University of Zurich[UZH]

Department of Banking and Finance

Centre of Competence for Sustainable Finance

The Performance and Flows of Sustainable Actively Managed Mutual Funds During the COVID-19 Crisis

Master Thesis in Business and Finance

Alberto Notaro

Full Text Version

CCSF Thesis Series no. 38 (2022)

December 2022

CCSF Thesis Series (formerly CSP/CMF Thesis Series)

The Performance and Flows of Sustainable Actively Managed Mutual Funds during the COVID-19 Crisis
Master Thesis in Banking and Finance

Alberto Notaro
Advisor: Dr. Annette Krauss
Professor: Professor Dr. Marc Chesney

Full Text Version
CCSF Thesis Series no. 38 (2022)

Zurich: University of Zurich, Department for Banking and Finance / Center of Competence for Sustainable Finance, Plattenstrasse 14, 8032 Zurich, Switzerland

Abstract

This study investigates how sustainability ratings affect the performances and flows of actively managed funds during a period of unprecedented market disruption, namely the COVID-19 crisis. This is achieved through two regression analyses on an extensive panel data. Despite some of the results with difficult interpretation, the overall research shows a positive impact of sustainability ratings on funds' performances, and a negative one on funds' flows. These findings are partially in line with the theory that wants sustainability to be a "luxury good", relinquishable during a crisis, but could at the same time represent evidence of how sustainable approaches offer potentially better performances during market crashes.

Executive Summary

In this study, research about the impact of sustainability ratings on funds' performances and flows during the COVID-19 pandemic crisis are conducted. The dataset consists of US-based actively managed mutual funds, with a focus on funds investing in Equity. The primary objectives are the following: first, to determine the abnormal returns of actively managed mutual funds and determine whether funds with a higher sustainability rating outperform poorly rated ones. Second, to analyse the net fund inflows or outflows to determine whether funds with higher sustainability ratings are more resilient than those without such ratings. The exogenous shock of the economic crisis caused by the COVID-19 pandemic serves as a context for this research: the time horizon covered in the analysis spans from the 3$^{\text{rd}}$ of February to the 29$^{\text{th}}$ of May 2020, with daily data observations. This study also allows for the testing of a common theory that consider sustainability as a "luxury good", something only sophisticated investors care about. This would imply that, in the context of a severe market shock like the one developed in the first half of 2020, average investors would not allocate their money based on environmental or social concerns, but prioritise other, more traditional objectives. After a first introduction where the relevance of this topic is addressed, a literature review is undertaken in which the most prevalent perspectives on ESG-based ratings are offered. The central part of the study focuses on the empirical methodology employed in the research, with an emphasis on data processing. Sustainability and financial data are provided by Morningstar through its proprietary platform, Morningstar Direct. Two sustainability ratings are used in the analysis, issued by Morningstar and based on Sustainalytics data: the first is the Morningstar Sustainability Rating, that allows investors to assess the relative environmental, social, and governance risks present in a portfolio and ranges from one to five "globes", and the second is the Low Carbon Designation, a label that identifies companies, portfolios, and funds with minimal exposure to fossil fuels and involved in the transition to a low-carbon economy. Financial data about funds require a considerable effort to be processed, as most of them is at share class level (a "fraction" of a fund that entitles its holders to different privileges). After building a preliminary dataset through the platform, the data are cleaned, processed, and aggregated at fund level using Python. The dependent variables of the study are also computed: on one hand, four sets of abnormal returns are calculated in order to analyse the performances of the funds, and on the other hand the percentage flows are computed as daily dollar flows over the total asset under management of each fund. The final dataset is composed of 8'618 share classes, eventually aggregated to 2'396 funds. As a last step, the regression analysis is performed, employing a Pooled OLS and a Difference-in-Difference models. The time horizon of the study is sliced into three temporal subsets to test the same models also on the Pre-Crisis, the Crash and the Recovery periods. The results do not allow to draw any definitive conclusions as their statistical significance is very weak, although they are in agreement with the articles taken as main references. Specifically, it would appear that high sustainability is a positive predictor of larger abnormal returns, while, at the same time, being related to a sharper decrease in net fund flows as a reaction to the pandemic shock.

Contents

List of Figures

List of Tables

List of Abbreviations

AUM Asset Under Management

CBOE Chicago Board Options Exchange

CSR Corporate Social Responsibility

DID Difference-in-Differences

ESG Environmental, Social and Governance

ETF Exchange Traded Fund

GFC Global Financial Crisis

LCD Low Carbon Designation

NAV Net Asset Value

NBER National Bureau of Economic Research

OVB Omitted Variable Bias

TNA Total Net Assets

WHO World Health Organization

Chapter 1

Introduction

The purpose of this study is to provide a methodical empirical analysis able to shed light on how investors' choices were affected by the sustainability rating of actively managed mutual funds during a period of significant market stress. In particular, the effects of sustainability rating on performances and flows of such funds are investigated. In this introduction, the relevance of this subject is addressed by using four major concepts: the definition and importance of exogenous shocks; the developments of sustainable finance; the COVID-19 pandemic evolution and its effects on the economy; and finally, the relevance of the active mutual funds sector.

1.1 Exogenous shocks

The study exploits a so-called "exogenous shock", namely the COVID-19 pandemic, to provide a baseline situation where virtually every agent and organisation in the market experienced a similar unexpected disruption of their activities. There is no univocal definition for such an event, but multiple literature sources, at least in the field of economics, agree on describing it as an unforeseen event with a low likelihood but a potentially high influence, originated outside of the organization, industry, or environment under examination. Just to give a few examples, according to IMF Policy Development and Review Department (2003, p. 4) exogenous shocks are characterised as “a sudden event beyond the control of the authorities that has a significant negative impact on the economy”; another definition is provided by Miklian and Hoelscher (2022), which refers to an exogenous shock “as an unpredictable and/or unexpected event not initiated by a given market, community or country that carries a significant negative impact upon that market, community or country”; last but not least, it’s possible to draw some similarity with the concept of “black swan” depicted by Taleb (2007), referring to highly improbable events with high impact. Prior economic crises had consequences that were often at least partially endogenous to corporate decisions and generally had a higher and predicted likelihood of occurring. A spontaneous comparison can be drawn between the Pandemic crisis of 2020 and the Global Financial Crisis (GFC) of 2007-08: the latter was caused by mismanagement of financial institutions, financial event causing a financial crisis and government intervention and bailouts were directed mainly towards financial institution. On the other hand, the COVID-19 pandemic may be easily associated with a “natural disaster”: the subsequent crisis was not caused by a particular sector or any structural problem, and companies and industries that suffered the most were not necessarily mismanaged. More in general, studies involving events such as the GFC may be biased by the very same events in unknown way, as they are at least somewhat endogenous to corporate decisions, are frequently subject to extensive discussion and have a significant probability of happening (e.g., political events), and/or do not strike quickly but rather take some time to unfold completely (e.g., regulatory changes). The COVID-19 crisis offers instead a more controlled and, to some extent, unique environment for this analysis: in this study, this exogeneity is used to quantify the effects of environmental, social, and governance (ESG) laws on the performance and flow of actively managed mutual funds.

1.2 Sustainable finance developments

Sustainability ratings represent the central independent variable in this study: can they drive or influence investors' decisions? Can they predict performances? If yes, by how much? Recent years have seen an increasing focus on sustainability issues, including in the financial industry. Nowadays, a variety of actions fall under the umbrella of sustainable finance and sustainable investment, from contributing money to green energy initiatives to buying stock in businesses that uphold ethical ideals or that respect good corporate governance principles and compliance rules. Governments and international institutions have recognised on many occasions the importance of sustainable finance and sustainable investing and that more needs to be done regarding these topics in order to meet the climate goals set by the Paris Agreement of 2015 and avoid a climatic catastrophe. The European Union claims that sustainable finance has a critical role to play in the world's transition to net zero by directing private wealth towards carbon-neutral initiatives (European Commission, 2020). Also professionals have claimed for quite some time that ESG efforts provide value for organisations and their shareholders. For instance, according to McKinsey's 2019 Global Survey on ESG programmes, the vast majority of CEOs and financial experts concur that ESG policies boost shareholder value (Delevingne et al., 2020). But although sustainability is a relatively well-known concept in the economy and its influence is still expanding, a market of finance that is expressly focused on producing social and environmental benefit in addition to financial gain has just lately evolved, and despite a multitude of competing reporting standards and concepts (such as the UN Principles for Responsible Investment (PRI[1]), the Global Reporting Initiative (GRI[2]), and the Social Accounting Standards Board (SASB[3])), this sector is still under-institutionalised and characterised by a lack of common terminology, consolidated financial or impact performance data sets, and minimal disclosure regulation. In corporate finance, much emphasis has been placed

[1] https://www.unpri.org/about-us/what-are-the-principles-for-responsible-investment
[2] https://www.globalreporting.org/
[3] https://www.sasb.org/standards/

on Corporate Social Responsibility (CSR), which can be seen as a business model that allows a corporation to apply policies, practices, and behaviours to maximise profit while respecting the social and environmental concerns of the community and aiming to have a net positive impact on both aspects. This concept helps organisations show their stakeholders they desire to be more environmentally, socially, and organisationally sustainable. In the financial sector instead, it is becoming increasingly popular the notion of Environmental, social, and governance criteria (ESG), a set of indicators for a company's conduct that socially responsible investors use to evaluate possible investments. When the first, disastrous wave of COVID-19 struck, ESG has been described as an "equity vaccine" (Willis, 2020), with ESG holdings appearing to have held up better than the rest and outperformed, and a very optimistic view on the future of these stocks was popular. Hale (2021) also presented arguments along these lines, expanding the analysed time-frame to the whole of 2020 with a focus on sustainable equity funds. Both CSR and ESG have been jointly used in many studies as indicators of the sustainability of companies (Yoon et al. 2018, Pollman 2019, Gerard 2019 among others), and even if this study focuses mainly on ESG, other papers referring to CSR are taken into account in chapter 2. The main sustainability ratings used for the mutual fund sector are issued by Morningstar, a leader firm in independent investment research. The company provides a wide range of online solutions and services for financial institutions, asset managers, consultants, and private customers. In addition to real-time data, Morningstar provides information and research on a wide range of investment tools, including stocks, mutual funds, and other managed products, as well as private markets, indices, futures, options, commodities, and precious metals. Its platform dedicated to data analysis is called Morningstar Direct. The two sustainability ratings issued by Morningstar are the Low Carbon Designation and the Morningstar Sustainability Rating, and they rely mostly on Sustainalytics' data, a leader in ESG ratings and research provider. In July 2020, Morningstar has successfully acquired

Sustainalytics, with the aim of further integrating ESG data with its research tools [4]. Both ratings are described more in depth in section 3.1.1.

1.3 COVID-19 first wave and its economic consequences in the US

Never before has a pandemic outbreak had such a massive impact on the worldwide financial markets, due to both the severity of the pandemic and the globalisation that characterises the current economic environment. The reactions governments, companies and institutions needed to put in place to face such an emergency disrupted dramatically the global economy, providing an exceptional opportunity to study the behaviour of investors in such harsh and unpredictable market conditions.

In the next part, the timeline of the first outbreak and its repercussions on the economy are briefly summarised. Both are relevant to this study, as the timeline allowed to establish on which time span to focus the analysis while the economic repercussions qualified the pandemic event as an external and unprecedented shock, contributing to the relevance of this study.

On the 31st of December 2019, multiple cases of pneumonia in Wuhan, China, with symptoms like shortness of breath and fever, are reported to the WHO Country Office in China. The Huanan Seafood Wholesale Market appears to be involved in all of the early instances. As cases surge and the first deaths are confirmed, the Chinese government decides to isolate the city on the 23rd of January 2020, and the following week the World Health Organization declares a global health emergency, identifying a new sort of Coronavirus as the cause of the illness. Within a short time, neighbouring countries first and western nations soon afterwards start to report cases of people infected by the same virus; on the 11th of February, the WHO named the disease they developed as COVID-19, an acronym that stands for Coronavirus

[4] `https://www.sustainalytics.com/esg-news/news-details/2020/07/05/morningstar-inc.-completes-acquisition-of-sustainalytics`

Disease 2019. The virus that causes it is named the following day as SARS-CoV-2. Despite the efforts put in place by authorities by monitoring airports and travel routes, the virus spreads all over the world by the end of February. As the number of confirmed cases in Italy increases from fewer than five to more than 150, Europe sees its first significant outbreak with Italy being in the forefront. On the 23$^{\text{rd}}$ of February, authorities shut down 11 municipalities between Lombardy and Veneto regions after a dramatic rise in cases. This date is crucial because after that, the markets will start to experience very high volatility and strong negative trends. On the 9$^{\text{th}}$ of March, the government extends containment measures to the entire country; Italy is now in lockdown, the first among western countries to adopt such severe measures. As fear rises, it pressures governments to react, and a growing number of countries adopt similar restrictions as Italy. On the 11$^{\text{th}}$ of March, the WHO officially declares that COVID-19 has become a pandemic[5]. On the 17$^{\text{th}}$ of March, the European Union prohibits non-essential travel from outside the bloc. 26 countries are now virtually closed to visitors from the rest of the world for at least 30 days, but soon after that, air transport is suspended in most regions of the world. On the 26$^{\text{th}}$ of March, with almost 1,000 fatalities and at least 81,321 confirmed cases, the United States is the pandemic's hardest-hit nation. At the time, there are more cases reported in this country than in China, Italy, or any other nation. By April 2, the pandemic had sickened more than one million people in 171 countries across six continents. By the end of April, over one million cases are confirmed in the US alone. In many of European countries, the emergency is tamed in May and the healthcare situation is back under control; governments that imposed more or less severe restrictions on businesses and private citizens begin to cautiously lift or ease them, while the pandemic still rages in North and South America and part of Asia. Improving epidemiological outlook and more effective prevention measures are driving more and more governments to a careful return to normality after the spring, foreshadowing a summer safe from COVID-19. As it is known today, more

[5] `https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020`

waves would have followed starting from the autumn 2020, but for the matter this study is going to analyse it is sufficient defining the first wave as the one occurred between February and June 2020.

The effects of the first wave on the economy were massive and, in terms of speed and magnitude of the shock, never before seen. This study focuses on US mutual funds, therefore, the effects of the pandemic on key US economic indicators are summarised below. As authorities instituted lockdowns, the first effect was a shock in the employment rate: as shown in figure 1.1, in April 2020, the unemployment rate rose by 10.3 percentage points to 14.7%. In the history of the statistics (available back to January 1948), this is both the highest rate and the biggest month-over-month rise. In April, there were 15.9 million more jobless people than in March. The dramatic increases in these metrics are a reflection of the COVID-19 pandemic's consequences and containment attempts.

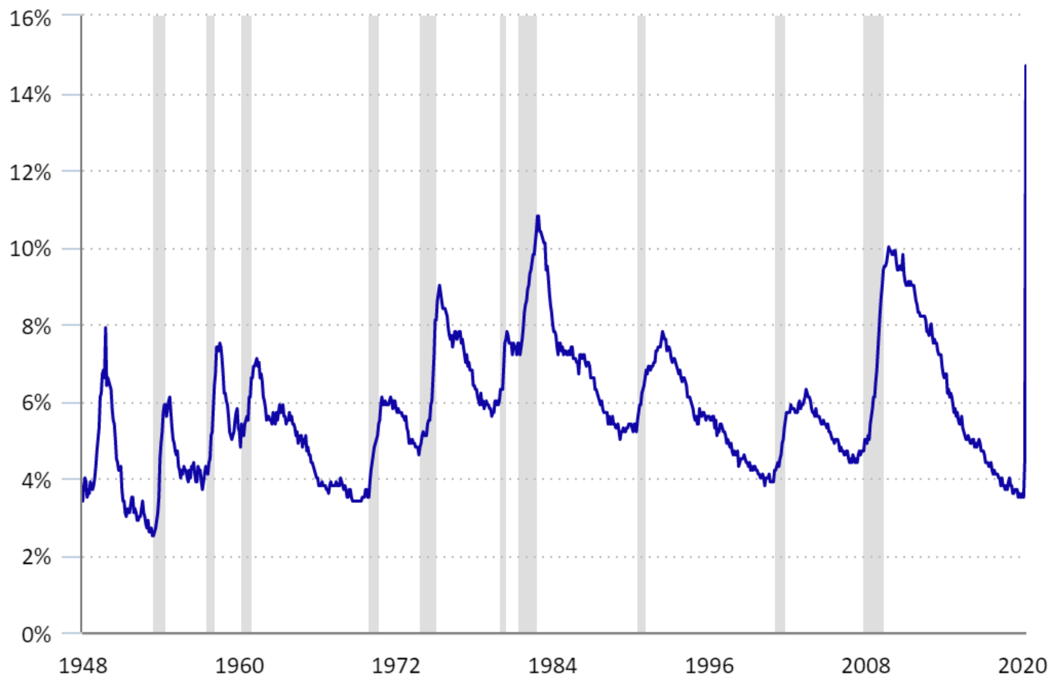


Figure 1.1 – The figure shows the monthly unemployment rate in the US. The shaded areas correspond to recessions as determined by the NBER. The months of 2020 are not shaded because when this chart was published NBER had not yet determined a start and endpoint for the recession. Source: US Bureau of Labour Statistics

The stock market's panic reaction was analysed by Ramelli and Wagner (2020) and Mazur et al. (2021), which noticed that markets moved frantically as the virus swept throughout Europe and the United States, forcing lockdowns in major countries, but the cross-section of returns, however, showed distinct patterns. Investors and analysts started to worry about high corporate debt levels and the likelihood that companies with little cash would survive. Overall, market participants anticipated that the effect of the COVID-19 health crisis would have been amplified by financial channels, and this eventually translated into the fastest 20% correction in the S&P500 ever recorded (Wells, 2020), with only 16 sessions before entering the bear market, and a massive increase in the VIX index[6] as shown in figure 1.2, almost reaching the levels of 2008-2009 GFC.

Figure 1.2 – Historical chart showing the daily level of the CBOE VIX Volatility Index back to 2000. The VIX index measures the expectation of stock market volatility over the next 30 days implied by S&P 500 index options. Source: CBOE

[6] The VIX Index is a calculation that uses the mid-quote values of call and put options on the S&P 500 in real time to provide a measure of constant, 30-day expected volatility of the U.S. stock market. It is one of the most extensively reported indicators of volatility on a worldwide scale and is closely watched as a daily market indicator by a variety of financial players.

According to Capo McCormick et al. (2021), global indebtedness, both private and public, reached historic levels, reaching 123.9% of GDP in 2020, second only to the Post-World War II levels. In the US only, domestic debt[7] jumped up to 394% in Q2 2020, due to both the sudden drop in GDP (-32.4% annualised, largest fall ever recorded) and the issuing of new debt. As shown in figure 1.3, the most significant increase as a percentage of GDP was that of Federal Government debt, associated with the extensive fiscal measures employed to address the pandemic recession (Faria e Castro, 2021): in Q2 2020 US Debt-to-GDP ratio hit a new record of 135.9%.

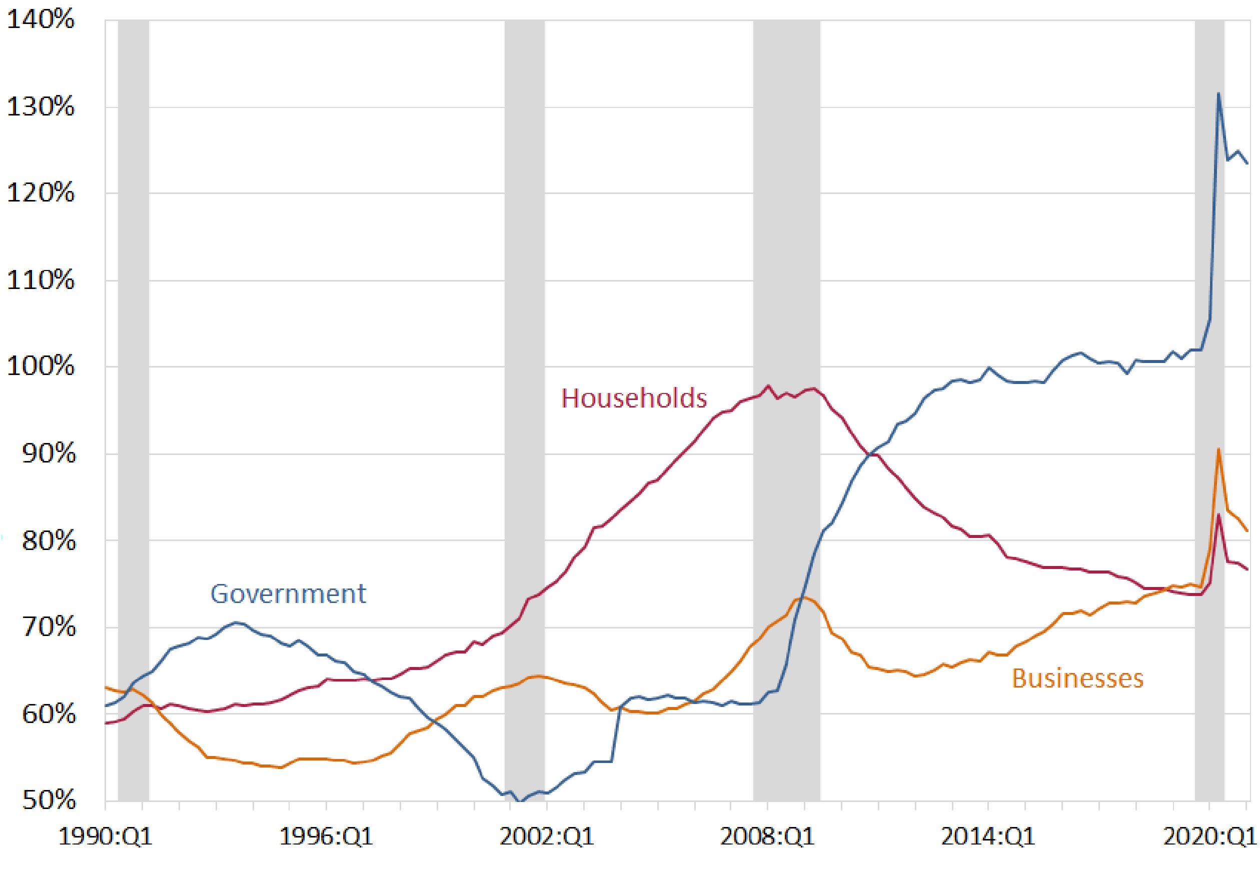


Figure 1.3 – US outstanding domestic debt of the nonfinancial sector decomposed further into its three main components: households, nonfinancial businesses and government debt. Source: Faria e Castro (2021)

[7] Total domestic debt includes both financial and nonfinancial sector debt, with the latter including households, businesses and the government.

1.4 Actively managed funds in today's economy

Finally, the relevance of the active mutual funds industry in the US economy needs to be highlighted. As underlined in the report by Bari et al. (2022), the mutual fund industry has been struggling in a very competitive landscape in the recent past. Pressures on margins, the rise of ETFs and passive investments, and increasing market consolidation resulted in a long-term outlook under stress, despite the rise in global financial markets. Economic literature, beginning with Jensen (1968) and Carhart (1997) and continuing on to Pástor and Stambaugh (2002) and Fama and French (2010), generally acknowledges that active equity mutual funds have underperformed passive benchmarks, net of costs (despite recent studies challenging such evidence, see Cremers et al. 2019). The presence of a significant underperforming sector would be unexpected given that passive funds are readily accessible to investors, and in fact multiple sources reported a considerable shift from active to passive investment. Bari et al. (2022) highlighted that from 2010 to 2019, passive investing's share of total AUM on the mutual fund market climbed from 20% to 39%, and it is expected to be 55% by 2025. The same report also anticipates that new products (such as a 30% increase in ETF funds) will partially balance the closure of around 15% of current mutual funds between now and 2025, resulting in a 6% net decline in the total number of registered fund products. Ultimately, the trends are not in favour of active management, however, despite its persistent underperformance, this business continues to manage trillions of dollars, making it a sizable market. Traditionally, active management has been a resource used by investors that tolerate this underperformance, trusting that active funds will outperform the market in crucial periods, such as economic downturns or shocks. This hypothesis has been formalised in literature by Glode (2011), which designs a model where a fund manager provides active returns that depend on the market environment. In equilibrium, the management chooses to work harder during times in which the marginal utility of consumption for investors is greater, since investors are ready to pay for this insurance. If active funds generate strong returns when investors need them the most, then the performance of these products understates their genuine potential. The COVID-19 crisis thus provides an ideal framework to test such hypotheses and more generally to check whether active management really offers a hedge against economic downturn.

The study is structured as follows. In Chapter 2, a literature review is performed, discussing major papers written around the same topics, their contribution to this study, and what further findings this study can bring to the research in the area. Chapter 3 is the most articulated: the methodology followed by this study is be explained, delving into details of the dataset composition, the data processing approach, and the regression models. Chapter 4 focuses on the empirical results achieved by the analysis on both performances and flows of the analysed funds and finally the conclusions are drawn.

Chapter 2

Literature Review

Economic research is still debating the role of ESG and sustainability in general and their relationship with financial performances and risk hedging. In the context of the COVID-19 crisis, despite the proximity of the events, a growing number of studies investigating these same topics have been published, but there is no unanimity in their findings. In this study, the evidences observed in Ramelli and Wagner (2020) regarding the market reactions are taken into account, but when broadening the readings to other studies, some research claims that increasing ESG ratings successfully enhanced the resilience and performances of stocks, while others suggest that there is no link between the two, if not even a detrimental correlation.

Garel and Petit-Romec (2021) used a dataset consisting of 1'626 large US listed firms, for which all key ESG data and economic indicators were obtained through Thomson Reuters database (EIKON). According to their study, during the COVID-19 crisis, investors have started to reward firms with responsible approaches to climate change to a greater extent, as shown by the data indicating that companies with strong environmental scores and overall better environmental policies achieve somewhat higher returns. The results of the regressions demonstrated that a one-standard deviation higher environmental score was associated with 1.41% higher stock returns during the crisis; furthermore, the environmental score's economic relevance in explaining cross-sectional returns is able to justify to a considerable extent the significance of cash holdings and long-term debt, two well-established drivers of the cross-sectional returns during the COVID-19 crisis. Because the stock market provides a glimpse of what investors anticipate for the future, this could potentially imply that businesses with sustainable policies on climate concerns would actually do better over the long run. Using a slightly different time period to conduct the analysis did not affect the results. Finally, consistently with Ramelli et al. (2018), when investors have a long-term perspective, environmentally responsible strategies appear to be rewarded to a greater degree. Albuquerque et al. (2020) found out that stocks with a focus on sustainability performed better during the crisis and in the following months with respect to other stocks. During a market decline, the performance of high-ES (Environmental and Social)[8] equities with substantial advertising is particularly positive. Even though sales were down in Q1 of 2020, businesses with high ES scores saw a rise in their operating profit margin. This was consistent with a customer loyalty mechanism, and considering the lower volatility of stocks with higher ES scores, it would be fair to assume that customer loyalty contributes, along with other factors, to increasing corporate resilience. After performing a cross-sectional regression for quarterly abnormal returns, the magnitude of the estimated coefficient reflects that a one standard deviation rise in ES ratings is connected with an average 1.8% increase in first-quarter stock returns. The difference-in-differences estimation revealed, with a remarkably high level of economic significance, that high ES-rated companies received an abnormal daily return of 0.45% relative to other firms during the studied time period, for an overall effect of 7.2%. Also Ding et al. (2020) gets to comparable conclusions. Using a global sample of over 6'000 companies from 56 economies, the study demonstrated that prior to the pandemic, companies with better CSR policies and programmes saw stronger stock price performance. These findings are consistent with the concept that CSR increases stakeholders confidence, which in turn makes workers,

[8] The paper neglects governance effects, in order to focus on the environmental and social aspects. The low correlation between governance score and environmental and social ones assures, according to the authors, that the results are not influenced by good corporate governance effects (Albuquerque et al., 2020). This approach is used also in other cited researches, and appears to be a relatively common practice.

suppliers, and consumers more susceptible to making adaptations to help the business when it is under strain. Moreover, evidence shows that the impact that COVID-19 has on the stock prices of companies that had better CSR strategies as well as higher ES scores prior to 2020 is remarkably less severe. The analysis revealed that the stock price reaction to COVID-19 for firms with a high CSR score would be on average 19% less than for firms with a poor CSR score.

On the other hand, Glossner et al. (2020) does recognise the positive impact of ESG on stocks' performance during the COVID-19 crisis, but at the same time argues that institutional investors, which seem to exacerbate market collapses whenever an exogenous tail event takes place, placed a higher premium on "hard" metrics of company resilience (such as large cash and low leverage) than they did on "soft" measures (ES issues), considered of lesser importance, and thus do not appear to have generated extra interest from institutional investors in aggregate. A more opposing perspective comes from Demers et al. (2021), which elaborated a study on the basis of Albuquerque et al. (2020), arguing that the above-mentioned analyses may have been affected by omitted variable bias. Although the findings of this research do not explain the longer-term creation of shareholder value through corporate social responsibility policies, they do indicate that companies with higher ESG scores did not have superior returns either during the pandemic-induced selloff in the first quarter of 2020 or for the entire COVID 2020 year, after industry affiliation and accounting- and market-based drivers of returns have been adequately accounted for. The firm's stock of investments in internally-generated intangible assets turned out to be particularly economically significant in explaining returns throughout both the Q1 2020 market crisis and the full 2020 year. This finding suggests that the flexibility derived from a large stock of innovative assets was more important than the firm's social capital in achieving stronger share price resilience, leading to the conclusion that ESG factors did not protect stock prices during the COVID-19 crisis, but investments in intangible assets did. Bae et al. (2021) analysed CSR data from two sources, MSCI ESG Stats and Refinitiv ESG, on a sample of 1,750 US companies. The study showed no evidence that CSR influenced stocks' returns during the pandemic-induced stock market crisis, and this observation remains true in the post-recession period and across sectors. Finally, taking a step back from the pandemic and analysing the ESG relationship with risk and return, Cornell (2021) finds that investors that lean their portfolios toward firms with strong ESG scores may be dissatisfied. Although the company and the society as a whole will benefit from it, if more and more investors choose companies with high ESG scores, the result will inevitably be higher prices and therefore lower returns. Uncertainty remains over the existence of an ESG risk element. There are no ESG ratings that are uncontroversial (Berg et al., 2022), and the sample period for which ESG data is accessible is brief. Even if ESG ratings are associated with an underlying risk issue, they cannot be applied to detect better investments. In conclusion, the increasing emphasis on ESG in stock investment may have societal advantages; however, these advantages are accompanied by reduced expected returns for investors.

As for the main topic of this study, two papers in particular explored what was the impact of COVID-19 crisis on actively managed mutual funds, namely Pástor and Vorsatz (2020) and Döttling and Kim (2020). The purpose of this study is to highlight the different approaches and conflicting results of the two previously mentioned studies, as well as to elaborate on their methods and incorporate them into the analysis. A prevalent belief in neoclassical economics wants that environmental concerns are a "luxury good", likely to be of interest primarily to individuals whose more fundamental needs for food, shelter, and survival are already addressed (Baumol et al., 1979). Both the above-mentioned studies challenge this belief and pose the same questions: is sustainability a luxury item, relinquished in the event of a crisis, or is it a necessity for modern investors? Examining the exogenous shock caused by the pandemic crisis, both articles analysed changes in the mutual fund industry in terms of both flows and performance. Pástor and Vorsatz (2020) constructed a dataset of over 3'600 US-based actively managed equity mutual funds following Pástor et al. (2015) and analysed the funds' returns net of the expense ratio (i.e., what is ultimately delivered to the clients after fees) using various indexes as benchmarks and their cash in- or outflows. They employed Morningstar Direct's daily data from the 1^{st} of January 2017 through the 30^{th} of April 2020. Results of the regression models show that during the pandemic crisis, the average active fund underperformed the S&P500 by a significant margin, and even if the subsequent recovery followed roughly the same path, at the end of April 2020, funds were on average almost 4% behind the index. This again reinforces the idea that actively managed funds are incapable of achieving results that would justify their choice over passive investing. Results varied depending on the benchmark, but even using different indexes and factor models, the fraction of funds with negative alphas ranged from 60.4% to 80.2%. At the same time, integrating the Morningstar Sustainability Rating data, the research demonstrates that funds with more "globes" and higher star ratings performed better. Investors chose funds with superior sustainability ratings and exclusion criteria when reallocating capital and recognised sustainability as a need, not a luxury, as evidenced by the fact that they continued to prioritise it throughout a significant economic and health crisis. On the flows side, a positive flow-rating relation is observed (with high statistical significance), with some evidence linking the outflows from certain funds and fire sales induced by panic in the market[9], even if the statistical analysis is less reliable. It is unclear to what degree the effect on flow is attributable to the sustainability rating rather than to the superior performance of the funds.

The second most important paper addressing the issues of this study is Döttling and Kim (2020), which, similarly to Pástor and Vorsatz (2020), gathered a sample of open-end domestic US equity mutual funds from Morningstar Direct and collected daily data on the returns and flows of such funds from January 2019 to April 2020. The hypothesis to challenge was the same: the movement in investor demand away from sustainable investments is caused by retail investors' perception of ESG as a luxury item that becomes expensive as a result of the COVID-19 financial and economic crisis. This belief would be backed by the fact that

[9] A similar exercise conducted by Falato et al. (2021), but while it studied funds in corporate bond markets, Pástor and Vorsatz (2020) focused on equity funds

retail investors are also economically powerful, accounting for approximately 61% of aggregate net assets and close to 80% of aggregate absolute net flows. Using a difference-in-difference specification with net flow on a given week as dependent variable, the study estimates the relationship between the net flows themselves and the different levels of sustainability ratings. The results of such analysis, however, disagree with the ones obtained by Pástor and Vorsatz (2020): in response to the COVID-19 shock, mutual funds with high sustainability ratings suffered a sharper decline in net fund flows and a greater likelihood of net outflows than both the average and the low-sustainability funds, erasing the relative appeal of retail flows these funds enjoyed prior to the pandemic-induced downturn. Even after the COVID period has ended and the stimulus package has been approved by US government, there is still a substantial difference in net flow. The authors consequently argued what possible channels could explain these results: as already mentioned, the shift in demand away from sustainability is the most commonly accepted hypothesis, but other relevant factors discussed are the "buying the dip" strategy, where retail investors invest in funds that have depreciated substantially in value in expectation of greater future expected returns (considered unlikely to have had an influence by the authors), or the assumption that significant structural changes were triggered by the COVID crisis and therefore driving an increasing number of investors away from sustainable investing (although even including a set of controls accounting for such structural shifts the results do not differ considerably from the baseline analysis).

An honourable mention must be made for Hartzmark and Sussman (2019), which not only focused on many of the topics related to this study but also served as both an inspiration and a reference for the two previously mentioned articles. Hartzmark and Sussman (2019), in fact, provided evidence that investors as a whole place a high value on sustainability, ruling out the possibility that investors do not care about this information or that they would punish a fund for keeping a portfolio of sustainable investments by demonstrating that they do value sustainability. According to their findings, the funds that have the greatest Globe Ratings see an increase in fund flows of more than $24 billion, while the funds that have the lowest globe ratings see a decrease in fund flows of more than $12 billion. This provides support for the idea that a sizable proportion of customers regards sustainability as a desirable quality in a business. In spite of the fact that investors are given detailed information about the percentile rank of sustainability within Morningstar categories, they largely ignore this information and instead respond to the simpler and more influential globe ratings. This is consistent with the psychological literature on categorization as well as the literature showing that the choice of how information is displayed influences financial decisions. Although there is some experimental evidence indicating that sustainability is seen as favourably forecasting future performance, the article does not find any data to suggest that funds with a high level of sustainability outperform funds with a low level. The reason why Hartzmark and Sussman (2019) is not recognised in the same manner as Pástor and Vorsatz (2020) and Döttling and Kim (2020) is the absence of the external shock caused by the pandemic; nevertheless, it represents a consistent source for what concerns data handling and empirical methodologies, being among the most distinguished and well-structured articles in this literature review.

Chapter 3

Methodology

As in every empirical study, a considerable effort is dedicated to data gathering and processing. In this chapter, the methods and sources used to set out the data needed for the empirical analysis is described. This chapter is structured as follows: in the first section the dataset composition is thoroughly described; in the second section, the details of the returns and flows, the dependent variables of the analysis, are discussed; then, the articulated operations of data processing are explained, with emphasis on the steps of data cleaning and aggregation; lastly, the regression models are formulated.

3.1 Data

3.1.1 Sustainability Ratings

This section is meant to explain what the sustainability ratings used in the analysis are and how they work. These two ratings, mentioned in the introduction, were chosen both for their being consistently and analytically structured and for their simplicity of access, as they are issued by Morningstar (based both on Sustainalytics and proprietary data), which is the main source of data in this study. Through them, it will be possible to establish a relationship between performance, flows and sustainability of the mutual funds analysed later on.

Issued for the first time in May 2018, the Low Carbon Designation recognises investment funds with minimal environmental risk in relation to the securities in their portfolio. According to Morningstar itself, "the designation is an indicator that the companies held in a portfolio are in general alignment with the transition to a low-carbon economy" (Hale, 2018). In order to receive this designation, a fund must have:

- a *Morningstar Portfolio Carbon Risk Score* below 10
  - To calculate the portfolio carbon-risk scores, Morningstar uses Sustainalytics' company carbon-risk ratings, which indicate the risk that companies face from the transition to a low-carbon economy.
  - Sustainalytics' assessment of a company's carbon risk is based on three points: (1) its own view on how the company's activities will be affected by the energetic transition, (2) the firm's exposure to carbon-related risks throughout the value chain and (3) the firm's ability to decrease its carbon risks.
  - The formula is a simple weighted average, as follows:

$$\sum_{i=1}^{n} w_i \cdot CCRR \tag{3.1}$$

    where n is the number of securities in the portfolio, w_i is the asset weight of security i and $CCRR$ is the Company Carbon Risk Rating issued by Sustainalytics.

- a *Morningstar Portfolio Fossil Fuel Involvement* less than 7% of assets
  - It represents the portfolio's percentage exposure to fossil fuels, averaged over the trailing 12 months. Companies with fossil-fuel involvement are those generating at least 5% of their revenue from thermal coal extraction, thermal coal power generation, oil and gas production, and oil and gas power generation. Companies

deriving at least 50% of their revenue from oil and gas products & services are also included.

All Morningstar Portfolio Carbon Metrics are calculated quarterly, therefore also the Low Carbon Designation updates are delivered with the same frequency. Chen et al. (2019) contends that many managers of fixed-income mutual funds misreport the creditworthiness of their assets to Morningstar in order to influence its evaluations, specifically the Stars ratings. In contrast to the credit quality of fixed-income assets, the just described measurements underpinning the Low Carbon Designation are not self-reported by fund managers but rather calculated by Morningstar based on the holdings of the funds' portfolios. Obviously, it is not possible to dismiss categorically the possibility that certain funds may misrepresent their holdings. Nevertheless, such misreporting carries with it significant legal and reputational dangers. Overall, misrepresentation does not appear to be a significant issue in this scenario. The Low Carbon Designation was also central in the research of Ceccarelli et al. (2021) which offered an important foundation for the methodology and data structuring of this study.

The second sustainability rating is the Morningstar Sustainability Rating, introduced in 2016 and designed to support investors in evaluating the relative environmental, social, and governance risks within portfolios. Since then, it has undergone several updates until reaching its current stage where the rating is considered a measure of the financially material ESG risks in a fund when compared with similar funds (Barr et al., 2021). Each fund that qualifies for a Morningstar Sustainability Rating receives a rating that ranges from 1 to 5 "globes", with a greater number of globes implying a lower ESG risk in the portfolio. The number of globes a fund obtains is compared to other funds in the same Morningstar Global Category[10]. This implies that even if two funds are in different global categories and have different definitions

[10] Morningstar assigns global categories based on a variety of aspects, among which there are: familiarity with the portfolio managers' strategy and the fund family, the Morningstar Retail category assigned to the fund, and an intention to present the most accurate picture of economic exposure possible. More information at `https://www.morningstar.com/content/dam/marketing/shared/research/methodology/860250-GlobalCategoryClassifications.pdf`

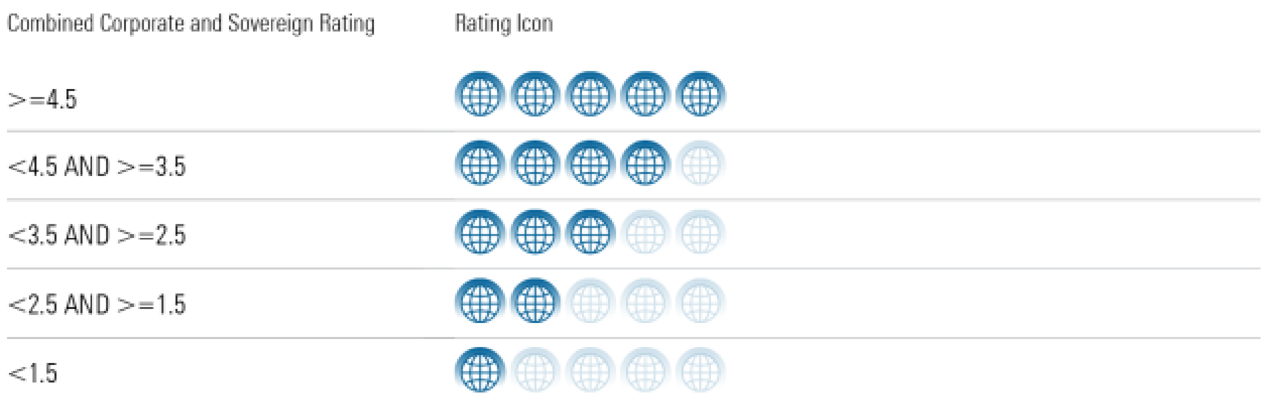


Figure 3.1 – Morningstar Sustainability Rating expressed as a the weighted sum of corporate and sovereign sustainability rating and its equivalent number of globes a fund receives. Higher Sustainability Ratings represent lower ESG risk relative to a fund's peer group. Source: Barr et al. (2021)

of what constitutes a relatively low or relatively high degree of ESG risk, one fund may have greater ESG risk than the other and yet obtain a better rating. Just like the above-mentioned designation, the Morningstar Sustainability Rating is calculated relying on Sustainalytics' data, namely:

- the ESG Risk Ratings for corporate issuers
  - they measure the magnitude of a company's unmanaged ESG risks. This is calculated through the ESG Exposure to and ESG Management of material ESG issues[11]. "Unmanaged Risk" refers to any risk posed by an ESG issue that the firm is not adequately managing or is unable to control. Corporate entities are categorised into one of five ESG risk categories based on their Unmanaged Risk scores: Negligible, Low, Medium, High, and Severe. Unlike relative risk assessments, which evaluate performance in relation to peers but may not be directly comparable to nonpeers, the ESG Risk Ratings

[11] Materiality is a wide concept that incorporates many distinct features and characteristics that exist in an organisation, such as how it controls its supply chain, its financial status, and so on. The idea of materiality in ESG research refers to the amount of priority that an organisation assigns (or should assign) to certain environmental or social problems, and it is one of the most important aspects in determining a company's commitment to addressing ESG issues (Jebe, 2019). More information at `https://www.sasb.org/standards/materiality-finder/?lang=en-us`

are deemed an absolute risk assessment, which implies that the output is comparable across sectors, industries, and subindustries.

- the Country Risk Ratings for sovereign issuers
  - They analyse the threats to the socioeconomic well-being of a sovereign entity by combining an evaluation of the government's existing stock of capital with an assessment of its ability to handle wealth in a sustainable manner. The rating combines two factors to assess risk: Wealth and ESG performance. The first is measured as the value of assets within a country, as calculated by the World Bank and it is inversely proportional to vulnerability to ESG risks (the higher the Wealth, the lower the vulnerability), while the second offers an assessment of how well a country is managing key ESG factors, whereby sound ESG performance indicates that wealth stocks are likely to improve, while weak ESG performance indicates the opposite.

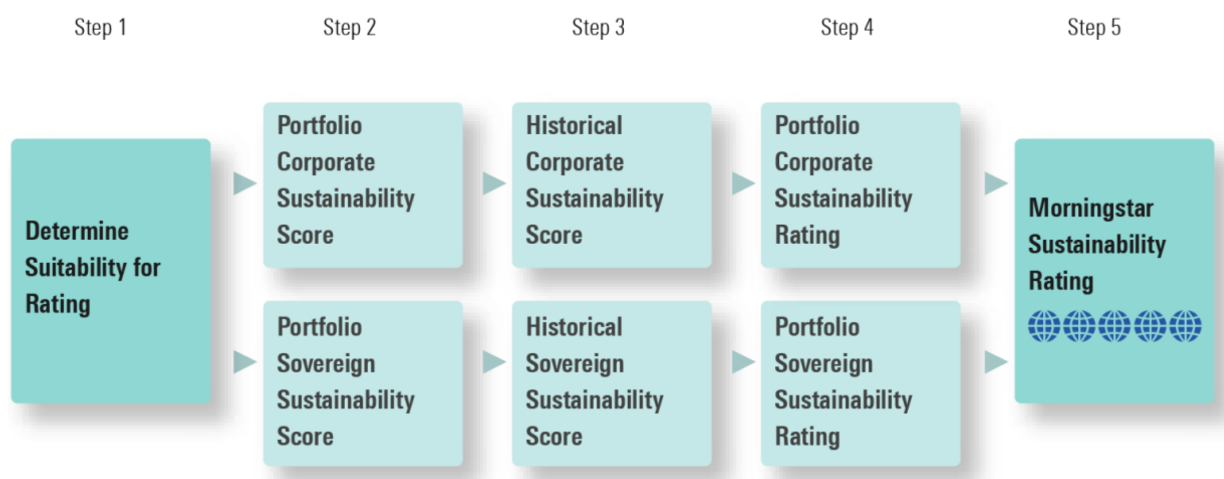


Figure 3.2 – Morningstar Sustainability Rating as the result of this five-step process, where the ESG risks and Country risks are put together. Source: Barr et al. (2021)

Once the two input data are assessed, Morningstar conducts a 5-step process illustrated in Figure 3.2, where the sustainability scores are asset-weighted average of Sustainalytics' company-level ESG Risk Rating and Sustainalytics' Country Risk Rating. The result of this process is a rating, expressed with a number from 1 to 5, and a correspondent number of globes as showed in figure 3.1. The Morningstar Sustainability Rating is released monthly, one month and six business days after the reported as-of date for company and country data from Sustainalytics.

3.1.2 Funds

Taking into account Pástor and Vorsatz (2020), Döttling and Kim (2020) and Ramelli and Wagner (2020) as well as the chronological development of the pandemic described in the introduction, the time horizon that is used for this analysis starts on February 3, 2020 and ends on May 29, 2020[12]. The choice has been made because in February 2020 the US stock market reached a peak, before entering a period of severe turbulence, while in May 2020 the market had already rebounded and the authorities around the world were lifting or at least easing the restrictions posed to their citizens. As stated previously, Morningstar Direct is the main source of data for this analysis: extracting data from this platform implies specifying first some Search Criteria, and then building an Investment List with funds that fulfil such criteria. The Search Criteria used for building the preliminary dataset are the following:

- **Type of Funds:** the funds analysed in this study are open-end, actively managed mutual funds.

  As explained in the introduction, this study aims to test, among other things, the capability of active management to adapt to a sudden market crash. Because of this, it is clear that this analysis should focus on actively managed funds. An open-end fund is a diversified portfolio of pooled investor money that can issue an infinite number of shares. These shares are valued daily in accordance with their present net asset value (NAV). Open-end funds include some mutual funds, hedge funds, and exchange-traded funds (ETFs). Since one of the main features of open-end funds is to receive a NAV on a regular (typically daily) basis, the choice of such a type of funds appeared straightforward. Moreover, the pool of open-end

[12] The first two days of February and the last two of May are not business days.

funds provided by Morningstar is vast, to say the least, and has a reasonably high number of frequently updated data points. Nevertheless, building a well-organised and structured dataset is no easy task, and the process of doing so is described later on.

- **Domicile:** the funds analysed are all domiciled in the United States.
  As in the case of most of the articles used as references, this study focuses on US-based funds. Generally, it is fair to say that US financial markets provide a high level of competitiveness, but most importantly, the data providers (in this case, Morningstar) can offer more detailed data with higher frequency and transparency for such a market. It would definitely be an interesting analysis to research the European and Asian markets as well, but they are outside the scope of this study.

- **Fund Size at specific date:** the funds must have at least a size (TNA) of 15 million US dollars as of the 3rd of February 2020, namely the first trading day within the studied time horizon.
  The choice of 15 million USD is not random, but it was made taking into consideration Pástor and Vorsatz (2020), which, using also Elton et al. (2001) as a reference, rightly points out that filtering by size is particularly relevant in order to investigate fund flows. It is plausible to assume funds with lower TNA levels would have considerable swings in percentage flows even with relatively small dollar flows. A similar approach is also adopted by Döttling and Kim (2020), which in turn used 20 million USD as a threshold. Less restrictive with this condition is the analysis done by Hartzmark and Sussman (2019), which considers funds with a TNA greater than or equal to 1 million USD.

- **Percentage of Equity in the Asset Allocation (Net):** it is defined by Morningstar as the percentage of the fund's assets invested in stocks, and in this study is set to be at least 90%. This figure is calculated separately for the short and long positions of the portfolio, and the sum of the asset allocation of each will not necessarily equal 100%. The net value is derived by subtracting the short postions from the long one. The long and short positions can be rescaled as well: rescaling ensures that the sum of the asset allocation breakdown will sum to 100%.

The choice of using funds with almost all their assets invested in stocks is also taken from existing literature. If on one hand is true that an exogenous shock should impact all areas of the economy, on the other hand, it is empirically proven that it is the stock market that suffers the most severe disruption. Furthermore, the sustainability ratings are widely applied to equities, while they are less frequently assigned to other asset classes.

- **Index and Fixed Income funds:** funds with the word "Index", "Bond" or "Fixed Income" in their name are excluded from the preliminary sample. The same filter has been applied by Pástor and Vorsatz (2020) and represents an additional way, together with the specification just above, to ensure no funds investing in asset classes different from equities are present in the dataset. Passive investments, such as ETFs, are also excluded from the sample by this filter, even if they have bypassed the first condition, namely being open-end active funds.

- **Survivorship bias:** also known as survivor bias, it is the propensity to ignore those companies or funds that have failed and instead focus on the performance of those that are still trading. A fund's or market index's basic characteristics as well as previous performance may be overstated as a result of that. To address any potential survivor bias, the sample used for the analysis includes funds that no longer exist, that have been shut down or fully liquidated.

Most importantly, the data provided by Morningstar Direct are on a share class level rather than a fund level. This significantly increased the analysis's complexity and called for extensive data reworking, which is covered in detail in section 3.3.

3.2 Returns and Flows

Returns and flows, which are going to be the dependent variables of the regression models, are provided by Morningstar Direct but in an unprocessed state, and thus they need some reworking before they can be used in the research.

Throughout the study, returns are expressed as percentage points. Morningstar Excel Add-in is used in order to download the relevant return data from the platform. After inserting the SecID[13] of all the share classes that matches the criteria specified in the previous paragraph, the platform is able to provide daily returns both as "Gross" and "Total", which respectively correspond to returns gross and net of fees. Following Pástor and Vorsatz (2020), the returns extracted from the platform are net of the expense ratio, as the study focuses on the returns delivered to clients after fees. The data is of reasonable quality, with very few missing entries. As previously stated, the majority of the data provided by the platform is at the share class level and must be aggregated at the fund level before any type of analysis can be performed: returns make no exception. Once aggregated (the process is described in the next section), the resulting returns are used to compute the abnormal or excess returns, which will in turn be the dependent variable. Four type of excess returns are computed:

- *Abnormal* or *Benchmark-adjusted returns*

  These are the easiest to compute, as they are calculated subtracting the market return from the actual fund return. To put it formally:

$$\alpha_t^i = R_t^i - R_{m,t} \tag{3.2}$$

  Where α_t^i is the benchmark-adjusted return for fund i at time t, R_t^i is the return for fund i at time t and $R_{m,t}$ is the market return at time t.

  These set of returns are called benchmark-adjusted precisely because the market is being

[13] The SecID is the Morningstar identifier for a share class of an investment. Each Fund has its own FundID, but within each fund there may be several share classes, differentiated by their SecID.

represented by a benchmark index. Several indexed are taken into consideration like the S&P500 and the MSCI World, but both of them presented flaws that could bias or reduce the precision of the analysis. The MSCI World index was considered at the beginning, when the scope of the study was still broader and not focused only on the US. Although this index is among the most commonly used in research and gives a reasonable representation of how the global stock market is moving, after narrowing the research scope, it is clearly not an ideal benchmark given the presence of stocks of completely different geographical areas in its composition. The next candidate turned out to be also a sub-optimal selection: in fact, despite being the most commonly cited index for the US stock market, the S&P500 cannot be the ideal benchmark for all funds due to its emphasis on large-cap firms. The S&P500 is difficult to beat since a number of large-cap technology stocks overperformed during the crisis. As noted in Sensoy (2009), also the prospectus benchmark[14] is not an ideal alternative, as the choice made by the mutual fund issuer can be strategic and aimed at giving a better representation of the fund performances.

Lastly, the Russell 3000 is identified as the optimal benchmark. That is due to both its geographical focus on the US and the fact that, differently from the S&P500, it tracks the performance of the largest 3'000 US companies representing approximately 96% of the investable US equity market[15]. The influence of few large companies is thus mitigated by the presence of basically all the US publicly listed companies, whatever their size might be. Hence, Russell 3000 represents the "market" throughout the whole study.

- *Factor-adjusted returns*

  Alongside with the benchmark-adjusted returns, other three set of alphas are computed

[14] A prospectus for a mutual fund is a booklet or brochure that contains details about a mutual fund. Before making an investment, mutual fund companies are required to provide comprehensive prospectuses to potential investors. In the prospectus, the mutual fund company also indicates an ideal benchmark against which to compare its own performance.

[15] This information is provided by the factsheet of Russell 3000, available at the following website: `https://www.ftserussell.com/products/indices/russell-us`. All the factsheets are provided by FTSE Russell, the financial group that mantains this and other indexes.

using the following models:

- CAPM

  Attributed to William Sharpe who developed it based on the work of Harry Markowitz, the CAPM is one of the most popular financial models ever produced. It is used in this study to derive excess returns, using the following formula:

$$\alpha_t^i = R_t^i - [R_{f,t} + (R_{m,t} - R_{f,t}) \times \beta_t^i] \tag{3.3}$$

  α_t^i, R_t^i, and $R_{m,t}$ are the same as before. The new factors are $R_{f,t}$, which represents the risk-free rate and corresponds to the 1-month rate of US treasury bills, and β_t^i, which is a measure of the systematic risk of the fund. The data for the market risk premium ($R_{m,t} - R_{f,t}$) and the risk-free rate are downloaded from the data library of Kenneth French[16], while betas, being unique for each fund, are provided by Morningstar but on a share class level, and thus need to be aggregated at fund level before calculating any alphas.

- Carhart model

  This model, also know as the 4-factor model, is an expansion of the CAPM, and was developed by Eugene Fama, Kennet French and Mark Carhart. In particular, in Carhart (1997) the latter expanded the model invented by the first two researchers (3-factor model), and that is why this model is associated with his name. The formula used to extract excess returns is the following:

$$\alpha_t^i = R_t^i - [R_{f,t} + (R_{m,t} - R_{f,t}) \times \beta_t^i + SMB + HML + MOM] \tag{3.4}$$

  Where SMB (Small minus Big) is a factor that accounts for smaller companies outperforming larger ones over the long-term, and HML (High minus Low) takes into

[16] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

consideration value stocks trend to outperform growth stocks. These two factors were the innovation introduced by Fama and French to the CAPM in their notorious article Fama and French (1993). The last component was instead an introduction by Carhart, who also wanted to include the cross-sectional momentum (MOM) factor, which roughly describe the excess return of "winner" stocks (stock with the highest return in the past year) over past year "loser" stocks. All thre factors SMB, HML and MOM are taken directly from Kenneth French data library.

- Fama-French 5-factor model

  This is the latest model in chronological order. It was developed in Fama and French (2015), and introduced two new factors to the Fama-French 3-factor model. Despite attracting some criticism (Blitz et al., 2022), the model proved to be able to expand the explanatory power of its predecessor. The equation to compute alphas is as follows:

$$\alpha_t^i = R_t^i - [R_{f,t} + (R_{m,t} - R_{f,t}) \times \beta_t^i + SMB + HML + RMW + CMA] \tag{3.5}$$

  Where RMW (Robust minus Weak) is a coefficient that take into account the return spread of the most profitable firms minus the least profitable, while the CMA (Conservative minus Aggressive) includes in the model the return spread of firms that invest conservatively minus aggressively (hence, different investment portfolios).

Together with the different set of alphas, also the net flows will be the dependent variables of the regression analysis. Similarly to the returns, flows variables also need to be processed before using them in the models. Estimated flows are one of the few data that Morningstar offers at fund level[17]. However, the data points obtained are expressed as inflows or outflows of USD on a daily basis (Dollar flows). The USD amount of flows is hardly significant, being the sample formed of funds with very different TNA from each other, therefore this data

[17] The data point is populated by Morningstar with aggregated share class-based flow if all are available, otherwise populated using flow computed from surveyed fund size.

point is transformed into a new set of data, namely the Percentage flows. The Percentage flows variable is computed as the percentage of daily net dollar flows divided by the fund's total net assets in the same day. The data for the fund size is already available at fund-level. This methodology is found in Sirri and Tufano (1998) and it is applied in most of the relevant literature used to write this study.

3.3 Data processing and Final sample

3.3.1 Preliminary sample and data cleanup

As discussed above, most of the data made available by Morningstar are on a share class level. A single mutual fund with a single investment portfolio may provide investors with shares in more than one "class". Each class reflects a comparable ownership in the portfolio of the mutual fund. Depending on the fund's class, different costs and fees will be applicable. Some classes offer privileges or advantages that others do not[18]. The preliminary sample extracted from Morningstar Direct consists of 14'107 share classes that meet the criteria specified in section 3.1.2. The data points considered valuable to the analysis are the following[19]:

- FundID (which identifies the fund)
- SecID (which identifies the share class)
- Morningstar Category
- Inception date
- Fund Size comprehensive (daily)
- Net Assets (daily, at share class level)
- Estimated fund-level net flow - comprehensive (expressed in USD amount, daily)

[18] More information on Wallace (2017)

[19] Most of the data point definitions can be found following this link to Morningstar Excel Add-In Data Dictionary: `https://addin.morningstarcommodity.com/`; alternatively, each data point exact definition and calculation is available using the platform Morningstar Direct. A thorough description of variables and data points is produced in the Appendix.

- Low Carbon Designation (Q4 2019 and Q1 2020)
- Morningstar Sustainability Rating (from February 2020 to May 2020)
- Rating Overall (that is, Morningstar Star Rating; from February 2020 to May 2020)
- Returns (net of expenses, daily)

This data are essential to the models that will be shown in the next section. An initial basic data cleanup is carried out, during which the following items are deleted:

- all the funds with no Morningstar Sustainability Rating during the whole time horizon of the analysis
- all the funds with no Net returns available
- all the load waived share classes[20], which lack of most of the relevant data

After this step, the dataset still contains 8'700 share classes, but many of these share classes show missing data points within the different variables. A new, deeper data cleanup is thus performed using Python: the step-by-step procedure is described in the study annex dedicated to the Python code. In this stage, after analysing how many missing data there are in the variables Net Assets and Estimated net flows, a function is setup so that all the share classes belonging to a fund with more than 10% of missing data across the entire time horizon of the study are eliminated. This eventually results in the elimination of 1.20% of the funds. Finally, the final sample is built, counting 8'618 share classes belonging to 2'396 funds. Its graphical representation broken down by sustainability rating category is showed in Figure B.1.

[20] Load-waived funds are a share class of a mutual fund that don't charge its investors the usual load fees (such as front-end loads). Investors gain from owning shares in load-waived funds since they can keep the entire return on their investment rather than losing a portion of it to fees. The quantity of load-waived share classes is typically restricted by mutual fund firms, who also restrict access to them to a select group of investors.

3.3.2 Aggregation at fund level and proxy data

Despite the effort to implement valid aggregation methodologies, many share classes and funds present missing values that complicate the study. Therefore in a first step, for each variable with significant missing data issues, a different approach is considered, and sometimes proxy data are generated.

- Low Carbon Designation: many share classes had no information regarding the Low Carbon Designation. Hale (2021) explains that the designation is given only at fund level, therefore, if one share class within a fund has the LCD, the fund is recognized as LCD achiever. On the contrary, if no share classes within a fund has information regarding the LCD, then the fund is assumed not to qualify for the designation.

- Fund Size: the sporadic missing observations still present after data filtering are filled with values generated through linear interpolation along the time axis of observation of the values. If missing values span for 3 days, for examples, the day before and the day after this interval are used to linearly interpolate missing values in between. This may not be the most rigorous approach, but still preferable than using a simple average.

- Net Assets and Estimated Net Flows: the approach is the same as above, with the only difference that these data points are on share class level and are subsequently aggregated.

- Returns: sporadic observations of returns have value of exactly zero. The most reasonable assumption is that Morningstar has a missing data point for that specific observation. Returns of exactly 0% are thus proxied using the return of the benchmark index (Russell 3000) on the same day.

- Betas: these are provided by Morningstar at share class level for the specified time horizon (in this case, from February 3 to May 29, 2020). 79 share classes within the dataset have no betas. This strategy is used to overcome this issue: if a fund has at least one share class with a beta, the same beta is applied to all the other share classes linked to the same fund. If more than one share class has a beta, then their average is applied to the share class(es) with a missing beta. If there are no betas at all within a fund, then the data entry is substituted using a beta of 1 (which implies that this fund moves proportionately to the market). Eventually, only 18 funds in the entire dataset are assigned a beta of 1.

- Controls: some of the control variables also have missing data. These data are provided by Morningstar at fund level; therefore, when there are some missing data, it means that the whole fund is affected. Fortunately, only very few funds were affected by this issue (around 0.25%). In such cases, the missing value is substituted with the median value of that control variable across the entire dataset.

Aggregating the data at fund level requires a considerable amount of programming. Many variables, such as the share class returns, are transformed into fund level variables using a weighted average, in which the weights are given by the Net Assets inside each share class. It is easy to understand the logic of this aggregation process by taking the example of the fund returns, which are computed as follows:

$$R_t^i = \frac{\sum_{j=1}^{n} R_t^j \times NA_t^j}{\sum_{j=1}^{n} NA_t^j} \tag{3.6}$$

Where R_t^j is the return, and NA_t^j are the Net Assets of share class j at time t.
Kacperczyk et al. (2014), Pástor and Vorsatz (2020), and Hartzmark and Sussman (2019), among others, substantially follow the same methodology. Once applied to the returns, the exact same aggregation approach is replicated for other variables (mostly controls), also delivered by Morningstar at share class level.
However, not every variable can be aggregated through this weighted average. An example may be the variable "Rating Overall", namely the Morningstar Star Rating. This is another rating, in all likelihood the most famous, issued by Morningstar but once again provided by

the platform at share class level. While sustainability ratings tend not to vary between share classes, the Star Rating is relatively less consistent. Aggregating the Star Rating at fund level using the weighted average method would have resulted in each fund having a Star Rating with decimals, presumably very different from each other. That would have made very complicated to include the Star Rating as a control in the regression models: to get around this issue, each fund is given the same Star Rating as its largest share class (in terms of Net Assets).

3.4 Models

Once the data are structured as a panel data, the two regression models on which this study is focused can be performed. The first one is a standard Pooled OLS regression model, built as shown below:

$$y = \beta_0 + \beta_1 \times HighESG + \beta_2 \times AbAvgESG + \beta_3 \times BelAvgESG + \beta_4 \times LowESG + \beta_5 \times LCD \tag{3.7}$$

Where

- the variables *HighESG*, *AbAvgESG* (above average), *BelAvgESG* (below average) and *LowESG* are dummy variables that identify funds rated respectively with five, four, two and one Morningstar Sustainability Globes. The funds with three globes are less relevant as they express an average sustainability rating which would provide no answer to the fundamental questions of the study.
- the variable *LCD* is also a dummy variable that identifies funds that achieved the Low Carbon Designation.

In this model, the Morningstar Global Category is used as fixed effects, and a set of control variables is included in order to exclude potential Omitted Variable Bias (OVB). The controls used are the Inception date of the fund, log of the Fund Size, the Turnover Ratio, the Net Expense Ratio, the Morningstar Star Rating and the Excess Return for the past 12-month as computed by Morningstar. An in-depth description of such variables is provided in the Appendix A. The regression is performed on the Benchmark-adjusted and the Factor-adjusted returns as well as on the Percentage flows. The different periods used are the Pre-Crisis (February 3 to February 21), Crash (February 22 until March 23) and Recovery (March 24 until May 29), and obviously the entire time span of the analysis (February 3 to May 29). However, when running the regressions on the temporal subsets, the model would not be able to incorporate Fixed effects and therefore a regular Pooled OLS is conducted instead. The reason is that the variation within each fund is simply not enough. Consider, for instance, the Sustainability Rating, which is issued on a monthly basis: in the Pre-Crisis period, the daily observations are exactly the same for each day of the studied horizon.

The second regression model is of the Difference-in-Differences (DID) type, and is designed as follows:

$$y = \beta_0 + \beta_1 \times \mathit{HighESG} \times \mathit{COVID} + \beta_2 \times \mathit{LowESG} \times \mathit{COVID} \tag{3.8}$$

Where

- $\mathit{HighESG}$ is a dummy variable that identifies funds that have received four or five Morningstar Sustainability Globes.
- LowESG is also a dummy variable that identifies funds with one or two Globes.
- COVID is a dummy variable used as interaction term, which will be equal to 1 from the 24$^{\text{th}}$ of February onwards, to signal the beginning of the Crash period and the panic caused by the first pandemic wave, and 0 otherwise.

β_1 and β_2 calculate the difference between the average fund and high and low sustainability funds in terms of how much higher (lower) flows or better (worse) performances they experience following the start of the COVID-19 crisis, and thus represent the key elements of the analysis. Funds with average sustainability rating are the control group, while the funds with either high or low sustainability are considered as the treatment groups. Once again, the Morningstar Global Category is used as fixed effects, and the same set of control variables is included to mitigate potential OVB. This time, the LCD dummy variable is included as a control. The dependent variables are the same shown above, as well as the time periods analysed, with the exception of the Pre-Crisis period which, given the nature of the COVID dummy variable, would give insignificant results.

Chapter 4

Empirical results

The following chapter describes and explains the results of the regression analysis discussed above. It is divided in two sections, the first focusing on fund performances and the second on fund flows, the two dependent variables studied.
The regressions are performed using R, and their outputs are shown in the Appendix B. Summary statistics of the most important variables in the analysis are shown in Table B.1.

4.1 Performances

As shown in Figure B.2, when comparing the distributions of cumulative densities of net returns, funds with four and five globes ("High ESG") performed, on average, better than funds with one or two globes ("Low ESG") across the entire time horizon of the study. This finding is in line with Pástor and Vorsatz (2020), although the distributions are clearly different.
Another point of view is outlined in Figure B.5: here, the average returns are broken down by their Morningstar Sustainability Ratings. Despite their evident differences mainly related to the factors used to compute the factor-adjusted returns, for every set of alphas, the higher the rating, the higher the abnormal returns. The substantial variation of the calculated returns is even more evident in Figure B.3: here it is easy to notice that there is almost no difference

between the Benchmark-adjusted returns and the ones computed through the CAPM. As a matter of fact, the only differences between the two sets are the Beta, that has an average of 1.0024, and the Risk free rate, which ranges from 0.6% to 0%: this explains the almost perfect overlapping of the two lines. Conversely, Carhart and Fama-French abnormal returns show significantly more variability, which in turn is reflected in the coefficients of the regressions. Also in the case of the LCD, the higher the sustainability, the higher the (average) excess returns, as shown in Figure B.6. Here being only two categories (with LCD or without it), the gap is more evident than in the previous cases. The graphical visualization of these results would be a glaring indication that sustainability was a major predictor of returns over the pandemic crisis, but only 24.92% of funds qualify for the LCD at the outset of the analysis, while the remaining percentage do not.

To look for more meaningful results, the regression analysis is carried out, but the results turn out to be relatively surprising. In Table B.2, the first regression model is applied to the dataset for the entire time period of the study, including the Global Category fixed effects and the set of control variables specified in section 3.4 and in Appendix A. In addition to its statistical significance being extremely low, Table B.2 presents mixed results that make it very cumbersome to draw unambiguous conclusions. In column (1), where the Benchmark-adjusted returns are examined, the coefficient for HighESG (5 Globes) shows a positive impact on returns, while all the other ratings, including the LCD, have a negative impact. The latter is the only coefficient with considerable significance. The first results are thus counterintuitive, to say the least. A possible explanation for such results would be that, given their different calculations and methodologies, the Morningstar Sustainability Rating and the Low Carbon Designation do not equally represents the sustainability of a fund (or, in general, of a security) from an investors point of view, which would partially explain such different coefficients. Similar results, both in terms of magnitudes and significance, are offered by the regression on CAPM (2). In column (3), instead, a different point of view is offered by the analysis of Carhart model's returns: this column has the highest level of statistical significance (both for the coefficients and for the R^2) and shows a negative effect of all the sustainability ratings except for the lowest rated category (Low ESG, 1 Globe) which has a positive sign. Regression on Fama-French abnormal returns is displayed in column (4), and shows a positive coefficient for LCD, as well as for 4 and 5 Globes. To provide a reasonable explanation to such results, the factors used for the calculation of the abnormal returns and the stocks or indices used for their derivation should be investigated in depth.

In Tables B.3, B.4, and B.5, the same regression is performed, respectively, on the Pre-Crisis, Crash and Recovery temporal subsets. As specified before, these models do not include the Global Category fixed effects and apply a regular Pooled OLS regression approach. Despite the overall models' still relatively modest explanatory power, the coefficients' significance improves considerably, and their magnitude is more consistent across the four columns.
In particular, in Table B.3 the LCD is associated with a positive effect for all sets of alphas, while having 4 or 5 Globes is associated with a negative effect on funds on average. Conversely, it seems the lowest rated categories predict positive performances. The coefficients nearly never change across different sets of alphas: this may also be explained by Figure B.5, which shows fairly similar performances of funds with different sustainability ratings until the onset of the pandemic-induced market panic (last week of February).
The situation shifts substantially in Table B.4, where, once entered in the Crash period, higher sustainability ratings predict higher abnormal returns, while funds with low and below average sustainability experience the opposite.
Lastly, in the Recovery periods, while the beneficial effect of LCD persists, the pattern for Morningstar Sustainability Ratings once more reverses, with higher ratings predicting lower abnormal returns and lower ratings doing the opposite. This is entirely reasonable, given the

broad and substantial market's rebound that characterised such period; presumably, securities that experienced significant underperformances amid the most turbulent phases of the crisis had been reevaluated, while the most secure (and most expensive) highly sustainable assets were being sold to profit or simply reassessed in the context of the improved financial situation.

The Difference-in-Differences analysis, however, offers utterly unequivocal results. All the regressions done with this approach, regardless of the time period taken into account, shows a positive impact of high sustainability and at the same time a negative one of low sustainability on alphas.

Once again, the regression is performed on the entire time horizon of the study in Table B.8, including Fixed Effects, and according to the findings having high sustainability (four or five Globes) is associated, on average, with 0.32% higher abnormal returns compared with an average sustainability (three Globes), while belonging to the lower categories (one or two Globes) implies, on average, 0.004% lower returns. However, unfortunately enough, also this regression is characterised by an extremely low statistical significance.

As in the case of the first model, also in the DID model when disregarding the fixed effects and switching to a pooling model for the temporal subsets, the significance of coefficients increase noticeably, although the relevance of the model itself is still well below expectations. The pattern elaborated above repeat itself in Table B.9 and B.10, where sustainability is associated positively with abnormal returns during the Crash and inversely during the Recovery.

Interestingly, the only set of alphas that has contrasting findings is the one computed with the Carhart model, in column (3). There is no straightforward evidence to justify such a discrepancy, but as written above, this result is presumably due to the factors (in particular, the MOM) used in the calculation of abnormal returns.

4.2 Flows

The Flows variable employed in the regression analysis is, as mentioned before, the Percentage flows. In Figure B.4 a comparison of the average Dollar and Percentage flows is presented. The lines representing the two sets of flows are almost overlapping for most part of the time horizon, with, as easily predictable, ample swings during the apex of the crisis and relatively less varying flows in the aftermaths.

Running the first regression model on the Percentage flows variable for the entire time span of the crisis results once again in extremely low statistical significance of the model, and coefficients with difficult interpretations. From Table B.6, it appears that only the LCD has a significant - and positive - impact on the percentage flows, while all the categories of the Morningstar Sustainability Rating shows negative coefficients. Their magnitude suggests larger cash outflows for funds with two Globes in particular.
Table B.7 breaks down the regression outcomes by the three different subperiods used earlier. Similarly to what was done previously, when running the regression on these temporal subsets, fixed effects are dropped, and a regular Pooled OLS regression is performed. Also in this case, identifying a pattern is not straightforward. In the Pre-Crisis period - column (1) -, funds with LCD and five and two Globes experience greater daily cash outflows than three-globes funds, whereas funds with four and one Globes experience the opposite. With the exception of the LCD coefficient, which stays negative, the results in column (2) are reversed when the analysis is moved to the Crash period. Finally, in column (3), during the Recovery period, the extreme categories - five and one globes - seem to attract more investments compared to funds with four and two globes, which suffer further cash outflows.

A DID approach is used also with flows: according to the results shown in Table B.11, funds with four and five Globes experience roughly -0.000269% net cash outflows per day than funds with three Globes. Funds rated worst in terms of sustainability, instead, experience a 0.0000468% larger daily net cash inflows. Such results are most consistent with the theory that deems sustainability as "luxury item". Investors would therefore consider the non-financial benefits deriving from sustainable investments as expensive and unnecessary in the context of severe financial constraints, shifting their preferences towards traditional objectives instead. It is also possible that market participants would simply pay less attention to sustainability issues as a consequence of the significant economic shock brought on by COVID-19, but there is no evidence to back this hypothesis.

Slicing the time horizon in the two subperiods after the beginning of the crisis offers another picture (Table B.12). During the worst time of the pandemic (Crash), highly rated funds have larger cash inflows than both the control group (three-Globes funds) and the less sustainable funds (one and two globes), and the situation reverses once again during the Recovery period, when both highly and poorly rated funds in terms of sustainability experience negative flows compared to the ones rated with average sustainability. A plausible interpretation is that, in the midst of the financial market meltdown, panic-stricken investors started to withdraw their capital from funds, resulting in greater outflows from funds with three globes (which represent the vast majority both in the market and in the dataset used for this study, as shown in Figure B.1), ideally deeming the funds with high sustainability safer. However, why funds with one and two globes also have better flows remains a difficult question to answer, especially when combined with the evidence from the performance analysis conducted earlier which sees *LowESG* funds having lower excess returns. A possible answer is given by Döttling and Kim (2020), which consider the possibility that market participants put in place a "buy the dip" strategy, therefore shifting their capital allocation in favour of assets that experienced a sharper decline in value over the course of the pandemic; this would be a reasonable explanation, but so far there is no evidence to back it.

Chapter 5

Conclusions

Several reasons have been advanced in support of sustainability positively or adversely predicting performance and flows, or having no relationship at all with such variables. Although it is beyond the scope of this study to fully answer the questions regarding how investors or funds react to sustainability ratings during a market shock, here are discussed different explanations for fund performances and flows as a function of sustainability ratings.

On a more general note, given their weak statistical significance, the findings of this study do not, by themselves, adequately address the key issues raised at the beginning of the research. This may be attributable to many factors, such as misspecified models (although they are largely similar to the ones applied in some of the relevant literature examined), or issues with the main identifying assumptions (referring to the missing counterfactual parallel trends assumption for the DID models) or even the unreliability of ESG databases (Berg et al. 2020 and Berg et al. 2022). The fact that even the articles taken as main references struggle with statistical significance is consoling only to a certain extent, since it suggests that the actual issue could be related to the data or to the empirical methodology. However, the lack of significance does not rule out the possibility of an actual underlying relationship, but it implies that the data do not support the existence of an effect. In line with this concept, a summarised interpretation of the results is provided below.

For what concern the analysis conducted on funds' performances, the results generally show a positive impact of sustainability on returns during the COVID-19 crisis. This is in line with the findings of Pástor and Vorsatz (2020) and hints that sustainability actually can combine financial (returns) and non-financial (social benefits) objectives even during an unprecedented crisis. The average fund in the dataset still underperforms the benchmark index (Russell 3000) for all the computed sets of alphas (except for the Fama-French 5 factors model), providing additional proof that active management is not necessarily a reliable protection in the event of a severe market shock.

The analysis on flows is instead more complex and is neither satisfactory in terms of statistical significance nor for the results themselves, which are puzzling and difficult to decipher. The DID regression models provide plausible outcomes, with highly rated funds experiencing a sharper decline in daily net fund flows (in line with the findings of Döttling and Kim 2020), but yet again the statistical relevance of the coefficient and the overall model is not sufficient to make any claim regarding a relation between sustainability ratings and flows. A possible issue may lie precisely within the choice of the studied variable. In fact, despite the Percentage flows being no longer an absolute measure of fund daily flows but rather a relative measure expressed as a percentage of TNA, its significance can be questioned. As pointed out by Hartzmark and Sussman (2019), even percentage flows are noisy and can change systematically depending on various factors like fund size. Hartzmark and Sussman (2019) tries to solve this issue by creating a new set of flows, the Normalized flows[21]. The results of the

[21] The same approach is used in Döttling and Kim (2020) and Ceccarelli et al. (2021). Normalized flows are created following a two steps procedure: first, the fund universe is divided into deciles based on fund size. Once these ten buckets are obtained, funds are ranked according to their net flows within their size bucket; this net flow ranking is in turn computed as percentile. The percentiles associated with each fund correspond to the Normalized flows variable.

papers using such variable appear to be more robust and statistically significant, therefore future research focusing on the same topics should take the Normalized Flows into account, alongside the more common Percentage Flows. Another issue is certainly related to the frequency of data: in Döttling and Kim (2020) and Hartzmark and Sussman (2019), respectively weekly and monthly data are considered, while in this study data are acquired on daily basis. This eventually result in Percentage Flows - and consequently regression coefficients - very close to zero and with many decimals. This does not represent by itself a serious problem, but combined with the weak significance of the regressions' outcomes, dealing with such small numbers makes it even more difficult to draw some conclusions.

The comparison between the two different regression models - Pooled OLS and DID - employed in this study favours the second one, as its outputs are more comprehensible and consistent with the articles used as references. Some concerns remain regarding the use of Fixed Effects, as once applied to the models, they do not seem to give the desired contribution in terms of improved model accuracy. That may be caused by the presence of other time invariant variables (controls) in the dataset, but this would need further investigation to prove.

The fact that during the COVID-19 crisis funds' high (low) sustainability is associated with better (worse) performances but at the same time with net cash outflows (inflows) is counterintuitive, but consistent with some of the literature explored in chapter 2. Positive ESG ratings are indicators of performance resilience during shocks, but investors seem not to entirely see it, privileging other - less sustainable - funds in the midst of the pandemic-induced panic.

Finally, some further suggestions for future research are made below.
In order to improve or elaborate further the findings of this study, an analysis on the sector a

fund is focusing on could be performed. This was one of the original objectives of the study, but it was impossible to research due to a lack of access to the GICS data[22]. With regard to the geographical area, it would be interesting to know more about how the same relationships studied in this thesis developed across other markets, both developed and emerging ones, also in relation to the very different reactions of governments and authorities to the spread of the virus.

[22] Some evidence in this regard is presented in Mazur et al. (2021)

Appendices

Appendix A

Details of Variables and Data Points definitions

The variables used in this study are extracted from three data sources: Morningstar Direct is the main one, followed by Kenneth French Data Library and FTSE Russell website.
Below, a review of the main variables and data point used in the analysis is provided.

Main fund level variables:

Benchmark-adjusted Returns or Abnormal Returns

Also called "alphas" or excess returns, they are a set of returns computed subtracting the daily returns of the Russell 3000 index to the daily net returns of each fund. They are described more in details in section 3.2. They are among the dependent variables of this study.

Dollar Flows

These correspond to the original set of flows provided by Morningstar Direct. Morningstar defines the data point as being populated with aggregated share-class-based flow if all are available, otherwise populated with flow computed from surveyed fund size. Dollar flows provide a measure of how much capital (USD) joined or left the fund on a given day. The variable definitely has some value, but it is not very useful to the analysis since funds with very different sizes would almost certainly have very different dollar flows each day.

Factor-adjusted Returns

These are three different sets of returns, computed through the use of three different models: CAPM, Fama-French 5 factors and Carhart models. They are described more in details in section 3.2. They are among the dependent variables of this study.

Fund Size

This variable represents the fund's entire amount of money managed across all share classes and subaccounts as a standalone portfolio. This may be larger than or equal to the net assets of the respective share class or subaccounts. The amount of the net assets and the fund size will coincide if only one share class is offered or the fund only appears in one policy. The data is provided by Morningstar already aggregated at fund level (sum across the net assets of all share classes), with daily observations.

Morningstar Low Carbon Designation

It is one of the two sustainability ratings and one of the independent variables use in this study. It is described in details in section 3.1.1. For more information, please refer to Hale (2018).

Portfolios with low carbon-risk scores and little exposure to fossil fuels are given this label. The designation serves as a sign that the companies in a portfolio are key in the development of the transition to a low-carbon economy.

Morningstar Sustainability Rating

It is one of the two sustainability ratings and one of the independent variables use in this study. It is described in details in section 3.1.1. For more information, please refer to Barr et al. (2021).

Through this Rating, investors may assess the relative environmental, social, and governance risks present in a portfolio. Sustainalytics' methodology for evaluating corporate and sovereign ESG risk serve as the foundation for the bottom-up evaluations of a portfolio's underlying holdings that determine ratings. The Morningstar Sustainability Rating is calculated in various processes in order to accurately reflect the relative risk within each portfolio, but the result is a category of 1 to 5 "globes" for each eligible portfolio.

Net Returns

Funds' returns, net of the expense ratio. They are provided as daily data at share class level by Morningstar, and aggregated at fund level using a weighted average calculation, when the weights are represented by the Net Assets.

Percentage Flows

This variable has been computed as the percentage of Dollar flows over the Fund Size (AUM), daily. Percentage flows are a much better variable to include in the regression analysis, as they represent a measure relative to the fund size and thus allow the results to account for the heterogeneity of size within the studied dataset.

Other variables and controls:

Beta

Beta is is a measure of the volatility (or systematic risk) of a security or portfolio compared to the market as a whole, which in this study is represented by the Russell 3000 index. It is generally used in the CAPM model. A security is considered to be theoretically less volatile than the market if its beta value is less than 1; conversely, a beta greater than 1 is associated with more volatile securities. In this study, Beta is used to compute the Factor-adjusted returns. It is provided by Morningstar for a specified time horizon (in this case, from February 3 to May 29, 2020), but at share class level. Before using this factor in the calculations, it is essential to ensure that there is no missing data. Hence, missing data are filled using the following method: if a fund has only one share class with a beta, the same beta is applied to all other share classes linked to the same fund. If there are many share classes with a beta, the average of their betas is applied to the share class or share classes that lack of it. If a fund has no betas at all, the data is filled with the value of 1. After that, beta is aggregated at fund level, using the weighted average approach seen in section 3.3.2.

Excess Returns

This variable is a measure of a fund's return in excess of a benchmark for an entire year before the analysed period (i.e. from February 3, 2019 to February 2, 2020). Basically, they are the same thing as the Benchmark-adjusted or Abnormal returns; the only difference is that, while the latter are computed on a daily basis and thus are viable dependent variables for the regression analysis, this variable is used as a control because it is not a daily data. Morningstar is able to provide Excess returns for whatever benchmark for a specified time period, but only as annualized returns (i.e. Annualized fund return - Annualized benchmark return). This eventually result in a time-invariant set of returns, which are impossible to use as y in a panel data regression model. Nevertheless, they add some information to the model used as controls, providing a coefficient that may explain the investor propensity to choose funds with determined previous excess returns.

Inception date

Control variable included in the regressions. It is computed as the number of days passed from the inception of the fund to the day of the analysis. Since many share classes have different inception dates, to aggregate at fund level, only the oldest date is considered, which normally coincides with the inception date of the oldest/first share class ever issued by the fund. This variable is included to take into account the possibility that, compared to newly constituted funds, a long-established fund may have a higher likelihood of gaining new capital or retaining the ones that already manages.

Morningstar Global Category

The Morningstar Global Category system groups investment vehicles across the globe that invest in similar asset classes. The categories are based on the investment vehicles' underlying local Morningstar Category assignments. Entries in the Morningstar Global Category system may be broader or more granular than the local categories that constitute the global category. More information at this clickable link. Including this variable would have the effect of clustering heterogeneity withing the global categories made by Morningstar, and thus improving the precision of the models.

Morningstar Star Rating

Quoting Morningstar itself, the Morningstar Rating, or "star rating," is a purely quantitative, backward-looking measure of a fund's past performance, measured from one to five stars. Star ratings are calculated at the end of every month. To determine a fund's star rating for a given time period (three, five, or 10 years), the fund's risk-adjusted return is plotted on a bell curve: If the fund scores in the top 10% of its category, it receives 5 stars (Highest); if it falls in the next 22.5% it receives 4 stars (Above Average); a place in the middle 35% earns 3 stars (Average); those lower still, in the next 22.5%, receive 2 stars (Below Average); and the bottom 10% get only 1 star (Lowest). The Overall Morningstar Rating is a weighted average of the available three-, five-, and 10-year ratings. More information are available in Morningstar (2021).

The Morningstar Star Rating is used in the analysis as control variable, and data are extracted for the time horizon of the analysis. Also this rating is issued at share class level, and it is aggregated at fund level taking the one of the largest share class, as described in section 3.3.2. Investors choose where to allocate their capital also according to the Star Rating of the funds (Pástor and Vorsatz, 2020), therefore including it in the regression models should allow for higher quality of the analysis.

Net Assets

This variable returns the daily share class level of total net assets, measured in currency units (in this case, USD). The size, adaptability, and popularity of a fund can be determined by looking at its net assets. They assist in figuring out whether a small firm fund, for instance, can stay in its investment-objective category if its asset base grows to a problematic size. In this study, net assets are used mainly in the aggregation process at fund level, as the weights of the weighted average. They are not part of the regression models.

Net Expense Ratio

This variable specifies the percentage of the fund's assets are used to cover management fees, 12b-1 fees, administrative charges, and any other asset-based expenditures, with the exception of brokerage fees. The expense ratio does not take sales charges into account. The underlying fund costs are not included in the expense ratio for a fund of funds; only the wrap or sponsor fees are. Including this variable in the analysis is supposed to account for the fact that investors also calibrate their decisions according to the expense ratio of funds.

In the analysis, net expense ratios of 2019 and 2020 are considered, and used as controls. They are expressed at share class level, and aggregated through the weighted average approach.

Turnover Ratio

This metric represents the fund's trading activity and is calculated by dividing the lesser of the fund's acquisitions or sales (excluding any securities with maturities of less than one year) by the average monthly net assets. A turnover ratio of at least 101% does not guarantee that every security in the portfolio has been exchanged. Practically, the percentage that results roughly corresponds to the portion of the portfolio's holdings that have changed during the past year. Low turnover (between 20% and 30%) is a sign of a buy-and-hold strategy. More than 100% turnover would suggest a significant amount of buying and selling of securities in an investment plan. Morningstar does not compute turnover ratios; instead, they are derived from the financial highlights of the fund's annual report.

In the analysis, turnover ratios of 2019 and 2020 are considered and used as controls. They are already provided at fund level.

Appendix B

Tables and Figures

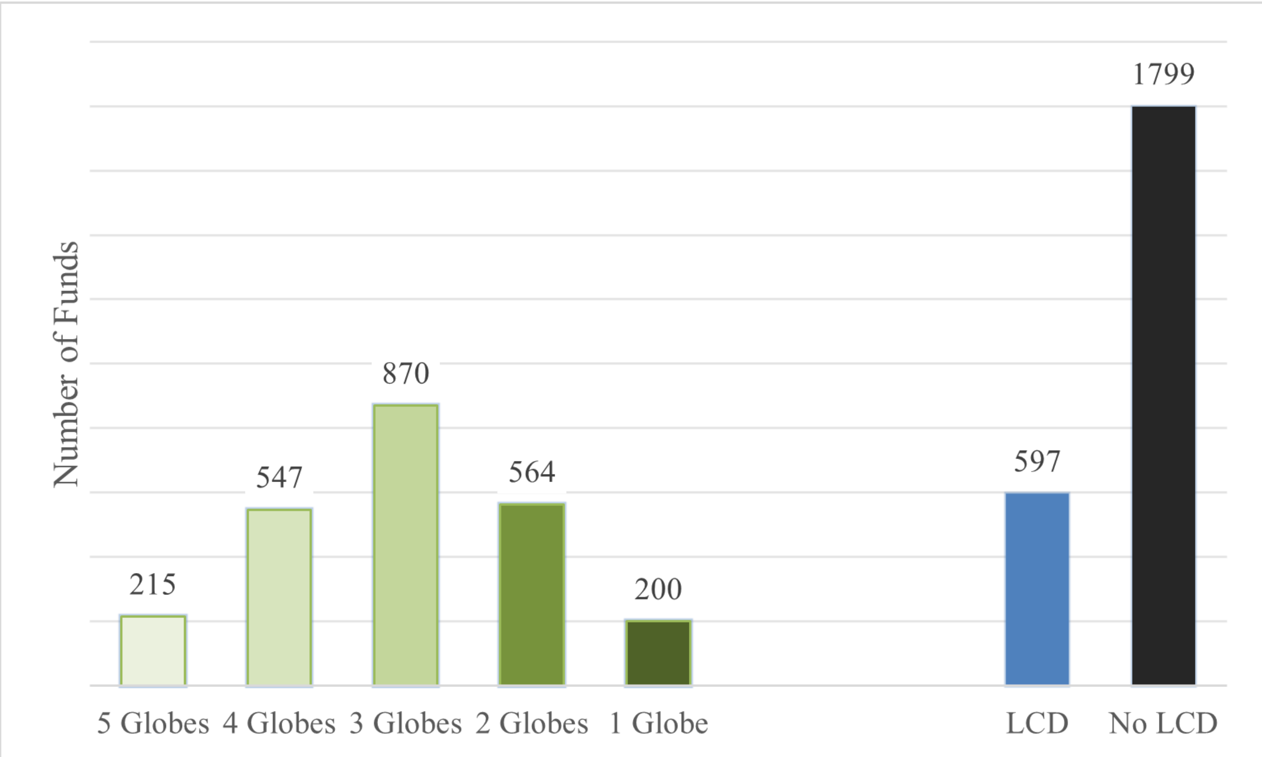


Figure B.1 – Dataset composition by Sustainability Ratings
This graph shows the number of funds in each sustainability rating category: on the left, the Morningstar Sustainability Rating; on the right, the Low Carbon Designation. Source: own research.

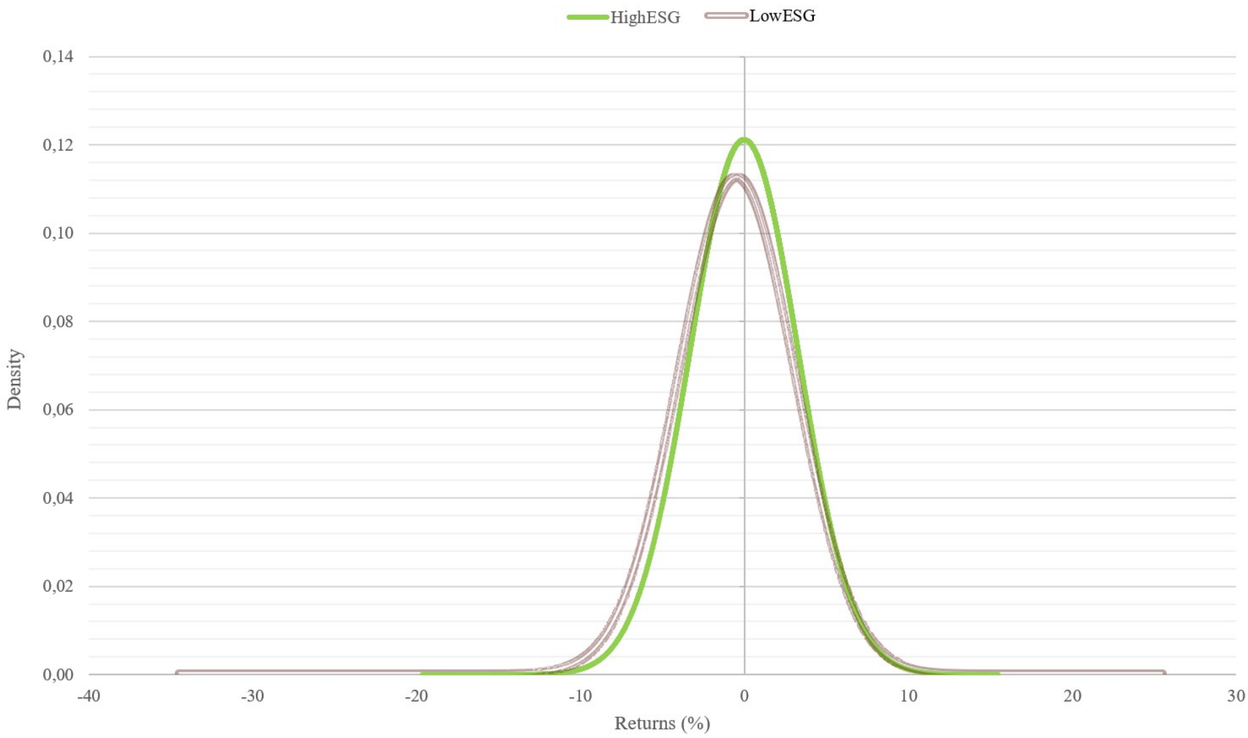


Figure B.2 – Cumulative Return Densities across Sustainability Ratings
This figure plots densities of funds' net returns from February 3 to May 29, 2020 for two categories of sustainability: High ESG (four or five globes in the Morningstar Sustainability Rating) and Low ESG (one or two globes). As ratings are assigned each quarter, this figure takes into account the ratings issued on December 2019 and March 2020. The net returns, instead, are daily data provided by Morningstar. Source: own research.

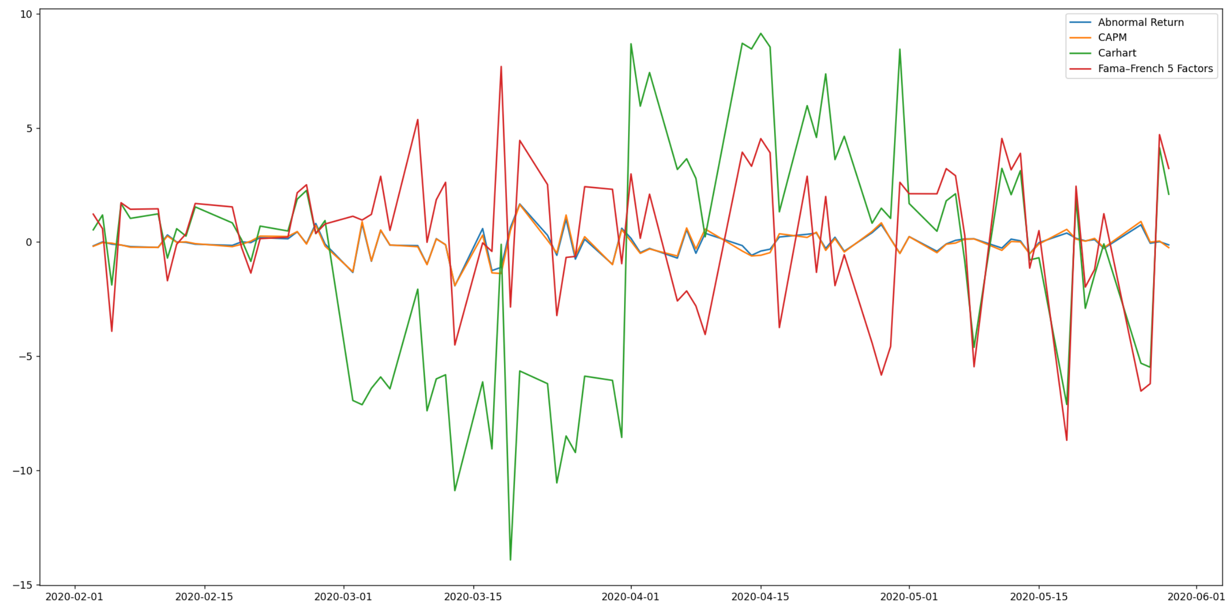


Figure B.3 – Average fund's alphas over the studied time horizon
This figure shows the evolution of the average fund's sets of alphas calculated and used in this study throughout the entire time horizon (February 3 to May 29, 2020). Source: own research.

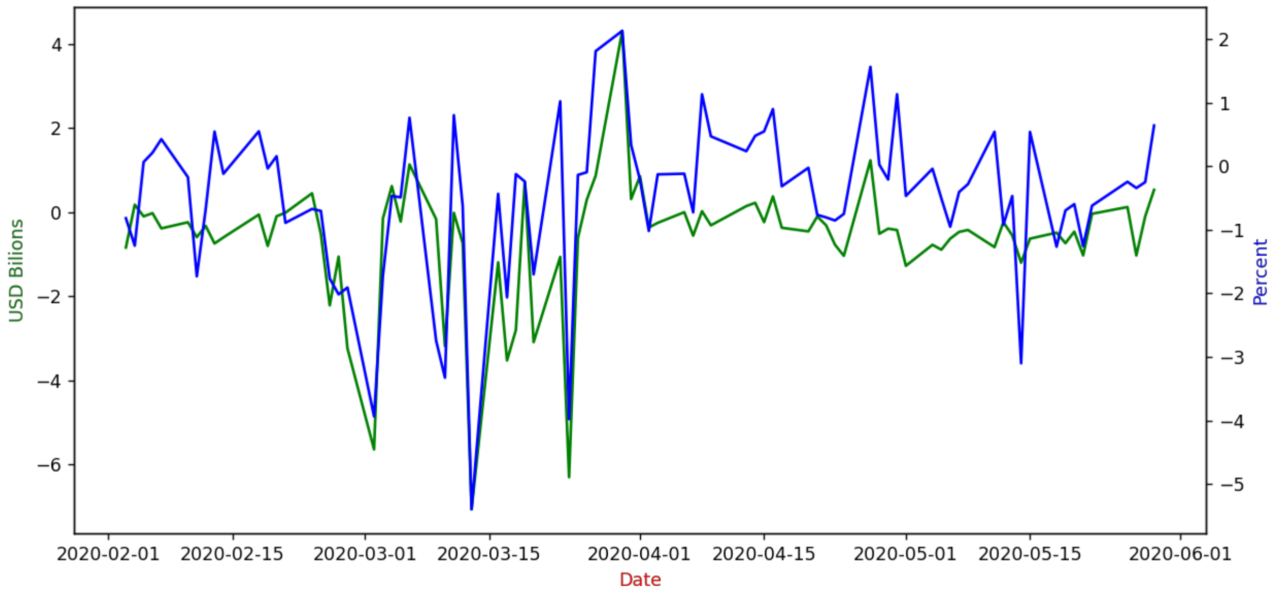
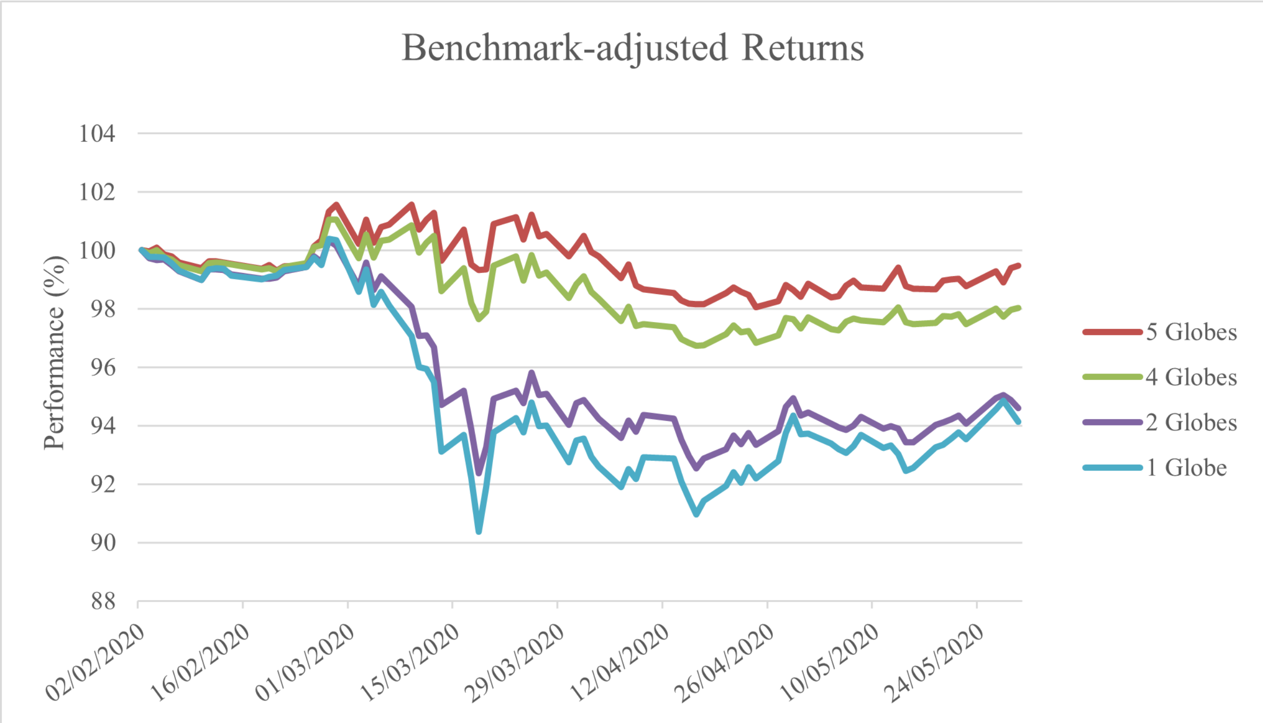
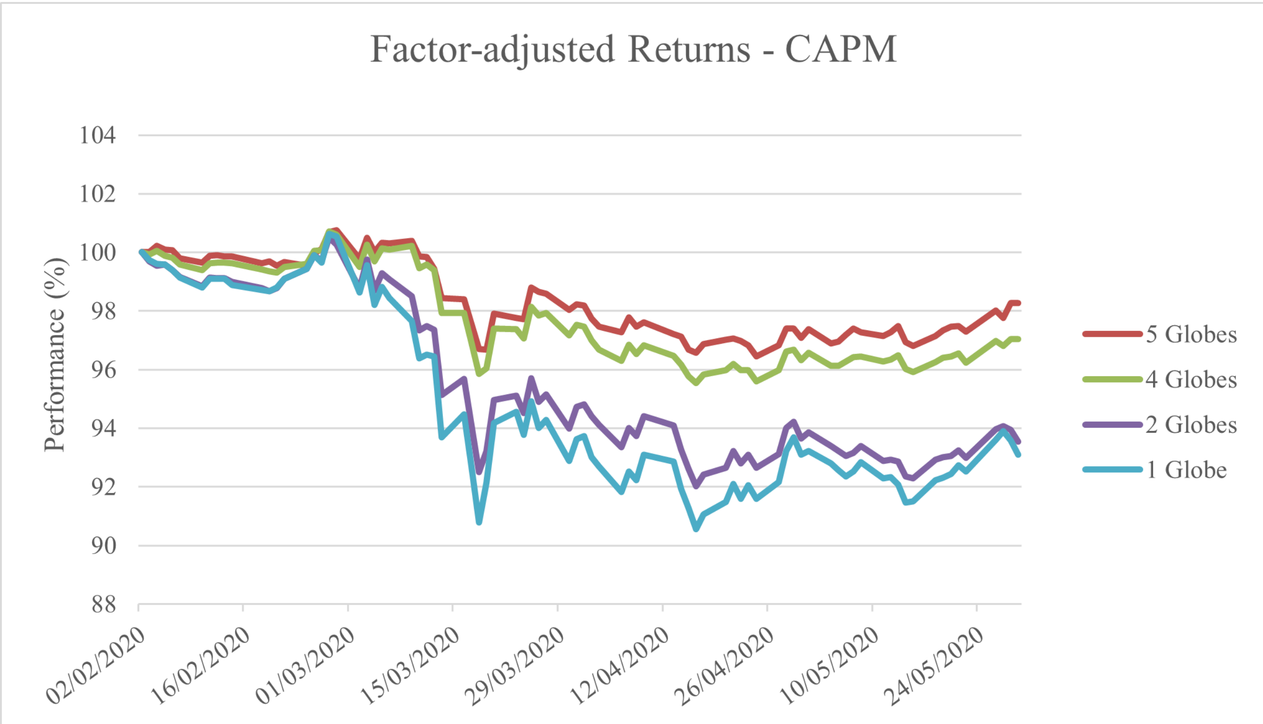


Figure B.4 – Average Dollar vs Percentage flows comparison
This figure represents the average dollar and percentage flows of the funds in the dataset over the analysed time horizon. The left axis measures flows in USD Billions, the right one in percentage (over the TNA). Source: own research

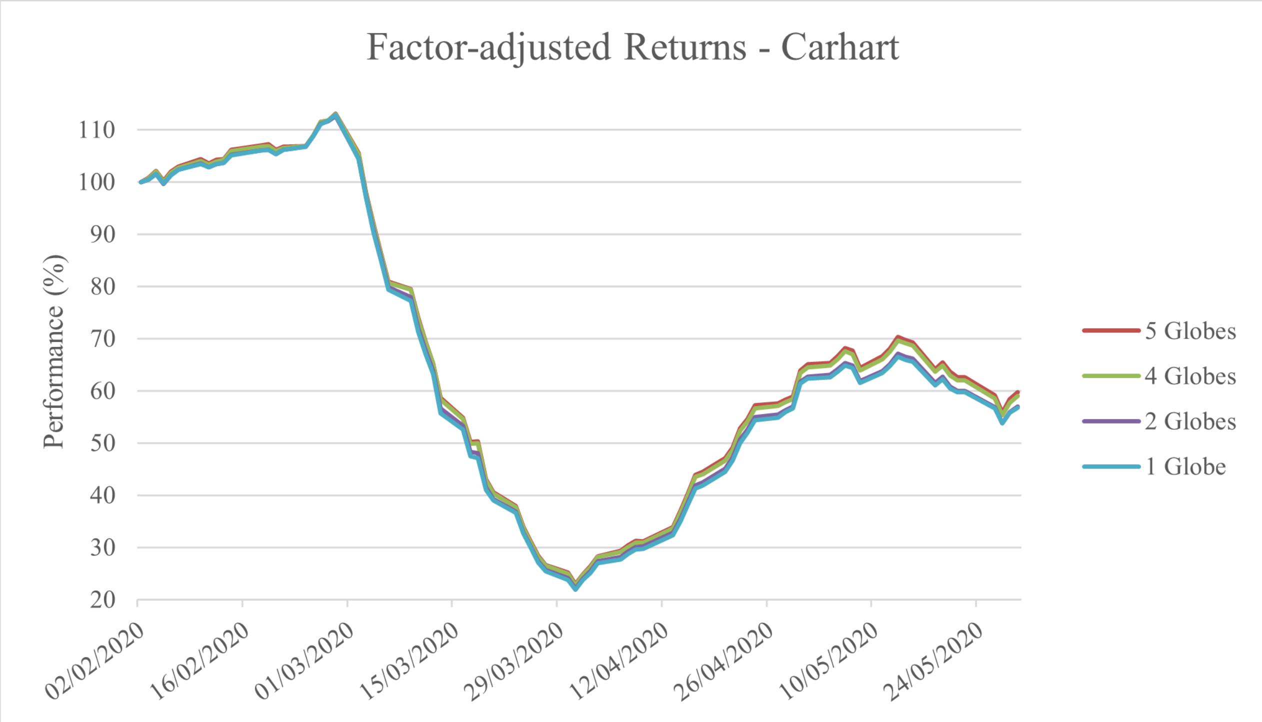


(a)

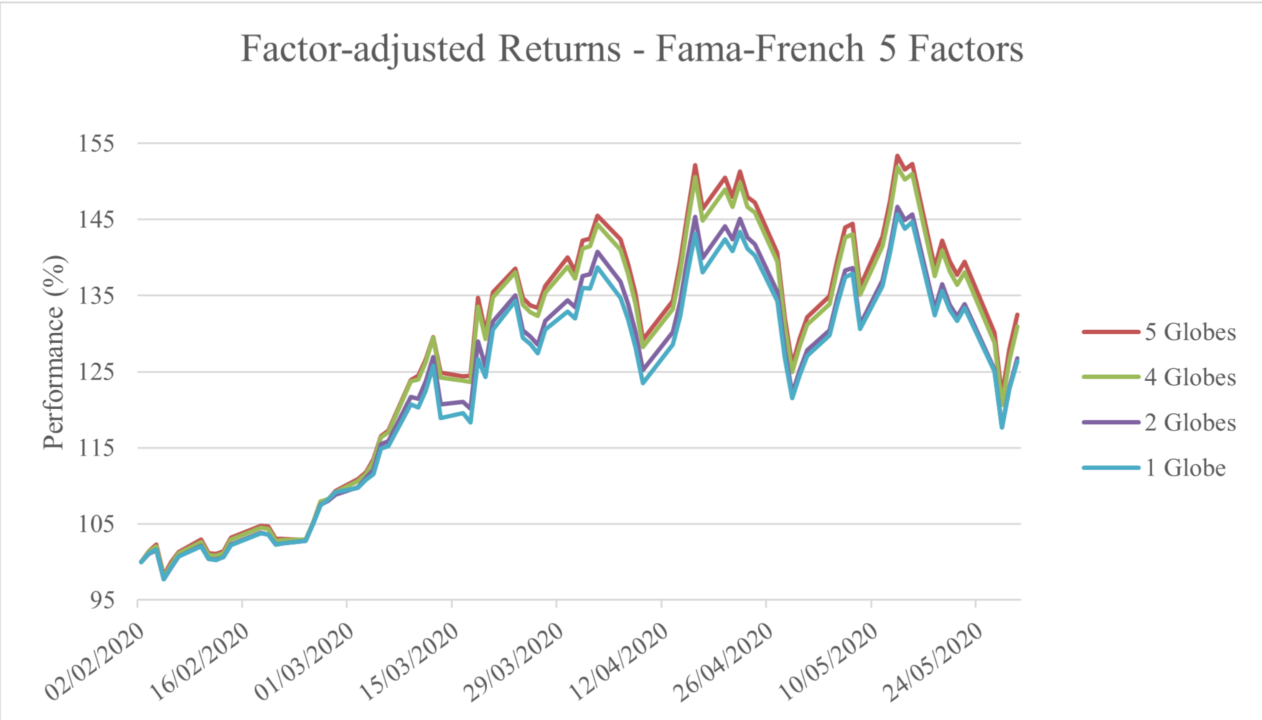


(b)

Figure B.5 – Performances broken down by Sustainability Ratings
In these two graphs, the average returns for Benchmark-adjusted and CAPM model are broken down by Morningstar Sustainability rating. Source: own research.



(c)



(d)

Figure B.5 – Performances broken down by Sustainability Ratings (cont.)
In these two graphs, the average returns for Carhart and Fama-French 5 factors models are broken down by Morningstar Sustainability rating. Source: own research.

Figure B.6 – Performances (Benchmark-adjusted Returns) broken down by Low Carbon Designation
This figure represents the average benchmark-adjusted performance of funds with (or without) the Low Carbon Designation over the studied time horizon. Only 597 funds out of the 2'396 that compose the dataset have the LCD at the beginning of February 2020, while 1'799 do not have the designation. Source: own research.

Table B.1 – Summary Statistics[a]

Statistic	Mean	St. Dev.
Fund Size	1’654’579’842.00	5’885’760’226.00
Dollar Flows	−269’763.20	11’836’294.00
Percentage Flows	−0.0002	0.015
Net Returns	−0.052	3.439
Abnormal Returns	−0.044	1.281
CAPM	−0.058	1.226
Carhart	−0.527	5.282
Fama-French 5 Factors	0.355	3.226

Statistic	Min	Pctl(25)	Median	Pctl(75)	Max
Fund Size	15’000’000.00	89’554’960.00	335’804’248.00	1’137’834’610.00	132’358’972,700.00
Dollar Flows	−1’117’997’642.00	−428’131.30	−31’130.39	89’473.77	1’112’158’134.00
Percentage Flows	−1.883	−0.001	−0.0002	0.0004	0.890
Net Returns	−34.116	−1.637	0.161	1.633	30.748
Abnormal Returns	−26.276	−0.565	−0.028	0.499	24.738
CAPM	−23.132	−0.539	−0.038	0.475	24.042
Carhart	−31.432	−5.109	0.467	2.461	17.517
Fama-French 5 Factors	−25.052	−1.532	0.830	2.476	20.870

[a] All the tables in the study are produced using R and the stargazer package (Hlavac, 2022).

Table B.2 – Returns: Pooled OLS regression for the entire time horizon

This table shows the outcomes of running the first regression model on all the different sets of alphas (or abnormal returns) for the entire time horizon of the study. The dependent variables are, respectively, the Benchmark-adjusted returns (1) and the Factor-adjusted returns with CAPM (2), Carhart (3) and Fama-French 5 factors models (4). Returns are already expressed in percentage and net of expense. Global category fixed effects are applied. Control variables include the log of fund size, the days passed from the fund's inception date, the turnover ratio of 2019 and 2020, the net expense ratio of 2019 and 2020, the Excess returns from the past year (as computed by Morningstar) and the Morningstar Star Rating.

	Dependent variable:			
	α	α^{CAPM}	$\alpha^{Carhart}$	α^{FF5}
	(1)	(2)	(3)	(4)
LCD	-0.079^{**}	-0.041	-1.981^{***}	0.304^{***}
	(0.035)	(0.034)	(0.141)	(0.088)
High ESG	0.062	0.003	-0.719^{***}	0.0005
	(0.057)	(0.054)	(0.227)	(0.142)
Above Average ESG	-0.008	-0.016	-0.117	0.010
	(0.029)	(0.027)	(0.115)	(0.072)
Below Average ESG	-0.034	-0.019	-0.112	-0.005
	(0.028)	(0.027)	(0.114)	(0.071)
Low ESG	-0.066	-0.014	0.028	-0.117
	(0.063)	(0.060)	(0.250)	(0.157)
Observations	196,472	196,472	196,472	196,472
R^2	0.002	0.004	0.071	0.022
Adjusted R^2	0.010	0.008	0.060	0.009
Controls	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table B.3 – Returns: Pooled OLS regression for the Pre-Crisis period

This table shows the outcomes of running the first regression model on all the different sets of alphas (or abnormal returns) for the first subperiod of the studied time horizon, namely the Pre-Crisis. The dependent variables are, as before, the Benchmark-adjusted returns (1) and the Factor-adjusted returns with CAPM (2), Carhart (3) and Fama-French 5 factors models (4). Returns are already expressed in percentage and net of expense. This model does not incorporate Fixed effects. Control variables include the log of fund size, the days passed from the fund's inception date, the turnover ratio of 2019 and 2020, the net expense ratio of 2019 and 2020, the Excess returns from the past year (as computed by Morningstar) and the Morningstar Star Rating. The Pre-Crisis period includes data from February 3 to February 21, 2020.

	Dependent variable:			
	α	α^{CAPM}	$\alpha^{Carhart}$	α^{FF5}
	(1)	(2)	(3)	(4)
LCD	0.014*	0.046***	0.047***	0.047*
	(0.007)	(0.008)	(0.016)	(0.024)
High ESG	−0.037***	−0.028*	−0.027	−0.035
	(0.011)	(0.011)	(0.023)	(0.035)
Above Average ESG	−0.020***	−0.013	−0.013	−0.013
	(0.008)	(0.008)	(0.016)	(0.024)
Below Average ESG	0.003	−0.001	−0.001	−0.001
	(0.007)	(0.008)	(0.016)	(0.024)
Low ESG	0.018*	0.015	0.015	0.014
	(0.011)	(0.011)	(0.023)	(0.035)
Observations	33,544	33,544	33,544	33,544
R^2	0.013	0.015	0.004	0.002
Adjusted R^2	0.012	0.014	0.003	0.001
Controls	Yes	Yes	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table B.4 – Returns: Pooled OLS regression for the Crash period

This table shows the outcomes of running the first regression model on all the different sets of alphas (or abnormal returns) for the second subperiod of the studied time horizon, namely the Crash. The dependent variables are, as before, the Benchmark-adjusted returns (1) and the Factor-adjusted returns with CAPM (2), Carhart (3) and Fama-French 5 factors models (4). Returns are already expressed in percentage and net of expense. This model does not incorporate Fixed effects. Control variables include the log of fund size, the days passed from the fund's inception date, the turnover ratio of 2019 and 2020, the net expense ratio of 2019 and 2020, the Excess returns from the past year (as computed by Morningstar) and the Morningstar Star Rating. The Crash period includes data from February 22 to March 23, 2020.

	Dependent variable:			
	α	α^{CAPM}	$\alpha^{Carhart}$	α^{FF5}
	(1)	(2)	(3)	(4)
LCD	0.234***	0.015	−0.278***	0.048
	(0.021)	(0.020)	(0.053)	(0.034)
High ESG	0.023	0.032	0.094	−0.047
	(0.031)	(0.030)	(0.078)	(0.051)
Above Average ESG	0.054**	−0.004	0.063	−0.014
	(0.022)	(0.021)	(0.055)	(0.036)
Below Average ESG	−0.031	−0.018	−0.018	−0.018
	(0.021)	(0.021)	(0.054)	(0.035)
Low ESG	−0.011	−0.002	0.027	−0.005
	(0.031)	(0.031)	(0.079)	(0.052)
Observations	50,316	50,316	50,316	50,316
R^2	0.015	0.007	0.003	0.003
Adjusted R^2	0.015	0.007	0.003	0.002
Controls	Yes	Yes	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table B.5 – Returns: Pooled OLS regression for the Recovery period

This table shows the outcomes of running the first regression model on all the different sets of alphas (or abnormal returns) for the third subperiod of the studied time horizon, namely the Recovery. The dependent variables are, as before, the Benchmark-adjusted returns (1) and the Factor-adjusted returns with CAPM (2), Carhart (3) and Fama-French 5 factors models (4). Returns are already expressed in percentage and net of expense. This model does not incorporate Fixed effects. Control variables include the log of fund size, the days passed from the fund's inception date, the turnover ratio of 2019 and 2020, the net expense ratio of 2019 and 2020, the Excess returns from the past year (as computed by Morningstar) and the Morningstar Star Rating. The Recovery period includes data from March 24 to May 29, 2020.

	Dependent variable:			
	α	α^{CAPM}	$\alpha^{Carhart}$	α^{FF5}
	(1)	(2)	(3)	(4)
LCD	-0.012	0.078^{***}	0.081^{**}	0.082^{***}
	(0.009)	(0.008)	(0.041)	(0.027)
High ESG	-0.026^{*}	-0.006	-0.042	-0.010
	(0.013)	(0.013)	(0.062)	(0.042)
Above Average ESG	-0.022^{**}	0.004	0.026	-0.002
	(0.009)	(0.009)	(0.043)	(0.029)
Below Average ESG	0.017^{*}	0.015^{*}	0.016	0.015
	(0.009)	(0.009)	(0.042)	(0.028)
Low ESG	0.034^{**}	0.034^{***}	0.033	0.032
	(0.013)	(0.013)	(0.062)	(0.042)
Observations	112,612	112,612	112,612	112,612
R^2	0.001	0.003	0.0001	0.0003
Adjusted R^2	0.001	0.003	0.00002	0.0002
Controls	Yes	Yes	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table B.6 – Flows: Pooled OLS regression for the entire time horizon

This table shows the outcomes of running the first regression model on the percentage flows variable for the entire time horizon of the study. Percentage flows are computed daily as Dollar flows over TNA. Global category fixed effects are applied. Control variables include the log of fund size, the days passed from the fund's inception date, the turnover ratio of 2019 and 2020, the net expense ratio of 2019 and 2020, the Excess returns from the past year (as computed by Morningstar) and the Morningstar Star Rating.

	Dependent variable:
	Percentage Flows
LCD	0.001^{**}
	(0.0004)
High ESG	-0.0004
	(0.001)
Above Average ESG	-0.0003
	(0.0003)
Below Average ESG	-0.001^{*}
	(0.0003)
Low ESG	-0.0005
	(0.001)
Observations	196,472
R^2	0.003
Adjusted R^2	-0.010
Controls	Yes
Fixed Effects	Yes
Note:	*p<0.1; **p<0.05; ***p<0.01

Table B.7 – Flows: Pooled OLS regression for the three different subperiods

This table shows the outcomes of running the first regression model on the percentage flows variable for the three different subperiods of the study. Percentage flows are computed daily as Dollar flows over TNA. These regressions do not incorporate Fixed effects. Control variables include the log of fund size, the days passed from the fund's inception date, the turnover ratio of 2019 and 2020, the net expense ratio of 2019 and 2020, the Excess returns from the past year (as computed by Morningstar) and the Morningstar Star Rating.

	Dependent variable: Percentage Flows		
	Pre-Crisis	Crash	Recovery
	(1)	(2)	(3)
LCD	−0.0002	−0.0003	−0.0001
	(0.0002)	(0.0002)	(0.0001)
High ESG	−0.0001	0.001*	0.00001
	(0.0003)	(0.0003)	(0.0002)
Above Average ESG	0.0002	−0.0002	−0.0001
	(0.0002)	(0.0002)	(0.0001)
Below Average ESG	−0.0001	0.0004	−0.0001
	(0.0002)	(0.0002)	(0.0001)
Low ESG	0.0003	−0.0003	0.0001
	(0.0003)	(0.0003)	(0.0002)
Observations	33,544	50,316	112,612
R^2	0.002	0.004	0.003
Adjusted R^2	0.001	0.004	0.003
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.8 – Returns: Difference-in-Differences regression for the entire time horizon

This table shows the outcomes of running the second regression model on the abnormal returns for the entire time horizon of the study. As before, the dependent variables are, respectively, the Benchmark-adjusted returns (1) and the Factor-adjusted returns with CAPM (2), Carhart (3) and Fama-French 5 factors models (4). COVID represent a dummy variable equal to 1 starting in the week ending on February 22 and 0 otherwise, and used as interaction term. High ESG is a dummy equal to 1 for funds with 4 and 5 Globes and 0 otherwise, and the same applies to Low ESG for funds with 2 and 1 Globes. Returns are already expressed in percentage and net of expense. Global category fixed effects are applied. Control variables are the same as in the previous regression model, including the LCD this time.

	Dependent variable:			
	α	α^{CAPM}	$\alpha^{Carhart}$	α^{FF5}
	(1)	(2)	(3)	(4)
High ESG × COVID	0.032*	0.011	−0.023	0.031
	(0.019)	(0.018)	(0.073)	(0.047)
Low ESG × COVID	−0.004	−0.010	0.048	−0.016
	(0.019)	(0.018)	(0.073)	(0.047)
Observations	196,472	196,472	196,472	196,472
R^2	0.002	0.004	0.120	0.032
Adjusted R^2	−0.010	−0.008	0.109	0.020
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table B.9 – Returns: Difference-in-Differences regression for the Crash period

This table shows the outcomes of running the second regression model on the abnormal returns for the second subperiod of the studied time horizon, namely the Crash. As before, the dependent variables are, respectively, the Benchmark-adjusted returns (1) and the Factor-adjusted returns with CAPM (2), Carhart (3) and Fama-French 5 factors models (4). COVID represent a dummy variable equal to 1 starting in the week ending on February 22 and 0 otherwise, and used as interaction term. High ESG is a dummy equal to 1 for funds with 4 and 5 Globes and 0 otherwise, and the same applies to Low ESG for funds with 2 and 1 Globes. Returns are already expressed in percentage and net of expense. Global category fixed effects are applied. Control variables are the same as in the previous regression model, including the LCD this time. The Crash period includes data from February 22 to March 23, 2020.

	Dependent variable:			
	α	α^{CAPM}	$\alpha^{Carhart}$	α^{FF5}
	(1)	(2)	(3)	(4)
High ESG × COVID	0.046**	0.011	0.071	0.022
	(0.020)	(0.020)	(0.051)	(0.033)
Low ESG × COVID	−0.026	−0.014	−0.006	−0.015
	(0.020)	(0.019)	(0.050)	(0.033)
Observations	50,316	50,316	50,316	50,316
R^2	0.015	0.007	0.003	0.003
Adjusted R^2	0.015	0.007	0.003	0.002
Controls	Yes	Yes	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table B.10 – Returns: Difference-in-Differences regression for the Recovery period

This table shows the outcomes of running the second regression model on the abnormal returns for the third subperiod of the studied time horizon, namely the Recovery. As before, the dependent variables are, respectively, the Benchmark-adjusted returns (1) and the Factor-adjusted returns with CAPM (2), Carhart (3) and Fama-French 5 factors models (4). COVID represent a dummy variable equal to 1 starting in the week ending on February 22 and 0 otherwise, and used as interaction term. High ESG is a dummy equal to 1 for funds with 4 and 5 Globes and 0 otherwise, and the same applies to Low ESG for funds with 2 and 1 Globes. Returns are already expressed in percentage and net of expense. Global category fixed effects are applied. Control variables are the same as in the previous regression model, including the LCD this time. The Recovery period includes data from March 24 to May 29, 2020.

	Dependent variable:			
	α	α^{CAPM}	$\alpha^{Carhart}$	α^{FF5}
	(1)	(2)	(3)	(4)
High ESG × COVID	−0.023***	−0.002	0.009	−0.004
	(0.009)	(0.008)	(0.040)	(0.027)
Low ESG × COVID	0.022***	0.020**	0.020	0.019
	(0.008)	(0.008)	(0.039)	(0.026)
Observations	112,612	112,612	112,612	112,612
R^2	0.001	0.003	0.0001	0.0003
Adjusted R^2	0.001	0.003	0.00003	0.0002
Controls	Yes	Yes	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table B.11 – Flows: Difference-in-Differences regression for the entire time horizon

This table shows the outcomes of running the second regression model on the percentage flows variable for the entire time horizon of the study. Percentage flows are computed daily as Dollar flows over TNA. COVID represent a dummy variable equal to 1 starting in the week ending on February 22 and 0 otherwise, and used as interaction term. High ESG is a dummy equal to 1 for funds with 4 and 5 Globes and 0 otherwise, and the same applies to Low ESG for funds with 2 and 1 Globes. Global category fixed effects are applied. Control variables are the same as in the previous regression model, including the LCD.

	Dependent variable:
	Percentage Flows
High ESG × COVID	−0.000269
	(0.000220)
Low ESG × COVID	0.0000468
	(0.000221)
Observations	196,472
R^2	0.003
Adjusted R^2	−0.009
Controls	Yes
Fixed Effects	Yes
Note:	*p<0.1; **p<0.05; ***p<0.01

Table B.12 – Flows: Difference-in-Differences regression for Crash and Recovery subperiods

This table shows the outcomes of running the second regression model on the percentage flows variable for the two remaining subperiods of the study, excluding the Pre-Crisis one. Percentage flows are computed daily as Dollar flows over TNA. COVID represent a dummy variable equal to 1 starting in the week ending on February 22 and 0 otherwise, and used as interaction term. High ESG is a dummy equal to 1 for funds with 4 and 5 Globes and 0 otherwise, and the same applies to Low ESG for funds with 2 and 1 Globes. Global category fixed effects are applied. Control variables are the same as in the previous regression model, including the LCD.

	Dependent variable: Percentage Flows	
	Crash	Recovery
	(1)	(2)
High ESG × COVID	0.0000331	−0.000072
	(0.000214)	(0.000102)
Low ESG × COVID	0.000196	−0.000070
	(0.000209)	(0.000100)
Observations	50,316	112,612
R^2	0.004	0.003
Adjusted R^2	0.003	0.003
Controls	Yes	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01

Bibliography

R. Albuquerque, Y. Koskinen, S. Yang, and C. Zhang. Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash. *The Review of Corporate Finance Studies*, 9(3):593–621, 07 2020. ISSN 2046-9128. doi: 10.1093/rcfs/cfaa011. URL `https://doi.org/10.1093/rcfs/cfaa011`.

K.-H. Bae, S. El Ghoul, Z. J. Gong, and O. Guedhami. Does CSR matter in times of crisis? Evidence from the COVID-19 pandemic. *Journal of Corporate Finance*, 67:101876, 2021. ISSN 0929-1199. doi: https://doi.org/10.1016/j.jcorpfin.2020.101876. URL `https://www.sciencedirect.com/science/article/pii/S0929119920303205`.

T. Bari, R. Bradley, P. Finnerty, B. Geis, M. Spiryda, and M. VanDemark. Mutual Fund industry outlook. `https://www.pwc.com/us/en/industries/financial-services/library/pdf/awm-mutual-fund-outlook-6-22-22.pdf`, 2022. PwC. Accessed on: 12.09.2022.

C. Barr, D. Doman, and V. Redensek. Morningstar® Sustainability Rating Methodology, 2021. URL `https://www.morningstar.com/content/dam/marketing/shared/research/methodology/744156_Morningstar_Sustainability_Rating_for_Funds_Methodology.pdf`.

W. J. Baumol, W. E. Oates, et al. *Economics, environmental policy, and the quality of life.* Englewood Cliffs, N.J., Prentice-Hall, 1979.

F. Berg, K. Fabisik, and Z. Sautner. Is history repeating itself? The (un)predictable past of ESG ratings. *The (Un)Predictable Past of ESG Ratings (August 24, 2021). European Corporate Governance Institute – Finance Working Paper*, 708, 2020.

F. Berg, J. F. Kölbel, and R. Rigobon. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6):1315–1344, 2022.

D. Blitz, P. van Vliet, and H. Matthias. Fama-French 5-factor model: why more is not always better. https://www.robeco.com/ch/en/insights/2022/03/fama-french-5-factor-model-why-more-is-not-always-better.html, 2022. Robeco, Insights. Accessed on: 20.11.2022.

L. Capo McCormick, C. Torres, M. Benhamou, and D. Pogkas. The Covid-19 Pandemic has added $19.5 trillion to global debt. https://www.bloomberg.com/graphics/2021-coronavirus-global-debt/?sref=qPle0K9J, 2021. Bloomberg. Accessed on: 10.09.2022.

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

M. Ceccarelli, S. Ramelli, and A. F. Wagner. Low-carbon mutual funds. *Swiss Finance Institute Research Paper*, (19-13), 2021. doi: http://dx.doi.org/10.2139/ssrn.3353239. European Corporate Governance Institute – Finance Working Paper No. 659/2020.

H. Chen, L. Cohen, and U. Gurun. Don't take their word for it: The misclassification of bond mutual funds. Working Paper 26423, National Bureau of Economic Research, November 2019. URL http://www.nber.org/papers/w26423.

B. Cornell. ESG preferences, risk and return. *European Financial Management*, 27(1):12–19, 2021. doi: https://doi.org/10.1111/eufm.12295. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/eufm.12295.

K. M. Cremers, J. A. Fulkerson, and T. B. Riley. Challenging the conventional wisdom on active management: A review of the past 20 years of academic literature on actively managed mutual funds. *Financial Analysts Journal*, 75(4):8–35, 2019.

L. Delevingne, A. Gründler, S. Kane, and T. Koller. The ESG premium: New perspectives on value and performance. `https://www.mckinsey.com/business-functions/sustainability/our-insights/the-esg-premium-new-perspectives-on-value-and-performance`, 2020. McKinsey & Company. Accessed on: 12.09.2022.

E. Demers, J. Hendrikse, P. Joos, and B. Lev. ESG did not immunize stocks during the COVID-19 crisis, but investments in intangible assets did. *Journal of Business Finance & Accounting*, 48(3-4):433–462, 2021. doi: https://doi.org/10.1111/jbfa.12523. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/jbfa.12523`.

W. Ding, R. Levine, C. Lin, and W. Xie. Corporate Immunity to the COVID-19 Pandemic. Working Paper 27055, National Bureau of Economic Research, April 2020. URL `http://www.nber.org/papers/w27055`.

R. Döttling and S. Kim. Sustainability preferences under stress: Mutual fund flows during COVID-19. *VoxEU.org*, 19, 2020. Working Paper.

E. J. Elton, M. J. Gruber, and C. R. Blake. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and morningstar mutual fund databases. *The Journal of Finance*, 56(6):2415–2430, 2001.

European Commission. Financing the green transition: The european green deal investment plan and just transition mechanism, 2020. URL `https://ec.europa.eu/commission/presscorner/detail/en/ip_20_17`.

A. Falato, I. Goldstein, and A. Hortaçsu. Financial fragility in the covid-19 crisis: The case of investment funds in corporate bond markets. *Journal of Monetary Economics*, 123:35–52, 2021.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. ISSN 0304-405X. doi: https://doi.org/10.1016/0304-405X(93)90023-5. URL https://www.sciencedirect.com/science/article/pii/0304405X93900235.

E. F. Fama and K. R. French. Luck versus Skill in the Cross-Section of Mutual Fund Returns. *The Journal of Finance*, 65(5):1915–1947, 2010. doi: https://doi.org/10.1111/j.1540-6261.2010.01598.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2010.01598.x.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2014.10.010. URL https://www.sciencedirect.com/science/article/pii/S0304405X14002323.

M. Faria e Castro. Domestic Debt Before and After the Pandemic Recession. https://www.stlouisfed.org/on-the-economy/2021/september/domestic-debt-pandemic-recession#authorbox, 2021. Federal Reserve Bank of St. Louis. Accessed on: 10.09.2022.

A. Garel and A. Petit-Romec. Investor rewards to environmental responsibility: Evidence from the COVID-19 crisis. *Journal of Corporate Finance*, 68:101948, 2021.

B. Gerard. ESG and Socially Responsible Investment: A Critical Review. *Beta*, 33(1):61–83, 2019. doi: 10.18261/issn.1504-3134-2019-01-05. URL https://www.idunn.no/doi/abs/10.18261/issn.1504-3134-2019-01-05.

V. Glode. Why mutual funds “underperform”. *Journal of Financial Economics*, 99(3):546–559, 2011.

S. Glossner, P. Matos, S. Ramelli, A. F. Wagner, et al. Where Do Institutional Investors Seek Shelter when Disaster Strikes?: Evidence from COVID-19. Technical report, Swiss Finance Institute Zurich, 2020.

J. Hale. Morningstar® Low Carbon Designation™, 2018. URL `https://s21.q4cdn.com/198919461/files/doc_news/2018/Morningstar-Low-Carbon-Designation-Methodology-Final.pdf`.

J. Hale. Sustainable Equity Funds Outperform Traditional Peers in 2020. `https://www.morningstar.com/articles/1017056/sustainable-equity-funds-outperform-traditional-peers-in-2020`, 2021. Morningstar, Sustainability Matters. Accessed on: 15.09.2022.

S. M. Hartzmark and A. B. Sussman. Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. *The Journal of Finance*, 74(6):2789–2837, 2019. doi: https://doi.org/10.1111/jofi.12841. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12841`.

M. Hlavac. *stargazer: Well-Formatted Regression and Summary Statistics Tables.* Social Policy Institute, Bratislava, Slovakia, 2022. URL `https://CRAN.R-project.org/package=stargazer`. R package version 5.2.3.

IMF Policy Development and Review Department. Fund assistance for countries facing exogenous shocks, 2003. URL `https://www.imf.org/external/np/pdr/sustain/2003/080803.pdf`.

R. Jebe. The Convergence of Financial and ESG Materiality: Taking Sustainability Mainstream. *American Business Law Journal*, 56(3):645–702, 2019. doi: https://doi.org/10.1111/ablj.12148. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/ablj.12148`.

M. C. Jensen. The Performance of Mutual Funds in the Period 1945–1964. *The Journal of Finance*, 23(2):389–416, 1968. doi: https://doi.org/10.1111/j.1540-6261.1968.tb00815.x.

URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1968.tb00815.x.

M. Kacperczyk, S. V. Nieuwerburgh, and L. Veldkamp. Time-varying fund manager skill. *The Journal of Finance*, 69(4):1455–1484, 2014.

M. Mazur, M. Dang, and M. Vega. COVID-19 and the march 2020 stock market crash. Evidence from SP1500. *Finance Research Letters*, 38:101690, 2021. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2020.101690. URL https://www.sciencedirect.com/science/article/pii/S1544612320306668.

J. Miklian and K. Hoelscher. SMEs and exogenous shocks: A conceptual literature review and forward research agenda. *International Small Business Journal*, 40(2):178–204, 2022. doi: 10.1177/02662426211050796. URL https://doi.org/10.1177/02662426211050796.

Morningstar. The Morningstar Rating™for Funds, 2021. URL https://www.morningstar.com/content/dam/marketing/shared/research/methodology/771945_Morningstar_Rating_for_Funds_Methodology.pdf.

L. Pástor and R. F. Stambaugh. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63(3):315–349, 2002.

L. Pástor and M. B. Vorsatz. Mutual fund performance and flows during the COVID-19 crisis. *The Review of Asset Pricing Studies*, 10(4):791–833, 2020.

L. Pástor, R. F. Stambaugh, and L. A. Taylor. Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45, 2015. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2014.11.008. URL https://www.sciencedirect.com/science/article/pii/S0304405X14002542.

E. Pollman. Corporate Social Responsibility, ESG, and Compliance. *Forthcoming, Cambridge*

Handbook of Compliance (D. Daniel Sokol & Benjamin van Rooij eds.), Loyola Law School, Los Angeles Legal Studies Research Paper, (2019-35), 2019.

S. Ramelli and A. F. Wagner. Feverish Stock Price Reactions to COVID-19*. *The Review of Corporate Finance Studies*, 9(3):622–655, 07 2020. ISSN 2046-9128. doi: 10.1093/rcfs/cfaa012. URL https://doi.org/10.1093/rcfs/cfaa012.

S. Ramelli, A. F. Wagner, R. J. Zeckhauser, and A. Ziegler. Investor rewards to climate responsibility: Evidence from the 2016 climate policy shock. Technical report, National Bureau of Economic Research, 2018. Available at SSRN: https://ssrn.com/abstract=3294878.

B. A. Sensoy. Performance evaluation and self-designated benchmark indexes in the mutual fund industry. *Journal of Financial Economics*, 92(1):25–39, 2009. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2008.02.011. URL https://www.sciencedirect.com/science/article/pii/S0304405X08002109.

E. R. Sirri and P. Tufano. Costly search and mutual fund flows. *The journal of finance*, 53 (5):1589–1622, 1998.

N. N. Taleb. *The black swan: The impact of the highly improbable*, volume 2. Random house, 2007.

K. Wallace. How to Choose Among Fund Share Classes. https://www.morningstar.com/articles/788979/how-to-choose-among-fund-share-classes, 2017. Morningstar, The Short Answer. Accessed on: 20.11.2022.

P. Wells. S&P 500 suffers its quickest fall into bear market on record. https://www.ft.com/content/d895a54c-64a4-11ea-a6cd-df28cc3c6a68, 2020. Financial Times. Accessed on: 06.09.2022.

A. Willis. ESG as an equity vaccine. https://www.morningstar.ca/ca/news/201741/esg-as-an-equity-vaccine.aspx, 2020. Morningstar, Market Insights. Accessed on: 15.09.2022.

B. Yoon, J. H. Lee, and R. Byun. Does ESG Performance Enhance Firm Value? Evidence from Korea. *Sustainability*, 10(10), 2018. ISSN 2071-1050. doi: 10.3390/su10103635. URL https://www.mdpi.com/2071-1050/10/10/3635.