Where is the Carbon Premium? Global Performance of Green and Brown Stocks*

Michael D. Bauer[1,2,3], Daniel Huber[1], Glenn D. Rudebusch[4,5], and Ole Wilms[1,6]

[1]Universität Hamburg
[2]CEPR
[3]CESifo
[4]Brookings Institution
[5]New York University
[6]Tilburg University

January 23, 2023

Abstract

The relative equity pricing of more climate-friendly ("green") versus less climate-friendly ("brown") companies is an open question in climate finance. Previous research comes to conflicting conclusions, documenting either a "carbon premium" with brown stocks yielding higher returns, or the opposite, with green stocks outperforming brown. This paper provides new international evidence on this issue for a range of methodologies. Using carbon dioxide (CO_2) emissions as reported by companies to measure their greenness, we document that green stocks across the G7 have generally provided higher returns than brown stocks for much of the past decade. We also try to reconcile our findings with previous work, and we provide some results for early 2022 that show that brown stocks outperformed green ones during the energy crisis.

Keywords: climate risk, transition risk, carbon emissions, green stocks, brown stocks
JEL Classifications: G11, G12, Q54

*We thank Dries Laurs and Luke Taylor for sharing their factor series and two anonymous referees, Sanjay Patnaik, Gian Maria Milesi-Ferretti, and David Wessel for helpful comments. Bauer and Wilms acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project no. 509502478.

1 Introduction

An important question at the forefront of recent financial investment, research, and regulation is whether private business enterprises are appropriately positioned for climate change. One aspect of this multi-faceted inquiry considers how a company's financial position may be affected by the prospect of climate-related changes in the physical, economic, financial, and policy environments. Companies may have assets at greater or lesser risk as sea levels or extreme temperatures rise, or they may have business models that will become more or less profitable given new carbon taxes or climate regulations that promote decarbonization. Firms that are considered more environmentally sustainable or climate-friendly—for example, those better positioned to succeed in a low-carbon world due to their lower CO_2 emissions—are often referred to as green firms, and less climate-friendly ones are brown firms. This paper provides new international empirical evidence on the relative equity returns of green and brown firms. We find that, over the past decade, green portfolios defined using firm-level reported emissions data have performed better than brown ones. Our empirical approach tries to ensure that these superior returns could have actually been achieved by investors, using portfolio methods and publication lags that avoid ex-post biases to the extent possible.

The specific question we examine—the green versus brown relative equity performance—has been much debated by climate finance researchers, the investment industry, and policymakers. A variety of different mechanisms have been proposed by which a firm's environmental characteristics may affect its financial performance and cost of capital. Many finance researchers (e.g., Bolton and Kacperczyk, 2021, 2022) use traditional efficient capital markets theory and apply a standard risk and expected return tradeoff to green and brown firms. Brown firms arguably face greater climate-related financial, liability, and regulatory risks associated with fossil-fuel energy use.[1] For example, high-carbon emitters are subject to elevated "climate transition risk," so future increases in carbon prices or similar climate policies would disproportionately affect these firms, devaluing their assets (potentially resulting in "stranded assets") and business models. As a result, investors in brown firms would require compensation in the form of higher expected returns for holding additional climate risk—a "carbon risk premium." This premium implies that brown firms face a higher cost of capital and lower valuations (price multiples) on projected earnings.

[1] Understanding the potential climate-related physical and transition risks faced by different assets and businesses has been center of much recent analysis by financial regulators and central banks worldwide (e.g., NGFS, 2021). In the United States, the Securities and Exchange Commission (SEC) has proposed a requirement that publicly-traded companies would disclose their climate-related risks in their financial reports.

By contrast, the investment industry promotes the potential for higher average returns of green firms—in effect, a low-carbon alpha—rather than their reduced riskiness. The broad category of ESG funds, which take into account environmental, social, and governance factors in shaping investment strategies, has seen some evidence of superior returns in recent years, and an investor survey by BNP Paribas (2019) reported that 52% of respondents cited "improved long-term returns" as a motivation for ESG investing while only 37% cited "lower investment risk." The survey also reported that some investors have non-pecuniary preferences for environmentally-friendly companies.[2] The potential dual appeal of ESG investing with financial gain and social mission has been summarized under the rubric of "doing well by doing good." However, such a combination is difficult to reconcile with a simple asset pricing model. Any investor demand for green assets solely for their greenness—and a distaste for carbon-dependent assets—will tend to bid up prices of green firms relative to brown ones and reduce the green expected returns. Consequently, similar to a carbon risk premium, a "carbon aversion premium" based on investor preferences would also tend to lower brown firm valuations and raise their expected equity returns and cost of capital.

That is, in theory, if green assets provide both a hedge against climate risk and have appealing non-financial characteristics, they would seem unlikely to provide higher expected returns and thus realized returns. However, realized green returns can differ substantially from expected returns for a period of time if risk perceptions or preferences shift unexpectedly over time, as described by Ardia et al. (2022) and Pástor et al. (2022). Notably, Pástor et al. (2021) provide a theoretical framework that explains such a divergence via increasing investor demand for green assets and increasing consumer demand for green products. With such unexpected shifts, green stocks' realized returns could have exceeded those of brown stocks despite having lower expected returns. For example, unexpected increases in preferences for green assets or in public knowledge and attention to corporate carbon footprints could have pushed up green asset prices. Alternatively, as argued by Lontzek et al. (2022), a sequence of large negative climate shocks can raise perceived climate risks, boosting the share of green investors and driving up green asset valuations and returns over time. Such scenarios appear consistent with the trend toward greater climate concerns over the past decade and the sizable volume inflows that resulted in a remarkable increase in the amount of ESG assets under management. Especially with widely differing initial views and a slow learning by investors about the extent and importance of climate change and the climate policy response, there may be a lengthy transitional period with continuing sizable demand growth for sustainable investments.

[2] In the extreme, some may consider high-carbon firms as "sin stocks" and avoid them altogether. However, catering to such investor green preferences has become politically controversial in the United States recently.

While theoretical models have described several possible factors affecting the relative performance of green and brown stocks, the related empirical research has made slow progress in sifting among them. Instead, it has produced a broad set of seemingly contradictory results that are far from a consensus. Some empirical studies find that green stocks have lower returns, confirming the existence of a carbon premium in line with the predictions of basic asset pricing theory. Prominent examples include two papers by Bolton and Kacperczyk (2021, 2022), who use panel regressions of equity returns on CO_2 emissions and find that higher-emitting firms do appear to have higher returns. Also, Bansal et al. (2021a) argue that climate change risk is priced in the stock market by showing that low-frequency variations in global temperature have a negative effect on asset valuations and carry a positive risk premium. On the other hand, several studies have documented substantially positive returns for portfolios that go long in green stocks and short in brown stocks (Garvey et al., 2018; In et al., 2019; Huij et al., 2021; Pástor et al., 2022). However, as noted above, this evidence for green outperformance may not rule out the existence of a carbon premium. Indeed, Pástor et al. (2022) and Ardia et al. (2022) provide some evidence that green stocks may have had higher realized returns than brown stocks because of increased concerns about climate change, but similar or even lower expected returns. As a whole, the current body of climate finance research has not consistently established the recent relative performance of green and brown stocks on an ex ante or ex post basis.

Many different choices about empirical design likely contribute to the very mixed results, and three of these stand out.[3] First, the greenness of firms can be measured in various ways, using the level of CO_2 emissions, or emissions scaled by firm size (emission intensity), or the growth rate of emissions, or constructed environmental scores, say, the "E" component of the ESG ratings provided by various financial data providers. In the previous literature, Bolton and Kacperczyk (2021) find a carbon premium for the emissions level but not for emission intensity, while In et al. (2019) and Cheema-Fox et al. (2021) find positive returns for portfolios sorted on emission intensity, and Pástor et al. (2022) find the same for portfolios sorted on E-scores. A related issue is the common use of *estimated* CO_2 emissions—again provided by various third-party data sources—for the many companies that do not report their actual emissions. Aswani et al. (2022) find that using either reported or estimated emissions can have pronounced effects on the empirical results.

A second important choice is the sample—the set of firms and time period for analysis. While Bolton and Kacperczyk (2021) analyze a relatively long sample from 2005 to 2017, Pástor et al. (2022) suggest that use of a shorter and later sample from 2012 to 2020 may

[3]See also Bolton et al. (2022) for a discussion of potential reasons for the divergence in existing estimates of the carbon premium.

account for the divergent results. These and most other related studies consider only publicly-traded firms in the United States.

Third, methodological choices appear to be quite consequential in this debate as well. The strongest findings for a carbon premium are from the panel regressions of Bolton and Kacperczyk (2021, 2022). However, as emphasized by Aswani et al. (2022), these results seem to depend to some degree on the regression specification and included control variates. By contrast, most of the reports of better returns for green versus brown firms are not based on regression analysis but on portfolio sorts and portfolio returns, using classic methods from empirical asset pricing.

We assess evidence on the carbon premium using an empirical design that carefully addresses these three choices. Our baseline methodology follows established practice from empirical finance: We document the performance over time of portfolios of firms, based on actually reported emissions data and appropriate publication lags. In particular, to better understand the gap between previous empirical research that relies on panel regressions or portfolio methods, we also investigate the sensitivity of our findings by using a parallel panel regression approach, going step by step from mean portfolio returns to various panel regression specifications.

We begin our empirical analysis with data from the United States, where we can compare our results with those of previous studies and examine some methodological variations. Using a range of different portfolio methods and both emission levels and emission intensities, we find that over the period from 2010 to 2021 green stocks had higher average returns than brown stocks. Instead of being concentrated in specific episodes, the outperformance was rather steady and persisted over almost the entire sample period. We measure greenness using *reported* scope 1 and scope 2 CO_2 emissions and emission intensities to avoid issues associated with third-party estimated emissions or E-scores.[4] Our evidence suggests that the green outperformance in the U.S. is a robust result in the data based on portfolio sorts with either emission levels or emission intensity and for all factor methods. Using emission levels, we find that from 2010 to 2021 green portfolios had a cumulative return up to 70% percent higher than brown portfolios. Using emission intensity, the differences in cumulative returns are even higher. Furthermore, quintile portfolios based on emission intensities show a monotone decrease in average returns going from the greenest to brownest portfolio.

With few exceptions, most studies focus exclusively on the United States.[5] We extend and confirm the robustness of our findings by estimating green and brown portfolio returns

[4]Estimates of ESG scores from different providers can have low correlations and are subject to revisions (Berg et al., 2022). In additional analysis, we also include estimated emissions, but these are based on proprietary estimation methods that differ across data vendors.

[5]Notable exceptions include Bolton and Kacperczyk (2022) and Zhang (2022).

for all of the other G7 countries as well. Our multi-country analysis reveals that in most G7 countries green stock returns were higher than brown returns—similar to the United States. Using emission intensity as a proxy for greenness, we find a green outperformance for six of the G7 countries ranging from 23% for Canada to 110% for France. The notable exception is Italy, where brown stock returns have been higher on average than green stock returns over our sample period.

From an asset pricing perspective, the natural next question is whether the higher returns of green portfolios reflect a higher amount of associated risk. A first step to answering this question is to investigate the Sharpe ratio, the simplest measure of risk-adjusted return, for the green and brown portfolios. We find that the return of brown portfolios tend to be slightly less volatile than green portfolios, both in the U.S. and internationally. However, the differences in average returns are larger than can be accounted for by differential volatility; that is, green portfolios typically have higher Sharpe ratios than brown ones. We also report estimates for the classic three-factor asset pricing model of Fama and French (1993). The results show that a portion of the negative average returns of brown-minus-green spread portfolios arises from the higher exposure of brown stocks to the value factor. Brown-minus-green portfolios also have negative alphas for most measures of greenness and methods to construct spreads, even though alphas are not statistically significant at conventional significance levels. In sum, the green outperformance in our U.S. sample appears to be driven by a combination of lower risk and higher risk-adjusted returns of green vs. brown stocks.

We also estimate panel regressions and try to reconcile the broad differences in results obtained with the two empirical methodologies. After including fixed effects and controlling for firm characteristics and past returns similar to Bolton and Kacperczyk (2021), we still find that brown stocks have lower returns than green stocks. This result holds across a number of different specifications, even though the return differences in panel regressions tend to be smaller than for portfolio methods. Our results suggest the importance of measuring greenness relative to other firms, instead of estimating regressions of returns on the level or intensity of emissions across all firms and time periods.

While the carbon premium hypothesis implies that green stocks have lower expected returns than brown stocks, we find that green stocks have higher realized returns than brown stocks. Assuming that average realized returns are good proxies for expected returns—a common assumption in empirical asset pricing—our findings appear inconsistent with the existence of a carbon premium. However, as noted above, the observed green outperformance could be consistent with higher expected returns for brown stocks if, over the past decade, unexpected changes in investor preferences or risk pricing provided support for price increases of green stocks. The evidence in Ardia et al. (2022) and Pástor et al. (2022) provides some support for the view that a steady rise in climate concerns in recent years has led to higher realized green stock returns. We do not provide a definitive interpretation of whether our evidence can be reconciled with the presence of a carbon premium, as the focus of our paper is on empirical equity returns historically conditional on carbon emissions and the surrounding methodological issues. But our evidence suggests that the carbon premium was unlikely very large at the beginning of our sample, because in that case, substantial unexpected shocks—and a large wedge between realized and expected returns—would have been required to overcome expected brown outperformance.

Finally, although the sample for our central empirical analysis ends in 2021, we also briefly examine green and brown stock performance in 2022 in the midst of the energy shock caused by the Russian invasion of Ukraine. For the first half of 2022, we find that brown stocks have posted higher returns than green ones—in stark contrast to our results for the previous decade. This marked reversal in brown stock performance holds for different measures for greenness as well as across all of the G7 countries. This dramatic shift in green vs. brown performance likely reflects unexpected demand in the energy and defense industries, two particularly brown sectors. As the crisis is still ongoing, more time and data are needed to understand whether this is a transitory setback in relative green performance or a more long-lasting shift.

Following a short overview of the related literature, the remainder of this paper is organized as follows. Sections 2 and 3 describe the data and empirical methodology. The empirical results for the United states and all G7 countries are presented in Sections 4 and 5, respectively. Section 6 tries to reconcile portfolio methods and panel regressions, which have both been used in the literature with conflicting results. Section 7 discusses the impact of unexpected shocks including the current energy crisis. Section 8 concludes.

Related Literature

Our paper connects with a large empirical literature that studies the relationship between environmentally sustainable practices of corporations and financial performance. For example, El Ghoul et al. (2011) find that firms with stronger corporate social responsibility, including environmental policies, tend to have a lower cost of capital, and Plumlee et al. (2015) connect better environmental disclosures to higher firm value. Focusing on greenhouse gas emissions, Delmas et al. (2015) documented a positive relationship between environmental performance (reduced emissions) and long-run financial performance, measured by Tobin's q, but a negative relationship with short-run financial performance, measured by returns on assets (ROA). These results were revisited and updated by Busch et al. (2022a), who found that environmental performance was negatively related with both short-run and long-term financial performance.

The influential studies by Bolton and Kacperczyk (2021, 2022), which provide evidence in favor of a carbon premium, are consistent with the results in Delmas et al. (2015) and Busch et al. (2022a) showing a positive relationship between carbon emissions and financial performance. This evidence is generally based on panel regressions in which measures of financial performance, such as stock returns, are regressed on measures of carbon emissions. A related example is Görgen et al. (2020), who regress returns on their custom "brown-minus-green" score and find a positive coefficient.

Other work has investigated this issue with the more traditional method in asset pricing of investigating the returns of portfolios of stocks defined by some company characteristic. Applying this approach in the context of climate finance, the portfolio sorts are based on firms' past environmental record. Garvey et al. (2018), In et al. (2019), Huij et al. (2021), and Pástor et al. (2022) show that, in the United States, green stocks—measured by the level of carbon emissions, emission intensity, or E-scores—have on average outperformed brown stocks by a significant margin.[6] In other words, portfolio-based estimates generally suggest a negative relationship between carbon emissions and financial performance. Ardia et al. (2022) show that green stocks tend to outperform brown stocks in response to unexpected increases in climate change concerns, measured from newspaper articles. Their results on the importance of climate change concerns provide a possible rationale for green outperformance.[7] Aswani et al. (2022) challenge the results of Bolton and Kacperczyk (2021) on the grounds that using estimated carbon emissions makes them unreliable. Bolton et al. (2022) respond to this criticism and acknowledge that the estimated carbon premium is smaller in a sample of companies that disclosed emissions.

An important but unresolved question is whether average realized returns differ significantly from expected returns of green-minus-brown portfolios, i.e., whether substantial, persistent surprises drive a wedge between the two. Pástor et al. (2022) address this issue and provide evidence that higher returns for green stocks were in part driven by unanticipated increases in environmental concerns, using the text-based measure of climate change concerns of Ardia et al. (2022). van der Beck (2022) uses fund-flow data to show that in the absence of flow driven price pressure, green portfolios would not have outperformed brown

[6] Consistent with this, Görgen et al. (2020) find a negative average return for a brown-minus-green portfolio based on their own environmental score but note that it is not statistically significant. For related evidence see Cheema-Fox et al. (2021).

[7] Some previous studies, including Huij et al. (2021) and Görgen et al. (2020), have investigated whether carbon is a priced risk factor in the cross section of stock returns. Our focus is somewhat different, as our main goal is to document relative green and brown equity performance.

portfolios from 2015–2021.

There are various alternative ways to model climate risk exposures, beyond using emissions data. Engle et al. (2020) use ESG scores to model firms' climate risk exposures and develop a dynamically hedged portfolio that has a negative premium attached to climate change risk. But ESG scores have been criticized based on the observation that they often differ widely across different data providers, and they can be revised retroactively (Berg et al., 2022; Gibson Brandon et al., 2021). Instead, Sautner et al. (2022a) use earning calls to determine firm-level climate change exposures, and Sautner et al. (2022b) find that climate change exposures carry a positive and increasing risk premium since the financial crisis.

2 Data

We obtain firm-level accounting, equity return, and carbon emissions data from Refinitiv: Accounting data are from Refinitiv's Worldscope database, stock market returns are from Datastream, and carbon emissions are from the Asset4/ESG database. Refinitiv has extensive global coverage of emissions from various providers (including the Carbon Disclosure Project), and these emissions data have been used in a number of related studies including Ardia et al. (2022) and Rohleder et al. (2022). In our empirical analysis, we include companies from the G7 countries, i.e., the United States, Canada, France, Germany, Italy, Japan, and the United Kingdom. We concentrate on these seven large, developed economies because they have the best coverage of emissions data. Currently, Refinitiv contains emissions data for about 3,500 U.S. firms and about 5,800 G7 firms. However, we omit companies with unusual equity quotes by applying a variety of standard filters in the literature.[8]

Instead of relying on ESG scores—or just E-scores as in Pástor et al. (2022)—we only use reported data for CO_2 emissions.[9] Emissions data are more consistent than E-scores as documented in Busch et al. (2022b), and CO_2 emissions are an important aspect of a company's environmental footprint for investors and consumers.

A number of previous studies have investigated the relationship between CO_2 emissions and financial performance, as noted above. One potential problem is that even for companies with no reported emissions or emissions disclosures, data providers can provide "estimated"

[8] These filters are: The type of security must be common equity (e.g., Ince and Porter, 2006); only primary equity quotations are included (Fong et al., 2017); only firms located (Ince and Porter, 2006) and securities listed (Griffin et al., 2010) in the relevant countries are considered; securities with a price quoted in a currency different from the one of the associated country are disregarded (Griffin et al., 2010); non-ordinary shares are disregarded; securities whose name fields indicate non-common equity affiliation are removed (Ince and Porter, 2006; Karolyi et al., 2012); and duplicates based on the firm name are manually excluded.

[9] As a robustness check, we also provide results using estimated emission by the proprietary model of Refinitiv.

emissions based on proprietary models of varying levels of sophistication.[10] Using these imputed emissions data can lead to biases in empirical results as discussed by Aswani et al. (2022). Consequently, we focus on the actually reported, or "disclosed" emissions. While this choice does not come without a cost as it reduces the number of firms in our sample, it may boost the reliability of our emissions data. Another advantage is that reported emissions generally do not get revised later.

We focus on scope 1 and scope 2 emissions, that is, the direct emissions from company-owned and controlled resources (scope 1) and the indirect emissions from the power generation of the energy purchased by the company (scope 2). These emissions are relatively straightforward to measure and mandatory to report according to the comprehensive global Greenhouse Gas Protocol Corporate Standard.[11] As in Huij et al. (2021) and other studies, we exclude scope 3 emissions—the emissions generated by a company's upstream and downstream activities—because these are hardest to monitor, voluntary to report, and very large in magnitude.

We calculate two measures of a firm's CO_2 footprint. The first is simply the total level of a company's scope 1 and scope 2 emissions in a given year. Emission levels are the simplest proxy for the greenness of a firm, but they depend on a firm's scale. Two firms with the same energy efficiency but different sizes would have different reported levels of greenness. To reduce the impact of size on the measured greenness of a firm, the second measure we consider is emission intensity, which normalizes emissions by the size of the company. Specifically, we define emission intensity as the ratio of the level of emissions and total revenues in the same year. Emission intensity has been the industry standard to determine the carbon exposure of stock indices (e.g., S&P Global (2020) and MSCI (2022)) and it is also a widely used measure for the greenness of a stock in finance research (e.g., Garvey et al. (2018), In et al. (2019)) and Cheema-Fox et al. (2021)).

Table 1 reports summary statistics of our emissions data for the sample period from 2010 to 2021. The data include a little over 12,000 company-year observations with non-missing emissions data, which is about one-fourth of the roughly 50,000 observations in our sample. There is a small number of very large emitters for both, the level and intensity of emissions. As a result, the emissions data are very right-skewed, as mean emissions exceed not only median emissions but even the third quartile.

In addition to annual calendar-year emissions and fiscal-year balance sheet data, we use monthly return data. Specifically, we use the total return index in U.S. dollars (USD)

[10]For a description of the carbon emissions estimation methods of Thomson Reuters Refinitiv, see https://www.refinitiv.com/content/dam/marketing/en_us/documents/fact-sheets/esg-carbon-data-estimate-models-fact-sheet.pdf (accessed 10/19/2022).

[11]See https://ghgprotocol.org/corporate-standard (accessed 10/19/2022).

for each stock in our sample, so that our performance results apply to the perspective of a U.S. investor who continually reinvests dividends.[12] When constructing value-weighted portfolio returns, we also use monthly observations of market capitalizations.

To illustrate the coverage of firms in our data, Figure 1 plots the number of U.S. firms in our sample.[13] Between 2003 and 2010, the number of available firms increases from near-zero to 135, and towards the end of the sample the number of firms reaches a level around 750. This increase reflects a greater tendency of firms to report emissions data and the improving coverage of the Refinitiv database.[14] For comparison, the figure also shows the number of firms used in Huij et al. (2021), based on emissions data from Trucost, as well as in Pástor et al. (2022), based on ESG data from MSCI. Generally, these samples include many more firms than ours, mainly because they do not require the availability of reported emissions, relying instead on estimated emissions or environmental scores. It is also striking that the number of firms in these other samples at some point jumps up dramatically, in 2012 for Pástor et al. (2022) when MSCI expanded its ESG coverage, and in 2017 for Huij et al. (2021). Such dramatic changes in sample coverage may be problematic, and indeed Pástor et al. (2022) start their sample immediately after the jump in coverage to enhance the consistency of their data. Our firm coverage gradually increases, but to ensure sufficient sample size, we start our sample in January 2010 when we arguably have a reasonable degree of firm coverage to reliably calculate portfolio returns. Our baseline analysis ends in December 2021, which yields a sample of 12 years. As an out-of-sample addition, we also discuss the first half of 2022, which is a unique period due to the global energy crisis caused by the Russian invasion of Ukraine (currently unfolding as we write). We consider this extraordinary episode and discuss the role of unexpected shocks in Section 7.

[12] We perform a series of dynamic filters that are recommended in the literature with respect to the calculation of returns and market capitalizations of a firm: We delete zero returns and market capitalizations from the end of the sample up to the first non-zero return to correct the false continuation of stale prices in case of delistings (Ince and Porter, 2006); we drop returns and market capitalizations in the case of unadjusted prices that are larger than one million USD; we delete monthly returns and market capitalizations if the monthly return is larger than 990% to remove outliers (Griffin et al., 2010).

[13] In each month t, the number of firms is determined by the availability of return data for month t as well as emissions data for the year containing month $t-18$, given the publication lag we use in matching returns and emissions, as discussed below in Section 3.

[14] At its inception in 2003, the Refinitiv emissions data only covered companies in the major indices in the U.S., Germany, France, U.K., and Switzerland. The coverage expanded substantially in 2008 with the companies in the DJ STOXX and MSCI World, and continued to grow thereafter. For details, see https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf (accessed, 10/19/2022).

3 Empirical methodology

There is a sizable literature in empirical asset pricing that compares the stock returns of different portfolios of companies that are distinguished by various firm characteristics.[15] Similarly, our empirical approach is to analyze portfolios constructed of green and brown firms that are defined by their relative emission levels or intensities. For this exercise to be informative, it is important to classify firms and measure their performance in a way that is consistent with what investors could have achieved in real time. To this end, we classify firms into portfolios based on emissions (and accounting) data that was publicly available at the time.

Emissions data are available on an annual basis and tend to be published by the middle of the following year. Therefore, at the end of each June, we construct new green and brown portfolios using emissions data from the previous year. Specifically, our portfolios are constructed using contemporaneous market data as of the end of June and emissions data from the previous year. Then, portfolio returns are calculated for the subsequent twelve months until the end of June of the following year when a new sorting process takes place. In other words, returns for month t are matched to emissions data from the year containing month $t-18$. For example, the portfolio sorts used for the returns in each month from July 2020 to June 2021 are based on emissions data for 2019. With this publication lag, we aim to avoid look-ahead bias, similar to Ardia et al. (2022) and Ilhan et al. (2021). By contrast, Bolton and Kacperczyk (2021, 2022) do not lag their emissions data and relate returns to current-year emissions.

For robustness, we employ three different methods for constructing green and brown portfolios. The first method is a "size-adjusted spread" that closely follows Huij et al. (2021). As in Fama and French (1993), who use two-way sorts by market cap and book-to-market in order to construct a return series for a value factor, we use a two-way sort by market cap and emissions (level or intensity). First, firms are classified as "small" or "big" based on their current (end-of-June) market capitalization in comparison to the median market capitalization.[16] Then, in each size group, firms are divided into three carbon emissions groups "low/green", "medium", and "high/brown", based on the 30th and 70th percentiles of the emissions variable. For the resulting six portfolios, we calculate value-weighted returns. The green portfolio return is the equal-weighted average of the returns for the "small-green" and "large-green" portfolios, and similarly for the brown portfolio return. The size-adjusted spread return is the difference between the brown and green portfolio returns (see equation

[15] See Banz (1981), De Bondt and Thaler (1985), Fama and French (1993) and many others.

[16] While Huij et al. (2021) use the median NYSE firm size for this classification, we use the median firm size in our sample.

(1) in Huij et al. (2021)).

Our second and third portfolio methods employ a "simple spread," in the sense of the capital market anomaly literature. Pástor et al. (2022) use a similar approach to construct their "green-minus-brown" factor, again, using E-scores. Instead, we sort stocks into quintiles based on their reported emissions level or intensity and calculate the average return for each quintile portfolio.[17] The simple spread return is the difference between the quintile portfolio returns with the highest and lowest emissions. Whereas Pástor et al. (2022) use the spread of tertile portfolios, we follow the more common approach in empirical asset pricing and use the quintile spread.[18] In our second portfolio method, we calculate this simple spread using value-weighted returns within each quintile portfolio. As a third method, we calculate a simple spread using equal-weighted returns within each quintile. The use of equal-weighted returns effectively downweights the very large firms within each portfolio that tend to dominate value-weighted portfolio returns. In this way, the equal-weighted measure is able to control for firm size similar to a size-adjusted spread.

4 Green outperformance in the United States

We begin our investigation by examining the relative performance of green and brown stock portfolios in the United States. This initial U.S. analysis allows us to compare how the results vary across our three methods for constructing spread portfolios, facilitating the choice of a single baseline method for subsequently examining returns in the G7 countries. In addition, a deep dive on the U.S. experience will also allow us to assess how our results using reported emissions compare to previous estimates obtained in Huij et al. (2021) with estimated emissions and in Pástor et al. (2022) with E-scores.

Figure 2 shows results for portfolios formed using the level of emissions. It plots the cumulative log returns for brown-minus-green (BMG) portfolios constructed with each of our three different factor methods, that is, (i) size-adjusted spread, (ii) simple spread, value-weighted (VW) and (iii) simple spread, equally-weighted (EW). For all three methods, green portfolios have outperformed brown portfolios on a fairly consistent basis since about 2012. There are a few temporary partial reversals of this experience, notably in early 2016. Nevertheless, the cumulative return of BMG portfolios over the entire sample is about -50% for both the size-adjusted spread and simple-EW spread. The portfolio difference is even more

[17] The requirements on data availability (e.g., market cap) are the same as for the size-adjusted spread to ensure comparability.

[18] See, for example, Cohen and Frazzini (2008) and Jacobs (2016). Notably, the use of decile spreads is even more common in the analysis of U.S. data (e.g. Hou et al., 2020). Since decile spreads tend to yield even larger return differences, our quintile spreads can be considered as somewhat conservative.

pronounced for the simple-VW spread, with a cumulative green outperformance of around -75%.

For comparison to earlier results in the literature, Figure 2 also plots the cumulative return for the size-adjusted spread of Huij et al. (2021), which uses the total estimated level of scope 1 and 2 emissions, and the negative of the green-minus-brown spread of Pástor et al. (2022), which is based on E-scores. Both of these existing measures are available over the period from January 2010 to December 2020.[19] Generally, these earlier spread portfolios show similar patterns as our own spread measures: little net movement from 2010 to 2012 and then pronounced declines, indicating persistent green outperformance, until the end of the sample period. The cumulative return of the simple spread of Pástor et al. (2022) by the end of 2020 is -63%, roughly in line with our value-weighted simple spread. This concordance is particularly noteworthy in light of the fundamentally different approaches taken to measuring greenness. Also, at the end of 2020, the size-adjusted spread portfolio of Huij et al. (2021) has a cumulative return of about -50%, very close to the performance of our size-adjusted spread. This correspondence suggests that in this context, the use of estimated carbon emissions (from Trucost) or only reported carbon emissions (from Refinitiv) does not materially affect the estimated green outperformance, despite the differences in data construction and sample sizes.

Figure 3 plots cumulative returns for the three spread portfolios based on emission intensity. The performance of these portfolios also shows sideways movements over 2010-2012 and a relatively steady green outperformance over the rest of the sample—again, excepting early 2016. The use of emission intensity tends to generate even larger gains for the green portfolio relative to the brown one, compared to the results based on emissions levels in Figure 2. The size-adjusted and VW simple BMG spreads now have cumulative returns that are 13 and 32 percent lower at the end of the sample, respectively, than when using emissions levels to form portfolios. One notable consequence of using emission intensity is that the EW and VW simple spread returns diverge during 2020 and 2021. It appears that early in the pandemic, the VW spread falls in 2020 and the EW spread rises in 2021. This pattern is likely driven by a differential performance of large and small companies. In 2020, it appears that large green firms had higher returns than large brown firms, so the VW spread performed well. In 2021, small brown firms outperformed small green ones, so the EW spread underperformed.

Table 2 provides details with mean returns and t-statistics that summarize the performance of the different spread portfolios as well as their components (i.e., the long "brown leg" and the short "green leg"). The mean spread return is always negative, ranges from -0.65% to -0.18% per month across approaches and is statistically significant at the 10%-

[19]We would like to thank the authors for providing us with their portfolio return series.

level in four out of eight specifications. The value-weighted simple spread based on emission intensity has the lowest mean return of -0.65% per month with a t-statistic of -2.37. This finding is in line with Garvey et al. (2018), In et al. (2019) and Cheema-Fox et al. (2021) who document green outperformance using emission intensity as a proxy for the greenness of a firm. Focusing on the long and short legs of the portfolios, we find that both green and brown portfolios have positive average returns for all measures of greenness reflecting a strong bull market during our sample with an average monthly market return of 1.27%. All brown portfolios have underperformed, and all green portfolios have outperformed the market.

Table B.1 in Appendix B shows the corresponding results using *estimated* emissions, based on the proprietary models of Refinitiv for firms that did not report emissions. We find that the results for emission intensity are very similar, both qualitatively and quantitatively, to the results using only reported emissions with an outperformance of green over brown stocks between 0.23% and 0.65% per month for the different spread portfolios. However, in contrast to our results for reported emissions, and in contrast to the results in Huij et al. (2021), we do not find a green outperformance when using the level of estimated emissions as a proxy for the greenness of a firm. The difference with Huij et al. (2021) may be due to the different proprietary models of Trucost and Refinitiv used to estimate firm-level emissions. We view this discrepancy as an indication that the use of actually reported emissions is preferable to the use of estimated emissions, even though the latter approach increases the available number of firms. Hence, we establish our main results using reported emissions to measure greenness, avoiding a dependence upon proprietary models of data vendors for estimation of CO_2 emissions.

Table 3 shows details for all quintile portfolios and their (value-weighted) returns. Specifically, we report average monthly returns, t-statistics, volatilities, and Sharpe ratios for each portfolio, for the spread (corresponding to the VW-simple spread), and for the total market portfolio. Panels A and B show results for the United States; we defer the discussion of the results for the G7 average, shown in the other panels, to Section 5. For emission intensity, we observe a monotone decrease in mean returns going from the green portfolio to the brown portfolio. While the green portfolio has a slightly larger volatility than the brown portfolio and hence is more risky ex-post, the Sharpe ratio of the green portfolio is 0.27 and thus substantially larger than that of the brown portfolio, which is 0.17. This pattern is consistent but weaker for the portfolios based on emission levels with green and brown Sharpe ratios of 0.29 and 0.26, respectively and for emission levels, average returns are not monotonically decreasing across quintile portfolios. Finally, we note that the market portfolio tends to have a higher Sharpe ratio than each individual quintile portfolio, due to a higher degree of diversification and reduction in risk.

Table 4 reports correlations between our factors as well as the factors of Pástor et al. (2022) and Huij et al. (2021). We observe correlations among the factors based on emission levels between 0.76 and 0.82. Compared to the spreads reported in the literature, we find the largest correlation for the size-adjusted spreads with the Huij et al. (2021) factor as both are based on the same methodology. However, the correlation of only 0.72 shows that including estimated emissions instead of only actual reported emissions has a significant impact on the performance of the spread portfolio. Correlations with the factor from Pástor et al. (2022) are even lower due to differences between emission levels and E-scores as proxies for greenness. Spreads based on emission intensities also have correlations with the spreads based on levels that are only 0.57 for the two simple spreads. These different measures of greenness yield only moderately correlated spreads, but, as shown above, do provide broadly similar performance results.

Our performance results raise the question whether the additional return for green portfolios arises from exposure to standard risk factors, such as the market return, value, or size. In Appendix A, we show estimation results for the classic three-factor model of Fama and French (1993). Consistent with previous results, we find that the BMG spread returns load positively on the HML (value) factor, meaning that brown firms are more likely to be value stocks, and green firms are more likely to be growth stocks.[20] The risk exposures, mainly to the value factor, explain a large share of the negative mean spread returns: We estimate alphas that are generally still negative, but they tend to be smaller in magnitude and have lower t-statistics than the mean returns in Table 2.

Overall, in the United States, we find broadly better performance of green relative to brown stocks over the past decade using different measures of the greenness of a firm as well as different portfolio construction methodologies. We find the strongest results using emission intensities but obtain solid results using emission levels as well. As a whole, there is robust evidence for a sustained green outperformance in the U.S. stock market over the past decade.

5 Global evidence

We now turn to new international evidence on the performance of green and brown stocks in the G7 countries. These seven countries are Canada, France, Germany, Italy, Japan, the

[20] In particular, Huij et al. (2021) report that returns for their pollutive-minus-clean portfolio and the HML factor are positively correlated, and note that "the most pollutive firms tend to be value firms, while cleaner firms tend to be growth firms" (p. 16). And Pástor et al. (2022) mention that their "green factor and HML are negatively correlated, as value stocks are more often brown than green" (p. 4).

United Kingdom and the United States, which, as of 2020, collectively accounted for about half of global net wealth. Each country has a somewhat different mix of green and brown companies and industries, so this G7 analysis provides a good environment to assess the robustness of our U.S. evidence for a green overperformance.

In this analysis of G7 data, we rely exclusively on the "simple spread" (value-weighted) method to construct portfolio returns. The various methods that we explored for U.S. data gave broadly similar results, so here we conduct a more parsimonious analysis that uses our preferred method.

Figure 4 shows G7 cumulative log returns for simple spreads using emission levels, and Figure 5 shows the corresponding results using emission intensities. Besides the individual countries, the figures also include results for the average across countries, which is shown by the solid black lines in each figure. This global BMG return is calculated as the average of the country-specific spread returns, instead of constructing a global spread portfolio based on pooling all G7 firms, in order to account for country-specific effects regarding the greenness of a firm. These two figures also have a slightly expanded sample that ends in June 2022. Here, as for our U.S. results, we focus on the period ending in December 2021, which is denoted by the vertical black dashed line, and defer a discussion of the returns in 2022 until Section 7.

In Figure 4 using total emissions levels, green stocks provided about 30% higher returns than brown ones on average across the G7 from January 2010 to December 2021. The general pattern over time is very close to our earlier U.S. results but about half as large in magnitude. There are also notable differences across countries with four of the seven countries showing an outperformance of green over brown stocks for emission levels (Figure 4). France reveals an impressive cumulative brown minus green return difference of about -115%. For the United Kingdom, the United States, and Canada, the spread portfolios exhibit cumulative returns between -80% and -30%. In Germany and Japan, there has been an outperformance of green stocks until late 2020, but by the end of 2021 the cumulative spread return is close to 0%. The clear outlier is Italy with an outperformance of brown stocks of about 45%.

The G7 results for emission intensities, shown in Figure 5, reveal a somewhat larger green outperformance. All countries, except Italy, show a green outperformance ranging from -40% to -110%. Again, France shows the strongest outperformance closely followed by the United States. Germany also shows a strong green outperformance of almost -80%, even though there was no green outperformance when emission levels were used as a greenness measure. We leave for future research the source of the striking differences between Italy and the other G7 countries, which may reflect different carbon disclosure practices among firms, a weaker regulatory environment highlighting climate risk, or less growth in investor enthusiasm for green funds.

Table 5 shows average monthly returns together with t-statistics for the spread portfolios as well as the corresponding long and short legs.[21] For emission levels, we find that four of the G7 countries have negative average BMG returns, three of which are significant at least at the 10% level, and for the G7 average the mean spread return is -0.20%. When using emission intensities, the relative performance of green portfolios tends to become stronger. In this case, six countries have negative average GMB returns, and the G7 average falls to -0.33%.

Panels C and D of Table 3 report average monthly returns, t-statistics, volatilities and Sharpe ratios for each of the five quintile portfolios as well as the market for the G7 average. In line with the results for the United States (shown in panels A and B), we find that after the first, greenest quintile, mean returns are monotonically decreasing across quintiles, and brown portfolios are slightly less volatile than green portfolios. However, the differences in average returns are larger than can what can be accounted for by differences in volatilities, implying that Sharpe ratios of green portfolios are larger than those of brown portfolios.

To sum up, the results show that the green outperformance documented for the United States is also a broadly consistent feature among the other G7 countries. Especially using emission intensities as a proxy for greenness, we find evidence for green outperformance in six out of the seven countries.

6 Bridging the gap: portfolios or panel regressions

Our results for green outperformance are consistent with earlier analyses of U.S. portfolio returns (Garvey et al., 2018; In et al., 2019; Görgen et al., 2020; Huij et al., 2021; Pástor et al., 2022). By contrast, the panel regressions in a number of papers generally point in the other direction, with a positive relationship between carbon emissions and financial performance (Bolton and Kacperczyk, 2021, 2022; Delmas et al., 2015; Busch et al., 2022a; Görgen et al., 2020). This naturally raises the question to which extent portfolio vs. panel methods yield different results, and whether we can bridge the gap between these two sets of evidence. To attempt some reconciliation of these apparently disparate analyses, we again narrow our consideration to the U.S. data.

A useful first step is to compare the estimated mean returns for BMG spread portfolios with estimated panel regressions that are as similar as possible to the portfolio method. This comparison will help elucidate at a basic level how portfolio returns and regression estimates

[21] In line with the evidence for the United States, the brown portfolio and the green portfolio both have positive average returns reflecting the bull market over our sample period.

compare. We start our comparison in an environment as close as possible to our baseline specification and then add the usual additional ingredients to the regression specification—such as controls and fixed effects—and investigate how they affect the results.

We first consider the mean returns of the simple BMG spreads with equally-weighted portfolios, that is "simple spread, equally-weighted" in panels A and C of Table 2, with mean spread returns of -0.22 and -0.18. Since regressions by default also weigh each observation equally, this is a natural starting point. We define an indicator variable, D_{it} that takes the value +1 if stock i in month t belongs to the brownest fifth of stocks and -1 if it belongs to the greenest fifth (based on either emission level or intensity, and using the same publication lag as before). In a cross-sectional regression of stock returns in month t, the coefficient on D_{it} would yield the spread return in that month, and the time-series average of the regression coefficients would reproduce the mean spread return in Table 2.[22]

Instead of estimating separate regressions in each month, a panel regression pools all firm-month observations. Columns (1) and (4) of Table 6 show the resulting estimates for regressions of stock returns on D_{it}, based on either emission levels or intensities. The estimated coefficients are negative, in line with the mean spread returns in Table 2. But they are less than half as big as the portfolio-based estimates. The differences must be due to the different number of observations in each month, which results in changes in the effective weights for each monthly return. That is, for the portfolio method, each month is weighted equally. Instead, in panel regressions, the weight of each month effectively depends on the number of observations in this month, and since our panels are unbalanced, these weights are generally not equal. Apparently, the changes in the data coverage lead to more observations, and a higher weight, for months with less pronounced green outperformance or even brown outperformance.

Next, we make two important modifications to the regression specification: We add time fixed effects and firm-specific controls, following Bolton and Kacperczyk (2021). The control variables we include are the ratio of book equity to market equity, 1-year sales growth, the natural log of property plant and equipment, leverage (i.e., total assets divided by book equity), the 1-month past return, the natural log of the 1-month past market capitalization, the cumulative past return from $t-12$ to $t-2$, return on equity (i.e., income before extraordinary items divided by lagged book equity), and investment-to-assets (i.e., the change in total assets divided by lagged total assets). The estimates for these specifications are in columns (2) and (5) of Table 6. The coefficients on the brown-green indicator become larger, in absolute magnitude, than in the OLS regression without controls. In the case of emission intensity (column 5), the coefficient is now statistically significant at the 1%-level.

[22]This method is similar, at least in spirit, to the second step of the Fama-MacBeth procedure.

In the case of emissions level (column 2), the t-statistic is around 1.5, which does not quite clear the bar for conventional levels of significance.

Our final specification also adds industry fixed effects, which are potentially important to account for since greenness of firms likely differs systematically with industry (Aswani et al., 2022). Adding industry fixed effects indeed lowers the absolute magnitude of the coefficient on the brown-green indicator. The coefficient remains negative, but is only marginally significant in the case of emission intensity and statistically indistinguishable from zero for the level of emissions.

With these panel regressions, we have bridged part of the gap between portfolio-based estimates and the panel regression approach. In particular, we control for time and industry fixed effects and other key firm characteristics, similar to Bolton and Kacperczyk (2021). We still find that brown stocks had lower returns than green stocks, in contrast to the positive carbon premium documented in Bolton and Kacperczyk (2021, 2022).

One key difference between the panel regressions in Table 6 and those in previous papers is that we rely on an indicator variable that measures whether a certain stock was considered brown or green relative to its peers (akin to a ranking of stocks). By contrast, previous work has generally used the level or intensity of emissions as the key independent variable in the regression. We view our approach as preferable for several reasons. First of all, it measures the relative brown and green performance based on classifications *at that particular point in time*. This approach is therefore more closely aligned with the perspective of stock investors who need to decide on their portfolios based on cross-sectional comparisons of firms. Furthermore, with the use of a brown-green indicator, firm size and other firm characteristics become less important for the regression results, as these characteristics are much less strongly correlated with the indicator than with unscaled emissions or even emission intensity. As a result, regressions with a brown-green indicator are generally less sensitive to the exact choice of regression specification as in the case with emissions level or intensity.

Table 7 shows results for panel regressions that instead use the level or intensity of emissions, as in previous studies (Bolton and Kacperczyk, 2021, 2022). For the (log) level of emissions, the results are qualitatively similar to those in Table 6, with a negative coefficient on the emissions variable that becomes smaller in magnitude and less statistically significant when we add fixed effects and firm controls. However, the results are quite different when using emission intensity as a regressor (columns 4-6 of Table 7) as opposed to using a brown-green indicator based on emission intensity (columns 4-6 of Table 6). In the former case, the coefficient is not statistically significant and switches signs depending on the specification. This result is in line with the estimates of Bolton and Kacperczyk (2021) who also find insignificant, sign-switching coefficients for emission intensity (see their Table 8, panel C).

It stands in contrast to the consistently negative and in some cases strongly statistically significant coefficient that we obtain when using a brown-green indicator.

Portfolio methods, like those we have used in Sections 4–5, reveal the relative performance of strategies that investors could have pursued in real time, which is an important reason for their popularity. However, we agree with Bolton et al. (2022) that simple portfolio methods may also reflect "the effects of differences in industry and technological characteristics that are correlated with, but do not directly contribute to, carbon intensity" (p. 19). The type of panel regressions we have shown in Table 6 can partly address this criticism. While the estimated effects are not always statistically significant, they suggest that green stocks have provided higher returns to investors than brown stocks, controlling for various differences in industry and firm characteristics.

7 Unexpected shocks, realized and expected returns

We have provided extensive evidence of green outperformance based on observed stock returns from 2010 to 2021. What are the implications of this evidence for the hypothesis of a carbon premium, or more generally, a climate risk premium? If climate change has the potential to reduce economic output and asset values, then the associated climate risks should be priced by markets. As a consequence, brown stocks, which are more exposed to climate risks, should carry a larger risk premium—and have compensatorily higher expected returns—than green stocks. By contrast, we have shown that brown stocks had lower realized returns than green stocks. The classical asset pricing approach to estimating expected returns is to use sample averages of observed returns, based on the presumption that shocks should average out over sufficiently long sample periods. This perspective suggests that our findings based on realized returns are evidence against the existence of a carbon premium during the past decade.

However, the better performance of green stocks can be reconciled with a positive carbon premium if, over the period of interest, there has been a sequence of unexpected shocks in favor of green stock prices. These predominantly one-sided shocks could be unexpected shifts in perceptions and beliefs about the size of climate risk, the pricing of that risk, or the green preferences and information sets of investors—any type of change that shifts demand from brown to green assets. Certainly, over the past decade, even a casual observer would have noted that general perceptions of the magnitude of climate change damages and risks has steadily risen over time as more has been learned about the hazards of even relatively moderate levels of warming. At the same time, there has been an increasing trend in the number and size of sustainable investment funds reflecting greater investor interest and demand for green assets. Therefore, the underlying averaging assumption required to make the jump from realized to expected returns may not serve as a good approximation to the recent past.

As noted above, a helpful theoretical framework for this issue is provided in Pástor et al. (2021). Their key assumption is that investors have a preference for green assets that may shift over time. As investors derive utility from green assets, they require lower expected returns for holding them compared to brown assets. In turn, if preferences for green assets and hence their demand increases, green stocks will outperform brown stocks even though the model implies a positive carbon premium. Empirical evidence in support of this channel includes the estimates by Ardia et al. (2022), who show that green stocks outperform brown stocks in response to unexpected changes in climate change concerns, and Pástor et al. (2022), analyze the performance of green stocks after accounting for the presence of climate concern shocks over the past decade.

Another theoretical model with a similar effect is provided in Lontzek et al. (2022) who argue that changes in the disagreement among investors about the impact of climate change on the economy can reconcile a positive carbon premium with the observed outperformance of green stocks. In their framework, brown stocks still earn a higher premium because they are more exposed to climate risk. However, shocks to climate damage assessments, measured by unexpected changes in global temperature, can increase the market share of green investors who believe in a larger impact of climate change on the economy. This increased market share drives up the demand for green stocks and hence leads to higher realized returns for green stocks relative to brown stocks.

These models thus give possible explanations for how past observed better green asset performance could be consistent with a positive carbon risk premium. Over the past decade, growing attention regarding increasingly salient evidence of the risks of climate change may have devalued prices of brown stocks and increased the carbon premium. However, the persistent and strong green outperformance in our data suggest that any carbon premium during our sample was unlikely very large. In the presence of a sizable carbon premium, very large shocks in favor of green stock prices would have been needed to more than offset the higher expected returns of brown stocks.[23]

More data and research are needed to draw a complete picture of the relation between expected and realized returns of green and brown assets. But one recent example that highlights the crucial role of unexpected shocks for the performance of green and brown stocks is the energy crisis just before and following the Russian invasion of Ukraine in early

[23] Relatedly, Lontzek et al. (2022) argue that until climate tipping points have been crossed investors require only a modest carbon premium for holding brown assets.

2022. As noted earlier, Figure 4 shows the performance of the brown-minus-green carbon emissions portfolio for the G7 countries until June 2022. The vertical dashed black line marks the beginning of 2022 and the subsequent energy crisis. Corresponding results using carbon intensity are shown in Figure 5. Using either the level of carbon emissions or carbon intensity as a measure for greenness, the first half of 2022 shows much stronger returns for brown over green stocks. The retrenchment in cumulative gains is consistent across all G7 countries and is sizable, though not completely reversing earlier green gains.

At this point (in late 2022), it is hard to anticipate if this episode will leave a transitory imprint—similar to 2016—or portends a more permanent shift. The extraordinary energy and defense crisis associated with the Russian invasion of the Ukraine represents a major unexpected shock to the economy that is still unfolding and one that investors are still adapting to. There are various reasons why this shock may have led to an outperformance of brown over green stocks. Foremost, of course, is the increased demand for products from mostly brown energy producers and the defense industry, two particularly high carbon emitters. In addition, investor preferences for green assets may have declined at the same time. For example, Bansal et al. (2021b) argue that green assets could underperform in times of crisis because social responsible investing (SRI) is a luxury good and hence depends on wealth. While it is too early to draw final conclusions regarding the impact of the energy crisis on green and brown assets, it is an important task for future research.

8 Conclusion

This paper documents a persistent outperformance of green over brown stocks since 2012. We measure greenness using the level or intensity of CO_2 emissions as reported by emitters—not imputed or estimated by third-party providers. Green portfolios based on either measure generally exhibited higher returns than brown portfolios, and the green return advantage is larger using emission intensity. The historical pattern of higher returns for green over brown stocks is evident not only for the United States but also for almost all other G7 countries, with Italy being the exception. However, the reversal of this pattern in 2022, with much stronger returns for brown over green stocks, also introduces a major area of investigation.

Our finding of higher realized returns for green stocks is at least a partial challenge to the hypothesis of a carbon premium, which implies that expected returns are higher for brown stocks. One way to reconcile our evidence with the carbon premium is to assume that this premium has gradually increased over the course of our sample, possibly due to changes in risk perceptions or risk pricing. We leave it to future research to assess whether such divergence between realized and expected returns is plausible.

References

Ardia, David, Keven Bluteau, Kris Boudt, and Koen Inghelbrecht (2022) "Climate change concerns and the performance of green versus brown stocks," forthcoming in *Management Science*.

Aswani, Jitendra, Aneesh Raghunandan, and Shivaram Rajgopal (2022) "Are carbon emissions associated with stock returns?" research paper, Columbia Business School.

Bansal, Ravi, Dana Kiku, and Marcelo Ochoa (2021a) "Climate Change Risk," Working paper.

Bansal, Ravi, Di Wu, and Amir Yaron (2021b) "Socially Responsible Investing in Good and Bad Times," *The Review of Financial Studies*, 35 (4), 2067–2099, 10.1093/rfs/hhab072.

Banz, Rolf W (1981) "The relationship between return and market value of common stocks," *Journal of Financial Economics*, 9 (1), 3–18.

van der Beck, Philippe (2022) "Flow-Driven ESG Returns," Working paper.

Berg, Florian, Julian F Kölbel, and Roberto Rigobon (2022) "Aggregate Confusion: The Divergence of ESG Ratings," *Review of Finance*, forthcoming.

BNP Paribas (2019) "The ESG Global Survey," Available at `https://securities.cib.bnpparibas/the-esg-global-survey-2019/` (1/2/2023).

Bolton, Patrick, Zachery Halem, and Marcin Kacperczyk (2022) "The Financial Cost of Carbon," *Journal of Applied Corporate Finance*, 34 (2), 17–29.

Bolton, Patrick and Marcin Kacperczyk (2021) "Do investors care about carbon risk?" *Journal of financial economics*, 142 (2), 517–549.

——— (2022) "Global Pricing of Carbon-Transition Risk," July, forthcoming in *Journal of Finance*.

Busch, Timo, Alexander Bassen, Stefan Lewandowski, and Franziska Sump (2022a) "Corporate carbon and financial performance revisited," *Organization & Environment*, 35 (1), 154–171.

Busch, Timo, Matthew Johnson, and Thomas Pioch (2022b) "Corporate carbon performance data: Quo vadis?" *Journal of Industrial Ecology*, 26 (1), 350–363.

Cheema-Fox, Alexander, Bridget Realmuto LaPerla, George Serafeim, David Turkington, and Hui Stacie Wang (2021) "Decarbonization factors," *Journal of Impact and ESG Investing.*

Cohen, Lauren and Andrea Frazzini (2008) "Economic links and predictable returns," *The Journal of Finance*, 63 (4), 1977–2011.

De Bondt, Werner FM and Richard Thaler (1985) "Does the stock market overreact?" *Journal of Finance*, 40 (3), 793–805.

Delmas, Magali A, Nicholas Nairn-Birch, and Jinghui Lim (2015) "Dynamics of environmental and financial performance: The case of greenhouse gas emissions," *Organization & Environment*, 28 (4), 374–393.

El Ghoul, Sadok, Omrane Guedhami, Chuck CY Kwok, and Dev R Mishra (2011) "Does corporate social responsibility affect the cost of capital?" *Journal of banking & finance*, 35 (9), 2388–2406.

Engle, Robert F, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel (2020) "Hedging climate change news," *The Review of Financial Studies*, 33 (3), 1184–1216.

Fama, Eugene F. and Kenneth R. French (1993) "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33 (1), 3 – 56, https://doi.org/10.1016/0304-405X(93)90023-5.

Fong, Kingsley Y L, Craig W Holden, and Charles A Trzcinka (2017) "What Are the Best Liquidity Proxies for Global Research?" *Review of Finance*, 21 (4), 1355–1401, 10.1093/rof/rfx003.

Garvey, Gerald T., Mohanaraman Iyer, and Joanna Nash (2018) "Carbon footprint and productivity: does the "E" in ESG capture efficiency as well as environment," *Journal of Investment Management*, 16 (1), 59–69.

Gibson Brandon, Rajna, Philipp Krueger, and Peter Steffen Schmidt (2021) "ESG Rating Disagreement and Stock Returns," *Financial Analysts Journal*, 77 (4), 104–127, 10.1080/0015198X.2021.1963186.

Görgen, Maximilian, Andrea Jacob, Martin Nerlinger, Ryan Riordan, Martin Rohleder, and Marco Wilkens (2020) "Carbon risk," working paper.

Griffin, John M., Patrick J. Kelly, and Federico Nardari (2010) "Do Market Efficiency Measures Yield Correct Inferences? A Comparison of Developed and Emerging Markets," *The Review of Financial Studies*, 23 (8), 3225–3277, 10.1093/rfs/hhq044.

Hou, Kewei, Chen Xue, and Lu Zhang (2020) "Replicating Anomalies," *The Review of Financial Studies*, 33 (5), 2019–2133, 10.1093/rfs/hhy131.

Huij, Joop, Dries Laurs, Philip A. Stork, and Remco C. J. Zwinkels (2021) "Carbon Beta: A Market-Based Measure of Climate Risk," working paper.

Ilhan, Emirhan, Zacharias Sautner, and Grigory Vilkov (2021) "Carbon tail risk," *The Review of Financial Studies*, 34 (3), 1540–1571.

In, Soh Young, Ki Young Park, and Ashby Monk (2019) "Is 'Being Green' Rewarded in the Market? An Empirical Investigation of Decarbonization and Stock Returns," working paper, Stanford Global Project Center.

Ince, Ozgur S. and R. Burt Porter (2006) "Individual equity return data from Thomson Datastream: Handle with care!," *Journal of Financial Research*, 29 (4), 463–479.

Jacobs, Heiko (2016) "Market maturity and mispricing," *Journal of Financial Economics*, 122 (2), 270 – 287, https://doi.org/10.1016/j.jfineco.2016.01.030.

Karolyi, G. Andrew, Kuan-Hui Lee, and Mathijs A. van Dijk (2012) "Understanding commonality in liquidity around the world," *Journal of Financial Economics*, 105 (1), 82–112, https://doi.org/10.1016/j.jfineco.2011.12.008.

Lontzek, Thomas, Walter Pohl, Karl Schmedders, Marco Thalhammer, and Ole Wilms (2022) "Asset Pricing with Disagreement about Climate Risks," Working paper.

MSCI (2022) "MSCI Low Carbon Indexes,"Technical report, available at `https://www.msci.com/low-carbon-indexes`.

NGFS (2021) "NGFS Climate Scenarios for central banks and supervisors," Available at `https://www.ngfs.net/en/ngfs-climate-scenarios-central-banks-and-supervisors-june-2021` (1/4/2023).

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor (2021) "Sustainable investing in equilibrium," *Journal of Financial Economics*, 142 (2), 550–571.

——— (2022) "Dissecting green returns," *Journal of Financial Economics*, 146 (2), 403–424.

Plumlee, Marlene, Darrell Brown, Rachel M Hayes, and R Scott Marshall (2015) "Voluntary environmental disclosure quality and firm value: Further evidence," *Journal of accounting and public policy*, 34 (4), 336–361.

Rohleder, Martin, Marco Wilkens, and Jonas Zink (2022) "The effects of mutual fund decarbonization on stock prices and carbon emissions," *Journal of Banking & Finance*, 134, 106352.

Sautner, Zacharias, Laurence van Lent, Grigory Vilkov, and Ruishen Zhang (2022a) "Firm-level Climate Change Exposure," *Journal of Finance*, forthcoming.

——— (2022b) "Pricing Climate Change Exposure," *Management Science*, forthcoming.

S&P Global (2020) "Index Carbon Metrics Explained,"Technical report, available at https://www.spglobal.com/spdji/en/documents/additional-material/spdji-esg-carbon-metrics.pdf.

Zhang, Shaojun (2022) "Carbon Premium: Is It There?" Working Paper 2022-03-006, Fisher College of Business.

Table 1: Summary statistics for the G7 countries (total)

	Mean	Median	25th	75th	Obs.
Scope 1	2,921,550	66,653	9,008	526,035	12,674
Scope 2	561,833	97,130	21,327	402,473	12,458
Scope 1+2 level	3,425,962	222,858	42,860	1,193,117	12,851
Scope 1+2 intensity	0.31	0.03	0.00	0.12	12,820
Market cap	8,036	1,375	329	4,976	50,636

Summary statistics—mean, median, 25th quantile, 75th quantile, and number of observations—for emission variables and market capitalization. Emissions are measured in tons of CO_2, emission intensity is measured in tons of CO_2 divided by thousands of USD in firm revenues, and market capitalization is measured in millions of USD. Sample period: January 2010 to December 2021.

Table 2: Average monthly returns in the US

Factor/Portfolio	brown	green	brown-green
Panel A: Level of emissions			
Size-adjusted spread	1.15	1.40	−0.25
	(3.17)	(3.45)	(−1.30)
Simple spread, value-weighted	1.02	1.47	−0.45
	(3.27)	(3.63)	(−1.84)
Simple spread, equally-weighted	1.15	1.37	−0.22
	(2.93)	(3.06)	(−1.00)
Panel B: Related literature			
Size-adjusted spread - Huij et al.	1.22	1.57	−0.34
	(2.75)	(3.51)	(−1.64)
Simple spread - Pastor et al.	0.84	1.28	−0.45
	(2.13)	(3.49)	(−2.61)
Panel C: Emission intensity			
Size-adjusted spread	1.04	1.37	−0.33
	(2.68)	(3.27)	(−1.70)
Simple spread, value-weighted	0.79	1.45	−0.65
	(2.11)	(3.38)	(−2.37)
Simple spread, equally-weighted	1.24	1.43	−0.18
	(2.65)	(3.25)	(−0.75)

Monthly average returns for spread portfolios as well as the corresponding long (brown) and short (green) portfolios, based on the level or intensity of CO_2 emissions. In parentheses are t-statistics based on heteroskedasticity-robust (White) standard errors. Panels A and C show results for the size-adjusted spread, a simple value-weighted quintile return spread, and and a simple equal-weighted quintile return spread. Panel B shows results for the size-adjusted spread by Huij et al. (2021) based on estimated emission levels and the simple tertile spread based on E-scores of Pástor et al. (2022). Sample period: January 2010 to December 2021.

Table 3: Portfolio returns and Sharpe ratios for U.S. and G7

	Quintiles					brown-	
	1 (green)	2	3	4	5 (brown)	green	Market
United States Panel A: Level of emissions							
Mean return	1.47	1.26	1.32	1.31	1.02	−0.45	1.27
	(3.63)	(3.27)	(3.60)	(3.87)	(3.27)	(−1.84)	(3.69)
Volatility	4.86	4.66	4.41	4.07	3.75	2.92	4.13
Sharpe ratio	0.29	0.26	0.29	0.31	0.26	−0.17	0.30
Panel B: Emission intensity							
Mean return	1.45	1.31	1.17	1.04	0.79	−0.65	1.27
	(3.38)	(4.08)	(3.96)	(3.05)	(2.11)	(−2.37)	(3.69)
Volatility	5.16	3.86	3.57	4.10	4.53	3.33	4.13
Sharpe ratio	0.27	0.33	0.32	0.24	0.17	−0.21	0.30
G7 average Panel C: Level of emissions							
Mean return	0.80	0.91	0.82	0.75	0.59	−0.20	1.05
	(1.84)	(2.21)	(1.98)	(1.83)	(1.54)	(−1.20)	(3.07)
Volatility	5.22	4.97	4.96	4.93	4.66	2.05	4.11
Sharpe ratio	0.15	0.18	0.16	0.14	0.12	−0.12	0.25
Panel D: Emission intensity							
Mean return	0.80	0.86	0.74	0.71	0.48	−0.33	1.05
	(1.71)	(2.10)	(2.01)	(1.86)	(1.16)	(−1.60)	(3.07)
Volatility	5.66	4.93	4.45	4.61	4.93	2.48	4.11
Sharpe ratio	0.14	0.17	0.16	0.15	0.09	−0.15	0.25

Monthly average returns, t-statistics (in parentheses) based on heteroskedasticity-robust (White) standard errors, volatilities, and Sharpe ratios for value-weighted quintile portfolios sorted according to the level and intensity of carbon emissions, the "simple spread" portfolio, and the total market. Panels A–B show results for the United States, panels C–D show results for the average across the G7 countries. Sample period: January 2010 to December 2021.

Table 4: Correlations of monthly spread returns in the U.S.

	Size-adjusted spread	Simple spread, VW	Simple spread, EW
Panel A: Correlations among our spreads based on emission levels			
Size-adjusted spread	1		
Simple spread, VW	0.82	1	
Simple spread, EW	0.87	0.76	1
Panel B: Correlations with spreads from the literature			
Size-adjusted spread - Huij et al.	0.72	0.62	0.71
Simple spread - Pastor et al.	0.46	0.32	0.31
Panel C: Correlations with spreads based on emission intensity (EI)			
Size-adjusted spread, EI	0.84	0.64	0.77
Simple spread, VW, EI	0.67	0.57	0.64
Simple spread, EW, EI	0.77	0.50	0.77

Correlations between different spread portfolios, based on either emission levels or emission intensities (EI), using three different methods: size-adjusted spread, a simple value-weighted (VW) quintile return spread, and a simple equal-weighted (EW) quintile return spread. Also shown are correlations with spreads from the literature: the size-adjusted spread by Huij et al. (2021) based on emission levels, and the spread based on E-scores of Pástor et al. (2022). Sample period: January 2010 to December 2021.

Table 5: Average monthly returns in the G7 countries

	Level of emissions			Emission intensity		
Factor/Portfolio	brown	green	brown-green	brown	green	brown-green
Canada	0.65	0.89	−0.24	0.46	0.84	−0.38
	(1.38)	(1.98)	(−0.93)	(0.93)	(1.86)	(−1.25)
France	0.31	1.06	−0.75	0.40	1.12	−0.71
	(0.64)	(1.91)	(−2.45)	(0.82)	(1.91)	(−2.43)
Germany	0.49	0.45	0.04	0.09	0.56	−0.47
	(0.85)	(0.96)	(0.11)	(0.16)	(1.05)	(−1.34)
Italy	0.64	0.15	0.49	0.68	0.35	0.33
	(1.16)	(0.19)	(0.99)	(1.18)	(0.44)	(0.71)
Japan	0.61	0.58	0.03	0.42	0.66	−0.24
	(1.84)	(1.50)	(0.14)	(1.15)	(1.53)	(−0.83)
United Kingdom	0.44	0.98	−0.54	0.47	0.65	−0.18
	(1.15)	(2.05)	(−1.91)	(0.96)	(1.19)	(−0.52)
United States	1.02	1.47	−0.45	0.79	1.45	−0.65
	(3.27)	(3.63)	(−1.84)	(2.11)	(3.38)	(−2.37)
G7 average	0.59	0.80	−0.20	0.48	0.80	−0.33
	(1.54)	(1.84)	(−1.20)	(1.16)	(1.71)	(−1.60)

Average monthly returns for the (value-weighted) simple spread as well as the corresponding long (brown) and short (green) portfolios, based on either the level or intensity of CO_2 emissions. Results are shown for all G7 countries and an (equal-weighted) average across these seven countries. In parentheses are t-statistics based on heteroskedasticity-robust (White) standard errors. Sample period: January 2010 to December 2021.

Table 6: Panel regressions based on brown-green allocations (United States)

	Level of emissions			Emission intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.537***			1.535***		
	(0.05)			(0.05)		
Brown-green indicator	−0.130*	−0.146	−0.046	−0.062	−0.315***	−0.194*
	(0.07)	(0.10)	(0.11)	(0.08)	(0.10)	(0.11)
Book-to-market		0.240	0.218		0.191	0.190
		(0.21)	(0.22)		(0.21)	(0.22)
Sales growth		0.000	0.000		0.000	0.000
		(0.00)	(0.00)		(0.00)	(0.00)
Log(PPEGT)		0.050	0.199***		0.102*	0.235***
		(0.06)	(0.07)		(0.06)	(0.07)
Leverage		0.014	0.006		0.007	0.003
		(0.02)	(0.02)		(0.02)	(0.02)
$r(t-1)$		−0.036***	−0.036***		−0.035***	−0.036***
		(0.01)	(0.01)		(0.01)	(0.01)
Log(market cap)		−0.201***	−0.347***		−0.270***	−0.383***
		(0.06)	(0.08)		(0.07)	(0.08)
$r(t-12, t-2)$		−0.003*	−0.003*		−0.003	−0.003*
		(0.00)	(0.00)		(0.00)	(0.00)
ROE		0.053	0.012		0.048	0.011
		(0.24)	(0.24)		(0.23)	(0.24)
Investment-to-assets		−0.238	−0.147		−0.222	−0.148
		(0.30)	(0.30)		(0.30)	(0.30)
Number of observations	47418	47418	47418	47400	47400	47400
R^2	0.00	0.27	0.27	0.00	0.27	0.27
Time fixed effects	No	Yes	Yes	No	Yes	Yes
Industry fixed effects	No	No	Yes	No	No	Yes

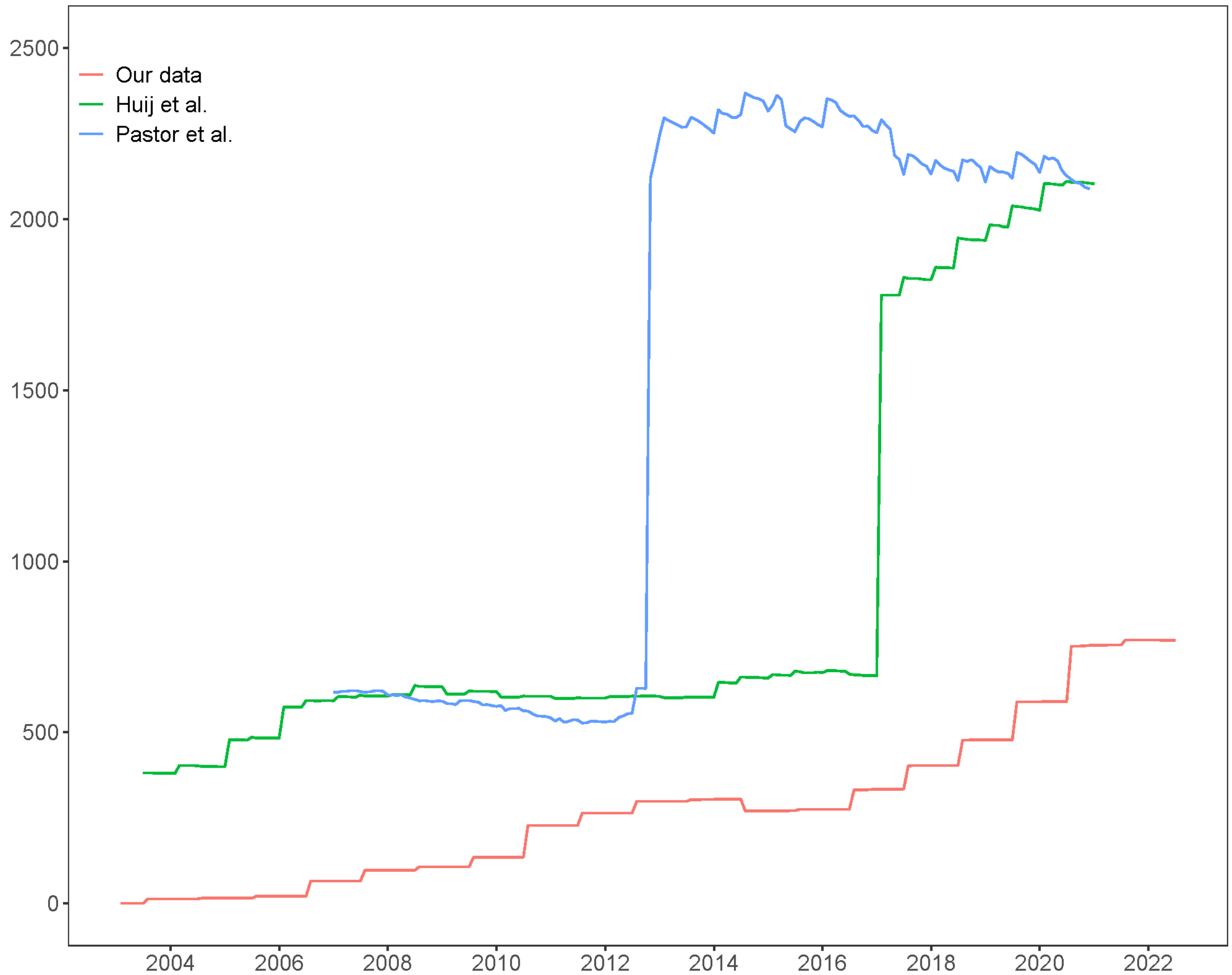
Panel regressions for monthly stock returns on an indicator variable ("Brown-green"), which indicates if a stock is in the "green" portfolio (−1), or in the "brown" portfolio (+1), or anywhere between (0), based on quintile portfolio sorts using either the level (columns 1-3) or intensity (columns 4-6) of CO_2 emissions. Returns are mapped to emissions data using an 18-month publication lag. Controls are book-to-market, sales growth, log property plant and equipment (PPEGT), leverage, the past 1-month return, log market capitalization of the previous month, the cumulative past return from $t-12$ to $t-2$, return on equity (ROE), and investment-to-assets. Standard errors are clustered by time and given in parentheses. ***, **, and * indicate statistical significance at the 1%-, 5%-, and 10%-level, respectively. Sample period: January 2010 to December 2021.

Table 7: Panel regressions based on CO_2 emissions (United States)

	Level of emissions			Emission intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	2.775***			1.530***		
	(0.27)			(0.05)		
Emissions measure	−0.094***	−0.057*	−0.022	0.000	−0.019	0.054
	(0.02)	(0.03)	(0.04)	(0.06)	(0.05)	(0.06)
Book-to-market		0.277	0.248		0.285	0.239
		(0.23)	(0.24)		(0.23)	(0.24)
Sales growth		0.000	0.000		0.000	0.000
		(0.00)	(0.00)		(0.00)	(0.00)
Log(PPEGT)		0.067	0.203**		−0.004	0.172***
		(0.07)	(0.08)		(0.05)	(0.06)
Leverage		0.019	0.010		0.020	0.009
		(0.02)	(0.02)		(0.02)	(0.02)
$r(t-1)$		−0.036***	−0.036***		−0.036***	−0.036***
		(0.01)	(0.01)		(0.01)	(0.01)
Log(market cap)		−0.192***	−0.335***		−0.175***	−0.323***
		(0.06)	(0.08)		(0.06)	(0.07)
$r(t-12, t-2)$		−0.002	−0.002		−0.002	−0.002
		(0.00)	(0.00)		(0.00)	(0.00)
ROE		−0.075	−0.094		−0.085	−0.094
		(0.14)	(0.14)		(0.14)	(0.14)
Investment-to-assets		−0.174	−0.134		−0.169	−0.131
		(0.19)	(0.19)		(0.19)	(0.19)
Number of observations	47388	47388	47388	47388	47388	47388
R^2	0.00	0.27	0.27	0.00	0.27	0.27
Time fixed effects	No	Yes	Yes	No	Yes	Yes
Industry fixed effects	No	No	Yes	No	No	Yes

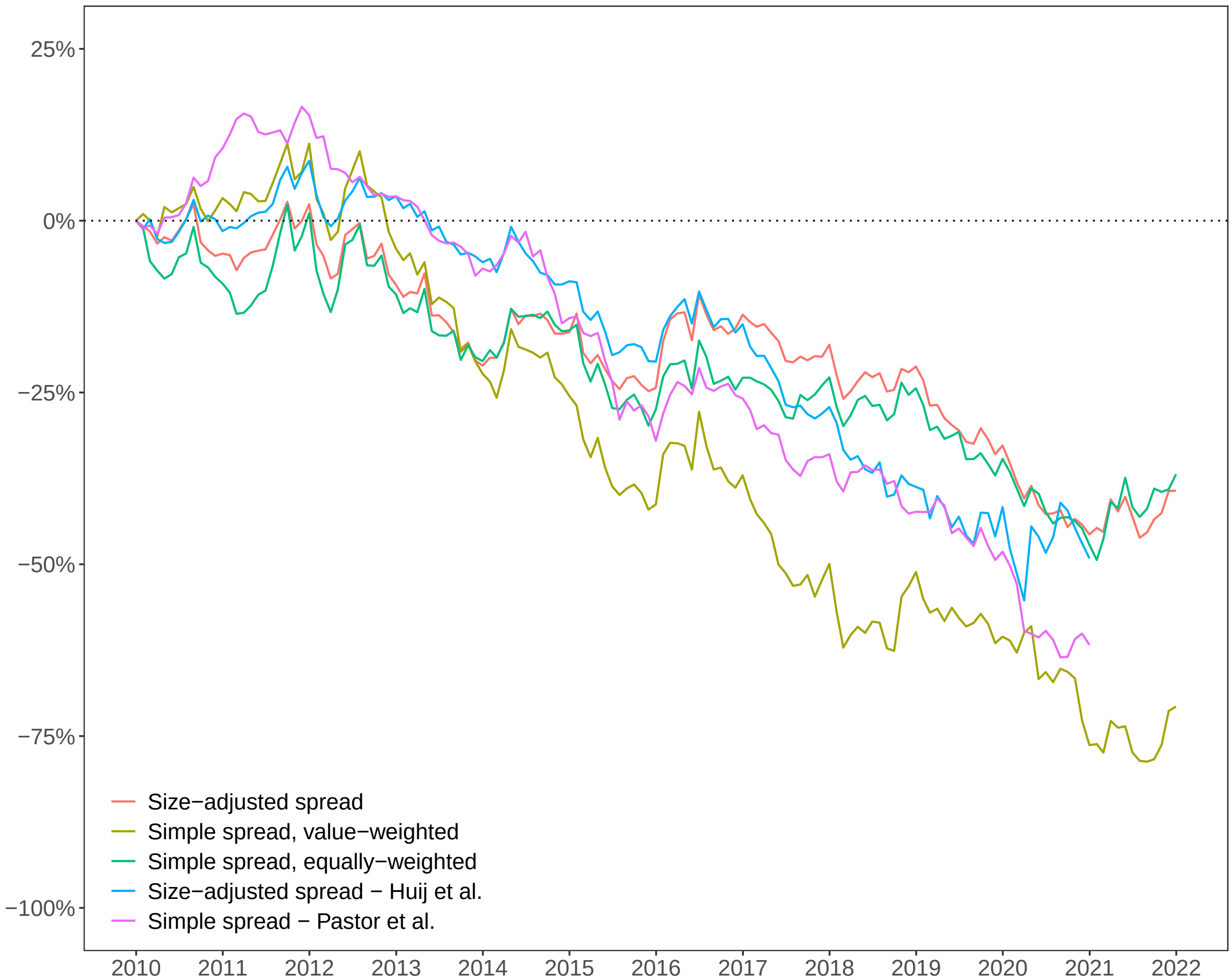
Panel regressions of monthly stock returns on two different emission measures: the (log) level of emissions (columns 1–3) and emission intensity (columns 4–6). Controls are book-to-market, sales growth, log property plant and equipment (PPEGT), leverage, the past 1-month return, log market capitalization of the previous month, the cumulative past return from $t-12$ to $t-2$, return on equity (ROE), and investment-to-assets. Standard errors in all cases are clustered by time and given in parentheses. ***, **, and * indicate significance at the 1%-, 5%-, and 10%-level, respectively. Sample period: January 2010 to December 2021.

Figure 1: Number of firms



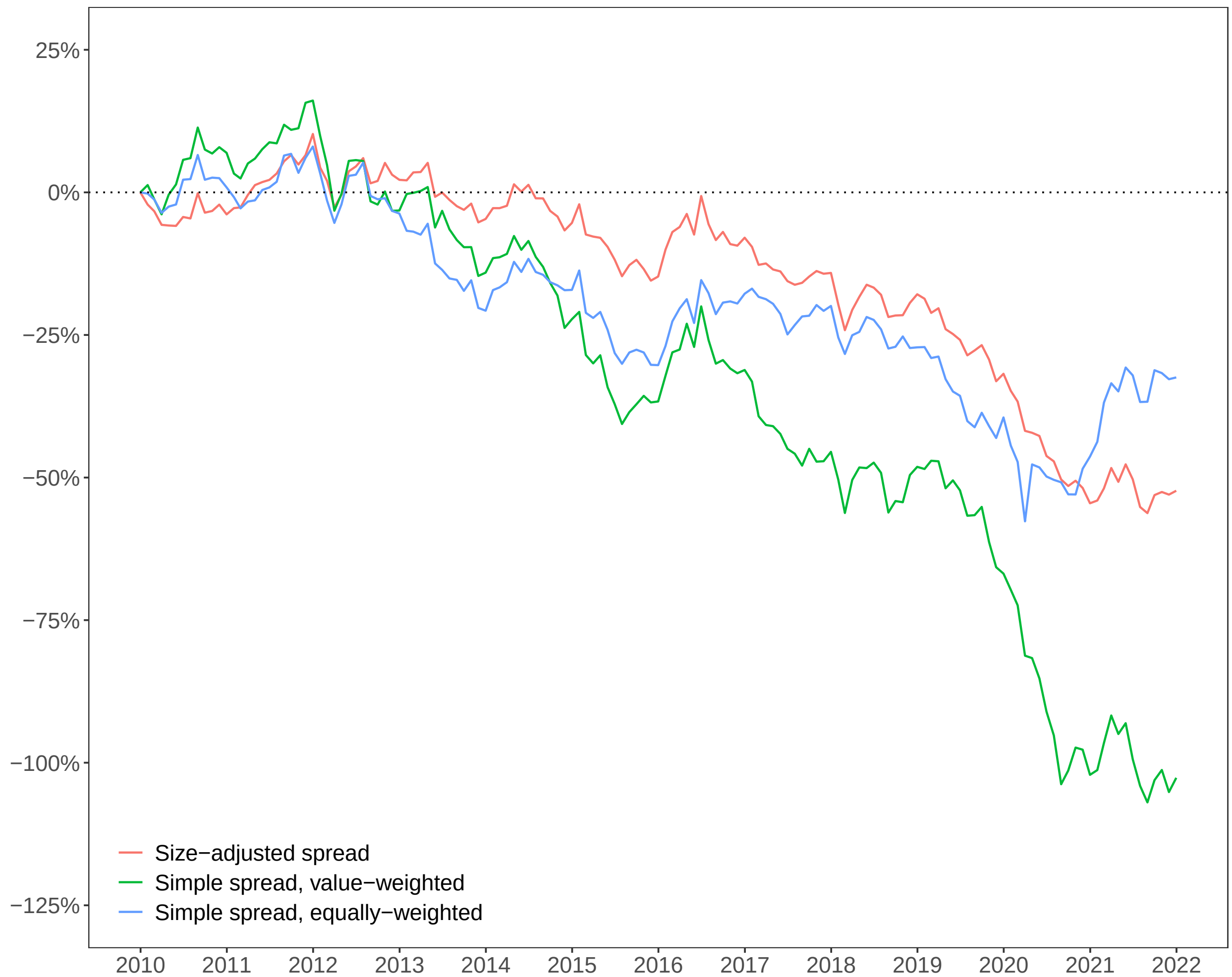
Monthly number of firms in three data samples: our own data, which includes firms with reported emissions data from Refinitiv (using an 18-month publication lag), the data used by Huij et al. (2021), which includes firms with reported and estimated emissions data from Trucost, and the data used by Pástor et al. (2022), which includes firms with E-scores from MSCI.

Figure 2: Brown vs. green performance in the U.S., total emissions



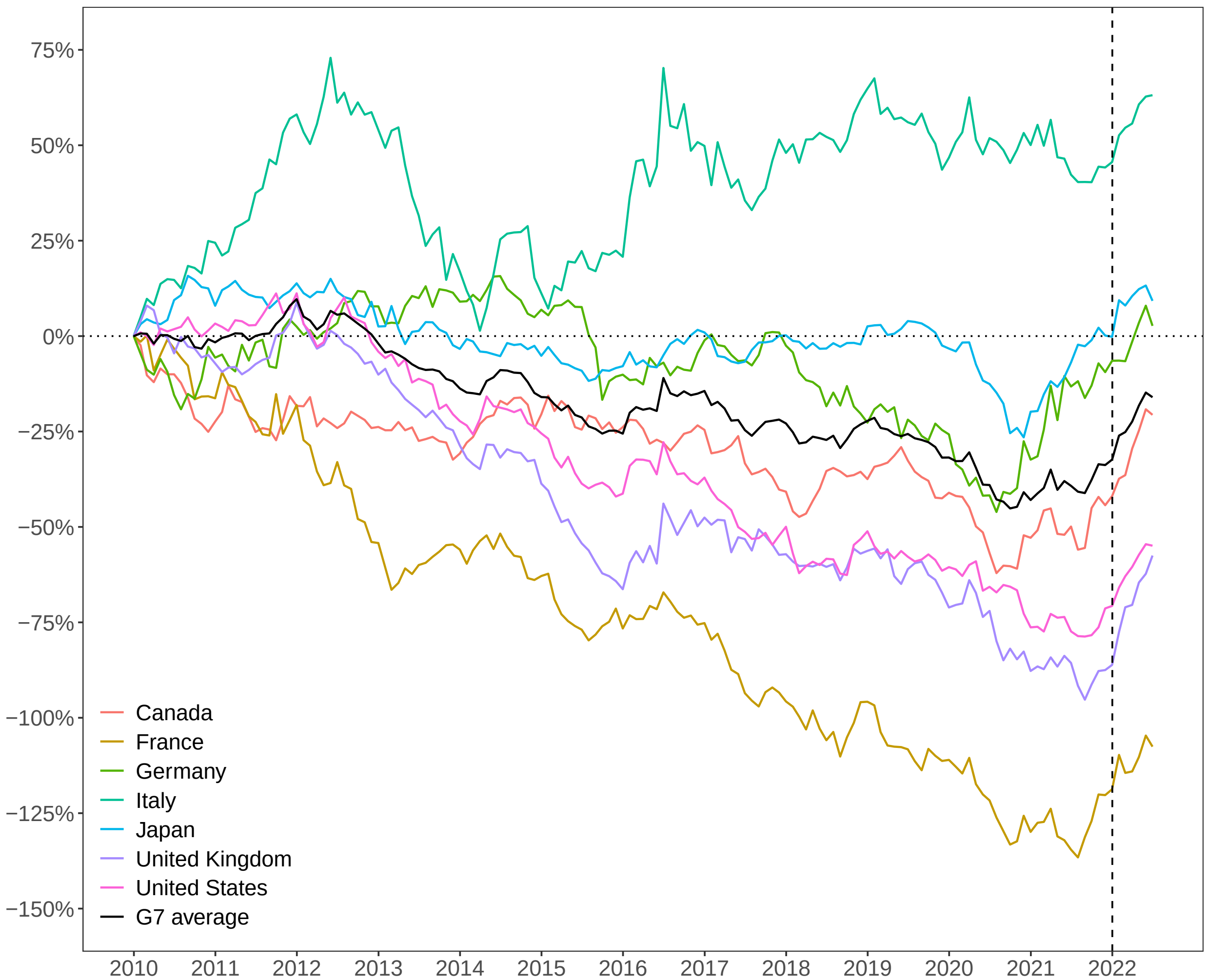
Cumulative log returns for different brown minus green spread portfolios, based on the level of CO_2 emissions: a size-adjusted spread, a simple value-weighted quintile return spread, and a simple equal-weighted quintile return spread. Also shown are cumulative returns for the size-adjusted spread of Huij et al. (2021), based on estimated emissions, and the simple spread of Pástor et al. (2022), based on E-scores. Sample period: January 2010 to December 2021.

Figure 3: Brown vs. green performance in the U.S., emission intensity



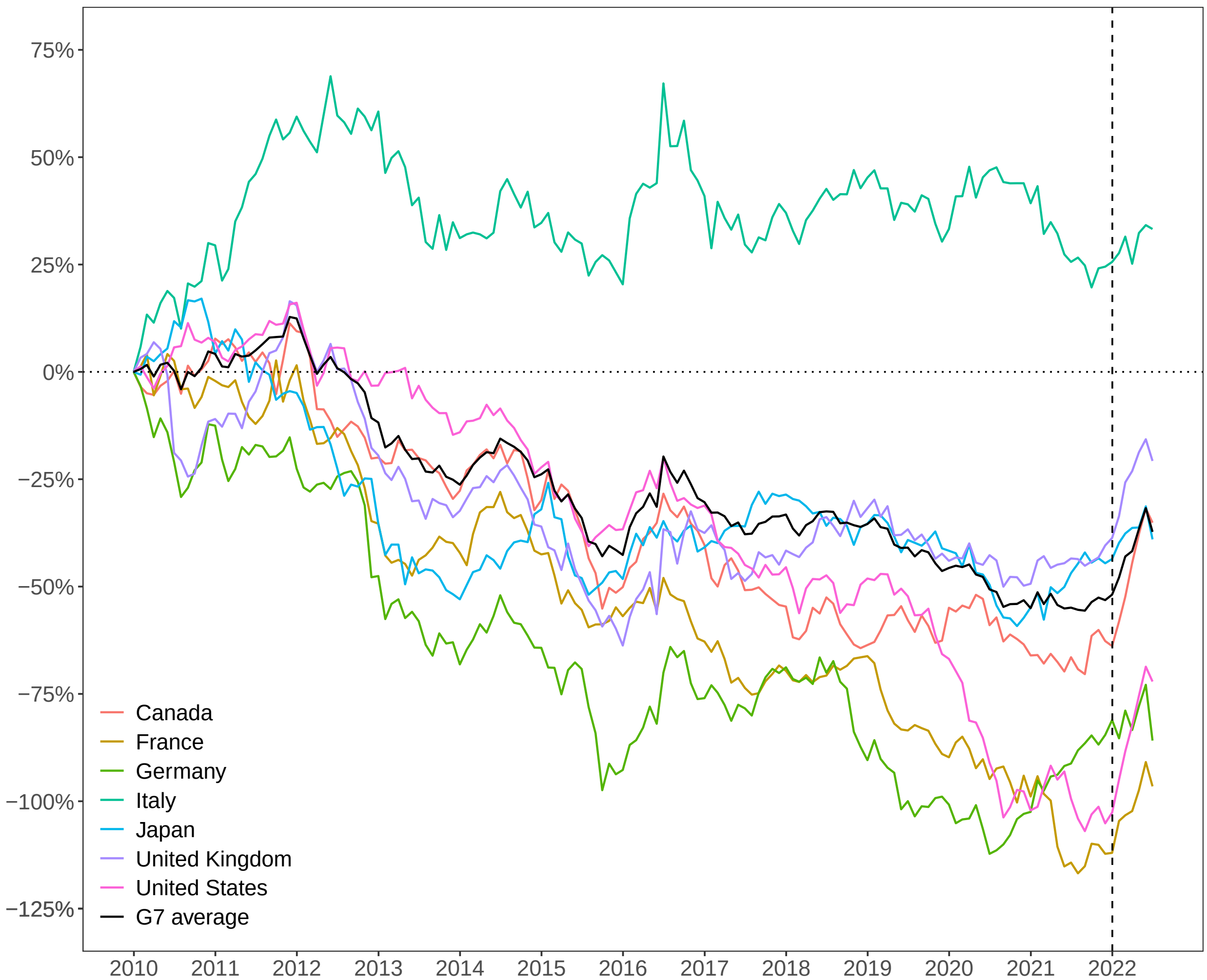
Cumulative log returns for different brown minus green spread portfolios, based on CO_2 emissions scaled by total scales: a size-adjusted spread, a simple value-weighted quintile return spread, and a simple equal-weighted quintile return spread. Sample period: January 2010 to December 2021.

Figure 4: Brown vs. green performance in the G7 countries, total emissions



Cumulative log returns for value-weighted simple spreads, based on emission levels, for all G7 countries as well as the average across these countries. Sample period: January 2010 to June 2022. Vertical dashed line marks the beginning of 2022, after the end of our baseline sample period.

Figure 5: Brown vs. green performance in the G7 countries, emission intensity



Cumulative log returns for value-weighted simple spreads, based on emission intensities, for all G7 countries as well as the average across these countries. Sample period: January 2010 to June 2022. Vertical dashed line marks the beginning of 2022, after the end of our baseline sample period.

Appendix

A Performance and factor risks

In the following we estimate the exposure of our brown-minus-green spread portfolios to standard risk factors. To this end, we estimate a Fama-French three-factor model for each of the spread return series in the United States.

The details for this analysis are as follows: We require a non-missing market capitalization in June of year t, a non-missing market capitalization in December of year $t-1$, and positive book equity at the fiscal year ending in $t-1$ for the construction of the SMB (small-minus-big) and HML (high-minus-low valuation). More specifically, at the end of June of every year, we classify firms into two size groups: "small" and "big", based on the median market capitalization. Independently, we divide firms into three value groups: "low", "medium", and "high" based on the 30th and 70th percentiles of book-to- market, the ratio of book equity and market equity, with the former at the fiscal year ending in calendar year $t-1$ and the latter at the end of December of year $t-1$. At the intersections of the two size and the three value groups, six portfolios are created and the associated value-weighted or equally-weighted returns are computed from July of year t until June of year $t+1$, when a new sorting process takes place. SMB is constructed each month as the difference in average returns between the three small stock portfolios and the three big stock portfolios. HML is created every month as the difference in average returns between the two high value portfolios and the two low value portfolios. In addition, RMRF is computed every month as the value-weighted market return less the risk-free rate. We use the one-month US Treasury bill rate as the risk-free rate, taking the perspective of a US investor. In the construction of RMRF, we include all stocks in the US that represent common equity. In addition, for the computation of the alphas associated with the size-adjusted spreads of Huij et al. (2021) and Pástor et al. (2022), we use the three factors for the US provided by Ken French.[24]

Table A.1 reports exposures of the different spread portfolios to the three Fama-French risk factors. All brown-minus-green (BMG) spread portfolio have a large and significant exposure to HML. This positive exposure to the value factor is consistent with the notion that brown stocks are on average value stocks. The increased demand for green stocks reflected in the outperformance of our BMG portfolios might be one of the reasons why the HML factor has performed significantly worse in our sample period from 2010-2021 compared to the decades before.

[24]For more info, see: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html (accessed 12/26/2022).

Table A.1: Alphas

	Intercept	RMRF	SMB	HML
Panel A: Level of emissions				
Size-adjusted spread	0.02	−0.15	−0.11	0.27
	(0.12)	(−2.98)	(−1.91)	(4.35)
Simple spread, value-weighted	−0.04	−0.25	−0.22	0.26
	(−0.18)	(−4.28)	(−3.05)	(3.42)
Simple spread, equally-weighted	−0.12	−0.01	−0.16	0.37
	(−0.59)	(−1.22)	(−2.22)	(5.52)
Panel B: Related literature				
Size-adjusted spread - Huij et al.	−0.23	0.01	−0.32	0.23
	(−1.32)	(0.13)	(−4.60)	(2.79)
Simple spread, E-score - Pastor et al.	−0.32	−0.03	0.14	0.23
	(−1.79)	(−0.67)	(1.71)	(3.31)
Panel C: Emission intensity				
Size-adjusted spread	−0.05	−0.18	0.03	0.27
	(−0.27)	(−3.95)	(0.46)	(3.83)
Simple spread, value-weighted	−0.24	−0.26	0.10	0.45
	(−0.94)	(−3.88)	(1.33)	(4.73)
Simple spread, equally-weighted	−0.25	−0.01	0.19	0.51
	(−1.16)	(−2.50)	(2.11)	(6.14)

Alphas (i.e., intercepts) in % and coefficients from regressing the respective returns of the spread portfolios (RET) on the excess return of the market (RMRF), small minus big (SMB), and high minus low (HML), relying on the following model: $RET_{it} = \alpha_i + \beta_i \cdot RMRF_t + s_i \cdot SMB_t + h_i \cdot HML_t + \epsilon_{it}$. The t-statistics are based on heteroskedasticity-consistent (White) standard errors, reported in parentheses. Sample period: January 2010 to December 2021.

Exposures to SMB are negative for the spread portfolios based on the level of emissions. Hence, brown firms based on emission levels are large firms on average. This is expected as this measure for greenness does not account for firm size which we argue in Section 2 is one of the main weaknesses of using emission levels as a proxy for greenness. In turn, the spread portfolios based on emission intensities have positive exposures to SMB. Hence, using this measure for greenness, green firms are large firms on average, although this effect is only significant for the equally-weighted spread at the 5% level.

In line with the reported outperformance of green stocks, all intercepts, except for the size-adjusted spread based on emission levels are negative, although insignificant. This also holds for the spread portfolios by Huij et al. (2021) and Pástor et al. (2022).[25] However, as outlined above, the increased demand for green stocks could have negatively affected the performance of HML and hence, the outperformance might not show up in significant alphas but is simply reflected in the positive exposure to HML. We leave it for future research to further analyze the impact of climate related risks on the standard risk factors.

B Mean returns based on estimated emissions

In the following, we analyze the relative performance of green and brown stock portfolios in the United States using both, reported and estimated firm-level emissions from Refinitiv. Over the period from 2010 to 2021, the share of firms that report emissions was around 40-45%. Hence, by also using estimated emissions, the number of firms more than doubles. Table B.1 shows monthly average returns for spread portfolios as well as the corresponding long (brown) and short (green) portfolios, based on the level or intensity of estimated and reported CO_2 emissions. The results for emission intensity are very similar, both qualitatively and quantitatively, to the results using only reported emissions (see Table 2) with an outperformance of green over brown stocks between 0.23% and 0.65% per month for the different spread portfolios.

In contrast, for emission levels we do not find a green outperformance when using both, estimated and reported emissions to construct the spread portfolios. In turn, Huij et al. (2021) report a green outperformance for their size-adjusted spread which they construct using both, reported and estimated emission levels provided by Trucost. The difference in the results might be explained by the different proprietary models of Trucost and Refinitiv to estimate firm-level emissions.[26] Hence, to establish our main results we use only reported

[25] Note that Pástor et al. (2022) report a larger and statistically significant alpha in their paper which can be explained by the shorter sample period used in their paper which only starts in 2012.

[26] The difference might also arise because the coverage of firms with reported emissions might be different for the two data providers.

emissions so that our results are not dependent on model-based estimates of firm-level emissions.

Table B.1: Average monthly returns in the US based on reported and estimated emissions

Factor/Portfolio	brown	green	brown-green
Panel A: Level of emissions			
Size-adjusted spread	1.31	1.26	0.05
	(3.15)	(2.86)	(0.21)
Simple spread, value-weighted	1.16	0.95	0.21
	(3.61)	(2.10)	(0.77)
Simple spread, equally-weighted	1.19	1.30	−0.11
	(2.74)	(2.77)	(−0.41)
Panel B: Emission intensity			
Size-adjusted spread	1.04	1.39	−0.35
	(2.52)	(3.42)	(−2.06)
Simple spread, value-weighted	0.86	1.51	−0.65
	(2.35)	(3.73)	(−2.78)
Simple spread, equally-weighted	1.11	1.34	−0.23
	(2.30)	(3.07)	(−1.08)

Monthly average returns for spread portfolios as well as the corresponding long (brown) and short (green) portfolios, based on the level or intensity of reported and estimated CO_2 emissions. The estimates are provided by Refinitiv. In parentheses are t-statistics based on heteroskedasticity-robust (White) standard errors. Panels A and B show results for the size-adjusted spread, a simple value-weighted quintile return spread, and and a simple equal-weighted quintile return spread. Sample period: January 2010 to December 2021.