TI 2023-042/IV
Tinbergen Institute Discussion Paper

The pricing of climate transition risk in Europe's equity market

Philippe Loyson [1]
Rianne Luijendijk [2]
Sweder van Wijnbergen [3]

1 VU Amsterdam

2 DNB

3 University of Amsterdam and Tinbergen Institute

The pricing of climate transition risk in Europe's equity market*

Philippe Loyson, † Rianne Luijendijk‡ Sweder van Wijnbergen§

July 7, 2023

Abstract

We assess whether climate transition risk is priced in Europe's equity market by analysing relative equity returns of high versus low CO_2-emitting firms. We use a panel data set covering firm-specific carbon emissions of 1,555 European companies over the period 2005-2019. We add to the existing literature by addressing problems in carbon data and by using various econometric methods ranging from panel data analysis to the SCM. Fama-French style panel regressions at both the individual firm level as well as portfolio level suggest that carbon intensity is negatively related to stock returns. Treatment effect models, however, provide some evidence for increased pricing of climate transition risk after the Paris Agreement.

JEL codes: G12, Q54

Keywords: Climate Change,Carbon Emissions Intensity, Paris Agreement, Transition Risk Premia.

*We thank Siem Jan Koopmans and Kristy Jansen for many helpful suggestions and seminar participants at DNB for useful comments

†VU Amsterdam, philippeloyson@live.nl

‡DNB, r.a.h.luijendijk@dnb.nl

§University of Amsterdam, Tinbergen Institute, CEPR, s.j.g.vanwijnbergen@uva.nl

1 Introduction

The Paris Agreement (PA) signed in 2015 was the first binding agreement between countries to avoid dangerous climate change by limiting global warming to well below 2 degree Celsius.[1] Ever since, concerns on climate change have grown. The window to limit global warming to 1.5 degree Celsius above pre-industrial levels is closing rapidly (IPCC (2023)). Governments are being called upon to set more ambitious climate policies to limit the amount of CO_2 companies and individuals will emit in the coming decades. Stricter climate policies feed into significant expenses for polluting firms as they would have to alter production processes to reduce emissions, or could end up having to discontinue operations. The costly adjustment towards a low-carbon economy prompted by the anticipation of stricter climate policies is expected to disproportionately impact polluting firms' asset valuations. Under the assumption that investors are rational and demand to be compensated for bearing additional risk in their portfolios, companies that emit a lot of CO_2 would have to generate higher expected returns as a result of growing climate policy risk.

The implications of climate policy risk on equity returns have been widely debated, especially after the PA, yet empirical evidence is mixed. Most research focuses on U.S. or global equity markets. Some papers argue in favor of the existence of a carbon risk premium, as they find that clean stocks generate lower returns (e.g. Bolton and Kacperczyk (2021); Hsu et al. (2022); Huij and Zwinkels (2021)). Others, however, do not find any significant indication of a carbon risk premium, and instead conclude that portfolios that go long in clean stocks and short in polluting stocks tend to generate higher returns (e.g. Park et al. (2017); In and Monk (2019); Bauer and Wilms (2023)). Various papers assess how stock prices respond to news about climate change, and suggest that the higher returns on clean stocks in recent years are driven by rising climate concerns (e.g. Ardia and Inghelbrecht (2022); Pástor et al. (2022); Engle et al. (2020); Pedersen et al. (2021)). As empirical findings are mixed, it is important to better understand what causes this carbon risk premium puzzle. Do the heterogeneous results point to insufficient pricing of climate policy risk in global equity markets, or can they be traced back to specific methodological choices?

In this paper we assess to what extent climate policy risk is priced in Europe's equity market. EU countries are legally obliged to reach the climate goals embedded in the PA.[2] We add to the existing literature, by making use of four econometric models to assess whether the mixed empirical findings thus far are a result of methodological choices. Furthermore, we also account for problems in carbon data, such as the effect of inflation and exchange rates on carbon intensity which expresses a firm's total carbon emissions relative to its revenues. We shed light on whether investors consider and price climate transition risk by comparing relative equity returns of high versus low CO_2-emitting European firms. We also compare the period pre and post the PA in 2015. The main research question is: to what extent is carbon risk priced in Europe's equity market, and did the PA have any effect on the size of this carbon risk premium? We expect to find a positive impact: the PA has likely led to growing awareness of climate policy risk among investors and may well have brought effective climate policies forward.

We use a panel data set consisting of 1,555 European companies over the period 2005-2019. A five-step approach is used to estimate the carbon risk premium. First, to analyse the short-run relationship between individual firm's equity (excess-) returns and carbon intensity, we perform a straightforward Fama-French style panel regression with fixed effects. Second, to study whether climate policy risk manifests itself at a portfolio level, we use a widely diversified long-short portfolio to construct a

[1] https://unfccc.int/process-and-meetings/the-paris-agreement

[2] In June 2021, the Council adopted the European climate law – a key element of the European Green Deal. With it, EU countries are legally obliged to reach both the 2030 and 2050 climate goals. These targets include at least 55 percent fewer emissions by 2030, and climate neutrality by 2050.

carbon factor. We then use the Fama-Macbeth two-step procedure to determine whether investors regard carbon intensity as a risk to their investment. This method, however, is sensitive to an Errors-in-Variables problem. Therefore, as a third and a fourth step, we use treatment effect models to assess whether there was a shift in the carbon risk premium after the Paris Agreement was concluded in 2015. We first use the classical Difference-In-Difference (DID) approach which requires that unobserved differences are constant over time. We next apply the Synthetic Control Method (SCM) which allows the unobserved heterogeneity to vary over time. The DID and SCM test the effect of the PA on highly polluting companies as classified under the Climate Action 100+ initiative.[3] As a final and fifth step, we allow for different short-run and long-run effects of the relationship between carbon emissions and total returns by estimating an Error Correction Model (ECM).

Our results suggest that carbon risk is not priced in Europe's equity market, implying that investors do not value the risk they bear of investing in polluting companies. The Fama-French style panel regressions indicate that there is no direct significant short-term relationship between stock (excess-) returns and carbon intensity. The long-short portfolio also generates negative cumulative returns over the period 2005 to 2019, and the carbon factor has no additional explanatory value over the three traditional Fama-French factors. The Fama-Macbeth procedure actually finds evidence for a negative (although insignificant) carbon premium. These findings are all counter intuitive to the idea that investors in European stocks require a premium for their exposure to climate policy risk. The two treatment effect models (DID and SCM) do indicate that the PA had a positive effect on stock returns of companies with high carbon intensity, although the premium is insignificant. The treatment model results suggest at least a rising carbon premium, which would be in line with our hypothesis. Finally, using the Error Correction Model, so as to allow for richer dynamics, again displays a this time even significant but negative long-run relationship between CO_2 emissions and total returns. The absence of a priced-in carbon factor suggests the need for more ambitious climate policies, in order for investors to allocate more capital to cleaner firms. Intriguingly, as long as investors fail to properly price carbon risk, "free" hedging possibilities exist.

The remainder of this paper is set up as follows. In Section 2, the most important literature is briefly reviewed. Section 3 explains the sample composition and analyzes the data set. Section 4 explains the econometric models used for the empirics. In Section 4.1 we set up the panel regression approach. Section 4.2 constructs long-short portfolios with the carbon factor and explores the existence of a carbon premium using the two-step Fama and Macbeth procedure. Section 4.3 explains the two different treatment effect models we use; a DID model (4.3.1) and the SCM (4.3.2). Section 4.4 allows for richer dynamics by using an ECM approach. Section 5 shows the results of applying this battery of econometric methods to the data. Section 6 concludes.

2 Literature Review

There is an ongoing academic debate on the extent to which climate-related risks influence asset prices. A diversified body of work exists by now, studying different types of assets with various methodological approaches feeding into highly heterogeneous results. Our research adds to this debate as it focuses on the asset pricing implications of climate policy risk in the European equity market. In this section we first discuss how climate risk can impact asset prices (for a detailed and recent survey, see Giglio et al. (2021), Campiglio (2022)), where after we briefly review the main empirical findings on the pricing of

[3]Climate Action 100+ is focused on companies that are key to driving the global net zero emissions transition. 166 focus companies have been selected for engagement, accounting for up to 80 percent of global corporate industrial greenhouse gas emissions. The full list of focus companies can be found via https://www.climateaction100.org/

carbon risk in equity markets. We make three observations that help to further understand what we might call a carbon risk premium puzzle.

2.1 Climate risks and asset prices

We can distinguish two main types of climate-related risk that could, when materializing, negatively impact returns on financial assets: transition and physical risk (Campiglio et al. (2019)). *Physical risks* on the one hand are related to the direct impact of a changing climate on a company's operations, built capital and assets. When discussing physical risks, a distinction is often made between acute risk events (e.g. extreme weather events) and chronic risk events (e.g. sea-level rise), both of which can negatively affect returns on assets linked to exposed economic activities (Hultman et al. (2010)). *Transition risks* on the other hand do not relate to the direct impact of climate change, but arise as a result of the shift to a low-carbon economy, and is mostly about risks created by stricter future climate policies, as well as risks related to the emergence of new, clean technologies or behavioral changes of consumers and investors (TCFD (2017)). *Climate policy risks* are a type of transition risk, and specifically relate to the impact of the policies likely to be implemented to combat or adapt to climate change. Both physical and transition risk should be priced by investors, but climate change is too diverse a concept to capture in a simple indicator, so data measuring the exposure of individual firms to physical climate risks do not exist. For this reason we focus exclusively on transition risk, or, as we will refer to it from here on, to climate policy risk or carbon risk. When climate policy risks materialise, valuations of polluting firms can erode rapidly, leading to balance sheet deterioration through reduced collateral values and stranded assets. As a result, polluting firms will likely have to reduce the dividend they can pay to their shareholders.

The asset pricing implications of carbon risk can be studied by examining price mutations in the underlying securities of companies most affected by stricter future climate policy. Higher (perceptions of) risk can then be expected to either lead to a decrease in the price of exposed assets or to force the issuers of the asset to provide higher returns for investors as a compensation for the additional risk (Cochrane (2009)). Either way a risk premium will emerge and we can say that the relevant risks are priced in. Failure to find any pricing in of climate policy risk is not just an academic puzzle but should also concern policy makers. Inability to reflect climate policy risk in asset prices may threaten financial stability and will likely delay or even prevent efficient capital reallocation. And only when capital is allocated away from polluting firms towards clean firms, governments and policymakers will be able to avoid climate disasters.

There are two main channels through which climate policy can affect investors' perception of climate policy risk: market segmentation and uncertainty. As regards the first channel, if there is extensive divestment from polluting companies by significant investor constituencies, market segmentation may increase. As a result, a smaller number of investors will end up having to absorb more stocks from polluting companies, and may require a higher risk premium to compensate for what is called *exclusion risk*. The market segmentation theory is supported by empirical literature showing for example that so called sin stocks (alcohol, tobacco, and gaming) generate positive abnormal returns (Hong and Kacperczyk (2009)). Concerning the second channel, investors do not know at which pace the world is transitioning to a carbon neutral economy, nor the impact of this shift or the precise timing and nature of the policy measures that will be taken to bring this about. This fundamental uncertainty over timing, nature and effect of climate policy is what should give rise to a carbon risk premium and this premium is what we focus on in this paper.

2.2 Empirical evidence on the carbon risk premium in equity markets

To assess whether climate policy risks are priced in equity markets, researchers have worked along different lines. Some compare the risk premium of polluting stocks to clean stocks, to see whether investors price the difference in carbon risk exposure (Soh (2017); Alessi (2021); Bolton and Kacperczyk (2020); Hsu et al. (2022)). Others asses whether stock prices react to news on climate policy risk (Faccini), or a change in climate sentiment estimated via textual analysis of news sources (Ardia and Inghelbrecht (2022); Engle et al. (2020); Pástor et al. (2022)). Some papers deploy difference-in-difference approaches around key climate policy events (e.g. Nguyen (2020)). We briefly review the empirical findings documented in the literature thus far, and subsequently discuss three observations that help to further understand the drivers of the so-called carbon risk premium puzzle: the failure to unambiguously find evidence of a carbon risk premium.

A few papers that compare relative equity returns of clean versus polluting firms put forward evidence of a carbon risk premium (including Bolton and Kacperczyk (2019); Alessi (2021); Hsu et al. (2022); Pástor et al. (2022)). Bolton and Kacperczyk (2020) use a panel data set spanning the period 2005 to 2018 and regress returns of over 14,400 individual firms on total carbon emissions. Their results suggest that high-emitting firms generate higher returns, pointing to a carbon risk premium in all sectors in Asia, Europe, and North America, which has been rising in recent years and can in their view be traced to exclusionary screening by institutional investors. More recent research contributions by the same authors suggest that voluntary disclosure of carbon emissions by companies results in lower stock returns relative to non-disclosing firms (Bolton and Kacperczyk (2021)). Hsu et al. (2022) use a more wide ranging measure of toxic emissions and construct a long-short portfolio based on U.S. firms with high versus low toxic emission intensity (including carbon emissions). They also find that high-polluting firms on average earn 4.42% more per year than low-polluting firms and attribute this to the risk of a future regime shift in environmental regulation. Alessi (2021) also use a long-short portfolio to assess relative equity returns of clean versus polluting EU firms, and find evidence of a negative risk premium for clean stocks with good environmental disclosures, suggesting a ceteris paribus lower return on these investments.

A large number of papers follow a similar approach but find no evidence of a priced carbon risk factor (including Choi et al. (2018); Soh (2017); Cheema-Fox (2019); Rohleder et al. (2022); Gimeno and Gonzalez (2022); Bauer and Wilms (2023)). Soh (2017) construct a long-short portfolio based on annual carbon intensity of 700 U.S. firms over a period of ten years. The authors estimate a Fama-French model to which they add carbon risk as a fourth factor, and conclude that low-emitting firms generate higher abnormal returns. Choi et al. (2018) come to a similar conclusion by assessing the returns variation of a long-short portfolio in low and high emitting U.S. stocks over a 41-year period. The authors find that low-emitting firms earn an average abnormal annual return of 3.6-5.3%, inconsistent with the view of the existence of a carbon risk premium. Görgen et al. (2020) construct a carbon risk mimicking portfolio and assess the return variation of 1600 globally listed companies over the period 2010 to 2017. They do not find evidence of a carbon risk premium and attribute this to opposing price movements of polluting firms becoming cleaner.

Various papers assess how policy events impact the relationship between stock returns and firm-level emissions (Ramiah; Hsu et al. (2022); Hengge (2023); Nguyen (2020)). Firms may be positively or negatively exposed to different types of climate risk (physical or transition). While coal companies would likely suffer from realizations of transition risks, renewable energy companies might benefit. And while climate change will negatively affect the value of coastal real estate, it might increase the value of farmland in colder regions of the world. Giglio et al. (2021) therefore argue that investors' perceptions of climate risk are influenced by various climate risk categories, implying that it is difficult to isolate carbon policy risk in a widely diversified long-short portfolio. To better capture the effect of climate policy, many authors also assess whether stock prices react to specific policy events. Hsu et al. (2022) perform an event study to provide evidence that the emission-return relation is linked to uncertainty about environmental policy. They analyze stock price reactions on the date of Trump's U.S. presidential election win, and find a positive effect on polluting stock prices as investors expected environmental regulations to be relaxed. Hengge (2023) find that carbon policy surprises have a significant negative impact on stock returns of high-emission firms. The carbon policy surprises were approximated via price changes in the EU ETS futures price around 98 regulatory events regarding the supply of emission allowances. Nguyen (2020) employ a DID approach to assess the effect of the ratification of the Kyoto protocol on Australian firms' cost of equity. Again they find that this policy shock led to higher cost of capital for polluting firms. These empirical results suggest that polluting stocks respond more negatively to climate policy events than clean stocks, which does suggest or at least is consistent with the emergence of a carbon risk premium.

More recent papers assess how stock prices react to a change in "climate sentiment" (Engle et al. (2020); Ardia and Inghelbrecht (2022); Pástor et al. (2022); Barnett et al. (2020); Noailly (2021); Faccini). Engle et al. (2020) and Faccini use textual analysis to construct a climate concern index based on news about climate change published by major U.S. newspapers, and find that stocks of clean firms have higher returns during periods with negative news about the future path of climate change. Ardia and Inghelbrecht (2022) add a risk component to this climate concern index, and conclude that on days with an unexpected increase in climate change concerns, clean firms' stock prices tend to increase, whereas brown firms' prices decrease. Pástor et al. (2022) use the same media index and confirm that clean stocks outperformed in recent years. The authors also find evidence for lower expected returns for clean stock, using both implied cost of capital and realized returns. They argue that outperformance caused by strengthened climate concerns today is followed by lower expected performance of clean stocks going forward. Once again the evidence is mixed.

The first observation we make is that there is no consensus in the literature on how to capture firm exposure to climate policy risk. On the one hand, future increases in carbon prices would disproportionately impact firms with high levels of CO_2-emissions today. Firms' capacity to absorb such price increases depends on their ability to generate revenues in a carbon efficient manner versus their peers. On the other hand, firms with high levels of CO_2-emissions may face lower risks if they manage to set, disclose and meet decarbonization targets, in anticipation of stricter climate policies. All research contributions that find evidence of a priced carbon risk premium use either total carbon emissions (Bolton and Kacperczyk (2019); Bolton and Kacperczyk (2020)), toxic emission intensity (Choi et al. (2018)), environmental ratings (Pástor et al. (2022)), or a combination of carbon intensity and environmental disclosures (Alessi (2021)). The documented findings are not robust to other metrics such as carbon intensity. This relative carbon metric is used to distinguish clean from polluting firms in various papers that find no evidence of a priced carbon factor (Park et al. (2017); Bauer and Wilms (2023)). Aswani (2023) re-examine the existing carbon data, and conclude that unscaled emissions are correlated with stock returns while relative carbon metrics are not. Furthermore, the authors also find that estimated emissions tend to correlate with financial fundamentals, suggesting that prior evidence for a carbon risk premium mostly captures the association between fundamentals and returns.

The second observation we make is that it is unclear from the literature whether carbon risk is priced *within sectors*. Research contributions based on diversified long-short portfolios constructed by means of a best-in-class ranking within sectors, generally find no evidence of a carbon risk premium. Papers following this approach assume investors refrain from excluding polluting sectors all together to reduce their exposure to climate policy risk. At the same time, the costs associated with climate change will probably impact some industries more than others. Papers that narrow their analysis to high carbon sectors (Batten et al. (2016)) or a set of climate-policy-relevant sectors (Alessi (2021)) appear to find some evidence of a carbon risk factor.

The third observation we make is that papers that isolate the carbon risk premium by assessing stock prices around climate events, rather than focusing directly on (excess-)returns, generate more consistent results in favor of a carbon risk premium. The significant price movements in polluting stocks in response to unexpected changes in climate sentiment point to a connection between carbon return predictability and changes in environmental policies and regulations. These findings also suggest that the relative outperformance of clean stocks in recent years is likely to be followed by lower expected performance of clean stocks going forward (Pástor et al. (2022)). However, it is difficult to accurately measure expected stock returns, especially when there is only a short time series on carbon data available (Giglio et al. (2021)).

A different angle starts from the view that markets may price the anticipation of stricter climate policy in gradually, as different investor groups have diverging preferences and reaction functions. This approach builds on the empirical finding of Bebchuk and Wang (2013), who found that market participants have gradually learned about the usefulness of governance which led to it being priced into financial assets only gradually over time. Pedersen et al. (2021) and Pastor and Taylor (2021), building on this idea, make a distinction between different types of climate-aware investors with various preferences. They predict that clean assets have lower expected returns than polluting assets for two reasons: investors have clean tastes, and clean assets are a better hedge against climate policy risk. Thus, clean stocks' lower expected returns reflect both a taste premium, as well as a risk premium. In the long run, decarbonization should reduce the portfolio's relative performance, if investors indeed have clean preferences (Pastor and Taylor (2021)). But gradual phasing in of climate policy risk into asset prices will confound the measurement of excess returns: when asset prices fall gradually rather than instantaneously, measured returns based on actual price changes instead of anticipated changes will have a downward bias.

In conclusion, we identify multiple problems within the existing literature that could feed into the carbon risk premium puzzle. To account for these in our empirical set-up, we take three steps. First, we use various metrics to capture a firm's exposure to climate policy risk, including carbon emissions and carbon intensity. Second, we construct multiple long-short portfolios based on firm's emission profiles, using different sorting methods that account for sector, size and value exposures. Third, we combine various econometric models often seen in the empirical literature, including panel regressions as well as carbon risk mimicking portfolios, to assess whether we find evidence of a carbon risk premium. We add to the existing literature by employing a DID and SCM method to compare the period pre- and post PA, to see whether this event affected investors' perception of climate policy risk. We focus specifically on the European equity market, as the European Commission has formulated an ambitious climate policy agenda and EU countries are legally obliged to reach the decarbonization targets embedded in the Paris Agreement.

3 Data and Sample

Our primary database covers the period 2005-2019 and is a result of a matching of two data sets by respectively Trucost and Bloomberg. We performed the matching using ISIN as the main identifier. The ultimate matching produced 1,602 unique non-financial European companies for which CO_2 emission data is available. We removed null values and identified outliers using box plots which can be found in in Annex A in figures 6 - 11 [4]. The final data set includes observations for 1,555 European companies.

[4] One other possibility is to winsorize the data, as done by Bolton and Kacperczyk (2021). Winsorization means changing the value of an outlier to the nearest value of the observation that is not an outlier. This method reduces the effect of an outlier by replacing them with less extreme values. Winsorization can be done at different levels, i.e., at 1.0%,

As the availability of carbon data changes over time, the resulting data set is an unbalanced panel.

3.1 Carbon data

Carbon data is characterised by three well-documented problems. First, carbon data is inconsistent across data providers, especially if a large share of the datapoints is estimated on the basis of proprietary models. Second, relative carbon metrics such as carbon intensity express a firm's total carbon emissions per unit of revenue, but do not correct for inflation and/or exchange rate effects. Third, carbon information is provided with a lag of 6 months after the closing of the financial year, which means that carbon data is fed into market prices with a delay. The following paragraphs will discuss the carbon data from two dataproviders (MSCI and Truost), and describe how we account for the problems in carbon data. We prefer Trucost carbon data for this research, as coverage in terms of companies, countries, years, and reported values is the highest. Most of the existing literature also uses this data (including Park et al. (2017), Bolton and Kacperczyk (2019), Bolton and Kacperczyk (2020), Pedersen et al. (2021) and Campiglio et al. (2019)).

Trucost and MSCI source the majority of their data on carbon emissions from the Carbon Dislcosure Project (CDP), rather than it being a measurement. Thousands of companies voluntarily report into this initiative via questionnaires (circa 9,000). The CDP assesses the quality of the information by using models to fill data gaps. The questionnaires and reports used by the CDP follow guidance from the Greenhouse Gas Protocol. This Protocol specifies how to disclose and/or calculate carbon emissions, and distinguishes between three different types of emissions: Scope 1 refers to direct emissions from owned or controlled sources, scope 2 refers to emissions from the generation of purchased electricity, and scope 3 refers to all other indirect emissions from up- and downstream activities along the value chain.

Since scope 1 and 2 emissions are often reported by companies, this data is relatively comparable. Scope 3 data, however, is often estimated by dataproviders on the basis of proprietary models. Owing to a lack of methodological clarity, this data is still very noisy and inconsistent across data providers (Klaassen (2021)). We therefore only look at direct emissions, represented by the sum of scope 1 and 2 carbon emissions ($CE12TC_{i,t}$ for Trucost and $CE12MSCI_{i,t}$ for MSCI).

We use a relative carbon metric to compare companies with different characteristics (i.e. size, industry), while accounting for inflation and exchange rate effects. Carbon intensity is constructed by dividing a company's total carbon emissions by its revenues in millions of US dollars (tCO_2e/USD 1M revenue). This new metric provides information on a company's carbon dependency when generating revenues and, therefore, on its transition risk ($CI12TC_{i,t}$ for Trucost and $CI12MSCI_{i,t}$ for MSCI). Carbon intensities are generally reported in tCO_2/\$M, while we only focus on stock returns of European companies. As a result, inflation and exchange rate effects influence the relative carbon metric, especially when assessing developments over time. When price levels rise, carbon intensity will also rise simply because real revenues are lower than nominal revenues (see Janssen et al. (2021)).

In line with Janssen et al. (2021), we correct our carbon intensity metric in two steps:

1. The carbon intensity values are deflated to 2005 equivalents (starting point of the data set) using the Consumer Price Index (CPI) from the World Bank [5].

2. All deflated values are converted to their corresponding currency using the average exchange rate

2.5%, 5.0% etc. However, this is a risky method since the carbon emission data can vary substantially between companies in different sectors. A coal company probably emits more carbon than a company active within the Communication Services sector.

[5] https://datacatalog.worldbank.org/

in that corresponding year. This data comes from the OECD database[6].

Firm-level data on carbon emissions is available on an annual basis, and reflects the amount of CO_2 emitted over the course of one full financial year (scope 1, 2 and 3). MSCI and Trucost generally provide the carbon information with a lag of circa 6 months, which means that these providers publish data by the middle of the following year. As a result, carbon data is generally fed into market prices with a lag of one year, depending on the time of publication. To prevent relating returns to current-year emissions, we include a lagged variable of carbon intensity ($CI12TC_{i,t-1}$), similar to Bauer and Wilms (2023), Ardia and Inghelbrecht (2022), Ilhan et al. (2021).

The carbon data within our sample is relatively consistent between Trucost and MSCI. A correlation matrix suggests that the datapoints are highly comparable, especially for reported values. Both data providers offer data on absolute carbon emissions as well as carbon intensities (scope 1 and 2). When companies fail to report on their direct emissions the data providers use in-house models to estimate the emissions. Table 1 shows that the correlation is lower when carbon emissions are estimated, both in absolute as well as in relative values (0.86 and 0.72), compared to the reported values (0.91 and 0.91). Table 21 in Annex A indicates that correlations are highest for reported emissions in the sectors Energy and Materials (0.97 and 0.94), probably because many of these companies report under the CDP. In other sectors, however, correlations are still relatively low, which means that the data is not consistent across providers.

Table 1: Correlations Trucost and MSCI (carbon emissions and intensity)

	Estimated emissions				Reported emissions			
	CE12TC	CE12MSCI	CI12TC	CI12MSCI	CE12TC	CE12MSCI	CI12TC	CI12MSCI
CE12TC	1.0000				1.0000			
CE12MSCI	0.8622	1.0000			0.9084	1.0000		
CI12TC	0.3790	0.4360	1.0000		0.3648	0.3943	1.0000	
CI12MSCI	0.2288	0.3359	0.7150	1.0000	0.3323	0.4299	0.9130	1.0000

Furthermore, the panel is unbalanced as many companies only started disclosing their carbon emissions over the last couple of years. Trucost and MSCI provide carbon data from respectively 2005 and 2008 onwards, and both data sets include estimated and reported emissions. Reported data is preferred as this data is generally more reliable and consistent across providers. Table 2 shows the coverage for both data providers and indicates that Trucost offers more data on European companies, of which a larger share is reported instead of estimated.

Table 2: Coverage carbon data from Trucost and MSCI

Coverage	Trucost	MSCI
Companies	1,555	1,105
Geography	All countries	All excl. EE, LT and SI
Time period	2005 - 2019	2008 - 2019
Mean carbon emissions	2.2 mln ton CO2	2.3 mln ton CO2
Mean carbon intensity	199.1 tCO2e/USD $1M	206.0 tCO2e/USD $1M
# estimated emissions	4,219	4,272
# reported emissions	9,525	6,790

[6] https://data.oecd.org/. The OECD database does not provide data for Romania and Bulgaria, so therefore these countries were deleted from our initial dataset of 1602 unique non-financial European countries.

3.2 Financial data

Bloomberg provides data on stock returns and corporate fundamentals. Total stock returns are estimated on the basis of Bloomberg's total return index gross dividends ($totalreturns_{i,t}$), which accumulates movements in the stock price plus dividend payouts during the holding period. The index assumes that dividends are reinvested and is, therefore, helpful in comparing stock returns. The variable $totalreturns_{i,t}$ is non-stationary and trending, so the dependent variable $ARET_{i,t}$ is constructed which is the percentual difference of $totalreturns_{i,t}$ from year $t-1$ to year t.

$$ARET_{i,t} = \frac{(totalreturns_{i,t} - totalreturns_{i,t-1})}{totalreturns_{i,t-1}} * 100\% \tag{1}$$

We subsequently calculate annual excess returns $ER_{i,t}$ by subtracting the risk free rate (R_f) from $ARET_{i,t}$. As a proxy for the risk-free rate the 3-month Euribor rate is used, and as a proxy for the market return rate (R_m), the return on the EuroStoxx600 Index is used. Furthermore, we use size ($logsize_{i,t}$) measured in total assets on a copmany's balance sheet, and book-to-market ($BTM_{i,t}$) data for listed equities from Bloomberg to control our results for other known risk factors in a Fama-French-type framework. Figures used to analyse the data are presented in Annex A.

3.3 Sample composition

The time period spans from 2005 to 2019, and thus includes the signature of the Paris agreement in 2015. This allows us to analyse potential changes in expectations on materialisation of climate policy risk, as may be reflected in market prices. As the availability of carbon data changes over time, the resulting dataset is an unbalanced panel. Frequency of firm-level carbon emissions is yearly, and frequency of stock returns is monthly.

We focus on non-financial listed companies from the Eurozone, Switzerland, Denmark, Great Britain, Hungary, Norway, and Sweden for which Trucost provides carbon data (scope 1+2). Companies that have their country of risk outside Europe were excluded. Romania and Bulgaria were also excluded from the sample, as we could not perform the inflation correction for these countries. In total the dataset contains 22 countries and 1,555 companies. The distribution of firms in our sample across 10 GICS sectors is shown in Table 3. Financials are excluded because it is difficult for banks and insurance companies to measure carbon emissions as these are linked to the composition of their loan books instead of their operational activities. In 2019 the sectors Utilities, Materials and Energy were responsible for most of the absolute carbon emissions. On the contrary, the sectors Health Care, Information Technology and Real Estate emitted relatively little. As regards carbon intensities, the range within sectors is high which suggests that some firms manage to operate much more carbon-efficiently than their peers (see Figure 15 in Annex A).

Table 3 shows the data composition by geography and sector. Great Britain, France and Germany are over-represented in the data set as they cover 20.15%, 14.27% and 13.63% of the data set. Estonia, Hungary, Lithuania, Luxembourg, Malta, Portugal, and Slovenia are all under-represented in the data set as they all cover less than 1% of the data set. The Industrials and Consumer Discretionary sectors are over-represented within the sample as they cover 26.49 % and 16.00%, respectively. The Utilities and Energy sector are underrepresented as they cover 4.10% and 3.39%. The general structure of a panel data set is shown in Annex A in table 19, where in this case there are $N = 1, ..., 1,555$ companies, $T = 1, ..., 15$ years and $m = 1, ..., M$ different variables.

Table 3: Data composition by geography, year and sector

Countries	Obs.	Years	Obs.	Sectors	Obs.
AT	31	2005	478	Communication Services	112
BE	52	2006	508	Consumer Discretionary	249
CH	127	2007	569	Consumer Staples	111
DE	213	2008	566	Energy	50
DK	40	2009	537	Health Care	155
EE	2	2010	597	Industrials	412
ES	76	2011	642	Information Technology	167
FI	4	2012	670	Materials	108
FR	223	2013	594	Real Estate	128
GB	315	2014	922	Utilities	63
GR	24	2015	984		
HU	5	2016	1,349		
IE	25	2017	1,405		
IT	90	2018	1,533		
LT	2	2019	1,512		
LU	6				
MT	4				
NL	53				
NO	56				
PT	15				
SE	144				
SI	3				
Companies	1,555	Observations	12,849	Companies	1,555

Over the whole time period, absolute carbon emissions decreased for all sectors. In 2005 the average company emitted 3.4 million tons of CO_2 per year, whereas in 2019 this was only 1.27 million tons of CO_2. This is a reduction of more than 50%. This could result from energy efficiency improvements, technological innovation, and increased reliance on renewable energy sources. At the same time, however, carbon intensity levels actually increased in most sectors. Only in the Consumer Discretionary, Real Estate and Utilities sector a decrease in carbon intensity is observable. The average company in 2005 had a carbon intensity of 252.7 tCO2e per USD mln in revenue, and in 2019 a carbon intensity of 212.7 tCO2e per USD mln in revenue. The lower reduction pace in carbon intensity compared to total carbon emissions could stem from increased reporting among less polluting firms, as the number of companies in the sample increased over the period.

3.4 Testing for multi-collinearity and unit roots

We test for two conditions before moving to the panel data analysis: multi-collinearity and the presence of unit roots in the variables. Of course perfect multi-collinearity implies a singular data matrix which would make the inversion of that matrix, necessary to estimate the covariance matrix, impossible. But even imperfect but high collinearity makes covariance estimates vulnerable to numerical rounding errors, with imprecise estimates of the covariance matrix as a result. To assess that risk we construct a correlation matrix to test for multi-collinearity. Table 4 shows that the correlation between the explanatory variables is in fact low, which means that we do not need to worry about multi-collinearity problems.

Table 4: Correlations between independent variables

	GDP	inflation	CI12tc	BTM	logsize
GDP	1.0000				
inflation	-0.0316	1.0000			
ci12tc	-0.0353	-0.0034	1.0000		
BTM	-0.0547	0.0400	0.1538	1.0000	
logsize	-0.0709	0.0097	0.1434	0.0858	1.0000

Note: The independent variables are not perfectly correlated, so no multi-collinearity is present in the data.

The presence of unit roots can cause problems in statistical interference involving time series. If a variable shows a trend in the mean, this most likely implies violation of the stationarity condition, possibly because of the presence of unit roots. Pesaran (2007) constructed a unit root test in a panel setting where the null hypothesis H_0 is that all the panels contain a unit root, whereas the alternative hypothesis H_1 states that some panels are stationary. But for the purpose of our panel data analysis we are more interested in a slightly different hypothesis, namely whether all panels are stationary. We therefore use the Hadri unit root test, which has that hypothesis as H_0.

The Hadri (2000) unit root test tests as null hypothesis H_0 whether all panels are stationary, versus the alternative hypothesis H_1 that at least one panel contains a unit root. The test statistic is constructed using a residual-based Lagrange multiplier (LM) with the residuals taken from the regression:

$$y_{it} = \delta_{mi} d_{mt} + \varepsilon_{it}, \quad m = 2, 3 \tag{2}$$

for $i = 1, \ldots, N$. The test statistic is then given by:

$$H_{LM,m} = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{S_{it}^2}{\hat{\sigma}_{ei}^2}, \quad with \quad \hat{\sigma}_{ei}^2 = 1/T \sum_{t=1}^{T} \hat{e}_{it}^2 \tag{3}$$

A problem is that the Hadri unit root test assumes normality and, more importantly, requires a strongly balanced panel data set. So we can only apply the test to the variables $ARET_{i,t}$, $totalreturns_{i,t}$, $GDP_{i,t}$, $logsize_{i,t}$, $inflation_{i,t}$ and $BTM_{i,t}$.

The test statistic is not rejected for the variables $ARET_{i,t}$, $GDP_{i,t}$, $logsize_{i,t}$ and $BTM_{i,t}$ so we can accept the claim that these series are mean stationary I(0)[7]. The test does reject the null hypothesis for the variables $totalreturns_{i,t}$ and $inflation_{i,t}$ with $p = 0.000$. This implies that these series contain unit roots. Because the models will use the variable $GDP_{i,t}$ as control variable, $inflation_{i,t}$ is dropped. For the difference component of total returns, i.e. $\Delta totalreturns_{i,t}$, the test statistic is not rejected, implying that $totalreturns_{i,t}$ is I(1), which is not surprizing given that inflation apparently also is I(1).

The bigger problem is that the carbon variables $CI12TC_{i,t}$ and $CE12TC_{i,t}$ are characterised by incomplete time series for most of the companies in the sample so we cannot include them in the Hadri test: Only 278 companies have carbon data available over the full period under investigation (15 years). So we need a different approach for the unit root test of these two variables. We regress both $CI12TC_{i,t}$ and $CE12TC_{i,t}$, as well as $LogCI12TC_{i,t}$ and $LogCE12TC_{i,t}$ on their own lag and possible company and time fixed effects, as shown in equation 4 for $CI12TC_{i,t}$.

$$CI12TC_{i,t} = \lambda_1 CI12TC_{i,t-1} + \alpha_i + \zeta_t + u_{i,t} \tag{4}$$

[7] For the dependent variable $ARET_{i,t}$ we can clearly accept the H_0 with a test-statistic of -5.3511 and a p-value of 1.000, leading us to conclude that $ARET_{i,t}$ is I(0)

Table 5 and table 6 show the outcome of the regressions and lists the different $\hat{\lambda}$ coefficients, estimated with OLS. All the $\hat{\lambda}$ coefficients are significantly smaller than unity, which suggests that the variable $CI12TC_{i,t}$ is I(0). Since we use the logarithms of $CI12TC_{i,t}$ and $CE12TC_{i,t}$ in the regressions, and we can see in table 6 that the $\hat{\lambda}$ estimates of $logCE12TC_{i,t}$ are admittedly significantly below one but still very close to one, it makes the assumption of a unit root more plausible. Therefore we explore in section 4.4 whether there is a co-integrating relationship between $logCE12TC_{i,t}$ and $totalreturns_{i,t}$ to see whether we can explore dynamics by applying the ECM approach to these variables.

Table 5: Regressions results for $CI12TC_{i,t}$ and $CE12TC_{i,t}$ on their own lag (equation 4)

	$CI12TC_{i,t}$		$CE12TC_{i,t}$	
Type of regression	$\hat{\lambda}$	S.E.	$\hat{\lambda}$	S.E.
No FE, OLS	0.87	0.005	0.85	0.005
Company FE	0.70	0.007	0.13	0.010
Company and time FE	0.70	0.007	0.13	0.010

Table 6: Regressions results for $logCI12TC_{i,t}$ and $logCE12TC_{i,t}$ on their own lag

	$logCI12TC_{i,t}$		$logCE12TC_{i,t}$	
Type of regression	$\hat{\lambda}$	S.E.	$\hat{\lambda}$	S.E.
No FE, OLS	0.83	0.006	0.93	0.004
Company FE	0.54	0.008	0.23	0.009
Company and time FE	0.54	0.009	0.23	0.009

The $CI12TC_{i,t}$ variable is auto-correlated and pretty persistent with a coefficient on the lagged term of 0.87; apparently companies change their carbon emissions intensity only slowly. While the fixed effects (FE) estimator wipes out the firm fixed effects η_i by applying within demeaning, $y_{i,t-1} - \bar{y}_{i,-1}$ is still correlated with $u_{it} - \bar{u}_i$. This is the well-known Nickell Bias (Nickell, 1981). In micro panels where T is small and N is large, this FE estimator is inconsistent with a bias of $O\left(\frac{1}{T}\right)$. But since $T = 15$ and this regression is only used to test for unit roots, we ignore the Nickel Bias in estimating $\hat{\lambda}$ in equation 4. These findings imply that we meet the condition that all variables in the panel regression are integrated of the same order. The independent variable $CI12TC_{i,t}$ and the dependent variable $ARET_{i,t}$ are both of order $I(0)$. The results underlying Equation 4 do not point to a unit root in $CI12TC_{i,t}$, so we feel confident to follow theory and include the level variable in the main regression.

4 Methodology

In this section we outline and motivate the models and estimation methods used. Section 4.1 explains the panel regressions we use to analyze the short-run relationship between stock returns and carbon intensity. Section 4.2 explains how the carbon factor is constructed and highlights the main regressions used in the two-step Fama and Macbeth procedure to determine whether a carbon risk premium exists in Europe's equity market. Section 4.3 explains two treatment effects models, the DID model and SCM, to analyze the effects of the PA on the pricing of carbon risk in Europe's equity market. Lastly, Section 4.4 shows the error correction model (ECM) used to search for any long-run relationship between carbon emissions and total returns.

4.1 Panel regressions

Our first step is a straightforward Fama-French (FF) style panel regression where we add carbon related variables to the standard set of FF factors. The panel data set has time-varying observations on scope 1 + 2 carbon emissions ($CE12TC_{i,t}$), scope 1 + 2 carbon intensities ($CI12TC_{i,t}$), annual returns ($ARET_{i,t}$), log size ($logsize_{i,t}$), book-to-market ratio's ($BTM_{i,t}$) and GDP growth ($GDP_{i,t}$), and time in-varying observations on companies ($company_i$), sectors ($sectors_s$), and countries ($country_c$). One issue in many panel data studies is cross-sectional dependence, whereby all companies in the same cross-section may be correlated due to unobserved common factors. As this data set is on companies, a possibility for such inter-dependencies could be that firms' carbon intensities are affected by the same supply and demand shocks like for example during the COVID-19 crisis. Another issue in panel data sets is unobserved heterogeneity, where the errors are non-independent. Unobserved heterogeneity implies the existence of unmeasured differences between companies. In our case for example, the carbon intensity of company i at time t can be affected by a country's time-invariant policy, such that $cov(CI12TC_{i,t}, country_c) \neq 0$. Unobserved heterogeneity implies that least squares estimation may be biased and inconsistent with asymptotic distributions that have nuisance parameters (Bun et al. (2015)).

We use fixed effects (FE) in the regression model to remove unobserved heterogeneity. An alternative would be to use Random Effects; in contrast to the FE model, the RE model assumes that variation between different effects is uncorrelated with the independent variables and also random: $cov(CI12TC_{i,t}, country_c) = 0$. However, the assumption that all factors control for the unobserved heterogeneity is unlikely. We test which model is preferred in our setup through a Hausman-Wu test (Hausman (1978)), which came out in favor of the FE model [8].

The annual excess stock returns are regressed onto a breakdown of the carbon intensity level, its lagged value, the market risk premium($R_{Mt} - R_{ft}$) and in addition the well known Fama and French (1993) factors. The lagged carbon factor is included because carbon emissions are often reported with a lag. The carbon intensity is based on yearly data, so we conduct a yearly panel regression. Since $ARET_{i,t}$, R_{ft}, R_{Mt} and $GDPi,t$ are in percentages, $BTM_{i,t}$ is around 1 and size is in logarithmic scale, the carbon intensity variable ($CI12TC_{i,t}$) is also in logarithmic scale. A log transformation reduces the skewness of variables and makes the error term more likely to be normally distributed, which makes statistical inference more reliable. Equation 5 shows the regression:

$$\begin{aligned}&ARET_{i,t} - R_{ft} = \beta_1(R_{Mt} - R_{ft}) + \beta_2 GDP_{i,t} + \beta_3 BTM_{i,t} + \beta_4 log(size_{i,t}) + \\ &\beta_5 log(CI12TC_{i,t-1}) + \beta_6(log(CI12TC_{i,t-1}) - log(CI12TC_{i,t-2})) + \gamma^{'} \boldsymbol{X}_{i,t} + \varepsilon_{i,t}\end{aligned} \tag{5}$$

where $\boldsymbol{X}_{i,t}$ is a vector of fixed effects for $sector_s$, $country_c$, and $timeperiod_t$ and $\varepsilon_{i,t}$ is the error term. In order to account for cross-sectional dependence in the FE model, sector cluster-robust standard errors are used. Section 5.1 elaborates on the results obtained from the fixed effects regressions with different combinations of fixed effects for country, sector and time periods.

Finally in Table 25 in Annex C we show the results for the same regression, but with with country fixed effects and a dummy variable distinguishing companies with low and companies with a high carbon intensity. The dummy variable $LH_{i,t}$ equals 1 for companies with a carbon intensity below the median in that corresponding sector and 0 for those with a carbon intensity above the median in that

[8] the Hausman statistic: $W = \frac{(\hat{\beta}^*_{FE} - \hat{\beta}^*_{RE})^2}{\mathrm{Var}(\hat{\beta}_{FE}) - \mathrm{Var}(\hat{\beta}_{RE})} \overset{H_0}{\sim} x_i^2$ testing the RE model (H_0) against the FE model (H_1) equals $W = 2429.17$ with p-value is $p = 0.000$, so H_0 is rejected and the FE model is preferred.

corresponding sector. By including the market risk premium, the regression already accounts for some part of the time fixed effects, so only country fixed effects are included.

4.2 Constructing and testing the carbon factor

In our next step we go beyond simply assessing whether a carbon intensity factor has an impact on a firm's risk premium by trying to explicitly assess whether carbon risk is priced in. To do this we follow the Fama-Macbeth's two-step procedure (Fama and MacBeth (1973)). The basic idea of the Fama and Macbeth two-step procedure is to first assess whether the assets' excess returns are influenced by the factors that explain some risks in the cross-section in each time period. This constitutes the first step. The second step consists of a regression to directly assess whether these factor exposures (the coëfficiënts of the first step regression) are significantly priced [9]. This section explains how we set up the Fama and MacBeth (1973) two-step procedure to estimate parameters based on the usual three Fama-French factors and an added carbon factor.

In the first step, we augment the standard Fama-French three-factor model with a fourth "carbon return factor". To develop the carbon return factor, we construct a carbon risk-mimicking portfolio: the "Intense-Minus-Less intense" portfolio (IML). Intense stands for companies with high values for $CI12TC_{i,t}$, and less intense stands for companies with low values for $CI12TC_{i,t}$. The IML is long in companies with high carbon intensity and short in companies with low carbon intensity. Constructing portfolios based on CO2 intensity levels is useful to examine whether there is a systemic return from investing in a CO2-efficient European equity portfolio. The method of weighting matters for the risk premiums. Here, the portfolios have been constructed using equal weighting, which is the average of all the stock returns. An alternative procedure would be to weigh the stocks based on relative market equity, i.e., value weighting. In FF applications, portfolios are usually re-balanced monthly. However, carbon data is reported annually and the ranking of the companies by carbon intensity does not change much from year to year, so there is no need for constant re-calibrating of the portfolios. This is in line with Gimeno and Gonzalez (2022), who also do not re-calibrate due to the stationarity of carbon data. Because the availability of carbon data changes over time, the year with the most carbon data was chosen to perform the ranking, which was 2018 in this data set (see table 3).

In line with Fama and French (1993), the portfolios are first sorted on size and value using quantiles. In order to account for heterogeneity, the portfolios are then double sorted on carbon intensity, where companies with the lowest carbon intensity are placed in portfolio 1 and companies with the highest carbon intensity are placed in portfolio 5. However, when using this double sorting, there is a risk that the IML factor will be biased towards certain sectors. Table 7 shows the distribution of companies per portfolio. The portfolios add up to 1,192 companies, as Bloomberg provided complete monthly return data from 31-10-2008 to 31-08-2020 for 1,192 out of 1,555 companies. The table shows that double sorting based on size/value and carbon intensity leads to biased portfolios. The Information Technology sector is highly over-represented in portfolio 1 when using double sorting on size, while the Utilities sector is over-represented in portfolio 5. When using double sorting on value, the Real Estate sector is not even included in portfolio 1.

[9] see Cochrane (2009) for a brief exposition of this approach

Table 7: Company distribution per portfolio using double sorting on size and value

	Double sorted on size and carbon intensity					Double sorted on value and carbon intensity				
Portfolio	1	2	3	4	5	1	2	3	4	5
Utilities	4	3	9	9	38	3	11	19	18	12
Industrials	62	93	89	89	77	70	96	95	83	66
Consumer Discretionary	48	60	54	47	34	60	36	45	53	49
Materials	9	23	19	28	28	12	21	19	30	25
Energy	6	5	10	15	17	1	5	6	12	29
Real Estate	7	16	38	37	24	0	3	12	48	61
Communication Services	25	13	20	20	32	26	26	24	15	18
Information Technology	80	34	22	15	12	50	44	39	16	13
Consumer Staples	7	25	20	24	31	19	27	20	16	24
Health Care	54	34	24	19	16	62	35	29	17	5

In order to reduce the sector bias, portfolios are constructed based on three grouping variables. First, the firms are classified per sector, leading to 10 classifications in line with the GICS sector classification (without financials). Second, per sector, the firms are split into two equally weighted shares by taking the median size of the firms per sector. This ensures that half the companies included in each sector portfolio are small and half are big. Third, the firms are split into five equally weighted portfolios based on their carbon intensity to compare the returns across portfolios. This approach takes into account the variation in size and carbon intensity and rules out the possibility that results are entirely driven by sector exposures. This method is also followed when in the second step, firms are split based on value (book-to-market ratio) so that half the companies included in each sector portfolio have a high value and half have a low value. Figure 18 shows the cumulative returns for all portfolios when triple sorted on sector and on value. Both ways of sorting show that the less carbon-intense portfolio produces higher cumulative returns than the more carbon-intense portfolio. This is counter-intuitive, as riskier firms (companies with high carbon intensity) are expected to yield higher returns than less risky firms (companies with low carbon intensity). This first insight does not indicate a carbon premium. Pástor et al. (2022) also found this and stated that the greater focus on global warming had sparked more interest in environmentally-friendly investments, resulting in improved returns. In order to calculate the IML factor, the weighted average of size and value sorting is calculated.

$$IML_t = 0.5(IML_{t,size}) + 0.5(IML_{t,value}) \tag{6}$$

It is important that the IML factor is not correlated with the other factors. The SMB (Small Minus Big) and HML (High Minus Low) factors are downloaded from the official Fama and French website[10] and are offered monthly for the European market. As a proxy for the risk-free rate (R_f), the 3-month Euribor rate is used, and as a proxy for the market return rate (R_m), the return on the EuroStoxx600 Index is used. Table 8 shows that the factors are not correlated.

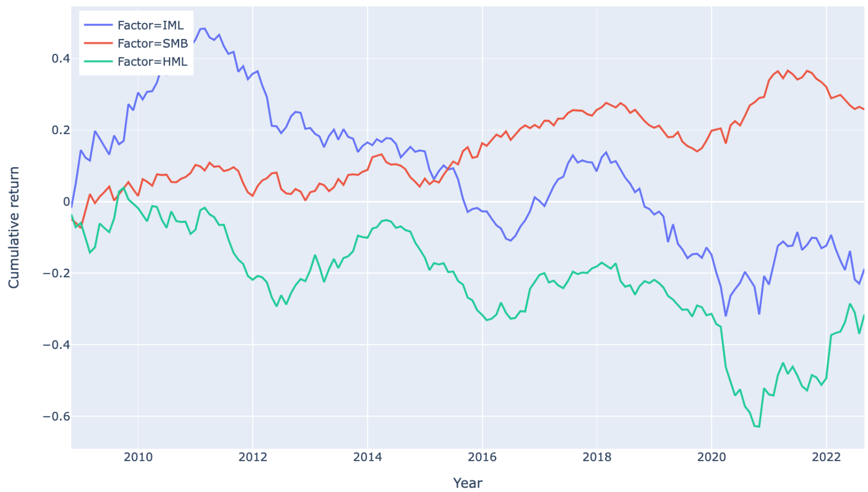
Table 8: Correlations between the different factors

	IML	SMB	HML	Mp
IML	1.0000			
SMB	-0.0395	1.0000		
HML	0.3460	-0.1404	1.0000	
Mp	0.3301	0.0339	0.4171	1.0000

[10] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Figure 1 plots the cumulative returns of the IML, SMB and HML factors for the period from December 2008 to October 2022. The cumulative returns of the IML factor is positive until 2016, which is in line with figure 18, where companies with high carbon intensity performed worse than companies with low carbon intensity. Then, the IML is again positive for two years and then from 2019 onwards, the IML drops all the way to -31% at the end of 2020. Hence, companies with high carbon intensity performed worse than companies with low carbon intensity in the last two years.

Figure 1: IML performance compared to SMB and HML



Now that all factors have been constructed, we run rolling regressions for each of the assets monthly excess returns (here named as $MRET_{i,t} - R_{f,t}$) using Fama and French's three risk factors plus the carbon factor (Mp, SMB, HML and IML):

$$MRET_{i,t} - R_{f,t} = \alpha_{i,t} + \beta_{0,i,t} Mp_t + \beta_{1,i,t} IML_t + \beta_{2,i,t} SMB_t + \beta_{3,i,t} HML_t + \varepsilon_{it} \tag{7}$$

$MRET_{i,t}$ are the monthly t returns of each company i. IML_t, SMB_t and HML_t are the carbon -, size- and value factors, respectively and $\varepsilon_{i,t}$ is the pricing error. Here, a window of 24 months of returns is rolled forward one month at a time, and therefore the beta estimates are estimated from the end of October 2010 onwards. Rolling regressions are useful for examining whether there might be a 24-month period where the Fama French three factors + carbon factor were particularly strong. These regressions analyze the changing relationships among variables over time. Section 5.2 elaborates on the results.

Next we use the Fama and MacBeth (1973) two-step procedure. In the first step, the procedure regresses $i = 1, ..., 1,192$ asset excess returns on the risk factors to estimate factor exposures (the beta's):

$$\begin{aligned} &MRET_{1,t} - R_{f,1,t} = \alpha_1 + \beta_{1,Mp} Mp_{1,t} + \beta_{1,IML} IML_{1,t} + \beta_{1,SMB} SMB_{1,t} + \beta_{1,HML} HML_{1,t} + \epsilon_{1,t} \\ &\vdots \\ &MRET_{i,t} - R_{f,i,t} = \alpha_i + \beta_{i,Mp} Mp_{i,t} + \beta_{i,IML} IML_{i,t} + \beta_{i,SMB} SMB_{i,t} + \beta_{i,HML} HML_{i,t} + \epsilon_{i,t} \end{aligned} \tag{8}$$

The second step regresses all $i = 1, ..., 1,192$ assets' excess returns for each period with the estimated betas from step one now as independent variables so as to directly estimate the risk premia ($\hat{\lambda}$) for each risk factor beta. This premium is expected as a reward for exposure to certain risks. The second step is then used to test if the risk factors from step one are significantly priced:

$$\begin{aligned}
&MRET_{i,1} - R_{f,i,1} = \gamma_{1,0} + \lambda_{1,1}\hat{\beta}_{i,M_p} + \lambda_{1,2}\hat{\beta}_{i,IML} + \lambda_{1,3}\hat{\beta}_{i,SMB} + \lambda_{1,4}\hat{\beta}_{i,HML} + \epsilon_{i,1} \\
&\vdots \\
&MRET_{i,T} - R_{f,i,T} = \gamma_{T,0} + \lambda_{T,1}\hat{\beta}_{i,M_p} + \lambda_{T,2}\hat{\beta}_{i,IML} + \lambda_{T,3}\hat{\beta}_{i,SMB} + \lambda_{T,4}\hat{\beta}_{i,HML} + \epsilon_{i,T}
\end{aligned} \tag{9}$$

Section 5.2 shows the results we obtained for the factor exposurers and the respective risk premia.

4.3 Treatment effect models

The FM approach is an ingenuous attempt to assess whether specific risk factors are priced in, but it has one clear problem, already noted by Fama and MacBeth themselves (Fama and MacBeth, 1973): since the second stage regressors are the parameters of the first round, the procedure suffers from an errors-in-variables problem. And its rolling regression nature makes it more difficult to assess whether new information about a relevant specific event had a significant impact on risk premia. This is an issue because of the Paris Agreement (PA), in which a large number of countries committed to undertake yet to be specified policies to reach sufficiënt carbon reduction to keep the global temperature increase below 2 degrees, and preferably below to 1,5 degrees. The PA took place in 2015, and therefore this event can be seen as a certain treatment to companies. We therefore use treatment effect models to investigate its impact on stock prices and risk premia before and after the PA between companies with high carbon intensity and companies with low carbon intensity. Since signatories of the PA committed themselves to specific emission targets and to an annual assessment of the adequacy of their climate policy measures, it is plausible to assume that climate policy risk increased after that date, with presumably a corresponding shift in risk premia. We distinguish explicitly between companies with high carbon intensity and companies with low carbon intensity, as the former are more likely to be targeted (or at least will affected more) by the measures required to achieve the PA goals.

To perform correct inference using treatment effects and to have sufficient pre-and post-intervention information, companies without complete data for the carbon intensity variable between 2005-2019 were deleted in this section, leading to 305 unique companies with complete data from 2005 to 2019. All sectors and 17 countries are still represented in the data set.

The purpose of treatment effect models is to anticipate the result that the treated group would have yielded if the event would not have happened, utilizing the control group's result as a reference. The approach relies on the asssumption that any difference between the control group and the treated group are constant in the absence of treatment. An obvious problem is the selection of the control group. We first apply the traditional Difference-In-Difference (DID) method where the control group is chosen subjectively (see 4.3.1). In Section 4.3.1 we elaborate on the choice of a control group, and show that this choice is not clearcut and the assumption of constant differences between the two groups in the absence of treatment may not be warranted. Therefore we also apply a more data-driven method, the Synthetic Control Method (SCM), that does not rely on such an assumption. Section 4.3.1 shows the setup for the empirical analysis using the traditional DID method and Section 4.3.2 explains the original SCM and how we extend it for use with multiple treated units.

4.3.1 Difference-in-difference estimation

This part of the analysis uses the econometric implementation of the DID approach by Wooldridge (2010). It is not straightforward to identify a good control group, as all companies face some kind of transition risk towards a low-carbon economy. Companies are classified as 'affected by the PA' when they have a high carbon intensity and classified as 'not affected by the PA' when they have a low carbon intensity. Section 5.3.1 shows exactly how the treatment and control groups are classified. The variable indicating this classification is called $treatment_i$, which is 0 for a company i not affected by the PA and 1 for a company i that is affected by the PA. Next we define the time effect. The dummy variable $time\_effect_t$ takes the value 0 if the observation is before or in 2015 (as the PA took place in November 2015) and 1 if the observation is after 2015. Finally we construct the $DID_{i,t}$ variable, which is the interaction term between the time and treatment group dummy variables.

In the first model to measure the magnitude of this DID effect, a complete OLS regression for yearly excess returns ($ARET_{i,t} - R_{ft}$) is employed without fixed effects. This model assumes that there are two groups (companies with high carbon intensity and companies with low carbon intensity) and two periods (before and after 2015):

$$\begin{aligned} ARET_{i,t} - R_{ft} &= \beta_1(R_{Mt} - R_{ft}) + \beta_2 BTM_{i,t} + \beta_3 log(size_{i,t}) + \beta_4 log(CI12TC_{i,t-1}) \\ &+\beta_5 treatment_i + \beta_6 time\_effect_t + \beta_7 DID_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{10}$$

In the second model used to assess the magnitude of the DID effect, the regression for yearly excess returns ($ARET_{i,t} - R_{ft}$) is run on only fixed effects for companies and time. In this model, no other variables are included so they are absorbed by the fixed effects:

$$ARET_{i,t} - R_{ft} = \alpha_i + \lambda_t + \beta_1 DID_{i,t} + \varepsilon_{i,t} \tag{11}$$

One crucial assumption in DID estimation is the parallel trend assumption. This is crucial because the estimation becomes biased when this assumption is violated. The assumption means that in the absence of the PA, both the control group and treatment group would have followed the same trend, i.e., the difference between the two groups would have been constant over time in the absence of treatment. No statistical tests exist to confirm this assumption, but a visual representation is useful to inspect parallel trends. Furthermore, since we base our assessment on annual excess returns instead of monthly or daily returns, we assume that the anticipation effect of the Paris Agreement has little effect on our dependent variable pre-2015 . Section 5.3.1 shows the results.

4.3.2 Synthetic Control Method

The synthetic control method (SCM) was introduced by Abadie and Gardeazabal (2003) and Abadie et al. (2010) with the aim of investigating the effects of aggregate interventions. The DID method in the previous section assumes that unobserved differences are constant over time, whereas the SCM allows the unobserved differences to vary over time. The treatment effects are estimated by constructing a combination of control units with different weights instead of a combination of control units with equal weights (DID). Therefore, the DID can be considered as a special case of the SCM where all cross-sectional units have equal weights. One advantage of the SCM method over the DID method is that the control group is objectively chosen, whereas in the DID method the control group is chosen subjectively. Zeng et al. (2021) state that the data-driven selection of the control units compensates for the defect of the DID procedure. The SCM seeks to adjust the control group's weighting such that it shares the same pre-event attributes as the treated group and is most like the treated units. The estimated counterfactual is the path of the synthetic control group outcome if the intervention would not have happened. The original set-up by Abadie and Gardeazabal (2003) has one treated unit, usually a city, region or country. Annex B provides a detailed model description of the SCM if only one company would have been used as treated unit. We extend this approach to multiple treated units.

Since there are 276 control units, this may lead to overfitting. Therefore, following Abadie and Gardeazabal (2003), the donor pool is limited to units that have similar characteristics as the treated units. The model also assumes that sufficient pre-and post-treatment information is available, which in our case is ten years pre-treatment (2005-2015) and four years post-treatment (2016-2019). Advantages of using the SCM are that the estimators preclude extrapolation outside the data, i.e., the weights are within 0 and 1. Furthermore, synthetic controls do not need access to post-treatment outcomes when designing the model, and it safeguards against p-hacking because it is completely data driven. The results are shown and discussed in Section 5.3.2.

4.4 Error Correction Model

Finally we apply an error correction model (ECM) to assess whether there is a long term relationship between carbon emissions and stock returns. An ECM is able to work with non-stationary data to estimate the short-run and long-run relationship between variables. All the variables in the ECM must be co-integrated with each other. In order to meet this condition, all variables should contain unit roots and be integrated of the same order. The evidence on whether the carbon emissions variable has a unit root is mixed. We already saw that the simple regression of the variable on its own lag just falls short of supporting the presence of a unit root, which is somewhat of a surprise at least for the unscaled emissions variable $CE12TC_{i,t}$. However the coefficient of the $logCE12TC_{i,t}$ variable in Table 5 is close to one at a value of 0.93. We therefore do test whether there is a cointegrating relation between $logCE12TC_{i,t}$ and $totalreturns_{i,t}$. Pedroni (2004) claimed that the normal Dickey-Fuller (DF) test is not appropriate when working with panel data. Instead they construct a panel co-integration method which computes three different test statistics, namely the Modified PP t-statistic, the PP t-statistic and the Augmented Dickey–Fuller t-statistic. The null hypothesis H_0 is no co-integration versus the alternative hypothesis H_1 that the panels are co-integrated. Table 9 below shows that all three tests reject the null hypothesis so we can conclude that $totalreturns_{i,t}$ and $logCE12TC_{i,t}$ are indeed co-integrated.

Table 9: Co-integration tests between $totalreturns_{i,t}$ and $logCE12TC_{i,t}$

	Test statistic	p-value
Modified Phillips-Perron test	6.3914	0.0000
Phillips-Perron test	-1.857	0.0316
Augmented Dickey-Fuller test	-5.5902	0.0000

Now that we have established a co-integrating relationship between $totalreturns_{i,t}$ and $logCE12TC_{i,t}$, the ECM is set up as follows:

$$\begin{aligned}\Delta totalreturns_{it} = \mu + \textstyle\sum_{j=1}^{T_{11}} \alpha_{1j} \Delta totalreturns_{i,t-j} + \sum_{j=1}^{T_{12}} \beta_{1j} \Delta logCE12TC_{i,t-j} \\ + \lambda ECT_{i,t-1} + u_{1it}\end{aligned} \tag{12}$$

Here, $T_{lm}, l, m = 1, 2$ denotes the number of lagged values of $\Delta totalreturns_i$ and $\Delta logCE12TC_i$. Furthermore, μ, α and β are the regression parameters, λ is the speed of adjustment, $u_{lit}, l = 1, 2$ are the disturbance terms and $ECT_{i,t-1} = totalreturns_{it} - \hat{\phi}_0 - \hat{\phi}_1 logCE12TC_{it}$ is the error correction term. Section 5.4 shows the results of the ECM.

5 Results

In the following sections we show the results obtained using the various methods explained in Section 4. We briefly summarise the results upfront: first, Section 5.1 shows that there is no short-term relationship between stock returns and carbon intensity. Section 5.2 elaborates on the rolling regressions using the constructed IML factor and finds a negative although insignificant carbon premium in Europe's equity market. So these results when combined reject the hypothesis that there is a carbon policy risk premium in Europe's equity market. We then switch to methods incorporating the possibility of a shift in the premium after the Paris Agreement was concluded in 2015 on the presumption that only then the threat of carbon policy became real. Section 5.3.1 and 5.3.2 do show that the PA had a positive although still insignificant effect on stock returns of companies with high carbon intensity. This indicates a rising carbon premium, which would be partially supportive of the carbon risk premium hypothesis, although insignificantly so. And Section 5.4 reportss a significant but *negative* long-run relationship between carbon emissions and total returns.

5.1 Panel regressions

Table 10 shows the FE model (equation 5) with fixed effects for country, sector, time periods, or a combination of these. The estimated coefficients for $R_{M,t} - R_{f,t}$, which is the excess return on the market portfolio, show a significant positive parameter estimate in all regressions, so an increase in the market portfolio return is correlated with an increase in excess stock returns. Because in all six regressions these estimates are positive and significant, this is a robust result, in line with standard CAPM. The estimates for $BTM_{i,t}$ are also statistically significant at a significance level of 1% or 10%, meaning that if a company's book-to-market grows, a company's expected excess returns is likely to decrease. The variable $logsize_{i,t}$ is positively associated with stock returns. So bigger companies produce higher returns. The Fama French factors are robustly significant.

Table 10: Panel regressions including carbon intensity and combinations of fixed effects

	Annual return (1)	Annual return (2)	Annual return (3)	Annual return (4)	Annual return (5)	Annual return (6)
$R_{M,t} - R_{f,t}$	1.024***	1.030***	1.411***	1.430***	1.023***	1.396***
	-64.73	-64.85	-14.37	-14.44	-64.67	-14.26
GDP	-1.238***	-1.214***	0.683*	0.963**	-1.289***	0.614*
	(-6.99)	(-6.63)	-2.35	-2.97	(-7.07)	-2.11
Book-to-market	-0.193***	-0.180***	-0.168***	-0.169***	-0.194***	-0.180***
	(-29.43)	(-28.46)	(-27.25)	(-27.39)	(-29.44)	(-28.14)
Log Total Size	0.00656***	0.00691***	0.00511**	0.00522**	0.00728***	0.00578**
	(-3.59)	(-3.75)	(-2.95)	(-2.85)	(-3.78)	(-3.23)
Log CI12TC t-1	-0.00167	0.00208	0.0011	0.00102	-0.00188	-0.00229
	(-0.62)	-0.93	-0.52	-0.47	(-0.69)	(-0.87)
Δ Log CI12TC t-1	-0.0163*	-0.0185**	-0.0229***	-0.0224***	-0.0158*	-0.0204**
	(-2.54)	(-2.90)	(-3.70)	(-3.63)	(-2.46)	(-3.29)
Constant	0.0591**	0.100***	-0.0904**	-0.0728*	0.0779*	-0.111***
	-2.93	-3.59	(-3.08)	(-1.96)	-2.57	(-3.55)
Country fixed effects	n	y	n	y	y	y
Sector fixed effects	y	n	n	n	y	y
Time fixed effects	n	n	y	y	n	y
R2	0.408	0.404	0.442	0.444	0.41	0.448
N	8916	8916	8916	8916	8916	8916

Note: This table shows panel regressions of yearly excess returns as the dependent variable on the CI t-1 and delta CI t-1 and control variables and country, sector and time fixed effects for the period from 2005 to 2019. *, **, *** denote significance on the 10%, 5%, and 1% level, respectively. Country cluster robust t-statistics are displayed in parentheses. Significance tests are based on two-sided t- tests.

However, the parameter estimates of the carbon intensity variable $logCI12TC_{t-1}$ are not significant in any of the six combinations of fixed effects: the carbon intensity variable has no significant impact on (excess) returns. This suggests that investors, at least between 2005 and 2019, do not yet seem to require a carbon risk premium; transition risks are not priced in the European equity market. Investors apparently do not require higher returns for bearing more transitional risk. The rate of change $\Delta logCI12TC_{t-1}$ however has a significant although *negative* effect on excess stock returns for all combinations of fixed effects. It is not entirely clear how this result should be interpreted, as it suggests that firms that are becoming less carbon intense tend to generate higher returns. Görgen et al. (2020) and Pástor et al. (2022) present similar results, and suggest that the relative outperformance of low carbon stock is driven by a heightened sense of urgency to act on climate change. Investors increasingly demand low carbon stock thereby inflating stock prices and returns, especially when these price gains are not instantaneous and thus show up as positive capital gains over the period of time prices take to realign. At the same time, one may also interpret the rate of change variable as capturing the expectation of a future increase in total excess returns. Either way, the negative sign does not conform with our hypothesis.

Table 11: Panel regressions including carbon emissions and combinations of fixed effects

	ARET (1)	ARET(2)	ARET (3)	ARET (4)	ARET (5)	ARET (6)
$R_{Mt} - R_{ft}$	1.021***	1.027***	1.380***	1.400***	1.021***	1.377***
	-64.59	-64.72	-14.05	-14.13	-64.53	-14.06
GDP	-1.223***	-1.211***	0.643*	0.953**	-1.275***	0.607*
	(-6.90)	(-6.62)	-2.2	-2.93	(-6.99)	-2.08
Book-to-market	-0.193***	-0.180***	-0.168***	-0.169***	-0.194***	-0.181***
	(-29.80)	(-28.78)	(-27.66)	(-27.80)	(-29.78)	(-28.53)
Log Total Size	0.00915**	0.0120***	0.0106***	0.0113***	0.0105***	0.00985***
	-3.12	-4.48	-4.21	-4.35	-3.48	-3.47
Log CE12TC t-1	-0.00268	-0.00468*	-0.00530**	-0.00594**	-0.00332	-0.00431
	(-1.12)	(-2.43)	(-2.89)	(-3.17)	(-1.36)	(-1.84)
Δ Log CE12TC t-1	-0.00168	-0.000625	0.000495	0.000867	-0.00131	-0.0000638
	(-0.52)	(-0.20)	-0.16	-0.28	(-0.40)	(-0.02)
Constant	0.0619**	0.127***	-0.0630*	-0.0375	0.0823**	-0.102**
	-3.06	-4.56	(-2.14)	(-1.01)	-2.71	(-3.25)
Country fixed effects	n	y	n	y	y	n
Sector fixed effects	y	n	n	n	y	y
Time fixed effects	n	n	y	y	n	y
F-test	407.65 (0.0000)	250.83 (0.0000)	414.42 (0.0000)	202.86 (0.0000)	186.92(0.0000)	276.64 (0.0000)
Df	8915	8915	8915	8915	8915	8915
R2	0.407	0.404	0.442	0.444	0.41	0.447
N	8916	8916	8916	8916	8916	8916

Note: This table shows panel regressions of yearly excess returns as the dependent variable on the CE and control variables and country, sector and time fixed effects for the period from 2005 to 2019. *, **, *** denote significance on the 5%, 1%, and 0.1% level, respectively. Significance tests are based on two-sided t- tests. Trucost data is used.

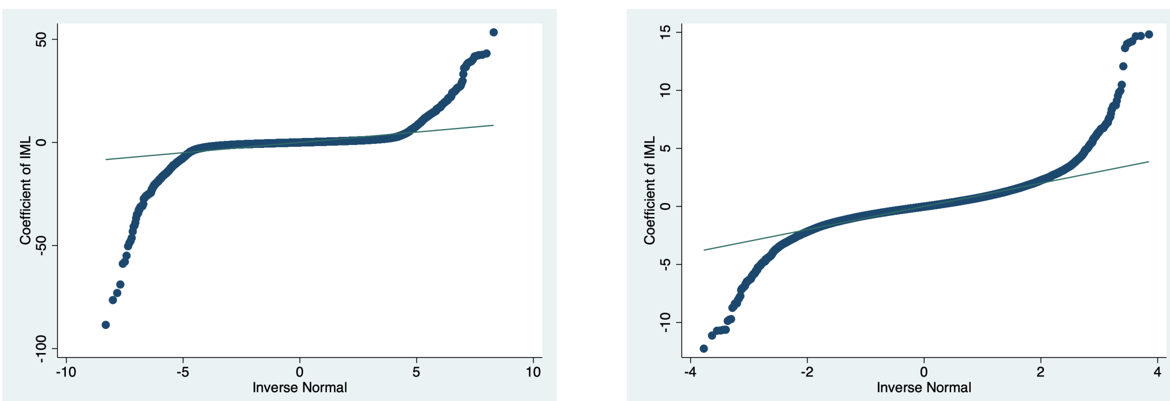
Table 11 again presents results of panel regressions, but now with unscaled carbon emissions as explanatory variable. We once again find that carbon emissions as well as growth in carbon emissions have an insignificant effect on total returns. Notably, the coefficients of both the level $logCE12TC_{t-1}$ and the difference component $\Delta logCE12TC_{t-1}$ have a negative sign, which suggests that an increase in carbon emissions actually weighs on excess returns. This is in stark contrast with Bolton and Kacperczyk (2019), who find a positive and significant relationship between stock returns and the (growth in) unscaled carbon emissions, while they do not find any evidence for a relationship between carbon intensity and stock returns. The authors include scope 1,2 and 3 emissions in their cross-sectional regression model and apply a large number of controls while focusing predominantly on the U.S. market. Bolton and Kacperczyk (2021) extend the analysis to a global level and again conclude that stock returns are positively related to the (growth in) unscaled carbon emissions. When performing the analysis per region, however, the authors find less clear results as for Europe none of the emission metrics is highly significant, which is in line with our findings. All this taken together suggests considerable geographic variation in the carbon premium throughout the world.

We next check whether our results are similar for low and high carbon intensive companies by including a dummy variable in the panel specification. If the carbon intensity level of a company lies above the median level in the sample, the dummy variable is set to 1 and to 0 otherwise. Table 25 in Annex C shows the results for regression equation 5, but with country fixed effects. The parameter estimates for the other variables are consistent with the findings in table 10. The dummy estimate $\hat{\gamma}$ is negative and statistically significant. This suggests that firms with high carbon intensity have a more negative influence on stocks' excess returns, which is not conform our hypothesis. To better understand the relationship between equity prices and carbon risk, the next section shows the results obtained from the Fama and MacBeth two-step regressions with an extra carbon factor where portfolios are constructed using the Fama and French (1993) framework.

5.2 Constructing and testing the carbon factor

Equation 7 represents the rolling regressions performed on 24-month windows to allow for different beta estimates for different periods. Rolling regressions are useful because investors' awareness and hence incorporation of climate change into required returns may vary over time. Since the regression uses monthly data for 14 years (2008-2022), a window of 24 months is used, as longer rolling windows usually yield smoother estimates. Figure 2 shows the estimated values of the coefficient of the Intense-Minus-Less intense (IML) factor for each company in the 24-month rolling regressions, where the excess monthly return per stock is regressed on the market, size, value and carbon factor. The IML is long in companies with high carbon intensity and short in companies with low carbon intensity, and is useful to examine whether there is a systemic return from investing in a CO2-efficient European equity portfolio.

Figure 2: Estimated IML coefficients before and after the PA



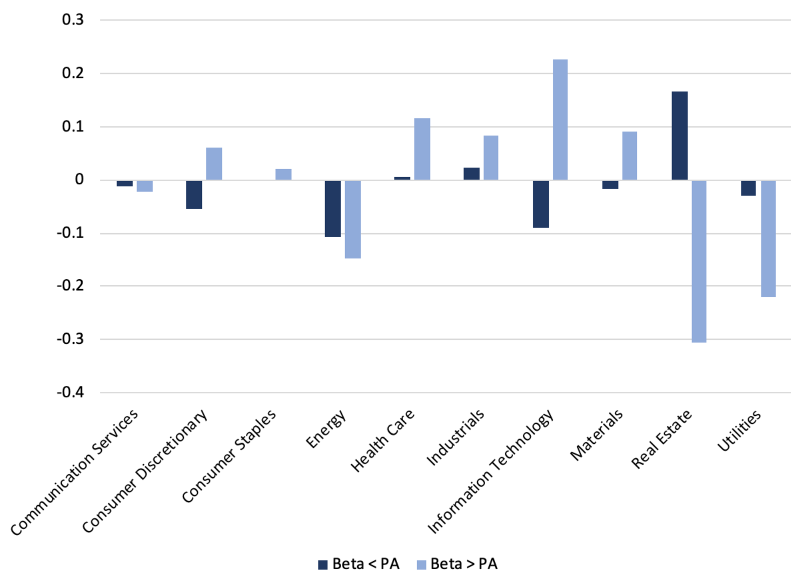
(a) Before the PA **(b) After the PA**

Note: Values of the IML coefficient on the regression of the excess monthly return of each company on all the factors (Mp, SMB, HML, IML).

The estimated coefficient for the IML factor varies from $\hat{\beta}_1 = -88.48$ to $\hat{\beta}_1 = 53.38$ before the PA and from $\hat{\beta}_1 = -12.26$ to $\hat{\beta}_1 = 14.82$ after the PA. After the PA, the IML factor estimates fall in a smaller range than before the PA. Therefore, after the PA, going long in companies with high carbon intensity and short in companies with low carbon intensity will result in lower monthly excess returns because the estimates for IML are smaller. The tails of the normal quantile plot are fatter before the PA, meaning that the possibility for extreme values is greater than after the PA. So, portfolios with companies with high carbon intensities were more volatile before the PA than after the PA. Also, both figures show that the left tails are fatter than the right tails. This indicates that more extreme negative estimates exist, meaning higher extreme losses on portfolio's that are long in companies with high carbon intensity.

Figure 3 shows the mean of the different estimated beta coefficients of the IML factor per sector before and after the PA. Before the PA, the estimated beta coefficient for the IML factor was negative for the Communication Services, Consumer Discretionary, Energy, Information Technology and Materials sectors. After the PA, more beta estimates became positive, namely the Consumer Discretionary, Consumer Staples, Health Care, Information Technology and Materials sectors. These sectors seem to respond positively to the PA because the average betas go from negative to positive. Higher values for beta mean that the IML factor is positive. So, portfolios with more companies with high carbon intensity produce higher monthly excess returns. This indicates that the PA seems to have had a positive effect on the IML factor in different sectors, and thus on the carbon risk premium in those sectors. However, looking at the sectors with higher carbon intensities, i.e. the Energy and Utilities sectors, the average betas after the PA are more negative than before the PA. These sectors respond in the opposite way to the PA agreement when compared to the other sectors.

Figure 3: Average beta IML before and after PA per sector



However, only 10.77% of the listed EU companies show an IML coefficient that is statistically significant at a 5% significance level before the PA, and only 10.57% show an IML coefficient that is statistically significant at the 5% significance level after the PA. The IML factor therefore has low significance in the rolling regressions used to analyze the changing relationships among the different factors and monthly excess returns over time. The percentages of significant p-values from the IML factor are also lower than those of the SMB and HML factors, which are respectively 19.52% and 16.98% before the PA and 16.70% and 19.12% after the PA.

We next apply the Fama and MacBeth (1973) two-step regression, because it allows us to estimate the carbon risk premium directly. The first step (equation 8) estimates 1,192 different stocks monthly excess returns on each of the factors Mp, SMB, HML and IML. This regression is done by a Weighted Least Squares approach, as Hou et al. (2015) suggests that this approach is more suitable than OLS, because the estimated coefficients will not be biased toward smaller companies. Since monthly asset returns do not fluctuate much from one month to another, the returns could be correlated. In order to test for this autocorrelation, the Durbin Watson Test is performed on the residuals:

$$DW = \frac{\sum_{t=1}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad (13)$$

where e_t are the residuals. The test statistic is $DW = 2.06912$, which indicates there is little to no autocorrelation in monthly excess returns. To check the robustness of the first regression and to see if there are no omitted variables, every two years the mean of the residuals is calculated to see if there are any structural effects. Table 12 shows there are no clear shift in average residuals, so we can perform step 2 of the Fama and MacBeth regression.

Table 12: Average residuals calculated every two years

Years	2008-2010	2011-2012	2013-2014	2015-2016	2017-2018	2019-2020	2021-2021
Average residuals	0.005149	0.000781	0.001068	0.000298	-0.00321	-0.00058	0.00253

In the second step (equation 9), the monthly cross-sectional regressions are run on the estimated betas obtained from step 1 to directly estimate risk premia. The Fama and MacBeth procedure is prone to the errors-in-variables (EIV) problem (Görgen et al. (2020)), and is known to only provide standard errors corrected for cross-sectional correlation and not for time-series auto-correlation (Fama and French (1988)). Table 13 shows the results of the second step where Newey and West (1986) standard errors are calculated. This procedure is used because it adjusts for autocorrelation and heteroskedasticity. However, it is still uncertain if these standard errors fully correctfor any autocorrelation in the carbon intensities.

Table 13: Fama and MacBeth results

Risk premia:	Mp	SMB	HML	IML	R^2
Monthly Excess returns	0.0076	0.0035	-0.0057	-0.0029	0.2966
Standard errors	0.0047	0.0026	0.0034	0.0047	
T-statistic	1.6094	1.3519	-1.6756	-0.6235	
P-value	0.1075	0.1764	0.0938	0.5330	

Note: The table shows that all risk premiums lack significance. Newey and West standard errors are used and shown in the second row. P-values are calculated using a 95% confidence level.

We find that none of the estimated risk premia are significant. The carbon risk premium is -0.29%, but insignificant and therefore lacks the ability to explain the cross-section variation in excess returns. This finding, together with the panel regressions in Section 5.1, indicate that investors do not seem to require additional compensation for their exposure to climate policy risk. This is not conform our hypothesis. Furthermore, the outcome of this assessment is not in line with Bolton and Kacperczyk (2019) who also performed the Fama and MacBeth procedure and found a significant positive risk premium for U.S. companies. Figure 19 in Annex C shows how the risk premia vary over time, and even suggests that the risk premium does not significantly change after the PA.

We subsequently assess whether the inclusion of the carbon factor improves the explanatory power of the traditional Fama French three factor model. Gibbons et al. (1989) developed a test to measure which asset pricing model performs best[11]. The test is performed on the standard CAPM model (Mp

[11] The test is developed to measure the efficiency of a given portfolio by testing if the intercept of α_i in $r_{it} = \alpha_i + \sum_j \beta_{i,j} f_{jt} + \epsilon_{it}$ is zero for all assets excess monthly returns on the monthly factor returns, where $f_{j,t}$ are the different factors Mp, SMB, HML and IML. One advantage of this test is that it takes into account the covariance matrix

as only factor), the standard Fama and French three factor model (Mp, SMB and HML as factors) and on the Fama and French three factor + carbon factor model (Mp, SMB, HML and IML as factors). Furthermore, 20 different portfolios are used which are sorted on value, size, carbon intensity, and a combination of these as a diversified number of portfolios increases the power of the test. From table 14, it is clear the original Fama and French three-factor model performs the best as this model has the lowest p-value and mean $|\alpha|$. The model with the IML factor has a lower GRS test statistic and it produces a higher mean adjusted R^2. However, the higher mean adjusted R^2 can be caused by the extra IML factor. The IML factor, therefore, does not seem to explain the portfolio returns any better than the original Fama and French model. When only looking at the mean $|\alpha|$, the model with IML factor does outperform the original CAPM model. We conclude that adding the IML factor to the original Fama and French three factor model does not enhance the explanatory power of the model, which suggests that carbon emissions are not a significant driver of stock returns.

Table 14: GRS test results

Asset pricing model	GRS statistic	p-value	Mean adj. R^2	Mean $\|\alpha\|$
CAPM	1.8558275	.01985318	.74978411	.00163167
Fama and French	2.1076563	.00628381	.80223577	.0012858
Fama and French + IML	2.1042333	.00641737	.84111811	.00133693

5.3 Treatment effect models

The results from the panel regressions and the Fama Macbeth procedure do not suggest a strong relationship between (un)scaled carbon emissions and stock returns. The carbon parameters are insignificant and anyhow have the wrong sign; if anything the results sofar suggest that firms with lower carbon emissions tend to generate higher returns. These results obviously do not conform to our carbon policy risk premium hypothesis. A possible explanation is that the panel analysis yields unreliable results due to the highly unbalanced sample. We therefore assess whether treatment effect models are better suited to capture the effect of climate policy risk on stock returns.

5.3.1 Difference-in-difference estimation

The DID estimation is performed for three different classifications of the treatment and control groups. First, the treatment group is determined by the number of Climate Action 100+ companies included in the data set. In 2017, MSCI and the Carbon Disclosure Project (CDP) identified the 100+ companies responsible which collectively are responsible for over 2/3 of GHG emissions worldwide[12] [13]. In our data set 29 companies with complete data for 2005-2019 were classified as high polluters under the Climate Action 100+ initiative. We do not account for the other stock return drivers, $BTM_{i,t}$ and $logsize_{i,t}$, and base the composition of the first treatment group purely on the Climate Action 100+ classification. The control group is created by closely mimicking the sector distribution of the treatment group but with lower carbon intensities. As the polluting sectors only consist of a small number of similar companies with lower carbon intensities, it was not possible to fully mimic the sector distribution.

of the different factors. The GRS test statistic is calculated by $f_{GRS} = \left(\frac{\tau-n-k}{n}\right)\frac{\boldsymbol{\alpha}'\Sigma^{-1}\boldsymbol{\alpha}}{1+\boldsymbol{\mu}_f'\Sigma_f^{-1}\boldsymbol{\mu}_f}$, where τ is number of months, n is the number of assets, k is the number of factors, Σ_f is the covariance matrix of the different factors, Σ is the covariance matrix of the residuals and $\boldsymbol{\alpha}$ are the different alpha's.

[12] https://www.theguardian.com/sustainable-business/2017/jul/10/100-fossil-fuel-companies-investors-responsible-71-global-emissions-cdp-study-climate-change

[13] Climate Action 100+ is an initiative that brings together investors worldwide to engage with polluting companies with the aim to lower their emissions. see: https://www.climateaction100.org/

As a result, for the first classification only 58 companies are used out of the 305 available companies, and therefore the second classification takes into account more companies. The second classification is done by calculating quantiles for the carbon intensity variable and classifying companies that are in the lowest 40% in terms of carbon intensity as the control group and companies in the highest 60% as the treatment group. With this classification, all companies in the sample are used. However, using this classification, a sector bias is observed as the average carbon intensity differs substantially between different sectors. In the third approach, the high-emitting sectors (Energy, Industrials, Materials and Utilities) are classified as the treatment group and the low-emitting sectors (Communication Services, Consumer Discretionary, Health Care and Information Technology) are classified as the control group. The Consumer Staples and Real Estate sector are excluded from this classification. Table 15 shows the sector distribution as well as the descriptive statistics of both the treatment and control groups for the three different classification methods.

Table 15: Sector distribution and descriptive statistics for the treatment- and control groups

Classification method:	100+ companies		quantiles		low vs. high sectors	
Group: control = C, treatment = T	C	T	C	T	C	T
Communication Services	3	0	19	6	25	0
Consumer Discretionary	4	4	20	21	41	0
Consumer Staples	3	1	9	19	0	0
Energy	3	5	0	15	0	15
Health Care	4	2	14	7	21	0
Industrials	5	5	42	47	0	89
Information Technology	0	0	17	2	19	0
Materials	4	6	2	27	0	29
Real Estate	0	0	3	15	0	0
Utilities	3	6	0	20	0	20
Total	29	29	126	179	106	153
Mean carbon emissions (tCO2)	994,180	2.58e+07	501,131	7,233,645	697,316	8,172,191
Mean carbon intensity (tCO2/$M)	40.33	855.08	21.07	394.48	47.60	414.41
Mean BTM	.47	.73	.49	.63	.53	.56
Mean totalsize	20,832	96,598	23,932	24,078	26,214	24,731
Mean E_r	13.59%	5.22%	11.73%	11.01%	11.91%	10.92%

To observe the possible effects of the PA, the DID estimation from equations 10 and 11 is run for all three classification methods; the results are given in table 16. The treatment effect is not significant for any of the classification methods, which suggests that stock returns are not significantly higher for companies that are more affected by the PA. From table 16, it also becomes clear that the PA had no significant effect on treated companies' stock returns. In the first classification method, the estimated DID coefficient is $\hat{\beta_{DID}} = 0.063$ for the OLS regression with no fixed effects and $\hat{\beta_{DID}} = 0.053$ for the OLS regression with company and time fixed effects. However, both are insignificant, with t-statistics substantially smaller than $|1.96|$. This also holds for the classification methods where quantiles and sectors are used to classify companies as control or treatment group. The positive estimate for the DID coefficient means that the PA may have had a positive effect on the yearly excess returns for treated companies, but the evidence is not significant in the statistical sense.

Various econometric problems may still feed into the results. To assess whether the results of the DID change when using other carbon metrics, we also ran the regressions in equation 10 and 11 using the unscaled carbon emission variable $CE12TC_{i,t}$ as an independent variable. We again find insignificant estimates for the DID coefficient in the regressions with the unscaled carbon emissions data. In terms of R^2, all models fit the regression with comparable goodness of fit.

Table 16: DID estimation results, inflation corrected carbon intensity data

Classification method:	100+ companies		quantiles		low vs. high sectors	
DID model:	OLS	OLS with FE	OLS	OLS with FE	OLS	OLS with FE
$R_m - R_f$	.983***		1.077***		1.080***	
	(-13.42)		(-22.83)		(-19.94)	
Book-to-market	-.137***		-.172***		-.184***	
	(-6.35)		(-7.79)		(-6.95)	
Log Total Size	-.011		-.00598		-0.00552	
	(-0.95)		(-1.74)		(-1.41)	
Carbon intensity (CI t-1)	-0.00468		-.00383		-0.00557	
	(-0.56)		(-0.83)		(-1.16)	
Treatment effect	-.00705		-0.000602		-.01513	
	(-0.31)		(-0.04)		(-1.40)	
Time effect	-.0359		-.0489**		-.0476**	
	(-1.06)		(-2.60)		(-2.30)	
DID	.0593	.05269	.00178	.0099	.0236	.0271
	(-0.83)	(1.36)	(-0.8)	(0.57)	(-0.91)	(1.33)
Constant	.233*	.312***	.170***	.432***	.169***	.418***
	(-2.32)	(9.57)	(-6.78)	(24.79)	(-7.25)	(21.36)
Time fixed effect	n	y	n	y	n	y
Company fixed effect	n	y	n	y	n	y
R^2	0.4882	0.3829	0.4769	0.4649	0.4708	0.4528
F-stat	529.32	36.60	1858.92	159.28	3066.72	131.06
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	560	560	4004	4004	3388	3388

Note: This table shows the DID panel regressions of yearly excess returns as the dependent variable on the CI and control variables. Robust t-statistics are displayed in parentheses. *, **, *** denote significance on the 10%, 5%, and 1% level, respectively. Significance tests are based on two-sided t- tests.

Figure 20 in Appendix C suggests that the three classification methods using EU companies, at least graphically, seem to display parallel trends, so the estimates in table 16 can be considered unbiased.

An important qualification stems from the fact that all European companies are treated by the Paris Agreement, as the EU's commitment to reduce carbon emissions affects the entire economy, including our control group. We therefore also perform the DID with U.S.companies as control group. Former U.S. president Donald Trump in June 2017 communicated his intent to withdraw from the PA, feeding into lower climate policy risks for U.S. companies. Three years after the announcement, the U.S. officially withdrew from the PA in November 2020, due to Article 28 of the PA [14]. As a result, U.S. companies presumably felt less need to transition toward a low-carbon economy and were less likely to be confronted with carbon pricing or similar climate policies. The data from Trucost for the U.S. companies only included data from 2008 to 2019, so only EU companies with complete data for 2008 to 2019 are included. To compare them with companies with high carbon intensity in Europe, companies with carbon intensity < 300 were deleted, leading to 58 U.S. control companies and 35 EU treated companies. The results are shown in table 26 in Annex C. Table 26 does show significantly positive estimates of the DID term: so we can conclude that European companies had to pay a higher (or less negative) carbon premium in the EU after the PA when compared to U.S. companies.

[14]This article states that during the initial three years of the PA's implementation in a particular country, notice of withdrawal cannot be given. As a result, the official withdrawal period of the U.S. only represented 107 days.

5.3.2 Synthetic Control Method

The SCM allows for unobserved differences to vary over time. This means that, contrary to the DID approach, the parallel trend assumption does not have to hold to generate unbiased results. The results of the SCM are shown in the graphical analysis in figure 4. We use the bias correction version of the synthetic control proposed by Abadie and L'Hour (2021) to model the results and to adjust for differences in predictor variable values between a treated unit and its donor pool. Companies with an average carbon intensity of 30 tCO_2/\$M or less were deleted from the data set to reduce the chance of overfitting and to create a donor pool that resembles the treated units. This leads to a data set consisting of 22 (22 Climate Action 100+ companies) treated units and a donor pool of 102 companies.

Figure 4a displays the excess yearly returns trajectory for the 2005-2019 period. Up to 2011, the results show there is little difference in yearly excess returns between the synthetic companies and the treated companies. From 2011 to 2014, the synthetic control group did not fully reproduce the excess yearly returns of the treated companies. After the PA, the synthetic control group yields lower returns, indicating that if the PA would not have taken place, treated companies would have had lower returns than they have now. This first indication is in line with the positive sign of the DID estimate in Section 5.3.1. Table 17 shows that the average company characteristics are similar between the treated units and the synthetic companies. This suggests that the synthetic control group provides a better comparison for the treated companies than the average of the donor companies.

Table 17: Means of company characteristics measured before the PA

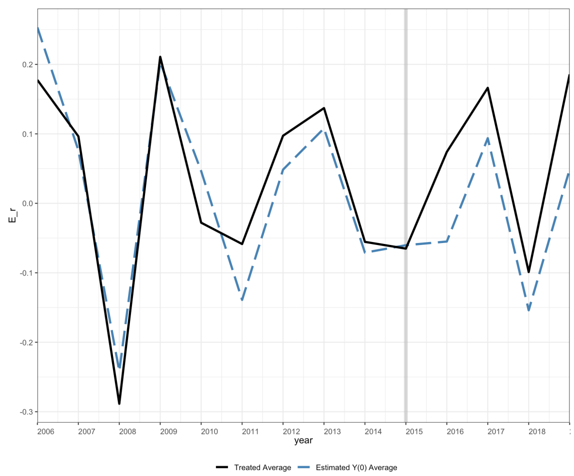
	Treated companies	Synthetic control group	Donor sample
$R_m - R_f$	.0898044	.0898044	.080003
Book-to-market	.7691699	.7803094	.8300189
Log size	11.19285	10.63189	9.099082
Log carbon intensity (CI)	6.703892	6.756034	6.842612
RMSPE	0.1953		

Figure 4b displays the gap between the predicted outcomes of the synthetic control group and the treated units. The gap between the predicted outcomes of the treated units and the synthetic control group before 2015 indicates the quality of the synthetic control group. A line close to 0 means the predictions are of good quality. Unfortunately, the gap before 2015 is *not* close to 0, indicating that the fit is actually not good. The gap after the PA indicates the average treatment effect on the treated companies (ATT): 0.0983. The positive sign indicates that the PA did have a positive effect on the returns of the 22 Climate Action 100+ companies.

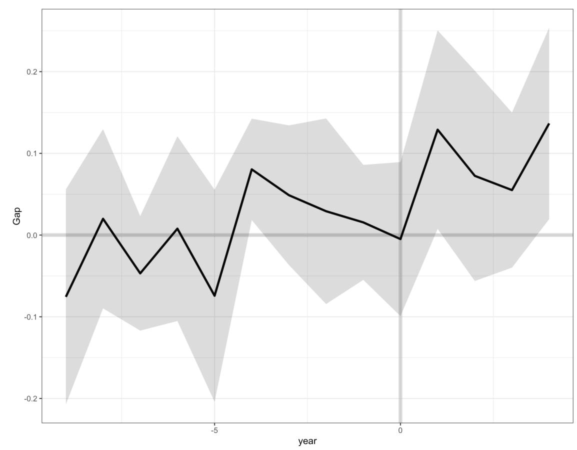
Figure 4c makes clear that the estimated ATTs are not extreme compared to the 102 placebo companies. Lastly, the p-values, which also indicate goodness-of-fit, indicate that the average treatment effect is significant, namely p = 0.0074 (at a 95% significance level). However, following Abadie et al. (2010), doing inference on permutation methods is superior to looking at the p-values. Looking at the placebo effects using multiple treated units indicates that the PA did not have significant effects on the returns of high-emitting companies. The DID also found a positive but insignificant sign; in that sense both DID and SCM arrive at the conclusion that this data set does not provide significant evidence that high-emitting companies started to gain higher yearly excess returns than low-emitting companies after the PA . We have to conclude that at least for this dataset the PA had a positive but insignificant effect on companies with high carbon intensity.

One big limitation in this model is that the donor pool consists of similar companies as the Climate Action 100+ treated companies in terms of carbon intensity. Therefore it is questionable if, apart from the investor's initiative to lower carbon emissions in the Climate Action 100+ companies, any differences remain between the treated units and the donor pool. A better donor pool would be companies not located in Europe, preferably companies located in countries that were not committed to the PA. In future research we intent to also include U.S. companies in the donor pool, and perform a robustness check similar to the one performed within the DID analysis.

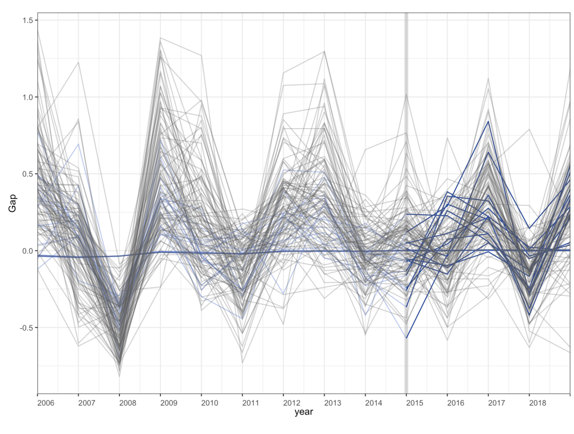
Figure 4: SCM graphical analysis



(a) Trends excess yearly returns



(b) Gap figure



(c) The treated units and 102 placebo gaps (grey lines)

Note: SCM results for multiple treated units. Figure a shows that the fit before 2015 is not perfect and figure b confirms this as the gap before 2015 is not close to 0. Lastly, figure c makes it clear that most treated units are not extreme relative to the placebo effects. The PA seems to have a positive but insignificant effect on the companies' yearly excess returns.

5.4 The Error Correction Model

In our final approach to estimation we focus on short-run and long-run dynamics of the total returns and carbon emission relationship. In 4.4 we established that total returns and unscaled carbon emissions are co-integrated with one another. This means that although the variables move independently over time, the average distance between them remains relatively constant. When this condition is met, we can estimate the ECM model given in equation 12 using Pesaran's Pooled Mean-Group Model. The pooled mean-group model constrains the long-run effects to be equal across all panels, whereas the short-run coefficients are allowed to vary across all panels. The short-run coefficient represents the current average relationship between CO_2 emissions and total returns. These short-run relationships may vary as different countries have different policies. The long-run effects affect all companies, as all companies in the European economy are moving to a low-carbon economy. Therefore, this specification may provide better estimates than the dynamic fixed effects estimation, where all parameters are constrained to be equal across panels. Table 18 shows the results for the long run model and the short run model averaged over all cross sections. In the long-run model, the coëffiënt for the carbon emission variable again gets a significant but negative value (p-value < 0.05), indicating that in the long run, higher carbon emissions will have a negative rather than a positive effect on a company's total returns.

Table 18: ECM results

	Coefficient	Standard Error	t-statistic	p-value
Long run relationship				
Log carbon emissions	-.5838821***	.1314492	-4.44	0.000
Short run relationship				
Δ Log carbon emissions	-.4966753	.5475755	-0.91	0.364
Error Correction term	-.3505087***	.0208201	-16.84	0.0000
Intercept	12.59243***	1.019924	12.35	0.000
Log likelihood	-8954.776			

Note: This table shows the estimated parameters of the ECM of yearly total returns as the dependent variable and carbon emissions as independent variable. *, **, *** denote significance on the 10%, 5%, and 1% level, respectively.

The estimated error correction coefficient $\hat{\lambda} = -0.3554$ and statistically significant. This value shows the short-run dynamics of $totalreturns_{i,t}$ and $ARET_{i,t}$ when these deviate from the long run equilibrium values. The error correction coefficient shows that the total returns adjust to about 35% of the imbalance in the previous period. The short-run dynamic between carbon emissions and total returns is again negative and statistically insignificant. The negative value is in line with the findings in Section 4.1, which implied that a company's carbon intensity had a negative but insignificant short-run effect on the annual excess returns of an individual stock i from year t to $t-1$. The difference in significance between the short run and the long run coefficients may be due to short run noise and volatility. In the long run, this noise may well be smoothed out over time which could explain why the long run coefficient is significant.

6 Conclusion

Governments worldwide still fail to set effective climate policies and meet decarbonisation targets. As a result, the window to limit global warming to 1.5 degree Celsius above pre-industrial levels is closing rapidly (IPCC (2023)). Market participants are slowly recognizing the financial risks associated with climate change, but it is unclear to what extent this growing awareness is reflected into prices of financial assets. This research has shed light on whether investors consider and price climate policy risk when investing in Europe's equity market, and specifically compared the period pre and post-PA in 2015. We assessed the effect of carbon intensity (tCO_2/\$M) on relative stock returns of clean versus polluting firms using a panel data set consisting of 1,555 European companies over the period 2005-2019.

We did not find empirical evidence that carbon risk is being priced in a diversified European equity portfolio, implying that investors do not seem to be aware of or at least do not require a risk premium for the risk they bear by investing in polluting companies. The Fama-French style panel regressions indicates no significant short-term relationship between stock returns and carbon intensity, and to the extent it finds such a relation at all, it is negative rather than positive. The long-short portfolio also generated negative cumulative returns over the period 2005 to 2019, and the carbon factor had no additional explanatory value over the three traditional Fama-French factors. The Fama-Macbeth procedure also provided some evidence for a negative (although admittedly) insignificant carbon premium. The ECM approach allowed for richer dynamics but also displayed a significant negative long-run relationship between CO_2 emissions and total returns. These findings are all counter intuitive to the idea that investors in European stocks require a premium for their exposure to climate policy risk. An explanation for the lack of apparent market pricing of carbon risk could be that investors are not sure how to capture and quantify climate policy risk.

We did find some preliminary indication that the Paris Agreement (PA) had a positive effect on stock returns of companies with high carbon intensity, which suggests at least a rising carbon risk premium. Two treatment effect models were employed to measure the effects of the PA on pricing of climate policy risk, again within Europe's equity market. Using three different classification methods for treated and non-treated companies, the DID model estimates for the PA effect were positive but insignificant. However, when we use U.S. companies as the control group, the DID estimates for the PA is positive and significantly so. The SCM estimates, where the weights for the control group were assigned objectively and driven by data instead of subjectively as in the DID, also found a positive treatment effect for companies with high carbon intensity. However, the placebo gaps indicated the effects on the treated companies were once again insignificant.

In conclusion, we do not find empirical evidence that investors are actively pricing climate policy risk: They do not seem to demand higher compensation for investing in polluting firms. However, we find some preliminary indication that the PA positively impacted polluting firms' stock returns, which could point to a gradual market pricing of climate policy risk. More research needs to be done to further explore this hypothesis. The absence of a priced-in carbon risk factor suggests that governments should signal more ambitious future climate policies more clearly if they want to see capital reallocated to lower emission activities by firms.

Similar to other research contributions in this field, our empirical analysis has its limitations related to the carbon data used, the time horizon under investigation as well as the portfolio construction process. First of all, there are a number of data issues. Firm-level data on carbon emissions is only being recorded and estimated yearly since 2005, making it difficult to assess the reliability of this data. Carbon data, even on scope 1 and 2 emissions, is noisy and not comparable across providers. Therefore, it may well be that carbon intensity is not the right metric to capture firm's climate policy risk exposure. Second, our research is based on a relatively short time period during which interest rates were historically low. This has led to increased investor demand for equities, feeding into higher stock prices. A longer time horizon is usually required to reduce the effect of cycles in financial market and accurately measure return driving risk factors. Third, the perception of risk is affected by high levels of uncertainty. The required pace to safely transition to a carbon neutral economy is unknown, which means it is unclear when climate policy risk are likely to materialize - in the short, medium or long run. It is furthermore difficult to isolate and thus price climate policy risk, as the risk is highly intertwined with physical risk. The longer it takes to install good climate policies, the higher the chances of an increase in extreme weather events and, thus, physical risk materializing. Finally, our long-short portfolios were constructed using equal weighting and we refrained from annual re-balancing in the Fama and Macbeth regressions as firms' ranking based on carbon intensity did not alter much over time.

We point out three avenues for future research. The first is methodological: the portfolios in the Fama and Macbeth regressions in our current setup are not rebalanced and not wieghted by market share. It is arguable that rebalancing and weighting by market shares would lead to more accurate results. The second is to use more forward-looking variables to assess a firm's climate policy risk exposure, as such variables would better capture the efforts a company is taking to bring down its emission levels. Recent research has indicated that the quality of disclosures and forward-looking emission reduction goals (Alessi (2021); Bolton and Kacperczyk (2021)) are relevant. Other interesting variables to gauge climate policy risk that are not necessarily forward-looking could be carbon prices in emissions trading schemes or the effective level of carbon taxes. The third possible extension is to further explore the idea of markets pricing in climate policy risk only gradually. Gradual price adjustment confounds attempts to isolate a risk premium since actual price changes are unavoidably used to incorporate capital gains and thus cause a downward bias in the data if higher risk premia are produced by a preceding fall in prices upfront rather than higher returns.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132.

Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834.

Alessi, L., O. E. . P. R. (2021). What greenium matters in the stock market? the role of greenhouse gas emissions and environmental disclosures. *journal of financial stability*, 54(100869).

Ardia, David, K. B. K. B. and Inghelbrecht, K. (2022). Climate change concerns and the performance of green versus brown stocks. *Management Science*.

Aswani, J., R. A. . R. S. (2023). Are carbon emissions associated with stock returns. *Oxford University Press*.

Barnett, M., Brock, W., and Hansen, L. P. (2020). Pricing uncertainty induced by climate change. *The Review of Financial Studies*, 33(3):1024–1066.

Batten, S., Sowerbutts, R., and Tanaka, M. (2016). Let's talk about the weather: the impact of climate change on central banks.

Bauer, M. Huber, D. R. G. and Wilms, O. (2023). Where is the carbon premium? global performance of green and brown stocks. *Working paper*, (83).

Bebchuk, L.A., A. and Wang, C. (2013). Learning and the disappearing association between governance and returns. *Journal of Financial Economics*, 2(108):323–348.

Bolton, P. and Kacperczyk, M. (2019). Do investors care about carbon risk? *Journal of financial economics*, 142(2):517–549.

Bolton, P. and Kacperczyk, M. (2021). Global pricing of carbon-transition risk. Technical report, National Bureau of Economic Research.

Bolton, P. and Kacperczyk, M. T. (2020). Carbon premium around the world.

Bun, M. J., Sarafidis, V., et al. (2015). Dynamic panel data models. *The Oxford handbook of panel data*, pages 76–110.

Campiglio, E., Monnin, P., and von Jagow, A. (2019). Climate risks in financial assets. *Council on Economic Policies, Discussion.*

Campiglio, Daumas, M. v. J. (2022). climate-related risks in financial assets. *journal of economic survey.*

Cheema-Fox, A., L. B. R. S. G. T. D. . W. H. (2019). Decarbonization factors. *Harvard Business School Working Paper*, 20-037.

Choi, J. J., Jo, H., and Park, H. (2018). Co2 emissions and the pricing of climate risk. *Social Science Research Network: Rochester, NY, USA.*

Cochrane, J. (2009). *Asset pricing: Revised edition.* Princeton university press.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.

Faccini, R., M. R. . S. G. Dissecting climate risks: Are they reflected in stock prices? *SSRN Electronic Journal*.

Fama, E. F. and French, K. R. (1988). Permanent and temporary components of stock prices. *Journal of political Economy*, 96(2):246–273.

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

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57(5):1121–1152.

Giglio, S., Kelly, B., and Stroebel, J. (2021). Climate finance. *Annual Review of Financial Economics*, 13:15–36.

Gimeno, R. and Gonzalez, C. I. (2022). The role of a green factor in stock prices. when fama & french go green.

Görgen, M., Jacob, A., Nerlinger, M., Riordan, R., Rohleder, M., and Wilkens, M. (2020). Carbon risk. *Available at SSRN 2930897*.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *The Econometrics Journal*, 3(2):148–161.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, pages 1251–1271.

Hengge, Panizza, V. (2023). carbon policy and stock returns. Technical report.

Hong, H. and Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of financial economics*, 93(1):15–36.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Hsu, P.-H., Li, K., and Tsou, C.-Y. (2022). The pollution premium. *Journal of Finance, Forthcoming*.

Huij, Joop, D. L. P. A. S. and Zwinkels, R. C. J. (2021). Carbon beta: A market-based measure of climate risk.

Hultman, N. E., Hassenzahl, D. M., Rayner, S., et al. (2010). Climate risk. *Annual Review of Environment and Resources*, 35(1):283–303.

Ilhan, E., Sautner, Z., and Vilkov, G. (2021). Carbon tail risk. *The Review of Financial Studies*, 34(3):1540–1571.

In, S., P. K. and Monk, A. (2019). Is 'being green' rewarded in the market? an empirical investigation of decarbonization and stock returns. *working paper.*

IPCC (2023). Ar6 synthesis report: Climate change 2023. Technical report.

Janssen, A., Dijk, J., Duijm, P., et al. (2021). Misleading footprints. inflation and exchange rate effects in relative carbon disclosure metrics. Technical report, DNB.

Klaassen, S. (2021). harmonizing corporate carbon footprints. *Nature communication*, 12(6149).

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Nguyen, J H, T. C. . Z. B. (2020). The price of carbon risk: Evidence from the kyoto protocol ratification. *SSRN Electronic Journal (September).*

Noailly, J., N. L. . v. D. H. M. (2021). Heard the news? environmental policy and clean investments. *Centre for International Environmental Studies, The Graduate Institute.*, 70-2021.

Park, S. Y., In, K. Y., and Monk, A. (2017). Is "being green" rewarded in the market? an empirical investigation of decarbonization risk and stock returns. *International Association for Energy Economics (Singapore Issue)*, 46(48).

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics*, 146(2):403–424.

Pastor, L., S. R. and Taylor, L. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142:550–571.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: The esg-efficient frontier. *Journal of Financial Economics*, 142(2):572–597.

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric theory*, 20(3):597–625.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of applied econometrics*, 22(2):265–312.

Ramiah, V., M. T. M. I. G. M. . P. L. The effects of announcement of green policies on equity portfolios: Evidence from the united kingdom. *Managerial Auditing Journal*, 31(2):138–155.

Rohleder, M., Wilkens, M., and Zink, J. (2022). The effects of mutual fund decarbonization on stock prices and carbon emissions. *Journal of Banking & Finance*, 134:106352.

Soh, B., I. Y. P. K. Y. . M. A. (2017). Is "being green" rewarded in the market? an empirical investigation of decarbonization risk and stock returns. i. *International Association for Energy Economics (Singapore Issue)*, 46(48).

TCFD (2017). Recommendations of the task force on climate-related financial disclosures. Technical report.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data.

Zeng, J., Zhao, R., and Dai, W. (2021). Extending synthetic control method for multiple treated units: an application to environmental intervention. *Economic Research-Ekonomska Istraživanja*, 34(1):311–330.

A Figures and tables

Figure 5: GICS sector descriptions

GICS Sector	Description
Energy	The Energy Sector comprises companies engaged in exploration & production, refining & marketing, and storage & transportation of oil & gas and coal & consumable fuels. It also includes companies that offer oil & gas equipment and service.
Materials	The Materials Sector includes companies that manufacture chemicals, construction materials, glass, paper, forest products and related packaging products, and metals, minerals and mining companies, including producers of steel.
Industrials	The Industrials Sector includes manufacturers and distributors of capital goods such as aerospace & defense, building products, electrical equipment and machinery and companies that offer construction & engineering services. It also includes providers of commercial & professional services including printing, environmental and facilities services, office services & supplies, security & alarm services, human resource & employment services, research & consulting services. It also includes companies that provide transportation services.
Consumer Discretionary	The Consumer Discretionary Sector encompasses those businesses that tend to be the most sensitive to economic cycles. Its manufacturing segment includes automotive, household durable goods, leisure equipment and textiles & apparel. The services segment includes hotels, restaurants and other leisure facilities, media production and services, and consumer retailing and services
Consumer Staples	The Consumer Staples Sector comprises companies whose businesses are less sensitive to economic cycles. It includes manufacturers and distributors of food, beverages and tobacco and producers of non-durable household goods and personal products. It also includes food & drug retailing companies as well as hypermarkets and consumer super centers.
Health Care	The Health Care Sector includes health care providers & services, companies that manufacture and distribute health care equipment & supplies, and health care technology companies. It also includes companies involved in the research, development, production and marketing of pharmaceuticals and biotechnology products.
Financials	The Financials Sector contains companies involved in banking, thrifts & mortgage finance, specialized finance, consumer finance, asset management and custody banks, investment banking and brokerage and insurance. It also includes Financial Exchanges & Data and Mortgage REITs.
Information Technology	The Information Technology Sector comprises companies that offer software and information technology services, manufacturers and distributors of technology hardware & equipment such as communications equipment, cellular phones, computers & peripherals, electronic equipment and related instruments, and semiconductors.
Communication Services	The Communication Services Sector includes companies that facilitate communication and offer related content and information through various mediums. It includes telecom and media & entertainment companies including producers of interactive gaming products and companies engaged in content and information creation or distribution through proprietary platforms.
Utilities	The Utilities Sector comprises utility companies such as electric, gas and water utilities. It also includes independent power producers & energy traders and companies that engage in generation and distribution of electricity using renewable sources.
Real Estate	The Real Estate Sector contains companies engaged in real estate development and operation. It also includes companies offering real estate related services and Equity Real Estate Investment Trusts (REITs).

Note: This table also shows the sector "Financials". This sector is excluded from the data set as this sector consists of the industries "banks", "diversified financials", and "insurance companies". For banks and insurance companies, emissions are challenging to measure as these are linked with the composition of their loan books and thus, not a reflection of their operational activities. Source: MSCI.

Table 19: General structure panel data set

Company (ISIN) i	Time t	Variable y	Variable x_1	Variable x_2	Variable x_M
Company 1	t_1	$y_{1,1}$	$x_{1,1,1}$	$x_{2,1,1}$	$x_{M,1,1}$
...	...	...	...	...	...
Company 1	t = T	$y_{1,T}$	$x_{1,1,T}$	$x_{2,1,T}$	$x_{M,1,T}$
Company 2	t_1	$y_{2,1}$	$x_{1,2,1}$	$x_{2,2,1}$	$x_{M,2,1}$
...	...	...	...	...	...
Company 2	t = T	$y_{2,T}$	$x_{1,2,T}$	$x_{2,2,T}$	$x_{M,2,T}$
Company 3	t_1	$y_{3,1}$	$x_{1,3,1}$	$x_{2,3,1}$	$x_{M,3,1}$
...	...	...	...	...	...
Company 3	t = T	$y_{3,T}$	$x_{1,3,T}$	$x_{2,3,T}$	$x_{M,3,T}$
Company N	t_1	$y_{N,1}$	$x_{1,N,1}$	$x_{2,N,1}$	$x_{M,N,1}$
...	...	...	...	...	...
Company N	t = T	$y_{N,T}$	$x_{1,N,T}$	$x_{2,N,T}$	$x_{M,N,T}$

Note: Here there are $N = 1, ..., 1,555$ companies in the Trucost data set, $T = 1, ..., 15$ years and $m = 1, ..., M$ different variables.

Table 20: Average GDP growth and inflation between 2005 and 2019

Country code	Average GDP growth	Average Inflation
AT	1.50%	1,88%
BE	1.54%	1,93%
BG	2.93%	3,46%
CH	2.11%	1,49%
DE	1.50%	3,25%
DK	1.34%	1,44%
EE	2.90%	1,26%
ES	1.26%	1,44%
FI	1.11%	1,44%
FR	1.2%	3,36%
GB	1.54%	1,08%
GR	-1.06%	1,45%
HU	2.04%	3,06%
IE	4.61%	1,84%
IT	0.040%	1,84%
LT	3.33%	1,61%
LU	2.61%	2,10%
MT	4.83%	1,41%
NL	1.48%	4,11%
NO	1.42%	1,829%
PT	0.720%	1,71%
RO	3.65%	1,2%
SE	2.10%	2,20%
SI	2.11%	0,34%

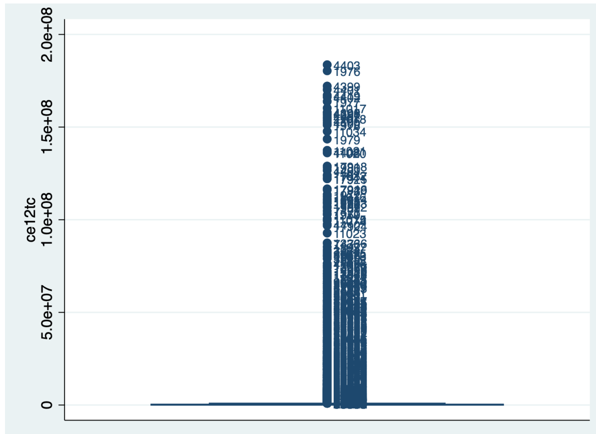
Note: Ireland has the largest GDP growth and Italy the lowest. In terms of inflation, most countries lie between 1% and 4%. Source: https://databank.worldbank.org

Table 21: Correlations between Trucost and MSCI per sector

		Estimated				Reported			
		CE12TC	CE12MSCI	CI12TC	CI12MSCI	CE12TC	CE12MSCI	CI12TC	CI12MSCI
Communication Services	CE12TC	1.0000				1.0000			
	CE12MSCI	0.6785	1.0000			0.8752	1.0000		
	CI12TC	0.7929	0.6135	1.0000		0.4185	0.3592	1.0000	
	CI12MSCI	0.1397	0.3398	0.4936	1.0000	0.1818	0.2423	0.6440	1.0000
Consumer Discretionary	CE12TC	1.0000				1.0000			
	CE12MSCI	0.2351	1.0000			0.6140	1.0000		
	CI12TC	0.2638	0.0731	1.0000		0.2780	0.4108	1.0000	
	CI12MSCI	0.0591	0.9137	0.1234	1.0000	0.2937	0.5008	0.9233	1.0000
Consumer Staples	CE12TC	1.0000				1.0000			
	CE12MSCI	0.1323	1.0000			0.3151	1.0000		
	CI12TC	0.4472	-0.1725	1.0000		0.0479	0.1628	1.0000	
	CI12MSCI	0.1277	0.1207	0.6154	1.0000	0.0616	0.2563	0.6455	1.0000
Energy	CE12TC	1.0000				1.0000			
	CE12MSCI	-0.0290	1.0000			0.9685	1.0000		
	CI12TC	-0.0421	0.3588	1.0000		-0.1557	-0.1605	1.0000	
	CI12MSCI	-0.3522	0.8898	0.6046	1.0000	-0.1891	-0.1908	0.9867	1.0000
Health Care	CE12TC	1.0000				1.0000			
	CE12MSCI	0.9393	1.0000			0.6631	1.0000		
	CI12TC	-0.0336	-0.0670	1.0000		0.0760	0.1273	1.0000	
	CI12MSCI	-0.0091	0.0806	0.1287	1.0000	0.0808	0.1212	0.9839	1.0000
Industrials	CE12TC	1.0000				1.0000			
	CE12MSCI	0.8513	1.0000			0.6972	1.0000		
	CI12TC	0.5036	0.5528	1.0000		0.4277	0.5439	1.0000	
	CI12MSCI	0.2067	0.2493	0.7645	1.0000	0.3705	0.5352	0.9148	1.0000
Information Technology	CE12TC	1.0000				1.0000			
	CE12MSCI	0.8913	1.0000			0.0624	1.0000		
	CI12TC	0.0165	-0.0377	1.0000		0.0109	0.3974	1.0000	
	CI12MSCI	-0.0939	0.0350	0.5045	1.0000	0.0023	0.4167	0.8689	1.0000
Materials	CE12TC	1.0000				1.0000			
	CE12MSCI	0.9425	1.0000			0.9369	1.0000		
	CI12TC	0.2495	0.3276	1.0000		0.3241	0.3475	1.0000	
	CI12MSCI	0.1682	0.3585	0.8551	1.0000	0.3055	0.3336	0.9506	1.0000
Real Estate	CE12TC	1.0000				1.0000			
	CE12MSCI	0.1887	1.0000			0.0844	1.0000		
	CI12TC	-0.0296	0.2131	1.0000		0.1973	0.0199	1.0000	
	CI12MSCI	0.0874	0.3962	0.7492	1.0000	0.1451	0.4806	0.7462	1.0000
Utilities	CE12TC	1.0000				1.0000			
	CE12MSCI	0.2034	1.0000			0.8614	1.0000		
	CI12TC	0.5560	0.1479	1.0000		0.1467	0.1795	1.0000	
	CI12MSCI	0.0503	0.8639	0.3763	1.0000	0.0742	0.3402	0.7455	1.0000

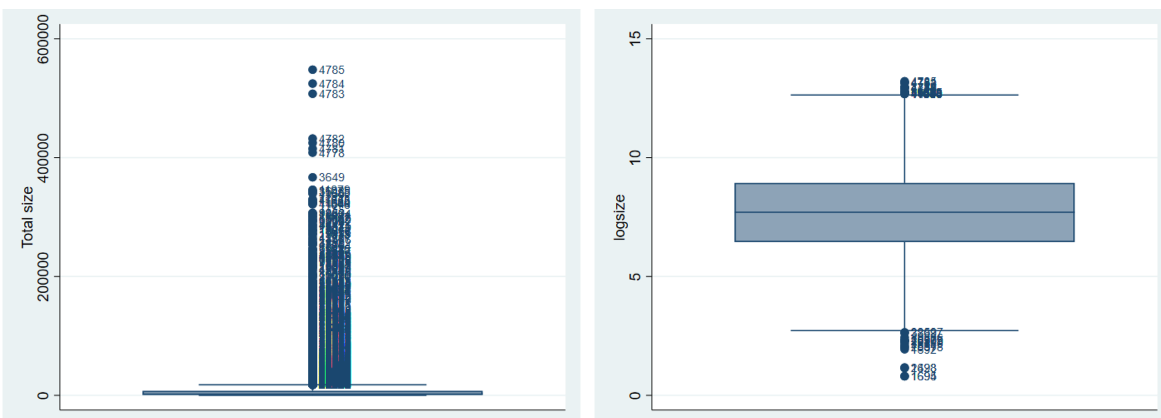
Note: For the estimated carbon emissions, the sectors Health Care and Materials had the highest correlation for carbon emissions (0.94 and 0.94). For the reported carbon emissions, the sectors Energy and Materials had the highest correlation for carbon emissions (0.97 and 0.94).

Figure 6: Boxplot $CE12TC_{i,t}$



Note: Since there are 1,555 companies included in the data set, the carbon emissions vary substantially between different companies. Therefore, no companies were deleted based on their carbon emissions.

Figure 7: Boxplots $totalsize_{i,t}$ and $logsize_{i,t}$

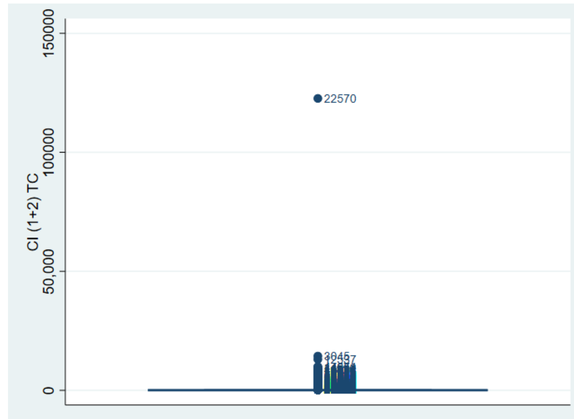


(a) Boxplot $totalsize_{i,t}$ **(b) Boxplot $logsize_{i,t}$**

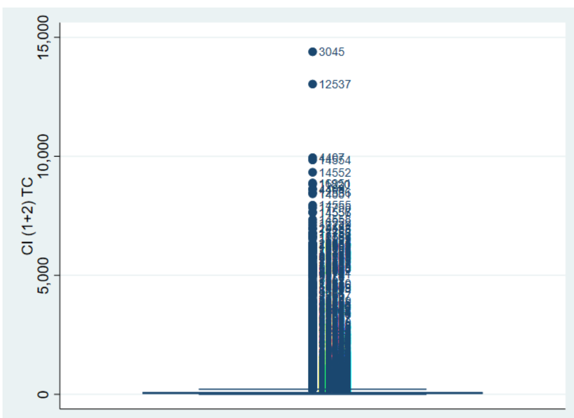
Note: Since there are 1,555 companies included in the data set, the size varies substantially between different companies. Therefore, no companies were deleted based on their size.

Figure 8: Box plots $CI12TC_{i,t}$

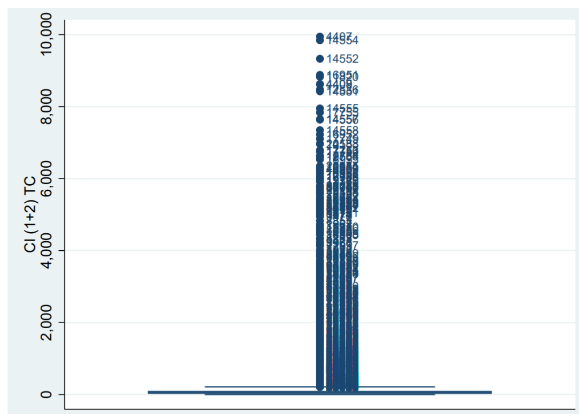
(a) First boxplot



(b) Second boxplot, first outlier deleted



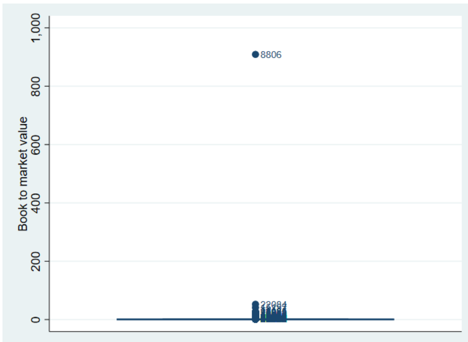
(c) Third boxplot, two more outliers deleted



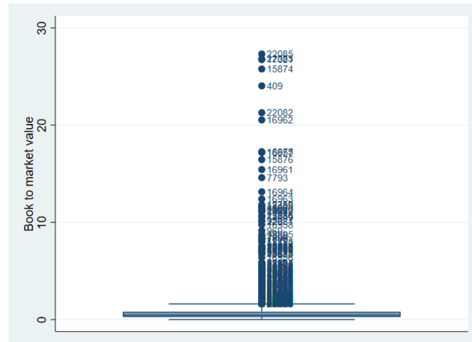
(d) Fourth boxplot, no outliers

Note: If the carbon intensity of a company was > 10,000 tCO2e/$1M, the observation was classified as an outlier and removed from the data set.

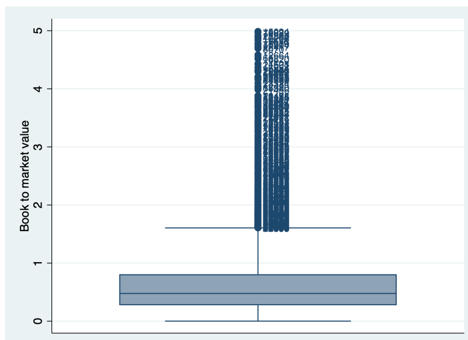
Figure 9: Boxplots $BTM_{i,t}$



(a) Boxplot with outlier



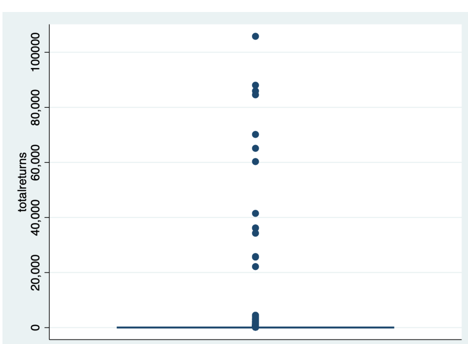
(b) Boxplot with BTM values > 5



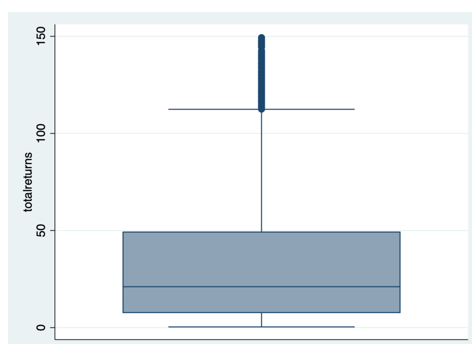
(c) Boxplot without outliers

Note: Since book-to-market values higher than 5 are unrealistic and probably indicate bankruptcy, these values are deleted.

Figure 10: Boxplots $totalreturns_{i,t}$.

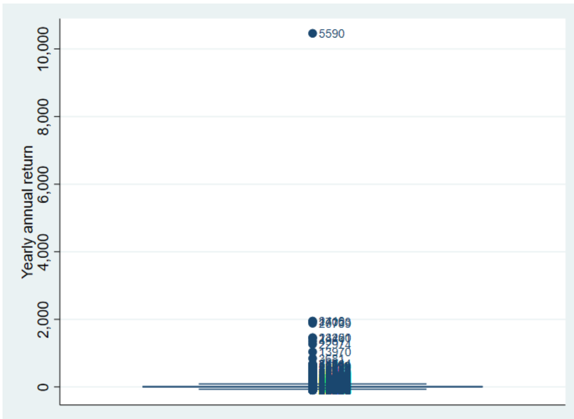


(a) Boxplot with outliers

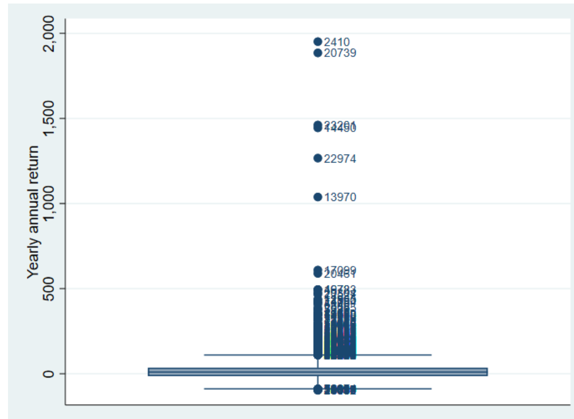


(b) Boxplot with no outliers

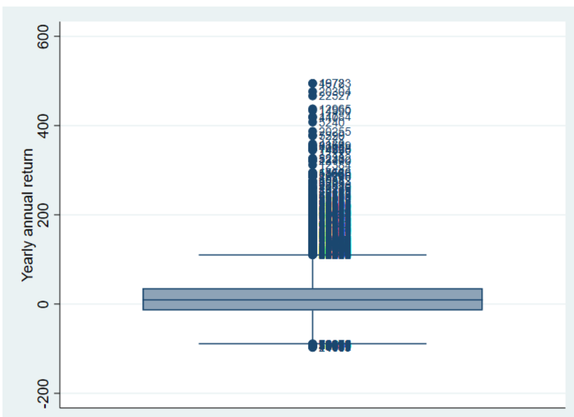
Figure 11: Boxplots $ARET_{i,t}$



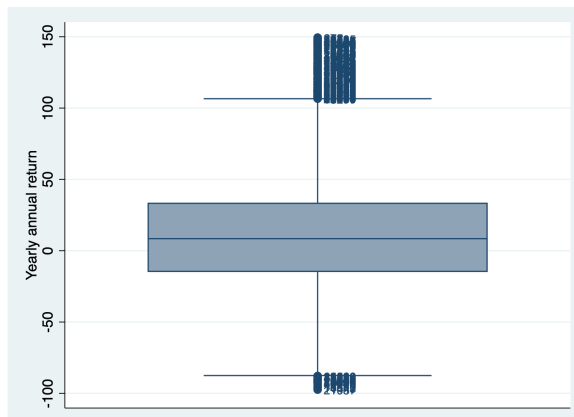
(a) Boxplot with one extreme outlier



(b) Boxplot with eight more outliers



(c) Boxplot with values > 150%

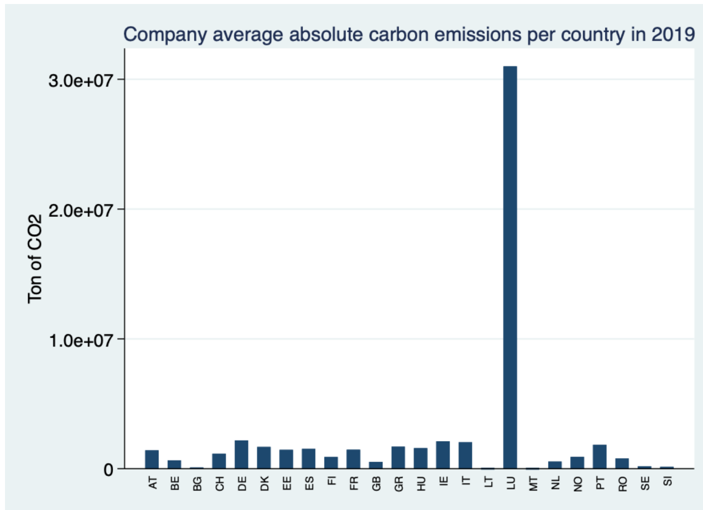


(d) Boxplot with no outliers

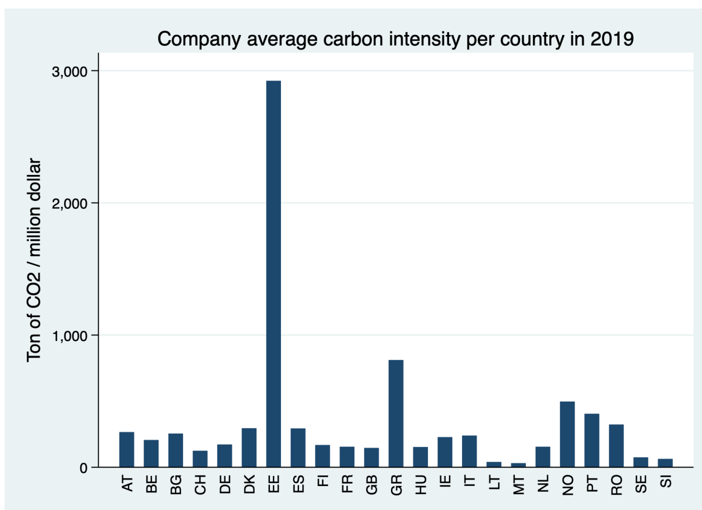
Note: Annual returns above 150 % are considered unrealistic, so these observations are deleted.

Data analysis: geography

Figure 12: Average carbon emissions and carbon intensities per country in 2019



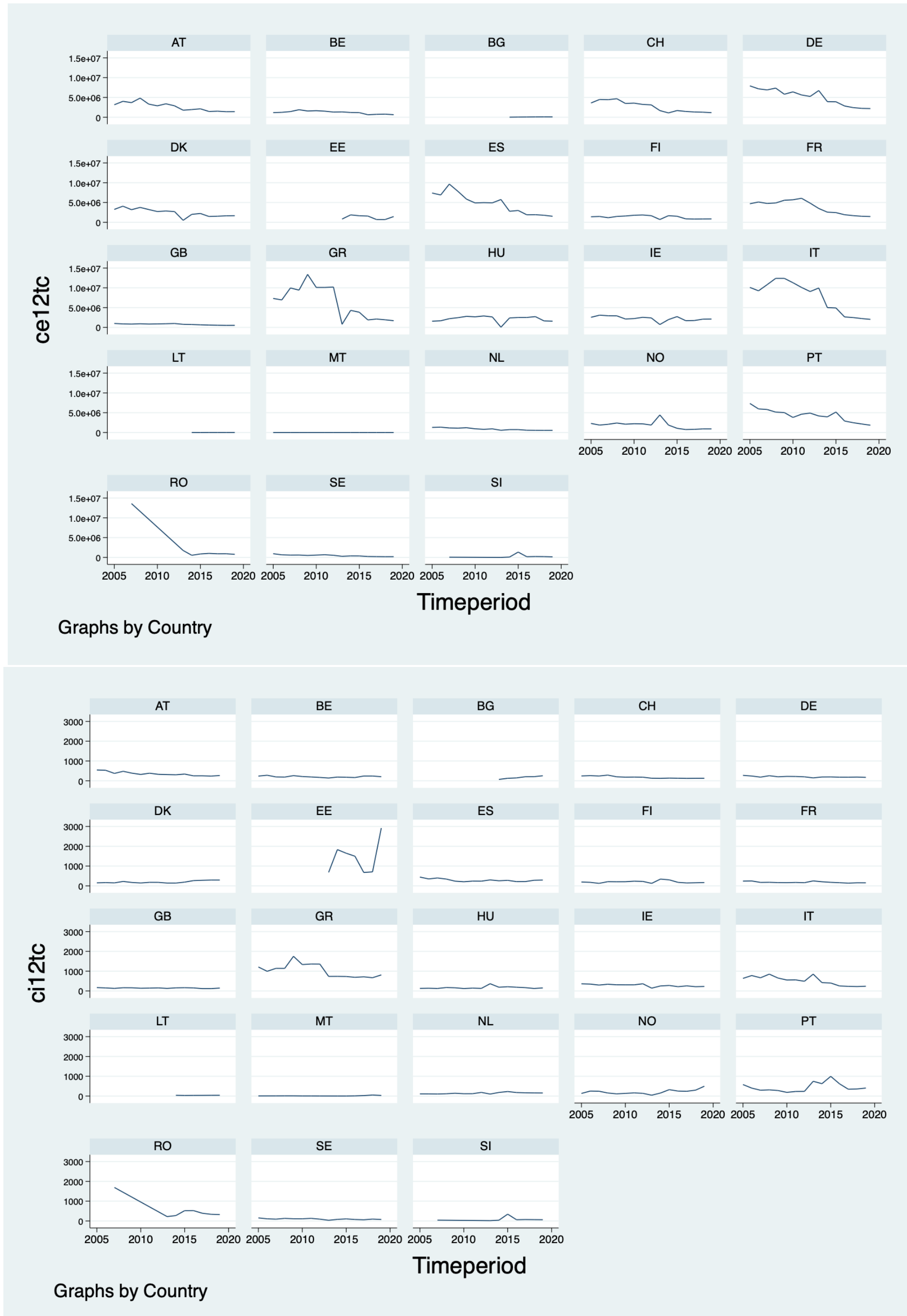
(a) Average carbon emissions per country in 2019



(b) Average carbon intensity per country in 2019

Note: On average, companies in Luxembourg, Germany, and Ireland emit the most with respectively 31.0, 2.2, and 2.1 million tons of CO2. The three lowest contributors in carbon emission are Bulgaria, Lithuania, and Malta, where companies, on average, emit 92.5, 10.0, and 3.6 thousand tons of CO2, respectively. The three biggest emitters in carbon intensities (tCO2e/$M) are Estonia, Greece, and Luxembourg, with respectively 2,923.51, 652.83, and 628.92 tCO2e per $ million revenue. It must be noted that there are only two Estonian companies in the data set, so this might not be representative. The three lowest contributors in terms of carbon intensity are Malta, Lithuania, and Slovenia, with respectively 31.17, 40.05, and 62.92 tCO2e/$M.

Figure 13: Average carbon emissions and intensities per country per year



Note: The top figure shows that the average company emitted 50% less in 2019 compared with 2005. The figure also shows the changes in carbon intensities. Here the changes are not as significant as for carbon emissions.

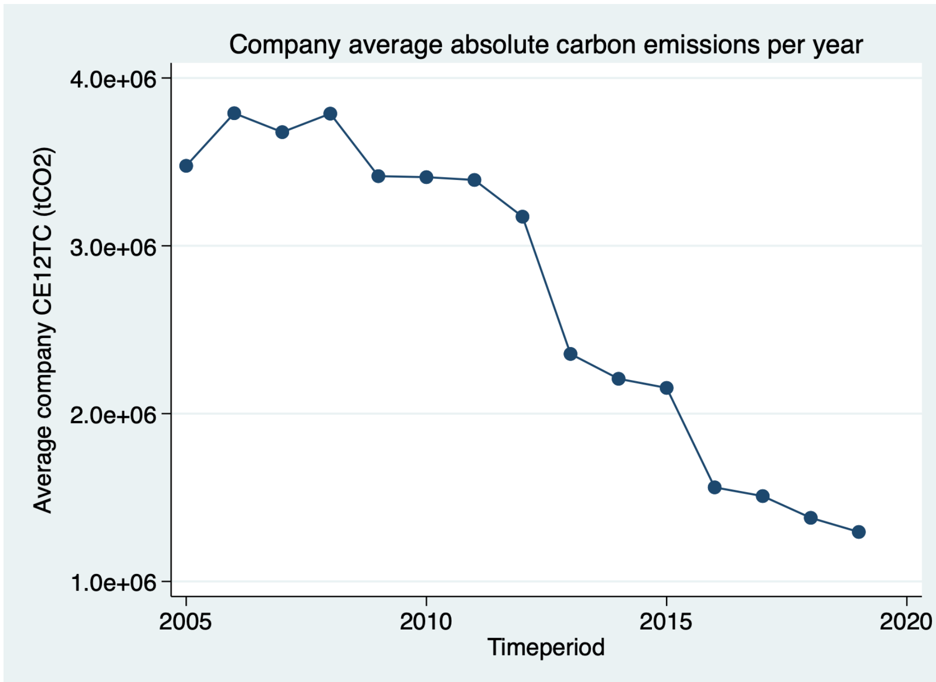
Table 22: Average company size, book-to-market value and yearly returns per country in 2019

Country	mean $totalsize_{i,t}$	std. dev $totalsize_{i,t}$	mean $BTM_{i,t}$	std. dev $BTM_{i,t}$	mean $ARET_{i,t}$	std. dev $ARET_{i,t}$
AT	5951.85	8539.059	.732115	.3517595	19.0401	20.61178
BE	8156.45	32980.71	.797551	1.283386	16.534	30.45623
BG	2210.04	3018.979	1.97644	1.585724	-13.9022	8.620061
CH	4782.88	13005.4	.554957	.4642487	26.1465	33.26106
DE	16488.8	53320.72	.616821	.4565464	27.8359	52.30999
DK	4988.26	10300.97	.492487	.4481944	23.3041	48.73097
EE	1008.26	1008.479	.876088	.5408052	16.6756	15.1978
ES	11597.1	23856.08	.710894	1.000346	12.2656	28.12116
FI	4517.76	7802.51	.63168	.7224842	24.1559	33.64861
FR	16027	40405.43	.692072	.4961551	16.6831	34.59245
GB	6781.13	23892.83	.586726	.6201595	31.9312	42.7181
GR	3053.6	3305.37	1.15877	1.137674	41.0983	44.2102
HU	5348.75	6897.317	.802312	.4781394	15.4247	50.17076
IE	6280.67	12654.79	.643896	.5562566	18.298	38.82343
IT	9271.32	27594.06	.632335	.4332393	22.8466	33.64043
LT	430.579	366.2976	.471744	.0425499	30.103	13.06813
LU	20147.4	33595.4	.838812	.7078817	-.868145	26.02488
MT	423.326	436.7817	.336674	.2213196	-46.1717	19.56671
NL	7635.91	13915.08	.469953	.3736205	28.8676	51.80473
NO	5427.08	16254.85	.982457	1.730591	16.726	53.49413
PT	7658.21	11856.9	.833858	.9158992	14.7539	17.74365
RO	3282.64	4449.162	.877011	.5481979	23.15	24.8816
SE	3103.83	6454.23	.481597	.531178	32.3236	45.88061
SI	1981.7	532.6589	1.12879	.4877545	17.6077	17.74626

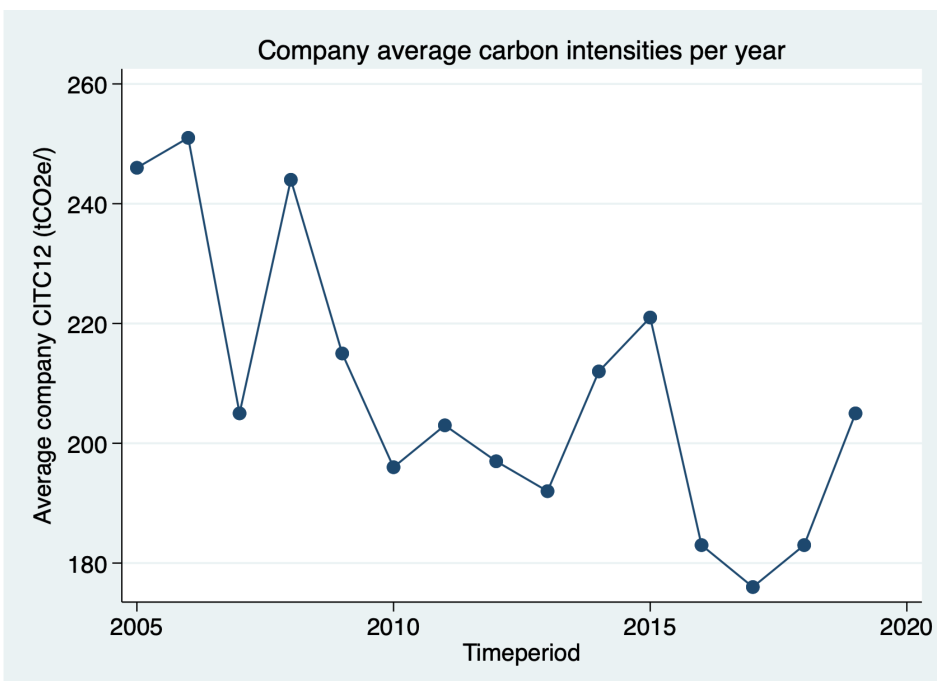
Note: In terms of size, companies in Luxembourg are the biggest with an average size of 20,147.4 assets, followed by German and French companies. Estonian, Lithuanian and Maltese companies are, on average the smallest. Regarding the book-to-market value, Bulgaria, Greece, and Slovenia are the only companies with a book-to-market value greater than 1 in 2019, with 1.98, 1.16, and 1.13. Companies in these countries might be underpriced, as these values are above 1. Stock prices in all other European companies included in the dataset are slightly overpriced as their book-to-market values are below 1. Lithuanian, Dutch, and Maltese companies had the lowest book-to-market values. Greek, Swedish, and Great-British companies have, on average, the highest total annual return in 2019. In terms of standard deviation, Norwegian companies have a standard deviation of 53.5%, which implies that the Norwegian market is very volatile. Norway has many oil companies and oil prices are volatile, this could explain the high standard deviation. Luxembourgian, Bulgarian, and Maltese companies have the lowest annual returns.

Data Analysis: time

Figure 14: Average carbon emissions and carbon intensities per year



(a) Average carbon emissions per year



(b) Average carbon intensity per year

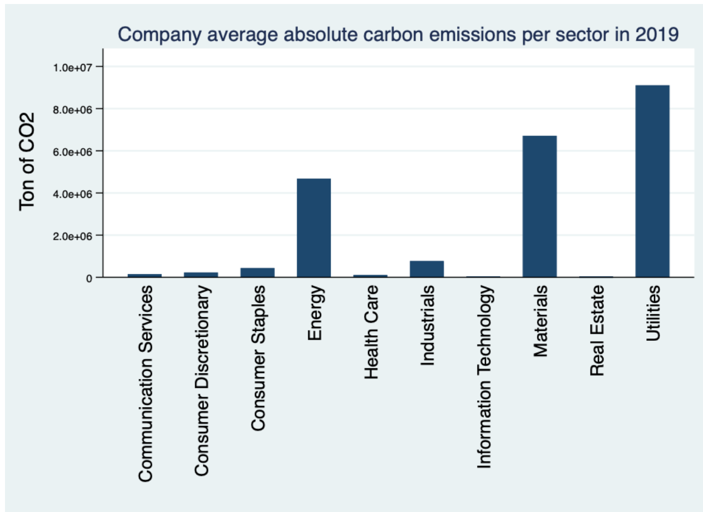
Table 23: Average company size, book-to-market value and yearly returns per year

Year	mean $totalsize_{i,t}$	std. dev $totalsize_{i,t}$	mean $BTM_{i,t}$	std. dev $BTM_{i,t}$	mean $ARET_{i,t}$	std. dev $ARET_{i,t}$
2005	12571.7	29966.14	.475151	.5612617	15.4329	42.67393
2006	13975.3	32623.83	.44599	.9674085	49.2296	37.3796
2007	15139.3	34962.34	.54175	1.195392	9.80396	38.78224
2008	16228.2	36593.7	1.17007	1.910569	-48.8698	20.5556
2009	16076.2	36143.2	.748561	1.439945	69.1752	97.80432
2010	15742.3	36009.63	.663794	.936807	22.5173	40.16109
2011	15459.6	36498.47	.875154	1.053487	-13.1957	25.49405
2012	15361.7	37547.19	.763418	.9401501	29.9382	32.75932
2013	12281	32204.94	.579451	.5491263	41.4269	44.69524
2014	11196.3	31481.38	.662625	.8116125	-2.35499	67.61832
2015	10175	29018.02	.648154	1.131744	9.89869	55.98137
2016	8001.5	26206.96	.62671	.8667226	6.50708	37.56929
2017	8598.4	28711.79	.547231	.634106	38.3715	50.8005
2018	8662.73	28826.94	.744665	.8957574	-17.6182	33.51386
2019	9087.85	30022.32	.639686	.6995153	24.5917	41.73769

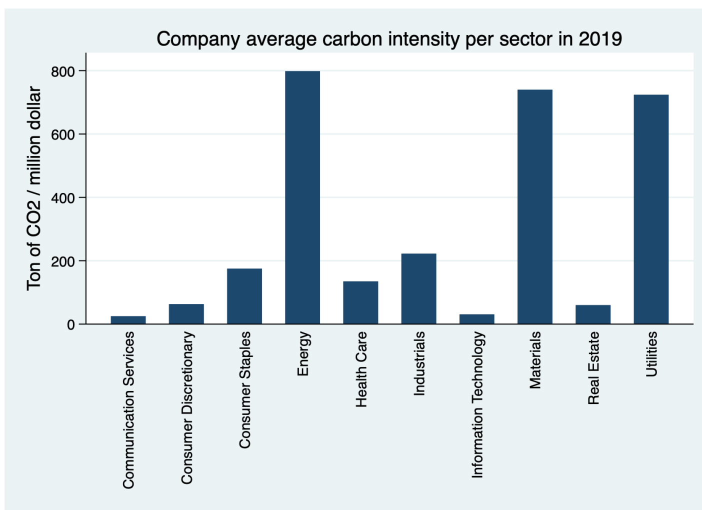
Note: In terms of size, European companies have shrunk from 2005 to 2019. In 2005, the average European company had 12,571.7 assets while the average European company in 2019 had 9,087.85 assets on its balance sheets. The number of assets grew from 2005 to 2009 and then started to shrink, possibly due to the financial crisis in 2008. From 2016 onwards, the number of assets is growing again. In contrast with the total size, the average book-to-market value did not increase over the years. In most years, companies are overpriced as their book-to-market values are below 1. Only in 2008, the book-to-market value was more significant than 1. After 2010, the average European company is slightly overpriced each year. The mean book-to-market value between 2005 and 2019 is 0.67. The highest annual returns were achieved in 2009, one year after the financial crisis, probably due to the low stock prices in 2008. Here the average total annual return was 69.2%. The annual return was lower in 2005 (15.4%) than in 2019 (24.6%).

Data analysis: sectors

Figure 15: Average carbon emissions and carbon intensities per sector



(a) Company average carbon emissions per sector in 2019



(b) Company average carbon intensity per sector in 2019

Note: On average, companies in the Utilities, Materials, and Energy sectors emitted 9.0, 6.5, and 4.82 million tons of CO2 in 2019. The three lowest contributing sectors in terms of carbon emission in 2019 were the Health Care, Information Technology, and Real Estate sector, where companies, on average, emitted 107.4, 36.7, and 21.9 thousand tons of CO2, respectively. The figure also shows that the three biggest emitters in terms of carbon intensities (tCO2e/$M revenue) are again Energy, Materials, and Utilities (but in a different order) with a respective average carbon intensity of 801.6, 739.9, and 724.03 tCO2e per $ million revenue. The Communication Services, Information Technology, and Real Estate sectors are the three lowest carbon intense contributors. Companies in these sectors had an average carbon intensity in 2019 of 25.1, 30.7, and 58.3 tCO2e per $ million revenue.

Figure 16: Company average carbon emissions per sector per year

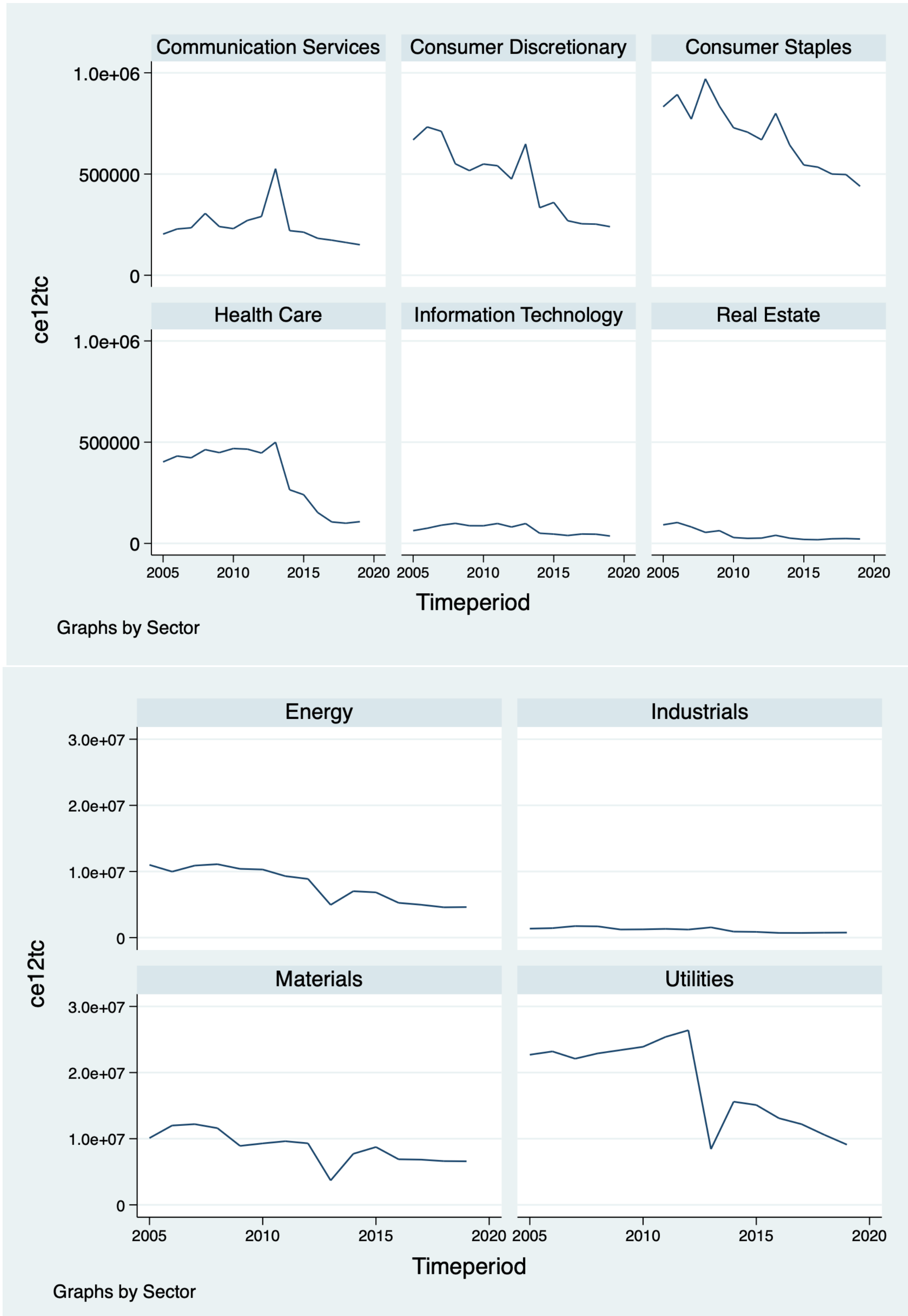
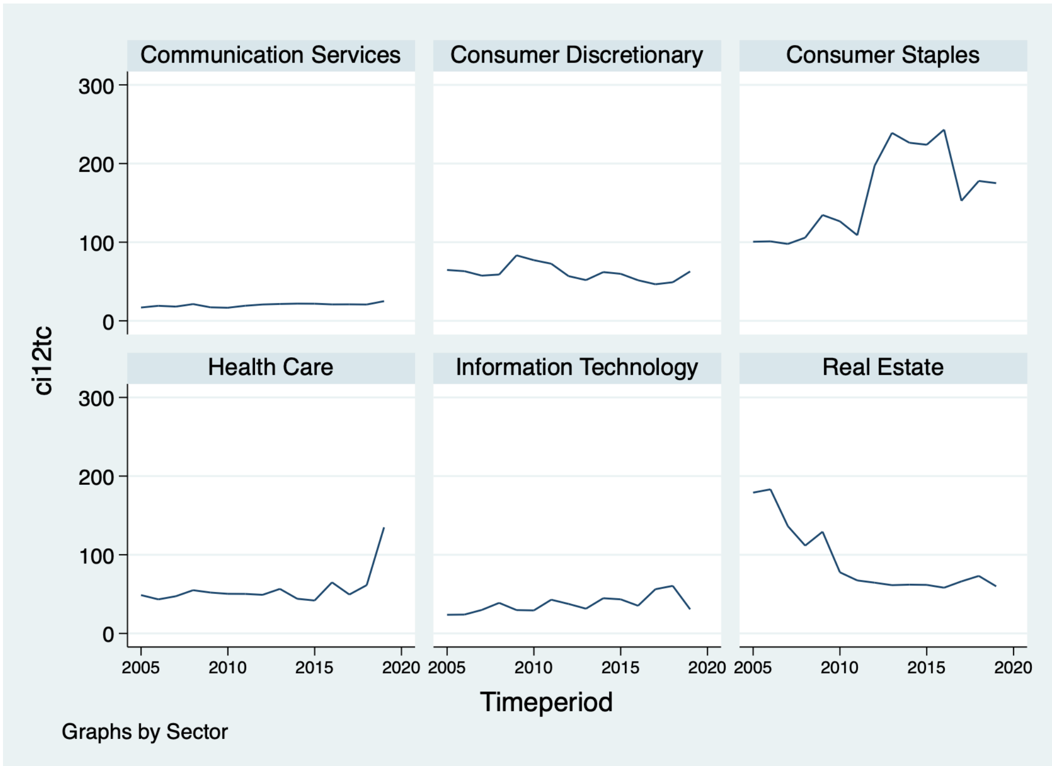


Figure 17: Company average carbon intensity per sector per year



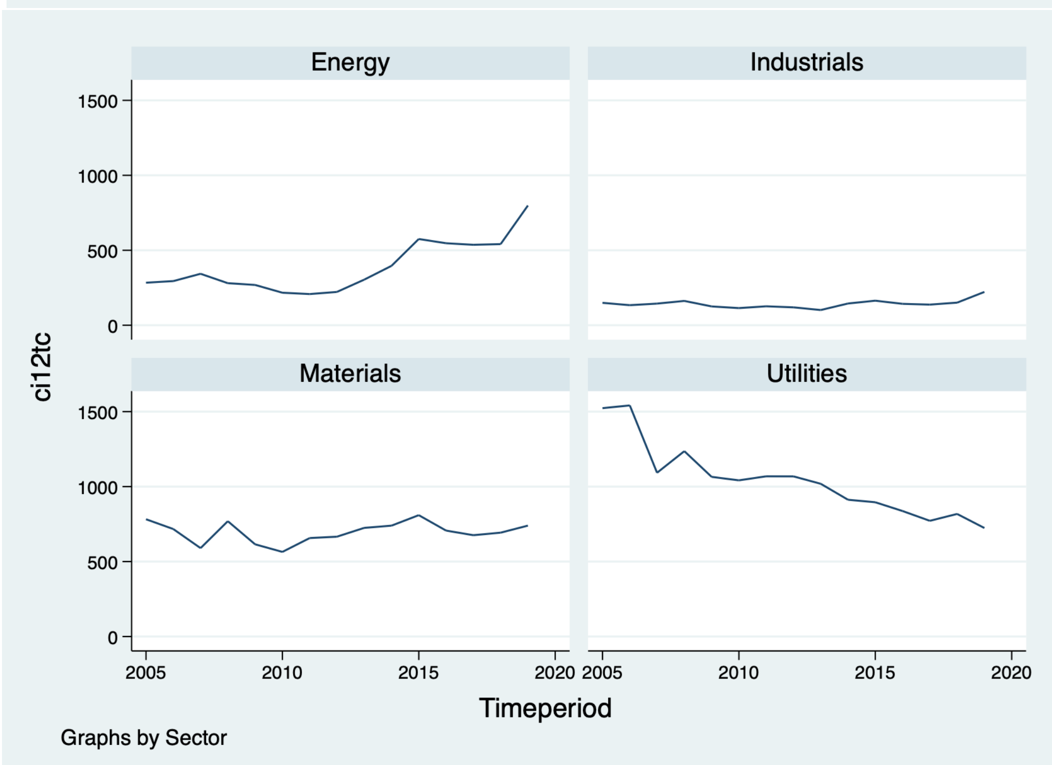


Table 24: Average company size, book-to-market value and yearly returns per sector in 2019

Sector	mean $totalsize_{i,t}$	std. dev $totalsize_{i,t}$	mean $BTM_{i,t}$	std. dev $BTM_{i,t}$	mean $ARET_{i,t}$	std. dev $ARET_{i,t}$
Communication Services	13421.2	31411.84	.708285	1.038268	14.5519	42.81839
Consumer Discretionary	9949.49	46565.45	.623143	.5441074	19.3712	41.58484
Consumer Staples	12388.4	31049.43	.656426	.5069081	11.156	31.69412
Energy	22373.4	61461.87	1.37875	1.770734	15.0671	45.83487
Health Care	6058.02	20131.95	.34453	.3022882	23.827	45.92323
Industrials	6566.24	15539.75	.610175	.7090054	22.7455	44.17936
Information Technology	2543.1	7574.171	.401574	.3273636	31.2180	49.58702
Materials	8433.18	15924.59	.753892	.5882844	16.3989	28.98579
Real Estate	5977.68	10128.11	.949951	.3724002	30.176	26.72214
Utilities	29851.8	55754.84	.650533	.4776695	31.3624	34.91113

Note: Companies in the Utilities sector are the biggest followed by companies in the Energy and Communication Services sector. On average, companies in Health Care, Real Estate, and Information Technology sectors are the smallest in 2019. In terms of the book-to-market value, the sectors Energy, Real Estate, and Materials have the highest average book-to-market values with respectively 1.38, 0.95, and 0.75. Companies in the Energy sector in 2019 seem underpriced, whereas companies in the sectors of Real Estate and Materials seem overpriced in 2019. In all other sectors, the book-to-market values are below 1, so companies in these sectors might also be overpriced. Companies in Information Technology and Health Care sectors are highly overpriced as their book-to-market values are respectively 0.40 and 0.30. Companies in the Utilities sector yielded the highest average annual return in 2019, namely 31.4%. Other well performing sectors in 2019 were Information Technology (31.2%) and Real Estate (30.2%). The Energy and Consumer Staples sectors have the lowest annual returns in 2019, namely 15.1% and 11.2%.

B Original set-up SCM

If the treated unit was one single company, the model would look as follows: $J+1$ units (companies) are observed in periods $1, 2, \ldots, T$. Unit "one" is exposed to the intervention (PA), i.e. treated, during periods $T_0+1, \ldots, T$. This research uses sample data during the period between 2005 and 2019. The other J are untreated potential control units i.e. the donor pool (all other companies). The result for unit i at time t, given no PA, would be Y_{it}^N; the outcome for the same unit when the intervention is applied between T_0+1 to T would be Y_{it}'. The goal is then to estimate the effect of the PA on the treated unit:

$$\tau_{1t} = Y_{1t}' - Y_{1t}^N = Y_{1t} - Y_{1t}^N \tag{14}$$

for $t > T_0$, and Y_{1t} is the outcome for unit one at time t. Next, $\boldsymbol{W} = (w_2, \ldots, w_{J+1})'$ is defined as a vector of the different weights for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Then, the model defines $\boldsymbol{X}_1$ as a $(k \times 1)$ vector that contains the characteristics of the treated unit before the intervention and $\boldsymbol{X}_0$ as a $(k \times J)$ matrix that includes the same variables for the non-treated units. The goal here is to minimize $\|\boldsymbol{X}_1 - \boldsymbol{X}_0\boldsymbol{W}\|$ subject to the weight constraints. $\|\boldsymbol{X}_1 - \boldsymbol{X}_0\boldsymbol{W}\|$ can also be written as $\sqrt{(X_1 - X_0W)' V (X_1 - X_0W)}$. The synthetic control estimator becomes:

$$\widehat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}. \tag{15}$$

The observed outcome of the SCM can be written as the sum of a treatment-free potential outcome $({Y_{it}}^N)$ and the effect of the treatment τ_{it}:

$$Y_{it} = Y_{it}^N + \tau_{it} D_{it} \tag{16}$$

$$Y_{it}^N = \delta_t + \lambda_t \mu_i + \theta_t Z_i + \varepsilon_{it} \tag{17}$$

where D_{jt} represents the treatment effect with 1 for a treated unit after T_0, and is 0 otherwise. δ_t is the time fixed effect, μ_j are unobserved factor loadings, θ_t are observed common factors, Z_i are observed factor loadings and λ_t are unobserved common factors. ε_{it} is the idiosyncratic error term. However, in equation 17, Abadie et al. (2010) established a bias bound. Only if the covariates of the control would equal those of the treated group i.e. $\sum_{j=2}^{J+1} w_j Z_j = Z_1$ and $\sum_{j=2}^{J+1} w_j Y_{jt} = Y_{1t}$ for $t = 1, \ldots, T_0$, $\widehat{\tau}_{1t}$ is an approximately unbiased estimator of τ_{1t}.

C Empirical results

Table 25: Panel regression results with dummy variable for companies with low and high carbon intensity

	ARET
$R_{Mt} - R_{ft}$	1.046***
	-71.52
GDP	-0.707***
	(-4.15)
Book-to-market	-0.179***
	(-30.51)
Log Total Size	0.00619***
	-3.59
Log carbon intensity (CI)	0.00454
	-1.81
First difference log CI	-0.0240***
	(-4.17)
Dummy variable LH (low = 0, high = 1)	-0.0188**
	(-2.59)
Constant	0.115***
	(4.42)
Country fixed effects	y
Sector fixed effects	n
Time fixed effects	n
F-test	252.78
Df	10,792
R2	0.397
N	10,793

Note: This table shows panel regressions of yearly excess returns as the dependent variable on the CI and control variables. The country fixed effects and dummy variable for high and low emitting companies are also included for the period from 2005 to 2019. *, **, *** denote significance on the 10%, 5%, and 1% level, respectively. Robust t-statistics are displayed in parentheses. Significance tests are based on two-sided t- tests.

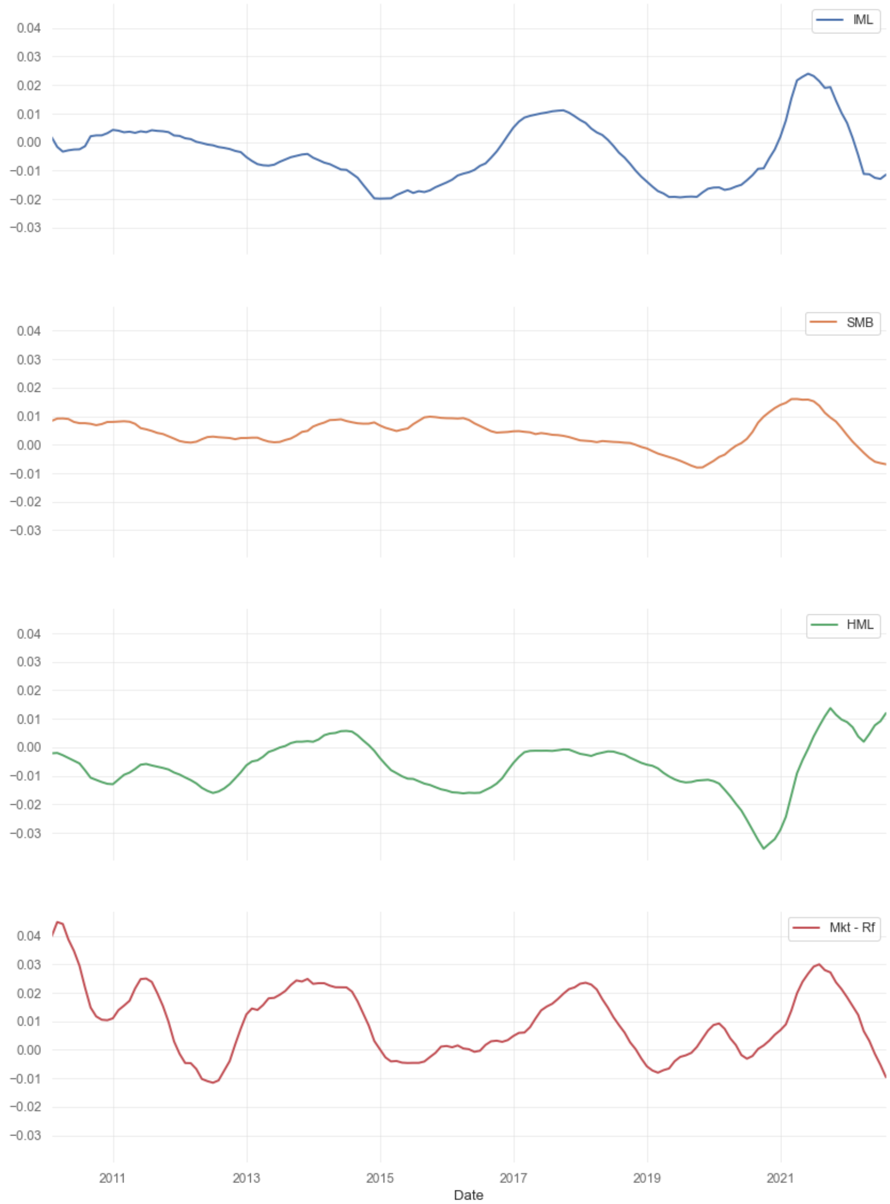
Figure 18: Cumulative returns different portfolios

(a) Triple sorted on sector, size and carbon intensity

(b) Triple sorted on sector, value and carbon intensity

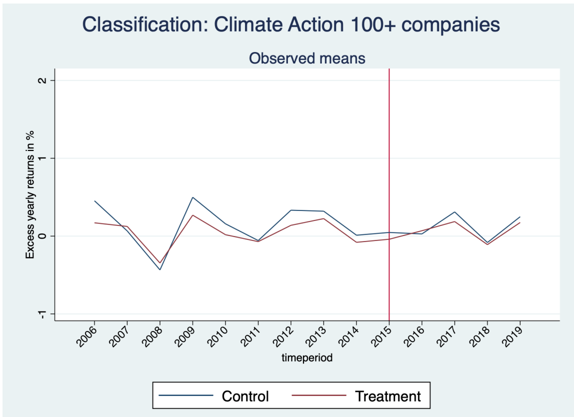
Note: Cumulative return for each portfolio when triple sorted on sector, size/value and carbon intensity. The top figure shows the cumulative returns for all five portfolios when sorted on sector, size, and carbon intensity, where portfolio 1 (blue line) is the least carbon-intense portfolio and portfolio 5 (orange line) is the most carbon-intense portfolio. From 2008 until 2015, portfolio 1 and portfolio 5 follow a similar trend. However, the more carbon-intense portfolio keeps outperforming the less carbon-intense portfolio. Then, between 2015 and 2016, the two portfolios interchanged in terms of performance, and from 2019 onwards, the less carbon-intense portfolio yielded higher returns than the more carbon-intense portfolio. Portfolios 1, 2, and 3 all outperform portfolios 4 and 5. So, portfolios consisting of less carbon-intense firms yield higher cumulative returns than portfolios consisting of more carbon-intense firms. The bottom figure shows the cumulative returns for all five portfolios when sorted on sector, value, and carbon intensity, where portfolio 1 and 5 follow a very similar trend where from 2019 on, the less carbon-intense portfolio yields higher returns than the more carbon-intense portfolio. The only big difference is portfolio 4. This portfolio performs better now compared to when it was sorted on size.

Figure 19: Risk premia from 2008 to 2022

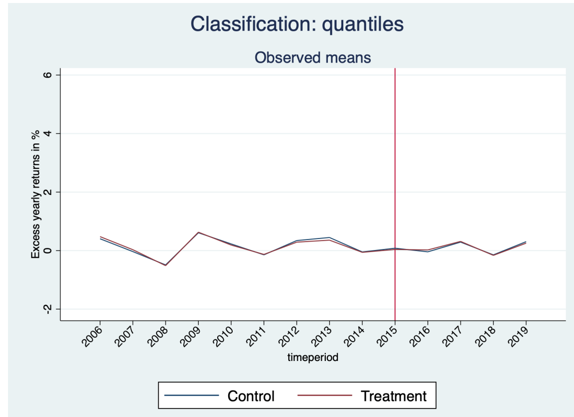


Note: The IML and HML risk premium follow similar trends whereas the SMB factor is less volatile. No clear change is observed in or around 2015.

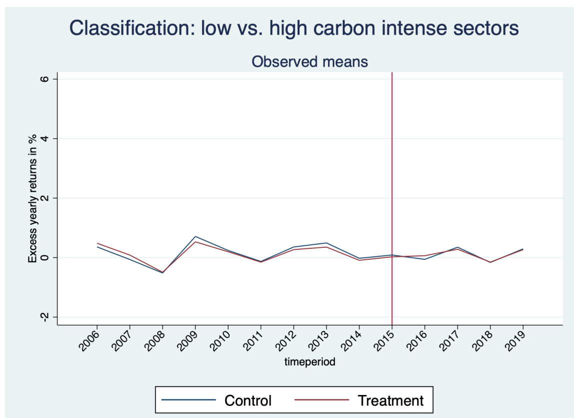
Figure 20: Graphical diagnostics DID estimates



(a) Climate Action 100+ companies



(b) Quantiles



(c) Low vs. high sectors

Note: Graphical diagnostics for parallel trends of the PA effect with three different classification methods for treatment and control group. All classification methods show parallel trends.

Table 26: DID estimates with U.S. companies as control group

	OLS	OLS with FE
$R_m - R_f$	0.318***	
	-4.53	
Book-to-market	-0.174***	
	(-7.67)	
Log Total Size	-0.00424	
	(-0.95)	
Carbon intensity (CI)	-0.00331	
	(-0.62)	
Treatment effect	-0.0258	
	(-1.09)	
Time effect	-0.00394	
	(-0.20)	
DID	0.0644**	0.0648*
	-2.97	-2.26
Constant	0.237***	-0.181***
	-5.71	(-5.49)
Time fixed effect	n	y
Company fixed effect	n	y
R^2	0.1435	0.1674
F-stat	69004.01	20.04
(p-value)	(0.000)	(0.000)
Observations	1116	1116

Note: This table shows the DID panel regressions of yearly excess returns as the dependent variable on the CI and control variables. The control group consists of 58 U.S. companies and the treated group consists of 35 EU companies. Robust t-statistics are displayed in parentheses. *, **, *** denote significance on the 10%, 5%, and 1% level, respectively. Significance tests are based on two-sided t- tests.