Norwegian School of Economics
Bergen, Spring 2020

The ESG Puzzle

A Meta-Analysis Exploring the Academic Dissensus on the Link Between ESG and Financial Performance

Julie Marie Hushovd Markussen and Thea Fossland Blom

Supervisor: Francisco Santos

Master thesis, Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

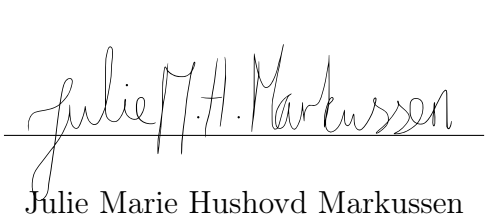
This thesis is a part of our MSc in Economics and Business Administration, with a major in Financial Economics. It marks the end of five educational years at the Norwegian School of Economics (NHH).

We would like to extend our gratitude to our supervisor, Francisco Santos. His insights and advice this semester have been of great value throughout the process of writing this thesis. We also want to thank NHH for providing access to databases and software that have been highly useful for our study.

Finally, we would like to thank our friends and fellow students for making our time at NHH a memorable one.

Norwegian School of Economics

Bergen, June 2020



Julie Marie Hushovd Markussen

Thea Fossland Blom

Abstract

In this thesis, we aim to untangle the lack of consensus among previous studies on the subject of ESG investing. We replicate four articles that focus on the Global, U.S. and European markets, at different time periods, and following different methodologies. These articles are then the basis of a meta-analysis where we consider three main explanatory factors, namely sample selection, time period and methodology. We find that the sample selection of a study affects obtained results. The Global sample exhibits more negative results in terms of the relationship between ESG and financial performance than the other two samples, and the U.S. seem to lag behind Europe. Further, we find that the time period in focus can lead to differing results. Our study exhibits more negative results in previous years compared to a more recent time period, suggesting that ESG investing is becoming increasingly more beneficial. Related to methodology, we find that the choice of weight allocation in portfolios, namely value-weighted vs equal-weighted, affect obtained results. Equal-weighted ESG portfolios seem to perform poorer. Lastly, we remark that the choice of ESG score provider might impact the results of a study. In the replications where we substitute the original ESG data with a different provider, we obtain different results than the original study. Thus, we conclude that the lack of consensus within previous research on the subject of ESG investing, can be explained by the sample selection, the time period in focus, the methodology used, and the ESG score provider.

Contents

1 Introduction **1**

2 Data **6**
2.1 Refinitiv . . . 6
2.2 Article Replications . . . 7
2.2.1 Article 1 . . . 7
2.2.2 Article 2 . . . 8
2.2.3 Article 3 . . . 9
2.2.4 Article 4 . . . 9
2.3 Meta-Analysis . . . 10

3 Empirical Analysis **11**
3.1 Article Replications . . . 11
3.1.1 Article 1 . . . 11
3.1.2 Article 2 . . . 14
3.1.3 Article 3 . . . 16
3.1.4 Article 4 . . . 19
3.2 Meta-Analysis . . . 21
3.2.1 Sample . . . 22
3.2.2 Time Period . . . 25
3.2.3 Methodology . . . 28
3.2.4 ESG Score Provider . . . 31
3.2.5 Meta-Analysis Summary . . . 32

4 Conclusion **35**

References **39**

Appendix **42**
A1 Time Period Analysis . . . 42
A2 Methodology Analysis . . . 44

List of Tables

2.1 Article Overview 7
3.1 Article 1 Replication 14
3.2 Article 2 Replication 16
3.3 Article 3 Replication 19
3.4 Article 4 Replication 21
3.5 Sample Analysis: Global vs. USA vs. Europe (2008-2018) 23
3.6 Sample Analysis: Global vs. USA vs. Europe (2004-2012) 25
3.7 Time Period Analysis: 2008-2018 vs. 2004-2012 vs. 2002-2011 (Global) . 27
3.8 Methodology Analysis: Method of Article 3 applied on the Global sample, 2008-2018 29
A1.1 Time period comparison (USA): 2002-2011 vs. 2008-2018 42
A1.2 Time period comparison (Europe): 2008-2018 vs. 2004-2012 43
A2.1 Method of Article 3 applied on the U.S. sample, 2002-2011 44
A2.2 Method of Article 1 applied on the European sample, 2004-2012 45

1 Introduction

Climate change, rapid technological changes and socioeconomic differences are all impacting the way business is done (Schiano, 2018). Now, more than ever, investors and companies are looking to do good, not only through investing in companies with solid profits and prospects, but also companies that show strong adherence to Environmental, Social and Governance (ESG) issues (Eccles & Klimenko, 2019). Since the introduction of the UN Principles for Responsible Investment (PRI) in 2006, which aim to promote incorporation of ESG factors into investment decisions, the number of committed companies has grown from 100 to more than 2300 (Principles for Responsible Investment, 2020). ESG-oriented investing has also grown steadily throughout the years, with an estimated $30 trillion invested worldwide in 2019, a 68% increase since 2014 (Henisz, Koller, & Nuttall, 2019).

The growing demand from both institutional and retail investors have led the asset management industry to shift its focus towards more sustainable industries, leaving companies that do not meet their investment requirements behind. The motivation being the idea that responsible investments will deliver better risk-adjusted returns in the long run, as companies that pay attention to ESG will have stronger cash flows and are subject to less reputational risk (FactSet Insight, 2020; Henisz et al., 2019). The growing interest in ESG has also spiked interest among researchers and has led to the publication of more than 2000 papers investigating the relationship between ESG and financial performance (Friede, Busch, & Bassen, 2015). The research is ambiguous, as different papers display different results, and consensus among scholars is far from established.

The current relevance and the importance of the subject, as well as the lack of harmony among scholars, is what spiked our interest in the subject. The chance to research how the financial industry may play a role in the drive towards a more sustainable future, is something we find highly intriguing. Therefore, we aim in this thesis to investigate the lack of consensus among researchers with regard to the effect ESG incorporation has on financial performance. For this, we replicate one thesis and three articles, all presenting different results, to gain insight as to where the lack of consensus may arise. The thesis establishes a negative relationship, one article finds a neutral relationship, and the two last articles find a positive relationship between ESG and return. These are replicated

and then further analyzed in a meta-study to uncover the main drivers behind the lack of consensus. Hereby, the four papers analyzed in this thesis are referred to as Article 1 through Article 4, and these are the following:

- **Article 1**: ESG Investments: Exploring the Impact of Sustainability on Financial Performance (Johannesen & Tveiterås, 2019)
- **Article 2**: The Wages of Social Responsibility - Where are They? A Critical Review of ESG Investing (Halbritter & Dorfleitner, 2015)
- **Article 3**: Do Socially Responsible Investment Policies Add or Destroy European Stock Portfolio Value? (Auer, 2016)
- **Article 4**: Establishing ESG as Risk Premia (Pollard, Sherwood, & Klobus, 2018)

From the replications and meta-analysis based on these articles, we find explanatory factors for the lack of consensus in research on the subject of ESG investing. The differences in conclusions can be explained by the sample market in focus, the time period of the sample, the methodology used to analyse the relationship, and finally, the choice of ESG score provider. From the sample-analysis, we find that conducting studies on the European market leads to more positive conclusions on the relationship between ESG and return, than when the focus is on the U.S. or the Global market. For the time period factor, we find that analyzing data from more recent years leads to more positive conclusions than what is the case when focusing the study just ten years back in time. When considering the methodology, we find that ESG portfolios that are equal-weighted yields more negative returns than the value-weighted portfolios for the U.S. and Global sample.

Considering the choice of ESG score provider, it is worth mentioning that different providers employ different methodologies for constructing the ESG scores. Which provider is used in the analysis might therefore affect the obtained results. In our study, we only use the Refinitiv scores, as we do not have access to data from other score providers. For the replicated articles using other providers, we obtain different results than the original articles. This might be related to the difference in ESG score input, as the methodologies and samples we use are similar to those of the original articles. Thus, it is reasonable to believe that the choice of ESG score provider is another explanatory factor for the lack of consensus in previous studies on the subject of ESG investing.

For the replications, we employ the methodologies and samples used in the original articles. For Article 1, we construct both equal-weighted and value-weighted portfolios based on a Global sample of stocks. We then employ a high-low strategy where we buy companies with high ESG scores and short companies with low ESG scores. To analyze the performance of the portfolios, the following asset pricing models are deployed: CAPM, the Fama French three factor model (Fama & French, 1993) with and without Momentum, the Fama French five factor model (Fama & French, 2015) with and without Momentum, as well as the Fama French five factor model with Momentum and the Liquidity factor. Like the article, with a strategy of buying high ESG score companies and shorting low ESG score companies, we obtain negative abnormal returns. Hence, we find a negative relationship between ESG and financial performance.

Article 2, which looks at the U.S. market from 2002-2011, finds no evidence of neither superior nor inferior performance when incorporating an ESG strategy. Similar to Article 1, the authors apply a high-low strategy. We construct two value-weighted portfolios for each year based on the top and bottom 20% ESG performers. This article goes one step further as it also constructs portfolios based on the individual Environmental, Social and Governance (ESG) pillars. To evaluate the performance of these portfolios, we deploy the Carhart model (Carhart, 1997). Contrary to the article, we establish a somewhat negative relationship between the various ESG factors and financial performance.

Article 3 looks at the European market with ESG scores obtained from Sustainalytics, and presents a positive relationship between ESG and financial performance (Auer, 2016). Given our limited access to data providers, we substitute the ESG scores from Sustainalytics with the ones available through Refinitiv. Based on Stoxx Europe 600 in the time period from June 2004 to October 2012, we construct portfolios based on the ESG score, as well as the individual pillars. Constructed portfolios include a benchmark comprising the whole investment universe, a rated only (RO) portfolio comprised of companies with an ESG score, as well as negatively screened portfolios constructed for the combined ESG score and the individual pillars. We then estimate Sharpe ratios and evaluate the differences between the portfolios by conducting a Bootstrap test. Contrary to the article, we find no evidence that the screened portfolios outperform the benchmark.

Article 4 presents a positive relationship between ESG and financial performance when

considering the Global market (Pollard et al., 2018). The scores are originally retrieved from MSCI, but substituted with the scores from Refinitiv in our replication. Three benchmark portfolios and three ESG portfolios, consisting of thirty randomly selected stocks, are constructed in 2007. The portfolios are then rebalanced every quarter. The 20% companies with the lowest returns are substituted with new random companies for the benchmarks, and the companies with the highest ESG score increase for the ESG porfolios. The ESG portfolios are then compared to their associated benchmarks. In contradiction to the original article, we do not find evidence that the ESG portfolios outperform the benchmarks. Instead, we obtain inconsistent results. Due to the questionable methodology and the ambiguous results, we choose to exclude this article from the meta-analysis.

After replicating the articles, we further deploy these (Article 1-3) in a meta-study where we investigate potential drivers behind the lack of consensus on the subject of ESG investing. We test three factors, namely sample selection, time period and methodology. We control for each of these factors separately, making the only difference between a set of regressions the input of one factor. Thus, when comparing the results, we can ascribe the variation to the factor in focus. For example, we test the sample factor by comparing results of analyses focused on the Global, U.S. and European markets, by holding the time period and methodology constant. Here, varying results will then be due to the sample input. Similar analyses are conducted for the time period and methodology factors. This allows us to more clearly identify where differences in conclusions may arise. Thus, we can pinpoint factors that help explain the ambiguity in previous research on the subject of ESG investing.

The four articles analyzed in our thesis, are just some examples of studies where different conclusions are reached when considering the link between ESG and return. In the following, we will look further into the literature on the subject. Some papers find a negative relationship between ESG and financial performance, and some find no relationship at all. Orlitzky (2013) concludes in his study that CSR will have unintended negative consequences on stock market volatility and pricing due to asymmetrical information and investor sentiment. Furthermore, Revelli and Vivani (2015) find in their study no evidence that incorporating CSR issues into an investment strategy yield neither superior nor inferior results. Another study that finds no evidence that ESG leads to abnormal

returns, is a paper by Breedt et al. (2018). Here, the authors incorporate the three ESG pillars into a worldwide equity portfolio, and finds a negative relationship between ESG and performance in the U.S., while Europe exhibits slightly positive results. The results are however not significant, and hence the authors cannot infer a relationship between ESG and financial performance.

Other studies instead find a positive relationship between financial performance and ESG. A study by Clark, Feiner, and Viehs (2015) investigates more than 200 published papers on the subject. They find that incorporating sustainability issues into investment decisions will, 80% of the time, result in better stock performance. The study also highlights how ESG practices increase the operational performance of firms. Another study, published by MSCI Research Insights, found that incorporating ESG criteria into passive strategies generally led to higher risk-adjusted returns (Melas, Nagy, & Kulkarni, 2016).

As is evident from the presented research, there still exists uncertainty as to the exact relationship between ESG and financial performance. In this thesis, we therefore attempt to identify the drivers behind this uncertainty. In the following, we present the structure of the thesis. Section 2 presents the data and samples for each of the four replicated articles separately, as well as the additional data collection for the meta-analysis. In section 3, we present the empirical analysis and results of each of the replications, followed by the methodology and results of the meta-analysis. Finally, we summarize our findings and present our conclusion in the fourth section.

2 Data

In this section, we present the data collection and data cleaning for the four articles in our replication. The first part focuses on the data provider Refinitiv and the methodology behind the ESG rankings. In the second part, we explain, for each article separately, what data is collected and how it is cleaned in accordance with the original article. Finally, we discuss the additional data collection and processing needed for the meta-analysis.

2.1 Refinitiv

This section presents the data utilized in the paper as well as the methods used to prepare the data for the empirical analysis. Data for all four articles is retrieved from Refinitiv Datastream, an integrated feature of Eikon formerly known as Thomson Reuters Datastream. Datastream is a financial database covering 70 years of historical data across 175 countries. It provides detailed information on bonds, equities, convertibles, and stock market indices, as well as ESG data (Refinitiv, 2019).

We employ the Refinitiv ESG scores in all four article replications. These scores are based on verified data that is publicly available. This allows for differentiation between companies that implement and execute ESG measures, and those who just proclaim to do so. This ensures objectivity. Article 1 and 2 use the Refinitiv scores in their analyses, while Article 3 obtains ESG scores from Sustainalytics, and Article 4 from MSCI. As neither ESG scores from Sustainalytics nor MSCI is accessible to us, we choose to replace these scores with the Refinitiv score in the replication, while the sample and methodology remain the same.

Refinitiv gathers and computes over 450 company indicators related to ESG. The company scoring and assessment process consists of the 186 indicators that are the most relevant and applicable for comparison. These indicators are then grouped into 10 categories that compile the three pillars of ESG, namely the Environmental, Social and Governance pillars (Refinitiv, 2020).

The companies are assigned scores based on their relative performance compared to other companies in the 10 categories. The category scores within the Environmental

and Social pillars are benchmarked against the industry each company belongs to. For the Governance categories, companies are rated relative to the performance within their respective countries. The reasoning behind the different benchmarks is that environmental and social categories are more consistent within industries, while governance practices are more consistent within countries. Each category is weighted depending on the number of indicators it consists of, so that categories with more indicators are assigned higher weights than those with few. The category scores range from 0 to 100, with 100 being the best possible score. The pillar scores are then constructed based on the scores and weights of their underlying categories, while the final overall ESG score is a product of the total scores and weights of the pillars (Refinitiv, 2020).

2.2 Article Replications

In the following, we present the samples and data processing of the four articles we replicate. Table 2.1 shows an overview of each article, while the next subsections describe the samples for each of the articles in more detail. As can be seen from the table, the articles differ in terms of samples and time periods, and exhibit varying ESG score characteristics. It is evident that the mean of the European sample is higher than the other samples. When looking at the two global samples of Article 1 and 4, we see that a one year shift in time period leads to a higher mean, as well as a higher minimum and maximum ESG score.

Table 2.1: Article Overview

	Sample Descriptives			ESG Score		
	Sample	Time Period	Company Count	Min	Max	Mean
Article 1	Global	2008-2018	8474	5.11	97.66	50.11
Article 2	U.S.	2002-2011	3242	0.41	95.38	33.43
Article 3	Europe	2004-2012	826	1.43	95.19	50.48
Article 4	Global	2007-2017	8771	0.06	95.38	40.26

This table presents an overview of the four article samples, as well as the ESG score descriptives for the samples.

2.2.1 Article 1

In accordance with Article 1, we make use of the Asset4 Universe List available through Datastream. We obtain yearly ESG scores, monthly total returns, market value and

market-to-book data for each individual company in the list. For the factors applied in the regression models, we retrieve the Pastor-Stambaugh liquidity risk factor (Pastor & Stambaugh, 2020), while the remaining factors are collected from the Kenneth French Data Library (French, 2020a).

When preparing the data for the analysis, a reported ESG score in January is necessary for portfolio construction. Hence, companies without a score in January are excluded from portfolio construction that year. It is also worth mentioning that there are no requirements as to having continuous years with an ESG score, in order to avoid survivorship bias. Furthermore, for a company to be eligible for picking, return is required in January as this is the month in which the portfolios are constructed. In accordance with the article, we retrieve monthly stock returns based on Datastream's Total Return Index (RI), where dividends are assumed to be reinvested. The RI represents the cumulative return of a stock. To calculate monthly returns (r_t) of a stock, we use the following formula, where RI_t is the Total Return Index of a stock at time t:

$$r_t = \Delta RI_t = \frac{RI_t - RI_{t-1}}{RI_{t-1}} \tag{2.1}$$

2.2.2 Article 2

To replicate Article 2, we retrieve the full Asset4 rating universe for U.S. companies in the period 2002 to 2011. From Datastream, we also obtain financial information such as monthly total returns, market capitalization data and book-to-market ratios. We calculate monthly returns in accordance with the above formula (Formula 2.1). Further, we collect ESG measures, including data on the individual pillars, Environmental, Social, and Governance, as well as the yearly ESG scores. As for the risk premia applied in the analysis, the authors of the article point to certain weaknesses with the HML factor available through Kenneth French's data library, and so we apply the risk premia available from AQR (2020) based on the work by Asness and Frazzini (2013). As this article is focused on the U.S. market, the one-month U.S. Treasury bill represents the risk-free interest rate (Halbritter & Dorfleitner, 2015).

2.2.3 Article 3

The sample for Article 3 consists of the historic constituents of the STOXX Europe 600 Index in the period June 2004 to October 2012 (Auer, 2016). In total, we gather data for 925 companies for this period. The authors of the article collected ESG scores from the Sustainalytics database for their analysis. Due to lack of access to this database, we replace the ESG scores with the ones available through Refinitiv. In total, 826 companies had ESG scores available in the relevant period. In accordance with the article, we exclude companies where ESG scores are available for less than six months. In addition to the combined ESG score, we also collect scores for the three pillars of ESG, namely the Environmental score, Social score and Governance score.

We obtain monthly total returns for the companies from Datastream's Total Return Index (see Formula 2.1). In the article, stock prices and dividends were retrieved to calculate the stock returns (Auer, 2016), these are however integrated into the calculation of the Total Return Index we retrieve from Datastream. To calculate excess return, we must also retrieve risk-free rates for the period. In correspondence with the article, we extract the monthly European Interbank Offered Rate (EURIBOR) for one-month deposits from the German central bank's database (Deutsche Bundesbank, 2020). These rates are subtracted from the monthly stock returns to obtain the monthly excess returns.

2.2.4 Article 4

Article 4 is based on the full sample of MSCI ESG rated companies from January 2007 to January 2017. It comprises 6400 companies globally. As we do not have access to the MSCI database, we instead use the Asset4 full universe list which is comprised of more than 8000 companies. In accordance with the methodology in the article, we construct three sets of portfolios with 30 randomly selected companies from the full universe of stocks. Considering that the sample used in the article and the sample we collect are both comprised of large, mid and small-cap companies on a global basis, it is likely that retrieving data from a "different" sample will not considerably affect the results. The ESG rating may thus be considered as the main difference between the analysis in the article and our analysis. The article uses quarterly ESG scores from the MSCI database, while we employ quarterly ESG scores from Refinitiv.

In addition to the ESG score, we retrieve monthly return data for the companies, which are then aggregated and transformed to quarterly returns, in accordance with the article. When turning over the portfolios, which we will come back to in the empirical analysis section, we need the quarterly growth in ESG score per company. This measure is calculated based on the quarterly ESG score we retrieve from Refinitiv. For this article, the randomly selected benchmark portfolios operate as the risk free rate when calculating Sharpe ratios for the ESG portfolios.

2.3 Meta-Analysis

To execute the meta-analysis, we collect additional data beyond what is utilized in the replications. To compare and contrast the samples, methodologies and conclusions of the articles, we test each of them by integrating aspects of the other articles. Thus, for each of the three articles included in the meta-analysis we collect data not only for the time periods of each respective article, but for the combined time period 2002-2018. To test the methodologies across the different samples, we also collect the various risk factors used in Article 1 and 2 that match the other articles' samples and time periods. We further need to collect the ESG pillar scores for all samples. Additionally, we retrieve financial information such as market value and return for the full samples and the full combined time period. This information is retrieved in the different formats and time laps needed for each method. As for the data cleaning, we apply the same steps as for the individual articles, depending on the methodology in focus for the different meta-analyses we conduct.

3 Empirical Analysis

This section presents the methodologies and results of the replications and the meta-analysis. First, we replicate the selected articles to see whether we find the same results as the original articles. We then examine the reasons for the the lack of consensus among researchers by applying a meta-analysis methodology where we control for different factors, namely sample selection, time period and methodology. We also discuss the impact the choice of ESG score provider might have on the obtained results. Finally, we summarize the results of the meta-analysis.

3.1 Article Replications

3.1.1 Article 1

In this article, a high minus low strategy is utilized to test the effect ESG incorporation has on financial performance. In other words, we construct a high portfolio consisting of the companies with the highest ESG scores, and a low portfolio consisting of the lowest scored companies. We then obtain the excess return of the strategy of buying the high portfolio and selling the low portfolio, by subtracting the low portfolio returns and risk-free rates from the high portfolio returns. We create portfolios consisting of the 10% firms with the highest and lowest ESG score, and similar with a 25% threshold. The portfolios are picked in January each year and are held for one year, after which they are rebalanced according to the new ESG scores. Furthermore, we reallocate the portfolio weights when a company is delisted, which is further explained in the following paragraph.

For each of the thresholds of 10 and 25%, we construct both equal-weighted and value-weighted portfolios. In the equal-weighted portfolios, each stock is assigned the same weight, while in the value-weighted portfolios, the stocks are assigned relative weights based on their market value. Bigger companies are thus assigned bigger weights, and smaller companies are assigned smaller weights. If a company is delisted during a year, it is dropped from the portfolio. In the value-weighted portfolios, the company is automatically given a weight of zero, while for the equal-weighted portfolios, we redistribute the weight of the delisted company among the remaining companies in the portfolio.

To evaluate the performance of the constructed portfolios, six different asset pricing models are utilized on the four portfolios. These are the following:

(1) CAPM
(2) Fama French three factor
(3) Fama French three factor with Momentum
(4) Fama French five factor
(5) Fama French five factor with Momentum
(6) Fama French five factor with Momentum and Liquidity.

The main findings from the part of Article 1 we replicate, is that all the models have negative alphas that are significant at a 1 and 5% level for the equal-weighted portfolios. This means that the portfolios consisting of high-rated companies significantly underperform compared to the portfolios with the low-rated companies. As for the value-weighted portfolios, the alphas are also negative, however, they are not significant. The lack of significance for the value-weighted portfolio alphas can be explained by the fact that small companies often generate more alpha than large companies (Banz, 1981), and that small companies are given smaller weights in these portfolios.

When replicating the sample and methodology of Article 1, we find that the alphas in all the regressions run for the equal-weighted portfolios are significantly negative at a 1 and 5% level (Table 3.1). This is in line with the results from the article. However, our results exhibit slightly more negative alphas for the three different five factor regressions of the equal-weighted portfolios compared to the article. In our replication, the equal-weighted quartile portfolio exhibits negative alphas of between 0.38 and 0.49%. This implies that monthly returns of the low ESG portfolio on average outperforms the high ESG portfolio by this same range. In the article, this estimate is approximately 0.4%. As for the decile portfolio, our results exhibit alphas between 0.63 and 0.73%. This implies that the low ESG portfolio outperforms the high ESG portfolio with an even bigger percentage for the decile portfolio than for the quartile portfolio.

For the value-weighted portfolios, our results exhibit non-significant negative alphas for all the models, similar to the article. The alphas for our value-weighted portfolios range from negative 0.06% to negative 0.16%, while in the article they range from negative 0.26% to negative 0.69%. Furthermore, we find that the Small-minus-Big (SMB) coefficient is

significantly negative for all regressions where it is included. This means that our strategy is putting more bets on large companies and is more exposed to the risk of these companies. In Article 1, they find significant SMB-coefficients only for the three factor models of the value-weighted decile portfolio, and positive coefficients as such. The authors of this article therefore remark that, with their strategy, small market cap companies have a risk premium (Johannesen & Tveiterås, 2019).

Further, the CMA factor is significant and positive for model (4), (5) and (6) for all four portfolios. The CMA factor represents the difference between returns of companies that invest conservatively, and of those that invest aggressively (French, 2020b). Our positive coefficients therefore imply that the high ESG portfolio consists of more companies that invest conservatively rather than aggressively. The HML coefficient is positive and significant at a 1% level for model (2) and (3) for both value-weighted portfolios. This suggests that investors rather buy value companies than growth companies in the high ESG portfolio. Finally, we obtain negative liquidity factor coefficients in model (6) that are significant for both the value-weighted portfolios.

Similar to the original Article 1, we find a negative relationship between ESG and financial performance. Our results exhibit significantly negative abnormal returns for the high-low strategy for all asset pricing models for the equal-weighted portfolios. For the value-weighted portfolios, we obtain insignificant alphas for all models. This is in accordance with the findings in Article 1.

Table 3.1: Article 1 Replication

Panel A: Quartile Portfolios

	Equal-Weighted Portfolios						Value-Weighted Portfolios					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Mkt-rf	0.01	-0.03	-0.02	0.03	0.03	0.03	0.02	-0.04*	-0.02	0.02	0.02	0.03
SMB		-0.49***	-0.48***	-0.42***	-0.42***	-0.41***		-0.54***	-0.53***	-0.47***	-0.48***	-0.46***
HML		0.10	0.14*	0.06	0.07	0.06		0.20***	0.26***	0.12	0.15*	0.14*
WML			0.05		0.00	0.01			0.08**		0.03	0.04
RMW				0.17	0.17	0.17				0.11	0.11	0.09
CMA				0.33***	0.33***	0.31***				0.39***	0.36***	0.31***
LIQ						-0.02						-0.06*
α	-0.43***	-0.38***	-0.39***	-0.49***	-0.49***	-0.49***	-0.13	-0.06	-0.08	-0.16	-0.16	-0.16
N	132	132	132	132	132	132	132	132	132	132	132	132
R2	0.00	0.21	0.22	0.25	0.25	0.26	0.00	0.39	0.42	0.47	0.47	0.49
Adj. R2	-0.01	0.19	0.19	0.22	0.22	0.21	0.00	0.38	0.40	0.45	0.45	0.46

Panel B: Decile Portfolios

	Equal-Weighted Portfolios						Value-Weighted Portfolios					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Mkt-rf	0.02	-0.03	-0.02	0.05	0.05	0.05	0.02	-0.05*	-0.04	0.03	0.02	0.03
SMB		-0.54***	-0.53***	-0.47**	-0.47**	-0.48**		-0.49***	-0.48***	-0.43***	-0.43***	-0.41***
HML		0.22	0.25	0.03	-0.03	-0.02		0.27***	0.31***	0.09	0.07	0.06
WML			0.04		-0.04	-0.05			0.05		-0.02	-0.01
RMW				0.01	0.02	0.03				0.01	0.01	0.00
CMA				0.57***	0.63***	0.65***				0.52***	0.55***	0.49***
LIQ						0.03						-0.07*
α	-0.70***	-0.63**	-0.64**	-0.73***	-0.73***	-0.73***	-0.15	-0.06	-0.07	-0.15	-0.15	-0.16
N	132	132	132	132	132	132	132	132	132	132	132	132
R2	0.00	0.09	0.09	0.13	0.13	0.13	0.00	0.28	0.29	0.39	0.39	0.40
Adj. R2	-0.01	0.07	0.06	0.10	0.09	0.08	-0.01	0.27	0.27	0.36	0.36	0.37

Note:*<0.1;**p<0.05;***p<0.01

This table exhibits the findings of the replication of Article 1, based on the Asset4 universe list from 2008-2018. Panel 1 presents the quartile portfolios divided into an equal-weighted and a value-weighted portfolio. Panel 2 contains the equal-weighted and value-weighted decile portfolios. The dependent variable is the monthly excess return of the high-low portfolio, were we go long in companies with high ESG scores, and short in those with low ESG scores. The intercept, alpha, is the achieved abnormal return from the strategy and Mkt-rf represents the market risk premium. The SMB-factor represents the outperformance of small versus big firms, while the HML-factor captures the exposure to high book-to-market stocks. WML, known as the momentum-factor, captures the exposure to winners versus losers. RMW captures the exposure to firms with robust versus weak operating profitability. CMA represents the outperformance of companies investing conservatively as opposed to aggressively (French, 2020b). Finally, the LIQ-factor captures the exposure to liquidity shocks. The table consists of six different regression models: (1) CAPM, (2) Fama French 3 factor, (3) Fama French 3 factor with Momentum, (4) Fama French 5 factor, (5) Fama French 5 factor with Momentum, and (6) Fama French 5 factor with Momentum and Liquidity.

3.1.2 Article 2

The method applied in this article is similar to the one utilized in Article 1. In accordance with the article, we construct value-weighted portfolios for each year from 2002 to 2011 for the U.S. companies in the Asset4 universe. The portfolios are constructed on the basis of the ESG score that is available in the prior year, and where the top and bottom 20% ESG performers are assigned to a high and low portfolio, respectively. Similar portfolios are also constructed based on the scores of each of the individual ESG pillars. In the original article, portfolios are also created for an economic sustainability score (ECN).

This pillar was a part of the former Asset4 score, but is however no longer a part of the ESG scoring process of Refinitiv (2020), and is thus excluded from our analysis. We use a high-low strategy with a long position in the high portfolio and a short position in the low portfolio. To evaluate the performance of the portfolios, we use an asset pricing model, namely the Carhart model (Carhart, 1997), to see whether the high portfolio outperforms the low portfolio.

When creating the high and low ESG portfolios based on the Asset4 universe of scores, the authors of Article 2 find that neither of the portfolios of the three ESG pillars nor the combined ESG score exhibit significant alphas. This is the case for both the high portfolios, the low portfolios and the high-low portfolios. Consequently, they conclude that there are no clear differences between the portfolios based on high scores, and the ones based on low scores (Halbritter & Dorfleitner, 2015).

In our replication of the Article 2 analysis (Table 3.2), similar to the original, we find that none of the high and low portfolios for the pillars and the combined ESG score exhibit significant alphas. The two high-low portfolios based on the Environmental pillar and Governance pillar also have insignificant alphas in accordance with the article. We however experience different results when looking at the high-low portfolios based on the Social score and the combined ESG score. Here, the results of our analysis show significant negative alphas at a 5% level. This implies that the high-score portfolios underperform compared to the low-score portfolios. The Social score portfolio exhibit a negative alpha of 0.50%, while the combined ESG portfolio have a negative alpha of 0.40%.

When considering the factor coefficients, our results exhibit significant negative coefficients for the MKT factor for the ESG, Environmental and Social high-low portfolios. In the original article, this coefficient is only significant for the Social high-low portfolio. For the SMB factor, we see that the high-low portfolios based on both the ESG and Social score exhibit significant positive coefficients. This entails that both of these portfolios are more exposed to smaller companies, and are putting more bets on these. In the original article, the SMB factor is significantly positive for all high-low portfolios.

Our results point to a negative relationship between Social score and abnormal returns, and similar for the relationship between the combined ESG score and abnormal returns. We therefore conclude that the relationship between ESG and financial performance is

not completely neutral, as expressed by the article we replicate, as our results exhibit a somewhat negative relationship.

Table 3.2: Article 2 Replication

		Alpha	MKT	SMB	HML	WML	R2
ASSET4							
ESG	High	0.005	0.215*	-0.226	-0.414*	-0.266*	0.066
	Low	0.007	0.339***	-0.407*	-0.299	-0.217	0.088
	High-low	-0.004**	-0.125**	0.191*	-0.117	-0.053	0.092
ENV	High	0.003	0.172	-0.230	-0.413*	-0.280*	0.063
	Low	0.005	0.364***	-0.307	-0.433*	-0.239	0.088
	High-low	-0.003	-0.194***	0.087	0.018	-0.045	0.086
SOC	High	0.004	0.230**	-0.288	-0.443*	-0.278*	0.049
	Low	0.008	0.381***	-0.433*	-0.450*	-0.309*	0.056
	High-low	-0.005**	-0.153***	0.215**	0.001	0.027	0.109
GOV	High	0.005	0.246**	-0.285	-0.343	-0.225	0.073
	Low	0.006	0.317**	-0.277	-0.416	-0.263	0.077
	High-low	-0.003	-0.073	0.002	0.071	0.034	0.023

Note:*<0.1;**p<0.05;***p<0.01
This table presents the findings of the Article 2 replication. The sample consists of the U.S. companies in the Asset4 universe list from 2002-2011, on which we apply Carhart asset pricing regressions. The dependent variables are the monthly excess returns of the high portfolio, low portfolio and high-low portfolio. These portfolios are created based on each of the three ESG pillar scores (Environmental, Social and Governance) and for the combined ESG scores. The intercept, alpha, represents the achieved abnormal returns for each portfolio. MKT stands for the market risk premium, SMB captures the exposure to small versus big firms, HML captures exposure to high book-to-market stocks, and WML represents the momentum factor (French, 2020b).

3.1.3 Article 3

The methodology of this article focuses on negative screening based on ESG scores, and three sets of pillar scores. The sample consists of the constituents of the STOXX Europe 600 from June 2004 to October 2012. The full set of companies from this sample represents the benchmark (BM) portfolio. The first step of the screening process is to construct a rated only (RO) portfolio, which only consists of the companies where ESG scores are available during the period. As a next step, we apply further negative screens on the remaining companies. We screen based on the combined ESG score and each of the three pillar scores with cut-off rates at 5, 10, 15 and 20%. For example, for the combined ESG screened portfolio with a 5% cut-off rate, we exclude the 5% worst ESG rated companies, and are thus left with a portfolio consisting of the 95% highest rated companies. The

different ESG, Environmental, Social and Governance screened portfolios are held for a year, before they are rebalanced each January, as the scores of the companies are updated once a year but at different times. In total we have 18 portfolios, namely the BM portfolio, the RO portfolio, four ESG screened portfolios, and four screened portfolios for each of the three pillars. Within each portfolio, the stocks are assigned equal weights.

To measure the performance of the portfolios, we calculate Sharpe ratios in accordance with the article. The ratios are calculated by dividing the mean excess return by the standard deviation of the excess return of each portfolio. We then compare the BM Sharpe ratio to that of the RO portfolio and the negative screened portfolios, and then compare the RO Sharpe ratio to that of the negative screened portfolios. To make these comparisons, we make use of bootstrap tests. We test the null hypothesis that the difference between two portfolios' Sharpe ratios (SR) are equal to zero (H0 : SRi – SRj = 0) (Ledoit & Wolf, 2008).

The original article finds that the RO portfolio significantly outperforms the benchmark portfolio at a 5% significance level. This indicates that holding a portfolio of only rated stocks yields higher risk-adjusted returns. Further, the article reports higher Sharpe ratios for all screened portfolios, however, only some are significantly different from the benchmark. The article presents a significant outperformance of the benchmark for the Environmental and Social selection at the 5% cut-off, and for the Governance and Combined selection at all cut-off rates. The best-performing portfolios, as presented in the article, are the Governance screened portfolios as these portfolios outperform the benchmark at a 5% significance level for the 5 and 10% cut-off, and at a 1% significance level for the 15 and 20% cut-off. The Governance screening also significantly outperforms the RO portfolio for cut-off rates at 10 to 20% (Auer, 2016).

Table 3.3 shows the results from the replication of Article 3, with descriptive statistics of the excess returns for the various portfolios, as well as their Sharpe ratios and their significance levels. The RO portfolio generates monthly excess return of 0.536% leading to a monthly Sharpe ratio of 0.101, whereas the benchmark realizes a monthly excess return of 0.551% with a Sharpe ratio of 0.102. Contrary to the article, the RO portfolio delivers a lower Sharpe ratio than the benchmark. This does, however, not result in a significant difference between the benchmark and the RO portfolio. Hence, there is no evidence that

only investing in rated stocks yield better risk-adjusted returns as opposed to holding the benchmark portfolio.

For the Environmentally screened portfolios, our replication exhibits higher Sharpe ratios as opposed to the benchmark, however none that are significant. The article displays a positive significant difference for the 5% cut-off when compared to the benchmark. The differing results may arise because of the different ESG scores, where we employ the ones from Refinitiv and not Sustainalytics. Similar to the article, there is no indication that an investor is able to realize excess returns by employing Environmental screens as opposed to holding the RO portfolio. A similar conclusion holds true also for the Socially screened portfolios, as our replication does not exhibit significant differences in any of the Sharpe ratios, whereas the article displays a significant difference at the 5% cut-off.

The results of the Governance screening replication are perhaps the most different when compared to the article's findings. The article reports a significant outperformance of the benchmark with the 5 and 10% cut-off rate at a 5% significance level, and the 15 and 20% cut-off rate at a 1% significance level. Our results do not display a significant outperformance of the benchmark. Again, this may be the result of different ESG scores applied in the replication and the article. For the Governance screening, the article also reports a significant outperformance of the RO portfolio of the 10% through 20% cut-off portfolios. In our replication, no such relationship is found. As for the combined ESG score screening, our replication returns higher Sharpe ratios than the benchmark, but reports no significance.

Because our findings suggest no evidence that the rated portfolio significantly outperforms the benchmark, a similar conclusion as to the original Article 3 cannot be drawn. Where the article reports a significant positive relationship between return and various ESG screens, our replication suggests a neutral relationship. The differing conclusions may be the result of different ESG providers where we make use of Refinitiv's ESG score, whereas the article uses the scores from Sustainalytics.

Table 3.3: Article 3 Replication

	Descriptives				Portfolio Performance				
	Min	Max	Mean	SD	SR	Diff 1	p-value	Diff 2	p-value
Initial portfolios									
Benchmark	-19.250	20.403	0.551	5.391	0.102	-	-	-	-
Rated only	-19.353	20.147	0.536	5.287	0.101	-0.001	(0.839)	-	-
Environmental Selection									
5% cut-off	-19.882	20.709	0.561	5.473	0.102	0.000	(0.883)	0.001	(0.709)
10% cut-off	-19.683	20.628	0.591	5.469	0.108	0.006	(0.210)	0.007	(0.210)
15% cut-off	-19.601	20.590	0.589	5.463	0.108	0.006	(0.254)	0.006	(0.253)
20% cut-off	-19.460	20.754	0.584	5.487	0.106	0.004	(0.548)	0.005	(0.528)
Social Selection									
5% cut-off	-19.764	20.275	0.559	5.454	0.103	0.000	(0.891)	0.001	(0.789)
10% cut-off	-19.462	19.847	0.571	5.393	0.106	0.004	(0.428)	0.005	(0.487)
15% cut-off	-19.076	19.802	0.586	5.369	0.109	0.007	(0.315)	0.008	(0.360)
20% cut-off	-19.214	20.129	0.581	5.401	0.108	0.005	(0.465)	0.006	(0.487)
Governance Selection									
5% cut-off	-19.892	20.697	0.547	5.493	0.099	-0.003	(0.343)	-0.002	(0.637)
10% cut-off	-19.848	20.862	0.551	5.511	0.100	-0.002	(0.528)	-0.001	(0.765)
15% cut-off	-19.742	20.974	0.580	5.503	0.105	0.003	(0.491)	0.004	(0.476)
20% cut-off	-19.684	20.865	0.583	5.477	0.106	0.004	(0.372)	0.005	(0.367)
Combined Selection									
5% cut-off	-19.690	20.661	0.558	5.463	0.102	0,000	(0.991)	0.001	(0.875)
10% cut-off	-19.633	20.530	0.575	5.468	0.105	0.003	(0.515)	0.004	(0.516)
15% cut-off	-19.661	20.226	0.590	5.465	0.108	0.006	(0.301)	0.007	(0.308)
20% cut-off	-19.729	20.397	0.583	5.485	0.106	0.004	(0.532)	0.005	(0.519)

Note:*<0.1;**p<0.05;***p<0.01

This table presents the results of the replication of Article 3, based on STOXX Europe 600 from June 2004 to October 2012. The table shows descriptive statistics for the screened portfolios at all cut-off rates. The Portfolio Performance section displays the Sharpe ratios (SR) of all portfolios. *Diff 1* refers to the difference between the ESG screened portfolios and the benchmark, whereas *Diff 2* displays the difference between the screened portfolios and the rated only portfolio. Next to the difference columns, the p-values are listed.

3.1.4 Article 4

This article integrates quarterly ESG scores into a Global equal-weighted equity portfolio to establish ESG as a risk factor. In the original article, the authors use the full sample of MSCI ESG rated companies. Given that we do not have access to MSCI, we use the Asset4 full universe list for the replication. This entails a Global sample of more than 8000 companies from 2007-2017. Two portfolios that make the basis of our analysis, are constructed in January 2007 and are rebalanced every quarter until 2017. These two portfolios are comprised of the same 30 randomly selected stocks from the pool of companies, where one portfolio constitutes the benchmark and the other the ESG portfolio. Upon rebalancing, the benchmark substitutes the 20% lowest performing stocks with six new random stocks from the universe. For the ESG portfolio, we substitute the 20%

lowest performing stocks with the six stocks from the pool of companies which have the highest ESG score increase over the quarter. Portfolio returns are then calculated on a quarterly basis. The process of randomly selecting thirty stocks for the benchmark and the ESG portfolio, and then rebalancing these portfolios every quarter according to the abovementioned method, is then replicated twice. Finally we have three sets of benchmark and ESG portfolios.

To evaluate the performance of the three ESG portfolios compared to their respective benchmarks, we calculate yearly alphas, as well as Sharpe ratios for each ESG portfolio, in accordance with the article. The alphas represent excess returns for the ESG portfolios, and are calculated by subtracting the benchmark return from the ESG portfolios' return each year. For the Sharpe ratios, we calculate overall ratios per ESG portfolio for the whole period. These are calculated by subtracting the benchmark portfolio returns from the expected ESG portfolio returns, and dividing this by the cumulative standard deviation of the excess returns. To test whether ESG integration is a factor that affects the increase in alphas, we conduct a paired t-test. This test compares the cumulative ESG portfolio returns with the cumulative returns of the benchmark portfolios, to see if there is a significant difference.

Table 3.4 displays our results from the replication of Article 4. The article reports consistent positive alpha generation throughout the time period of the analysis (Pollard et al., 2018), whereas our findings suggest a somewhat random allocation. The article also establishes that there is a significant difference in Sharpe ratios favoring the ESG portfolios over the benchmark. Our replication reports both negative and positive Sharpe ratios, and we are not able to quantify an outperformance of the ESG portfolios over the benchmarks by employing the same t-test. Hence, our conclusion differs from the one in the article. This difference might be a result of the different ESG score providers, where the article make use of the scores from MSCI whereas our scores are retrieved from Refinitiv.

Moreover, we beg the question of the trustworthiness of the applied methodology and whether it is a valid method to establish an actual relationship between ESG and financial performance. The method is quite unusual. Considering previous research on the subject of ESG investing, we cannot find that a similar methodology has been deployed. Thus, to

further test the validity of this method, we conduct the same analysis with a sample size of 50 stocks, keeping the quarterly turnover at 20% constant.

With the increased sample, we obtain significantly higher Sharpe ratios for the three ESG portfolios as compared to their benchmarks. However, as the basis for the ESG portfolio is growth in ESG score and not the ESG score itself, we argue that this methodology is not relevant for measuring the impact of ESG investing on performance. A company with a low score will have a relatively higher growth in ESG score than a company that already has achieved a high score of for example 90 out of 100 possible. The selected companies in the ESG portfolio will thus not be the best ESG performers, but rather poor performers experiencing improvements. This, in addition to the inconclusive results achieved when replicating the original sample size, leads us to excluding this article from the following meta-analysis.

Table 3.4: Article 4 Replication

	ESG Portfolio 1 Cumulative alpha (%)	ESG Portfolio 2 Cumulative alpha (%)	ESG Portfolio 3 Cumulative alpha (%)
12/31/2007	-12.16	-9.69	-4.28
12/31/2008	-10.30	-6.61	2.46
12/31/2009	23.24	19.66	39.80
12/31/2010	-1.29	-9.63	-11.74
12/31/2011	-0.15	0.78	3.17
12/31/2012	8.00	10.58	12.04
12/31/2013	2.97	7.92	4.59
12/31/2014	-13.16	0.26	-2.87
12/31/2015	3.17	6.64	7.73
12/31/2016	-16.69	3.66	-2.94
Turnover (Quarterly, %)	20	20	20
Net Increase in ESG Rating	9.85	0.14	-0.01
Sharpe Ratio	-7.73	12.73	24.28

Note:*<0.1;**p<0.05;***p<0.01
The above table exhibits the results of the replication of Article 4, which is based on the Asset4 universe list from January 2007 to January 2017. For each of the three ESG portfolios in each year of the analysis, the cumulative alpha obtained in that year is displayed. The turnover refers to the % of companies that are replaced every quarter due to poor performance. The Net Increase in ESG Rating displays the % increase in total ESG Score for the portfolio from the beginning to the end of the analysis. Lastly, the overall Sharpe ratio of the portfolios are displayed.

3.2 Meta-Analysis

In this part of the empirical analysis, we explain the methodology and findings of the meta-analysis. The basis of this meta-analysis is Article 1, 2 and 3, while Article 4 is excluded due to the methodology being deemed questionable. To investigate the reason

behind the lack of consensus among these three articles on the subject of ESG investing, we conduct a meta-study where we test three main factors, namely sample selection, time period and methodology. Each of these factors are tested separately by holding the two other factors constant, while changing the factor in focus. More specifically, we test the sample of each of the articles by holding the methodology and time period constant, while conducting the same analysis on the European, U.S. and Global markets, to see if these studies yield different results. We then test the time period factor by holding the methodology and sample of the original articles constant, while comparing the results of multiple analyses with different time periods in focus. We finally test the effect of the choice of methodology by comparing the results of analyses utilizing different methodologies, with the sample and time period held constant.

In this section, we present each of the three factors, sample, time period and methodology, along with their associated findings. We then discuss how the choice of ESG score provider might have an effect on the obtained results, before summarizing the main findings.

3.2.1 Sample

The articles replicated in this thesis are all based on different samples, where Article 1 uses a Global sample, Article 2 looks at the U.S. market, whereas Article 3 looks at the European market. To investigate whether the sample of stocks is a contributing factor in determining the relationship between ESG and financial performance, we conduct a meta-analysis holding the methodologies and time periods constant while controlling for the sample. Given that the method of Article 1 and 2 are similar, additional regressions are only run for the methods of Article 1 and 3.

Table 3.5 displays the results of the meta-analysis of Article 1, where the time period 2008-2018 and the methodology is held constant. As is clear from the table, the Global sample seems to deliver the most negative relationship between ESG and abnormal returns. The Global sample exhibits significantly negative alphas for the equal-weighted portfolios. These alphas are all significant at a 1% level expect for model (2) and (3) in the decile portfolio, that are significant at a 5% level. Looking at the U.S. sample, it is clear that we obtain somewhat similar, however more neutral, results. For the equal-weighted portfolios, only the decile portfolio displays significant negative alphas for models (1) and (3) at

a 10% significance level, and a 5% significance level for models (4) through (6). As for the equal-weighted quartile portfolio, only models (4) through (6) display negative relationships between ESG and financial performance on a 10% significance level. As is also the case for the Global sample, both value-weighted portfolios display negative alphas, however none that are significant. The lack of significance for the value-weighted portfolios is most likely explained by the fact that smaller companies, who tend to deliver larger abnormal returns (Banz, 1981), are given smaller weights with this method.

As for the European sample, it is evident from Table 3.5 that the results are more neutral than that of the Global and U.S. samples. Neither of the decile portfolios exhibit any significant alphas. As for the quartile portfolios, only the model (4) alphas are significant and negative, while the other alphas are now insignificant. These differing results are similar to those found in the article by Breedt et al. (2018), where the U.S. sample seems to deliver poorer results than the European sample.

Table 3.5: Sample Analysis: Global vs. USA vs. Europe (2008-2018)

	(1)	(2)	(3)	(4)	(5)	(6)
Global						
Equal-Weighted 25%	-0.43***	-0.38***	-0.39***	-0.49***	-0.49***	-0.49***
Equal-Weighted 10%	-0.70***	-0.63**	-0.64**	-0.73***	-0.73***	-0.73***
Value-Weighted 25%	-0.13	-0.06	-0.08	-0.16	-0.16	-0.16
Value-Weighted 10%	-0.15	-0.06	-0.07	-0.15	-0.15	-0.16
USA						
Equal-Weighted 25%	-0.16	-0.10	0.00	-0.24*	-0.25*	-0.25*
Equal-Weighted 10%	-0.37*	-0.29	-0.32*	-0.5**	-0.5**	-0.47**
Value-Weighted 25%	-0.17	-0.02	-0.04	-0.17	-0.18	-0.19
Value-Weighted 10%	-0.36	-0.21	-0.25	-0.34	-0.33	-0.34
Europe						
Equal-Weighted 25%	-0.20	-0.20	-0.21	-0.35**	-0.46	-0.40
Equal-Weighted 10%	0.02	0.04	0.07	-0.16	-0.21	-0.15
Value-Weighted 25%	-0.22	-0.22	-0.22	-0.37**	-0.44	-0.39
Value-Weighted 10%	0.00	0.00	0.02	-0.14	-0.19	-0.11

Note:*<0.1;**p<0.05;***p<0.01

This table shows the results of the first sample meta-analysis, where the time period 2008-2018, as well as the methodology of Article 1, is held constant. The six utilized models are the same as the ones presented in Table 3.1. There are four different portfolios for each sample: two quartile portfolios (one equal-weighted and one value-weighted) and two decile portfolios (one equal-weighted and one value-weighted). The table displays the results from the original replication of Article 1 as well as the results obtained when replacing the sample. The coefficients represent the intercept of the regressions, i.e. the alphas which captures the achieved abnormal returns for the portfolios.

We conduct a similar meta-analysis with the time period and method of Article 3. The results shown in Table 3.6 signals a similar trend as to that found when conducting the

meta-analysis with the method from Article 1 (Table 3.5). Again, we see that it is the Global sample that exhibits the most negative results, where the benchmark significantly outperforms all screened portfolios at a 1 and 5% significance level. Thus, this implies that an investor will realize higher returns when holding the benchmark as opposed to applying various ESG screens to a portfolio. The results further indicate that an investor is also better off by investing in the RO portfolio as opposed to applying additional ESG screens. This holds true for the Environmental screening with a cut-off rate of 15 and 20%, the Governance screening at a 5, 10, and 20% cut-off, as well as for the ESG screening at a 20% cut-off. Screening the portfolios lead to a reduction in the number of companies, which might impact the diversification potential, and thus might affect the achieved risk-adjusted returns (Auer, 2016).

For the U.S. sample, all alphas are negative but not all are significant. Notably, the benchmark does not outperform the rated only portfolio, as was the case for the Global sample. However, the benchmark slightly outperforms the socially screened portfolio at a 10% significance level for the 15 and 20% cut-off, the Governance screened portfolios for the 10, 15 and 20% cut-offs, and for the ESG screened portfolios at a 15 and 20% cut-off. Once again, this indicates that an investor is better off financially when holding the benchmark as opposed to applying the abovementioned screening cut-off rates. As was the case for the Global sample, we also see that the RO portfolio in the U.S. slightly outperforms some of the screened portfolios, namely the Governance and ESG screened portfolios with a 15 and 20% cut-off at a 10% significance level. Once again, this outperformance may be explained by the reduced number of companies that comes with negative screening, which might impact the diversification potential (Auer, 2016).

The European sample, unlike the other samples, does not show evidence that the benchmark outperforms the screened portfolios. On the contrary, the screened portfolios mostly deliver higher Sharpe ratios as opposed to the benchmark, however none that are significant. The same reasoning holds true also for the RO portfolio, as the results clearly indicate that applying various ESG screens yields higher Sharpe ratios. However, as earlier, the differences are not significant. Hence, we are not able to infer any direct relationship between ESG and abnormal returns for the European sample. Nevertheless, the findings are in accordance with current research on the topic, suggesting that the U.S. is behind

Europe when it comes to integrating ESG standards (Gonçalves, 2020).

Table 3.6: Sample Analysis: Global vs. USA vs. Europe (2004-2012)

	Portfolio Performance Global			Portfolio Performance USA			Portfolio Performance Europe		
	SR	Diff 1	Diff 2	SR	Diff 1	Diff 2	SR	Diff 1	Diff 2
Initial portfolios									
Benchmark	0.223	-	-	0.162	-	-	0.102	-	-
Rated only	0.157	-0.066***	-	0.139	-0.023	-	0.100	-0.002	-
Environmental Selection									
5% cut-off	0.149	-0.074***	-0.008	0.136	-0.026	-0.002	0.103	0.001	0.003
10% cut-off	0.148	-0.075***	-0.009	0.139	-0.024	0,000	0.108	0.006	0.008
15% cut-off	0.146	-0.077***	-0.011*	0.137	-0.025	-0.002	0.108	0.006	0.008
20% cut-off	0.145	-0.078***	-0.012*	0.138	-0.025	-0.002	0.106	0.004	0.006
Social Selection									
5% cut-off	0.157	-0.065**	0.000	0.138	-0.024	-0.001	0.103	0.001	0.003
10% cut-off	0.156	-0.066**	-0.001	0.138	-0.024	-0.001	0.106	0.004	0.006
15% cut-off	0.158	-0.065**	0.001	0.135	-0.027*	-0.004	0.109	0.007	0.009
20% cut-off	0.156	-0.067**	-0.001	0.134	-0.028*	-0.005	0.108	0.006	0.008
Governance Selection									
5% cut-off	0.155	-0.068***	-0.002*	0.138	-0.024	-0.001	0.100	-0.002	0.000
10% cut-off	0.152	-0.070***	-0.005**	0.134	-0.028*	-0.004	0.100	-0.002	0.000
15% cut-off	0.152	-0.070***	-0.005	0.134	-0.028*	-0.005*	0.105	0.003	0.005
20% cut-off	0.153	-0.070***	-0.004*	0.131	-0.031*	-0.007*	0.107	0.005	0.007
Combined Selection									
5% cut-off	0.154	-0.069***	-0.003	0.137	-0.025	-0.001	0.102	0.000	0.002
10% cut-off	0.153	-0.070**	-0.004	0.137	-0.025	-0.002	0.105	0.003	0.005
15% cut-off	0.151	-0.072**	-0.006	0.132	-0.030*	-0.007*	0.108	0.006	0.006
20% cut-off	0.148	-0.074**	-0.009*	0.131	-0.031*	-0.008*	0.106	0.004	0.007

Note:*<0.1;**p<0.05;***p<0.01
This table exhibits results from regressions that follows the Article 3 methodology from 2004-2012, with the Global, the U.S. and the European samples.

In summary, the results indicate a clear trend where the European market seems to deliver better results than the Global and the U.S. market. Also knowing that the U.S. sample performs better than the Global, one might infer that most of the negative alpha generation for the ESG portfolios is due to the contribution of companies located outside of the U.S. and Europe. Many countries in Asia, Africa and South America are developing countries, where focus on ESG issues are normally less apparent (RobecoSAM, 2019). Hence, these companies might be assigned lower ESG scores while generating high returns. If these companies are removed as a result of the cut-off rate, this might affect the relative performance of the screened portfolios negatively, as these companies are still included in the benchmark.

3.2.2 Time Period

The main articles of our analysis focus on three different markets, and with three different time periods. While Article 1 is relatively new and focuses on the years from 2008 to 2018,

Article 2 and 3 focus on 2002-2011 and 2004-2012 respectively. In order to control for the time period factor, we run additional regressions where we hold the methodologies and samples of each of the three articles constant, while changing the time periods in focus.

For Article 1, we still focus on the Global market and use the same six asset pricing models as in the original analysis, but additionally run regressions on the time periods of the two other articles, i.e. 2002-2011 and 2004-2012. Our initial results for Article 1 exhibit negative and significant alphas for the equal-weighted portfolios, while the value-weighted portfolios are negative but insignificant (see Table 3.7). The lack of significance for the value-weighted portfolios might be explained by the tendency of smaller firms to have more abnormal returns (Banz, 1981), and in the value-weighted portfolios these smaller firms are assigned smaller weights. When considering the time period for Article 2, 2002-2011, we see that in addition to the equal-weighted alphas being significant, we now also experience significant negative alphas for the value-weighted portfolios. This means that even when the smaller companies, that often generate more alpha are given smaller weights, the high-ESG portfolio still experience significant negative abnormal returns. The same conclusion is drawn when controlling for the time period of Article 3, 2004-2012. While we experience neutral results for value-weighted portfolios in 2008-2018, we now find a negative result when looking at the years from 2004-2012. This indicates that, while holding a high-low ESG portfolio might have led to negative abnormal returns previously, in recent years, a focus on ESG does not significantly affect the expected abnormal returns. This applies for value-weighted portfolios.

When controlling for the time period factor in the Article 2 analysis, we again experience more significant negative results in previous years compared to more recent years. The original Article 2 focuses on the U.S. market from 2002-2011 and deploy the Carhart asset pricing model. In the time period analysis we hold the method and sample constant, but instead consider the years 2008-2018. In the original analysis of Article 2, we find that value-weighted high-low portfolios exhibit insignificant alphas for the Environmental pillar and Governance pillar, while for the Social pillar and the combined ESG portfolio we experience negative significant alphas (see Table A1.1 in Appendix). When running the regression for the more recent time period, the results exhibit insignificant alphas for all the three pillar portfolios, and same for the combined ESG portfolio. This again implies

that although ESG-focused investing might have negatively impacted expected excess returns previously, in more recent years, no negative effect can be found.

In the replication of Article 3, we find that there are no significant differences between the benchmark and the negatively screened portfolios. The same conclusion applies for the differences between the screened portfolios and the rated only (RO) portfolio (See Table A1.2 in Appendix). When instead looking at the period 2008-2018, we find that excluding the companies with the 5% lowest Environmental pillar scores leads to a significantly higher Sharpe ratio compared to the RO portfolio. This implies that, when screening for the worst performers within the Environmental indicators, an investor can obtain significantly higher risk-adjusted returns. Again, we see that we achieve more positive results when focusing the analysis on a more recent time period.

Table 3.7: Time Period Analysis: 2008-2018 vs. 2004-2012 vs. 2002-2011 (Global)

	(1)	(2)	(3)	(4)	(5)	(6)
2008-2018						
Equal-Weighted 25%	-0.43***	-0.38***	-0.39***	-0.49***	-0.49***	-0.49***
Equal-Weighted 10%	-0.70***	-0.63**	-0.64**	-0.73***	-0.73***	-0.73***
Value-Weighted 25 %	-0.13	-0.06	-0.08	-0.16	-0.16	-0.16
Value-Weighted 10%	-0.15	-0.06	-0.07	-0.15	-0.15	-0.16
2004-2012						
Equal-Weighted 25%	-0.53***	-0.54***	-0.56***	-0.63***	-0.63***	-0.63***
Equal-Weighted 10%	-0.62***	-0.64***	-0.65***	-0.71***	-0.71***	-0.71***
Value-Weighted 25 %	-0.34**	-0.37***	-0.40***	-0.46***	-0.46***	-0.45***
Value-Weighted 10%	-0.30*	-0.32**	-0.35**	-0.44***	-0.44***	-0.44***
2002-2011						
Equal-Weighted 25%	-0.63***	-0.61***	-0.61***	-0.69***	-0.69***	-0.72***
Equal-Weighted 10%	-0.81***	-0.78***	-0.77***	-0.81***	-0.81***	-0.84***
Value-Weighted 25 %	-0.46***	-0.45***	-0.44***	-0.45***	-0.46***	-0.45***
Value-Weighted 10%	-0.50***	-0.41***	-0.39*	-0.35**	-0.36**	-0.36**

Note:*<0.1;**p<0.05;***p<0.01

This table exhibits results from regressions that follows the Article 1 methodology with the Global sample (Asset4 universe list). The six types of regression models are the same as in Table 3.1. In addition to the original time period of Article 1 (2008-2018), we run the same regressions on the time periods of Article 2 (2002-2011) and Article 3 (2004-2012). There are four different portfolios for each time period: two quartile portfolios (one equal-weighted and one value-weighted) and two decile portfolios (one equal-weighted and one value-weighted). The coefficients represent the intercepts of the regressions, i.e. the alphas, which captures the achieved abnormal returns for the portfolios.

In conclusion, for all three methodologies and samples, we find that focusing on ESG scores when constructing portfolios has been more beneficial in recent years compared to previous years. For the Global market, we experience a shift from a negative to a neutral

relationship between ESG-investing and excess returns in value-weighted portfolios, when moving from 2002-2012 to 2008-2018. For the U.S. market, we experience the same shift from significantly negative to neutral abnormal returns, when comparing value-weighted portfolios in 2002-2011 with 2008-2018. For the European market, moving the focus of the analysis from 2004-2012 to 2008-2018, leads to a shift from a neutral result to a more positive result. In the original time period (2004-2012), we find no differences in risk-adjusted returns when comparing ESG-screened portfolios to a benchmark. While for 2008-2018, we obtain a significant positive improvement in risk-adjusted returns when operating with a cut-off rate of 5% for the Environmental pillar score. All three time period analyses support the implication that ESG-focused investing is more beneficial, or at least less disadvantageous, in recent years compared to previously. This is also in line with earlier research. Kell (2018) states that the first studies presenting a positive relationship between ESG performance and financial performance were published around 2013-2014, and that following this, growth of ESG investing accelerated.

3.2.3 Methodology

The third factor tested in the meta-analysis is the methodology. In this part, we are interested in investigating whether different methodologies affect the relationship between ESG and abnormal returns. Hence, we hold the sample and time period constant while controlling for the methodologies applied in the different articles. As was the case for the sample analysis, given the similarities between Article 1 and 2, we only control for the methodologies described in Article 1 and Article 3.

For the Global sample from Article 1, we hold the time period from 2008-2018 constant, but now change the methodology to that of Article 3. Hence, we construct a benchmark portfolio comprised of the full universe of stocks, as well as a RO portfolio. We also construct various negatively screened ESG portfolios based on the three pillar scores and the combined ESG score. These portfolios are equal-weighted and are rebalanced every year. From the equal-weighted section of Table 3.8, it is clear that the results are similar to the ones we obtain using the methodology of Article 1. The benchmark significantly outperforms both the RO portfolio as well as all other screened portfolios at all cut-off rates. This outperformance ranges from 4.6 to 5.1%, indicating the expected excess return an investor can achieve by holding the benchmark as opposed to the screened portfolios.

Further, the results also indicate that holding the RO portfolio, as compared to the screened portfolios, leads to higher Sharpe ratios and thus better risk-adjusted returns. However, this is only statistically significant for the Governance screened portfolios at a 5-15% cut-off. Once more, the results imply that an ESG focus does not result in positive abnormal returns. Hence, when holding equal-weighted portfolios with a negative screening methodology, we obtain similar results as to the high-low equal-weighted portfolios from Article 1.

Table 3.8: Methodology Analysis: Method of Article 3 applied on the Global sample, 2008-2018

	Equal-Weighted			**Value-Weighted**		
	SR	Diff 1	Diff 2	SR	Diff 1	Diff 2
Initial portfolios						
Benchmark	0.191	-	-	0.165	-	-
Rated only	0.145	-0.046***	-	0.169	0.004	-
Environmental Selection						
5% cut-off	0.141	-0.050**	-0.003	0.171	0.006	0.002
10% cut-off	0.142	-0.049**	-0.003	0.159	-0.006	-0.010
15% cut-off	0.142	-0.049**	-0.003	0.156	-0.009	-0.013
20% cut-off	0.142	-0.049**	-0.003	0.151	-0.014	-0.018
Social Selection						
5% cut-off	0.143	-0.047**	-0.001	0.167	0.002	-0.002
10% cut-off	0.143	-0.048**	-0.002	0.167	0.002	-0.002
15% cut-off	0.144	-0.047*	-0.001	0.166	0.001	-0.003
20% cut-off	0.143	-0.048*	-0.001	0.169	0.004	0.000
Governance Selection						
5% cut-off	0.142	-0.049***	-0.003**	0.170	0.005	0.001
10% cut-off	0.141	-0.050**	-0.004**	0.172	0.007	0.003
15% cut-off	0.140	-0.051**	-0.005*	0.171	0.006	0.002
20% cut-off	0.140	-0.051***	-0.004	0.177	0.012	0.008
Combined Selection						
5% cut-off	0.141	-0.050**	-0.003	0.166	0.001	-0.003
10% cut-off	0.142	-0.049**	-0.003	0.164	-0.001	-0.005
15% cut-off	0.142	-0.049**	-0.003	0.166	0.001	-0.003
20% cut-off	0.142	-0.049**	-0.003	0.166	0.001	-0.003

Note:*<0.1;**p<0.05;***p<0.01
This table exhibits the results of the methodology of Article 3 applied on the Global sample in the time period of 2008-2018, as well as the results obtained with value-weighted portfolios.

For the U.S. sample from Article 2, we construct portfolios in accordance with the methodology of Article 3, as described above. In this analysis (see Table A2.1 in the Appendix), there is no significant difference between the benchmark and the RO portfolio, hence we cannot conclude that holding the benchmark yields better returns, contrary to

above. This holds true also for the Environmentally screened portfolios at all cut-off rates, as well as for the Social and Governance screened portfolios at a 5% cut-off and for the ESG screened portfolios at a 5 and 10% cut-off. At the remaining cut-off rates however, the benchmark delivers significantly higher Sharpe ratios. The results, once more, is somewhat comparable to the high-low equal-weighted portfolios found in the replication of Article 2. There, we find a significantly negative relationship between the combined ESG score and return as well as for the Social score and return. For the RO portfolio, the results further suggest that an investor can generate higher returns by holding the RO portfolio as compared to applying negative screens at various cut-off rates. This applies for the Governance and Combined screens for the 15 and 20% cut-off. Again, this might be explained by the possible loss of diversification potential that a smaller portfolio might entail (Auer, 2016).

From the above discussion, there seems to be little difference between conducting a high-low strategy as opposed to a negative screening strategy when holding equal-weight portfolios for the Global and the U.S. sample. Both strategies yield similar results as to explaining the relationship between ESG and financial performance. Hence, we cannot conclude that the choice between these two methods significantly impacts the obtained conclusion concerning the relationship between ESG and financial performance, when holding the sample and time period constant.

To investigate the importance of the choice of weight-allocation strategy, we choose to perform an additional analysis. Once again, we use the Global sample with the methodology of Article 3, but we now construct value-weighted portfolios rather than equal-weighted. The Global sample is utilized as it has generated significantly negative alphas throughout the analysis. From the results of the value-weighted section displayed in Table 3.8, it is evident that by constructing value-weighted portfolios, all former significance levels now disappear. The benchmark no longer outperforms the RO portfolio, nor does it outperform any of the screened portfolios. On the contrary, the Social and Governance screened portfolios now deliver higher Sharpe ratios than the benchmark. These are, however, not significant. Hence, the results are again similar to the ones obtained with the replication of Article 1, where only the equal-weighted portfolios display significantly negative alphas. Thus, also with a value-weighted allocation for the Global sample, we

obtain similar results when comparing the high-low strategy and the negative screening strategy. This analysis therefore supports the finding that equal-weighted portfolios yield more negative alphas than value-weighted portfolios.

Now looking at the European market in the form of STOXX Europe 600 from 2004-2012, we change the methodology to that of Article 1. From Table A2.2 available in the Appendix, we see that all alphas for the equal-weighted portfolios are positive. These are however not significant, and hence we cannot infer a positive relationship. Similarly, for the value-weighted decile portfolio, none of the negative alphas are significant and hence a relationship cannot be inferred. For the value-weighted quartile portfolio however, the results now exhibit significantly negative alphas. For again to assess the importance of weight allocation, we perform an additional analysis using the methodology of Article 3 but with value-weighted portfolios. Contrary to above, we now obtain neutral results. For the European sample with value-weighted portfolios, conducting a high-low strategy yields disadvantageous results when compared to a negative screening strategy.

In summary, the above discussion indicates that the relationship between ESG and performance and the significance of the alphas, greatly depends on how the portfolio weights are allocated. It also indicates that the choice of methodology is less important in explaining the relationship between ESG and financial performance for the U.S. and the Global sample, as the high-low and the negative screening strategy result in somewhat similar conclusions. For the European sample however, performing a negative screening strategy yields better results. As for the weight allocation, the differing conclusions might be explained by the tendency of small firms to deliver greater abnormal returns (Banz, 1981). With an equal-weighted strategy, these companies are given bigger weights and hence have a bigger impact on portfolio returns. With a value-weighted strategy however, these abnormal returns are given smaller weights and thus contribute less to the overall portfolio performance.

3.2.4 ESG Score Provider

In this part of the empirical analysis, we consider the effects the choice of ESG score provider has on the drawn conclusion concerning the relationship between ESG investing and financial performance. As previously discussed, we use the Refinitiv ESG scores in

all four replications and again in the meta-analysis, due to limited access to ESG score data. The same score provider is used in the original Article 1 and 2, however, in Article 3 Sustainalytics scores are utilized, while Article 4 uses MSCI scores. In the original results of both Article 3 and 4, the authors find positive relationships between ESG-focused investing and financial performance. In our replications where the ESG scores are replaced with the Refinitiv scores, however, we find that the relationship is insignificant and thus neutral rather than positive. Considering that we follow the same methodology and use the same samples as the articles, it might be that the use of a different ESG score provider in the analyses has an impact on the achieved results.

Labella et al. (2019) highlights how various ESG providers view ESG-related issues differently. For example, the scores from Refinitiv and MSCI differ greatly and thus illustrates how investment strategies are influenced by the provider used. It is further mentioned that the correlation between the ESG scores set by different providers is 0.40, as opposed to credit rating scores for which the correlation is strong across different providers at 0.90. This emphasizes the difficulty of creating sound ESG investment strategies, as ESG considerations might not be properly reflected in the stock prices (Labella et al., 2019). This can explain why we obtain different results than the original Article 3 and 4 when using a different score provider. Furthermore, this lack of correlation between score providers might help explain the lack of consensus in the research on this topic. However, due to not having access to data from other providers, we cannot empirically test the impact of using different ESG score providers.

3.2.5 Meta-Analysis Summary

In this section, we have presented the findings of the conducted meta-analysis. Further, we have discussed the potential impact the choice of ESG score provider has on achieved results, when exploring the relationship between ESG and financial performance. From the sample analysis, we find that models involving a Global sample leads to significantly more negative results than when deploying a European sample. Also the results obtained from the U.S. sample exhibits more significant negative results than the European sample, although not to the same degree as the Global sample. This can imply that there is less focus on ESG standards in countries outside of the U.S. and Europe. As presented in the *Country ESG Ranking Update* (RobecoSAM, 2019), this is particularly prevalent in

countries located in Africa, South America and Asia. Hence, if companies with poor ESG scores from these countries are generating high returns, this might negatively impact the relative performance of the ESG portfolios.

In the time period analysis of the meta-study, we find that focusing a model towards recent years, i.e. 2008-2018, leads to less negative results than when an older sample is deployed (2002-2012). This applies for all three markets, namely the Global, the U.S., and the European market. This implies that the relationship between ESG and financial performance has become increasingly more positive in recent years as opposed to before. This is in accordance with the before-mentioned article by Kell (2018), highlighting that the ESG focus has accelerated in recent years.

From the analysis focused on the methodology factor, we find evidence that the choice of weight allocation for the portfolios impact the achieved results. Both when estimating alphas of a high-low strategy, and when calculating Sharpe ratios of ESG screened portfolios, we find that applying equal weights to the portfolio stocks yields significantly more negative results than when allocating weights based on the market value of the stocks. This applies for both the Global and the U.S. sample. This can be explained by the fact that when applying value-based weights, smaller companies, that often yield more abnormal returns (Banz, 1981), are assigned smaller weights. Accordingly, these abnormal returns have a lesser impact on the portfolio returns, than they would have in an equal-weighted portfolio. For the European sample, we however find no difference between value-weighted and equal-weighted portfolios when applying the negative screening methodology.

When considering the use of the negative screening methodology versus the high-low ESG score strategy, both methods yield similar results for the Global and the U.S. samples. The same is the case for the European sample when constructing equal-weighted portfolios. However, for the European sample when allocating weights based on value, we reach a neutral conclusion with the negative screening methodology, while when deploying the high-low strategy, the results exhibit negative alphas for the quartile portfolio. Thus, for the European sample, the choice between a negative screening strategy and a high-low strategy, impacts the achieved conclusion with respect to the link between ESG and financial performance. With that being said, the different conclusion does not necessarily

imply contradictory results, since the two portfolios are constructed based on vastly different terms.

Lastly, in addition to the three factors tested in the meta-analysis, it is also worth considering the impact the choice of ESG score provider has on explaining the relationship between ESG and return. The providers use different data and methods to construct their rating systems, and hence might assign different scores to the same companies. This can lead to varying results when estimating the relationship between ESG and financial performance, depending on which provider is deployed.

4 Conclusion

The aim of our thesis is to identify the main drivers behind the ambiguous research on the subject of ESG investing and financial performance. We replicate four articles that exhibit different links between ESG and return from a Global, a U.S., and a European perspective. To identify the drivers, we conduct a meta-analysis based on the replicated articles. In the meta-analysis, we test three factors: sample selection, time period and methodology. We find that the Global sample exhibits more negative results compared to the results obtained with the U.S. and the European samples. We also find that, when analyzing data from 2002-2012, the relationship between ESG and financial performance seems to be more negative compared to a more recent time period from 2008-2018. Lastly, we find that using a value-weighted approach to portfolio construction yields better results than an equal-weighted approach from a Global and a U.S. perspective. There is, however, little difference between a high-low strategy and a negative screening strategy for these samples. Contrary, with the European sample, performing a negative screening strategy yields better results than a high-low strategy.

For investors interested in ESG-investing, the key take-away from our findings is that investing in the European market is more likely to yield positive results when applying ESG strategies. For a portfolio consisting of only global or U.S. stocks, it is beneficial to apply weights to the stocks based on market value rather than applying equal weights. For investors investing in the European market, applying an ESG screening strategy is more beneficial than holding a high-low ESG portfolio. For the purpose of untangling the disharmony among previous studies on the subject of ESG investing, our research contributes by pointing to four key factors to explain differences in results, namely sample selection, time period, methodology, and ESG score provider.

To explain the lack of consensus among researchers on the subject of ESG investing, we first replicate four articles with different conclusions. Article 1 looks at a Global sample from 2008-2018, and finds that a high-low ESG score strategy yields significant negative alphas when holding equal-weighted portfolios. Value-weighted portfolios shows no significance. In our replication of this article, we come to the same conclusion. Article 2 considers the U.S. market from 2002-2011, and finds no significant abnormal returns when

holding value-weighted high-low portfolios. Contrary to the original Article 2 findings, we find a somewhat negative relationship between ESG and financial performance. This is evident from the negative alphas obtained from the high-low portfolios constructed based on the combined ESG score and the Social score.

In the third article, which considers the European market from 2004-2012, the authors conclude that ESG focused investing is beneficial. They remark that holding a portfolio consisting of only rated companies leads to higher risk-adjusted returns as opposed to holding the benchmark. Portfolios that are screened based on the combined ESG score and the Governance score, also significantly outperforms the benchmark. Contrary to the article, we find no significant differences between the benchmark and the ESG portfolios. This might be related to the use of ESG score provider, as we deploy Refinitiv scores, while the article utilizes Sustainalytics scores. Article 4 also finds a positive relationship between ESG investing and risk-adjusted returns. In our replication, however, we find no evidence of a positive relationship. A part of the explanation for the differing conclusions, might be that we use the ESG scores from Refinitiv, while the article uses the MSCI scores. However, we consider the methodology to be highly questionable, as we obtain different and contradictory results for every new random sample.

In the meta-analysis, when controlling for the sample selection by holding the methodology and the time period constant, we find that the Global sample performs poorer than the U.S. and the European sample. Further, the European sample seems to deliver the best results out of the three. Hence, it is reasonable to believe that ESG matters are of a greater focus in the U.S. and Europe as both deliver better results when compared to the Global sample. This finding is in line with research highlighting the importance of ESG in different countries, where developing markets perform worse in terms of ESG rankings. The findings are also in accordance with research suggesting that European companies is at the forefront when it comes to integrating ESG standards into everyday business. This further implies that the European investor values such matters greater than what is the case for its U.S. counterpart. Thus, the sample selection seems to be a relevant factor when it comes to explaining the ambiguous research on the topic of ESG and financial performance.

In the time period analysis we test the effect the choice of time period has on the obtained

results. This is achieved by keeping the sample and method constant, while applying different time periods. More specifically, we compare the time period from Article 1, 2008-2018, with that of Article 2 and 3, i.e. 2002-2011 and 2004-2012. We find that when studying the more recent time period, 2008-2018, we achieve less negative results than with the older time periods. This is the case for both the European, the U.S. and the Global sample. Accordingly, time period seems to be a factor that help explain the differing conclusion among researchers. Our findings suggest that ESG investing in previous years might not have been advantageous for investors, but looking at recent years the findings suggest otherwise. If this development continues, it is likely that ESG investing will become increasingly more beneficial for investors in the coming years.

To test for the methodology of the articles, we perform analyses where the sample and time period are held constant, while controlling for the methodology. Specifically, we test the methodologies of Article 1 and 3, as the methodology of Article 2 is similar to that of Article 1. From the obtained results, we can infer that a value-weighted strategy yields better results in terms of ESG and financial performance, as opposed to an equal-weight strategy. We also find that performing a high-low strategy versus a negative screening strategy yield similar conclusions as to explaining the link between ESG and return for the U.S. and the Global sample. For the European sample, however, we find that a negative screening strategy yields better results than a high-low strategy. Our findings thus imply that another contributing factor to the lack of consensus within academia, might be the choice of methodology. Specifically, the choice of weight allocation is deemed to be of particular importance to explaining the link between ESG and financial performance.

In addition to the abovementioned factors, we discuss the impact of the choice of ESG score provider. In our models we employ only the Refinitiv scores. Not deploying the same ESG score data as the original articles we replicate, might be seen as a limitation to the study. However, by deploying the same ESG scores in the meta-analysis, and thus holding the ESG score factor constant, we can more clearly see the effects of the other factors. We also see that when the original articles use another score provider, we do not reach the same conclusions in our replications when using the Refinitiv score. Seeing that we use the same samples and methodologies, this might indicate that replacing the original score provider might affect the obtained results. Thus, our replications infer that

one of the drivers behind the contradicting research on the subject of ESG investing, is related to the ESG data utilized in the different studies.

In regards to future studies, to take this research further, it could be interesting to include developing markets into the sample analysis. These are included in the Global sample, however they are not analyzed separately. Isolating a sample of developing countries in a similar meta-study, might capture some of the difference in results between studies of Global samples and those of U.S. and European samples. In our study, we also discuss the possible impact the choice of ESG score provider has on achieved results. However, in our replications and meta-study only the scores from Refinitiv are included. It would therefore be valuable to conduct a study where data from different ESG score providers are applied. This would allow to empirically test how the choice of ESG score provider impacts achieved results concerning the link between ESG and return.

References

AQR. (2020). The Devil in HML's Details: Factors, Monthly. *AQR Capital Management*. Retrieved 28.04.2020, from https://www.aqr.com/Insights/Datasets/The-Devil-in-HMLs-Details-Factors-Monthly

Asness, C., & Frazzini, A. (2013). The Devil in HML's Details. *Journal of Portfolio Management*, *39*(4), 49–68. doi: https://doi.org/10.3905/jpm.2013.39.4.049

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

Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*, 3–18.

Breedt, A., Ciliberto, S., Gualdi, S., & Seager, P. (2018). Is ESG an Equity Factor or Just an Investment Guide. *Capital Fund Management*. Retrieved from https://www.cfm.fr/assets/ResearchPapers/2018-Is-ESG-an-equity-factor-or-just-an-investment-guide.pdf

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Clark, G., Feiner, A., & Viehs, M. (2015). From the Stockholder to the Stakeholder: How Sustainability Can Drive Financial Outperformance. *University of Oxford and Arabesque Partners*. doi: http://dx.doi.org/10.2139/ssrn.2508281

Deutsche Bundesbank. (2020). *Statistics - EURIBOR.* Retrieved 10.05.2020, from https://www.bundesbank.de/dynamic/action/en/homepage/search/statistics/744770/title-search?query=euribor

Eccles, R., & Klimenko, S. (2019). The Investor Revolution. *Harvard Business Review*. Retrieved 04.06.2020, from https://hbr.org/2019/05/the-investor-revolution

FactSet Insight. (2020). *Corporate Governance, Reputational Risk, and Why ESG Matters.* Retrieved 26.05.2020, from https://insight.factset.com/corporate-governance-reputational-risk-and-why-esg-matters-key-takeaways

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

Fama, E. F., & French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economic*, *116*(1), 1–22.

French, K. R. (2020a). *Current Research Returns.* Retrieved 02.03.2020, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

French, K. R. (2020b). *Description of Fama/French 5 Factors for Developed Markets.* Retrieved 02.03.2020, from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5developed.html

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance and Investment*, *5*(4), 210–233. doi: http://dx.doi.org/10.1080/20430795.2015.1118917

Gonçalves, P. (2020). North America trailing behind Europe on ESG engagement. *International Investment*. Retrieved 04.06.2020, from https://www.internationalinvestment.net/news/4013144/north-america-trailing-europe-esg-engagement

Halbritter, G., & Dorfleitner, G. (2015). The Wages of Social Responsibility - Where are they? A Critical Review of ESG Investing. *Review of Financial Economics*, *26*, 25–35.

Henisz, W., Koller, T., & Nuttall, R. (2019). Five ways that ESG create value. *McKinsey & Company*. Retrieved from https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/five-ways-that-esg-creates-value

Johannesen, T. A., & Tveiterås, (2019). *ESG Investments: Exploring the Impact of Sustainability on Financial Performance.* Master's thesis, Norwegian School of Economics (NHH), Bergen.

Kell, G. (2018). The Remarkable Rise of ESG. *Forbes*. Retrieved 04.06.2020, from https://www.forbes.com/sites/georgkell/2018/07/11/the-remarkable-rise-of-esg/#2edb1b816951

Labella, M., Sullivan, L., Russell, J., & Novikov, D. (2019). The Devil is in the Details: The Divergence in ESG Data and Implications for Sustainable Investing. *QS Investors*. Retrieved from https://www.qsinvestors.com/insights/the-devil-is-in-the-details-the-divergence-in-esg-data-and-implications-for-sustainable-investing

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.

Melas, D., Nagy, Z., & Kulkarni, P. (2016). Factor Investing and ESG Integration. *MSCI Research Insights*. Retrieved from https://yoursri.com/media-new/download/research_insight_factor_investing_and_esg_integration.pdf

Orlitzky, M. (2013). Corporate Social Responsibility, Noise, and Stock Market Volatility. *Academy of Management Perspectives*, *27*(3). doi: https://doi.org/10.5465/amp.2012.0097

Pastor, L., & Stambaugh, R. F. (2020). *Liquidity Factors of Pastor and Stambaugh.* Retrieved 02.03.2020, from http://finance.wharton.upenn.edu/~stambaug/liq_data_1962_2019.txt

Pollard, J., Sherwood, M., & Klobus, R. G. (2018). Establishing ESG as Risk Premia. *Journal of Investment Management*, *16*(1), 1–12.

Principles for Responsible Investment. (2020). *About the PRI.* Retrieved 26.03.2020, from https://www.unpri.org/pri/about-the-pri

Refinitiv. (2019). *Datastream: The World's Most Comprehensive Financial Historical Database.* Retrieved 26.03.2020, from https://www.refinitiv.com/content/dam/marketing/en_us/documents/fact-sheets/datastream-economic-data-macro-research-fact-sheet.pdf

Refinitiv. (2020). *Environmental, Social and Governance (ESG) Scores From Refinitiv.* Retrieved 15.05.2020, from https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf

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

RobecoSAM. (2019). *Country ESG Ranking Update - December 2019.* Retrieved from https://www.robecosam.com/media/6/f/3/6f3ec361761ea4009e97e378d65d2d91_201912-csr-final_tcm1012-22420.pdf

Schiano, S. (2018). Climate Change is Transforming Business. *Forbes*. Retrieved 29.05.2020, from https://www.forbes.com/sites/forrester/2018/10/11/climate-change-is-transforming-business/#1abd4d672f47

Appendix

A1 Time Period Analysis

Table A1.1: Time period comparison (USA): 2002-2011 vs. 2008-2018

		2002-2011	2008-2018
ESG	High	0.005	0.006
	Low	0.007	0.007
	High-low	-0.004**	-0.001
ENV	High	0.003	0.006
	Low	0.005	0.006
	High-low	-0.003	-0.001
SOC	High	0.004	0.005
	Low	0.008	0.007
	High-low	-0.005**	-0.002
GOV	High	0.005	0.006
	Low	0.006	0.008
	High-low	-0.003	-0.002

Note:*<0.1;**p<0.05;***p<0.01
This table follows the Article 2 methodology and presents alphas for value-weighted portfolios consisting of companies with the 20% highest and lowest ESG scores, as well as a high-low portfolio for each of the ESG pillars and the combined ESG score. Portfolios are constructed for U.S. companies for two time periods, 2002-2011 and 2008-2018.

Table A1.2: Time period comparison (Europe): 2008-2018 vs. 2004-2012

	Portfolio Performance 2008-2018					Portfolio Performance 2004-2012				
	SR	Diff 1	p-value	Diff 2	p-value	SR	Diff 1	p-value	Diff 2	p-value
Initial portfolios										
Benchmark	0.115	-	-	-	-	0.102	-	-	-	-
Rated only	0.111	-0.004	(0.147)	-	-	0.100	-0.002	(0.423)	-	-
Environmental Selection										
5% cut-off	0.115	0.000	(0.917)	0.004**	(0.034)	0.103	0.001	(0.878)	0.003	(0.125)
10% cut-off	0.117	0.002	(0.567)	0.006	(0.222)	0.108	0.006	(0.210)	0.008	(0.120)
15% cut-off	0.117	0.002	(0.719)	0.006	(0.312)	0.108	0.006	(0.254)	0.008	(0.164)
20% cut-off	0.116	0.001	(0.924)	0.005	(0.527)	0.106	0.004	(0.548)	0.006	(0.419)
Social Selection										
5% cut-off	0.114	-0.001	(0.744)	0.003	(0.262)	0.103	0.001	(0.889)	0.003	(0.418)
10% cut-off	0.116	0.001	(0.899)	0.005	(0.446)	0.106	0.004	(0.426)	0.006	(0.297)
15% cut-off	0.116	0.001	(0.887)	0.005	(0.543)	0.109	0.007	(0.315)	0.009	(0.250)
20% cut-off	0.115	0.000	(0.961)	0.004	(0.633)	0.108	0.006	(0.464)	0.008	(0.362)
Governance Selection										
5% cut-off	0.113	-0.002	(0.470)	0.002	(0.346)	0.100	-0.002	(0.349)	0.000	(0.601)
10% cut-off	0.113	-0.002	(0.350)	0.002	(0.491)	0.100	-0.002	(0.533)	0.000	(0.891)
15% cut-off	0.114	-0.001	(0.713)	0.003	(0.432)	0.105	0.003	(0.485)	0.005	(0.278)
20% cut-off	0.115	0.000	(0.963)	0.004	(0.382)	0.107	0.005	(0.366)	0.007	(0.186)
Combined Selection										
5% cut-off	0.115	0.000	(0.905)	0.004	(0.323)	0.102	0.000	(0.977)	0.002	(0.563)
10% cut-off	0.116	0.001	(0.816)	0.005	(0.321)	0.105	0.003	(0.499)	0.005	(0.305)
15% cut-off	0.116	0.001	(0.809)	0.005	(0.368)	0.108	0.006	(0.293)	0.006	(0.196)
20% cut-off	0.116	0.001	(0.905)	0.005	(0.525)	0.106	0.004	(0.522)	0.007	(0.378)

Note:*<0.1;**p<0.05;***p<0.01

This table exhibits results from regressions that follows the Article 3 methodology with the European sample (Europe STOXX 600). In addition to the original time period of Article 3 (2004-2012), we run the same regressions on the time period for Article 1 (2008-2018).

A2 Methodology Analysis

Table A2.1: Method of Article 3 applied on the U.S. sample, 2002-2011

	Descriptives				Portfolio Performance				
	Min	Max	Mean	SD	SR	Diff 1	p-value	Diff 2	p-value
Initial portfolios									
Benchmark	-21.498	21.916	1.067	7.063	0.151	-	-	-	-
Rated only	-21.466	21.279	0.089	6.869	0.130	-0.021	(0.124)	-	-
Environmental Selection									
5% cut-off	-21.161	21.442	0.873	6.799	0.128	-0.022	(0.219)	-0.001	(0.886)
10% cut-off	-21.097	21.279	0.873	6.753	0.129	-0.021	(0.240)	0.000	(0.966)
15% cut-off	-21.162	21.429	0.866	6.745	0.128	-0.022	(0.232)	-0.001	(0.894)
20% cut-off	-20.826	21.401	0.864	6.729	0.128	-0.022	(0.194)	-0.001	(0.886)
Social Selection									
5% cut-off	-21.546	21.324	0.884	6.897	0.128	-0.023	(0.118)	-0.001	(0.371)
10% cut-off	-21.380	21.332	0.875	6.907	0.127	-0.024*	(0.090)	-0.003	(0.108)
15% cut-off	-21.414	21.341	0.852	6.893	0.124	-0.027*	(0.091)	-0.006*	(0.073)
20% cut-off	-21.244	21.598	0.849	6.905	0.123	-0.028*	(0.060)	-0.007*	(0.079)
Governance Selection									
5% cut-off	-21.260	21.239	0.885	6.846	0.129	-0.021	(0.125)	0.000	(0.794)
10% cut-off	-21.237	20.804	0.849	6.823	0.124	-0.026*	(0.088)	-0.005*	(0.091)
15% cut-off	-21.247	20.947	0.843	6.822	0.124	-0.027*	(0.082)	-0.006*	(0.063)
20% cut-off	-21.281	20.651	0.824	6.799	0.121	-0.030*	(0.081)	-0.008*	(0.066)
Combined Selection									
5% cut-off	-21.382	21.238	0.874	6.858	0.127	-0.023	(0.101)	-0.002	(0.131)
10% cut-off	-21.356	21.275	0.869	6.852	0.127	-0.024	(0.106)	-0.003	(0.230)
15% cut-off	-21.415	21.109	0.845	6.881	0.123	-0.028*	(0.076)	-0.007*	(0.094)
20% cut-off	-21.221	21.134	0.835	6.879	0.121	-0.029*	(0.058)	-0.008*	(0.086)

Note:*<0.1;**p<0.05;***p<0.01

This table presents the results of the Article 3 methodology applied on the U.S. sample from 2002-2011. SR exhibits the Sharpe ratios, Diff 1 represents the difference between the benchmark and the screened portfolios, while Diff 2 represents the difference between the rated-only portfolio and the screened portfolios.

Table A2.2: Method of Article 1 applied on the European sample, 2004-2012

Panel A: Quartile Portfolios

	Equal-Weighted Portfolios						Value- Weighted Portfolios					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Mkt-rf	-0.03	-0.05*	-0.04	-0.01	-0.01	-0.02	-0.06**	-0.07***	-0.06**	-0.03	-0.03	-0.04
SMB		-0.54***	-0.53***	-0.51***	-0.52***	-0.52***		-0.62***	-0.62***	-0.59***	-0.59***	-0.60***
HML		0.11	0.14*	-0.01	0.01	-0.00		0.05	0.08	-0.03	-0.02	-0.03
WML			0.04		0.02	0.01			0.03		0.01	0.00
RMW				-0.14	-0.14	-0.17				-0.07	-0.07	-0.08
CMA				0.29***	0.27**	0.31***				0.30***	0.30***	0.32***
LIQ						0.04						0.02
α	0.02	0.06	0.02	0.10	0.09	0.09	-0.35**	-0.31**	-0.34***	-0.29**	-0.30**	-0.30**
N	101	101	101	101	101	101	101	101	101	101	101	101
R2	0.01	0.44	0.45	0.50	0.50	0.51	0.05	0.54	0.54	0.58	0.58	0.59
Adj. R2	0.00	0.42	0.43	0.48	0.47	0.47	0.04	0.52	0.52	0.56	0.56	0.55

Panel B: Decile Portfolios

	Equal-Weighted Portfolios						Value-Weighted Portfolios					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Mkt-rf	-0.02	-0.01	-0.01	0.03	0.03	0.02	-0.06*	-0.05	-0.05	-0.00	0.00	0.01
SMB		-0.69***	-0.68***	-0.64***	-0.64***	-0.65***		-0.73***	-0.73***	-0.69***	-0.67***	-0.66***
HML		-0.01	0.01	-0.07	-0.08	-0.09		-0.01	0.03	-0.06	-0.07	-0.06
WML			0.03		-0.01	-0.02			0.04		-0.01	0.00
RMW				-0.00	-0.00	-0.03				0.06	0.06	0.08
CMA				0.36**	0.37**	0.42**				0.46***	0.46***	0.43***
LIQ						0.05						-0.02
α	0.12	0.16	0.13	0.14	0.15	0.15	-0.23	-0.19	-0.23	-0.25	-0.24	-0.25
N	101	101	101	101	101	101	101	101	101	101	101	101
R2	0.00	0.35	0.35	0.39	0.39	0.40	0.03	0.44	0.45	0.50	0.50	0.50
Adj. R2	-0.01	0.33	0.33	0.36	0.36	0.35	0.02	0.43	0.42	0.48	0.47	0.47

Note:*<0.1;**p<0.05;***p<0.01

This table exhibits the results of the Article 1 methodology applied on the European sample from 2004-2012. The intercept, alpha, shows the abnormal returns of the high-low ESG portfolios.