CFR Working Paper NO. 22-07

Does it pay to invest in dirty industries? – New insights on the shunned-stock hypothesis

T. Bauckloh • V. Beyer • C. Klein

centre for financial research
cologne

Does it Pay to Invest in Dirty Industries? - New Insights on the Shunned-Stock Hypothesis

Tobias Bauckloh[a] Victor Beyer[b] Christian Klein[b]

January 31, 2022

Abstract: This research examines whether stocks of firms operating in highly polluting industries ("dirty stocks") are treated like sin stocks. We assume that investors shun dirty stocks based on non-pecuniary preferences and employ screening approaches that lead to the exclusion of entire industries. Using emission data of the U.S. Toxics Release Inventory, we show that dirty stocks are held in lower proportions by institutional investors and are followed by fewer financial analysts than other stocks. The shunning leads to an outperformance of dirty stocks in cross-sectional and time-series return analyses. These observations affect all firms within a dirty industry, regardless of whether they have high or low TRI emissions. This means that comparatively clean firms are shunned by capital market participants simply because of their industry affiliation, which can result in financing disadvantages and low incentives to improve sustainability performance. Thus, our findings contribute to the understanding of environmental preferences of investors and their consequences for asset pricing.

[a] University of Cologne and Centre for Financial Research, bauckloh@wiso.uni-koeln.de
[b] University of Kassel, victor.beyer@student.uni-kassel.de, klein@uni-kassel.de
We thank the participants of the BVI-CFR-Event 2021 and the "3rd Workshop on Empirical Sustainable Finance (WESFi)" at the University of Kassel for their valuable comments and suggestions.

1 Introduction

The rise of environmental, social, and governance (ESG) investing[1] is accompanied by a debate on associated consequences for asset pricing. On the one hand, the majority of empirical evidence indicate a positive or at least non-negative relation between the consideration of ESG criteria in the investment process and financial performance (e.g., Friede et al., 2015; von Wallis and Klein, 2015). On the other hand, a number of equilibrium models argue that investor discrimination against firms with low ESG performances increases their costs of capital, suggesting a negative relation between ESG investing approaches and financial performance (Angel and Rivoli, 1997; Heinkel et al., 2001; Pástor et al., 2021). This study contributes to this debate by examining effects of norm-oriented and/or -constrained investors' (henceforth "sustainable investors") investing preferences on assets' financial performance. More precisely, we investigate whether sustainable investors exclude (i.e., shun) firms from high-polluting industries and whether this exclusion affects stocks' risk-adjusted returns.

According to the shunned-stock hypothesis, sustainable investors shun certain stocks based on non-pecuniary preferences, such as environmental preferences. In line with Merton (1987), this leads to a market segmentation and limited risk sharing among conventional investors (i.e., non-sustainable investors), which is compensated by higher risk-adjusted returns of shunned stocks. Thereby, it is implicitly assumed that sustainable investors are willing to forgo financial performance in exchange for the non-pecuniary utility they derive from shunning certain stocks. This assumption is supported by several studies on non-pecuniary utility and asset pricing (e.g., Barber et al., 2021; Fama and French, 2007). Hong and Kacperczyk (2009) are the first to find a shun-effect of social norms on stock markets, showing that sustainable investors avoid investments in stocks of socially controversial firms (i.e., "sin stocks", which are stocks of firms operating in the alcohol, tobacco, or gambling industry), resulting in a superior financial performance of sin stocks.

Similar to the exclusion of socially controversial firms in investment decisions, there is evidence that institutional investors exclude firms with a weak corporate environmental performance (CEP) from their investment universe.[2] Institutional investors, in particular,

[1] The Global Sustainable Investment Review 2020 reports that ESG-considering investment approaches reached \$35.3 trillion globally at the start of 2020, a 55% increase since 2016. As of March 2021, more than 3,400 institutions had signed the United Nations-supported Principles for Responsible Investment (PRI) and committed to incorporate ESG criteria into their investment process. From PRI's 2021 Annual Report, this represents a 26% increase in one year and a tripling since 2014.

[2] Krueger et al.'s (2020) survey confirms that negative screening is a commonly pursued strategy by

appear to be concerned about their "environmental footprint", suggesting that societal expectations or own tastes prevent them from funding environmentally harmful ventures (Gibson Brandon et al., 2021). Using KLD ESG ratings, Chava (2014) finds firms with environmental concerns to have lower institutional ownership and higher implied costs of capital, which can only be partly explained by a higher default risk of firms with low CEP. Fernando et al. (2017) suggest low- and high-CEP firms to be shunned by institutional investors but do not find an effect on their assets' financial performance. Measuring firms' CEP with their toxic emission intensity, Hsu et al. (2020) find a superior financial performance of stocks of firms with low CEP, but cannot relate this to institutional investors' emission preferences and instead suggest a systematic firm-level pollution premium. Similarly, considering carbon emissions, Bolton and Kacperczyk (2021) find only weak support for the shunned-stock hypothesis and link emission-related abnormal stock returns to a systematic carbon risk premium.

In sum, the aforementioned studies struggle to provide a clear picture of a shun-effect in the E of ESG investing. Consistently, they assume that investors are willing and able to evaluate environmental performance, e.g., the volume of toxic or carbon emissions, at firm-level. We challenge this assumption and suspect that at least a part of investors pay limited attention to firm-level emissions but concentrate on industry-level emissions. Such behavior can be explained by Gabaix' (2014) sparsity-based model of bounded rationality, which links irrational economic decisions to behavioural economics in equilibrium. Accordingly, decision makers form a sparse optimization problem that considers only first-order variables. In the context of emission evaluation, this could lead to the shunning of entire industries that are deemed dirty, rather than considering emissions at firm-level, since high-emitting firms are concentrated in a few salient industries. In addition to limited attention to individual firm's emissions, the lack of disclosed emissions or investors' inability to properly assess emission data may further exacerbate industry-level shunning. To identify high-emitting ("dirty") industries, we refer to toxic chemical emission data reported in the U.S. Environmental Protection Agency's (EPA) Toxics Release Inventory (TRI).

Our study is related to a number of theoretical models. Based on the work of Merton (1987), Angel and Rivoli (1997) suggest that investors' unwillingness to hold unethically firms leads to a segmentation of the market in which these firms face higher costs of capital. Following a related approach, Heinkel et al. (2001) derive an equilibrium model that accounts for the lack of risk sharing among shunned stock investors. The market

institutional investors in environmental risk management.

mechanism underlying both models is that shunned stocks are less in demand, resulting in a depressed market value, thus shunned stocks are expected to earn higher returns. This implicitly presumes that the shunning is relatively stable over time and not enough arbitrage capital has been brought to the market to eliminate the norm and preference driven mispricing (Hong and Kacperczyk, 2009; Shleifer and Vishny, 1997). Pástor et al.'s (2021) equilibrium model predicts a similar relationship between CEP and stock returns but links changes in asset prices primarily to investor tastes in the sense of Fama and French (2007). Accordingly, investors derive utility from holding "green" assets and are willing to accept lower returns when holding these assets. Additionally, Pástor et al. (2021) expand their model for capturing green assets' ability to hedge environmental risk.

We examine whether dirty stocks are shunned stocks by looking at the institutional ownership ratio of firms from high-emitting industries compared to firms from other industries. Based on Hong and Kacperczyk (2009), we suggest that dirty stocks are held in smaller proportions by institutional investors for the following reasons: first, institutional investors, such as endowments, might shun those stocks because of own taste, i.e., non-pecuniary environmental preferences. Second, since institutional investors managing more than \$100 million must disclose their holdings under Securities and Exchange Commission (SEC) regulations, their investment portfolios are under public scrutiny. Therefore, institutional investors, such as pension funds, insurance companies, and banks could be asked by norm-oriented contributors/customers to avoid investments in certain industries. Third, the increased demand for sustainable investments is leading asset managers, such as Blackrock, to offer more financial products that exclude investments in dirty industries. However, we assume that some institutional investors, such as hedge funds, act as natural arbitrageurs in the market and increase their investments in stocks of dirty industries in order to earn abnormal returns.

Consistent with our predictions, panel regression analyses indicate lower aggregate relative institutional ownership for firms from dirty industries compared to firms from other industries of about 5% over the period 1997 to 2020. In disaggregate ownership regressions, we find that the shunning is primarily driven by endowments, foundations, pension funds, and investment advisors and has increased in recent decades. Conversely, hedge funds, venture capital, and private equity firms tend to hold about 37% higher stakes in dirty stocks than in stocks from other industries. To assess the robustness of our institutional ownership analysis, we ensure that the shunning is conducted on industry-level rather than on firm-level and additionally examine firms' analyst coverage. Since analyst forecasts are primarily prepared for institutional investors, we expect dirty stocks to be less followed by financial analysts. Moreover, the consideration of analyst coverage can be understood as an investigation of information availability in the sense of Merton's (1987) model of incomplete information. Our finding of about 15% lower analyst coverage for dirty stocks relative to other stocks provides further evidence that dirty stocks are being shunned. Shunning appears to be greatest when financial uncertainty is high, i.e., during and in the years following the 2007-2009 financial crisis.

We subsequently assess whether the shunning of dirty stocks leads to differences in (risk-adjusted) returns across industries. Cross-sectional return analyses indicate significantly higher returns for dirty stocks, while controlling for a range of firm characteristics. The outperformance of dirty stocks is more pronounced when the degree of shunning is at its highest. Additionally, we present an implementable long-short trading strategy whose abnormal time-series returns cannot be explained by common risk factors. Both approaches indicate 2.9 to 3.3% higher annual risk-adjusted returns for dirty than for non-dirty stocks.

The contributions of our study are fourfold. First, we provide empirical evidence for theoretical models suggesting capital market segmentation and higher risk-adjusted returns for non-sustainable stocks as a consequence of some investors' non-pecuniary preferences. Second, we add to research on the shunned-stock hypothesis, which to date has focused predominantly on the effects of social preferences on stock markets and struggles to provide a clear picture on the effects of environmental preferences. Our results indicate similar albeit smaller effects for environmental preferences of investors, which are stronger in times of financial uncertainty. Third, we offer new insights into how investors conduct environmental screenings and find support for a wide application of crude industry-level emission assessment. Fourth, knowledge about the effect of non-pecuniary preferences on asset pricing is important for both corporate managers and regulators to understand and respond to market-based incentives and constraints to promote emission reduction. Shunning on industry-level might result in firms having less incentive to reduce their emissions, since they are excluded from investors' portfolios based on their industry affiliation and regardless of firm-level emission reduction efforts.

The remainder of this paper is organized as follows. Section 2 describes our sample and dirty industry classification scheme. Section 3 presents our empirical results, which is followed by a discussion in section 4. Section 5 concludes.

2 Data

We use a merged dataset derived from CRSP and Compustat that includes U.S.-based firms with common stocks listed on the NYSE, NYSE American (formerly Amex), or Nasdaq. Monthly stock returns, closing prices, shares outstanding, and trading volumes are obtained from CRSP. From Compustat, we receive a set of annual accounting variables and firms' NAICS industry classification. We merge the lagged Compustat data with fiscal year end in calendar year $t-1$ with CRSP data in July of year t and keep this data until June of year $t+1$ to ensure that accounting variables are available before the variables they are supposed to explain. Since necessary accounting data are not available prior to 1962 and some variables are based on 36 monthly observations, we restrict our longest sample period from July 1965 to December 2020. For a firm to be included in our sample, stock market data, accounting variables, and three-digit NAICS industry classification codes must be available. The resulting unbalanced panel contains 18,643 distinct stocks and a total of 2,273,858 firm-month observations. To avoid a possible survivorship bias caused by missing stock delisting returns, we follow the methodology for adjustments recommended by Bali et al. (2016) and Shumway (1997). If a delisting code is 500, 520, between 551 and 574 inclusive, 580, or 584, we set a stock's return during the delisting month to -30%. In case of a missing delisting return and any other delisting code, the return is set to -100%.

Our data on stock-level institutional holdings come from the Refinitiv Institutional Holdings (13F) database and is available from 1997 to 2020. Starting in 1976, we obtain data on analyst coverage from the Institutional Brokers Estimate System (IBES) Historical Summary File. One-month Treasury bill rates and factors employed in time-series return regressions are downloaded from Kenneth French's data library.[3]

2.1 Dirty Industry Classification

To assess environmental performance, we employ TRI emission data, which have been collected by the EPA since 1987. It includes toxic chemical release data reported by U.S.-based industrial and government facilities that handle TRI-listed chemicals in quantities above established levels. Generally, the reporting requirement applies to facilities of firms with ten or more full-time employees that manufacture or process more than 25,000 pounds or use more than 10,000 pounds of the listed chemicals per year. The TRI list of

[3]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (January 31, 2022).

chemicals is updated regularly and currently lists 770 individual chemicals.[4] Using TRI emission data as environmental performance measure offers the following three key advantages: first, TRI data are an objective and easily interpretable criterion for measuring the environmental performance of firms or industries. This distinguishes them from ESG ratings, which are subject to various biases, such as the sustainability rating agencies' understanding of sustainability (Berg et al., 2020; Chatterji et al., 2016; Dorfleitner et al., 2015) or the size (Drempetic et al., 2020) and the rewriting history bias (Berg et al., 2020). Second, unlike single environmental performance indicators such as carbon emissions, TRI data cover a large part of a firm's or industry's environmental performance, as several hundred different toxic emissions are taken into account. Third, reporting is standardized and the data are open access, can be downloaded from the EPA's website[5] and indicate facilities' parent companies' NAICS industry classification, making it more trustworthy and emission quantities easily comparable across industries. Consequently, it enables investors to identify dirty industries.

We define dirty industries as the industries with the highest level of aggregate TRI emissions. To link emission data to industries, we refer to the three-digit NAICS industry classification, which contains 95 different industries. Of these 95 industries, 82 have to report under the TRI. Sections 8.1 through 8.7 of the annual TRI data files report quantities of different types of emissions, which we sum up to receive total emissions in pounds. We do this for every year and industry from 1987 to 2020 and calculate the time-series average emissions per industry. To control for data errors, we winsorize firm-level emissions within each industry at the 99.5% level. Industries that contribute to more than 5% of all emissions in the time-series average are classified as dirty. This results in six dirty industries, which together account for almost 80% of all reported emissions on an annual average. The emission contribution of each dirty industry is shown in Table 1.

When working with TRI total emission data, three issues are worth to be discussed. First, reporting requirements change over time. To check the robustness of the dirty industry classification, we look at the total emissions per industry for each year, rather than taking their time-series mean. Until 1998, most NAICS 212 and 221 firms were not required to report emissions but the remaining four dirty industries collectively produced at least 74% of all reported emissions each year from 1987 to 1997. From 1998 to 2020, the six dirty industries were the most polluting industries in 21 out of 23 years and among

[4]The current list of chemicals and changes over time can be downloaded from the EPA's website: https://www.epa.gov/toxics-release-inventory-tri-program/tri-listed-chemicals (January 31, 2022).

[5]https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-present (January 31, 2022).

Table 1: Emission contribution of dirty industries

NAICS	Industry Name	Emission Contribution
325	Chemical Manufacturing	40.22%
331	Primary Metal Manufacturing	12.32%
322	Paper Manufacturing	8.84%
221	Utilities	6.53%
324	Petroleum and Coal Products Manufacturing	5.85%
212	Mining (except Oil and Gas)	5.39%
Total		79.15%

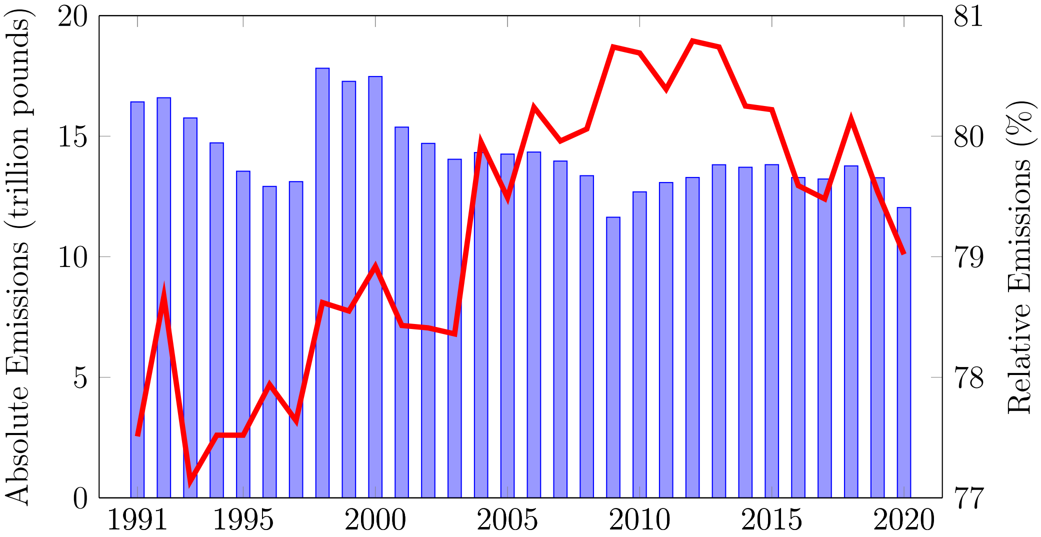
This table reports the time-series average of annual emission contribution of six dirty industries, measured as TRI total emissions relative to TRI total emissions of all industries from 1987 to 2020. Industries are denoted by their three-digit NAICS and sorted in descending order of emission contribution.

the top seven dirtiest industries in each year. As shown in Figure 1, absolute emissions of dirty industries have declined slightly over time, while their relative contribution to the total TRI emissions have increased until the early 2010s and only recently began to decrease.[6] Second, total emissions are not weighted for toxicity. Hsu et al. (2020) estimate toxicity with regional mortality rates to obtain more accurate results. Although listed chemicals might have different toxicities, the validity of such estimation methods is difficult to assess. We therefore opt for equal weighting, which suggests that investors do not evaluate different types of facility-based emissions with regard to their toxicity. Third, the emission data are self-reported and may not represent accurate emission numbers. For example, de Marchi and Hamilton (2006) note that despite EPA controls and penalties, reported air emissions often do not match actual emission levels. We assume that this data bias applies equally to all industries and therefore has no effect on the dirty industry classification. Since the selection of dirty industries has remained fairly stable over time, we assume that the identified industries can also be considered dirty before the EPA started collecting data in 1987, which allows for backward testing.

We also consider firm-level emissions by attributing the reported emission quantities of each facility to its parent firm. We then match the parent firm names in the TRI database with firm names in our sample of U.S.-listed stocks. To this end, the first step is to standardize the firm names by converting them to lowercase letters and removing suffixes, special characters, and punctuation. We also uniformly use common abbreviations such

[6]Figure 1 does not show the first four years (1987 to 1990) of TRI reporting due to inconsistent emission data collection methodology and missing data.

Figure 1: Dirty industries' emissions over time



This figure shows the aggregate emission contribution of dirty industries over the years from 1991 to 2020. Total TRI emissions in million pounds are plotted as blue bars on the left axis. The contribution of dirty industries relative to TRI total emissions of all industries in percentage is plotted as a red line on the right axis.

as "inc" for "incorporated" or "int" for "international". The matching process results in 1,681 perfect matches. The names of the remaining TRI firms are matched in a second step using a fuzzy algorithm. Subsequently, all matches with a Levenshtein distance fraction of less than 0.1 are checked by hand and adjusted if necessary. Finally, we obtain TRI firm-level emission data for 1,935 stocks in our sample.

2.2 Variables in the Institutional Ownership Regressions

In our institutional ownership regressions, the dependent variable $IO_{i,t}$ is the fraction of firm i's stocks held by institutional investors in the Refinitiv Institutional Holdings (13F) database at the end of year t. For stocks with no reported 13F holdings, $IO_{i,t}$ is set to zero. For stocks where the sum of all institutional investors' holdings exceeds 100%, we winsorize $IO_{i,t}$ to the maximum value of 1.[7] We continue to divide aggregate institutional holdings into five investor types according to Refinitiv's investor categories. Investor type 1 includes hedge funds (category ID 62), private equity (66), venture capital (67), and

[7] This affects a total of 7,266 or 6.54% of firm-year observations in our sample. Without winsorization, the maximum value of the ownership would have exceeded the shares outstanding by a factor of 26. Values greater than 100% in the 13F database occur, because short positions are not reported or due to data errors. Winsorization does not affect our findings, as we obtain similar regression results with the original values of $IO_{i,t}$.

hedge fund portfolios (80). Type 2 investors are endowment funds (58), foundations (60), pension funds (65), and sovereign wealth funds (69). Investment advisors (63) and investment advisors/hedge funds[8] (68) represent investor type 3. Banks and trusts (57), insurance companies (64), and insurance company portfolios (109) are grouped in type 4. The remaining investor categories excluding strategic entities are summarized in type “other”.

We construct a dirty industry dummy variable $DDUM_{i,t}$ that equals one if firm i is operating in the chemical, primary metal, paper, petroleum and coal products manufacturing, utilities, or mining industry at the end of year t, and zero otherwise. Additionally, several control variables that are known to have explanatory power for stocks' institutional ownership are included in the regressions. $BETA_{i,t}$ is the time-varying CAPM beta of firm i's industry. Industry betas are calculated using a rolling 36-month window for industries defined at the three-digit NAICS level. $LOGSIZE_{i,t}$ is the natural logarithm of the market capitalization of firm i at the end of year t, calculated as price times shares outstanding in \$million. $LOGMB_{i,t}$ is the natural logarithm of firm i's monthly market capitalization divided by its book value of common equity at the previous fiscal year-end. $LOGPRINV_{i,t}$ is the natural logarithm of firm i's inverse share price. $RET_{i,t}$ is firm i's geometric mean monthly return of the last 12 months in year t and $STD_{i,t}$ is the standard deviation of its last 12 monthly raw returns. To control for outliers, $RET_{i,t}$ is winsorized at the 1% and 99% level and $STD_{i,t}$ is winsorized at the 99% level. $LOGTURN_{i,t}$ is the average monthly share turnover of firm i in year t, measured as the natural logarithm of monthly shares traded divided by shares outstanding. $LOGAGE_{i,t}$ is the natural logarithm of the number of years for which stock market data are available for firm i up to year t. We include two more dummy variables in the set of control variables. $NASDAQ_{i,t}$ equals one if firm i's shares are traded at Nasdaq at the end of year t, and zero otherwise. $SP500_{i,t}$ is set to one if firm i is constituent of the S&P 500 index at the end of year t, and zero otherwise. Since the investor type categories for institutional investors are only available in the Refinitiv database from 1997 onwards, we restrict our sample period from 1997 to 2020. During this period, on average, 50.63% of a firms' shares were held by institutional investors. The average of $DDUM_{i,t}$ is 0.1488, indicating that 14.88% of the observations in our sample are attributable to dirty stocks. Summary statistics for the variables are provided in Panel A of Table 2.

[8]This investor category includes institutions that engage in both investment advisory and hedge fund activities. Since the holdings cannot be classified separately by activity, but investment advisory accounts for the larger part, we assign the category entirely to type 3.

Table 2: Summary statistics

Variable	Mean	Median	Std. Dev.
Panel A: Institutional ownership variables 1997-2020			
IO	0.5063	0.5277	0.3402
IO type 1	0.0651	0.0200	0.1074
IO type 2	0.0232	0.0100	0.0322
IO type 3	0.3996	0.3838	0.2961
IO type 4	0.0115	0.0030	0.0399
IO other	0.0127	0.0100	0.0291
DDUM	0.1488		0.3559
BETA	1.0650	1.0508	0.4040
LOGSIZE	5.7784	5.7115	2.2309
LOGMB	0.7141	0.6684	1.1034
LOGPRINV	−2.4934	−2.7107	1.3632
RET (%)	2.0223	1.7200	1.3503
STD	14.3898	11.6284	10.0008
LOGTURN	−0.0961	0.0223	1.1409
LOGAGE	2.4130	2.4849	0.9643
NASDAQ	0.6110	1	0.4875
SP500	0.1087	0	0.3113
Panel B: Analyst coverage variables 1976-2020			
COV	4.0462	1	6.3947
DDUM	0.1481	0	0.3552
BETA	1.0811	1.0874	0.3517
LOGSIZE	5.0986	5.0062	2.3070
LOGMB	0.6346	0.5688	1.0613
LOGPRINV	−2.3403	−2.6027	1.3852
RET (%)	1.9979	1.7078	1.3174
STD	13.8595	11.4399	9.1947
LOGTURN	−0.4770	−0.4382	1.1736
LOGAGE	2.3748	2.4849	0.9330
NASDAQ	0.5733	1	0.4946
SP500	0.1124	0	0.3158
Panel C: Cross-sectional return variables 1965-2020			
EXRF (%)	0.7952	−0.2800	18.0843
DDUM	0.1539		0.3609
LOGSIZE	5.0687	4.9536	2.2252
LOGMB	0.6279	0.5463	0.9812
RET (%)	1.9923	1.6897	1.4375

Continued on the next page

Table 2 continued

Variable	Mean	Median	Std. Dev.
STD	13.8730	11.3213	10.8437
BETA	1.0849	1.0960	0.3492
LOGTURN	−0.5604	−0.5212	1.1866
LOGAGE	2.3493	2.4423	0.9819
Panel D: Time-series return variables 1965-2020			
DIFF (%)	−0.0502	−0.0585	2.6686
MRP (%)	0.5618	0.9200	4.5302
SMB (%)	0.2168	0.1500	3.0907
HML (%)	0.2376	0.2200	2.9008
MOM (%)	0.6479	0.7250	4.2824

This table shows summary statistics (averages, medians, and standard deviations) for the variables used in the four regression sets. **Panel A** reports the variables used in the institutional ownership regressions from 1997 to 2020. *IO* is the fraction of a firm's stocks held by institutional investors at the end of year t. *IO type 1* is the fraction of stocks held by hedge funds, private equity, and venture capital firms. *IO type 2* is the fraction of stocks held by endowment funds, foundations, pension funds, sovereign wealth funds, and government agencies. *IO type 3* is the fraction of stocks held by investment advisors. *IO type 4* is the fraction of stocks held by banks and insurance companies. *IO other* is the fraction of stocks held by institutions in the remaining Refinitiv investor categories. *DDUM* equals one if a firm is operating in a dirty industry and zero otherwise. *BETA* is a firm's industry rolling CAPM beta calculated over the last 36 months. *LOGSIZE* is the natural logarithm of a firm's market capitalization. *LOGPRINV* is the natural logarithm of a firm's inverse share price. *RET* is the geometric mean of a firms monthly returns over the last 12 months. *STD* is the standard deviation a firm's monthly returns over the last 12 months. *LOGTURN* is the natural logarithm of a firm's average monthly share turnover over the last 12 months. *NASDAQ* equals one if a firm's shares are traded at Nasdaq and zero otherwise. *SP500* equals one if a firm is constituent of the S&P 500 index and zero otherwise. **Panel B** reports the variables used in the analyst coverage regressions from 1976 to 2020. *COV* is the number of analysts following a firm. The remaining variables are defined as in Panel A. **Panel C** reports the variables used in the cross-sectional return regressions from 1965 to 2020. *EXRF* is a firm's monthly stock return net of the risk-free rate. The remaining variables are defined as in Panel A, but are now calculated on a monthly basis. **Panel D** reports the variables used in the time-series return regressions from 1965 to 2020. *DIFF* is the monthly return of a portfolio that is long in dirty stocks and short in non-dirty stocks, excluding sin stocks. *MRP* is the monthly return of a value-weighted portfolio of all NYSE, NYSE American, or Nasdaq listed U.S. firms net of the risk free rate. *SMB* is the monthly return of a portfolio that is long in high market cap stocks and short in low market cap stocks. *HML* is the monthly return of a portfolio that is long in high book-to-market stocks and short in low book-to-market stocks. *MOM* is the monthly return of a portfolio that is long in stocks with high past 12 month returns and short in stocks with low past 12 month returns.

2.3 Variables in the Analyst Regressions

We measure analyst coverage by the number of available earnings per share (EPS) financial analyst estimates in the IBES database. The dependent variable $COV_{i,t}$ is defined as the number of analysts following firm i at the end of year t. For stocks with no IBES information in the fourth quarter,[9] the number of covering analysts is assumed to be zero. The average firm in our sample is covered by about four analysts over the 1976 to 2020 period. We use the same control variables as in the institutional ownership regression. Differences in the summary statistics of Panel A and Panel B of Table 2 result from different investigation periods and therefore different firm samples.

2.4 Variables in the Cross-Sectional Return Regressions

Panel C of Table 2 displays summary statistics for the cross-sectional return regressions. $EXRF_{i,t}$ as dependent variable denotes the monthly return of stock i net of the risk-free rate at the end of month t. The control variables $LOGSIZE_{i,t}$, $LOGMB_{i,t}$, $RET_{i,t}$, $STD_{i,t}$, $BETA_{i,t}$, $LOGTURN_{i,t}$, and $LOGAGE_{i,t}$ are defined as in the institutional ownership regression, but are now calculated on a monthly basis. Since the necessary accounting data are not available prior to 1962 and data going back at least 36 months are needed, we use the sample period from July 1965 to December 2020.

2.5 Variables in the Time-Series Return Regressions

In our time-series return regressions, we build two value-weighted portfolios. One is the dirty industry portfolio and the other is the non-dirty industry portfolio, excluding sin stocks. The dependent variable in our time-series regression is the monthly return of the difference portfolio $DIFF_t$, which is long in the dirty industry and short in the non-dirty industry portfolio. MRP_t, SMB_t, HML_t, and MOM_t are the well-known Fama & French (1992) and Carhart (1997) risk factors in month t. MRP_t is the excess return on the market calculated from a value-weighted portfolio of all NYSE, NYSE American, or Nasdaq listed U.S. firms with a CRSP share code of 10 or 11, minus the one-month Treasury bill rate. SMB_t is the return of a portfolio that is long in high market cap stocks and short in low market cap stocks. HML_t is the return of a portfolio that is long in value stocks and short in growth stocks, as measured by the book-to-market ratio. MOM_t is the return of a portfolio that is long in stocks with high past 12 month returns and short in stocks with low past 12 month returns. Summary statistics for these monthly factor portfolio returns over the sample period July 1965 to December 2020 are shown in Panel D of Table 2.

[9] If no analyst estimates are available for a stock at the end of December in year t, the number is taken from the most recent estimate in the last quarter of year t.

3 Results

3.1 Institutional Ownership

We start our exploration of a shun-effect by testing whether dirty stocks are held in lower proportions by institutional investors than comparable stocks from other industries. Attributing differences in ownership to industry membership is challenging because the stocks in our sample differ along many dimensions. For example, Gompers and Metrick (2001) study institutional investors' demand for stock characteristics and find that they prefer large caps for liquidity reasons. To control for systematic cross-industry differences unrelated to industry emissions, we include a set of firm characteristics as control variables that have been used earlier in related literature to predict institutional ownership (e.g., Fernando et al., 2017; Gompers and Metrick, 2001; Hong and Kacperczyk, 2009). Using pooled OLS, we estimate the following cross-sectional regression model:

$$IO_{i,t} = a_0 + a_{DDUM} \cdot DDUM_{i,t} + a_X \cdot X_{i,t} + \epsilon_{i,t}, \quad i = 1, \ldots, N_t \tag{1}$$

where $IO_{i,t}$ is the fraction of firm i's shares held by institutional investors at the end of year t, $DDUM_{i,t}$ is our dirty industry dummy variable, $X_{i,t}$ is a vector of control variables defined in section 2.2, and $\epsilon_{i,t}$ is the measurement error. We also include year and industry fixed effects and compute double clustered standard errors at the firm- and industry-level. Unlike our dirty industry dummy variable, industry fixed effects are defined at the one-digit NAICS level to capture broader industry effects. Thereby, we ensure that our regression results cannot be interpreted in a way that institutional investors prefer investments in certain one-digit NAICS industries over others.

Our coefficient of interest is a_{DDUM}. It measures whether the aggregate institutional ownership of dirty stocks differs from that of firms operating in non-dirty industries. If there were no ownership differences, the null hypothesis states that a_{DDUM} would equal zero. In contrast, we conjecture that dirty industry firms are held to a lesser extent by institutional investors and $a_{DDUM} < 0$.

Table 3 presents regression results for different specifications of the institutional ownership regression. In column (1) we include $BETA_{i,t}$, $LOGSIZE_{i,t}$, $LOGMB_{i,t}$, $NASDAQ_{i,t}$, and $SP500_{i,t}$ to the vector of control variables. The coefficient in front of $DDUM_{i,t}$ is -0.0230 and is statistically significant at the 5% level. The mean value of institutional ownership of a stock in our sample is 0.5063 and the coefficient a_{DDUM} tells that dirty stocks have on average 2.30 percentage points less institutional ownership, when control-

Table 3: Institutional ownership regressions 1997-2020

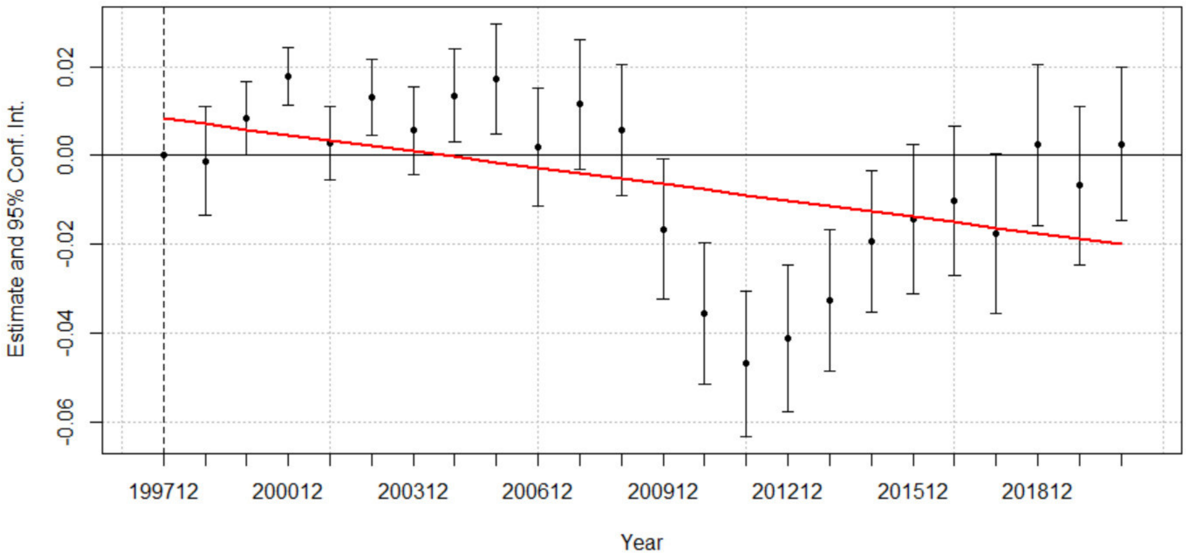
	(1)	(2)	(3)	(4)	(5)
DDUM	-0.0230**	-0.0182*	-0.0213**	-0.0232***	-0.0173**
	(0.0092)	(0.0093)	(0.0081)	(0.0081)	(0.0069)
BETA	0.0499***	0.0517***	0.0280***	0.0270***	0.0290***
	(0.0086)	(0.0086)	(0.0071)	(0.0070)	(0.0076)
LOGSIZE	0.1110***	0.1026***	0.0679***	0.0683***	0.0678***
	(0.0040)	(0.0052)	(0.0037)	(0.0037)	(0.0037)
LOGMB	-0.0266***	-0.0293***	-0.0338***	-0.0339***	-0.0336***
	(0.0041)	(0.0042)	(0.0032)	(0.0032)	(0.0034)
LOGPRINV		-0.0167***	-0.0325***	-0.0323***	-0.0325***
		(0.0047)	(0.0037)	(0.0037)	(0.0038)
RET		0.0023	0.0002	0.0002	0.0001
		(0.0020)	(0.0020)	(0.0020)	(0.0021)
STD		0.0004	-0.0043***	-0.0043***	-0.0043***
		(0.0004)	(0.0005)	(0.0005)	(0.0005)
LOGTURN			0.0960***	0.0955***	0.0962***
			(0.0059)	(0.0058)	(0.0061)
LOGAGE			-0.0052	-0.0052*	-0.0054*
			(0.0030)	(0.0030)	(0.0031)
NASDAQ	-0.0051	-0.0067	-0.0314***	-0.0314***	-0.0314***
	(0.0064)	(0.0065)	(0.0081)	(0.0080)	(0.0081)
SP500	-0.2055***	-0.1920***	-0.1574***	-0.1567***	-0.1572***
	(0.0196)	(0.0201)	(0.0154)	(0.0154)	(0.0154)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Controlling for sin stocks				Yes	
DDUM $\times$ year fixed effects					Yes
Observations	100,711	100,711	100,711	100,711	100,711
R^2	0.5156	0.5171	0.5768	0.5776	0.5771

This table shows institutional ownership regression results using pooled OLS. The dependent variable is *IO*, denoting the fraction of a firm's stocks held by institutional investors. The other variables are defined as in Table 2. The estimations are based on annual data during the period December 1997 to December 2020. Double-clustered standard errors (firm & year) are given in parentheses. Significance codes: *p<0.1, **p<0.05, ***p<0.01.

ling for a set of specific firm characteristics. This corresponds to a decrease of almost 5% compared to the sample mean. The coefficients of control variables indicate that size and industry beta positively affects institutional ownership and institutional investors seem to favour value stocks with low market-to-book ratios. Stocks that are constituents of the S&P500 have less institutional ownership than comparable stocks, while the effect of being Nasdaq-listed is insignificant.

With the successive inclusion of additional control variables in columns (2) and (3), the coefficient in front of a_{DDUM} ranges between -0.0182 and -0.0213 and is statistically significant at the 10% and 5% level, respectively. In column (4), we additionally control for the influence of sin stocks, which are known to be held by a smaller fraction of institutional investors, by adding a simple sin stock dummy variable that equals one for sin stocks and zero for other stocks (Hong and Kacperczyk, 2009). This increases the magnitude of our coefficient of interest a_{DDUM} to -0.0232 and remains statistically significant at the 1% level.

Figure 2: Changes in dirty stocks' institutional ownership over time



This figure shows coefficients of the interaction term of $DDUM_{i,t}$ with year dummy variables in the institutional ownership regression over the period 1997 to 2020 and their respective 95% confidence intervals. The OLS-fitted line is plotted in red.

To observe possible changes in non-pecuniary preferences of investors over time, we include an interaction term of $DDUM_{i,t}$ with year dummy variables. The regression results for the reference year 1997 are shown in column (5) and changes are plotted in Figure 2. For 1997, the dirty industry dummy coefficient is -0.0173 and statistically significant at the 5% level. The year-by-year interaction terms, yielding a negative OLS slope coefficient of -0.0013, which is statistically significant at the 5% level, indicate that the fraction of dirty stocks held by institutions decrease by an average of 0.13 percentage points per year over the sample period. However, Figure 2 shows that the shunning by institutional investors is particularly strong during and in the years following the 2007-2009 financial crisis. This indicates that non-pecuniary preferences of investors become stronger in times of financial uncertainty.

Table 4: Firm-level emissions and institutional ownership

	(1)	(2)	(3)	(4)	(5)
DDUM	-0.0213**		-0.0194**		-0.0229**
	(0.0081)		(0.0082)		(0.0082)
DFDUM1		-0.0272*	-0.0208		
		(0.0148)	(0.0149)		
DFDUM2				0.0145	0.0205
				(0.0121)	(0.0121)
Controls	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	100,711	100,711	100,711	100,711	100,711
R^2	0.5768	0.5766	0.5769	0.5765	0.5769

This table shows institutional ownership regression results using pooled OLS. The dependent variable is *IO*, denoting the fraction of a firm's stocks held by institutional investors. *DFDUM1* equals one if a firm is in the top quintile of firms with the highest total emissions and zero otherwise. *DFDUM2* equals one if a firm is in the top quintile of firms with the highest emission intensity and zero otherwise. The unreported control variables match the specification of column (3) in Table 3. The estimations are based on annual data during the period December 1997 to December 2020. Double-clustered standard errors (firm & year) are given in parentheses. Significance codes: *p<0.1, **p<0.05, ***p<0.01.

To investigate the possibility that our observations on industry-level are attributable to a shunning on firm-level, i.e., that sustainable investors shun stocks on firm- rather than on industry-level, we introduce two additional dummy variables in equation (1) to examine whether institutional holdings vary with the level of firm-level emissions. $DFDUM1_{i,t}$ is set to one if firm i is in the top quintile of firms that emit the highest quantities of TRI emissions in year t. $DFDUM2_{i,t}$ is set to one if firm i is in the top quintile of firms with the highest emission intensity, calculated as total emissions relative to sales. Regression results including both firm-level dummy variables are reported in Table 4. As shown in column (2), the $DFDUM1_{i,t}$ coefficient is -0.0272 and statistically significant at the 10% level, suggesting that firms with high total emissions tend to have less institutional ownership than firms with lower total emissions. Considering the firm-level dummy variable together with the industry dummy variable in column (3), the $DFDUM1_{i,t}$ coefficient is no longer significant, while the $DDUM_{i,t}$ coefficient remains similar in magnitude and significant at the 5% level. Thus, the shunning at the firm-level observed in column (2) seems to be the result of a shunning at the industry-level, since the exclusion of dirty industries involves the exclusion of the dirtiest firms. Conversely, the significant industry dummy variable coefficient indicates that the shunning at the industry-level observed in column (1) cannot be explained by institutional investors avoiding only the dirtiest firms. Consistently, in columns (4) and (5), we find no evidence for a shunning based on firm-level emissions intensity.

In a more in-depth analysis, we look at disaggregated institutional ownership. In doing so, we want to address the conjecture that not all institutional investor types exhibit the same aversion to dirty stocks. We divide institutional investors into five types in accordance with the Refinitiv investor categories. We hypothesize that type 1 investors, which comprise hedge funds, hedge fund portfolios, private equity, and venture capital firms, do not systematically shun dirty industries because they are expected to act as neutral arbitrageurs.

Table 5: Institutional ownership regressions by investor type

	Type 1	Type 2	Type 3	Type 4	Other
DDUM	0.0239***	-0.0014*	-0.0427***	-0.0011	-0.0001
	(0.0039)	(0.0008)	(0.0067)	(0.0009)	(0.0007)
Controls	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	100,711	100,711	100,711	100,711	100,711
R^2	0.1779	0.2639	0.5812	0.0513	0.0651

This table shows institutional ownership regression results by investor type using pooled OLS. The dependent variable *IO* is divided into subcategories, each of which denotes the fraction a firm's stocks held by the corresponding institutional investor type. Type 1 includes hedge funds, private equity, venture capital, and hedge fund portfolios. Type 2 includes endowment funds, foundations, pension funds, and sovereign wealth funds. Type 3 includes investment advisors and investment advisors/hedge funds. Type 4 includes banks and trusts, insurance companies, and insurance company portfolios. Other includes the remaining institutional investors according to the Refinitiv investor categories. The unreported control variables match the specification of column (3) in Table 3. The estimations are based on annual data during the period December 1997 to December 2020. Double-clustered standard errors (firm & year) are given in parentheses. Significance codes: *p<0.1, **p<0.05, ***p<0.01.

For other types of investors, such as endowments, foundations, pension funds, banks and insurances (type 2 and 4 investors), we assume that they have own non-pecuniary preferences or that they are subject to social norm pressure exerted by clients/contributors and have reduced their investments in the controversial dirty industries. For type 3 investors, such as Blackrock or State Street, we also expect a lower institutional ownership in dirty stocks due to an increasing demand for sustainable investment opportunities (e.g., sustainable equity funds) in recent decades. To test the institutional ownership of different investor types, we modify the dependent variable in regression equation (1). $IO_{i,t}$ now represents the fraction of firm i's shares held by each of the 5 investor types at the end of year t. The vector of control variables is composed as in column (3) of Table 3. Results of the five sets of regressions are shown in Table 5.

In the type 1 regression, we find a $DDUM_{i,t}$ coefficient of 0.0239, which is statistically significant at the 1% level. The sign of a_{DDUM} has switched to positive compared to the previous results on aggregate institutional ownership. In other words, type 1 investors, such as hedge funds or venture capital firms, hold a larger share of dirty stocks than they do of comparable stocks in other industries. Given that the mean fraction of a firm's stocks held by type 1 investors is 0.0651, our results imply that these investors hold almost 37% higher stakes in dirty stocks than in stocks from other industries. For type 2, 3, and 4 investors we find negative values for a_{DDUM}. This is line with our prediction that these three investor types drive our findings on aggregate institutional ownership, but only the first two values are statistically significant at the 10% and 1% level, respectively. The type 2 a_{DDUM} coefficient of -0.0014 indicates that dirty stocks are held 0.14 percentage points less by institutions like endowments or pension funds compared to other stocks. This corresponds to a shortfall of 6% relative to the mean fraction of type 2 holdings, which is 0.0232. The strongest industry shunning effect comes from the type 3 investors, which represent the largest investor type in terms of stock ownership. Investment advisors have a 4.27 percentage points lower institutional ownership in dirty stocks than in other stocks, which translates into a relative shortfall of about 11% when considering the average type 3 institutional ownership of 0.3996. For type 4 investors, such as banks and insurances, and other investor types, we do not find significant differences in dirty stock holdings.

Our overall results fit good with our hypothesis that dirty stocks are shunned by a substantial group of sustainable institutional investors. An exception are type 1 investors, who act as neutral arbitrageurs and hold higher stakes in dirty industry firms.

3.2 Analyst Coverage

In this part, we try to reinforce the observations made in the previous section. The basic idea is that financial analysts prepare their forecasts for stocks based on investor demand. Given the institutional ownership regression results, it can be suggested that dirty stocks have less analyst coverage. The IBES analyst data additionally offer the advantages that there is no threshold in reporting requirement and that data are available from 1976 onwards. This leads to a significant extension of the firm sample in comparison to the institutional ownership sample, which only includes data of institutional investors with over \$100 million assets under management from 1997 onwards.

To test for possible variations in analyst coverage, we use the same methodology as in the institutional ownership regression and estimate the following cross-sectional regression model:

$$COV_{i,t} = b_0 + b_{DDUM} \cdot DDUM_{i,t} + b_X \cdot X_{i,t} + \epsilon_{i,t}, \quad i = 1, \ldots, N_t \tag{2}$$

where $COV_{i,t}$ is the number of analysts following firm i at the end of year t, $DDUM_{i,t}$ is our dirty industry dummy variable, $X_{i,t}$ is a vector of control variables, and $\epsilon_{i,t}$ is an error term. The vector $X_{i,t}$ includes the same control variables as previously used in the institutional ownership regression in equation (1). This allows us to quickly summarize the key findings, which are presented in Table 6.

Our coefficient of interest b_{DDUM} measures whether the analyst coverage of dirty stocks differs from that of firms operating in non-dirty industries. It takes negative and statistically significant values in all regression specifications. To assess economic significance, we calculate the effect captured by b_{DDUM} on an average firm in our sample, which is followed by about 4.0462 analysts in a typical year. The coefficient in column (3) in front of $DDUM_{i,t}$ is -0.6167 and statistically significant at the 1% level. This number is equivalent to about 3.4273 analysts covering a typical firm in a dirty industry. In relative terms, this reflects a 15% decrease in analysts covering dirty stocks compared to the mean of comparable stocks.[10]

Using the same methodology as in the institutional ownership regressions, we analyse changes in dirty stocks analyst coverage over time. As shown in Figure 3, we observe a downward sloping coefficient of -0.0204, which is statistically significant at the 1% level,

[10]Prior literature often define analyst coverage as the natural logarithm of one plus the number of analysts following a firm (e.g., Fernando et al., 2017; Hong and Kacperczyk, 2009). We also run our regression with this definition of analyst coverage as the dependent variable and obtain robust results.

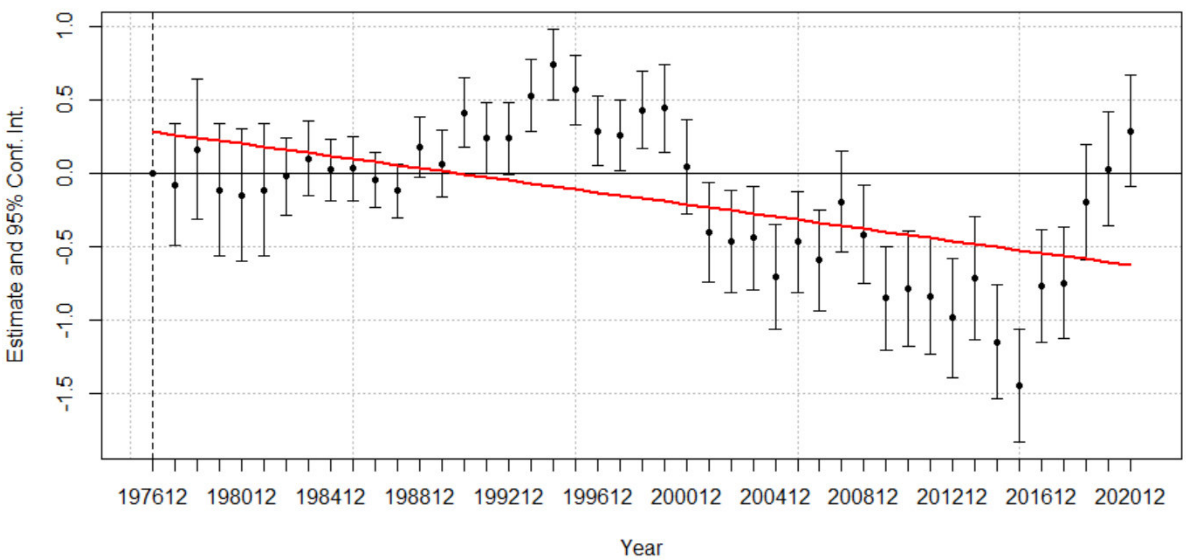
Table 6: Analyst coverage regressions 1976-2020

	(1)	(2)	(3)	(4)	(5)
DDUM	-0.4451**	-0.6234***	-0.6167***	-0.6348***	-0.4903**
	(0.1710)	(0.1750)	(0.1768)	(0.1769)	(0.1885)
BETA	0.5176***	0.3509***	0.1340	0.1246	0.1206
	(0.0997)	(0.0969)	(0.0948)	(0.0934)	(0.0999)
LOGSIZE	1.561***	1.924***	1.717***	1.721***	1.713***
	(0.0453)	(0.0593)	(0.0594)	(0.0597)	(0.0596)
LOGMB	-0.1382***	-0.0887*	-0.1447***	-0.1453***	-0.1400***
	(0.0484)	(0.0508)	(0.0432)	(0.0432)	(0.0451)
LOGPRINV		0.5461***	0.4731***	0.4749***	0.4769***
		(0.0503)	(0.0526)	(0.0526)	(0.0538)
RET		-0.1953***	-0.2340***	-0.2335***	-0.2324***
		(0.0268)	(0.0268)	(0.0266)	(0.0276)
STD		0.0233***	-0.0121**	-0.0118**	-0.0119**
		(0.0052)	(0.0047)	(0.0047)	(0.0048)
LOGTURN			0.7013***	0.6962***	0.7103***
			(0.0474)	(0.0472)	(0.0486)
LOGAGE			-0.2430***	-0.2434***	-0.2498***
			(0.0532)	(0.0530)	(0.0545)
NASDAQ	0.4403***	0.4399***	0.1629	0.1621	0.1583
	(0.0979)	(0.0985)	(0.1007)	(0.1006)	(0.1010)
SP500	5.141***	4.585***	4.884***	4.886***	4.896***
	(0.3352)	(0.3198)	(0.3218)	(0.3217)	(0.3221)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Controlling for sin stocks				Yes	
DDUM × year fixed effects					Yes
Observations	181,693	181,693	181,693	181,693	181,693
R^2	0.4862	0.4934	0.5039	0.5042	0.5047

This table shows analyst coverage regression results using pooled OLS. The dependent variable is *COV*, denoting the number of analysts following a firm. The other variables are defined as in Table 2. The estimations are based on annual data during the period December 1976 to December 2020. Double-clustered standard errors (firm & year) are given in parentheses. Significance codes: *p<0.1, **p<0.05, ***p<0.01.

indicating that analyst coverage of dirty stocks is decreasing on average over the period 1976 to 2020. Similar to the findings for the institutional ownership, analysts' shunning of dirty stocks is strongest during and in the years following the 2007-2009 financial crisis.

Figure 3: Changes in dirty stocks' analyst coverage over time



This figure shows coefficients of the interaction term of $DDUM_{i,t}$ with year dummy variables in the analyst coverage regression over the period 1976 to 2020 and their respective 95% confidence intervals. The OLS-fitted line is plotted in red.

In a robustness check, We test for differences in analyst coverage based on firm-level emissions. Table 7 reveals that both dirty industry affiliation and firm-level emission performance negatively affect analyst coverage.

Combined with our earlier findings on institutional ownership, we find strong support for the shunned-stock hypothesis at the industry-level. Sustainable investors avoid investing in stocks from dirty industries, while natural arbitrageurs like hedge funds are willing to buy these stocks. This raises the question of whether enough arbitrage capital is brought to market to offset lower demand for dirty stocks by certain investor groups, or whether their preferences to shun these stocks affect assets' financial performance.

3.3 Cross-Sectional Returns

In this section we start by examining dirty stock returns in the cross-section. Theory suggests that we can observe return premiums when stocks from dirty industries have been systematically shunned in recent decades. Hence, we estimate the following return forecasting regression model:

$$EXRF_{i,t} = c_0 + c_{DDUM} \cdot DDUM_{i,t-1} + a_X \cdot X_{i,t-1} + \epsilon_{i,t}, \quad i = 1, \ldots, N_t \tag{3}$$

Table 7: Firm-level emissions and analyst coverage

	(1)	(2)	(3)	(4)	(5)
DDUM	-0.7064***		-0.6220***		-0.6592***
	(0.1802)		(0.1741)		(0.1775)
DFDUM1		-1.373***	-1.197***		
		(0.3664)	(0.3612)		
DFDUM2				-0.9171***	-0.7615***
				(0.2236)	(0.2169)
Controls	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	141,924	141,924	141,924	141,924	141,924
R^2	0.5279	0.5278	0.5286	0.5273	0.5282

This table shows analyst coverage regression results using pooled OLS. The dependent variable is *COV*, denoting the number of analysts following a firm. *DFDUM1* equals one if a firm is in the top quintile of firms with the highest total emissions and zero otherwise. *DFDUM2* equals one if a firm is in the top quintile of firms with the highest emission intensity and zero otherwise. The unreported control variables match the specification of column (3) in Table 6. The estimations are based on annual data during the period December 1989 to December 2020. Double-clustered standard errors (firm & year) are given in parentheses. Significance codes: *p<0.1, **p<0.05, ***p<0.01.

where $EXRF_{i,t}$ is the monthly return of stock i net of the risk-free rate at the end of month t, $DDUM_{i,t-1}$ is our dirty industry dummy variable, and $X_{i,t-1}$ is a vector of control variables in month $t-1$, defined in section 2.4. Again, the selection of control variables is crucial because stock returns vary widely across firms with different characteristics. To best identify return differences between stocks from dirty industries and stocks from non-dirty industries, our strategy is to filter out as much of the cross-sectional variation as possible. We achieve this by including control variables that are commonly used in the literature to explain cross-sectional returns (Hong and Kacperczyk, 2009). We additionally include year/month and industry fixed effects to rule out the possibility that our results are driven by time or broader industry effects.

Since the time-varying industry beta $BETA_{i,t-1}$ is computed using a 36-month window of monthly returns and accounting data are available from July 1962 onwards, we obtain our first estimation on one month lagged control variables for July 1965. Our unbalanced panel includes 2,273,858 firm-month observations. The coefficient c_{DDUM} indicates whether firms operating in dirty industries earn different returns than those operating in non-dirty industries. If there are no return differences, c_{DDUM} is not significantly differ-

Table 8: Cross-sectional return regressions

	(1)	(2)	(3)	(4)	(5)
DDUM	0.2596**	0.2970***	0.2870**	0.2960**	0.4882*
	(0.1071)	(0.1083)	(0.1241)	(0.1243)	(0.2534)
LOGSIZE	-0.0565	-0.0508	-0.1157***	-0.1175***	0.0359
	(0.0465)	(0.0461)	(0.0408)	(0.0409)	(0.0641)
LOGMB	-0.4308***	-0.5256***	-0.4765***	-0.4762***	-0.3957*
	(0.0938)	(0.0775)	(0.0824)	(0.0824)	(0.2119)
RET		0.2098***	0.2982***	0.2982***	0.0650
		(0.0664)	(0.0699)	(0.0699)	(0.1080)
STD			-0.0276*	-0.0277*	-0.0088
			(0.0165)	(0.0165)	(0.0285)
BETA			-0.0680	-0.0627	-0.3380
			(0.2436)	(0.2437)	(0.3303)
LOGTURN			-0.0516	-0.0491	-0.1047
			(0.0781)	(0.0782)	(0.1565)
LOGAGE			0.1117***	0.1117***	0.0917
			(0.0350)	(0.0350)	(0.0792)
Year/month fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Controlling for sin stocks				Yes	
Observations	2,273,858	2,273,858	2,273,858	2,273,858	344,148
R^2	0.1014	0.1016	0.1018	0.1018	0.1255

This table shows cross-sectional return regression results using pooled OLS. The dependent variable is *EXRF*, denoting a stock's monthly return net of the risk-free rate. The control variables are defined as in Table 2. The estimations in columns (1) to (4) are based on monthly data for the period July 1965 to December 2020 and in column (5) for the period January 2008 to December 2015. Double-clustered standard errors (firm & year/month) are given in parentheses. Significance codes: *p<0.1, **p<0.05, ***p<0.01.

ent to zero. In contrast, our hypothesis states that dirty stocks are shunned by certain types of investors and earn higher return premiums. Therefore, we would expect to find a coefficient $c_{DDUM} > 0$. Pooled OLS regression results with standard errors clustered at firm- and year-level are presented in Table 8.

Specification (1) only includes size and market-to-book ratio as control variables and gives a dirty industry return exposure of 0.2596, which is statistically significant at the 5% level. When adding $RET_{i,t-1}$, the coefficient c_{DDUM} becomes larger and remains statistically significant in column (2). In line with the models of Fama and French (1992) and Carhart (1997), we find significantly negative coefficients for $LOGSIZE_{i,t-1}$ and $LOGMB_{i,t-1}$ and significantly positive coefficients for $RET_{i,t-1}$. Even with the inclusion of additional control variables in column (3), the coefficient for $DDUM_{i,t-1}$ remain of similar magnitude and statistically significant at the 5% level. That is, dirty stocks tend to earn higher returns than comparable stocks from non-dirty industries by about 26 to 30 basis points per month or 3.2 to 3.6% per year. As suspected, we observe a slightly increased $DDUM_{i,t-1}$ coefficient when including a sin stock dummy variable in column (4), which captures the return effect of shunned sin stocks.

Next, we examine the financial performance of dirty stocks in the years following the 2007-2009 financial crisis, when analyses of institutional holdings and analyst coverage revealed the highest degree of dirty industry shunning. Column (5) of Table 8 shows that c_{DDUM} almost doubles for the period 2008 to 2015 compared to the coefficients reported in columns (1) to (4) and is statistically significant at the 10% level. The coefficient of 0.4882 corresponds to dirty stocks outperforming other stocks by 49 basis points per month or 6.0% per year, after controlling for firm characteristics. This suggests that dirty stock returns are particularly high when the degree of shunning is at its highest.

3.4 Time-Series Returns

Based on the findings that dirty stocks tend to outperform non-dirty stocks in the cross-section, we test a trading strategy that takes advantage of these return differences. For this purpose, we build a value-weighted difference portfolio that is long in dirty stocks and short in non-dirty stocks. The time-series returns of the difference portfolio are regressed against commonly used risk factors in asset pricing theory. The simplest model studied is the CAPM:

$$DIFF_t = \alpha_{CAPM} + \beta_{MRP} \cdot MRP_t + \epsilon_t, \quad t = 1, \dots, T \tag{4}$$

where $DIFF_t$ denotes the return of the difference portfolio, MRP_t is the value-weighted market portfolio return net of the risk-free rate, and ϵ_t is an error term that is uncorrelated with MRP_t. The coefficient of interest is α_{CAPM}, which measures the excess return of the difference portfolio. In the further procedure, additional risk factors are included as independent variables in the regression equation (4). Adding size and growth/value factors gives the Fama and French (1992) 3-factor model. The forth model we estimate is the Carhart (1997) 4-factor model, which additionally includes a momentum factor:

$$\begin{aligned} DIFF_t = \alpha_{4F} + \beta_{MRP} \cdot MRP_t + \beta_{SMB} \cdot SMB_t + \beta_{HML} \cdot HML_t \\ + \beta_{MOM} \cdot MOM_t + \epsilon_t, \quad t = 1, \ldots, T \end{aligned} \tag{5}$$

where $DIFF_t$ again denotes the return of the difference portfolio, α_{4F} the portfolio's excess return, β's the independent factor loadings, and ϵ_t the measurement error.

The first time-series regression is run for the same period as the cross-sectional regression from July 1965 to December 2020 and the results are shown in Table 9. We find positive alphas across all models but they are statistically significant only in column (2) and (3). The return of the difference portfolio shows a negative exposure to the market return, indicating that the dirty stock portfolio is less market sensitive than the non-dirty stock portfolio. Statistically significant negative size factor loadings in all model specifications indicate that the difference portfolio is weighted toward large-cap stocks. When adding MOM_t in column (4), the alpha declines to 5 basis points and is no longer statistically significant, suggesting that the momentum factor explains part of our portfolio's excess return.

Table 9: Difference portfolio time-series return regressions 1965-2020

	(1)	(2)	(3)	(4)
ALPHA	0.1273	0.1554*	0.1278*	0.0534
	(0.0848)	(0.0795)	(0.0716)	(0.0663)
MRP	−0.3146***	−0.2570***	−0.2471***	−0.2293***
	(0.0406)	(0.0406)	(0.0379)	(0.0358)
SMB		−0.2787***	−0.2696***	−0.2680***
		(0.0305)	(0.0335)	(0.0310)
HML			0.0843	0.1178*
			(0.0699)	(0.0710)
MOM				0.0866*
				(0.0444)
Observations	666	666	666	666
R^2	0.2822	0.3759	0.3837	0.4010

This table reports time-series return regression results of the value-weighted difference portfolio for the period July 1965 to December 2020. The difference portfolio is long in dirty stocks and short in non-dirty stocks. Variables are defined as in Table 2. Newey-West (L=12) standard-errors are given in parentheses. Significance codes: *p<0.1, **p<0.05, ***p<0.01.

In our second time-series regression, we control for the fact that TRI reporting requirements were first introduced in 1987 and extended in 1998. It was not until 1998 that firms outside the manufacturing sector became subject to reporting requirements on a larger scale.[11] Additionally, annual TRI reports are usually published 6 months after the end of the respective reporting year.[12] To ensure that the necessary data are already available at the time the trading strategy is implemented, we re-run the time-series return regression for the period July 1999 to December 2020. Results are displayed in columns (1) to (3) of Table 10. The alphas generated by the difference portfolio are positive across the three asset pricing models tested. In the most conservative 4-factor model α_{4F} takes the value 0.2222 and is statistically significant at the 5% level. This corresponds to an abnormal return of the difference portfolio of 2.7% per year.

Table 10: Implementable trading strategy return in the time-series

		1999-2020			1989-2020	
	(1)	(2)	(3)	(4)	(5)	(6)
ALPHA	0.2296*	0.2708**	0.2222**	0.2558**	0.2451**	0.1847*
	(0.1393)	(0.1121)	(0.1073)	(0.1260)	(0.1117)	(0.1015)
MRP	−0.4451***	−0.4003***	−0.3566***	−0.3993***	−0.3504***	−0.3242***
	(0.0729)	(0.0625)	(0.0701)	(0.0586)	(0.0499)	(0.0506)
SMB		−0.2179***	−0.2361***		−0.2518***	−0.2533***
		(0.0507)	(0.0511)		(0.0404)	(0.0377)
HML		0.1261	0.1623		0.0588	0.0905
		(0.1048)	(0.1026)		(0.0971)	(0.1016)
MOM			0.0947			0.0783
			(0.0633)			(0.0605)
Obs.	258	258	258	378	378	378
R^2	0.3703	0.4398	0.4581	0.3295	0.4029	0.4156

This table reports time-series return regression results of the value-weighted difference portfolio for different time periods. For the period July 1999 to December 2020 (columns (1) to (3)), the long portfolio contains all six dirty industries. For the subperiod July 1989 to June 1999 of the period July 1989 to December 2020 (columns (4) to (6)), the long portfolio only contains the four dirty industries that were required to report during this time period (i.e., chemical, primary metal, paper, and petroleum and coal products manufacturing). The other variables are defined as in Table 2. Newey-West (L=12) standard-errors are given in parentheses. Significance codes: $^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

In columns (4) to (6) of Table 10 we show results for the period July 1989 to December 2020. For this purpose, we construct the difference portfolio in such a way that the long

[11] see 1998 TRI Public Data Release Report: https://www.epa.gov/sites/default/files/2018-12/documents/1998_pdr_complete_report.pdf (January 31, 2022).

[12] An exception is the first TRI report, which was not released until July 1, 1989.

portfolio contains only the four dirty industries that were already required to report in the period from July 1989 to June 1999. As of July 1999, we add the two remaining dirty industries to our long portfolio of similar magnitude. The regression results point in the same direction as before and we again find consistently positive and statistically significant alphas for the difference portfolio.

Overall, our time-series regression results indicate that dirty stocks identified from TRI data outperform non-dirty stocks. The fact that the alphas of the time series for the time period 1965 to 2020 are not consistently significant may have two possible causes. First, emission related screening was barely possible in the early years due to incomplete environmental information and the absence of TRI data. Second, the shunning of dirty stocks became stronger over the sample period and positive return effects did not emerge in the early years. The latter argument is strengthened by the observation that institutional ownership and analyst coverage of dirty stocks has declined relative to non-dirty stocks in recent decades. However, significant abnormal returns of dirty stocks for the time period 1965 to 2020 in cross-sectional regressions suggest that investors have shunned dirty stocks already before the introduction of the TRI.

4 Discussion

Our analyses find two main results. First, dirty stocks are held in lower proportions by institutional investors and receive less analyst coverage. Second, dirty stocks earn higher returns after controlling for commonly used risk factors. A key question concerning the second finding is whether these returns are abnormal and can be explained by non-pecuniary preferences of investors or whether they represent a fair compensation for systematic environmental risk that is not captured by common risk factors.

The results for the institutional ownership and analyst coverage speak in favor of a segmented capital market in which an important group of investors is unwilling to hold dirty stocks. Since we find shunning at the industry-level, we expect the return effect to become visible for the aggregate portfolio of all dirty stocks. The environmental risk-hypothesis, however, attempts to identify a risk factor that explains stock returns by firm specific risks across industries. By accounting for TRI emissions and emission intensities at the firm-level, we can mitigate the chance that part of the return effects we observe is explained by firm-specific risk factors, even though firms with high emissions are concentrated in some dirty industries. Still, the fact that we obtain higher risk-adjusted returns for entire dirty industries clearly points to a segmented capital market in the sense of the shunned-stock hypothesis. The observation that abnormal returns increase when the shunning of stocks becomes stronger (which is the case during and after the financial crisis of 2007-2009), stresses this impression. Finally, the fact that the return of the difference portfolio is actually negative during the investigation period 1965-2020 (see *DIFF* in Panel D of Table 2) even suggests that firms from dirty industries are less risky than firms from other industries.

The model of Heinkel et al. (2001) suggests that exclusionary environmental investments can create incentives for firms to reduce their emissions in order to gain access to a broader range of investors and benefit from lower costs of capital. Similarly, Pástor et al. (2021) argues that investors' tastes lead firms to become greener, thereby shifting real investments from low- to high-CEP firms. Both models assume that firms actually gain access to lower costs of capital when they reduce their emission levels. This assumption is questionable if investors do shun entire dirty industries regardless of firms' individual CEP performance. In a segmented capital market, firms operating in a dirty industry pay higher costs of capital, irrespective of their actual emissions. Even through strong CEP, a firm cannot escape the shunning of its entire industry. Hong and Kacperczyk (2009) find a related pattern for sin stocks, arguing that there is little these firms can do to be spared from the neglect of their industries. If there are no or only weak financial incentives to reduce emissions, firms might refrain from investments in environmental improvements.

5 Conclusion

This study investigates whether a shunned-stock effect is observable for the E in ESG investing. As opposed to previous studies on this topic, we avoid the use of ESG ratings and use TRI emission data as measure for environmental performance. Moreover, we assume that investors with non-pecuniary preferences shun stocks on industry-level rather than firm-level. In line with the shunned-stock hypothesis, we find firms operating in dirty industries are owned in lower proportions by institutional investors and receive less analyst coverage. We study financial effects of this market segmentation and find that stocks from dirty industries tend to outperform stocks from other industries in the cross-section from 1965 to 2020. The outperformance is particularly pronounced when the degree of shunning is high, as observed during and in the years following the 2007-2009 financial crisis. Given the high abnormal returns of dirty stocks, we test a trading strategy based on emission-related industry portfolios. Implementation becomes feasible from 1989 onwards, when the EPA published its first TRI report. Our long-short portfolio yields an abnormal annual return of 2.2 to 3.1% for the time period 1989 to 2020 that cannot be explained by common risk factors.

Previous studies struggle to provide a clear picture on the effects of environmental preferences on stock markets. We suspect that this is due to the fact that not all investors are willing and able to perform an accurate environmental performance assessment of the individual firm. Instead they might apply coarser evaluation schemes and judge firms based on their industry affiliation, which results in a systematic shunning of firms operating in dirty industries. Although we do not quantify the number of investors using such sparse models, our results suggest that their impact on firms' costs of capital is of economic significance. This is a crucial issue because firms operating in a dirty industry might lack the incentive for investments in CEP improvements if these won't be fully rewarded by the capital market.

References

Angel, J. J. and Rivoli, P. (1997). Does ethical investing impose a cost upon the firm? a theoretical perspective. *The Journal of Investing*, 6(4):57.

Bali, T. G., Engle, R. F., and Murray, S. (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Barber, B. M., Morse, A., and Yasuda, A. (2021). Impact investing. *Journal of Financial Economics*, 139(1):162–185.

Berg, F., Koelbel, J. F., and Rigobon, R. (2020). Aggregate confusion: The divergence of esg ratings. *Working Paper*.

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

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

Chatterji, A. K., Durand, R., Levine, D. I., and Touboul, S. (2016). Do ratings of firms converge? implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8):1597–1614.

Chava, S. (2014). Environmental externalities and cost of capital. *Management Science*, 60(9):2223–2247.

de Marchi, S. and Hamilton, J. T. (2006). Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory. *Journal of Risk and Uncertainty*, 32(1):57–76.

Dorfleitner, G., Halbritter, G., and Nguyen, M. (2015). Measuring the level and risk of corporate responsibility – an empirical comparison of different esg rating approaches. *Journal of Asset Management*, 16(7):450–466.

Drempetic, S., Klein, C., and Zwergel, B. (2020). The influence of firm size on the esg score: Corporate sustainability ratings under review. *Journal of Business Ethics*, 167(2):333–360.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2007). Disagreement, tastes, and asset prices. *Journal of Financial Economics*, 83(3):667–689.

Fernando, C. S., Sharfman, M. P., and Uysal, V. B. (2017). Corporate environmental policy and shareholder value: Following the smart money. *Journal of Financial and Quantitative Analysis*, 52(5):2023–2051.

Friede, G., Busch, T., and Bassen, A. (2015). Esg and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4):210–233.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gibson Brandon, R., Krueger, P., and Mitali, S. F. (2021). The sustainability footprint of institutional investors: Esg driven price pressure and performance. *Working Paper*.

Gompers, P. A. and Metrick, A. (2001). Institutional investors and equity prices. *The Quarterly Journal of Economics*, 116(1):229–259.

Heinkel, R., Kraus, A., and Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis*, 36(4):431–449.

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

Hsu, P.-H., Li, K., and Tsou, C.-Y. (2020). The pollution premium. *Working Paper*.

Krueger, P., Sautner, Z., and Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, 33(3):1067–1111.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3):483–510.

Pástor, Ľ., Stambaugh, R. F., and Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1):35–55.

Shumway, T. (1997). The delisting bias in crsp data. *The Journal of Finance*, 52(1):327–340.

von Wallis, M. and Klein, C. (2015). Ethical requirement and financial interest: a literature review on socially responsible investing. *Business Research*, 8(1):61–98.

cfr working paper series

centre for financial research cologne

cfr working papers are available for download from www.cfr-cologne.de.

2022

No.	Author(s)	Title
22-07	T. Bauckloh, V. Beyer, C. Klein	Does it Pay to Invest in Dirty Industries? – New Insights on the Shunned-Stock Hypothesis
22-06	J. Balthrop and G. Cici	Conflicting Incentives in the Management of 529 Plans
22-05	I. T. Ivanov, T. Zimmermann, N. W. Heinrich	Limits of Disclosure Regulation in the Municipal Bond Market
22-04	M. Ammann, A. Cochardt, S. Straumann, F. Weigert	Back to the Roots: Ancestral Origin and Mutual Fund Manager Portfolio Choice
22-03	A. Betzer, J. Gider, P. Limbach	Do Financial Advisors Matter for M&A Pre-Announcement Returns?
22-02	S. Lesmeister, P. Limbach, P.R. Rau, F. Sonnenburg	Indexing and the Performance-Flow Relation of Actively Managed Mutual Funds
22-01	T. Bauckloh, C. Klein, T. Pioch, F. Schiemann	Under Pressure: The Link between Mandatory Climate Reporting and Firms' Carbon Performance

2021

No.	Author(s)	Title
21-11	V. Agarwal, H. Ren, K. Shen, H. Zhao	Redemption in Kind and Mutual Fund Liquidity Management
21-10	N.C. Brown, W. B. Elliott, R. Wermers, R. M. White	News or noise: Mobile internet technology and stock market activity
21-09	V. Agarwal, B. Barber, S. Cheng, A. Hameed, A. Yasuda	Private Company Valuations by Mutual Funds
21-08	T.G. Bali, H. Beckmeyer, M. Moerke, F. Weigert	Option Return Predictability with Machine Learning and Big Data
21-07	F. Chabi-Yo, M. Huggenberger, F. Weigert	Multivariate Crash Risk
21-06	V. Agarwal, H. Aslan, L. Huang, H. Ren	Political Uncertainty and Household Stock Market Participation
21-05	G. Cici, P. Zhang	On the Valuation Skills of Corporate Bond Mutual Funds

21-04	V. Agarwal, P. Hanouna, R. Moussawi, C.W. Stahel	Do ETFs Increase the Commonality in Liquidity of Underlying Stocks?
21-03	M. Saglam, T. Tuzun, R. Wermers	Do ETFs Increase Liquidity?
21-02	C. Andres, D. Bazhutov, D. Cumming, P. Limbach	Does Speculative News Hurt Productivity? Evidence from Takeover Rumors
21-01	T.G. Bali, F. Weigert	Hedge Funds and the Positive Idiosyncratic Volatility Effect

2020

No.	Author(s)	Title
20-14	V. Agarwal, Y. Lu, S. Ray	Are hedge funds' charitable donations strategic?
20-13	M. Jagannathan, W. Jiao, R. Wermers	International Characteristic-Based Asset Pricing
20-12	E. Theissen, L. Zimmermann	Do Contented Customers Make Shareholders Wealthy? - Implications of Intangibles for Security Pricing
20-11	M. Hendriock	Implied Cost of Capital and Mutual Fund Performance
20-10	J. Fink, Stefan Palan, E. Theissen	Earnings Autocorrelation and the Post-Earnings-Announcement Drift – Experimental Evidence
20-09	E. Theissen, C. Yilanci	Momentum? What Momentum?
20-08	V. Agarwal, L. Jiang, Q. Wen	Why Do Mutual Funds Hold Lottery Stocks?
20-07	V. Agarwal, S. Ruenzi, F. Weigert	Unobserved Performance of Hedge Funds
20-06	M. Ammann, S. Fischer, F. Weigert	Factor Exposure Variation and Mutual Fund Performance
20-05	P. Limbach, P. R. Rau, H. Schürmann	The Death of Trust Across the U.S. Finance Industry
20-04	A. Y. Chen, T. Zimmermann	Open Source Cross-Sectional Asset Pricing
20-03	L. Ahnert, P. Vogt, V. Vonhoff, F. Weigert	Regulatory Stress Testing and Bank Performance
20-02	P. Schuster, E. Theissen, M. Uhrig-Homburg	Finanzwirtschaftliche Anwendungen der Blockchain-Technologie
20-01	S. Ruenzi, M. Ungeheuer, F. Weigert	Joint Extreme Events in Equity Returns and Liquidity and their Cross-Sectional Pricing Implications

This document only covers the most recent cfr working papers. A full list can be found at www.cfr-cologne.de.

centre for financial research
cfr/university of cologne
Albertus-Magnus-Platz
D-50923 cologne
fon +49(0)221-470-6995
fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de