CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of economic studies

Is ESG a resiliency factor for company stock returns during a crisis? Evidence from Europe during the covid-19 pandemic

Bachelor's thesis

Author: David Krameš

Study program: Economics and Finance

Supervisor: Jiří Novák M.Sc., Ph.D.

Year of defense: 2023

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title. The author further declares that, in line with IES and Charles University AI usage guidelines, he used generative artificial intelligence while writing this thesis in order to convert data tables to the TEX format, and that no text in this thesis was written by generative artificial intelligence.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, August 1, 2023

David Krames

Abstract

The goal of this thesis was to examine whether high ESG performance serves as a resiliency factor for company stock returns during times of crisis. Using a DID estimator for 3 different regions and treatment timings, I find that high ESG performance did serve as a resiliency factor for company stock returns in the short term during the covid-19 pandemic, with high-ESG firms having 1.125-4.785% higher stock excess log returns compared to low-ESG firms over a 15 day period. This is probably a result of their lower perceived riskiness. I also find this effect is primarily driven by the S pillar and for European companies, by firms belonging to the Financial and Healthcare industries. In the long term, I find that the effect reverses and ESG becomes a negative factor, which I believe is caused by investors starting to seek riskier investments again. Finally, for European and American firms, I find the effect of a high score in the G pillar is negative even in normal times.

JEL Classification	G01, G12, G32
Keywords	ESG, CSR, Stock returns, Investing, Crisis, Resiliency factor

Title	Is ESG a resiliency factor for company stock returns during a crisis? Evidence from Europe during the covid-19 pandemic

Author's e-mail	66507321@fsv.cuni.cz
Supervisor's e-mail	jiri.novak@fsv.cuni.cz

Abstrakt

Cílem této práce bylo zkoumat, zda vysoký výkon v oblasti ESG slouží jako faktor odolnosti pro výnosnost akcií firem během krizí. Použitím DID estimátoru pro 3 různé regiony a časování treatmentů jsem zjistil, že vysoký výkon v oblasti ESG skutečně sloužil jako faktor odolnosti pro výnosnost akcií ve krátkodobém horizontu během pandemie covid-19, přičemž firmy s vysokým ESG měly o 1,125 až 4,785 % vyšší přebytečné logaritmické výnosy akcií ve srovnání s firmami s nízkým ESG během 15 dnů. Toto je pravděpodobně důsledek jejich nižší vnímané rizikovosti. Zároveň jsem zjistil, že tento efekt je poháněn především pilířem S a pro evropské společnosti zejména finančnickými a zdravotnickými firmami. V delším období jsem zjistil, že se tento efekt obrací a ESG se stává negativním faktorem. Věřím, že důvodem pro tento jev jsou preference investorů, kteří opět začínají vyhledávat rizikovější investice. Finálně jsem zjistil, že pro evropské a americké firmy je efekt vysokého skóre v pilíři G negativní i během normálního období.

Klasifikace JEL	G01, G12, G32
Klíčová slova	ESG, CSR, Návratnost akcií, Investice, Krize, Faktor odolnosti
Název práce	Je ESG faktorem odolnosti pro akciové výnosy firem během krize? Důkazy z Evropy během pandemie covid-19

E-mail autora	66507321@fsv.cuni.cz
E-mail vedoucího práce	jiri.novak@fsv.cuni.cz

Acknowledgments

The author is grateful especially to Jiří Novák M.Sc., Ph.D. for his invaluable guidance, goodwill, and patience.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Krameš, David: *Is ESG a resiliency factor for company stock returns during a crisis? Evidence from Europe during the covid-19 pandemic.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of economic studies, Prague. 2023, pages 70. Advisor: Jiří Novák M.Sc., Ph.D.

Contents

List of Tables

List of Figures

Acronyms

CAPM Capital Asset Pricing Model

CHI Containment and Health Index

CO2 Carbon dioxide

CSR Corporate social responsibility

DID Difference-in-Differences estimator

ESG Environmental, Social, & Governance

ESI Economic Support Index

FFC Fama-French-Carhart

GFC Great Financial Crisis

HML High Minus Low

MLR Multiple Linear Regression

NFRF Non-Financial Reporting Directive

OECD Organisation for Economic Co-operation and Development

OLS Ordinary Least Squares

SMB Small Minus Big

VIF Variance Inflation Factor

Chapter 1

Introduction

In the current social, political, and economic landscape, the issues of Environmental, Social, and Governance (ESG) seem to be more relevant than ever. ESG-related actions and prospects of companies seem to be driving investment decisions more and more every day and many companies have been filling in a Chief Sustainability Officer position and their prominence is expected to increase in the following years (Oakey 2021). What's more, ESG as a topic in academic research has also been gaining on popularity and some ESG-focused journals have been gaining prominence and credibility. Even so, the amount of certain kinds of sustainability research is still relatively sparse, as is the case of the topic discussed in this thesis, namely the effects of ESG during crises.

While many support company ESG activities as they align with their own values, investors might wonder whether ESG spending brings about any value to them and whether or not it is an example of the agency problem. Using the recent exogenous shock brought about by the covid-19 pandemic, I set out to explore the question of whether high ESG performance serves as a resiliency factor during crises, following such shocks.

To empirically test this, I explore the shock in three different regions using three different event windows during the first quarter of 2020 and look at the difference in how high- and low-ESG firms' excess stock returns respond to it. I also perform a more long-term regression with two years of stock return data to see whether this effect stays, diminishes, or changes over time. I use the difference-in-differences estimator to try to capture the causal effect of ESG on company stock returns during the crisis period. Finally, I run numerous sets of modified regressions using which I test the robustness of my results.

The thesis is structured as follows: Chapter 2 provides a review of the current academic literature regarding ESG investing, the effect of ESG on stock returns, and the specific effect on stock returns during crises. Chapter 3 describes the datasets I use in the analysis and the methodology I use to perform the analysis, as well as states the three thesis hypotheses. Chapter 4 includes the analysis itself, a discussion of the results, several robustness analyses and their discussion as well as the limitations of this thesis and suggestions for future research. Chapter 5 summarizes all the findings and concludes the thesis.

Chapter 2

Literature review

In this chapter, I will summarize current research within the field of ESG investing and how its conclusions relate to the research question of this thesis, that is what current research says about ESG being a resiliency factor during crises. A core source for this chapter was the comprehensive ESG literature review by Gillan *et al.* (2021).

2.1 ESG as a risk-mitigating factor

A core assumption I make in explaining the results found in this thesis is the fact that investors consider ESG performance to be a risk-mitigating factor. While many topics within the field of ESG research are hotly debated, the perception of ESG as a factor mitigating all kinds of risk, including systematic or credit risk, as well as lowering firms' cost of capital is almost universally agreed on (Albuquerque *et al.* 2019; Chava 2014; Jiraporn *et al.* 2014). Furthermore, van Duuren *et al.* (2016) find that conventional asset managers adopt ESG investing practices for risk mitigation, and this finding is not in any way dependent on personal values. Important for this thesis, the study also finds a substantial difference in the perception and use of ESG by American and European asset managers, specifically, American perceive the ESG aspect as considerably less important and impactful than their European counterparts.

As for the reasoning behind this lowered risk of high-ESG firms, several authors put up and test many different hypotheses. For example, Albuquerque *et al.* (2019) find that lower systematic risk faced by high-ESG firms is the result of product differentiation. This means that the consumer demand these firms face is much less price-elastic than those of low-ESG (and therefore less-differentiated) companies. I theorize this product differentiation theory, if true, could strongly help these firms' performance during the pandemic, due to the less price-elastic demand (suggesting a lower drop in demand compared to other firms). From another point of view, two different papers come to the conclusion that the lowered cost of capital faced by high-ESG firms comes from a wider investor base of such firms, compared to low-ESG firms (El Ghoul *et al.* 2011; Hong & Kacperczyk 2009). I again theorize this effect could be beneficial during the pandemic, as during a time of market uncertainty, high-ESG firms could face a lower risk of position closing by either a few large investors, or a large number of specific kinds of investors.

One interesting exception to these conclusions comes from the paper by Breuer *et al.* (2018). They find that in countries with weak investor protection laws, ESG performance actually increases cost of capital. However, they also find lower cost of capital for high-ESG firms for firms in countries with strong protection laws. Given the measure of investor protection law strength used in the paper, I find the countries in my samples have generally higher amounts of investor protection laws. This means that the contrary finding in this paper should not be of concern for the risk-mitigation assumption.

As I said in the first paragraph, verifying the assumption that ESG lowers the risk faced by companies is important for my discussion later in this thesis. The almost universal acceptance of this idea, its practical application by asset managers, as well as the numerous papers that have been written about it, is strong and robust evidence for it being true.

2.2 ESG and stock returns

The research on ESG and stock returns is not nearly as conclusive, though the general sentiment is that during normal times, ESG portfolios do not provide any significant increase in stock returns compared to conventional portfolios in the long-term (Halbritter & Dorfleitner 2015; Landi & Sciarelli 2019). However, not all research comes to the same conclusion. For example, Hong & Kacperczyk (2009) find a negative relationship between ESG activities and stock returns. They conclude that this relationship could be the effect of so-called sin stocks (stocks of companies producing tobacco, alcohol, gambling, etc.). These companies have low ESG scores, yet they earn large returns. Similarly, Bolton & Kacperczyk (2021) find that firms with higher amounts "anti-ESG" activities (in this case CO2 production) have higher stock returns. This finding is also supported by previous research (Brammer *et al.* 2006; Heinkel *et al.* 2001). Bolton & Kacperczyk (2021) conclude this is likely caused by investors requiring a premium for investing in these "dirty" companies. However, save the sin stocks and especially low-ESG companies, the long-term effect of ESG on stock returns indeed seems to be neutral.

Some researchers also considered short-term returns of high-ESG companies following events such as green bond issuance, ESG news, or philanthropic donations and they generally find positive relationships. Krüger (2015) finds a strong negative response in stock returns to negative ESG announcements, though they also find no response for positive news, attributing this difference to how the two kinds of news are reported. Furthermore, Flammer (2021); Tang & Zhang (2020) both find a positive market reaction following green bond issuance. These findings overall suggest that markets do react to ESG events, and do so in the expected manner. Goldstein *et al.* (2022) also conclude that ESG and conventional investors react to ESG news differently, sometimes in the completely opposite direction. This is further supported by the findings of El Ghoul *et al.* (2011); Hong & Kacperczyk (2009) on different shareholder compositions between high- and low-ESG companies. While the covid-19 shock can hardly be considered an ESG event, the literature quite clearly shows responses to shocks can differ significantly for high- and low-ESG companies.

Since accounting performance also has an impact on stock returns, I find it important to shortly explore the relationship between ESG and accounting performance. Research generally agrees that ESG scores are positively correlated with both return-on-assets and return-on-equity (Borghesi *et al.* 2014; Cornett *et al.* 2016; Lins *et al.* 2017). Borghesi *et al.* (2014) also find that ESG activities are positively related to the levels of free cash flow. However, as Henriksson *et al.* (2018) note, there is a caveat to these findings. To cite the aforementioned paper, "One caveat to remember is that ESG expenditures and disclosures are voluntary. It is well known that profitable firms are more likely to voluntarily disclose more information and they are also in a better financial position to afford spending on ESG-related activities. This casts some doubts on the direction of causality". The research nevertheless never suggests a negative effect of ESG on accounting performance, meaning I can work with the assumption that ESG will not affect stock returns negatively through this channel.

2.3 ESG and stock returns during crises

Lastly, I would like to touch on the specific topic of this thesis, that is, how ESG affects companies' stock returns during crises. The issue with this kind of research is that there is quite a limited number of crises to work with, so the research is rather sparse. Lins *et al.* (2017) find a strong positive effect of ESG during the great financial crisis in 2007-8. Furthermore, they find this effect to be even stronger for companies in counties with higher levels of societal trust. Important to note is that as a proxy for societal trust, Lins *et al.* (2017) use the number of associations per capita in a given county, a measure which could be considered considerably flawed due to spurious correlation with company ESG activities. Also, as the authors themselves put it, the great financial crisis was also a crisis in trust in institutions, which could not necessarily be said about the covid-19 pandemic. Similarly, Albuquerque *et al.* (2020) also find a positive effect of ESG among American firms, specifically looking at the pandemic. They however lack longer-term data and only look at the immediate effect. Similar findings have been reported by numerous researchers in less reputable journals (Beloskar & Rao 2023; Engelhardt *et al.* 2021; Habib & Mourad 2023). While these findings may not be as trustworthy, their relative frequency does at the very least suggest the effect could be positive.

Contrary to these findings, Demers *et al.* (2021) find that using a detailed and fully specified regression, the positive effect of ESG on stock returns disappears. Instead, they find the level of internal innovation investment to be a much better resiliency factor for firms' stock returns during the pandemic. The issue with this finding is that I am not aware of any other study that would use this level of detail in its control variables, which makes the results difficult to judge. The question also is whether such a model would not be overspecified, and whether investors actually consider this level of detail when investing. Overall, current literature, examining the two major crises in recent years, suggests ESG does serve as a resiliency factor for company stock returns.

Chapter 3

Data & Methodology

3.1 Data

I obtained firms' ESG ratings as well as a variety of their financial and non-financial characteristics from Refinitiv's Eikon database, to which I got access thanks to the Institute of Economic Studies at Charles University. Refinitiv Eikon is a global database of firm characteristics, collecting large amounts of company data from a variety of credible sources annually, as well as gathering equity (and other financial instrument) pricing data on a daily basis. As Cardillo *et al.* (2020) note, Refinitiv has the best ESG rating coverage for European firms among the large ESG data providers, making it suitable for use in this thesis. I converted all the financial data into USD using the exchange rates for the particular time period available in the Refinitiv Eikon database. For daily stock prices, I use the closing price on a given day. For monthly returns, I use the end-of-month closing price for the current and preceding months. The vast majority of data preparation was done using Microsoft Excel and the Datastream plugin from Refinitiv. The rest of the preparation, as well as the data analysis itself, was done using R inside of RStudio.

I further obtained data on the number of covid-19 cases from the World Health Organization. For daily country-level policy changes in countries, I use two indices from the Oxford Covid-19 Government Response Tracker. The Containment and Health Index tracks the level of governmental response in measures such as movement restriction, healthcare availability, and vaccinations. The Economic Support Index tracks the level of governmental response in financial and other economic support of the country's citizens and businesses.

I also obtained the four Fama-French-Carhart factors for Europe, North America, and Asia (excluding Japan) from Kenneth R. French's website, I discuss them in detail in a later chapter. Lastly, I obtained measures of societal trust in different European countries from the European Social Survey website. The European Social Survey regularly surveys citizens of European countries and collects a vast range of data on their views on society, politics, economic situation, and others.

3.1.1 Refinitiv's ESG rating methodology

Refinitiv uses a total of 186 comparable metrics (collecting over 630) to calculate their ESG scores. These 186 metrics are split into 10 categories, which themselves fall under either one of the 3 ESG pillars. Under the environmental pillar (68 metrics), the categories are resource use, emissions, and innovation. Under the social pillar (62 metrics), it is workforce, human rights, community, and product responsibility. And lastly, under the governance pillar (56 metrics), they have management, shareholders, and CSR strategy. From these metrics, Refinitiv calculates separate pillar scores, an overall ESG score, and an overall ESGC[1] score. The ESGC score is not used in this thesis.

All categories include a number of specific metrics, using which Refinitiv calculates category scores. Refinitiv calculates each company's percentile rank for that metric among its peers, which makes it easier for companies in different industries to be comparable. For categories under the environmental and social pillars, a company's peers are those in the same TRBC industry group[2]. For categories under the governance pillar, a company's peers are those with the same country of incorporation. The final category score is an average of the scores of all metrics.

I also feel it important to touch on how Refinitiv handles missing data. In general, there are two types of data, boolean and numerical. For boolean data, such as "Does this company have a water treatment policy?", should the given data be missing, Refinitiv assigns it a negative value (that might be either 0 or 1 depending on the metric). For numerical data, such as "How many

[1]ESG score with an "ESG controversies overlay", a way of discounting ESG scores based on negative press coverage, developed by Refinitv

[2]Refinitiv's proprietary industry classification, see https://www.refinitiv.com/en/financial-data/indices/trbc-business-classification

tons of CO2 has this company emitted over the past year?", should the data be missing, Refinitiv will not use that metric in the score calculation for that company. This is important as it could artificially inflate the ESG scores of certain companies if they choose not to report a negative result.

Each of the 10 categories is weighted. These weights are different for each of the 62 industry groups. The weights are based on the materiality (i.e. real impact) of each category in each industry group. Note that in each industry group, of the 186 metrics, only relevant/material ones are taken into account. In practice, that means only between 70 and 170 metrics are actually used for any industry group. The materiality of a given category is based on the comparison of a median score within a given industry group compared to median scores in all other industry groups. The better a given industry group ranks in a given category, the higher the weight for that category for that industry group. This might again skew ESG ratings of certain companies upwards as if a given metric is strongly negative relative to other industries, companies within that industry will benefit from that metric not being weighted as heavily.

When all the category scores and weights are calculated, both the separate pillar and full ESG scores are just weighted averages of the categories. This allows for a rather easy combination of the pillar scores into, for example, ES scores, which I make use of in this thesis.

3.1.2 Descriptive statistics

In this chapter, I will present the descriptive statistics of my datasets. I will start by showing the summary statistics for firm variables (Tables 3.1, 3.2, 3.3). I use six separate datasets, one for daily returns during a part of Q1 2020 and one for monthly returns between Q1 2019 and Q2 2021 (excl. Q1 2020) for three separate regions. I will then show the number of firms per industry (Table 3.4) and per country (Appendix A). I will not show a table of country-specific variables, as that would require showing over 20 tables and bring little benefit. However, should you want to see these tables, feel free to contact me at the e-mail address written in the frontmatter of this thesis. I will conclude by mentioning amounts of European firms which fall into either one of two categories (essential and non-essential industries; high and low societal trust) which I use when examining additional resiliency factors. Due to the simple nature of these statistics, I see little reason in putting them into a space-demanding table. Details of all variables are described in later chapters.

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ESG	3.340	38.160	54.400	53.260	69.450	94.970
Raw returns$_{day}$	-89.470	-3.800	-0.930	-1.305	1.567	90.160
Excess returns$_{day}$	-89.480	-3.810	-0.940	-1.315	1.558	90.150
Raw returns$_{mth}$	-154.290	-3.080	2.320	2.707	7.990	345.730
Excess returns$_{mth}$	-154.480	-3.160	2.230	2.618	7.900	345.610
Tobin's Q$_{2018}$	0.060	0.670	0.980	1.539	1.680	10.870
Tobin's Q$_{2019}$	0.040	0.620	0.950	1.475	1.630	10.200
Size$_{2018}$	4.672	7.049	8.092	8.217	9.278	12.654
Size$_{2019}$	4.898	10.369	14.383	13.462	16.137	20.445
Cash ratio$_{2018}$	0.000	0.035	0.072	0.101	0.125	0.952
Cash ratio$_{2019}$	0.000	0.030	0.068	0.096	0.122	0.952
Leverage$_{2018}$	0.000	13.090	26.390	26.980	37.870	73.320
Leverage$_{2019}$	0.000	12.550	25.890	26.670	37.790	72.910

Table 3.1: Descriptive statistics of firm variables from the European sample

The European sample, shown in Table 3.1, is the main sample of interest. ESG scores are centered roughly around 50 (as per theory). Returns are measured in percentage points, size is a log of Assets in USD. Daily returns are symmetrically spread around 0, though the mean return is considerably below 0. Monthly returns are more asymmetrically and positively spread. There is little difference between raw and excess returns, both daily and monthly. Accounting data is winsorized at the 1% and 99% levels. Only Size differs significantly across the two years, becoming considerably larger on average.

For the American sample in Table 3.2, compared to the European sample, ESG is skewed to the left considerably more, centering around 32-36. Companies are considerably larger on average in 2018 compared to the European sample. The American sample also has several overleveraged companies, while the European one has none. Other statistics seem to be broadly similar to the European sample.

As for the Chinese sample in Table 3.3, just like with the American one, ESG scores seem to be skewed to the left. Returns are centered around 0,

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ESG	1.140	21.510	32.530	36.650	48.680	93.170
Raw returns$_{day}$	-114.180	-3.090	-0.360	-1.109	1.330	98.720
Excess returns$_{day}$	-114.190	-3.100	-0.370	-1.119	1.320	98.710
Raw returns$_{mth}$	-231.070	-3.910	2.980	3.473	10.303	303.900
Excess returns$_{mth}$	-231.250	-4.000	2.880	3.381	10.210	303.690
Tobin's Q$_{2018}$	-0.062	0.801	1.222	1.886	2.197	10.977
Tobin's Q$_{2019}$	0.012	0.820	1.272	1.983	2.381	11.138
Size$_{2018}$	9.442	12.836	14.211	14.250	15.606	19.213
Size$_{2019}$	9.549	12.960	14.285	14.338	15.685	19.268
Cash ratio$_{2018}$	0.000	0.030	0.084	0.156	0.200	0.995
Cash ratio$_{2019}$	0.000	0.030	0.081	0.151	0.190	0.997
Leverage$_{2018}$	0.000	5.705	23.840	26.619	39.540	116.700
Leverage$_{2019}$	0.000	5.220	23.770	26.180	39.260	112.420

Table 3.2: Descriptive statistics of firm variables from the American sample

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ESG	0.660	22.090	33.310	34.560	46.570	76.000
Raw returns$_{day}$	-15.460	-1.150	0.090	0.082	1.390	15.770
Excess returns$_{day}$	-15.470	-1.160	0.080	0.072	1.380	15.760
Raw returns$_{mth}$	-84.56	-4.47	1.515	2.444	8.470	109.120
Excess returns$_{mth}$	-84.770	-4.550	1.420	2.353	8.383	109.110
Tobin's Q$_{2018}$	0.089	0.567	0.898	1.367	1.620	7.650
Tobin's Q$_{2019}$	0.0890	0.5670	0.9535	1.4983	1.7672	8.5890
Size$_{2018}$	13.840	16.700	17.620	17.690	18.810	21.380
Size$_{2019}$	14.07	16.74	17.67	17.75	18.83	21.43
Cash ratio$_{2018}$	0.000	0.081	0.134	0.158	0.202	0.832
Cash ratio$_{2019}$	0.000	0.081	0.137	0.162	0.209	0.832
Leverage$_{2018}$	0.000	9.210	23.860	25.490	39.110	68.960
Leverage$_{2019}$	0.000	9.643	23.845	25.669	38.960	71.070

Table 3.3: Descriptive statistics of firm variables from the Chinese sample

though they do not reach nearly as high (or low) amounts as in the Western samples. Companies in the Chinese sample also seem to be much larger than their Western counterparts. In terms of leverage, the Chinese sample is very similar to the European sample. Other characteristics are broadly the same across all three samples.

Industry	Europe	America	China
Basic Materials	110	7	40
Consumer Cyclicals	195	113	46
Consumer Non-Cyclicals	79	109	23
Energy	49	124	18
Financials	98	161	41
Healthcare	101	432	35
Industrials	235	314	73
Real Estate	39	21	18
Technology	137	347	57
Utilities	46	66	17
Others		2	2

Table 3.4: Number of firms per industry in all three regions

In Table 3.4, we see that in Europe, Industrial, Consumer Cyclical, and Technology firms are the most common. In America, Healthcare, Technology, and Industrial firms are the most common. The dominance of Healthcare firms is especially of note. In China, most firms come from the same three industries as firms in Europe.

A table with the number of firms per country is available in Appendix A. When making the European dataset, I started with all EU27 countries and the United Kingdom and then removed observations from all countries with fewer than 5 firms with a recorded ESG score in 2018. After building the full dataset, I removed all firms without data for all relevant variables. I used the same approach when building the datasets for the other two regions. Lastly, of the 1,089 European firms in the final European sample, 294 ($\approx$ 27%) belong to essential industries, and 374 ($\approx$ 34.3%) are from a country with high societal trust.

3.2 Methodology

In this chapter, I will describe in detail the methodological approach I use to answer the research question as well as state the hypotheses I will be testing.

3.2.1 Regression equations

In the main regression of this thesis, I will regress daily excess returns of stocks during a thirty-day window (15 days pre-crisis, 15 days post-crisis) sometime in Q1 2020 on firm ESG performance as well as a variety of company, country, industry, and day-specific controls. In line with Cardillo *et al.* (2020), I will be running the main regression set using a difference-in-differences estimator, as it allows for comparatively easy identification of the causal effect of ESG during the covid-19 crisis, if its assumptions are fulfilled. While its assumptions are quite limiting, in a later chapter, I show that the model I specify fulfills these assumptions to a sufficient extent. I will further run two more regressions to test augmentative resiliency factors and several more sets of regressions to either support or contradict my findings in a robustness check. I talk about these in more detail further down in this chapter and in later chapters. The fully specified base regression can be found in Equation 3.1 below.

$$\begin{aligned} excess\_returns_{i,t} &= high\_ESG_i + post\_covid_t + high\_ESG_i * post\_covid_t + \\ lnc19c_{i,t} &+ ESI_{i,t} + CHI_{i,t} + tobins\_q_i + size_i + cash\_ratio_i + leverage_i + \\ & MOM_t + MKTRF_t + SMB_t + HML_t + \text{Industry dummies} + \epsilon_{i,t} \end{aligned} \tag{3.1}$$

To compute $excess\_returns_{i,t}$, I subtracted the market risk-free rate at time t from a stock i's raw return at time t, in line with the Fama-French-Carhart four factor model. I computed the raw returns as shown in Equation 3.2.

$$raw\_returns_{i,t} = \ln\left(\frac{P_{i,t} + D_{i,t}}{P_{i,t-1}}\right) \tag{3.2}$$

where $P_{i,t}$ is a stock i's price at time t and $D_{i,t}$ is stock i's dividend payout for time t. I computed dividend payouts with regard to the stock's ex-dividend date (as opposed to the actual payout date). I use the logarithmic return form as logarithmic returns are symmetrical for gains and losses. Imagine a given stock's price goes from 50 to 100 on day 1 and then from 100 back to 50 on day 2. Non-logarithmic returns would compute this as +100% on the first day and -50% on the second. For logarithmic returns, these values would be symmetrical around 0. Logarithmic returns also allow for additive computation of a compound return, meaning summing 3 daily log returns is equal to the three-day log return. These values are all multiplied by 100 for easier coefficient reading.

As for the difference-in-differences part of the regression, $high\_ESG$ is a dummy variable equal to 1 if a given firm's ESG score for the year 2018 was above the median for the final sample (that is, the Refinitiv Eikon ESG population per a given region). This value represents the treatment group within the difference-in-differences model. I chose the value for 2018 as that is the latest information generally available to investors at the beginning of the pandemic since the ratings for 2019 for the vast majority of firms had not yet come out. Next, $post\_covid$ is a dummy variable (start of treatment in difference-in-differences) equal to one from 11th of March onwards for the European sample, as that is the day World Health Organization declared covid-19 has the status of a pandemic[3], which also coincided with a general stock market drop. Lastly, $high\_ESG * post\_covid$ is an interaction term between the two aforementioned dummy variables, that is the treatment variable. In terms of intuition, the treatment is "having an above-median ESG score in the time of crisis". This also makes sense as several previous papers found that ESG performance does not significantly impact returns under normal circumstances (i.e. not during a crisis) (Broadstock *et al.* 2021; Hsu *et al.* 2021; Humphrey *et al.* 2012), suggesting that having high-ESG only becomes a differentiating factor for stock returns during a crisis.

The next 3 variables are country-level controls. $lnc19c_{i,t}$ is the natural logarithm of new covid cases per 100,000 inhabitants lagged one day (to account for delayed reporting) plus 1 in the country of company i at time t. Similarly, $ESI_{i,t}$ and $CHI_{i,t}$ are values of the Oxford Covid-19 Government tracker indices which track governmental response to the pandemic in terms of economic support and containment measures, respectively. Using these, I try to capture the local effects of the disease spread on life and business of different countries. For company-specific controls, I chose $tobins\_q$ to capture under-or-overvaluation, $size$ (computed as the natural log of total assets in millions USD plus 1 as in e.g. Albuquerque *et al.* (2019)) to capture company size (it could be easier to spend money on ESG related activities for bigger firms, as many are semi-fixed costs), $cash\_ratio$ (calculated as cash over total assets) to capture liquidity

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

(the economic crisis might require companies to have liquid assets available to fund a possible major decrease in cash inflows), and *leverage* (debt ratio) to capture indebtedness.

Lastly, I make use of the Fama-French-Carhart four-factor model with values for the European (the North American and Asia ex. Japan dataset are used for the other two regions) market ($MOM_t, MKTRF_t, SMB_t, HML_t$) to capture systematic risk. Fama & French (1992) show that the three Fama-French factors can explain most daily equity market price movements in diversified portfolios. Carhart (1997) adds a momentum factor to the three factors as he shows evidence that it also plays an important role in equity market asset pricing and this four-factor model performs better than the original three-factor alternative. This allows me to better separate the effect that ESG has on the excess returns from general market sentiment. I talk more about the Fama-French-Carhart four-factor model in a later chapter.

3.2.2 Additional factors of resiliency

It is possible that the effect of high ESG during a crisis could be even stronger under certain circumstances. I will therefore test two characteristics of environments the firms operate in that might increase the resiliency of high ESG firms during a crisis. Specifically, given the findings of Lins *et al.* (2017), I predict that the effect is going to be stronger in countries with larger amounts of trust among their citizens. My intuition here is that firms in such countries face somewhat higher expectations of ESG activities. As such, firms with lower ESG scores have an additional negative characteristic in the investors' eyes, as consumer loyalty might be lower and the effect of the crisis therefore stronger. It is important to note that Engelhardt *et al.* (2021) found the exact opposite effect. This ambiguity in existing research is interesting, and I believe it warrants further consideration. I will be using the data on societal trust from the European Social Survey. In line with Albuquerque *et al.* (2020), I will be testing this using a triple interaction term $post\_covid * high\_ESG * high\_ppltrst$, where $high\_ppltrst$ is a dummy for countries with an above average level of trust. I believe the average is a better cutoff than the median as this data does not suffer from having extreme outliers.

I also predict that the effect will be stronger for firms in "essential" industries. My intuition here is that governments presumably primarily supported essential businesses during the crisis. Then, if we consider high ESG performance to be a systemic risk-mitigating factor (Albuquerque *et al.* 2019; El Ghoul *et al.* 2016), and we assume being an essential business is also a risk-mitigating factor (given the government support), it follows that high-ESG essential firms face an even lower risk than firms with just one of either characteristic alone. Since prices generally fell during the crisis, it is possible that investors flocked to less risky stocks as the risk premiums required of riskier investments disappeared.

3.2.3 Robustness checking

I will test the robustness of my results in several ways. First, I will construct two other versions of the dataset, one for American and one for Chinese companies. Furthermore, each of these will use a different event window centered around a somewhat different exogenous shock. The methodology of these datasets will be identical to that of the European dataset. I will then test the robustness of my original results using data for all three regions, separately. Finally, for the European sample, I will explore possible sources of endogeneity in other events during Q1 2020, that could have influenced high- and low-ESG firms differently, which would bias the results.

I will build the same main difference-in-differences regression for both the United States and China and report the results. Important to note, in current literature on the topic, researchers examining different regions generally use different dates to indicate the post-covid period. I will therefore follow their example, which will add additional robustness to my results in answering the original research question, as I will test multiple different exogenous shocks. In line with Albuquerque *et al.* (2020), I will center my event window for American companies around February 24th. As Albuquerque *et al.* (2020) say, February 24th is the start of the so-called "fever" period in Ramelli & Wagner (2020), a date when American markets begin to strongly react to the unfolding pandemic situation. As for Chinese companies, in line with Broadstock *et al.* (2021), I will center my event window around February 3rd, which is the first time markets opened after the first lockdown in Wuhan, China. It is important to note that Chinese markets were not open between January 24th and 31st, as the country was celebrating the Lunar New Year holidays. Observations from these days were therefore removed.

For all three datasets, I will perform a cross-sectional regression for the entire thirty-day period. While this will provide additional backing to my results, it is important to mention that the cross-sectional regression will make it less as to what the direct effect of high ESG during covid is, as the entire event window, both pre- and post-covid, is considered at once.

It is possible that the markets will act differently immediately following the exogenous shock. To better capture this effect, I will run four additional regressions for each dataset, two panel and two cross-sectional. For each kind, I will test two narrower event windows, a four-day and a ten-day window around the treatment.

To further test the robustness of my results, I will perform a difference-in-differences regression for monthly returns during 2019 for the pre-covid period and between Q2 2020 and Q1 2021 for the post-covid period. While this regression will not directly examine the immediate effects of high ESG on stock returns when the pandemic "happens", it will provide useful insight into how long-lived these effects are.

Lastly, as is more closely described in the following chapter, I will run the original regression with modified samples of European companies to check for other events around the event period, which might influence the regression to a point where the parallel trends assumptions would be violated.

3.2.4 Difference-in-Differences estimator

As noted above, I decided to use the difference-in-differences estimator as the estimator for my main regression equation set. The difference-in-differences estimator, under the right settings given its assumptions, allows for a fairly simple yet credible examination of causality when only quasi-experimental data are available (Wooldridge 2012).

For one to be able to make causal inference from the regression generated by the difference-in-differences estimator using an OLS model, all multiple linear regression assumptions have to be fulfilled (Wooldridge 2012). An exception to this might be the normality assumption, I talk about this in more detail in a later chapter, where I will also be testing the fulfillment of all the assumptions. However, for difference-in-differences estimation, there is a special consideration within the no endogeneity assumption which I will talk about in the next few paragraphs.

A core assumption of the difference-in-differences estimator is the parallel trends assumption. It states that had the treatment not come into effect, the difference (in this case difference in stock returns) λ between the treatment group (in this case high-ESG firms) and the control group (in this case low-ESG firms) would have remained the same. While in a quasi-experimental setting such as this one, there is no perfectly accurate way to test this, one can find clues that suggest this assumption is fulfilled.

First, from a theoretical point of view, there is no reason to believe there would be a difference in stock returns between high- and low-ESG firms during normal times. As noted above, previous research has generally shown there is no significant difference between stock returns of high- and low-ESG firms during normal times, ceteris paribus (Broadstock *et al.* 2021; Hsu *et al.* 2021; Humphrey *et al.* 2012). Given this evidence, it is reasonable to assume that had the pandemic not happened, stock returns would have remained the same on average for both groups.

I can also look at the data I have available to test the assumption empirically. In Figure 3.1 below, you can find the average excess returns of firms in both the treatment and control groups both before and after treatment. The vertical dashed line signifies the treatment border. From a visual inspection of the data, it is apparent that the difference between the two groups is small in the pre-treatment period and becomes much larger on average during the post-treatment period. The data for China also suggests the prices difference rebounds extremely fast, compared to the other two samples, which will become important later. Empirical evidence therefore also suggests the parallel trends assumption is fulfilled.

One more possible source of endogeneity from a violation of the parallel trends assumption in the difference-in-differences estimator is another major event around the treatment period that would affect the stock returns of both

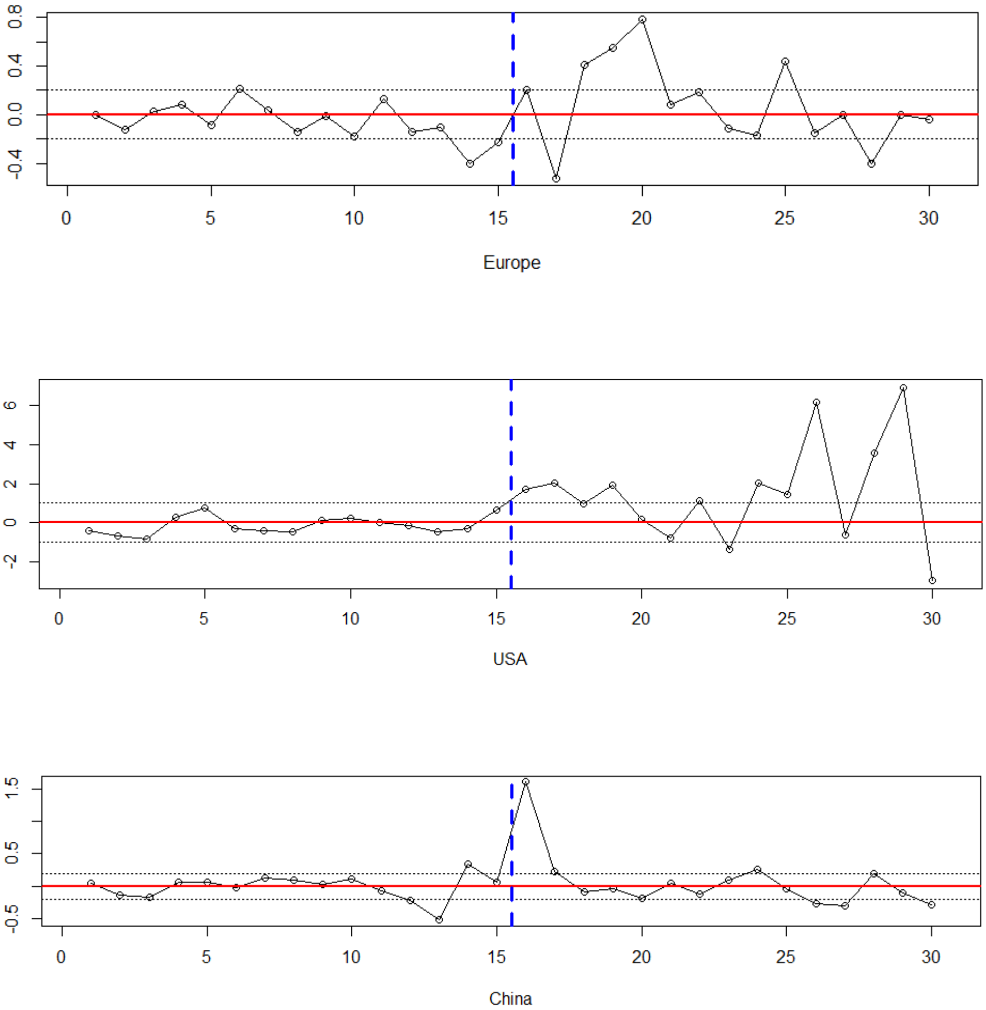

Figure 3.1: Difference in mean excess returns of high- and low-ESG firms within the event window

groups differently. Should that happen, one would not be able to make causal inferences from the results of the regression because of omitted variable bias, since the unobserved event would be correlated with the treatment variable and also affect the stock returns. Two such events of note that are not directly connected to the pandemic are the oil price crash in March 2020 and developments in the Brexit process. The possible effect of both of these events will be examined in the robustness chapter.

3.2.5 Fama-French-Carhart four-factor model

The Fama-French-Carhart four-factor model is a model used in investing research to control for the systematic risk of a given investment portfolio. As of now, the original three-factor model developed by Fama & French (1992) or one of its variations (four, five, or six-factor models) seems to be one of the most commonly used methods to control for systematic risk (Cardillo *et al.* 2020; Demers *et al.* 2021).

The model is an extension of the CAPM model, which only controls for systematic risk by considering the market risk-free rate and the market risk premium. Fama & French (1992) consider two more variables in the model, SMB and HML, which control for size and value respectively, as they noticed significant differences in returns of portfolios with differing size and value. Additionally, Carhart (1997) considered an additional "factor", momentum, when he was researching the role of skill in mutual funds performance. He found that the momentum of a given portfolio had significant explanatory power in predicting the mutual fund's performance in the short-to-medium term. The entire model then looks like the following.

$$E(r_p) = \alpha + r_{rf} + \beta * E(r_m - r_{rf}) + b_S * SMB + b_H * HML + b_M * MOM + \epsilon \quad (3.3)$$

where $E(R_p)$ is the expected return of the portfolio, r_{rf} is the risk-free rate, $E(r_m - r_{rf})$ is the market risk premium, SMB, HML are the two additional factors considered by Fama & French (1992), and MOM is the momentum factor added by Carhart (1997). As we can, as opposed to the CAPM, which only generates one β, the FF(C) three(four)-factor model generates 3(4) βs, which allows for a more complex systematic risk control.

I also feel it noteworthy to explain the three additional factors considered by Fama & French (1992) and Carhart (1997). Fama & French (1992) noticed that there is a difference in returns between companies with small and big market capitalizations and between companies with small and large book-to-market ratios. They therefore split stocks into portfolios and put their returns into a matrix sorted based on these two characteristics, that is market capitalization (size) and book-to-market ratio (value).

To calculate the factors, the portfolios are then further categorized into size categories big and small, based on being below or above the median market cap, and into value categories value, neutral, and growth, where "value" are the portfolios in top 30% of book-to-market ratios, "growth" in the bottom 30% and those in between are "neutral".

As previously noted, SMB is a variable that controls for systematic risk posed by size. Being an abbreviation for "Small Minus Big", SMB captures the difference in returns between portfolios with small market caps and large market caps for a given period. Mathematically, the calculation of the factor looks like so.

$$\begin{aligned} SMB = &\frac{1}{3}\left(\text{Small Value} + \text{Small Neutral} + \text{Small Growth}\right) \\ &- \frac{1}{3}\left(\text{Big Value} + \text{Big Neutral} + \text{Big Growth}\right) \end{aligned} \tag{3.4}$$

Similarly, HML (High Minus Low), controlling for systematic risk posed by the book-to-market ratio, is also calculated as a difference, specifically that between value and growth stocks. Mathematically, it is

$$HML = \frac{1}{2}(\text{Small Value} + \text{Big Value}) - \frac{1}{2}(\text{Small Growth} + \text{Big Growth}) \tag{3.5}$$

Lastly, Carhart (1997) introduced the momentum factor MOM as an addition to the three original factors. Similarly to Fama & French (1992), Carhart (1997) ranks stocks based on their previous (11 months') performance. He takes the stocks with returns in the top 30% and calls this group winners. He then takes the bottom 30% of stocks in terms of returns and calls this group losers. Momentum is then equal to the difference between the returns of winners and losers, mathematically

$$MOM = \text{Winners - Losers} \tag{3.6}$$

3.2.6 Alternative approaches

While I ultimately decided to use difference-in-differences estimation for this thesis, other approaches to examining causality exist. I will shortly summarise them and discuss why I did not find them suitable for use in this thesis in the next two paragraphs. Other, more advanced methods also exist, but due to my level of econometric knowledge and experience, I would not feel comfortable using these in an academic setting.

A natural extension of the difference-in-differences estimator is the difference-in-difference-in-differences estimator. Compared to the original difference-in-differences estimator, it works by using an additional control group in the regression analysis which leads to more robust results under its assumptions (Wooldridge 2007). However, I believe it is impossible to make use of this estimator in the case of researching the effects of the pandemic. For one to be able to use this estimator, one has to find a comparable group to the treatment group (in my case high-ESG firms, ideally in a Europe-like economic environment) not affected by the treatment effect (in my case the covid-19 pandemic). Given the, by definition, global scale of the pandemic, there were few countries unaffected by the pandemic by the end of March 2020. If I were to consider companies in non-European OECD countries to be generally comparable to European companies, I find that they were all affected by the pandemic by the end of March 2020 (World Health Organization 2020). Difference-in-difference-in-differences estimator is therefore not suitable for this use case.

Instrumental variable estimation can be said to offer a more robust approach to examining causality (Wooldridge 2012) compared to DID. In short, it works by finding a variable that only affects the dependent variable (in this case stock returns) through its effect on the variable of interest (in this case ESG). However, finding a good instrument can be very difficult and in some cases almost impossible, which I believe is the case here. A good instrument in my case would have to be something that influences ESG and not any of the control variables. The only possibly valid and attainable instrument I can think of for this case is a country-level sustainability-related policy. In theory, such a policy would raise ESG scores of all companies in a given country, regardless of their other characteristics, as they would be required to engage in ESG activities. Finding such policies is possible, and indices tracking the level of sustainability-related policies across countries also exist. However, when testing the idea, I found that these had little to no predictive power for ESG scores of companies. As I was not able to find a working instrument, I believe this method is also not suitable for this use case.

3.2.7 Hypotheses

Finally, at the tail end of this chapter, I will formally present three different hypotheses which I will be testing in this thesis.

Hypothesis 1: High ESG scores serve as a resiliency factor for stock returns during crisis.

$$H_0 : \beta_{high\_ESG*post\_covid} = 0,\ H_A : \beta_{high\_ESG*post\_covid} > 0 \tag{3.7}$$

I predict that during time of crisis, high-ESG scores will serve as a resiliency factor for company stock returns. The literature strongly suggests ESG performance serves as an indicator of lower risk of a given company for investors (Albuquerque *et al.* 2019; Hong & Kacperczyk 2009). Since investors require higher returns from riskier stocks, then during a time of market downfall, it would follow that investors would tend to buy less risky stocks, as the riskier securities would not offer the required risk premium. Therefore, following the beginning of the crisis, there should be a statistically significant difference in stock returns between low- and high-ESG companies, ceteris paribus.

Hypothesis 2: The effect in Hypothesis 1 is stronger in "essential" industries.

$$H_0 : \beta_{high\_ESG*post\_covid*ess} = 0,\ H_A : \beta_{high\_ESG*post\_covid*ess} > 0 \tag{3.8}$$

I predict that the effect described in Hypothesis 1 will be stronger for firms in industries that were considered essential during the covid-19 pandemic. Following the same line of thinking as in Hypothesis 1, given the "essentiality" of these industries, the companies within them pose lower risk for investors, which should make these stocks more appealing.

Hypothesis 3: The effect in Hypothesis 1 is stronger in countries with higher levels of societal trust.

$$H_0 : \beta_{high\_ESG*post\_covid*high\_ppltrst} = 0,\ H_A : \beta_{high\_ESG*post\_covid*high\_ppltrst} > 0 \tag{3.9}$$

I predict that the effect described in Hypothesis 1 will be stronger for firms in countries with higher levels of societal trust. This is strongly based on the finding of Lins *et al.* (2017), who found this exact effect during the great financial crisis for American firms. Assuming that trust rises in importance during crises, it would follow that more trustworthy firms would be preferred by members of a more trusting society. Furthermore, given the composition of the S pillar includes the "Employees" category, citizens of high-trust countries might have higher expectations of their employers and might therefore not be as inclined to stay loyal to the company during the crisis.

Chapter 4

Empirical results

In this chapter, I will talk in detail about the results of my main regressions, discuss their implications, and test them using a series of robustness tests. Lastly, I will touch on the limitations of this thesis and suggestions for future research.

Before delving into the results themselves, I feel it beneficial to talk about the expected results given the current literature. First, as per the literature review on ESG as a risk-mitigation factor, I expect the effect of ESG to be significant and positive for all three regions for the post-crisis period. This line of thinking extends to the augmentative effect of essential industries, whose effect I also expect to be positive and significant. As for the effect of societal trust, the question is largely whether the positive significant effect found by Lins *et al.* (2017) was driven by the fact that the GFC was also a crisis of trust in institutions. If it was, and assuming the pandemic was not (at least in the beginning), the effect should be insignificant, but the results largely depend on the correctness and fulfillment of the assumptions. Lastly, in the long-term regression, the literature suggests the effect should be insignificant. The question here is whether the effect will still be insignificant in the long term even when the economy is in crisis during that long term, which is the case for the pandemic (at least for Europe and America).

4.1 Main regression equation

The results of the main difference-in-differences regression for all three examined regions (Europe, USA, China) can be found in Table 4.1. The dependent variable is $excess\_returns$ and it is measured in percentage point units (100% = 100 units). The variable standard errors are clustered by firm and robust to heteroskedasticity. I removed the ESI from both the American and the Chinese regressions as it was equal to 0 during the entire examination period and CHI from the Chinese regression because of strong collinearity with $post\_covid$.

In both Europe and America, the variable of interest (treatment) is positive and statistically significant. In Europe, its coefficient value is approximately 0.271 and this value is significant at the 1% significance level. This suggests that excess log returns of high-ESG firms in Europe were (15*0.271) 4.065 p.p. higher compared to those of non-ESG firms over the post-crisis period. Furthermore, the fulfillment of the DID estimator assumptions (see later chapter) also suggests this relationship is causal. It is important to note, however, that the coefficient on $high\_ESG$ is also significant at the 1% level and it is negative, approximately equal to -0.196. This means that the increase in daily excess log returns of European high-ESG firms post-covid is only around (0.271-0.196=) 0.075 p.p., which is equal to 1.125% over the post-crisis period. This effect is still positive, but it is considerably (almost 4 times) smaller than the effect suggested by the coefficient on the interaction term.

In America, the coefficient on the interaction term is considerably higher than that of the European one, being equal to 0.451 and significant even at the 0.1% significance level. The coefficient of $high\_ESG$ is also negative and statistically significant, but the net effect is much higher at 0.319, which equals to 4.785% in excess log returns over the post-crisis period. Again, fulfilment of the DID assumptions suggests that there's a causal effect. An interesting aspect to note is that the median ESG score of the American sample is just 32.18 (instead of the theoretical 50). When running the regression again with the high-ESG cutoff being 50 instead of the sample median, I find that the coefficient on the interaction term rises all the way to 0.563 (0.430 net of the coefficient on $high\_ESG$) and remains significant at the 0.1% level (other results remain qualitatively unchanged).

The Chinese sample is the outlier here, with the $high\_ESG$ coefficient being insignificant and the coefficient on the treatment being considerably smaller at 0.161 (2.415% over the post-crisis period), though it is still significant at the 5% level and the net effect is actually larger than that in the European sample.

Table 4.1: Results of the main regressions

	(1) Europe		(2) USA		(3) China	
Variable	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	-0.787***	0.175	0.426	0.223	1.881***	0.296
high_ESG	-0.196**	0.064	-0.132**	0.044	0.012	0.058
post_covid	-0.990***	0.102	-0.186**	0.066	0.073	0.063
high_ESG*post_covid	0.271**	0.086	0.451***	0.161	0.161*	0.070
ln(covid-19 cases)	0.141**	0.043	-5.584***	1.418	0.456*	0.221
ESI	0.004**	0.001				
CHI	0.011***	0.002	0.002	0.006		
Tobin's Q	0.111***	0.015	0.097***	0.012	0.048*	0.024
Size	0.086***	0.019	0.050***	0.013	-0.081***	0.019
Cash ratio	-0.252	0.244	-0.047	0.128	0.118	0.201
Leverage	-0.008***	0.002	-0.003***	0.001	0.002	0.001
MOM	-0.155**	0.054	0.229***	0.049	-1.029***	0.107
MKTRF	1.162***	0.017	1.069***	0.012	2.582***	0.069
SMB	0.563***	0.045	0.941***	0.048	3.102***	0.135
HML	-0.037	0.059	0.118**	0.043	1.400***	0.124
Industry FE	Yes		Yes		Yes	
Adjusted R^2	0.416		0.415		0.224	
N	32,670		61,440		11,100	
No. of firms	1,089		2,048		370	

Notes: In the table, there are results of three separate regressions based on my main regression specification, one for each of the examined regions (EU, USA, China). All specifications use a sample with daily excess log returns for a 30-day window within Q1 2020. Accounting data is from 2018. All reported standard errors are clustered by Firm and robust to heteroskedasticity. ESI is not estimated for USA and China as it is equal to 0 during the entire event window. CHI is not estimated for China due to high collinearity with *post_covid*.

Once again, the fulfillment of the DID assumptions suggests the relationship is causal. The results of the main regression therefore strongly support my initial hypothesis.

Quite interesting, in my view, is how different the coefficients both on the intercept and the four factors look compared to the other two samples. While similar in terms of significance, their absolute values are much larger. I theorize that this vastly different behavior might somehow be a result of the interventions and the level of power the Chinese Communist Party has and exercises over the Chinese stock markets, as I see that as the core difference between how the Chinese and Western stock markets operate. Carpenter *et al.* (2021) found that while prices in Chinese stock markets have the same amount of predicting power for privately-run companies, the same is not the case for state-owned enterprises, whose price's predicting power is significantly lower, partially as a result of unpredictable subsidies and interventions. Furthermore, Ni *et al.* (2015) conclude that investor sentiment has a strong influence on stock prices in the short- to long-term (up to 2 years) and that Chinese investors have considerable cognitive bias and speculation tendency, which therefore makes stock prices less accurate. It could be that this influence of sentiment is what drives these market and other betas as well as the alpha to such high values. Examining these findings properly would nevertheless require much deeper research and is beyond the scope of this thesis.

4.1.1 Separate effects of the E, S, and G pillars

In this section, I will examine if and how the three pillars (Environmental, Social, and Corporate Governance) differ in how they affect stock return resiliency during a crisis. For the sake of brevity and readability, I will only show the coefficients for the three pillars in a single table (I run a regression for each pillar in each region separately), showing the values for both the *high_pillar* dummy as well as the treatment term (its interaction with *post_covid*). You can find the results in Table 4.2 below.

E(S, G)**post_covid* is the treatment for a given pillar. We see that both the effect in Europe and America is for the most part carried by the social pillar, being significant at 0.1% level in both cases, and equal to 0.320 and 0.429 respectively. For both regions, however, all three treatments have significant

	(1) EU		(2) USA		(3) China	
	Estimate	SE	Estimate	SE	Estimate	SE
E	-0.102	0.064	0.047	0.045	0.118*	0.055
E*post_covid	0.264**	0.086	0.182*	0.086	0.019	0.071
S	-0.063	0.063	-0.166***	0.045	0.078	0.056
S*post_covid	0.320***	0.086	0.429***	0.085	-0.006	0.071
G	-0.241***	0.059	-0.084*	0.040	0.005	0.054
G*post_covid	0.207*	0.086	0.242**	0.085	0.091	0.071

Table 4.2: Results of the main regression with separate pillars

Notes: In the table, there are results of nine separate regressions based on my main regression specification, one for each of the examined regions (EU, USA, China), and one for each separate pillar (E, S, G). Only the coefficients of interest are reported. All reported standard errors are clustered by Firm and robust to heteroskedasticity.

positive coefficients. Furthermore, the coefficient on the high S and high G dummies are also significant and negative, though the net effect is still positive.

The biggest surprise, however, is the 0.1% significant negative coefficient on the *high_G* dummy for the European sample. It is a surprise because it makes even the net effect post-crisis negative. This coincides with the surprising negative coefficient on $high_ESG$ in the first regression (similar can be said for the G dummy for the American sample, though the net effect is still positive). It would seem that basically the entire negative effect of high ESG during normal times is carried by high G scores in the European sample. While the governance pillar is an equal part of ESG ratings, it is qualitatively different from the other two pillars and some researchers have therefore decided to omit it when examining the effects of corporate social responsibility (Albuquerque *et al.* 2020), as it does not directly influence the larger society, but rather the way a given company is run. It could be that some companies try to increase their G scores in order to raise their entire ESG scores and appear more environmentally or socially friendly ("greenwashing"). Indeed, this breakdown does suggest that the composition of high-G companies is somewhat different from that of high-E or S companies, at least in Europe. In other words, companies with high-G scores do not necessarily have high E or S scores. Examining the data, I find that 134 companies from the European sample only have high G scores (and low E and S scores), compared to 48 and 50 for high E and S scores, respectively. Similarly, a combination of high G and E (S) scores and of a low S (E) score is considerably less common than high ES and low G (58 and 60 firms compared to 144). This strongly supports the claim that companies with high G scores are considerably different to those with high E and/or S scores, and also supports the claim that increasing own G score and not the other pillars could be a form of greenwashing and/or is not seen as a reliable signal by investors, at least during normal times.

The coefficients on Chinese treatments are all insignificant. A small surprise is the positive coefficient on E, significant at the 5% level. A part of Refinitiv's E scoring is the investment into innovation. China is a leader in renewable energy and electric mobility development and the industry has been on the rise for some time (Evans 2022), it could therefore be that stocks of these companies had been seen as more futureproof by investors, regardless of the crisis. Indeed, removing firms from the technology sector from the sample does make the coefficient on high E insignificant.

4.2 Assumption testing

In this chapter, I will test and provide commentary for all MLR assumptions for the Ordinary Least Squares model.

Linearity in parameters: The first assumption is fulfilled by how the regression equation is set up, that is, the dependent variable is calculated as a linear combination of all the independent variables.

Random sampling: Random sampling is difficult to test. The most important concern for me is the fact that I can only use companies in my regression for which I have access to all the data I use. While not an issue for accounting data, as I only needed to remove relatively few observations for missing data, I am severely limited by the availability of ESG scores. Even though no concrete data is available, I do believe Refinitiv, as one of the major ESG data providers, tracks the ESG data of the majority of companies who do report such data. A further concern may be the fact that the population of ESG-reporting companies itself is not representative. I believe this should not be a severe issue, however, as the thesis itself is built upon examining the effect of ESG activities. If a company does not do ESG reporting, one can assume that that company does not perform much ESG activity to begin with (as it can be very costly and if a company were to incur such costs, it would likely want to present its efforts through an ESG score) and their omission in the data should not skew the results. Moreover, large(r) companies in the European Union are required to report at least some ESG data through the Non-Financial Reporting Directive (Council of European Union 2014), and both the US and China have similar guidelines in place to at least encourage ESG reporting. However, it remains true that the approach of Refinitiv itself to ESG score calculation leaves room for foul play, as I discussed in the chapter about Refinitiv's ESG methodology. For any remaining concerns, I did include several company and country controls in my regression.

No endogeneity: Endogeneity is very difficult, if not impossible to test accurately. There are three main ways endogeneity could present itself. First, my model could suffer from omitted variable bias. I believe I controlled for all relevant aspects I could, including firm, day, and country controls as well as industry fixed effects. Second, there could be a simultaneous relationship between some of my independent variables and the dependent variable. I believe this is highly unlikely, as many of my independent variables are historical accounting data (and very little to no new accounting data appeared during the event period) and any other variables are extremely unlikely to be influenced by company stock returns. Third, a major source of endogeneity could be measurement error. Given I collected data from trustworthy sources and was making routine checks while building my dataset, I believe this also is not a concern. In the chapter about difference-in-differences estimator, I wrote in more detail about some aspects of endogeneity (incl. the parallel trends assumption) and why I believe endogeneity is not a severe concern in this thesis, especially after robustness testing.

As a simple empirical test, I plotted the residuals of my regressions against the row IDs. If there is no correlation between the independent variables and the error term, there should be no visible patterns or irregularities (firms are sorted alphabetically). In Figure 4.1, you can find the plot for the main European sample and as per my theoretical breakdown, there do not seem to be any patterns or major irregularities. The plots for the other two samples are qualitatively the same.

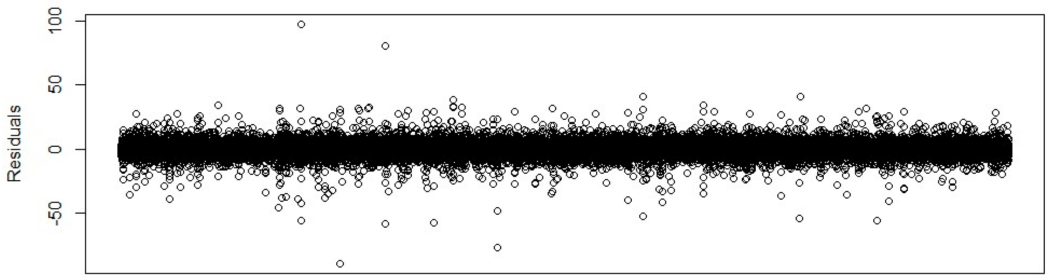

Figure 4.1: Plot of main regression residuals on row IDs for the sample of European firms

No perfect collinearity: Collinearity can be tested in R using the `vif()` command. A rule of thumb is that VIF scores above 5 start to raise suspicions, scores above 10 are considerable concerns. However, since multicollinearity does not affect other variables, it is not a considerable issue unless one of the variables of interest is highly collinear with the other. See the VIF scores with values above 4 (for brevity reasons) for all three samples (no variable for the Chinese sample has a VIF score larger than 4) in Table 4.3.

	EU	USA	China
lnc19c		5.025	
CHI		4.903	
MKTRF	4.602		
HML	4.493		

Table 4.3: VIF scores of variables for all three regions

We can see that for USA and Europe, two variables have scores around 5. For Europe, there are two of the Fama-French factors. As I am not interested in their coefficients and their scores do not even cross the 5 mark, I believe leaving them in as-is is not an issue. For America, the two variables are $lnc19c$ and CHI, that is the natural log of covid-19 cases per 100,000 people and the index of Containment and Health measures, respectively. Since a considerable correlation between these two variables is expected, the VIF score is not very large and I am not particularly interested in their coefficients, I believe they do not pose a problem that would violate the assumption.

No autocorrelation: Serial correlation can be tested using the Durbin-Watson test. A value close to 2 means little to no serial correlation, and a value near 0 or 4 means close to perfect autocorrelation. If we go by the random walk hypothesis for stocks, the value should indeed be near 2. The Durbin-Watson reports values as per this hypothesis, all are close to 2. See Table 4.4 for the results.

	D-W test result	p-value
European Union	1.923	6.376e-13
USA	1.949	8.334e-11
China	1.795	2.2e-16

Table 4.4: Results of the D-W test for autocorrelation for all three regions

Homoskedasticty: Heteroskedasticity is a serious concern, for which one can test using the Breusch-Pagan test. It is reasonable to assume that a wide portfolio of differing terms will suffer from heteroskedasticity when examining stock returns, and indeed, the Breusch-Pagan test does confirm this suspicion, as all three models suffer from heteroskedasticity. I correct for it by using firm-clustered robust standard errors (industry-clustered for cross-sectional regressions), which all regression results in this thesis include.

Normality: Drawing a QQplot of my regression residuals, I find that they are not normally distributed. Schmidt & Finan (2018) conclude that the normality assumption does not significantly impact results when the number of observations per variable is above ten. This is also a note presented in Wooldridge (2012), and a mantra in popular publications about econometrics. Since my sample sizes range from roughly several thousands to several tens of thousands of observations, I believe that a violation of this assumption is not an issue.

4.3 Further resiliency factors

In this chapter, I will further explore two possible factors which could augment the resiliency of high-ESG firms during a crisis even further. Specifically, I will examine factors described in Hypotheses 2 and 3, that is belonging to an essential industry and country-level societal trust. I will perform these analyses only for the European sample.

4.3.1 ESG as a resiliency factor in essential industries

In previous chapters, I describe the intuition behind why I believe the effect of high ESG as a resiliency factor during a crisis might be stronger for firms in industries that are considered essential. In short, being in an essential industry might be a risk-mitigating factor, as might having a high ESG score. Since market prices generally fell during the initial post-crisis period, riskier investments might not provide the required higher return premiums. As such, investors, wanting to protect their money, might prefer less risky investments. The regression equation is an extension of the main Equation 3.1 and can be seen in Equation 4.1.

$$er_{i,t} = \beta_0 + \beta_7 * high\_ESG * essential * post\_covid + \beta * \mathbf{X_{i,t}} + \epsilon_{i,t} \quad (4.1)$$

where $er_{i,t}$ are daily excess returns, *essential* is a dummy indicating essential industries, and $\mathbf{X_{i,t}}$ is a vector of all other dummy and control variables, including all combinations for the interaction term as well as all controls from Equation 3.1. Industries that I considered essential are Healthcare, Energy, Utilities, and Financials. All these industries (or sectors) can be considered a fundament without which modern society would not be able to function, and which cannot function "on its own" (such as Technology). In other words, these industries continuously provide core, essential services. See the results of the regression in Table 4.5. I only show the interaction term and its constituents, all other results are practically identical to the results in Table 4.1 for the EU sample. All standard errors are clustered by firm and robust to heteroskedasticity.

The results indicate rejection of the null hypothesis at the 5% level, that is, the coefficient on the triple interaction term is positive and significant, being equal to 0.607. That means high-ESG essential firms enjoyed over 9.105% higher excess log returns over the post-crisis period compared to non-essential and/or low-ESG firms. Indeed, the coefficient on the triple term is more than double the value of the one on the original treatment from Equation 3.1 (see Table 4.1). This, along with the insignificant coefficient on $high\_ESG * post\_covid$ suggests that the significant positive effect of the treatment found in Table 4.1 was (at least almost) entirely driven by firms in essential industries. What's more, the insignificant coefficient on $essential * post\_covid$ suggests

	Estimate	Std. Error
high_ESG	-0.113	0.095
essential	0.237	0.116
post_covid	-1.063***	0.195
high_ESG*essential	-0.330	0.173
high_ESG*post_covid	0.117	0.125
essential*post_covid	0.258	0.169
high_ESG*essential*post_covid	0.607*	0.241
Industry FE	Yes	
Adjusted R^2	0.417	
N	32,670	
No. of firms	1,089	

Table 4.5: Results of the essential industry triple interaction regression for the EU sample

Notes: In the table, there are results of the regression based on my main regression specification with a triple interaction term for industry essentiality for the European sample. Only the coefficients of interest are reported. All reported standard errors are clustered by Firm and robust to heteroskedasticity.

essentiality itself was not a resiliency factor during the crisis period, as only essential firms with high ESG scores saw the increase in excess returns.

Last thing to note is that coefficient of both *high_ESG* and *essential* were weakly significant at the 10% level. Both the direction and magnitude of the coefficient on *high_ESG* are consistent with the result of the regression of Equation 3.1 and I believe my discussion of it in chapter 4.1 still applies. As for the coefficient on *essential*, this one is a bit of a surprise. Since essential firms pose a lower amount of risk, I would expect them to have lower excess returns (risk premiums) during normal times, if anything. Removing Energy and Utilities firms from the sample leads to *essential* gaining significance at the 1% level and its coefficient value rising to circa 0.357. It is therefore Financial and Healthcare firms that seem to make the coefficient at least weakly significant. This is an interesting finding and would warrant further research as to the causes behind the causes. One can theorize that while all four industries are essential, certain firms in the Energy and Utilities sectors might face issues regarding their environmental practices in the future. If this expectation is prevalent enough within the market, it could be the cause behind the difference, as previous research has shown that market sentiment can affect asset pricing (Ni *et al.* 2015). Examining this issue properly, however, is beyond the scope of this thesis.

4.3.2 ESG as a resiliency factor in countries with high societal trust

Once again, in previous chapters, I describe the intuition behind why I believe the effect of high ESG as a resiliency factor during a crisis might be stronger for firms in countries with higher levels of societal trust. Previous research has shown this to be the case, such as Lins *et al.* (2017) (though Engelhardt *et al.* (2021) have shown the opposite, which makes the problem more interesting to examine). However, research on this effect during the covid-19 pandemic is sparse and the measure of societal trust itself seems to be far from standardized across papers. As in the essential industry effect testing, I will use a triple interaction term using the treatment from Equation 3.1 and a dummy ppltrst, which is equal to 1 when a given country's level of trust is higher than the average of the sample countries. This measure of societal trust is taken from the results of the European Social Survey. Again, the regression equation is an extension of the main Equation 3.1 and can be seen in Equation 4.2.

$$er_{i,t} = \beta_0 + \beta_7 * high\_ESG * ppltrst * post\_covid + \beta * \mathbf{X_{i,t}} + \epsilon_{i,t} \qquad (4.2)$$

where $er_{i,t}$ are daily excess returns, $ppltrst$ is a dummy indicating if the societal trust in a given country is above the average of the sample countries, and $\mathbf{X_{i,t}}$ is a vector of all other dummy and control variables, including all combinations for the interaction term as well as all controls from Equation 3.1. See the results of the regression in Table 4.6. I once again only show the interaction term and its constituents, as all other results are practically identical to the results in Table 4.1 for the European sample. All standard errors are clustered by firm and robust to heteroskedasticity.

In this case, the results are not as positive, though they are just as conclusive as in the case with essential industries. The results clearly show that the effect of the triple interaction is insignificant, and we therefore choose to not reject the null hypothesis.

	Estimate	Std. Error
high_ESG	-0.243**	0.079
ppltrst	-0.019	0.079
post_covid	-1.033***	0.115
high_ESG*ppltrst	0.107	0.114
high_ESG*post_covid	0.187	0.111
ppltrst*post_covid	0.125	0.121
high_ESG*ppltrst*post_covid	0.210	0.173
Industry FE	Yes	
Adjusted R^2	0.416	
N	32,670	
No. of firms	1,089	

Table 4.6: Results of the societal trust triple interaction regression for the EU sample

Notes: In the table, there are results of the regression based on my main regression specification with a triple interaction term for societal trust for the European sample. Only the coefficients of interest are reported. All reported standard errors are clustered by Firm and robust to heteroskedasticity.

I will shortly discuss why I believe these results are different from the previous research I cited. Lins *et al.* (2017) found the effect of societal trust to be additive. First, Lins *et al.* (2017) examined the data during the 2007-8 financial crisis as opposed to the pandemic. They specifically describe it as a crisis in trust in institutions (besides being a financial crisis). This is something that cannot be safely said about the covid-19 crisis, at least not at the very beginning, and at the same time, it seems to be the core driver of the effect in Lins *et al.* (2017). Furthermore, since they were only examining American firms, their measure of trust was a proxy made-up of the number of charitable organizations in a given county and similar. Using firms from multiple countries allows me to use what I believe to be a more accurate measure, specifically the results of the European Social Survey. This very different initial setup as well as the difference in trust identification is then what might have caused these differences.

Engelhardt *et al.* (2021) was examining the effect of ESG and trust during the pandemic using data on trust from the World Value Survey, and found a significant negative effect. However, their research was very different methodologically. They used trust as a dependent variable, split up their dataset by high and low returns, and compared the significance and magnitude of the coefficients on their ESG variable. They also used a cumulative return over a slightly different event window. Once again, it is these differences in research design that might be behind the differences in results.

4.4 Robustness and further ESG effect exploration

In this chapter, I will check the robustness of my results from the previous chapters as well as explore the specifics of the effect of ESG during a crisis using a series of several different tests and approaches.

4.4.1 Cross-sectional returns over the event window

As I described in the robustness checks section of the Methodology chapter, I will run a set of cross-sectional regressions for the entire event window. While these do not allow me to see the effect of high ESG specifically during the crisis period, they include a whole period return which includes both the pre- and post-crisis periods. As such, results consistent with my previous results add a significant amount of credibility to those previous results. To see what regression equation I will be using, see Equation 4.3 below.

$$\begin{aligned} er_i &= high\_ESG_i + ESI_i + CHI_i + tobins\_q_i + \\ &\quad size_i + cash\_ratio_i + leverage_i + \text{Industry dummies} + \epsilon_{i,t} \end{aligned} \tag{4.3}$$

where er_i are the excess returns, meaning the difference between the raw returns and the market risk-free rate of a given company for the entire 30-day period. Raw returns are a natural log of price on the last day of the event window plus all dividends throughout the 30-day period, divided by the price at the beginning of the event window. ESI and CHI are equal to their values on the last day in the main regression event window. For the European sample, I will also test the equation with $high\_ES$ instead of $high\_ESG$, given the findings in chapter 4.1.1. All the other variables are identical to those in the main regression. The variables of interest in all regressions do not suffer from multicollinearity or autocorrelation.

The results of the first set of regressions can be found in Table 4.7. It includes the full regressions as per Equation 4.3 in two variants, one with a high-ESG dummy and another with a high-ES dummy variable. Regressions for the American and Chinese samples can be found further down in this chapter.

	(1) ESG		(2) ES	
Variable	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	-74.907***	15.540	-72.864***	13.057
high_ESG	-0.924	2.425		
high_ES			1.600	0.248
ESI	-0.171***	0.049	-0.171***	0.047
CHI	0.535*	0.252	0.535**	0.198
Tobin's Q	3.159***	0.899	3.141***	0.530
Size	2.211***	0.536	1.776**	0.653
Cash ratio	-6.737	5.671	-7.261	6.960
Leverage	-0.248***	0.069	-0.247***	0.072
Industry FE	Yes		Yes	
Adjusted R^2	0.259		0.260	
N	1,089		1,089	
No. of firms	1,089		1,089	

Table 4.7: Results of the cross-sectional regression for the European sample

Notes: In the table, there are results of the cross-sectional regression specified in Equation 4.3 for the European sample. Specification (1) uses a dummy for a high ESG score, specification (2) uses a dummy for a high ES score. All reported standard errors are clustered by Industry and robust to heteroskedasticity.

The results table clearly shows the statistical insignificance of the coefficient on the variable of interest in both cases, which does not seem to significantly support (but not disprove either) my original finding. It is however still interesting to see the change in the coefficient from -0.924 in the high ESG case to 1.600 in the high ES case when comparing the two specifications. It is indeed what one would expect given the finding in chapter 4.1.1, as the negative coefficient on *high_ESG* seems to be driven by a high G score. In my original findings, the effect of high ESG during normal times in Europe was actually negative, and the net effect post-crisis was only a little larger than 0, which could explain the coefficient on *high_ESG* in this regression. Specifically, if the effect is negative during one half of the period and only slightly positive during the other, the effect over the entire period might be close to 0. Moreover, while it is far from being statistically significant, the coefficient on *high_ES* has a p-value of about 0.248, so it is not outside the realm of possibility for the effect to be real (and positive, compare it to the p-value of *high_esg*, which is 0.532). Perhaps the effect is not strong enough, but at the very least, we can see a move in the right (in terms of supporting my previous results) direction when the G pillar is removed.

	(1) USA		(2) China	
Variable	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	-46.120	8.798	42.127***	8.885
high_ESG	2.623	1.518	0.508	1.283
Tobin's Q	2.907***	0.319	2.016	0.890
Size	1.360*	0.585	-2.245***	0.443
Cash ratio	-3.589	3.852	2.560	8.654
Leverage	-0.094	0.061	0.114*	0.046
Industry FE	Yes		Yes	
Adjusted R^2	0.213		0.235	
N	2,048		370	
No. of firms	2,048		370	

Table 4.8: Results of the cross-sectional regression for the American and Chinese samples

Notes: In the table, there are results of the cross-sectional regression specified in Equation 4.3 for the American and Chinese samples. All reported standard errors are clustered by Industry and robust to heteroskedasticity.

The results for the cross-sectional regression for the American and Chinese samples can be found in Table 4.8. In both cases, both government response indices were removed as they provided no additional information to the model, given that each sample is only made up of observations from a single country. For the American sample, the coefficient on *high_ESG* is rather large, sitting at 2.554 and though it is just weakly significant at the 10% level.

As for the Chinese sample, somewhat consistent with my previous results, the coefficient on *high_ESG* is insignificant. There are still things of note in the results, however. First and foremost, the intercept for the Chinese sample is large, positive (equal to 42.127), and significant at the 0.1% level. Compare that to intercepts of around -75 for the European sample or the insignificant intercept of approx. -46 for the American sample. This tells us that excess returns are actually strongly positive over the event window. This increase is not the result of dividends as from the data, I know that the major dividend payout period is July and almost no dividends were paid out during the event window. It, therefore, follows that stock prices must have gone up considerably during this period. This is important as it suggests that for the Chinese market, the "market rebound" from the crisis came extremely fast. Indeed, excess returns in China only fell for about two trading days and then rebound to even higher levels than before, as I showed in an earlier chapter.

4.4.2 Other events near the event window

In the chapter on the difference-in-differences estimation, I mentioned a possible violation of the parallel trends assumption. This would happen if an event were to happen at roughly the same time as my cutoff for the post-treatment period, and that event would affect the treatment and control groups (high and low ESG firms) differently. Two such events might have occurred during the period, an oil price crash which was the result of a price war between Russia and Saudi Arabia in March 2020, and Brexit, as by the 1st of February 2020, the United Kingdom entered the transitionary period of leaving the EU.

It might be that firms whose stock prices heavily reflect oil prices as well as firms from Britain have substantially different ESG scores on average from the rest of the sample. If that were to happen, the parallel trends assumption would be violated, as described in the previous paragraph. I therefore test for this by running the regression from Equation 3.1 with two modified samples, once without Energy and Utility firms (affected by the oil price crash), and once without British firms (affected by Brexit).

Running the first regression, I find no significant differences between the original regression results (Table 4.1) and the results from the sample without Energy and Utilities firms. In terms of robustness checking, therefore, the results support my original findings. There are, however, interesting results found in the regression without British firms, I report the results of this regression as well as those from the original regression in Table 4.7.

First thing of note is the decreased negative effect of *post_covid* as well as the increased magnitude and significance of coefficient (from 0.271 to 0.310

	(1) EU (original)		(2) EU (no UK)	
Variable	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	-0.787***	0.175	-0.369*	0.163
high_ESG	-0.196**	0.064	-0.182**	0.062
post_covid	-0.990***	0.102	-0.636***	0.092
high_ESG*post_covid	0.271**	0.086	0.310***	0.085
ln(covid-19 cases)	0.141**	0.043	0.192***	0.045
ESI	0.004**	0.001	0.001	0.001
CHI	0.011***	0.002	0.001	0.002
Tobin's Q	0.111***	0.015	0.094***	0.014
Size	0.086***	0.019	0.058**	0.018
Cash ratio	-0.252	0.244	-0.482*	0.213
Leverage	-0.008***	0.002	-0.008***	0.002
MOM	-0.155**	0.054	-0.201***	0.057
MKTRF	1.162***	0.017	1.128***	0.018
SMB	0.563***	0.045	0.409***	0.045
HML	-0.037	0.059	-0.157*	0.064
Industry FE	Yes		Yes	
Adjusted R^2	0.416		0.456	
N	32,670		24,240	
No. of firms	1,089		808	

Table 4.9: Result of the main regression with and without British firms in the European sample

Notes: In the table, there are results of the original regression specified in Equation 3.1 for the European sample. Specification (1) is identical to that found in Table 4.1, specification (2) uses a modified sample without firms from the UK. All reported standard errors are clustered by Firm and robust to heteroskedasticity.

and from 1% to 0.1% significance level) on the treatment term. In terms of robustness checking, this tells us the effect remains qualitatively the same even when using the sample without British firms, also supporting my original finding. There are, however, other interesting differences. HML and the cash ratio coefficients both become significant at the 5% level. The change in HML probably speaks to the differences in the composition of firms in terms of market cap (or something following a similar price trend) between British and mainland firms, but it is otherwise of no interest to me in this thesis. More interesting is the negative effect of higher cash ratios for mainland firms, which is actually quite large in magnitude. I first assumed investors would prefer firms with higher amounts of liquidity during a crisis. However, the results suggest that the need for liquidity during the pandemic (which I believe to be a reasonable assumption) was not as important to investors as not having excess cash during normal times.

We can see that coefficients on both government response indices became insignificant in the reduced sample. This is odd as the values for the United Kingdom were not particularly high or low, compared to its mainland counterparts. The fact nevertheless remains that the significance of these indices dwindled when removing British firms from the sample, suggesting mainland firms' returns were not severely affected by the government measures during the event window. The aforementioned change in HML already hinted at a different company composition in the United Kingdom, compared to the European mainland. Perhaps the indices were significantly reflected only in the stock prices of those kinds of firms that are more prevalent in the United Kingdom, whatever these may be. This issue would certainly deserve further inquiry, however, this inquiry is outside of the scope of this thesis and I will have to leave it for future research. Lastly, I would like to note that the United Kingdom is the only country whose fixed effects significantly influence the results. Running the original regression with country fixed effects, the only significant factor is the United Kingdom, and the results still remain qualitatively the same. As such, I do not see a need to discuss country fixed effects as a robustness check in and of itself in a separate chapter.

4.4.3 Narrowing down the event window

In this chapter, I test both the original regression from Equation 3.1 as well as the cross-sectional regressions using narrower event windows. This approach will allow me to explore more immediate market reactions to the treatment. This will be useful especially for the Chinese sample, as the market rebound from the treatment was much faster than in the case of the other two samples. I specifically test two more event windows, a four-day (2 days pre- and 2 days post-crisis) and a ten-day (5 days pre- and 5 days post-crisis) window around the treatment. For the daily returns regressions, I will be using Equation 3.1. For the cross-sectional regressions, I will be using Equation 4.3. You can find the results of the daily excess return regressions for each of the samples in Table 4.10. For the sake of brevity, I only report the variables of interest. Given the findings from Chapter 4.1.1, I use *high_ES* instead of *high_ESG* for the European sample. All standard errors are clustered by firm and robust to heteroskedasticity. None of the models suffer from autocorrelation and none of the variables of interest suffer from multicollinearity. ESI and CHI were removed from the American and Chinese samples as they were constant during the event window.

	(1) EU 4-day		(2) EU 10-day	
	Estimate	Std.Error	Estimate	Std.Error
high_ES	-0.081	0.232	-0.324*	0.144
high_ES*post_covid	0.377	0.238	1.083***	0.227
Adjusted R^2	0.539		0.417	
N	4,356		10,890	
No. of firms	1,089		1,089	

	(3) USA 4-day		(4) USA 10-day	
	Estimate	Std.Error	Estimate	Std.Error
high_ESG	-0.116	0.123	-0.148*	0.070
high_ESG*post_covid	0.091	0.166	0.144	0.113
Adjusted R^2	0.195		0.167	
N	8,192		20,480	
No. of firms	2,048		2,048	

	(5) China 4-day		(6) China 10-day	
	Estimate	Std.Error	Estimate	Std.Error
high_ESG	0.565***	0.186	-0.064	0.120
high_ESG*post_covid	1.421***	0.357	0.765***	0.179
Adjusted R^2	0.495		0.428	
N	1,480		3,700	
No. of firms	370		370	

Table 4.10: Results of the main regressions for narrower event windows

Notes: In the table, there are results of the original regression specified in Equation 3.1 for all three sample regions (Europa, USA, China). For each region, two specifications are shown. Specifications (1), (3), (5) use modified samples with a 4-day window around the treatment. Specifications (2), (4), (6) use modified samples with a 10-day window around the treatment. Only variables of interest are shown. All reported standard errors are clustered by Firm and robust to heteroskedasticity.

For the European sample, the coefficient for both variables of interest is insignificant during the four-day event window, but both become significant during the ten-day window at least at the 5% level. In fact, the coefficient on the treatment is strongly significant at the 0.1% level and has a large value of 1.083. While the net effect is a little smaller at (1.083-0.324=) 0.759, this is still much larger than the coefficient. This means that high-ESG stocks had net (5*0.759=) 3.795% larger excess log returns over the five-day window after the treatment compared to low-ESG stocks. This strongly supports the idea that at least shortly after a crisis "comes into existence", investors seek stocks of high-ESG firms, perhaps because they perceive them as less risky. The results from the four-day window regression however suggest that this effect is not immediate.

For the American sample, none of the coefficients of interest are significant in either regression, with the exception of a negative coefficient on $high\_ESG$ in the ten-day window, significant at the 5% level. Given the result from the main (thirty-day window) regression, where the coefficient on the treatment is positive and significant, it is possible that the shock on February 24th was somehow different in nature to that on March 11th, as March 11th is still within the thirty-day window of the American sample regression and could therefore be a major source of the effect in the initial regression.

Lastly, the regression of the Chinese sample brings about the most new (and surprising) information. In the initial, thirty-day window regression, only the coefficient on treatment was significant and its magnitude was considerably lower than that of its Western counterparts. Here, however, we can see in both the four- and ten-day windows, the coefficients on the treatment are strongly positive and significant at least at the 1% level. What's more, in the four-day window, the coefficient on $high\_ESG$ is also positive, large, and strongly significant at the 0.1% level.

Following the logic we see in all three event windows for all three regions and knowing the quick recovery of the Chinese market post-treatment, I am thinking of the following. Imagine we split the post-crisis period into three general periods - immediate aftermath, crisis, and recovery. Given how much faster the Chinese market reached recovery compared to its Western counterparts post-treatment, it could be that the Chinese market simply went through the first two periods much faster. It would then follow that we observe the same effect in all three cases, only their timing is considerably different. We do not observe the effect immediately for the Western samples, but we do observe it after a few days, whereas in China, the effect can be seen basically immediately. It is also important to note that China, unlike Western countries, already had some containment measures in place by the start of even the thirty-day event window, suggesting that the overall perception of the crisis had different timing across the two general regions. However, even without this intuition, the overall evidence so far seems to strongly support the original hypothesis, that is that ESG is a resilience factor during a crisis.

As for the cross-sectional regressions, given that all but one coefficients of interest are insignificant, I will not be reporting the results in a table. The one exception is the $high\_ESG$ coefficient in the Chinese four-day regression, where the coefficient is strongly positive at 1.768 and significant at the 5% level, with standard errors clustered by industry and robust to heteroskedasticity. The coefficient values in the other regressions are broadly similar to the net effects found in Table 4.10, though as I noted, they are all statistically insignificant. This is a pattern similar to that in the panel and cross-sectional thirty-day regressions. It seems that while the effect of ESG during a crisis can consistently be found when using daily returns and accurately distinguishing the pre- and post-treatment periods, the effect does not seem to be strong enough to show when comparably long pre- and post-treatment periods are considered together.

4.4.4 Long-term monthly returns

To explore the more long-term effects of ESG during a crisis, I will run a variation on the difference-in-differences regression using three regional samples with long-term data. This is possible because the pandemic was still strongly in effect by the end of Q1 2021, which is the tail end of my long-term data. The regression equation used is identical to Equation 3.1. The results of the three regressions (one for each region) can be found in Table 4.11. All standard errors are clustered by firm and robust to heteroskedasticity. The regressions do not suffer from autocorrelation and the variables of interest do not suffer from multicollinearity.

I use monthly excess returns (calculated analogously to the daily excess returns) as a dependent variable. I have two years of data in total. The whole year 2019 is considered the pre-treatment period, Q2 2020-Q1 2021 is considered the post-treatment period, data during Q1 2020 is therefore left out, as treatments for all regions came into effect during this period and it would be impossible to accurately discern the pre- and post-crisis periods given the available data. $High\_ESG$ is assigned by 2018 ESG data. Covid-19 cases, as well as the government response indices, use the end-of-month value for a given month. Firm accounting data used is from 2018 for the pre-treatment period and from 2019 for the post-treatment period. Fama-French-Carhart factors are the monthly values. Data sources are identical to those used in the main regression.

Considering the variables of interest, results are broadly similar to the daily return regression in terms of significance, with the treatment variable coefficient being significant at the 0.1% level for both Europe and America, though it is insignificant for China. Opposite to the daily return regressions, both of the significant coefficients are negative and quite small in magnitude. Whereas in the thirty-day window regression, the effect of treatment for European firms was about ($15*0.271 \approx$) 4.065% higher excess log returns of high-ESG firms over 15 days during the post-crisis period, here, we find about ($12 * -0.84 \approx$) 10.08% lower excess log returns of high-ESG firms over a whole year. What's more, the initially negative and significant coefficient on $high\_ESG$ for European firms became significant, whereas the initially negative coefficient became positive for American companies.

The lack of significance on the treatment coefficient in the Chinese sample is quite interesting. I theorize this could have something to do with the way China handled the spread of the disease, compared to the US or European countries. As Lu *et al.* (2021) note that China's "covid elimination strategy" led to the containment of outbreaks and the return of normal life in most of the country. The approach generally used by Western countries meant that by the end of Q1 2021, many of these countries still had restrictive nationwide lockdown measures. In the eyes of investors, then, most of the post-treatment period could be considered normal times in China, which could explain the complete lack of an effect of ESG.

Table 4.11: Results of the long-term monthly excess returns regression for all three regions

	(1) Europe		(2) USA		(3) China	
Variable	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	-0.423	0.435	0.255	3.115	1.413	1.673
high_ESG	0.237	0.180	0.745***	0.186	0.298	0.254
post_covid	-4.017***	0.743	-13.407***	3.343	-15.125***	1.751
high_ESG*post_covid	-0.839***	0.250	-1.084***	0.281	-0.279	0.414
ln(covid-19 cases)	0.156*	0.072	-0.818***	0.217	-4.402	3.242
ESI	-0.003	0.004			-0.121***	0.20
CHI	0.055***	0.010	0.296***	0.064	0.279***	0.035
Tobin's Q	0.199***	0.047	-0.166***	0.040	0.478***	0.085
Size	0.089*	0.043	0.167***	0.043	-0.031	0.091
Cash ratio	0.974	0.713	-0.101	0.612	0.836	1.084
Leverage	-0.007	0.004	0.017***	0.003	0.009	0.009
MOM	-0.421***	0.034	0.046	0.044	0.246***	0.051
MKTRF	0.709***	0.025	1.067***	0.027	0.803***	0.051
SMB	0.217***	0.044	0.737***	0.038	0.368***	0.070
HML	-0.250***	0.040	0.315***	0.042	0.589***	0.072
Industry FE	Yes		Yes		Yes	
Adjusted R^2	0.162		0.142		0.113	
N	26,856		49,032		8,880	
No. of firms	1,119		2,043		370	

Notes: In the table, there are results of the original regression specified in Equation 3.1 for all three sample regions (Europa, USA, China). All specifications use a sample with monthly excess log returns between Q1 2019 and Q2 2021, excl. Q1 2020. Accounting data is from 2018 for the pre-treatment period and from 2019 for the post-treatment period. All reported standard errors are clustered by Firm and robust to heteroskedasticity. ESI is not estimated for the USA due to perfect collinearity with *post_covid*.

The data also shows that the magnitude of coefficients on the systematic risk factors decreases significantly. This is interesting as in the initial regression, the coefficients on these factors were relatively high compared to the other two samples (up to 2-3 times higher). Here, they are in line with the other two samples, suggesting the coefficients in the initial event window were largely inflated, though I have no tells as to the reason behind this occurrence. As I said before, this behavior warrants further research, but its close examination is outside of the scope of this thesis.

Overall, the results of this regression suggest a completely opposite effect in the long term compared to the short term, at least for the European and American samples. Specifically, it suggests that as the crisis lasts longer, the resiliency ESG provides turns into a negative factor for excess returns. Possibly, investors get used to the new status quo and start exchanging the risk-mitigating property of ESG for the higher risk premiums of more risky investments. That drives the returns of those investments up and the return of high-ESG investments down. The effect for the Chinese sample remains present only within a few days following the treatment, with no effect in as few as 15 days following the treatment, as well as in the long term.

4.5 Limitations and future research

While I hoped to cover the research question as comprehensively as possible, there are numerous issues raised within this thesis as well as in existing literature that I cannot hope to cover in the scope of academic work of this caliber, as well as methodological approaches I could not implement. In this chapter, I will cover some of the biggest limitations of this thesis as well as several suggestions for future researchers to consider examining.

4.5.1 Limitations of the thesis

Firstly, by far the biggest limitation of this thesis from my point of view is data availability. While I did manage to acquire enough data to examine the research question to an extent I believe is sufficient, there is no doubt that access to other or at least larger and more usable data libraries, especially regarding company data, would have made the results more robust. I could not, for example, access quarterly accounting data for companies. Accounting data is no doubt an important aspect of stock pricing, and being limited to annual data made it rather difficult for me to examine the causal effects of ESG in the medium term.

A special note has to be made about ESG scores themselves. While their use in research suggests they are reasonably reliable and used by ESG-oriented investors, the main issue is that ratings can be rather inconsistent across different ESG rating agencies, which is why researchers often do robustness testing using ESG data from multiple agencies (e.g. Albuquerque *et al.* (2020)). As I did not have access to any other ESG database, I was limited to examining the research question using a single ESG dataset. What's more, for the European sample specifically (which was the focal point of this thesis), there simply was not a large number of companies with all the data available. Currently, ESG reporting is only required of large companies by the NFRD (Council of European Union 2014), and I believe the amount of ESG reporting that is required is not enough to create a complex picture of a given company's sustainability practices. What's more, Refinitiv's own approach to handling missing data might skew the ESG scores of certain companies upwards if they do not report negative data. This issue specifically is, however, almost impossible to reconcile unless one were to build ESG scores for companies from scratch.

Secondly, probably the biggest limitation of this thesis from the reader's point of view is the methodological approach to examining the causal effect of ESG. In this thesis, I examined causality using the difference-in-differences estimator. While I do believe it has the power to uncover causality if its assumptions are fulfilled (which I believe they were), and I believe it was a fitting choice, a more robust approach to causality research would be to use more than one way of uncovering it, such as instrumental variable estimation or even just difference-in-difference-in-differences estimation. I talk about why I think these two methods specifically were not well suited for this thesis but had I had more knowledge and experience, I could have used more advanced techniques to add robustness to my results.

Furthermore, even when using the difference-in-differences estimator, a different regression specification could have been considered. For example, Demers *et al.* (2021) use a very large amount of control variables to come to a conclusion that high ESG offers no resiliency benefits. These variables include some fairly niche parameters, which I nevertheless expect to have an influence on a given firm's stock returns. Few researchers decide (or maybe manage) to include such detailed characteristics in their regressions, which might be why the aforementioned paper's conclusion is so different from most other research. As I was limited by data availability, I could not include these in my regression either.

Thirdly, many more and possibly more interesting additional resiliency factors could have been considered. Factors such as shareholder or customer composition (or countless others) could prove to be augmentative factors to high-ESG firm resiliency, however, these were not covered within this thesis, mostly due to data availability. As I have gained sincere interest in the topic during the writing of this thesis, I hope future research in this field continues to uncover these.

Lastly, while I believe my overall results are quite strongly suggestive of high ESG being a resiliency factor in certain contexts, they are by no means strongly conclusive. Besides the aforementioned robustness measures within a study, a method such as a meta-analysis of existing research in this field might come to more robust conclusions, which might be of more use in practice.

4.5.2 Suggestions for future research

Besides the possible extensions and betterments of this thesis I mentioned in the previous chapter, numerous interesting findings came to light during the writing of this thesis that were not necessarily directly tied to the research question. I believe these warrant further looking into (if they have not been already) by researchers in other fields, I will shortly talk about these in this chapter.

In my initial thirty-day window regression for the Chinese sample, I found unusually high betas on the Fama-French-Carhart systematic risk factors. These were unusual not just within the context of this thesis, but also within the context of the literature I have read and my general knowledge regarding the topic. What's more, this effect disappeared in the long-term regression. Future researchers might want to examine this further to see, whether this is a repeating pattern in the Chinese or other markets and what is the reason behind it.

When examining the industry essentiality as an augmentative resilience factor of ESG, I found that the effect found was only driven by the Financial and Healthcare sectors. Two other sectors I included as essential, Energy and Utilities, did not seem to enjoy this positive effect. I believe there is no doubt that these sectors are crucially important for Western societies to function. I briefly discussed my thoughts about this occurrence in the appropriate chapter, but as it lacked any scientific backing besides my own thoughts, it would be interesting to see a more scientifically robust approach to this question.

When examining other possible events within the event window, I came across results suggesting a significantly different composition of firms in the United Kingdom compared to the European mainland. This probably is not a surprise even from a layperson's point of view, given the United Kingdom's specific position both politically and economically, but I believe further inquiry into the specifics of these differences and the reasons behind them, historical, economical, or political, would make for a valuable addition to human knowledge.

Lastly, one of the most interesting findings I had in this thesis was the negative effect of high G scores on stock returns, even during the crisis. Multiple pieces of evidence pointed in this direction, but what is very interesting is that it only applied to the European sample, though the American sample also exhibited this behavior during normal times. I briefly mentioned my thoughts about this occurrence in the thesis, but once again, a more robust and conclusive examination of it would prove beneficial, I believe.

Chapter 5

Conclusion

In this thesis, I set out to examine the question of whether ESG activities serve as a resiliency factor during a crisis. Furthermore, inspired by previous research in the same and related fields, I examined whether this effect is strengthened by either belonging to an essential industry or being from a country with an above-average level of societal trust.

Having considered all my findings, I conclude that ESG does indeed serve as a resiliency factor during a short- to medium-term period following a crisis. I also find this effect is primarily driven by the S pillar of ESG. I believe this effect is caused by investors seeking less risky investments given the general market uncertainty. This effect is not only significantly strengthened by belonging to an essential industry, it is also primarily driven by it. Being from a high-trust society does not offer this strengthening benefit. Furthermore, I find that as the crisis lasts longer, investors become accustomed to the "new normal", they start selling off their safe high-ESG investments and buying riskier stocks with higher premiums. This leads to an effect reversal, where ESG becomes a negative factor for stock returns in a longer post-crisis period.

Specifically, using the difference-in-differences estimator over three different thirty-day event windows in three different regions, I found a significant positive effect of ESG, controlling for company characteristics, covid-19 spread, containment measures, and industry, resulting in 1.125-4.785% higher excess log returns for stocks of high-ESG firms over 15 days during the pandemic. All results are robust to heteroskedasticity. The causality is robust to a violation of the parallel trends assumption, as well as strongly supported over narrower event windows. For the negative long-term effect, I constructed the same regression as in the thirty-day regression using two years of monthly excess returns.

Unexpectedly, I find a strong negative effect during both crisis and normal times of having a high score in the G pillar. I find this effect both among European and American firms, though it is considerably stronger in Europe. I believe this finding warrants further research that would find the causal effect of this relationship.

Bibliography

Albuquerque, R., Y. Koskinen, S. Yang, & C. Zhang (2020): "Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash." *The Review of Corporate Finance Studies* **9(3)**: pp. 593–621.

Albuquerque, R., Y. Koskinen, & C. Zhang (2019): "Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence." *Management Science* **65(10)**: pp. 4451–4469.

Beloskar, V. D. & S. V. D. N. Rao (2023): "Did ESG Save the Day? Evidence From India During the COVID-19 Crisis." *Asia-Pacific Financial Markets* **30(1)**: pp. 73–107.

Bolton, P. & M. Kacperczyk (2021): "Do investors care about carbon risk?" *Journal of Financial Economics* **142(2)**: pp. 517–549.

Borghesi, R., J. F. Houston, & A. Naranjo (2014): "Corporate socially responsible investments: CEO altruism, reputation, and shareholder interests." *Journal of Corporate Finance* **26**: pp. 164–181.

Brammer, S., C. Brooks, & S. Pavelin (2006): "Corporate Social Performance and Stock Returns: UK Evidence from Disaggregate Measures." *Financial Management* **35(3)**: pp. 97–116.

Breuer, W., T. Müller, D. Rosenbach, & A. Salzmann (2018): "Corporate social responsibility, investor protection, and cost of equity: A cross-country comparison." *Journal of Banking & Finance* **96**: pp. 34–55.

Broadstock, D. C., K. Chan, L. T. W. Cheng, & X. Wang (2021): "The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China." *Finance Research Letters* **38**: p. 101716.

Cardillo, G., E. Bendinelli, & G. Torluccio (2020): "COVID-19, ESG investing, and the resilience of more sustainable stocks: Evidence from European firms." *Business Strategy and the Environment* **32(1)**: pp. 602–623.

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

Carpenter, J. N., F. Lu, & R. F. Whitelaw (2021): "The real value of China's stock market." *Journal of Financial Economics* **139(3)**: pp. 679–696.

Chava, S. (2014): "Environmental Externalities and Cost of Capital." *Management Science* **60(9)**: pp. 2223–2247.

Cornett, M. M., O. Erhemjamts, & H. Tehranian (2016): "Greed or good deeds: An examination of the relation between corporate social responsibility and the financial performance of U.S. commercial banks around the financial crisis." *Journal of Banking & Finance* **70**: pp. 137–159.

Council of European Union (2014): "Directive 2014/95/EU." *Official Journal of the European Union* **330**: pp. 1–9. Available at: (Accessed: 31 July 2023).

Demers, E., J. Hendrikse, P. Joos, & B. Lev (2021): "ESG did not immunize stocks during the COVID-19 crisis, but investments in intangible assets did." *Journal of Business Finance & Accounting* **48(3-4)**: pp. 433–462.

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

El Ghoul, S., O. Guedhami, C. C. Kwok, & D. R. Mishra (2011): "Does corporate social responsibility affect the cost of capital?" *Journal of Banking & Finance* **35(9)**: pp. 2388–2406.

El Ghoul, S., O. Guedhami, H. Wang, & C. C. Y. Kwok (2016): "Family control and corporate social responsibility." *Journal of Banking & Finance* **73**: pp. 131–146.

Engelhardt, N., J. Ekkenga, & P. Posch (2021): "ESG Ratings and Stock Performance during the COVID-19 Crisis." *Sustainability* **13(13)**: p. 7133.

Evans, J. (2022): "How China is Winning the Race for Clean Energy Technology." Available at: https://fairbank.fas.harvard.edu/research/blog/how-china-is-winning-the-race-for-clean-energy-technology/ (Accessed: 31 July 2023).

Fama, E. F. & K. R. French (1992): "The Cross-Section of Expected Stock Returns." *The Journal of Finance* **47(2)**: pp. 427–465.

Flammer, C. (2021): "Corporate green bonds." *Journal of Financial Economics* **142(2)**: pp. 499–516.

Gillan, S. L., A. Koch, & L. T. Starks (2021): "Firms and social responsibility: A review of ESG and CSR research in corporate finance." *Journal of Corporate Finance* **66**: p. 101889.

Goldstein, I., A. Kopytov, L. Shen, & H. Xiang (2022): "On ESG Investing: Heterogeneous Preferences, Information, and Asset Prices." Available at: https://www.nber.org/papers/w29839 (Accessed: 31 July 2023).

Habib, A. M. & N. Mourad (2023): "The Influence of Environmental, Social, and Governance (ESG) Practices on US Firms' Performance: Evidence from the Coronavirus Crisis." *Journal of the Knowledge Economy* .

Halbritter, G. & G. Dorfleitner (2015): "The wages of social responsibility — where are they? A critical review of ESG investing." *Review of Financial Economics* **26**: pp. 25–35.

Heinkel, R., A. Kraus, & J. Zechner (2001): "The Effect of Green Investment on Corporate Behavior." *Journal of Financial and Quantitative Analysis* **36(4)**: pp. 431–449.

Henriksson, R., J. Livnat, P. Pfeifer, & M. Stumpp (2018): "ESG Literature Review." Available at: https://cdn.pficdn.com/cms/pgimquantsolutions/sites/default/files/static_files/pdf/QMA_ESG_Literature_Review_June2018.pdf (Accessed: 31 July 2023).

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

Hsu, P.-H., H. Liang, & P. Matos (2021): "Leviathan Inc. and Corporate Environmental Engagement." *Management Science* **0(0)**.

Humphrey, J. E., D. D. Lee, & Y. Shen (2012): "Does it cost to be sustainable?" *Journal of Corporate Finance* **18(3)**: pp. 626–639.

Jiraporn, P., N. Jiraporn, A. Boeprasert, & K. Chang (2014): "Does Corporate Social Responsibility (CSR) Improve Credit Ratings? Evidence from Geographic Identification." *Financial Management* **43(3)**: pp. 505–531.

Krüger, P. (2015): "Corporate goodness and shareholder wealth." *Journal of Financial Economics* **115(2)**: pp. 304–329.

Landi, G. & M. Sciarelli (2019): "Towards a more ethical market: the impact of ESG rating on corporate financial performance." *Social Responsibility Journal* **15(1)**: pp. 11–27.

Lins, K. V., H. Servaes, & A. Tamayo (2017): "Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis." *The Journal of Finance* **72(4)**: pp. 1785–1824.

Lu, G., O. Razum, A. Jahn, Y. Zhang, B. Sutton, D. Sridhar, K. Ariyoshi, L. von Seidlein, & O. Müller (2021): "COVID-19 in Germany and China: mitigation versus elimination strategy." *Global Health Action* **14(1)**: p. 1875601.

Ni, Z.-X., D.-Z. Wang, & W.-J. Xue (2015): "Investor sentiment and its nonlinear effect on stock returns—New evidence from the Chinese stock market based on panel quantile regression model." *Economic Modelling* **50**: pp. 266–274.

Oakey, D. (2021): "The Future of the Chief Sustainability Officer." Available at: `https://www.deloitte.com/global/en/Industries/financial-services/perspectives/the-future-of-the-chief-sustainability-officer.html` (Accessed: 31 July 2023).

Ramelli, S. & A. F. Wagner (2020): "Feverish Stock Price Reactions to COVID-19." *The Review of Corporate Finance Studies* **9(3)**: pp. 622–655.

Schmidt, A. F. & C. Finan (2018): "Linear regression and the normality assumption." *Journal of Clinical Epidemiology* **98**: pp. 146–151.

Tang, D. Y. & Y. Zhang (2020): "Do shareholders benefit from green bonds?" *Journal of Corporate Finance* **61**: p. 101427.

Wooldridge, J. M. (2007): "Difference in Differences Estimation." Available at: https://youtu.be/ANSfhKOjyAU (Accessed 31 July 2023).

Wooldridge, 1960, J. M. (2012): *Introductory econometrics : a modern approach.* Fifth edition. Mason, Ohio : South-Western Cengage Learning.

World Health Organization (2020): "Coronavirus disease 2019 (COVID-19), Situation report 71." Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200331-sitrep-71-covid-19.pdf (Accessed: 31 July 2023).

Appendix A

Number of firms per country in the final samples

Country	Number of Firms
Austria	26
Belgium	37
Denmark	36
Finland	30
France	127
Germany	148
Greece	16
Hungary	4
Ireland	40
Italy	69
Luxembourg	21
Netherlands	50
Poland	24
Portugal	12
Spain	57
Sweden	111
United Kingdom	281
United States	2,048
China	370

Table A.1: Number of firms per country