of returns

Figure 11: QQ plots of median of returns from funds of sustainability rating 1 and 2

Figure 12: QQ plots of median of returns from funds of sustainability rating 3 and 4

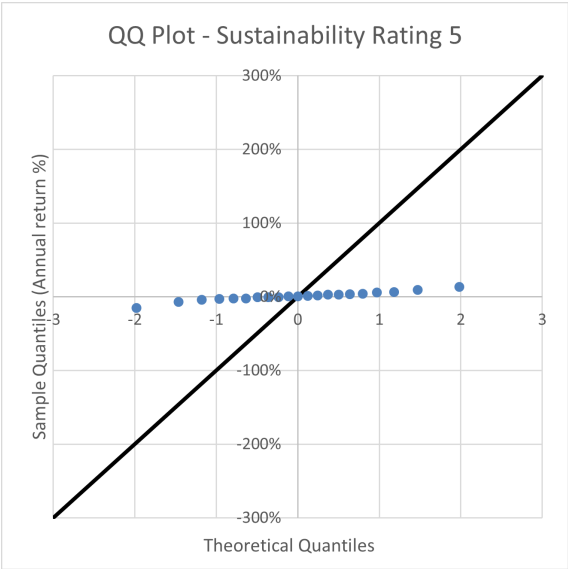


Figure 13: QQ plot of median of returns from funds of sustainability rating 5

## 7.2 Statistical Analysis of returns

Table 22: Full Wilcoxon Analysis for all fund grouping pairs

| Year | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Diff medians (5 - 4) | 7% | 11% | -30% | -23% | -35% | -33% | -3% | 41% | -55% | -37% | 7% | -21% | -21% | -9% | -17% | -9% | -12% | 0% | -28% | -19% | -27% |
| Statistical significance | 95%+ | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | | 99%+ | 99%+ | 99%+ | | 99%+ | 95%+ | | 99%+ | | 99%+ | 99%+ | 99%+ | 99%+ | 95%+ |
| Diff medians (5 - 3) | 11% | 22% | -32% | -23% | -33% | -31% | 0% | 42% | -58% | -40% | 9% | -21% | -28% | -13% | -18% | -11% | -13% | 2% | -31% | -21% | -32% |
| Statistical significance | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 95%+ | 99%+ | 99%+ | 99%+ | 99%+ | | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ |
| Diff medians (5 - 2) | 5% | 18% | -35% | -34% | -48% | -33% | -14% | 51% | -81% | -38% | 12% | -25% | -17% | -11% | -10% | -12% | -13% | 1% | -24% | -17% | -32% |
| Statistical significance | | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 95%+ | 99%+ | 99%+ | 99%+ | 99%+ | 99%+ | 95%+ | 99%+ | | 99%+ | 99%+ | 99%+ | 99%+ | | 99%+ |
| Diff medians (5 - 1) | 5% | 20% | -36% | -46% | -52% | -30% | -17% | 55% | -106% | -34% | 4% | -25% | -15% | -3% | -7% | -28% | -13% | -4% | -38% | -11% | -25% |
| Statistical significance | | | 95%+ | 99%+ | 99%+ | 95%+ | | 95%+ | 99%+ | 99%+ | | 99%+ | | | | 99%+ | | | 95%+ | | |
| Diff medians (4- 3) | 4% | 14% | -39% | -30% | -40% | -39% | -7% | 35% | -65% | -47% | 2% | -29% | -35% | -20% | -26% | -18% | -20% | -5% | -38% | -29% | -39% |
| Statistical significance | | | | | | | | | | | | | | 95%+ | | | | | | | |
| Diff medians (4- 2) | -3% | 11% | -42% | -42% | -55% | -41% | -21% | 44% | -88% | -45% | 4% | -33% | -25% | -18% | -17% | -20% | -20% | -6% | -32% | -25% | -40% |
| Statistical significance | | 95%+ | 95%+ | 99%+ | 99%+ | | 99%+ | 99%+ | 99%+ | | 95%+ | 99%+ | | 99%+ | | 95%+ | | 95%+ | | | 95%+ |
| Diff medians (4- 1) | -3% | 13% | -43% | -53% | -60% | -37% | -24% | 47% | -113% | -41% | -3% | -32% | -22% | -11% | -14% | -35% | -20% | -12% | -45% | -18% | -32% |
| Statistical significance | | | | 95%+ | | | | | 99%+ | | | | | | 95%+ | 95%+ | | 99%+ | | 95%+ | |
| Diff medians (3 - 2) | -6% | 7% | -46% | -45% | -59% | -44% | -25% | 40% | -92% | -49% | 0% | -37% | -29% | -22% | -21% | -24% | -24% | -10% | -36% | -29% | -43% |
| Statistical significance | | | | 99%+ | 99%+ | | 99%+ | 95%+ | 99%+ | | 95%+ | 99%+ | | | 95%+ | 95%+ | | | | | |
| Diff medians (3 - 1) | -6% | 9% | -47% | -57% | -63% | -41% | -28% | 44% | -117% | -45% | -7% | -36% | -26% | -14% | -18% | -39% | -24% | -16% | -49% | -22% | -36% |
| Statistical significance | 99%+ | | | 95%+ | 95%+ | | 99%+ | | 99%+ | | 95%+ | 99%+ | 99%+ | 95%+ | 95%+ | 99%+ | | 99%+ | | 95%+ | |
| Diff medians (2 - 1) | 0% | 2% | -1% | -12% | -5% | 3% | -3% | 3% | -25% | 4% | -8% | 0% | 3% | 8% | 4% | -15% | 0% | -6% | -13% | 6% | 8% |
| Statistical significance | | | | 95%+ | | | | | | | | | | 95%+ | | 99%+ | | | 95%+ | | |

## 7.3 Friedman test results

Table 23: Friedman Test results annual Equity Returns

| **Year** | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| **p value** | 3.8E-05 | 1.8E-05 | 9.90E-13 | 3.3E-07 | 1.3E-08 | 4.0E-07 |

| **Year** | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|
| **p value** | 5.2E-08 | 1.0E-06 | 3.7E-09 | 9.4E-11 | 1.8E-04 | 1.3E-07 |

| **Year** | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|
| **p value** | 1.4E-10 | 6.1E-11 | 4.7E-10 | 2.2E-05 | 2.0E-07 | 1.1E-03 |

| **Year** | 2019 | 2020 | 2021 |
|---|---|---|---|
| **p value** | 3.1E-07 | 4.4E-09 | 8.4E-10 |

Table 24: Friedman Test results annual Bond Returns

| **Year** | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| **p value** | 7.8E-03 | 2.5E-06 | 1.5E-02 | 3.3E-02 | 1.8E-02 | 6.8E-04 |

| **Year** | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|
| **p value** | 5.0E-02 | 8.1E-06 | 6.8E-02 | 4.7E-02 | 1.1E-02 | 7.1E-03 |

| **Year** | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|
| **p value** | 3.8E-04 | 1.0E-05 | 2.2E-04 | 5.0E-02 | 8.8E-04 | 2.0E-03 |

| **Year** | 2019 | 2020 | 2021 |
|---|---|---|---|
| **p value** | 1.5E-02 | 1.6E-05 | 1.5E-02 |