Greenwashing: Testing for portfolios' carbon footprint differences

Research Master's Thesis submitted by:
MURCILLO PETERSSON Luis Alejandro

Louvain School of Management
and
Aix-Marseille University

Master 120 ECTS in Finance with a major in Financial Risk Management
Master 120 ECTS in Management with a major in International Finance

Supervisor: HASSE Jean-Baptiste
20/07/2022

Acknowledgements

I am extremely grateful to Dr. Jean-Baptiste Hasse for agreeing to become my master's thesis supervisor, this research would not have been possible without him. His knowledge about the topic and enthusiasm of sharing the information have been invaluable and a source of inspiration for me these last months. I would also like to thank him for his guidance, support and above all his patience in every step of this project.

Contents

Chapter 1

Preface

Over the past decade, climate change has become one of the main talked-about topics on the world. From forums on the internet, TV broadcasts to elections, countries spend more resources trying to reduce their impact on the environment. For instance, the EU aims to become carbon neutral by 2050, whether it fails it is yet to be seen. On the other side of the Atlantic, president J. Biden pledges to reduce heavily the carbon footprint of the US. That two of the main economic powers of the world are working against climate change exemplifies the importance of the subject in domestic and foreign policy.

Moreover, people are increasingly interested not only on climate change but also on a tremendous account of social issues. Whether it is fighting for women's rights, gun policy or race related issues, communities now take a big part in social discussions. People are more invested and preoccupied nowadays. This culture has evidently permeated the financial sector to the point where an alternative way of investing has risen sharply. It is called Socially Responsible Investment (SRI) and its origins may date back to 1758. It is a investing strategy that aims to produce a social change but also a financial return for an investor.For example, this type of investment took an important role taking down the apartheid government in South Africa. Now, it can include companies that have a positive sustainable or social impact, such as a solar energy company. This type of investment strategy has been increasing with a steady pace. According to a 2019 survey, 85% of individual investors are interested on sustainable investing.Furthermore, according to Morningstar, there were 303 sustainable open-ended mutual funds in 2019, from 111 in 2014. This shows how important SRI has become in the current financial climate.

However, the main debate around SRI, arguing that there was no difference in performance between "conventional" and SRI has not given a clear consensus. Multiple researchers have indeed inquired about the performance of both alternatives and have compared them in multiple of ways. In this search, one point has become crucial. Some cases have been studied where asset managers make unsubstantiated or misleading claims about the commitment of their funds either to the environment or social governance. Evidently, these asset managers want to profit from this growing interest in socially responsible investing by misleading the investors. The practice has now been called as *greenwashing*. Its is defined generally as a marketing or advertising technique in which PR and green marketing are deceptively used to persuade customers. In this precise case, investors. This practice makes harder for individual investors to discern between really committed SRI and dubious ones.

The task is set for some researchers: find a way to distinguish the real SRI funds and "greenwashed" ones. Several researchers have been studying this and have come up with ways to distinguish between the two using econometric tools. Some have suggested that "greenwashing" occurs because of lack of regulation, hence making tougher laws in this sector of the economy would help to solve the problem. For instance, France is the European pioneer in regulations of this kind and a model country in this respect for others to follow.

Chapter 2

Introduction

In this dissertation, we are going to test the difference in terms of exposure to carbon emissions between two portfolios, allowing us in this way to identify greenwashing activities in the Financial Markets. In order to do this, we worked with a plethora of statistical and econometric tools that we used to compare the MSCI World and MSCI World Low Carbon index over one year period.

As the awareness over climate change due to the aggravation of environmental problems has grown in recent years, has led to the emergence of Socially Responsible Investment(SRI). This refers to an investment approach which takes into consideration the extra-financial performance of a company and not only the financial performance as was the case previously. At SRI, investors seek to align their values with their investments, trying not only to make a financial gain but also to make a positive impact for the planet. Nowadays, around 30% of global Assets under Management (AuM) are SRI and this number is expected to keep growing in the coming years.

The increased interest for Socially Responsible investing has led some actors to seize an opportunity to increase their profits improperly, mainly through marketing campaigns in which they provide misleading information or false ecological commitments in order to create a socially responsible image to the public, this phenomenon in called greenwashing.

Greenwashing has seen an important increase to a large extent because of the lack of regulations. However, things are changing recently. In 2021, France became a pioneer on the work of regulation reaching agreement on the "Climate and Resilience Law" which aims to contribute to the acceleration of ecological and social transition through strengthen the extra-financial transparency framework for financial players. More precisely, it aims to reach 40% greenhouse gas emissions reduction by 2030 compared to 1990 levels. France is now an example for the European Union. It has come up with a more ambitious proposal, the "European Climate Law", which seek to reduce greenhouse gas emissions by 55% by 2030 and reach climate neutrality by 2050.

Despite the efforts that have been made, in many cases investors keep having trouble identifying real SRI from "greenwashed" companies, funds or indexes. This encouraged us to create an econometric approach that an investor alone could use in an easy way and would help him to make a educated choice.

Consequently, the main goal of this work is to **identify greenwashing using a sensible and well-crafted econometric approach to spot it**. Moreover, we propose a series of statistical and econometric tools, for instance, independent two sample t-Test, Welch's-Test, Wilcoxon Mann-Whitney Test and the Overlapping Coefficient Index, allowing to compare Conventional and Socially Responsible Index to find out whether or not there was a significant difference in terms of exposure to carbon emissions.

This dissertation is structured as follows. Section 3 provides a literature review, in which the financial and econometric background relevant to the issue is reviewed. Section 4 exhibit the data choice and explain in detail the methodology that will be used in this analysis. In Section 5, we will discuss our results and main findings using RStudio. Lastly, section 6 provides a conclusion and mention either the limitations of our work and our recommendations.

Chapter 3

Literature review

In the wake of climate change and social instabilities around the world, several investors want to contribute to a positive social and environmental change being more socially responsible with their investments. Nevertheless, with the rise of greenwashing making the right choice might be a difficult challenge. So, how to identify greenwashing? To answer this question one must do a thorough analysis between the socially responsible investments and those who are not to determine if there is a substantial difference between them.

This chapter will be organized as follows: in the first section we will explain and discuss the most important aspects of the comparisons between SRI and common investments. We will present some important literature on the comparison of these two types of investments. The second section will be dedicated to explaining the history of our mathematical method we used to make this comparison.

3.1 Comparing investments: SRI and conventional

Socially Responsible Investments have a long history. However, recent social unrest and the environmental discussion, among many other socio-political themes, have inserted special social awareness in investors thus increasing the demand for socially responsible investments. Its importance in the market has risen accordingly. Hence the need for a closer look at the performance of these type of investments.

We have to note that this is not a recent investigation. Several other researchers have enquired about the performance of these two types of investments and their comparison. Our goal here is to present what has been done in the past and the conclusions researchers have come up with when doing their research.

We will present each contribution with no special order since the majority of these papers differ mainly on their methodology but not on the main question: Is there a difference in performance between SRI and conventional investments?

In 1993, S. Hamilton *et al.* investigated this question. While the investigation does not question the legitimacy of SRI, they asked explicitly what are the actual relative returns of socially responsible mutual funds and conventional funds and also tested three main hypotheses about the investments.[12] Using Jensen's alpha method, they analysized monthly returns of all equity mutual funds from 1981 to 1990. Their findings indicate that investors can expect to **lose nothing** by investing in SR mutual funds. On the same line, R. Bauer took interest on this problem and began enquiring in his 2005 paper.[3]Years before, Hamilton *et al.*(1993) and Statman(2000) both found no significant differences between the performances of these two types of investments using S&P 500 and DSI as data. Now, they extended the current research about the subject and used a multi-factor model called the Carhart model, suggested by other research, in order to analyse the performance of both SRI and common investments. In short terms, the Carhart model extends the Fama-French case of study to account for momentum anomaly. In this model, he took account for the size, book to market and momentum effects in order to control biasing influences in performance methods used before. His study was based on the performance of 103 US, UK and German mutual funds while characterizing in detail the investments styles of ethical mutual funds on the countries of study. The results were interesting not only because this approach improves the measurement of performance since it allows a more detailed analysis but also they found no evidence of statistically significant difference in return between ethical and common mutual fund returns.

Building upon the previous works, M.C. Cortez *et al.*,[11] expanded the study to the global market. In their paper, they wanted to investigate the performance of SR funds that invest **globally** using a multi-factor model that control for size and book-to-market. In that way they analysed investment styles and possible home bias. For that, they worked with single-index and multi-index models using among others, the MSCI Index as benchmark for both the SR and conventional fund in the US and European market. They found that in most European markets, the performance is comparable between conventional and SR mutual funds. In other words, the results show neutral performance compared with both SR and conventional benchamrks. Nevertheless, for the US and Austria, they results show an underperformance.

The findings sit under robust evidence and scrutiny to whatever performance model they used.

Contrarily to most of the literature and for instance, R. Bauer and S. Hamilton, Y. Belghitar also studied the comparison of performance of SRI and conventional ones. In his paper,[13] he re-examines the problem using a special method called the Marginal Conditional Stochastic Dominance (MCSD) that allows for any return distribution. This is interesting because the approximation of normally distributed return functions is not well-founded as Mandelbrot pointed out in 1963. Most importantly, he points out that the mean-variance method for measuring performance neglects potentially important information. He argues that a study limited to the first two moments of equity return should also include higher order moments such as kurtosis and skewness. In order to take that into account he uses MCSD and applies it to an analysis using the FTSE4Good Index Series to estimate investment performance. He compares four socially responsible indices and conventional indices. He foun a surprising result. Even though there is nothing to be gained nor lost from socially responsible investing in terms of mean-variance, his new study points out that there is a price to be paid in higher moments of the return distributions. In this way, investors can improve their expected utility by reducing holding more on conventional firms than rather on socially responsible ones.

Similarly, the question on how to classify SR mutual funds is of great interest. M. Statman studied how to classify and measure the performance of socially responsible mutual funds. He found some shortcomings on the methods used by Morningstar and others usually applying Sharpe's factor to classify the investment performance. He points out that different methods of classifying socially responsible mutual funds yielded different scores for a same mutual fund. To solve this inconsistency, in his paper,[9] he added two social responsibility factors, top-to-bottom (TMB) and AMS, to the conventional four-factor model in order to classify and evaluate the performance. These two factors represent the common criteria for classifying funds as socially responsible. His study is different from Sharpe's because the way he calculates returns. Factor returns are calculated as differences in returns between two indices. The interesting conclusion one can take from this study is that with this factor model, we could expand the criteria to not only the TMB and AMS. Thus constructing factors that reflect Islamic or Catholic values.

In a more technical note, K.H. Bae *et al.*,[10] investigate the role of corporate social responsibility and stock market returns during the market crash during the pandemic. In their study, they tested the notion that CSR protects firm value during crisis period. In order to test that, they examined a sample of 1750 US firms during the crisis and post-crisis using CSR, MSCI ESG Stats and Refinitiv ESG. They have found no evidence that CSR affected stock return during the market crash. However, the most important conclusion of the paper states that pre-crisis CSR was not effective at protecting shareholder wealth from the effects of the crisis suggesting a disconnection between firms' orientation with regards to social responsibility and actual actions.

This incoherence shows up in several ways in the industry. Growing evidence suggesting that there is an asymmetry information between asset managers and investors in SRI market can ultimately affect performances and mislead investors. In order to shed light on this subject, J.B. Hasse *et al.*[2] worked on this information asymmetry advocating for more regulation. This information asymmetry is related to the difficulty investors experience when they evaluate financial and extra-financial performance of mutual funds. The paper studied the identity of a panel of equity funds according to ESG signals sent by managers and ESG ratings attributed to the funds by rating agencies. Their findings show that information asymmetry has been growing in recent years vie the practise of "ESG-washing" related to "green-washing". Some managers mislead investors with unsubstantiated claims about their environmental, social and governance commitments. The ESG-washing is rooted in the lack of public governance of the mutual fund industry. Moreover, the study evidences that SR mutual funds names are not related to their non-financial performance in accordance with the literature. This study takes part on the debate for a need of regulation of the mutual fund industry and makes the case for such a regulation claiming that it could improve the efficiency of this financial market.

Now that we have a clear view on the current discussion about SRI and conventional investments, how they are and were compared, we move on to the next section where we are going to present our method for comparison.

3.2 The overlapping coefficient: an overview

The scope of this review is nevertheless limited. The mathematical background required to understand every detail in our references is non-negligible. I try to lay out the most important information required to understand what has been said about the use of the overlapping coefficient and its main uses. We will explain this in a chronological order explaining each step the development of this tool.

First of all, what is exactly the overlapping coefficient? To put it as simple as possible, the overlapping coefficient is just a measure of the agreement between two probability distributions. Mathematically, it is described as the common area of two probability density curves and it is written as,

$$\mathrm{OVL} = \int_{-\infty}^{\infty} dX \ \min\left[f_1(X), f_2(X)\right] \tag{3.1}$$

where $f_1(X)$ and $f_2(X)$ are the probability density functions of two distributions. Another formulation for this index has also been used in the past called the dissimilarity index but OVL has been preferred in order to asses the meaningfulness of the difference between two distributions. According to Bradley,[7] the first person to use the overlapping coefficient was Weitzman(1970) where he wanted to compare the income distributions of two distinct racial groups in the United States.

After this work, J.L. Gastwirth wanted to address the question of equality of pay between men and women.[4] Specially, he wanted to come up with a statistical method that would detect areas of labor in which the wages are not equal and this method could be followed over time. In his time, main statistical agencies used three different methods, one of which was the overlapping index. He found a flaw on the overlap measure and suggested a new measure called the $PROB$, related to the Mann-Whitney form of the Wilcoxon test. This measure consisted on randomly addressing the probability to select a woman that earns at least as much as a randomly chosen men. He comparing the measuring methods, he fond that the OVL and $PROB$ behave in a similar way and most importantly he found evident discrepancies concerning the wages of women compared to the men's.

However, the use of the overlap coefficient was not fully developed and its applicability fell short in practise. Then came, a series of work trying to change just that. In 1989, the properties of the overlapping coefficient were laid out and an extensive treatment for estimators of this measure was presented by H.Inman and E.Bradley.[7] They treated the special case where both distributions functions are univariate normal densities and the OVL must be estimated from sample date. In that moment, developments in the point estimation of the overlapping coefficient were either incorrect or ignored. In the 1989 paper, they suggested estimators for the value of OVL using the non central F-distribution. They lay out maximum-likelihood estimators for the overlap between two normal densities with equal variances that turned out to provide useful and reliable information when the sample sizes are large. They also noted that the OVL, thanks to its invariance under certain transformations, is easily implemented in computation programs. One important aspect, and in fact the one that will guide future research is the construction of confidence intervals estimates for OVL. From the paper,[7] we could construct these estimates however the methods were sparse and further developments should generalize the tool in order to use it widely. In 1994, the investigations on the overlap concentrated on tests of hypotheses and interval estimation of the OVL. These investigations being confusing, Inman and Bradley discussed a direct link interval estimation and tests of hypothesis. They devoted most of the paper in constructing and explaining the interval estimation procedures. They worked with the non-central F-distribution and other interval estimation options such as the non-central t-distribution under a normality approximaition.[6] At the end of the '94 paper,[6] they used a Monte Carlo simulation to compare the performance of the five interval estimators the described for the OVL. They found that the non-central t-distribution and the normal approximation perform quite similarly but because of the latter being more simple. they suggested to use it more when possible in computational implementations. In 1999, B.Reiser provided a more insightful view on the estimation of confidence intervals for the OVL when he investigated further the properties of the estimation intervals based on the non-central F- and t- distribution. Also, he provides analytical and simulations results in order to clarify the approaches and their limitations. His simulation study was carried out for several sample sizes and he denoted if the confidence interval contained or not the true value of the overlap. He found that the non-central F-distribution performs well for small sample sizes and it is recommended over the non-central t-distribution method because of its asymmetry in the results found which could be quite annoying when looking at sample data. He also suggests that the normal approximation should be preferred.

However, as pointed out by G.Anderson in 2012,[5] when encounter a discretized setting problem can lose information and raises inconsistencies. Besides that, the author points out that the lack of consensus on appropriate parametric models and the misspecification bias calls for a nonparametric approach which he explains in his 2012 paper. He lays out a robust nonparametrical method to assess the OVL and uses it to study the polarization within China in the recent decade proving that his method yields significant statistical results and can be used for different applications. After explaining the mathematical properties and asymptotic behaviour he studies the performance on small sample problem. He found that the performance of the method improves with sample size. The interesting aspect of this paper in our work is that he shows a robust and general theory for comparing non-independent samples. However, an important problem encounter for the use of the OVL has been the lack of methods for reliable estimation of confidence intervals.[14] In 2017, Wang and Tian exposed problems for both paramteric and nonparametric approaches and proposed robust methods for confidence interval estimation of OVL using multiple distributions. Their goal was to present multiple case scenarios in order to show the applicability of the OVL. They treated normal, a transformed normal and multimodal distributions. Also, they show generalized inference and parametric bootstrapping methods to construct confidence interval estimation for the OVL. Most importantly, in section 2 and 3 of the paper,[14] they lay out a well-structured guideline for constructing the confidence intervals using both approaches so it is easy to implement into code. At the end the did a simulation study to test the constructed confidence intervals from the proposed methods and also they applied their methods to a study a gene expression suspected in important processes in immunology or cancer. Finally, they showed that the OVL becomes a more suitable measure to use compared to the widely used method of AUC.

Finally, A. Franco-Pereira *et al.*[1] further studied the OVL and the construction of confidence interval estimation using parametric and nonparametric approaches. In the article, they presented readily applicable software and numerical methods were presented. The parametric approach exposed here was similar to the one presented by Wand and Tian.[14] The nonparametric approach was presented using the known kernel-based methods. When comparing all the available methods parametric approaches are preferable overall.They recommend a specific method and applied it to an HIV study. However, they point out that there is not a clear cutoff value which characterizes a good differentiation between the two distributions. They present a rough guideline to follow in consideration obviously with the confidence interval for the estimated OVL. To conclude, OVL can indeed serve as a simple measure for differentiation of two distribution density functions. It has been used in medicine, geology, sociological studies giving reliable results confirming it works. One must be cautious however, the approach has come a long way and the methods and computational implementations are now ready to use but there r work to do. When implementing this approach to our work of SRI, we must characterize the best we can the confidence interval and choose one of the many OVL approaches there are.

Chapter 4

Data and Methodology

4.1 Data

Since the aim of this dissertation is solely to put forward an econometric approach that could be used by an investor alone anywhere and anytime, allowing him to spot greenwashing activities, we considered that the best alternative for our data choice would be working with indices. On the one hand it gives us the opportunity to have a global approach and on the other hand it brings us a certain simplicity compared to working either with funds or with a basket of stocks considering the fact that in those cases it would be arduous and nonsense to carefully make a portfolio construction. This comes from the fact that most of the time funds manage their information privately and it will take a significant amount of time and effort to contact them for acquiring said information.

From this perspective, we selected the MSCI World and the MSCI World Low Carbon index, which are composed respectively of 1562 and 1357 companies of all developed markets in the world. Those indices provided two possible scopes: Carbon emissions and ESG environmental score. In order to answer precisely the question we have, we decided that the best choice would be working with carbon emissions. The entirety of our data comes from MSCI official website and has been analyzed over one year period, more precisely from 2020.

4.2 Methodology

4.2.1 *t*-test

This test is widely used in statistics and can be used, for instance, to determine if the means of two populations are significantly different from each other. In our case of study, we used the two-sample t-test because our two samples are independent.

The general formula for the t-test assuming equal variances is,

$$t = \frac{\overline{Y}_1 - \overline{Y}_2}{s_p\sqrt{1/N_1 + 1/N_2}} \tag{4.1}$$

where $\overline{Y}_i$ are the sample means, N_i, $i = 1, 2$, are the sample sizes and

$$s_p^2 = \frac{(N_1 - 1)s_1 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \tag{4.2}$$

4.2.2 *Welch*-test

The *Welch*-test is a generalization of the two-sample t-test, the formula is,

$$t_W = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s_1^2 + s_2^2}} \tag{4.3}$$

where $\overline{X}_i$ and s_i are the i^{th} sample mean and its standard deviation.

This test is designed for unequal population variances but as for the "standard", the normality condition is maintained.

4.2.3 Wilcoxon-Mann-Withney (WMW) test

This is a non-parametric test. It is designed to test the null hypothesis that the two populations are equal. It has the general formula of,

$$U = \sum_{i=1}^{n}\sum_{j=1}^{m} S(X_i, Y_j) \tag{4.4}$$

where X_i is an i.i.d sample from X and Y_i an i.i.d sample from Y. Also,

$$S(X,Y) = \begin{cases} 1, & \text{if } X > Y \\ \frac{1}{2}, & \text{if } Y = X \\ 0, & \text{if } X < Y. \end{cases} \tag{4.5}$$

For this kind of test, the theoretical range of U is from 0(complete differentiation between groups, the null hypothesis is most likely false and $\mathcal{H}_1$ is mostly likely true) to $n_1 \times n_2$, n_1, n_2 being the samples sizes.

4.2.4 Overlapping coefficient

The general formula of the overlapping coefficient is,

$$\text{OVL} = \int_{-\infty}^{\infty} dX \, \min\left[f_1(X), f_2(X)\right] \tag{4.6}$$

where $f_i(X)$ is the density distribution.

4.2.5 OVL with same variance

In this case, Eq.(4.6) becomes,

$$OVL = 2\Phi(-\frac{|\delta|}{2}) \tag{4.7}$$

where Φ is the cumulative function of the normal distribution and δ is defined as,

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \tag{4.8}$$

4.2.6 OVL with different variance

If different variances are assumed, 4.6 becomes,

$$\text{OVL} = 1 + \Phi(\frac{x_1 - \mu_1}{\sigma_1}) - \Phi(\frac{x_1 - \mu_2}{\sigma_2}) - \Phi(\frac{x_2 - \mu_1}{\sigma_1}) + \Phi(\frac{x_2 - \mu_2}{\sigma_2}) \tag{4.9}$$

assuming that $\sigma_1^2 < \sigma_2^2$ and

$$x_1 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) - \sigma_1\sigma_2\sqrt{(\mu_1-\mu_2)^2 + (\sigma_1^2-\sigma_2^2)\log(\frac{\sigma_1^2}{\sigma_2^2})}}{\sigma_2^2 - \sigma_1^2} \tag{4.10}$$

$$x_2 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) + \sigma_1\sigma_2\sqrt{(\mu_1-\mu_2)^2 + (\sigma_1^2-\sigma_2^2)\log(\frac{\sigma_1^2}{\sigma_2^2})}}{\sigma_2^2 - \sigma_1^2}. \tag{4.11}$$

4.2.7 Non-parametric OVL approach

Let X_{1n_1} and X_{2n_2} denote two random samples of sizes n_1 and n_2 taken from two independent distributions. The estimator for this distribution is,

$$\hat{f}_{X_1}(x) = \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{h}K(\frac{x - X_{in_1}}{h}) \tag{4.12}$$

where K is a kernel function and the bandwidth is,

$$h = (\frac{4}{3})^{1/5} s n_1^{-1/5}, \tag{4.13}$$

with

$$s = \sqrt{\frac{1}{n_1-1}\sum_{i=1}^{n_1}\left[X_{in_1} - \sum_{j=1}^{n_1}\frac{X_{jn_1}}{n_1}\right]^2}. \tag{4.14}$$

Chapter 5

Results and discussions

We begin our study using basic statistical tests. We worked using a ladder-approach, that is we began with simple tests then improve with more robust tests. This actually help us observe the precision of the tests and the consistency of our results and analysis. In this section we will present the results of the two-sample t-test, *Welch*-test and also the Wilcoxon-Mann-Whitney test. Finally, we will present three different versions of the OVL coefficient.

We used carbon emissions of two different populations. One being filled with conventional companies (MSCI World) and the other population is made of low-carbon emission companies (MSCI World Low Carbon). The populations samples are big, they have respectively 1562 and 1357 entries. Our first goal is to compare them and observe if there is a substantial difference in carbon emissions between the two samples.

In Fig.5.1 and Fig.5.2, we represented the frequency of the carbon emissions of each population on a histogram. The important thing to take from these histograms is that on a quick overview the populations are pretty similar almost indistinguishable. Also, we can point out that the dispersion is important for both cases. Moreover, making a *boxplot* diagram we can further see this similarity in Fig.5.3. The diagram shows two important aspects. First, the interquartile range of the conventional companies is greater than the one of low-carbon companies. More important, is that even though easily missed, the median of the two populations are not the same.

Now that we have made a quick presentation of our population and some of their properties we can begin presenting the main results. But before we do that, we must explain an important assumption made throughout this work mainly when we used the t-test and *Welch*-test. In order to use the these tests, there are several assumptions made for them to work properly. For instance, the two-sample t-test constrains ourselves to work with random variable following normal distributions. This assumption is also made by the

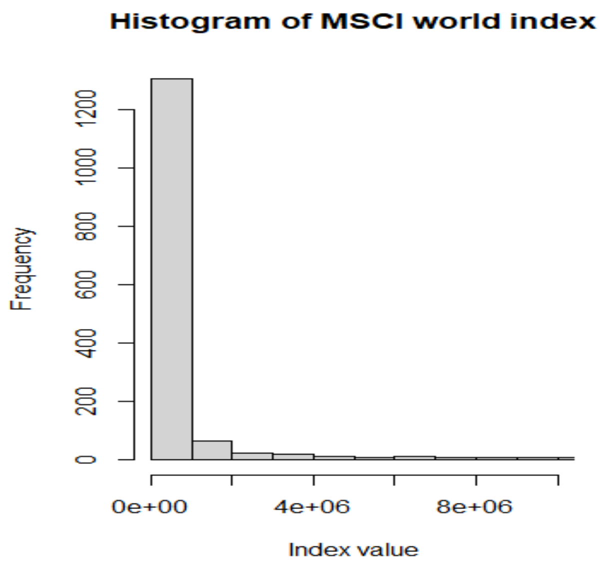

Figure 5.1: Histogram of conventional companies

Welch-test. Moreover, when using the **t**-test we assumed equal variance and that the two samples are independent. In fact, one can argue that they are independent because the low-carbon sample is a sample of the total sample. For the equal variance case, we wanted to test if this assumption made our results differ from the *Welch*-test, which does not have this assumption.

Concerning the assumption of normality, one can say two things. Firstly, normality tests always reject on the huge samples we work with today. When the number of variables gets bigger and bigger, the deviation from the perfect normality leads directly to a significant result. Thus, making a normality test such as the Shapiro-Wilks test will be uncalled for. Evidently, the null hypothesis will be reject by this test. Moreover, and perhaps most important is that the assumption is not erroneous. The Central Limit Theorem can be used in this case and thus the assumption is verified: we can say that these samples follow normal distributions.

Now, the first step in our work focused on answering a simple question: are the means of our two populations equal? To answer it we use the two-sample t-test. In order to perform this test we had assumed that the two samples are independent and they have been randomly sampled from normal populations. Moreover, we assumed equal variance for this test. This is our null-hypothesis: $\mathcal{H}_0$.

Some numerical properties are exposed on 5.1.

Using the **t.test()** function implemented in RStudio, the test shows us

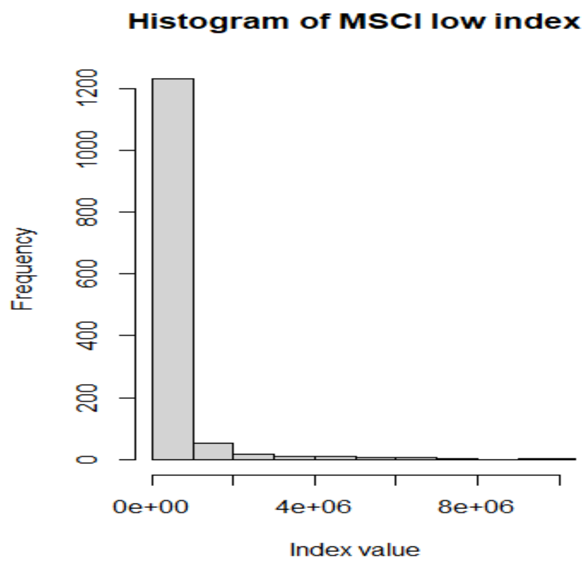

Figure 5.2: Histogram of low-carbon companies

Table 5.1: Obtained numerical mean and variances.

Results	MSCI world score	MSCI low-carbon score
Mean	2609262	693429.7
Variance	113039331441552	12845352942499

clear results. A quick reminder on how to interpret the results yielded by a t-test. We have the t-value and p-value and also the degrees of freedom. The t-value is made up of the difference between means and the variance between them. The degrees of freedom only relates to the size of the sample. Finally, the p-value tells us the probability of our t-value happening by chance.

The two-sample t-test showed that the difference is statistically significant, our t-value is,

$$t = 6.3326. \tag{5.1}$$

Moreover, our p-value yielded,

$$\text{p-value} = 2.782 \times 10^{-10} \tag{5.2}$$

which $p < 0.001$. Thus rejecting the null hypotheses. We conclude that **there is** a statistically difference between means.

However, is the assumption of equal variance well-founded? To test that, first we used the F-test to see if the variance are equal. This test gives us,

$$F = 0.11364. \tag{5.3}$$

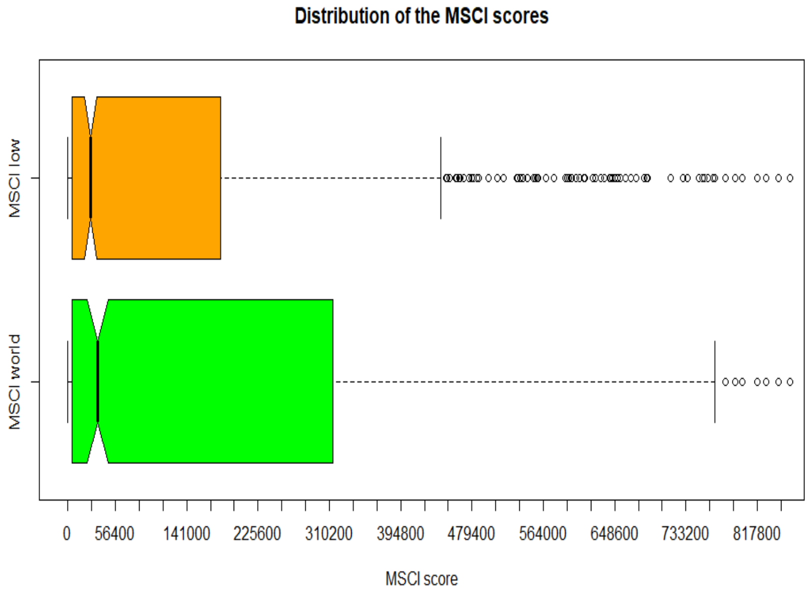

Figure 5.3: Boxplot of low-carbon and conventional emissions

This means that the equal variance is not well-founded. Because of this reason, now we used the *Welch*-test. It will help us see if there is a difference between the means of the two samples *without* the assumption of equal variance.

The results of the *Welch*-test are on the same line as those from the two-sample t-test. We found that there is a statistical difference between the two samples. Our results were,

$$t = 6.971 \tag{5.4}$$
$$\text{p-value} = 2.77E - 7. \tag{5.5}$$

We sum up the results from these two tests on 5.2

Table 5.2: Obtained numerical results of the T-test and Welch test.

Results	t-test	Welch test
t-value	6.6971	693429.7
p-value	6.3326	2.782e-10

We point out that the results from these two test differ very slightly. One can argue that if we could assume the normality of the population (via the central limit theorem), we could also assume the equal variance and thus have practically the same results as just happens here.

Before talking about the overlapping coefficient, we used a non-parametric test called the Wilcoxon-Mann-Withney test to add robustness to our previous results. The results are again unequivocal. The p-value from this test is,

$$\text{p-value} = 9.95 \times 10^{-4}, \tag{5.6}$$

thus being less than a significance level $\alpha\bar{0}.05$). We can conclude from this that there is a demonstrated statistical difference between our two samples.

Now, we want to improve our results. We want to use a more robust and general approach. This approach as explained on previous sections is called the overlapping coefficient (OVL). As pointed out before, this comparison method can be more precise than the previous methods and therefore could point to another answer. Let us see if the OVL yields the same results as the other methods.

First, we computed the OVL assuming equal variance and the normality of our samples. For that, we took our data and applied a normal distribution. It is noteworthy that this conditions are not necessary for this method to work. This means that one can have samples that do not have the same variance and means for it to work properly. This was pointed out by Reiser[8] and Pereira[1] on their respective findings. Now with this conditions, the OVL becomes,[8]

$$\text{OVL} = 2\Phi(-\frac{|\delta|}{2}) \tag{5.7}$$

where Φ is the cumulative function of the normal distribution and δ is defined as,

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \tag{5.8}$$

and σ is the weighted average of the standard deviations. The distributions for each population can be seen in Fig.5.4 Finally, the OVL with the normality approach gives us,

$$\text{OVL} = 0.89638. \tag{5.9}$$

In this approach we can compute the confidence interval (CI) in two different ways. Both methods are exposed on the research by Reiser.[8] First, we compute the CI using the normal distribution. Then, we computed the CI using the non-central t-distribution. The results of the confidence intervals for both methods are resumed in Tab.5.3. There are two main conclusion we can take from this first approach. The first one is that there is an important difference between computing the CI using normal or t-distribution. This difference accounts for 59% between lower and upper bounds. Also, we can

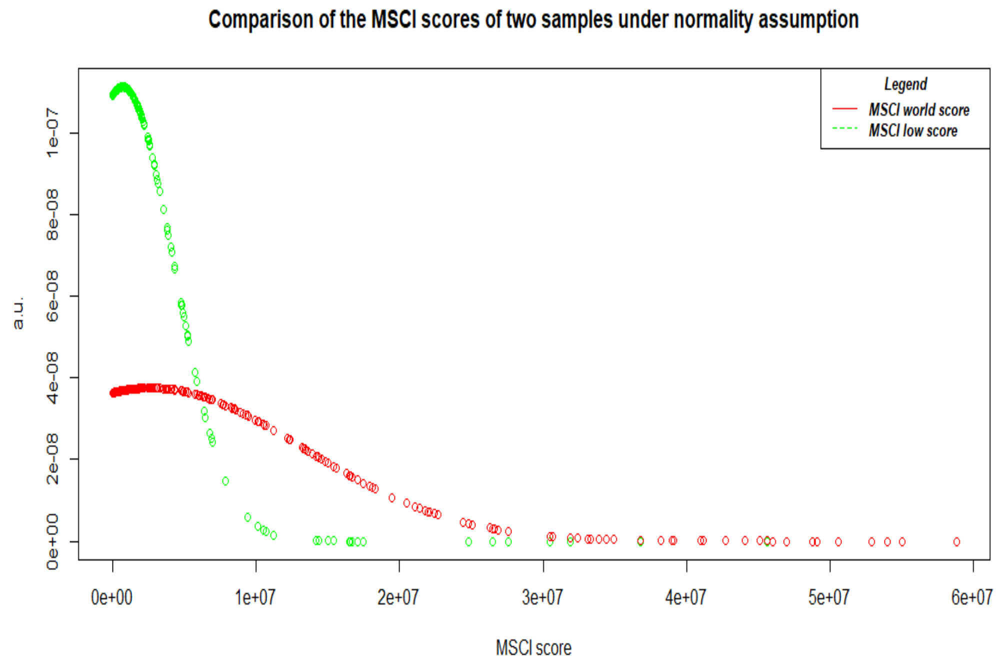

Figure 5.4: Normal distributions of conventional and low-carbon emissions companies

Table 5.3: Obtained numerical results for the confidence intervals using normal and t-distribution.

Results	Confidence intervals
t-distribution	[-14422509.9,14422511.7]
normal distribution	[-14416527.3,14416525.5]

point out that the OVL value is included on each of the 95% CI. This means that there is a big chance that the real OVL (not the estimator) is indeed the value found by our computations. Finally, since or OVL is lesser than one and greater than 0.75,[1] we can say that there is some differentiation between the two samples.

Now, we computed the OVL without the previous assumptions of equal variance and normality as was done in a parametric approach of Franco and Pereira.[1] The OVL in this case is defined as,

$$\mathrm{OVL} = 1 + \Phi(\frac{x_1 - \mu_1}{\sigma_1}) - \Phi(\frac{x_1 - \mu_2}{\sigma_2}) - \Phi(\frac{x_2 - \mu_1}{\sigma_1}) + \Phi(\frac{x_2 - \mu_2}{\sigma_2}) \quad (5.10)$$

assuming that $\sigma_1^2 < \sigma_2^2$ and

$$x_1 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) - \sigma_1\sigma_2\sqrt{(\mu_1-\mu_2)^2 + (\sigma_1^2-\sigma_2^2)\log(\frac{\sigma_1^2}{\sigma_2^2})}}{\sigma_2^2-\sigma_1^2} \tag{5.11}$$

$$x_2 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) + \sigma_1\sigma_2\sqrt{(\mu_1-\mu_2)^2 + (\sigma_1^2-\sigma_2^2)\log(\frac{\sigma_1^2}{\sigma_2^2})}}{\sigma_2^2-\sigma_1^2}. \tag{5.12}$$

The results of this approach are surprising. The OVL found was equal to one. This is highly unusual and even impossible for the type of data we worked with. This leads us to conclude that the result for this case is wrong.

In order to remediate this and confirm our previous result, we implemented the code presented by Franco and Pereira[1] using a **non-parametric approach**.The results found using this method were,

$$\mathrm{OVL}_{np} = 0.5261 \tag{5.13}$$

where np stands for non-parametric. The confidence interval using this approach was, Hence our OVL is also included on our 95% CI. Again using

Table 5.4: Obtained numerical results of the confidence interval using non-parametric approach.

Results	Confidence interval
non-parametric	[0.50002 ,0.552161]

the guideline presented by Franco *et al.*,[1] one may say that there is good differentiation.

Finally, for this first part of our work, which consisted on comparing the emissions of conventional and so-called low-carbon companies using MSCI data, we can conclude that there is in fact statistical evidence to support the premise that the two samples are indeed. Thanks to the non-parametric OVL calculation, we have found a value of 0.52. This means that both distributions do not overlap perfectly and indeed are very different.

Chapter 6

Conclusion and Recommendations

The main objective of this research was to compare a Conventional and a Low Carbon Index to find out whether or not there was a difference in terms of exposure to carbon emissions. With the increased interest for Socially Responsible investing over the last decades we have seen the rise of Greenwashing, in which some actors provide misleading information or false commitments to eco-friendly initiatives in order to create a socially responsible false image to the public. Thus hoping to increase earnings on the growing demand for low carbon. In that context, there are some indexes that claim to be Low carbon while having carbon footprints similar to conventional ones.

This study contributes to the literature by offering a methodology that allows to differentiate real SRI from "Greenwashed" Indexes. To fulfill this objective, we conduct an analysis of the MSCI World and the MSCI World low carbon over one year period using statistical and econometric methods. Based on our results of the two sample T-Test, Welch Test, Wilcoxon Man Whitney Test and the Overlapping coefficient index, we can conclude that there is in fact statistical evidence to support the premise that the two indexes differ in this aspect.

In this concern, an investor who is looking for a low carbon benchmark can rely on the MSCI World Low carbon, being sure that he will achieve a certain level of performance while generating a positive social impact. However, this analysis can be further improved by extending the period of collected data over one year. Also, in order to add robustness to this analysis one can work with different indexes. This method could be used in further research to examine conventional and SRI funds using a larger set of data.

Chapter 7

Appendix

7.1 Replication codes

```r
#import the library to read excel files
library(readxl)
library(carData)

# Read and store in lists from Excel file
world_carbon<-read_excel("Data_MSCI_Luis.xlsx",
                          sheet = 1, cell_cols("B"), na = "#N/
  A")
world_carbon[is.na(world_carbon)] = 0
low_carbon<- read_excel("Data_MSCI_Luis.xlsx",
                        cell_cols("B"), sheet = 2, na = "#N/A
  ")
low_carbon[is.na(low_carbon)] = 0

# Empty numeric vectors in which
# we are going to store the values from the lists
# This is necessary because the
#some functions take only numeric vector as input
vector_low <- c()
vector_world <- c()

# Store the values from the lists on the vectors.
for (i in 1:length(low_carbon[[1]])){
  vector_low[i] = low_carbon[[i,1]]
}
for (i in 1:length(world_carbon[[1]])){
  vector_world[i] = world_carbon[[i,1]]
}
N_low = length(vector_low)
N_world = length(vector_world)

```

```
#Plot Histogram
hist(vector_world,xlim=c(0,1e7),
     breaks=200, main="Histogram of MSCI world index",
     xlab="Index value")

#Barplot
names<-c("MSCI world","MSCI low")
dtlist<-lapply(names,get, envir=environment())
names(dtlist) <- names
boxplot(dtlist,ylim=c(0,max(vector_low)/70), notch=TRUE,
        xlab = "MSCI score",
        main="Distribution of the MSCI scores",
        col=c("green", "orange"),
        horizontal = TRUE,xaxp=c(0,max(vector_world),5000))
#------------------------------------------------------------

#Compute means and variances from each list
m_w = mean(vector_world)
m_l = mean(vector_low)
var_low = var(vector_low)
var_world = var(vector_world)

#Plot density functions
y_low<- dnorm(vector_low, m_l, sqrt(var_low))
y_world<- dnorm(vector_world, m_w, sqrt(var_world))
plot(vector_low, y_low,  xlab = "MSCI score",
     ylab = "a.u.", col = "green",
     main = "Comparison of the MSCI scores
     of two samples under normality assumption")
points(vector_world, y_world, col = "red")

# Add a legend to the plot
legend("topright",  legend=c("MSCI world score", "MSCI low
   score"),
       col=c("red", "green"), lty=1:2, cex=0.8,
       title="Legend", text.font=4)

# OVL same variance Reiser et al, equal variance
dif = (sqrt(var_world)*N_world +
      sqrt(var_low)*N_low)/(N_low+N_world)
delta = (m_l-m_w)/dif
ovl = 2*pnorm(-abs(delta)/2)
alpha = 0.05
#Confidence interval using normal distribution
CI_low<-ovl - qnorm(1-alpha/2)*dif
CI_high = ovl + qnorm(1-alpha/2)*dif
#Confidence interval using t-distribution
CI_low1= ovl-qt(1-0.05/2,df=N_world+N_low-2)*dif
CI_high= ovl + qt(1-0.05/2,df=N_world+N_low-2)*dif
```

```
#OVL Franco-Pereira et al
ovl_dif = 0
x1 = ((m_l*var_world-m_w*var_low)-
    var_low*var_world*sqrt((m_l-m_w)**2+
     (var_low-var_world)*log(var_low/var_world)))/(var_world-
    var_low)

x2 = ((m_l*var_world-m_w*var_low)+
    var_low*var_world*sqrt((m_l-m_w)**2+
     (var_low-var_world)*log(var_low/var_world)))/(var_world-
    var_low)

ovl_dif = 1 + pnorm((x1-m_l)/var_low) - pnorm((x1-m_w)/var_
    world) -
  pnorm((x2-m_l)/var_low) + pnorm((x2-m_w)/var_world)

#Evaluate t_test
t.test(vector_world,vector_low,conf.level = 0.95, var.equal =
     TRUE)
# Welch t_test
t.test(vector_world, vector_low, conf.level = 0.95)
#WMW test
res <- wilcox.test(vector_low, vector_world)
print(res)
```

7.2 Replication codes: Non-parametric approach

```
library(readxl)
# Read and store in lists both data
world_carbon<-read_excel("Data_MSCI_Luis.xlsx",
sheet = 1, cell_cols("B"), na = "#N/A")
world_carbon[is.na(world_carbon)] = 0
low_carbon<- read_excel("Data_MSCI_Luis.xlsx",
cell_cols("B"), sheet = 2, na = "#N/A")
low_carbon[is.na(low_carbon)] = 0

# Empty numeric vectors in which we are going to
store the values from the lists
# This is necessary because the
t_student function takes only numeric vector as input
vector_low <- c()
vector_world <- c()

# Store the values from the lists on the vectors.
```

```
for (i in 1:length(low_carbon[[1]])){
  vector_low[i] = low_carbon[[i,1]]
}
for (i in 1:length(world_carbon[[1]])){
  vector_world[i] = world_carbon[[i,1]]
}


ssdd=function(x){
  v=sqrt(var(x)(length(x)-1)/length(x))
  # this functions computes the standard deviation
                                          # of some vector x
  return(v)
}

#install.packages("ks")
library(ks) #the density function in R has some limitations.
 #This package has a better way
    #to estimate the PDF of a random variable.

#EPA kernel computes density function of the normal
   distribution
kernel=function(u){
  dnorm(u)
}

kernel.densidad=function(datos,puntos,h){
#estimator for the density functions
  ndatos=length(datos)
  # corresponds to the total number of data
  npuntos=length(puntos)
  # how many points are used to estimate the density function
    X_11...X_1n1
  matk=kernel((puntos%*%t(rep(1,ndatos))-
  t(datos%*%t(rep(1,npuntos))))/h) #compute the kernel
  as.vector((matk%*%rep(1,ndatos))/(ndatos*h))
  #gives the final density function
}
#
   ---------------------------------------------------------------


OVL89=function(x,y,alpha=0.05,B=100,h=1){
#this functions gives the CI in a list
  n1=length(x)
  n2=length(y)
  hx=ifelse(h==1,hscv(x),(4/3)^(1/5)*sd(x)*n1^-0.2)
  #equation (5) this gives us the bandwidth
  hy=ifelse(h==1,hscv(y),(4/3)^(1/5)*sd(y)*n2^-0.2)
```

```
  xo=x #auxiliary variables
  yo=y
  gridxy=seq(min(c(x,y)),max(c(x,y)),
  length.out=min(5*(n1+n2),1000)) #creates bootstrap samples
  fkx=kernel.densidad(x,gridxy,hx)    #density estimator f1
  fky=kernel.densidad(y,gridxy,hy)    #density estimator f2
  ff=pmin(fkx,fky)        # computes the min value used later
    in the ovl
  OVL=(gridxy[2]-gridxy[1])*sum(ff) #computes kernel based
   ovl
  OVL_ib=numeric(B)
  for (b in 1:B){   #loop iterates B times boostraping
   sampling and
  kernel based ovl
    # at the end it estimates the variance the OVL
    xn=sample(xo,replace=TRUE)
    yn=sample(yo,replace=TRUE)
    gridxy=seq(min(c(xn,yn)),max(c(xn,yn)),
    length.out=min(5*(n1+n2),1000))
    fkx=kernel.densidad(xn,gridxy,hx)
    fky=kernel.densidad(yn,gridxy,hy)
    ff=pmin(fkx,fky)
    OVL_ib[b]=(gridxy[2]-gridxy[1])*sum(ff)
  }

  var_OVL=var(OVL_ib)
  # confidence intervals of ovl
  IC1=OVL-qnorm(1-alpha/2)*sqrt(var_OVL)
  IC2=OVL+qnorm(1-alpha/2)*sqrt(var_OVL)
  return(list(IC1,IC2,OVL))
}

# Initalization of parameters

# generates normal random variates

x<- rnorm(length(vector_low), m_l, sqrt(var_low))
y<- rnorm(length(vector_world), m_w, sqrt(var_world))



OVL89(x,y,alpha=0.05,B=1000,h=1)
```

Bibliography

[1] Benjamin Reiser Alba M. Franco-Pereira, Christos T. Nakas and M.Carmen Pardo. Inference on the Overlap Coefficient: The binormal approach and alternatives. *Statistical methods in medical research*, 30(12):2672–2684, 2021.

[2] Bertrand Candelon, Jean-Baptiste Hasse, and Quentin Lajaunie. Esg-washing in the mutual funds industry? from information asymmetry to regulation. *Risks*, 9(11):199, 2021.

[3] R. Bauer et al. International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, 29:1751–1767, 2005.

[4] Joseph L. Gastwirth. Statistical Measures of Earnings Differentials. *The American Statistician*, 29(1):32–35, 1975.

[5] Yoon-Lae Whang Gordon Anderson, Oliver Linton. Nonparametric estimation and inference about the overlap of two distributions. *Journal of Economics*, 171:1–23, 2012.

[6] Henry F. Inman and Edwin L. Bradley. Hypotheses tests and confidence interval estimates for the overlap of two normal distributions with equal variances. *Environmetrics*, 5:167–189, 1994.

[7] Henry F. Inman and Edwin L. Bradley, Jr. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics-Theory and Methods*, 18(10):3851–3874, 1989.

[8] Benjamin Reiser and David Faraggi. Confidence intervals for the overlapping coefficient:the normal equal variance case. *The Statistician*, 48(3):413–418, 1999.

[9] Meir Statman and Denys Glushkov. Classifying and Measuring the Performance of Socially Responsible Mutual Funds. *Portfolio Management*, 42:140–151, 2016.

[10] Kee-Hong Bae *et al.* Does CSR matter in times of crisis? Evidence from the COVID-19 pandemic. *Journal of Corporate Finance*, 67, 2021.

[11] Marcia Ceu Cortez *et al.* Socially Responsible Investing in the Global Market: The Performance of US and European Funds. *International Journal of Finance Economics*, 2011.

[12] Sally Hamilton *et al.* Doing Well While Doing Good? The investment Performance of Socially Responsible Mutual Funds. *Financial Analysts Journal*, 1993.

[13] Yacine Belghitar *et al.* Does it pay to be ethial? Evidence from the FTS4Good. *Journal of Banking & Finance*, 47:54–62, 2014.

[14] Dan Wang and Lili Tian. Parametric methods for confidence interval estimation of overlap coefficients. *Computational Statistics and Data Analysis*, 106:12–26, 2017.