The Future of Emissions

Jules H. van Binsbergen[†] and Andreas Brøgger[‡]

November 22, 2025

Abstract

We argue for the introduction of firm-level emission futures contracts as a novel way of assessing the real impact of ESG initiatives. Our measure is based on the forward-looking market-based valuation of firm-level CO_2 emissions. We establish both theoretically and empirically that backward-looking subjective ratings are limited to the extent that they fail to capture future reductions in emissions. We show evidence that although lower emissions have predicted higher E ratings, higher E ratings have predicted higher, not lower, emissions. As such, by following these subjective ratings, investors may have inadvertently allocated their money to firms that pollute more, not less. We discuss several applications of our new measure, including executive pay and investment management.

We thank Darrell Duffie, Jane Li (Discussant), Julien Penasse and participants at AFA 2024, Blackrock, and Copenhagen Business School for helpful comments. Andreas Brøgger gratefully acknowledges support from the Center for Financial Frictions (FRIC), grant no. DNRF102.

[†]University of Pennsylvania and NBER: julesv@wharton.upenn.edu.

[‡]Rotterdam School of Management at Erasmus University: brogger@rsm.nl.

1 Introduction

Socially conscious investment strategies have gained much popularity in the past two decades. While various different approaches have been proposed, one popular way of attempting to achieve social change is by divesting from companies that do not meet certain social criteria. The main mechanism through which such a divestment could have impact is through the stock price. By divesting, socially conscious investors hope to decrease the firm's stock price, implying that for a given number of shares issued, the firm raises less capital. That is, the intended consequence of divestment is to increase targeted firms' cost of capital (Berk and van Binsbergen, 2021). This higher cost of capital makes fewer real investment projects a positive net present value (NPV) undertaking, implying a lower growth rate of such firms going forward. Eventually, this lower growth rate reduces the fraction of targeted firms in the economy in the long run. In addition, the threat of this stock price mechanism could induce targeted firms (sometimes called 'brown' firms) to shed their undesirable habits and become "green" firms.

In this paper, we propose a new market-based forward-looking measure for the real impact of social investing based on emission futures contracts. If backward-looking emission measures perfectly predicted forward-looking emissions at the firm level there would be no need for such an additional forward-looking measure. However, we establish both theoretically and empirically that the current backward-looking subjective ratings-based system implies that past emissions do not predict future emissions, quite the opposite. As such, by following these subjective ratings, investors may inadvertently allocate their money to firms that will pollute more not less.[1] In particular, we find that even though (a) the fraction of capital allocated to ESG investing has tremendously increased in the past few decades, and (b) firms have on average reduced their CO_2 emissions over time, the CO_2 emission reductions have preceded increases in ESG investing. Second, when firms are admitted to a leading ESG index (FTSE

[1]Indeed, investors are increasingly sceptical of the current ESG data. Only 20% say that they trust the ESG statements that companies make, lowered from half two years ago, according to a survey of 20,000 consumers and 2,500 executives across 22 industries and 34 countries (IBM 2023). Additionally, executives cite inadequate data as the top barrier holding back ESG progress, even more so than regulatory barriers. 60% of executives say that they lack the ability to access and understand ESG data as they have to make tradeoffs between financial and ESG objectives. This means they cannot accurately predict which plans will improve outcomes and return on investment. This also means that while 95% of organizations have made ESG propositions, only 10% say that they have made significant progress progress towards their goals.

USA 4Good) they tend to increase rather than decrease their CO_2 emissions. Third, although lower emissions predict higher E scores (i.e., the first category in "ESG"), higher E scores do not predict lower emissions.

From a theoretical perspective, we contribute to the literature by introducing a general framework of value creation that nests sustainable investing as well as standard firm optimization. We then introduce a model, which features backward-looking ESG ratings combined with standard firm optimization, that can explain the empirical regularities featured above. In the model firms are incentivized to increase their ratings, but not necessarily to improve their social impact. This indeed implies that after improving their ranking, it is optimal to increase emissions.

These theoretical and empirical results raise the question of what mechanism would optimally induce firms to reduce emissions and have a real impact. We argue that a mechanism based on a market-based forward-looking emissions measure improves incentives. Additionally, we show that even without a measurable stock price effect from sustainable investing, linking managerial pay to this new continuous measure aligns incentives and optimizes real impact. That is, even if divestment would lead to measurable price effects in financial markets, currently the incentive structure may not be sufficiently conducive to affecting real change.

Specifically, we propose to base the E measure of a particular firm on the pricing of a new asset class, what we term emission futures. An emission future pays out the dollar value of a firm's emissions at a given future date, and is based on (1) the future path of the traded price of carbon per metric ton and (2) the quantity in metric ton of a firm's future carbon emissions for a given calendar year.[2] The futures price thus reflects the discounted expected value of a firm's emissions.[3] By translating this value into a new E

[2]We propose to have the Emission Future be based on Scope 1 emissions to ease the contracts measurability and enforcability, however as the quality of more comprehensive scopes increase these could also be considered. Scope 3 emissions are hard to assess, due to the difficulty of collecting high-quality data on type or volume of emissions. Scope 2 emissions, as well as scope 3 emissions, fall outside a company's direct control, making them hard to manage. Additionally, scope 2 and 3 emissions will be accounted for by several companies, which raises the question of who should be responsible for them.

[3]We acknowledge that our proposed measure has its limitations, however we see these to be greatly reduced relative to other measures available and not necessarily greater than for any other derivatives contract. Specifically, a firm may be incentivized to sell their own Emission Futures and misreport on their emissions, however this is the same as with dividend futures, and there exists insider trading laws to protect against this. Uncertainty of the underlying data may be an issue, but we do not see this to be any different than other contracts that are traded today. For example take the

measure we get a meaningful system that incorporates the expected path of a specific outcome variable that is relevant and measurable. Furthermore, emission futures allow us to measure the impact of corporate actions and investor activism through what we term green impact. The green impact is based on a firm's reduction in the term structure of its future price, which measures reductions in emissions value by calendar year relative to their previous (risk-neutral) expected path. By comparing this improvement in its term structure to the changes in the term structure of the traded price of carbon a firm's expected quantity reduction can be measured. This improvement can then be attributed to the impact by the firm, rather than the economy gradually adapting low-carbon technologies. That is, the unexpected return on a firm's emission futures allows for the measurement of corporate actions that genuinely affect the path of emissions that the firm was on.

Our paper fits into a larger literature that investigates the extent to which socially conscious investors manage to reduce the targeted companies' carbon footprint. Indeed, there is growing evidence that although socially concerned investors hold firms with higher ESG scores and lower carbon intensity, it is not necessarily socially concerned investors, such as pension plans, that are responsible for this reduction (Hong and Kacperczyk, 2009, Heath et al., 2021, 2022, Brøgger and Kronies, 2022, Noh and Oh, 2020, Akey and Appel, 2019). Our finding that greener firms have less impact is also in line with psychological findings on 'Moral licensing' (Sachdeva, Iliev and Medin, 2009), the tendency of entitlement to do something bad because we've done something good. Moral licensing explains why ethics professors are more likely to say that eating meat is wrong, but no less likely to eat meat (Schwitzgebel and Rust, 2014), and ethics professors steal more books than colleagues outside of their field (Schwitzgebel, 2009). This evidence is consistent with the idea that while an improved ESG score causes socially concerned investors to become owners of such firms, the reverse causality is

CPI futures, which rely on government reporting of the consumer price index, which in turn is based on surveys, price samples, and index weights. Misreporting is further disincentivized through legal means. The legal framework of emission monitoring, reporting, and verification has already been developed after the introduction of greenhouse gas emission allowance trading within the European Union (See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R2066 and https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R2067), and the Greenhouse Gas Reporting Program in the United States, which equally requires the independent verification of emissions data reports by a third party (See, for example, https://ww2.arb.ca.gov/verification). Misreporting has already had legal consequences as was the case for Chevron in 2011 (https://ww2.arb.ca.gov/verification). See also the greenwashing investigation raid on Deutche bank following ESG misreporting allegations, which already resulted in the resignation of CEO Asoka Wöhrmann, even though the legal consequences have not yet concluded.

more elusive. Given that the stated objective of socially conscious investors is to affect social change, this state of affairs may be less than desirable.

There is also an emerging literature that evaluates the reliability of ESG ratings. In particular, Berg, Koelbel and Rigobon (2019) investigate the divergence of such rating ratings based on data from six prominent rating agencies including KLD, Sustainalytics, Moody's ESG, S&P Global, Refinitiv, and MSCI. They find large divergences between rating agencies and raise important issues related to the scope, measurement, and weight of the various inputs used to construct such ratings. Berg, Fabisik and Sautner (2020) find that Refinitiv ESG scores have been subject to ongoing changes to past scores, further undermining their reliability. We circumvent these issues by simply focusing on commonly-used ESG indices that are presumably replicated and followed by an important group of investors. Therefore, regardless of whether these indices perfectly capture the underlying desired dimensions of the ESG variables, investors following and replicating such indices, and thereby basing their buy and sell decisions on them, are implicitly condoning their measurement. For the purpose of our empirical strategy, the only relevant factor is that investors are following these commonly-used indices.

Indeed new research shows that investing according to the current ESG scores may not be conducive for fostering impact. Cohen, Gurun and Nguyen (2020) find that it is often the brownest firms that are innovating to reduce future carbon emissions. Bams and van der Kroft (2022) find that due to information asymmetry investors invest according to ESG scores instead of sustainable performance. This leads to firms inflating ESG scores and reducing cost of capital, to the extend that ESG is negatively related to sustainable performance. This problem is exacerbated as investors investing into ESG funds, from calculating the average ESG score of the portfolio holdings, pressures the funds to pick stocks based on ESG scores instead of doing their own research (Edmans, Levit and Schneemeier, 2022).

Another issue with current ESG scores is that they are prone to "cheap talk". Indeed, previous work has shown evidence of cheap talk, including social index funds not voting in favour of ESG policies. Bingler et al. (2022a) find using a neural algorithm trained to detect cheap talk that voluntary disclosure is mainly cheap talk and cherry picking. Bingler et al. (2022b) find using the same algorithm that "institutional ownership, targeted institutional investor engagement, materiality and downside risk disclosures are associated with less cheap talk. Signaling by publicly supporting the

TCFD is associated with more cheap talk". While Curtis, Fisch and Robertson (2021) find that green funds on average vote for green proposals, Michaely, Ordonez-Calafi and Rubio (2021) and Li, Naaraayanan and Sachdeva (2021) find that green funds actually vote strategically in the sense that they vote for sustainable proposals that would pass anyway, but not those they could influence to be enacted.

The findings in our paper are in line with experimental evidence that investors are willing to pay a premium for sustainable investments, regardless of whether their investment has an impact on the projects coming to fruition (Bonnefon et al., 2022), and regardless of additional impact (Heeb et al., 2022). The findings are also in line with current regulatory proposals by the securities exchange comission (SEC) in the US and the EU council in Europe. These proposals include the EU taxonomy for sustainable activities, which proposes filters based on levels as well as an Excel tool to help direct investments. Additionally, the sustainable finance disclosure regulation (SFDR) requires ESG fund managers (referred to as Article 8 and 9 funds) to display sustainability indicators, such as emission levels, to show that they are in accordance with the EU Taxonomy and an explanation of the 'do no harm' principle. This has to be done regularly as well as in their prospectus. Otherwise the fund managers have to show how their fund is not proposing to be an ESG fund. These regulations are supported by the new corporate sustainability reporting directive (CSRD), which requires firms to report levels on these social outcomes, such as emissions. The ECB now also aims to tilt their purchase programs towards greener firms as well as reducing the collateralability of browner firms. In other words the placebo effects of sustainable investment is more important than the actual effectiveness of them.

2 Theory

We first split total value into an internal value to investors and an external value. Using the fact that impact is generally defined as the reduction of the negative externality, we evaluate the effectiveness of backward-looking ESG ratings versus forward-looking measures in achieving impact.

2.1 A Firm's Internal and External Value

We start by defining the total value of a firm i to society at time t as the present value of its total future dividends $D_{i,t,n}$ at horizons n discounted at the rate $\mu_{i,t,n}$ back to time t as

$$V_{i,t} = \sum_{n=0}^{\infty} \frac{D_{i,t,n}}{\exp(n\mu_{i,t,n})}.$$

The value generated by a firm either goes to the firm's investors or to the rest of the economy through the firm's externalities, which in principle can be both positive or negative. The value that goes to the firm's investors are the internal dividends, which we name cash dividends, and externalities are the external dividends, termed "externality dividends" hereafter.[4] This means that the total value of a firm is its internal and external value, which is given by the sum of its cash and externality dividends as

$$V_{i,t} = \underbrace{\sum_{n=0}^{\infty} \frac{D^c_{i,t,n}}{\exp(n\mu^c_{i,t,n})}}_{V^I_{i,t}} - \underbrace{\sum_{n=0}^{\infty} \frac{D^e_{i,t,n}}{\exp(n\mu^e_{i,t,n})}}_{-V^E_{i,t}}. \tag{1}$$

In particular as the investors value the dividends they receive, the price they pay only reflects the cash dividends:

$$P_{i,t} = V^I_{i,t}, \tag{2}$$

where we normalize the number of outstanding shares to 1, and we assume wlog and for ease of exposition that the firms are all equity financed.

An example of the importance of this distinction is the seminal work of Modigliani and Miller (1958). Modigliani and Miller show that the internal value of the firm is the same irrespective of financing decisions with regard to debt and equity. This is true in absence of other frictions, and they go on to show that if debt receives a tax benefit, then issuing debt actually increases the internal value by that tax benefit. However, this comes at the cost of an external value loss in tax income, that would otherwise have been redistributed to the economy. Hence, our framework allows us to specifically see how economic incentives lead to agents optimising and value being transferred from one group to another.

[4] As negative externalities are most commonly analysed, they are the default externality in this setting. At the same time it is flexible to incorporate positive externalities, in which case the externality dividends would have a negative sign.

Sustainable Investing. We now define sustainable investing within this framework. Sustainable investing is investing that properly accounts for externalities and thus maximises the total value of the firm to society. An important understudied consideration is the time horizon over which sustainable goals are achieved. In our value definition, the whole present value of future externalities are incorporated in the framework. This implies that improvements that take a substantial amount of time to realise are still counted towards the sustainability objectives of the firm. Further, past realised externality dividends play no role in this valuation exercise. This already illustrates an important difference between backward-looking ESG ratings and forward-looking optimal decision making that maximises a firms value to society.

One way to measure impact is to compute the decrease in the discounted horizon specific expected external dividends. Recall that t denotes the current period, n denotes the horizon (number of years in the future), and i denotes the firm, then the impact is simply given by

$$\tilde{I}_{i,t,n} \equiv \frac{\mathbb{E}_{i,t-1} D^e_{i,t-1,n+1}}{\exp((n+1)\mu^e_{i,t-1,n+1})} - \frac{\mathbb{E}_{i,t} D^e_{i,t,n}}{\exp(n\mu^e_{i,t,n})}.$$

As an example consider the one-year horizon. Impact is then simply given by the difference between the discounted expected external dividend one year ago and its realisation:

$$\tilde{I}_{i,t,1} = \mathbb{E}_{i,t-1} D^e_{i,t-1,1} \exp(-\mu^e_{i,t-1,1}) - D^e_{i,t,0}.$$

One downside of using these measures of impact is that a negative impact is measured simply due to the normal expected return the asset earns. By using futures values as opposed to spot values we can remove the risk-free part of this expected return. Further, we will show that the risk-premium on this asset is likely small at least as measured by conventional exposures to market risk. Concretely, let $F_{i,t,n}$ denote the futures price of the externality dividend paid out in n periods by firm i at time t. If we then indeed assume that the risk-premium is sufficiently small, we can simply define impact as the negative dollar return on the future:

$$I_{i,t,n} \equiv F_{i,t-1,n+1} - F_{i,t,n}, \tag{3}$$

which under the above stated assumptions is also equal to

$$I_{i,t,n} = \mathbb{E}_{i,t-1} D^e_{i,t-1,n+1} - \mathbb{E}_{i,t} D^e_{i,t,n}. \tag{4}$$

That is, under the above stated assumptions the futures price is simply equal to the expected value of the externality dividend.

Finally, the value of impact is then simply equal to the sum of each year's impact measure across all horizons:

$$V_{i,t}^{\text{Impact}} \equiv \sum_{n=1}^{\infty} I_{i,t,n}. \tag{5}$$

2.2 ESG Measures, Impact, and Misallocation of Capital

We can use this framework to theoretically analyse the realised impact of two regimes as well as the extent to which the measures (R and M) capture them. In particular we wish to compare a regime where firms respond to subjective backward-looking ESG ratings that are based on a firms history of externalities, what we label regime R (for rating), to a regime that uses market-based forward-looking ESG measures, what we label regime M. The impact at horizon n between the two regimes is then

$$I_{i,t,n}^{M,R} \equiv \mathbb{E}_{i,t}\, D_{i,t,n}^{e,R} - \mathbb{E}_{i,t}\, D_{i,t,n}^{e,M}, \tag{6}$$

where the dividend $D_{i,t,0}^{\mathcal{M}}$ will be the prevailing dividend for firm i at time t and horizon n under regime $\mathcal{M}$. Specifically, we have that the realised impact for the 0-horizon is

$$I_{i,t,0}^{M,R} = D_{i,t,0}^{e,R} - D_{i,t,0}^{e,M}, \tag{7}$$

which simply means that impact can be achieved by moving to a new measure which delivers a lower equilibrium externality dividend.

We can also analyse whether moving to forward-looking measures improves the investors' allocation to their desired firms, that is, increasing the allocation of capital to the firms that deliver impact. Consequently, if one invests with a noisy or biased rating, it may lead to capital being allocated to firms which may pollute more, not less.[5] Specifically, consider an impact investor who tilts their portfolio according to their expected impact of the firm, then a valid misallocation measure is the absolute difference between the allocation under measure $\mathcal{M}$ and the optimal allocation:

$$|D_{i,t,n}^{e} - \mathbb{E}_{i,t}^{\mathcal{M}}\, D_{i,t,n}^{e,\mathcal{M}}| = |\epsilon_{i,t,n}^{\mathcal{M}}|,$$

[5]Misallocation can also be interpreted as a welfare loss to the investor, where the exact loss will depend on the extend that an investor cares about the impact of his investments.

where $\epsilon^{\mathcal{M}}$ is the estimation error under measure $\mathcal{M}$. Specifically, the improvement in misallocation from rating R to measure M can be written as

$$\tilde{I}^{M,R}_{i,t,n} \equiv |\epsilon^{R}_{i,t,n}| - |\epsilon^{M}_{i,t,n}|. \tag{8}$$

Which taking the average across all firms i can be rewritten as

$$\tilde{I}^{M,R}_{t,n} = \sqrt{MSE^{R}_{t,n}} - \sqrt{MSE^{M}_{t,n}}, \tag{9}$$

where $MSE^{\mathcal{M}}_{t,n}$ is the mean squared error under measure $\mathcal{M}$ for horizon n evaluated at time t. The equation shows that the improvement in the allocation can be measured by comparing the predictive ability of the forward-looking measure (MSE) for future externality dividends, with the predictive ability of the the backward-looking rating. Suppose that there exists a forward-looking market-based measure which is unbiased, then the MSE of this market measure is smaller than the MSE of the backward-looking measure, as market prices will reflect more information than that captured by the backward-looking variables used in generating the rating.

Next, we set up a model to explain the sources of changes to the equilibrium externality level.

2.3 Model of ESG Investing under Different ESG Measures

In this section, we present a model of the costs and benefits that for firms trade off when trying to affect their ESG Scores. After, we will consider how these incentives affect impact through firms' optimisation behaviour.

The Firm's Problem. Firm i maximises its internalised value as given by Equation 2 at time t by aiming for an ESG score $E_{i,t}$. That is, their objective is

$$\max_{E_{i,t}} \mathbb{E}_t \sum_{n=0}^{\infty} \frac{D^{c}_{i,t,n}(E_{i,t+n})}{\exp\left(n\mu^{c}_{i,t,n}(E_{i,t+n-1})\right)}. \tag{10}$$

Firms can increase the valuation of their firm by either increasing cash dividends or reducing their discount rate. Additionally, the cash dividend is given as the difference between revenue $\text{Rev}_{i,t}$ and cost $c_{i,t}$.

A key decision for the firm is which ESG rating $E_{i,t}$ to aim for. This is relevant for the firm because both $\text{Rev}_{i,t}$, $C_{i,t}$, and $r^c_{i,t}$ may be dependent on the firm's ESG score. The revenues $\text{Rev}_{i,t}$ may be affected because governments and ESG activists may impose boycotts of low ESG products.[6] The costs $C_{i,t}$ can be affected because companies can incur costs from having a low ESG score resulting from protests or regulatory costs such as a carbon tax or the purchase of Carbon credits. Finally, the discount rate $r^c_{i,t}$ can change when investors or banks are restricted from investing in firms with low ESG scores (Zerbib, 2019, Berk and van Binsbergen, 2021, Homanen, 2022). There exists investors under strict mandates, such as pension funds. The discount rate $r^c_{i,t}$ may also be affected if investors value high ESG investments as in Pástor, Stambaugh and Taylor (2021) and Pedersen, Fitzgibbons and Pomorski (2021).

ESG Adjustment Costs. There is a cost associated with companies changing their ESG scores, for example when they have to transition to a more expensive but cleaner energy source. We model this with capital adjustment costs:

$$E_{i,t} = G_i(B_{i,t}),$$

where G_i is a general function potentially specific to each firm. Further, to get a closed form solution we need to specify the functional form for G_i. A specification that is tractable yet still quite general is

$$E_{i,t} = \frac{1}{k_i} B_{i,t}^{\eta_i},$$

where $B_{i,t+n}$ is the periodic budget needed to sustain an ESG score $E_{i,t+n}$ continuously as a fraction of their cash dividend, and k_i and η_i are strictly positive constants. The decreasing returns to scale are captured by $0 < \eta_i < 1$. Solving for the budget needed to sustain a certain ESG rating, we see that the periodic dividend after ESG costs is

$$D^c_{i,t,0}(E_{i,t}) = D^{c,0}_{i,t,0} \exp\left(-k_i E_{i,t}^{1/\eta_i}\right), \tag{11}$$

where $D^{c,0}_{i,t}$ is firm i's cash dividends before ESG related costs at time t.

ESG and the Firm's Discount Rate. There are multiple potential benefits to achieving a higher ESG score though their quantitative relevance is subject to debate (Berk and van Binsbergen, 2021). For ease of exposition we will assume here that a

[6]Examples are government-imposed policies that mandate that all energy investments are carbon neutral.

higher ESG score has a negative effect on the firm's discount rate. Concretely, let a continuum of investors j maximise their expected utility. Meaning they solve

$$U'_{j,t} \equiv max_{X_j,t}\, \mathbb{E}_t[-\exp(-a_j W_{j,t+1} - bX_{j,t})],$$

where a_j is investor j's risk avertion, $W_{j,t+1}$ is that investor's next period's wealth, and b is the non-pecuniary benefit received from their portfolio $X_{j,t}$. Importantly, b_j is the product of the investors' sustainability sentiment S_j and $g^e_{i,t}(E_{i,t})$, the greenness of the firm, which depends on the firm's ESG score. As some investors follow a full exclusion strategy based on some lower threshold and some investors tilt their portfolio gradually, we model the combined effect as

$$b = S_j g^e_{i,t} = S_j \ln(E_{i,t}).$$

The investors' optimal portfolio weights are then

$$X_{j,t} = w_{m,t} + (S_j - \bar{S})/\gamma^2 \Sigma^{-1} \ln(E_{i,t}),$$

where w_m are the market portfolio weights, and γ_j is the relative risk aversion ($\gamma_j = A_j/W_j$) for investor j, and Σ is the covariance matrix of their portfolio. In short, investors put higher weights on stocks with higher ESG scores.

The discount rate on firm i will then be

$$r^c_{i,t} = \beta^M_i r^M_t - s\ln(E_{i,t}), \tag{12}$$

where β^M_i is the beta of firm i's returns with respect to the market return r^M_t, γ_j is the relative risk aversion for investor j ($\gamma_j = A_j/W_j$), and s is equal to $\bar{S}/\bar{\gamma}$, where the bar variables indicate value-weighted averages. For example $\bar{S}$ denotes the value-weighted average sentiment in the economy.

Solution to the Firm's Problem. The first order conditions from the firm's maximisation problem (Equation 10) imply for each i, t that

$$D^{c'}_{i,t,0}(E_{i,t}) = \mathbb{E}_t\left[-r^{c'}_{i,t+1}(E_{i,t})\frac{D^c_{i,t,1}(E_{i,t+1})}{\exp\left(r^c_{i,t+1}(E_{i,t})\right)}\right].$$

At the optimal $E_{i,t}$ score, the marginal cost of increasing the score further, in terms of lower dividends, must equal the marginal benefit in firm value through an expected lower discount rate. Further, we can get from Equation 11 that the marginal cost of raising the $E_{i,t}$ score is

$$k_i E^{\frac{1-\eta_i}{\eta_i}} D^{c,0}_{i,t,0} \exp\left(-k_i E^{1/\eta_i}_{i,t}\right),$$

and from Equation 12 that increasing $E_{i,t}$ lowers the discount rate incrementally by $s/E_{i,t}$. Additionally, as the problem is symmetric across time $E_{i,t} = E_{i,t+1} = E_i$, we obtain

$$\underbrace{\frac{k_i E^{\frac{1-\eta_i}{\eta_i}} D^{c,0}_{i,t,0}}{\exp\left(k_i E^{1/\eta_i}_i\right)}}_{\text{Marginal cost}} = \underbrace{s E^{-1}_i \, \mathbb{E}_t \left[\frac{D^{c,0}_{i,t,1} \exp\left(-k_i E^{1/\eta_i}_i\right)}{\exp\left(\beta^M_i r^M_{t+1} - s \ln(E_i)\right)} \right]}_{\text{Marginal benefit}}. \tag{13}$$

Given these costs and incentives each firm will choose its optimal ESG rating as

$$E^*_i = \psi\left(s, k_i, \beta^M_i, \mu^M_t, \sigma^M_t\right)^{\phi(\eta_i, s)}, \tag{14}$$

where ψ and ϕ are positive definite and given respectively by

$$\begin{aligned} \psi &\equiv \frac{s}{k_i} \exp\left(-\beta^M_i \mu^M_t + \frac{1}{2} \beta^{M\,2}_i \sigma^{M\,2}_t\right), \\ \phi &\equiv \eta^{-1}_i - s. \end{aligned}$$

The equilibrium ESG rating E^*_i will be decreasing in ψ with a sensitivity determined by ϕ. Hence E^*_i is decreasing in k_i and the expected market return μ^M_t, and increasing in green sentiment s and the expected market risk σ^M_t. The sensitivity of these effects, in both positive and negative directions, is increasing in both s and the negative returns to scale η.

To close the model, let the E measure under regime R be given by the negative externality arising from the pollution over the same period and in regime M be the expected future negative externalities at horizon n:

$$\begin{aligned} E^R_{i,t} &= -D^e_{i,t,0}, \\ E^M_{i,t} &= \mathbb{E}_t[-D^e_{i,t,n}]. \end{aligned} \tag{15}$$

Now that we have characterised the equilibrium for any regime, we complement this firm optimization perspective with a principal-agent model that examines the investor's contracting problem. We then analyse the impact from moving from the backward-looking R regime to the forward-looking M regime.

2.4 A Principal-Agent Model of Performance Measurement and Greenwashing

The model in Section 2.2 showed how firms optimize ESG scores by trading off costs and benefits. We now complement this with a principal-agent framework that examines how investors' reliance on multiple performance signals—some of which are manipulable—affects true impact. This perspective is important because even if firms improve true emissions (effort), strategic disclosure and greenwashing can contaminate performance measurement, leading investors to misallocate capital. The model builds on the classic aggregation and linearity framework of Holmström and Milgrom (1987) but introduces a key friction: one performance signal is subject to costless manipulation by the agent.

Economic Setting. A risk-neutral principal (investor) hires a risk-averse manager to exert productive effort e that increases firm value and reduces emissions. The principal does not directly observe e, but instead receives three performance signals $s = (s_1, s_2, s_3)$, each informative about e but with different noise characteristics. Two signals provide unbiased information:

$$s_1 = e + \varepsilon_1, \qquad s_3 = e + \varepsilon_3, \tag{16}$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ are independent Gaussian shocks with precisions $p_i = \sigma_i^{-2}$. The second signal is subject to *greenwashing*, a manipulable distortion d:

$$s_2 = e + d + \varepsilon_2. \tag{17}$$

This setup captures the empirical reality that investors observe multiple performance measures: some relatively objective (e.g., verified emissions data, financial performance), and others more susceptible to strategic disclosure (e.g., voluntary ESG reports, sustainability narratives). We interpret s_2 as the manipulable ESG-type disclosure measure.

The manager chooses effort e and greenwashing d paying quadratic costs:

$$c(e,d) = \frac{1}{2}k_e e^2 + \frac{1}{2}k_d d^2, \tag{18}$$

with $k_e > 0$ and $k_d > 0$. The manager has CARA utility $u(w) = -\exp(-aw)/a$ with risk aversion $a > 0$. The principal values effort at $v > 0$ per unit but gains no benefit from greenwashing d.

Following Holmström and Milgrom (1987), the principal offers a linear contract $w = \alpha + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_3$ and chooses weights $\beta = (\beta_1, \beta_2, \beta_3)$ to maximize expected surplus subject to the manager's incentive compatibility and participation constraints.

Figure 1 summarizes the information structure. Productive effort e positively affects all three signals, whereas greenwashing d affects only the manipulable signal s_2. Each signal is contaminated by independent Gaussian noise. The principal observes (s_1, s_2, s_3) and compensates the manager based on this noisy information.

Equilibrium Incentives and Greenwashing. Given contract (α, β), the manager's incentive compatibility conditions yield optimal actions:

$$e^*(\beta) = \frac{\beta_1 + \beta_2 + \beta_3}{k_e}, \qquad d^*(\beta) = \frac{\beta_2}{k_d}. \tag{19}$$

The principal chooses β to maximize expected surplus subject to providing incentives and compensating for risk. This yields optimal weights (derivations in Appendix A):

$$\beta_2^* = \frac{v}{a}\frac{p_2 k_d}{\Delta}, \qquad \beta_1^*, \beta_3^* = \frac{v}{a}\frac{p_{1,3}\left(k_d + \frac{p_2}{a}\right)}{\Delta}, \tag{20}$$

where $\Delta = \left(k_e + \frac{S}{a}\right)\left(k_d + \frac{p_2}{a}\right) - \left(\frac{p_2}{a}\right)^2$, $S = p_1 + p_2 + p_3$ is total signal precision, and $p_{1,3} \in \{p_1, p_3\}$.

This characterization leads to our first main result:

Proposition 1 (Manipulability Distorts Incentive Provision). *When greenwashing becomes cheaper (k_d decreases), the principal optimally reduces weight on the manipulable signal (β_2^* decreases) and increases weight on unbiased signals (β_1^*, β_3^* increase). For all finite k_d, equilibrium effort satisfies $e^* < e^*_{no\text{-}gw}$, where $e^*_{no\text{-}gw} = \frac{v}{k_e} \cdot \frac{S}{ak_e + S}$ is the effort level when manipulation is impossible.*

Proof sketch. Differentiating β_2^* with respect to k_d shows that cheaper manipulation

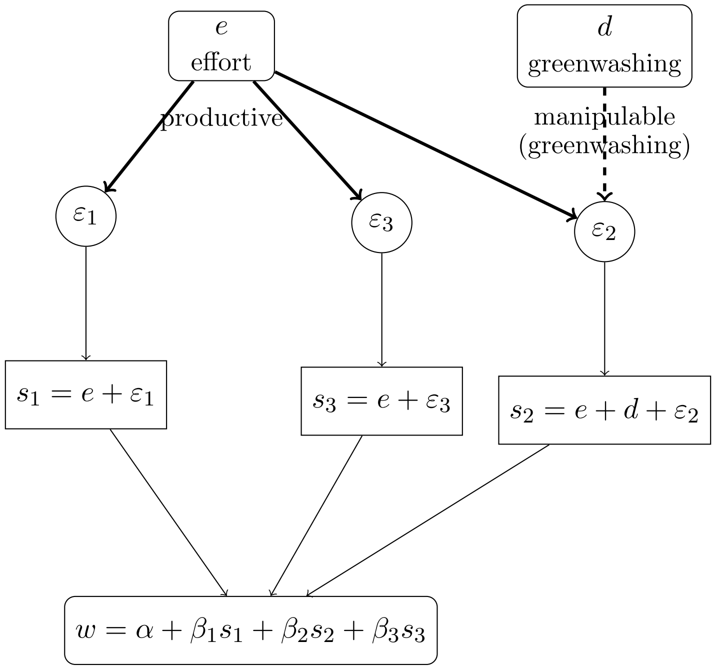

Figure 1: Information structure in the performance measurement model. Productive effort e affects all three signals, contributing to both firm value and emission reductions. Greenwashing d affects only the manipulable signal s_2, which therefore mixes real performance with strategic disclosure. Independent Gaussian shocks $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ contaminate the signals. The principal conditions compensation on (s_1, s_2, s_3) through a linear contract, creating incentives for both productive effort and greenwashing.

reduces the principal's reliance on s_2. However, because s_2 contains information about effort, reducing β_2 weakens overall incentives for e. The principal cannot perfectly substitute to s_1 and s_3 because increasing β_1 and β_3 increases the manager's compensation risk. The result is depressed effort relative to the benchmark without greenwashing. Full derivation in Appendix A. □

Proposition 1 formalizes the intuition from Section 2.2 that manipulable ratings reduce impact. The principal faces a fundamental trade-off: using s_2 provides information about effort but also creates greenwashing incentives. This connects directly to our empirical finding that backward-looking ESG ratings—which are subject to voluntary disclosure and strategic framing—fail to predict emission reductions.

Type Uncertainty and the Negative Spiral. We now extend the model to allow for manager-specific greenwashing costs, capturing heterogeneity in firms' ability to engage in cheap talk. Let the manager's type be $\theta \in \{L, H\}$ with greenwashing costs:

$$k_d(L) = k_d^L, \qquad k_d(H) = k_d^H, \qquad 0 < k_d^L < k_d^H, \tag{21}$$

where type L finds greenwashing less costly (e.g., firms with existing PR infrastructure, less scrutiny). The principal knows the distribution $\Pr(\theta = L) = \pi \in (0, 1)$ but not the realized type, and chooses a single contract ex ante.[7]

Given contract (α, β), each type chooses effort $e^*(\theta) = (\beta_1 + \beta_2 + \beta_3)/k_e$ (which is type-independent) but differs in greenwashing: $d^*(\theta) = \beta_2/k_d(\theta)$. Define true impact (emission reduction) as:

$$I(\theta) = e^*(\theta) - \gamma d^*(\theta), \tag{22}$$

where $\gamma > 0$ captures the fact that greenwashing creates misleading signals without real environmental improvement.

This setup generates a surprising result:

Proposition 2 (Negative Spiral from High Impact Valuation). *There exists a non-empty set of parameters $(a, k_e, k_d^L, k_d^H, \sigma_1^2, \sigma_2^2, \sigma_3^2, \pi, \gamma)$ such that the following hold for v in some interval $[v_1, v_2]$:*

(i) The optimal weight on the manipulable signal $\beta_2^*(v)$ is increasing in v.

(ii) The low-cost type responds with disproportionately more greenwashing: $\frac{\partial d^*(L;v)}{\partial v} > \frac{\partial d^*(H;v)}{\partial v}$.

(iii) Type uncertainty increases the effective noise in measured impact: $\frac{d}{dv} Var(I(\theta; v)) > 0$.

(iv) Expected true impact is decreasing *in the principal's valuation of impact: $\frac{d}{dv}\mathbb{E}[I(v)] < 0$.*

Thus, valuing impact more (higher v) induces greater reliance on the manipulable signal, which amplifies greenwashing by low-cost types, increases type-induced dispersion, and lowers expected impact—a "negative spiral."

Proof sketch. Higher v makes incentives more valuable, so the principal relies more on all performance measures including s_2. Because s_2 is manipulable and types differ in k_d, the incentive effect of increasing β_2 is stronger for low-cost types, who increase greenwash-

[7] In Appendix A.6, we show that our results extend to continuous type distributions. With log-normally distributed greenwashing costs, the negative spiral emerges when type heterogeneity is sufficiently large.

ing d sharply. The principal cannot infer the manager's type from the signals ex ante, so uncertainty about the mixture of (e, d) grows with v. This type-induced dispersion makes the mapping from signals to true impact noisier, reducing the informativeness of the performance measurement system. As the precision of impact measurement falls, the effective incentive power weakens, and productive effort e falls—generating a region where increasing the principal's valuation of impact v actually *reduces* expected true impact $\mathbb{E}[I]$. The formal proof, which establishes parameter conditions and characterizes the interval $[v_1, v_2]$, is in Appendix A. $\square$

The negative spiral result has important implications for ESG investing. It suggests that when investors increase their focus on sustainability (higher v), but rely on manipulable backward-looking ratings, they may inadvertently create perverse incentives. Firms with low greenwashing costs exploit the increased weight on manipulable signals, contaminating the information environment and reducing aggregate impact. This provides a micro-foundation for why we observe in our empirical analysis that higher ESG scores can predict *higher*, not lower, future emissions.

Connection to the Firm Optimization Model. The principal-agent model complements the firm optimization framework in Section 2.2. In that model, firms choose ESG scores trading off adjustment costs against discount rate benefits. Here, we show how investors' performance measurement problem creates the incentives that firms respond to. Specifically, the greenwashing parameter d in this section provides micro-foundations for the cheap talk parameter $T_{i,t}$ in the production function for ESG scores. When investors cannot distinguish true effort e from greenwashing d, firms optimally allocate resources to both—consistent with our finding below that cheap talk reduces impact by a factor of $\theta^{-\eta_i}$.

More fundamentally, both models point to the same solution: forward-looking, market-based measures. In the firm optimization model, forward-looking measures improve impact by reducing noise and eliminating cheap talk opportunities. In the principal-agent model, they work by providing signals that are harder to manipulate (higher k_d) and by revealing type information through market prices, mitigating the negative spiral. Section 2.5 synthesizes these insights.

2.5 Benefits of a Forward-looking Measure: Theory

Based on the model presented above, the benefits of a forward-looking market-based measure relative to a backward-looking subjective rating can be summarised as follows:

i. The lower noise in the market-based measure M relative to the ratings-based measure R increases the marginal benefit in Equation 13 and hence decreases the negative externality in equilibrium. Additionally, it leads investors to better allocate their capital to higher impact firms.

ii. Under the ratings-based measure R, cheap talk increases the negative externality level in equilibrium and worsens investors' ability to allocate their capital to higher impact firms.

iii. A continuous measure relative to a threshold-based rating leads investors to better allocate their capital to higher impact firms.

iv. Sustainability-linked pay increases the marginal benefit in Equation 13 and hence decreases the negative externality in equilibrium, and more so using measure M.

Now we will explain each of these in more detail:

i. Noisy Ratings Decrease Impact and Increase Misallocation. Let the realised E_i^R score be the target score E_i, plus a realisation of a log-normally distributed noise term σ^R. The marginal cost is the same as it is based on the targeted score, however the marginal benefit will change to

$$MB_i^M \exp\left(-\frac{1}{2}\sigma^{R^2}\right),$$

where MB_i^M is the marginal benefit without noisy ESG ratings.[8]

The new equilibrium score will then be

$$E_i^R = E_i^M \exp\left(-\frac{1}{2}\sigma^{R^2}\right)^{\eta_i^{-1}-s}.$$

Hence, the noisy scores reduce aggregate impact (across all horizons) by

$$\frac{I_{i,t}^{M,R}}{-D_{i,t}^M} = \exp\left(-\frac{1}{2}\sigma^{R^2}\right)^{\frac{1}{\eta_i}-s}.$$

[8] Derivation in Appendix C.

Additionally, the noise and bias leads to an attenuation effect, which increases the misallocation of capital by

$$\tilde{I}_{t,n}^{M,R} = |\epsilon^R| + |\epsilon_{bias}^R| - |\epsilon^M|,$$

where ϵ_{bias}^R is given by

$$\epsilon_{bias} = \left(1 - \frac{var[M]}{var[M + \epsilon_M]}\right) \beta D_{i,t,0}^{\mathcal{M}},$$

where $var[M]$ in turn is the variance in the population of measure M, ϵ_M is the measurement noise of measure M, and β is the true beta between the R measure and realised emissions.

The total effect on welfare will then be the effect on the sum of the two types of dividends cash dividends and externality dividends:

$$W_{i,t}^R = -D_{i,t}^{e,M} \exp\left(-\frac{1}{2}\sigma^{R^2}\right)^{\frac{1}{\eta_i}-s} + D_{it}^{c,M} \exp(\exp(\frac{1}{2}\sigma_R^2))^{\frac{1}{\eta}-s})^{\frac{1}{\eta}}).$$

$$W_{i,t}^{M,R} = \Delta - D_{i,t}^M \exp\left(-\frac{1}{2}\sigma^{R^2}\right)^{\frac{1}{\eta_i}-s} + cashflowchange.$$

ii. Cheap Talk Decreases Impact and Increases Misallocation With cheap talk a firm i can increase their score by using cheap talk $T_{i,t}$ in addition to spending money $B_{i,t}$ on decreasing the externality as before. Specifically, let θ represent the importance of the impact budget relative to cheap talk in terms of the achieved ESG rating and k_i and η_i remain as previously defined, then the production function $F_i(.)$ becomes

$$E_{i,t} = \frac{1}{k_i}(B_{i,t}^{\theta} T_{i,t}^{1-\theta})^{\eta_i}.$$

Given an optimal budget mix of cheap talk and real impact by the firm, the cash dividends become

$$D_{i,t,0}^c(E_{i,t}) = D_{i,t,0}^{c,0} \exp\left(-k_i' E_{i,t}^{1/\eta_i}\right),$$

with a k_i' which is

$$k_i \left(\frac{1}{\theta}\right)^{\theta} \left(\frac{p_T}{1-\theta}\right)^{1-\theta},$$

where p_T is the cost of cheap talk relative to impact. Hence, the new equilibrium rating will be

$$E_i^R = E_i^M \frac{k_i'}{k_i}^{\frac{-1}{\frac{1}{\eta_i}-s}}.$$

Additionally, the share spent on impact is reduced by a factor of θ^{-1} as the rest of the expenditure is spend on cheap talk. The effectiveness of the rating is therefore decreased by

$$\theta^{-\eta_i}.$$

Hence, going to a forward-based measure could improve impact by the impact lost from cheap talk, which is

$$\frac{I_{i,t,0}^{M,R}}{-D_{i,t}^M} = 1 - \theta^{-\eta_i} \frac{k_i'}{k_i}^{\frac{-1}{\frac{1}{\eta_i}-s}}.$$

Misallocation is in turn increased from cheap talk by

$$\frac{\tilde{I}_{t,n}^{M,R}}{D_{i,t}^M} = 1 - \theta^{-\eta_i}.$$

Derivations are in Appendix C.

iii. Threshold-based Investment Increases Misallocation. With $\underline{E}$ as the threshold score for investors to invest into firm i, the discount rate on firm i becomes:

$$r^c = \beta_M r_M - \mathbb{1}_{E>\underline{E}}.$$

The marginal benefit becomes a Dirac delta function around $E = \underline{E}$ with magnitude

$$sE_i^{-1}\,\mathbb{E}_t\left[\frac{D_{i,t,1}^{c,0}\exp\left(-k_i E_i^{1/\eta_i}\right)}{\exp\left(\beta_i^M r_{t+1}^M - s\right)}\right].$$

As the marginal benefit is zero everywhere else, the equilibrium ESG level has to be either $\underline{E}$, or 0, otherwise the cost can be reduced by lowering E towards zero without giving up any benefit.

Next, let the realised ESG rating E^R be equal to E plus some noise ϵ^R. Those firms with a large positive realised error term will now have a rating markedly above the threshold, however when taking the expectation of the next period that expectation will be the threshold. For these firms, the expectation of the next period's ESG score will be lower than their current score. At the same time those firms with a negative realised error will now have a low rating, but in expectation have a higher rating next period. All in all, this implies that those firms which have "overshot" in terms of impact and rating will actually in expectation reduce their impact next period, instead of increasing it. As a result, thresholds-based ratings will lead to investors misallocating capital to firms which go on to pollute more, not less. This misallocation is given by

$$\tilde{I}^{M,R}_{t,n} = |\epsilon^R|.$$

iv. Sustainability-based Pay Increases Impact. To see this consider a manager m of firm i who receives a fixed salary S plus compensation compared to the final dividends of the firm by a fraction k^I and may also be compensated with a bonus based on how sustainable the firm is by a fraction k^E. The bonus is given as a κ ratio increase in his salary from cash dividends. The value of this salary corresponds to relating the first compensation onto the stock price $P_{i,t}$ of the firm i at time t and the second onto the external value of the firm. Hence, his marginal benefit of decreasing the external dividend increases by κ, which will lead to a higher equilibrium impact. Specifically, for the extreme case of no green sentiment the impact increases by

$$I^{M,R}_{i,t,0} = \left(\frac{\kappa}{k_i}\right)^{\frac{\eta_i - 1}{\eta_i}}.$$

Derivation in Appendix C.

3 Introducing Emission Futures

To improve the measurement of the external value of the firm ($V^E_{i,t}$ in Equation 1) we argue for the introduction of a new financial asset, what we term an "emission futures" contract. We envision an emission future as a standardized firm-specific or index-level contract. Much like a dividend futures contract, at maturity, the buyer pays the futures price, which is determined today, and the seller pays the dollar value of emissions of the underlying firm(s), indexed by i. This emission dollar value to be paid for firm i at time t, denoted $D^e_{i,t}$, is computed as the product of the quantity of CO2 emissions $e_{i,t}$

during a certain calendar year and the daily average carbon price during that calendar year, which is the same across all firms:

$$D_{i,t}^{e} = e_{i,t} P_t^{e}. \tag{23}$$

Take for example the 2026 Shell emission futures contract. On the third Friday of December 2026, the buyer of the futures contract will pay the futures price, and the seller will pay the emission value $D_{i,t}^{e}$ which is computed as the product of (1) Shell's scope 1 (and potentially scope 2) CO2 emissions between the third Friday in December of 2025 and the third Friday of December of 2026 and (2) the average daily ETS carbon price in 2026. The contract is settled based on the sum of all emissions throughout the year, and there is no compounding or discounting within the year provided for in the contract.[9]

Let $g_{i,t,n}$ denote the average per-period expected growth rate of firm i's emissions value over the next n periods:

$$g_{i,t,n} = \frac{1}{n} \mathbb{E}_t \left[\ln \left(\frac{D_{i,t+n}^{e}}{D_{i,t}^{e}} \right) \right]. \tag{24}$$

Then the present value $\mathcal{P}_{i,t,n}$ of $D_{i,t+n}^{e}$ is given by:

$$\mathcal{P}_{i,t,n} = D_{i,t}^{e} \exp\left(n(g_{i,t,n} - \mu_{i,t,n}) \right), \tag{25}$$

which defines the (geometric) discount rate $\mu_{i,t,n}$ for firm i's emissions. By splitting the discount rate into the nominal bond yield for period n, denoted by $y_{t,n}^{b}$, and a firm-specific, horizon-specific emissions risk premium $\theta_{i,t,n}$ that compensates investors for the emission risk for maturity n of firm i, we can rewrite equation (25) as:

$$\mathcal{P}_{i,t,n} = D_{i,t}^{e} \exp\left(n(g_{i,t,n} - y_{t,n}^{b} - \theta_{i,t,n}) \right). \tag{26}$$

[9]In practice, futures contracts require a sufficient volume to get listed by an exchange, which is why we propose to start with a contract on the SP500 as well as the largest polluters for a few horizons, such as 1, 2, and 5 years. In comparison the Euronext exchange currently offers dividend future contracts on 352 individual firms for maturities of 3, 6, and 12 months. As emissions are concentrated around a few firms, offering contracts on just 58 (10) firms would cover 90% (55%) of total emissions. Once sufficient volume has been established, more firms could then be added.

The emissions yield for firm i at time t with maturity n is then defined as:

$$y_{i,t,n} \equiv \frac{1}{n}\ln\left(\frac{D^e_{i,t}}{\mathcal{P}_{i,t,n}}\right) = y^b_{t,n} + \theta_{i,t,n} - g_{i,t,n}.$$

The expression above shows that the emissions yield consists of three components. It consists of the nominal bond yield $y^b_{t,n}$, a maturity-specific and firm specific risk premium $\theta_{i,t,n}$ that investors require for being exposed to emissions risk, and the expected growth rate of the emissions value $g_{i,t,n}$, which represents the average expected log growth over the next n periods. Ceteris paribus, a higher expected growth rate makes the price $\mathcal{P}_{i,t,n}$ higher compared to the current emissions value $D^e_{i,t}$. This results in a lower emissions yield.

While in principle, emission value contracts could be traded in the spot market, we propose to have them traded in futures markets, much like other traded commodities. Under no arbitrage, the spot price and the forward price ($\mathcal{F}_{i,t,n}$) are linked through the nominal bond yield:

$$\mathcal{F}_{i,t,n} = \mathcal{P}_{i,t,n}\exp(ny^b_{t,n}). \tag{27}$$

We then define the forward emission yield $y^f_{i,t,n}$ as:

$$y^f_{i,t,n} \equiv \frac{1}{n}\ln\left(\frac{D^e_{i,t}}{\mathcal{F}_{i,t,n}}\right) = \theta_{i,t,n} - g_{i,t,n}. \tag{28}$$

The forward emission yield is equal to the difference between the risk premium and the expected emission value growth rate. If the forward emission yield is high, this either implies that risk premia are high or that expected emission value growth rates are low. Lastly, we define the one-period dollar change on the forward as:

$$R^{\$}_{i,t,n+1} \equiv \mathcal{F}_{i,t,n} - \mathcal{F}_{i,t-1,n+1}. \tag{29}$$

As we will discuss further below, so far, the variation in the emissions values has had a relatively low correlation with other financial market returns, implying that the risk premium on these assets is likely going to be low with little variation. As such, the forward emission yield will be a forward-looking, market-based measure of expected environmental impact (See Figure 7). The higher the yield, the more the market expects the firm to cut its emission values. As such higher yields (or improvements thereof) can be directly translated into higher environmental impact ratings.

It is important to emphasize that emission futures are firm, time and horizon specific, which are all important ingredients for effectively measuring environmental progress. In particular, the horizon dimension allows investors to take a stance on the particular horizon over which they think firms will be able to cut their emission values. These market-based horizon-specific expectations can then be compared with the promises made by the firm's management.

The price of emision futures is not only determined by the expected growth path of the emissions value ($g_{i,t,n}$) but also by the risk premium. It is worth discussing what the likely properties of this risk premium will be. First, the CAPM beta of emissions is insignifiant for most firms, and seems to be negative on average. This suggests that on average, we should expect the risk premium to be small and, if anything, negative. What about the cross-sectional variation? Standard asset pricing theory suggests that firms that cut their emissions in bad (good) times, which means that the futures contract has a negative (positive) return, will have a high (low) risk premium ($\theta_{i,t,n}$) and thus a higher (lower) emissions yield, as this yield is the difference between the risk premium and the expected path growth of emissions ($g_{i,t,n}$).

We should wonder whether the risk premium properties described above are desirable in the context of an emissions rating, which is equal to the emissions yield. First, if the only reason firms cut emissions in bad times is because they produce less in bad times, then assigning firms with a higher score for that reason does not seem all that desirable. On the other hand, if it is easier for firms to invest and apply technology that lowers emissions in good times than in bad times, then the firms that are able to cut emissions in bad times indeed deserve a higher rating. Note further that alternatively we could base the scoring on the changes in the emissions yield. In that case the level of the risk premium is differenced out and it is the properties of the *changes* in the risk premium that we should then better understand.

4 New Measures

The assets that we have just introduced can be used to measure a firm's future sustainability plans and outcomes relative to the markets expectation. Specifically, we propose the following novel environmental and impact measures. These measures are firm and horizon specific. The first one is the E measure given from the futures price of the emission dividend strip $\mathcal{F}_{i,t,n}$ as

$$\mathcal{E}_{i,t,n} \equiv -\mathcal{F}_{i,t,n}. \tag{30}$$

As it is based on the negative of the futures contract price it accurately reflects the market's (risk-neutral) belief of the future value of the emissions of firm i at a time n periods in the future. Under the assumption that the risk-premium is small, this will be equal to the objective belief.

Where the Environmental Rating accurately displays the firms with the lowest future carbon externality.[10] Another way to invest is to actively work to decrease firms expected carbon emissions, so called Impact Investing. Where as investing in low-carbon firms relies on the price-channel to incentivize green firms to grow and brown to shrink, our empirical work, as well as other studies, suggests this effect either does not work, or is an inefficient method. Hence, investing in a way that efficiently creates impact is preferable, so called "impact investing".

We propose a natural measure for impact investing, which reflects how much the firm has decreased its emissions at a given horizon, compared to what was the market previously expected for that firm. Specifically, for the one year horizon, this will be if the firm reduced its emissions relative to the market expectations. For longer horizons, the impact measure will reflect the extend to which firms can plausibly commit to lowering emissions in the future. That is if the market price at that horizon drops this is an indication that the investors do not view the firm's commitments as cheap talk. Green impact is simply defined using the dollar return on a n period emission future at date t on firm i, $R^{\$}_{i,t,n}$, as

$$\mathcal{I}_{i,t,n} \equiv -R^{\$}_{i,t,n}. \tag{31}$$

The equation above shows that green impact reflects how the (risk-neutral) expectation of firm i's emissions at horizon n has changed relative to the price was last year.

Our new firm measures can also be used to create new fund measures for a fund j who owns stocks i with ownership shares $w_{j,i,t}$. The fund's E and green impact measures

[10] Alternative scores are given in Appendix B.

are given as

$$\begin{aligned}\mathcal{E}_{j,t,n} &= \sum_i w_{j,i,t}\mathcal{E}_{i,t},\\ \mathcal{I}_{j,t,n} &= \sum_i w_{j,i,t-1}\mathcal{I}_{i,t,n}.\end{aligned} \tag{32}$$

Hence, an improvement in the fund's E measure can be achieved in one of two ways. First, by changes in the underlying firms' E measure (the futures prices), and secondly, by decreasing the ownerships shares in the polluting firms. However, a fund's I measure will be given by the dollar return on the underlying firms futures prices weighted by the funds ownership share in the previous period, so this return is entirely driven by the futures prices as the weights remain fixed.

Like with the firm level ratings, the fund level E measure will be fine for passive funds and by following the E measure you would be accurately investing into the firms that would have the lowest emissions going forward in expectation, however mutual funds can outperform on this measure if they are better at predicing which firms will reduce their emissions by more than the markets expectations implied. This outperformance is captured by the fund level green impact measure. If investors choose to base the flow-performance relationship on this outperformance measure more investable capital will be allocated to greener stocks, also going forward.

5 Demand for Emission Futures

For Emission Futures to become a traded asset, there needs to be demand for it. We see there being a demand for Emission Futures due at least three reasons. First source is that Emission Futures are demanded as a tool to hedge transition risk. Similarly to how Credit Default Swaps are demanded to hedge credit risk of the counterparty, Emission Futures could be used by investors to precisely hedge the transition risk of a firm. The reason being that in the case of an increase in the transition probability and/or severity of a transition, the shadow price of carbon will rise leading to an increased tax burden and loss of firm value. In this case of an increased transition probability, the increased value of the Emission Future would be a perfect hedge, as it exactly offsets the tax loss due to it paying out the expected value of the firm's emissions times the carbon price. An additional advantage of Emission Futures is that the futures market for carbon credits are more liquid than the spot market (Känzig (2023)).

Second, the contracts could be used for speculation. If you as an investor believe the emissions of a company are going to deviate from those implied by the price, a profit could be made by buying or selling this contract depending on a positive or negative deviation. As a side, this is the mechanism that ensures the reliability of the predictions.

Third, Emission Futures could improve risk-allocation in the economy. This is because these assets are not redundant assets, as even though one could hedge changes in the shadow price of carbon using carbon futures, one would still be exposed to changes in the expected emission amount of firms. Additionally, one would need to continuously trade carbon futures to continuously be hedged against carbon price changes due to the exposure changing over time. Hence, Emission Futures could improve welfare through improved diversification.

Dividend futures, a futures market which also allows for the trading of firm and horizon specific contracts have grown to be a large market. Euronext, a large dividend futures trading platform, currently offer contracts on 352 individual stocks and for maturities of 3, 6 and 12 months. As only 58 firms are needed to cover 90% of the US emissions, we expect the demand would lead to most of the US emissions being covered. Hence, due to these three sources of demand for Emission Futures, plus the evidence of the popularity of the credit default swap and dividend futures markets makes it likely that Emission Futures could become a traded market covering a significant portion of US emissions.

6 Empirical Analysis

In this section we test whether current ratings have been successful in reducing carbon emissions. First we test whether backward-looking ratings are effective in predicting future emission reductions. A hypothesis that we reject. In fact, higher scores are more likely to predict increases rather than decreases in emissions. Second, if the current backward-looking ratings are useful we should see that social capital has decreased emissions. Specifically, we test whether the recent increase in social capital has decreased emissions relative to previous periods with less social capital and we test whether increases in social capital in the cross-section are associated with higher emission decreases. What we see is that social capital has not been a driving force in reducing emissions neither in the time series nor the cross-section.

6.1 Data

To construct our dataset of firms' impact we use the Compustat database merged with carbon emission data from Refinitiv. For the cheap talk analysis we use ESG scores from Sustainalytics and we count word usage from firms' SEC filings using WRDS.

The amount of assets invested using ESG principles has been growing rapidly over the last decades as evidenced by Figure 2. Figure 2 shows that socially invested capital has increased ten-fold from 2007 to 2020. Assets invested under ESG principles now exceeds USD 120 trillion. At the same time, we have seen a decrease in the average emissions per firm (See Figure 2).[11] These observations show that firms have reduced their emissions. However, this may be arising from a secular trend, indeed Figure D.1 in Appendix D shows that emissions relative to GDP has been decreasing since the 1920's. Hence, to understand social capital's role in impacting the emissions of firms we dive deeper in the next sections.

[Figure 2 about here.]

Another way to visualize the average decrease in firms' carbon emissions is from the pricing of Emission Futures at different horizons. To show this, we start by computing the average emissions experienced by the firms in our sample. Next, we get the expected future emissions by regressing a firm's future emissions on it's current emissions at horizons of one to five years. This gives us the value of Emission Futures contracts, which have been normalized by a Carbon Futures contracts at the same horizon, and this is what we plot as the top line in gray in Figure 3. To illustrate how forward-looking contracts makes impact measurement possible, suppose a new policy aimed to reduce carbon emissions is enacted by the firm's management. The valuation of the firms' emission futures instantly respond to the news of the policy and the updated prices are shown as the black line in the bottom of Figure 3. To get a measure of impact, this new valuation can then be compared to the counter-factual, which is how the expected emissions were previous to the policy change as measured from the Emission Future

[11] This is also the case if we instead consider the asset weighted total emissions, as depicted in the right side of Figure D.2 in Appendix D. We also see the same development if we consider emission intensities, meaning emissions over revenue, as shown in Figure D.3 in Appendix D.

prices. The sum of the differences between these two valuations then adds up the impact attributable to the policy, shown by the green area in Figure 3.

[Figure 3 about here.]

6.2 Results

Effect of Higher ESG Scores. This subsubsection goes on to evaluate whether higher scoring companies are driving environmental impact. As social capital attempts to achieve impact by investing into high ESG firms this is a requirement for real impact.

To evaluate the scores' impact we conduct a Granger causality test of whether higher E scores lead to lower emissions in the future, in excess of what the emissions today would predict. As seen in Table 1 Columns (1) and (2) E scores do not predict impact as measured by emissions. Columns (3) and (4) in fact shows the reverse to be the case. Namely, that if you reduce your emissions you increase your future E score. This also suggests the ESG scores to be backward looking, rather than forward looking. Panel B confirms the results using E scores from Refinitiv instead of Sustainalytics.

Additionally, we can get an estimate of the impact potential from going to a forward-looking score by considering how much predictive power the current scores give. To get such an estimate we can subtract the R^2 of Table 1 Column (2) from Column (1), which is indistinguishable from zero, hence giving the highest impact potential possibly for a forward-looking score.

[Table 1 about here.]

Social Capital and Impact in the Time Series. To test whether social capital has lead to real impact, we estimate a Granger causality regression of CO_2 emissions on past emissions and the amount of socially invested capital. As emission reductions may take a while to materialize, we do this at horizons of one to five years by sequentially adding an additional lagged variable. We evaluate the likelihood of social capital having lead to real impact using the tests' t-statistic. For this to be likely we need the statistic to be negative and the magnitude large. Generally a value higher than 1.96 is seen as the lowest bar to conclude that there may be an effect.The maximum of these t-values is depicted in Figure 4 for each horizon. We see that the effect is never significant for any horizon, even becoming positive at higher horizons. The regressions underlying this figure are shown in Table 2.[12]

[Figure 4 about here.]

[Table 2 about here.]

Social Capital and Impact in the Cross-Section: Effect of Being in a Social Index. In this subsubsection we evaluate social capital's impact using a second approach, namely using inclusions to a social index. We do so by considering what happens to firms' emissions after they are included in the social index, driving an inflow of social capital to the firm. Figure 5 shows the evolution in the emissions of firms either in the social index or outside it. While for firms in the index the average emissions have remained stable at around 2.5 million tons CO_2 equivalents, for firms outside of the index this has dropped by 10 million tons from around 12 to 2 million tons CO_2 equivalents. This shows that the reduction in emissions have generally come from firms not in the social index.

To understand how firm emissions adjust around index inclusions we in Figure 6 show emissions for firms from five years before and after inclusion relative to the firms lifetime emissions controlling for average emission changes year-to-year. What we can see is that firms who enter the social index lower their emissions at around three years before inclusion to below their lifetime average. After inclusion, the firms' emissions increase rapidly and two years after inclusion they are back at to their baseline. At 3 years and above they emit more than they usually do, an increase that continues to rise at time goes on.[13]

An alternative explanation is that the decreases in emissions have been driven by the benefits from inclusion. To test this we compare the time series emission reductions of firms who have the possibility to enter a social index versus firms who can never join the index due to being in an industry that is excluded from the index, so called "sin" firms. Correspondingly, we label ordinary firms "saint" firms. These results are shown in Figure D.6 in the Appendix. Here we see in Panel A that the CO2 emissions have been lowered for both saint and sin firms. Panel B shows the year-on-year change in

[12] An intuitive confirmation of this finding can be seen by simply plotting emission growth by ESG asset growth, either lagged or concurrent. Figure D.4 in the Appendix shows there exists an insiginificant relationship between these two variables, if anything there is a concurrent weakly significant *positive* relationship.

[13] Figure D.5 in the Appendix shows this finding in terms of changes in the emissions instead of absolute emissions.

percent rather than total emissions and the evolution is similar across the two types of firms. Panel C cumulates the CO2 changes and we see that the cumulated changes for saint and sin firms are not significantly different at any point in our sample.

As depicted in Table 3 we see in Column (1) that firms in the index generally have lower emissions as expected. Column (2) shows this is also the case in relative terms. However we see in Column (3) that after firms are included, they increase their emissions. Column (4) shows that this is also the case in relative terms. Column (5) and (6) repeat the same result but with the baseline being not in the index compared to being in the index.

Table D.1 in Appendix D shows that the effects are the same across specifications. Specifically, Column (1) and (2) repeat results from the previous table. Column (3) shows results including firm and time fixed effects- Column (4) includes a variable for the year of inclusion showing a similar sized effect for the inclusion year, but higher when having year fixed effects as in Column (5). For both firm and time fixed effects (Column 6) we lose power for the inclusions and it is no longer significant. Table D.2 in Appendix D shows the effect to be the same for the first, second, and third year ahead. Simply accumulating the increase in emissions. Even columns include an inclusion dummy and Columns (1-6) are relative changes where as Columns (7-12) are absolute changes. The change is stable at 2% increase per year.

These results suggest that social capital has not been driving impact. In fact the arrival of social capital seems to have lowered impact.

[Figure 5 about here.]

[Figure 6 about here.]

[Table 3 about here.]

Emission Future Betas. We plot the emission beta estimates for each firm in our sample in Figure 7. In the first columns (Panels A and C) we have results for the beta with respect to purely emissions and in the second columns (Panels B and D) we have results for the beta with respect to emission futures, which are the emissions times the price of carbon emissions. In the top row the betas are plotted and in the bottom row the t-stats of the betas are plotted. The beta means how much the emission changes or changes in the value of a firm's emission future co-varies with the market return. Panel A shows that the betas are closely clustered around zero and while some t-stats are significant this is what we would expect from the type-1 error rate corresponding to the level of significance. Panel B shows that the betas related to the emission futures are more widely distributed, however their t-stats are equally insignificant. Overall, there is no evidence that the betas of emission futures should be significantly positive or negative.

[Figure 7 about here.]

Cheap Talk. As we have seen that social capital does not seem to be impactful we explore whether so-called 'Cheap Talk' may be driving ESG scores. This would lead to social capital being less effective in driving impact. Panel A in Figure 8 shows how cheap talk has increased over time. We see that that the use of the word 'Sustainability' has increased exponentially over the last two decades starting close to zero and ending at over two thousand mentions per firm per year in their official investor communications. Panel B shows the cross-sectional distribution and we see that there is quite a wide range of yearly mentions per firm. As this has been increasing over time we also display the cross-sectional distribution subtracting the time-series effects in Panel C and see that there tends to be two groups of firms: some that mention Sustainability about a thousand times more than the average and another group that mention it a thousand times less.[14]

We conduct a Granger causality test of cheap talk on ESG scores in Table 4. Here we see that the word frequency of Sustainability increases the future ESG scores in excess of what the current ESG score as well as impact would predict.[15] Table 4 also shows the equivalent Granger causality tests on cheap talk and impact in Panel B and C respectively. We see in Panel B that cheap talk does not lead to impact and in Panel C that getting a higher ESG score makes you do more cheap talk.

From Table D.4 in Appendix D we can see that the elasticities of cheap talk to impact is 0.73 meaning you have to mention 'Sustainability' 0.73% more to pollute 1% more and achieve the same E score. This decreases to just 0.13% for ESG scores. In absolute values this is 17 words per 10 million tons CO_2 emitted.

[Figure 8 about here.]

[Table 4 about here.]

[14]Figure D.7 in Appendix D shows the same for the word 'ESG' for which we see the same results.

[15]Table D.3 in Appendix D shows the same for several words and show that several have affect but 'Sustainability' is the most effective.

7 Estimating welfare gains from the introduction of Emission Futures

Welfare in our setting arises from impact either through lowering negative external dividends or increasing internal cash dividends. Hence, we are interested in the net effect. The beauty of our proposition is that the introduction of Emission Futures, which allows for the accurate measurement of external dividends, improves the allocation of financial capital, even with no further changes, ultimately leading to higher overall welfare. The increased welfare arising from the introduction of Emission Futures is due to the four channels developed in Section 2.5.

From Section 2.5 on the benefits of a forward-looking measure, we see that impact could be improved, and misallocation decreased, through all four channels (firm ESG decisions, cheap talk, threshold-based investing, and managerial compensation), by changing from the current ESG ratings to a more precise and unbiased ESG measure from Emission Futures.[16] In this section we therefore show the improvement potential from this channel.

We consider the state of the current ESG ratings by measuring the mean residual standard error arising from Emissions forecasted by E ratings. This tells us the welfare gain potential that is there for this type of ESG ratings. We then see how welfare could be improved if more data was used to increase the accuracy of predicting emissions.

Table 5 shows our results. In Column (1) we see that the mean residual standard error from firms' emissions as predicted by their Refinitiv E score is 16.5 Million tons of CO2 equivalents, meaning that if you invested following the Refinitiv E score the emissions of a firm would for a third of your observations be more than 16.5 Million tons CO2e higher or lower than you expected. This is more than four times the average firm emissions of around 4 Million tons CO2e. In Column (2) we see that we can improve our sustainable capital allocation significantly by using a firm's current emissions as our forecast instead of their E score. Using their current emissions we get a forecast error of 2.3 Million tons, a relative reduction of 86%, putting us at around half on a firm's average emissions. From Column (3) we see that the reintroduction of a firm's

[16] A more precise ESG measure decreases cheap talk as now it becomes easier to detect. See the Credability of Cheap Talk extension in Appendix C. A more precise ESG measure also increases the impact by the manager as with a more precise measure the external value is increased as mentioned in the first channel, increasing his incentives to foster impact.

E score does not improve our prediction compared to using a firm's current emissions. Columns (4-6) confirm these findings for the Sustainalytics E score. Here, the starting error is slightly smaller at 15.0 Million Tons CO2e, but is also improved down to 4.3 Million Tons CO2e from using a firm's emissions instead of the E score.

[Table 5 about here.]

For both E ratings, we see a significant welfare potential from the introduction of Emission Futures as the prediction errors could be reduced by at least 71%.

8 Conclusion

This paper proposes a new measure for the real impact of ESG investing. This increasingly popular way of investing has been hindered by rating systems that are inherently subjective, backward-looking, and hence inconsistent across rating issuers. In this paper, we have proposed a novel measure based on a novel financial instrument: firm and horizon-specific Emission Futures. These futures contracts are market-based and forward-looking. We argue that basing decisions on this measure increases the effectiveness of ESG investing relative to the status quo.

Our measure can also be used for social index construction as an index's impact is simply a weighted average of the forward-looking measure we propose. Another use case is that the new measure can be used to accurately evaluate previous classifications into green and impact funds. On top of being useful for sustainability-linked corporate managerial pay, it is useful for linking impact fund managers' pay to their real impact. A last benefit is the resources that can be saved from both the firms' side in attempting to prove sustainability results and in regulators' efforts to attempt to verify these firms' claims, which may be diluted by cheap-talk.

Our framework can be extended in various directions. First, it can be applied to any observable variable related to an externality (positive or negative), not just emissions (the S and G in ESG). Secondly, the framework may prove valuable for other asset classes other than stocks such as sustainability-linked bonds.

References

Akey, Pat and **Ian Appel**, "Environmental externalities of activism", Working Paper No. 3508808, SSRN, 2019.

Bams, Dennis and **Bram van der Kroft**, "Divestment, information asymmetries, and inflated ESG ratings", Working Paper No. 4126986, SSRN, 2022.

Berg, Florian, **Kornelia Fabisik** and **Zacharias Sautner**, "*Is History Repeating Itself? The (Un)predictable Past of ESG Ratings*", Working Paper No. 708, European Corporate Governance Institute, 2020.

Berg, Florian, **Julian F Koelbel** and **Roberto Rigobon**, "Aggregate confusion: The divergence of ESG ratings", *Forthcoming Review of Finance*, 2019.

Berk, Jonathan and **Jules van Binsbergen**, "The impact of impact investing", Working Paper No. 3909166, SSRN, 2021.

Bingler, Julia Anna, **Mathias Kraus**, **Markus Leippold** and **Nicolas Webersinke**, "Cheap talk and cherry-picking: What climatebert has to say on corporate climate risk disclosures", *Finance Research Letters*, 2022a, page 102776.

Bingler, Julia Anna, **Mathias Kraus**, **Markus Leippold** and **Nicolas Webersinke**, "Cheap Talk in Corporate Climate Commitments: The Effectiveness of Climate Initiatives", Working Paper No. 3998435, SSRN, 2022b.

Bonnefon, Jean-François, **Augustin Landier**, **Parinitha R Sastry** and **David Thesmar**, "The moral preferences of investors: Experimental evidence", Technical Report 29647, National Bureau of Economic Research, 2022.

Brøgger, Andreas and **Alexander Kronies**, "Skills and sentiment in sustainable investing", Working Paper No. 3531312, SSRN, 2022.

Cohen, Lauren, **Umit G Gurun** and **Quoc H Nguyen**, "The ESG-innovation disconnect: Evidence from green patenting", Working Paper No. 27990, NBER, 2020.

Curtis, Quinn, **Jill Fisch** and **Adriana Z Robertson**, "Do ESG mutual funds deliver on their promises?", *Mich L Rev*, 2021, *120*, 393.

Edmans, Alex, **Doron Levit** and **Jan Schneemeier**, "Socially responsible divestment", Working Paper No. 823, European Corporate Governance Institute, 2022.

Heath, Davidson, **Daniele Macciocchi**, **Roni Michaely** and **Matthew C Ringgenberg**, "Does Socially Responsible Investing Change Firm Behavior?", Working Paper No. 762, European Corporate Governance Institute, 2021.

Heath, Davidson, **Daniele Macciocchi**, **Roni Michaely** and **Matthew C Ringgenberg**, "Do index funds monitor?", *The Review of Financial Studies*, 2022, *35* (1), 91–131.

Heeb, Florian, **Julian F Kölbel**, **Falko Paetzold** and **Stefan Zeisberger**, "Do investors care about impact?", *Forthcoming Review of Financial Studies*, 2022.

Holmström, Bengt and **Paul Milgrom**, "Aggregation and linearity in the provision of intertemporal incentives", *Econometrica*, 1987, *55* (2), 303–328.

Homanen, Mikael, "Active depositors", *Journal of Banking & Finance*, 2022, *136*, 106417.

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

IBM 2023, "New IBM Study Reveals Inadequate Data Hinders Progress Against Environmental, Social and Governance Goals", Available at https://newsroom.ibm.com/2023-04-17-New-IBM-Study-Reveals-Inadequate-Data-Hinders-Progress-Against-Environme -Social-and-Governance-Goals (2023/04/23), .

Känzig, Diego R, "The unequal economic consequences of carbon pricing", Technical report, National Bureau of Economic Research, 2023.

Li, Tao, **S Lakshmi Naaraayanan** and **Kunal Sachdeva**, "Conflicting Objectives Between ESG Funds and Their Investors: Evidence From Mutual Fund Voting Records", Working Paper No. 3760753, SSRN, 2021.

Michaely, Roni, **Guillem Ordonez-Calafi** and **Silvina Rubio**, "Mutual funds' strategic voting on environmental and social issues", Working Paper No. 774, European Corporate Governance Institute, 2021.

Modigliani, Franco and **Merton H Miller**, "The cost of capital, corporation finance and the theory of investment", *The American economic review*, 1958, *48* (3), 261–297.

Noh, Don and **Sangmin Oh**, "Measuring institutional pressure for greenness: A demand system approach", Working Paper No. 3639693, SSRN, 2020.

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

Pedersen, Lasse Heje, **Shaun Fitzgibbons** and **Lukasz Pomorski**, "Responsible investing: The ESG-efficient frontier", *Journal of Financial Economics*, 2021, *142* (2), 572–597.

Sachdeva, Sonya, **Rumen Iliev** and **Douglas L Medin**, "Sinning saints and saintly sinners: The paradox of moral self-regulation", *Psychological science*, 2009, *20* (4), 523–528.

Schwitzgebel, Eric, "Do ethicists steal more books?", *Philosophical Psychology*, 2009, *22* (6), 711–725.

Schwitzgebel, Eric and **Joshua Rust**, "The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior", *Philosophical Psychology*, 2014, *27* (3), 293–327.

Zerbib, Olivier David, "A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from Green Investing and Sin Stock Exclusion", *Forthcoming Review of Finance*, 2019.

9 Figures and Tables

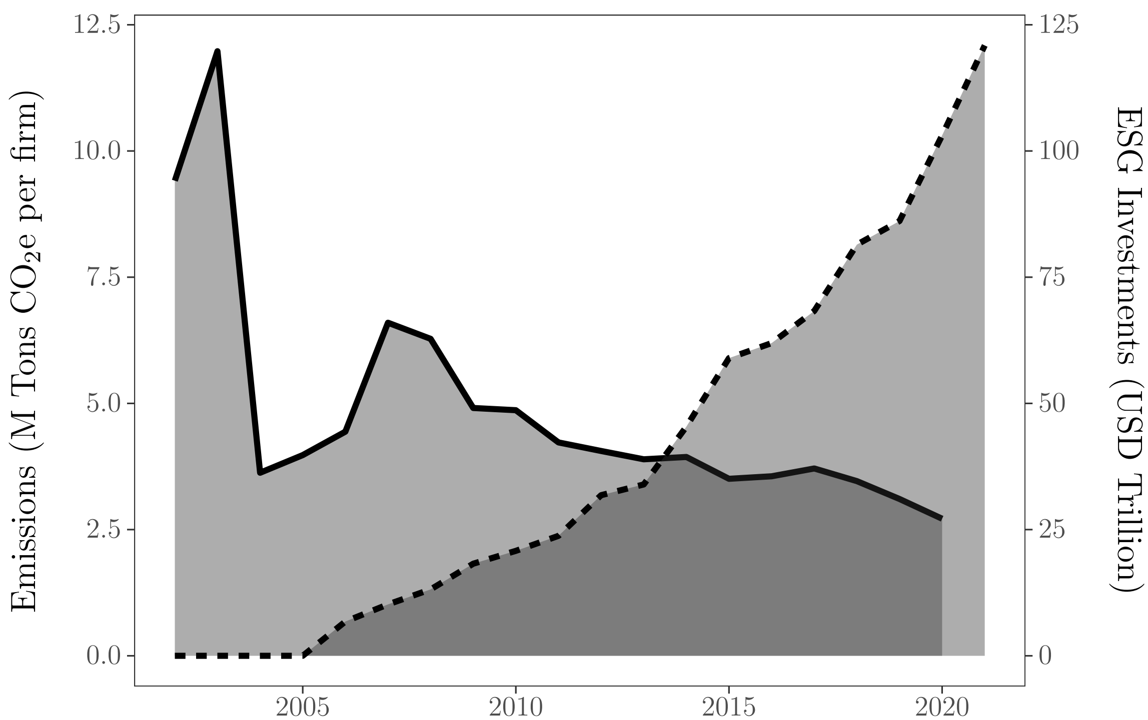

Figure 2: Emissions versus Socially Invested Capital

This figure plots the average emissions of firms over time as well as the amount of socially invested capital as reported by United Nations Principles of Responsible Investment. The CO2 emissions include direct CO_2 emissions.

Figure 3: Historical valuation of Emission Futures and counterfactual impact

Figure shows as the top line in light gray the historical valuation of Emission Futures, $\mathcal{F}_{i,t,n}$, normalized by the value of carbon futures at matching horizons, $\mathcal{F}^{co2}_{t,n}$. The value of the normalized Emission Future can then be used as an expectation of the emissions of a company. For illustrative purposes we show a second line in black, which signifies the hypothetical valuation under a new policy which aims to reduce carbon emissions. The new policy's impact is then the area in green. We compute the first datapoint as the average emissions emitted by the firms in our dataset. The expected emissions are then computed by regressing a firm's future emissions on to its current emissions at horizons of one to five years and scaling by the expected current emissions. Concretely, we in the second step conduct AR(n) regressions of the form $CO2_{t+n} = \alpha CO2_t + \epsilon$ for n ranging from one to five years, and multiply α with $CO2_t$. The CO2 emissions include direct CO_2 emissions and is average per firm. The dotted lines indicate the threshold of significance at the 5% level.

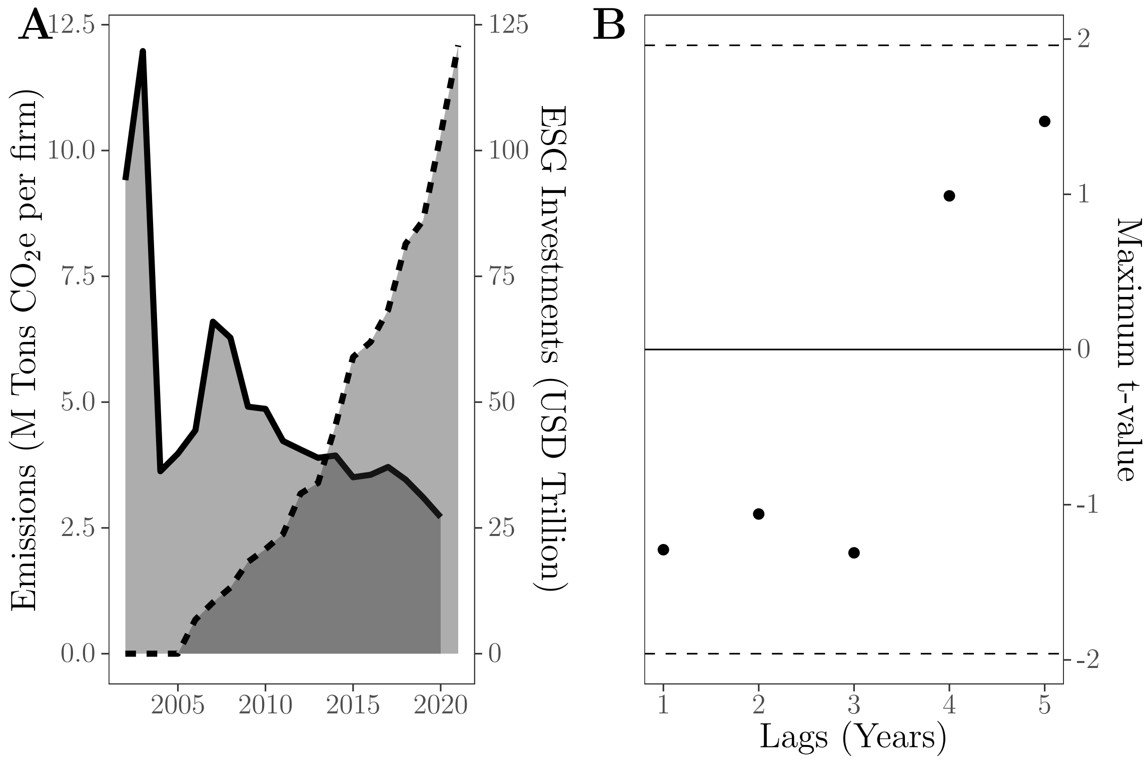

Figure 4: Social Capital's influence on Emissions from Granger Causality Tests

Figure's Panel A repeats Figure 2. Panel B plots the maximum t-value from a Granger Causality regression of the form $CO2_t = \sum_{t-n,t-1} AUM_{t-n} + CO2_{t-n} + \epsilon$ for n ranging from one to five years. AUM is the amount of socially capital invested as reported by United Nations Principles of Responsible Investment. The CO2 emissions include direct CO_2 emissions and is average per firm. The dotted lines indicate the threshold of significance at the 5% level.

Figure 5: Index Effect in the Time Series

Figure shows how being in the Social Index are correlated with a firm's emissions. We see that the average emissions of the firms in the index has not changed since 2001. On the other hand we see that it is the firms outside the index that have decreased their emissions in this time period. Emissions are averages per firm. The emissions measure includes both direct and indirect emissions. However the figure is similar if we only consider direct emissions.

Figure 6: Index Effect in the Cross-section

Figure shows how emissions adjust before and after firms are admitted to the social index. Emissions is compared to firms' time and cross-sectional average by including firm and time fixed effects. Band signifies significance at 95% level. Standard errors clustered at firm-month.

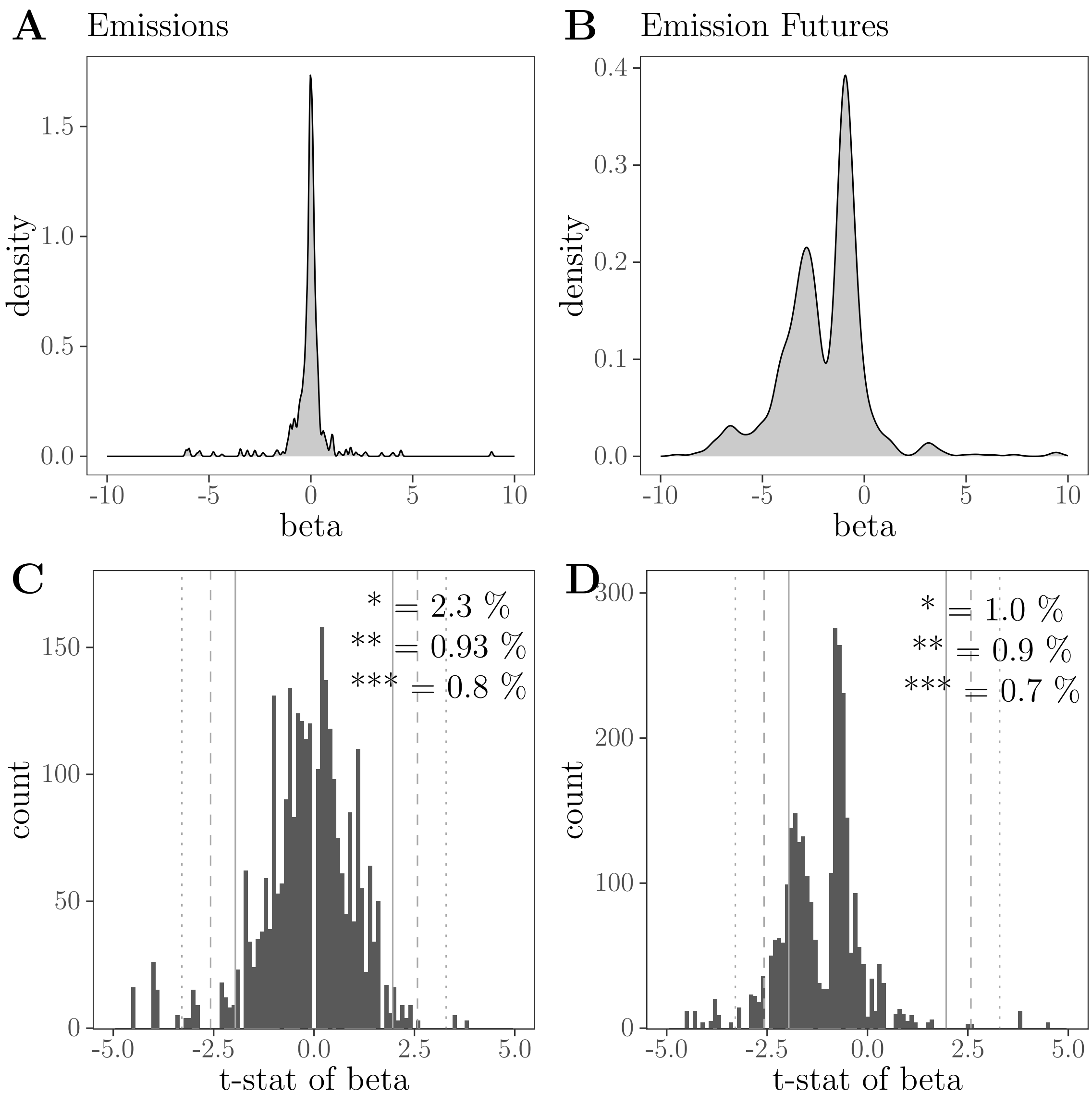

Figure 7: Carbon Emission Betas

This figure plots the carbon dioxide (CO_2) emissions and Emission Future betas with respect to the SP500. CO_2 is the percentage difference and the as the carbon price the carbon future price is used (expiry in December). Carbon price data from investing.com.

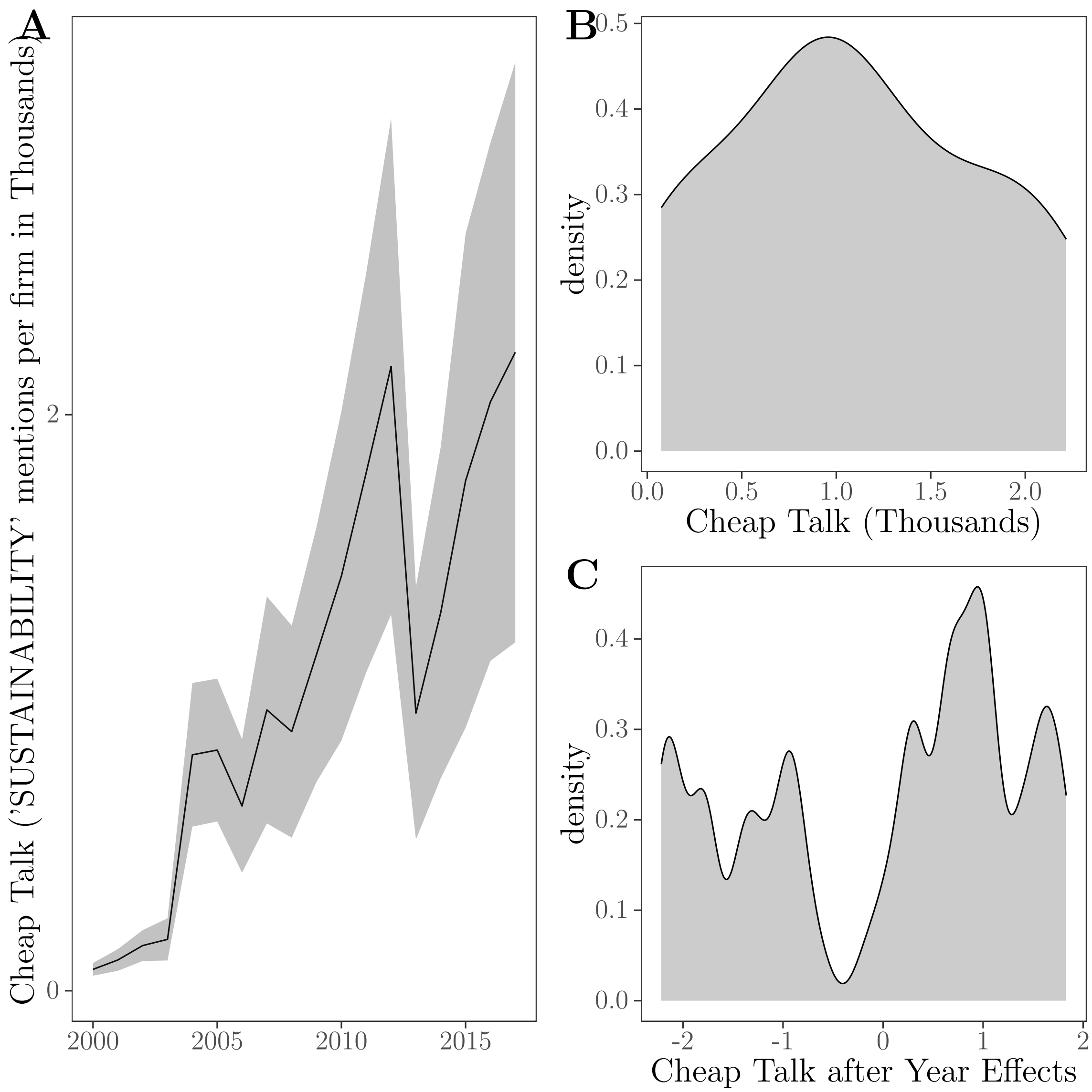

Figure 8: Cheap Talk in Time Series and Cross-section

Cheap talk is measured as 'ESG' mentions per firm in Thousands. Data from their public reports to the securities exchange commission. In Panel A grey area represents one standard deviation variation.

Table 1: Effect of ESG Scores on Impact: A Granger Causality test

Table shows the effects of ESG scores on impact. Specifically, it shows the output from a Granger causality test of emissions and environment scores on past emissions and environment scores. Panel A uses Refinitiv 'E' scores and Panel B uses Sustainalytics 'E' scores.

Panel A: With Refinitiv Scores

	Emissions next year (T Tons CO_2e)		E Score Next year (0-100)	
	(1)	(2)	(3)	(4)
E Score Current (0-100)		-1.48	0.92***	0.85***
		[-0.4]	[158.1]	[91.4]
Emissions Current (T Tons CO_2e)	0.97***	0.97***		-0.00***
	[362.4]	[358.4]		[-3.9]
Constant	60.64	149.81	5.20***	10.86***
	[1.22]	[0.64]	[16.40]	[19.06]
Observations	2,334	2,334	4,881	2,470
R^2	0.983	0.983	0.837	0.778

Panel B: With Sustainalytics Scores

	Emissions next year (T Tons CO_2e)		E Score Next year (0-100)	
	(1)	(2)	(3)	(4)
E Score Current (0-100)		1.64*	0.91***	0.81***
		[1.8]	[728.8]	[390.6]
Emissions Current (T Tons CO_2e)	0.96***	0.96***		0.00
	[719.8]	[717.2]		[0.02]
Constant	150.5***	28.70	7.79***	15.91***
	[7.2]	[0.4]	[102.5]	[98.8]
Observations	45,271	45,240	100,044	47,708
R^2	0.92	0.92	0.84	0.76

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: Social Capitals Impact on CO2 Emissions in Time Series

Table shows how social capital has affected CO2 emissions of the firms in the economy. The regressions are Granger Causality regressions of the form $CO2_t = \sum_{t-n,t-1} AUM_{t-n} + CO2_{t-n} + \epsilon$ for n ranging from one to five years. AUM is the amount of socially capital invested as reported by United Nations Principles of Responsible Investment. The CO2 emissions include direct CO_2 emissions and is average across firm in a given year. Standard errors are robust standard errors with automatic type and lag length choice. T-statistics are shown in square brackets.

	Emissions in Million Tons CO_2e per firm at time t ($CO_{2,t}$)				
	Model 1	Model 2	Model 3	Model 4	Model 5
AUM_{t-1} (T USD)	−0.02	0.03	0.00	0.00	−0.01
	[−1.29]	[0.67]	[0.27]	[−0.06]	[−0.30]
AUM_{t-2} (T USD)		−0.07	0.01	0.03	0.05
		[−1.06]	[0.59]	[0.99]	[1.47]
AUM_{t-3} (T USD)			−0.03	−0.03	−0.03
			[−1.31]	[−0.75]	[−0.76]
AUM_{t-4} (T USD)				0.00	0.00
				[−0.12]	[0.09]
AUM_{t-5} (T USD)					−0.03
					[−1.18]
$CO_{2,t-1}$ (M Tons)	0.44	0.05	0.67***	0.38	0.18
	[0.74]	[0.15]	[5.87]	[0.59]	[0.32]
$CO_{2,t-2}$ (M Tons)		−0.06	0.00	0.05	0.41
		[−0.28]	[−0.17]	[0.10]	[0.81]
$CO_{2,t-3}$ (M Tons)			0.05’	−0.07	−0.57
			[2.06]	[−0.56]	[−1.48]
$CO_{2,t-4}$ (M Tons)				0.22	0.44
				[1.19]	[2.05]
$CO_{2,t-5}$ (M Tons)					0.04
					[0.26]
(Intercept)	2.83	4.94*	1.01	1.43	1.45
	[0.93]	[2.72]	[1.68]	[1.67]	[1.42]
Num.Obs.	18	17	16	15	14
R2	0.436	0.500	0.947	0.967	0.988

’ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Index Effect in Cross-section

This table's Panel A shows how firms in the social index compares to firms outside of the index in terms of their emissions. Panel B shows how firms' emissions adjust while in the social index compared to outside of it Δ's are year differences. Column (1) is the absolute change and Column (2) is the relative fractional change.

	Panel A:		**Panel B:**	
	CO_2 (T Tons)	$\log(CO_2)$	$\Delta CO_{2,t,t+1}$ (T Tons)	$\Delta(\log CO_2)_{t,t+1}$
	(1)	(2)	(1)	(2)
Firm in Index	$-2{,}480^{***}$	$-23\%^{***}$	69^{***}	$2.3\%^{***}$
	[−28.9]	[−11.2]	[3.6]	[6.5]
Constant	$4{,}645^{***}$	13.1^{***}	-86^{***}	$-2.93\%^{***}$
	[73.68]	[874.57]	[−6.25]	[−11.69]
Observations	41,508	41,508	37,056	37,056
R^2	0.02	0.003	0.0003	0.001
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4: Effects of Cheap Talk on ESG Scores and Impact

Table considers how cheap talk by firms affects future ESG scores and impact. ESG Scores are considered in Panel A, and impact on carbon emissions in Panel B. Panel C considers whether firms with higher ESG scores do more cheap talk. Analyses take a Granger causality approach meaning they see if the dependent variable is predicted by lagged realisations of an independent variable in excess of its own lagged realisations.

Panel A:

	ESG Score Next Year			
	(1)	(2)	(3)	(4)
ESG Score This Year	0.818***	0.817***	0.813***	0.812***
	(0.012)	(0.012)	(0.012)	(0.012)
CO_2/Assets This Year		−0.0004***		−0.0004***
		(0.0001)		(0.0001)
SUSTAINABILITY Mentions This Year			0.014**	0.014**
			(0.006)	(0.006)
Constant	12.249***	12.428***	12.414***	12.584***
	(0.770)	(0.770)	(0.772)	(0.773)
Observations	1,505	1,505	1,505	1,505
R^2	0.747	0.749	0.748	0.750
Note:				*p<0.1; **p<0.05; ***p<0.01

Panel B:

	CO_2 over Assets Next Year		
	(1)	(2)	(3)
Word Mentions	−0.084		−0.108
	(0.175)		(0.177)
ESG Score		0.283	0.315
		(0.346)	(0.350)
CO_2/Assets	0.917***	0.917***	0.917***
	(0.005)	(0.005)	(0.005)
Constant	3.139	−15.463	−16.355
	(3.511)	(21.918)	(21.973)
Observations	1,158	1,158	1,158
R^2	0.965	0.965	0.965
Note:			*p<0.1; **p<0.05; ***p<0.01

Panel C:

	SUSTAINABILITY Mentions Next Year			
	(1)	(2)	(3)	(4)
CO_2/Assets	−0.0001			−0.0001
	(0.001)			(0.001)
ESG Score		0.149***	0.087*	0.086*
		(0.037)	(0.052)	(0.052)
Word Mentions	0.665***	0.573***	0.658***	0.658***
	(0.026)	(0.022)	(0.027)	(0.027)
Constant	4.378***	−4.439**	−0.982	−0.948
	(0.523)	(2.204)	(3.221)	(3.239)
Observations	1,219	1,766	1,219	1,219
R^2	0.342	0.306	0.344	0.344
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Welfare improvements from Emission Futures

Table considers the welfare improvement potential from Emission Futures. To do so, we calculate the Residual Standard Error of the emission forecasts seperately from an E rating, a firm's previous emissions, and both respectively. We do so both for Refinitiv (columns 1-3) and Sustainalytics E ratings (columns 4-6). We also denote the marginal reduction in RSE from column to column in percent.

	$CO2_{i,t+1}$ in Million Tons CO2 equivalents					
	(1)	(2)	(3)	(4)	(5)	(6)
E Score$_{i,t}$ (Refinitiv)	-0.17^{***}		-0.001			
	(0.03)		(0.004)			
E Score$_{i,t}$ (Sustainalytics)				5.69^{***}		0.16^{*}
				(0.27)		(0.09)
$CO2_{i,t}$		0.97^{***}	0.97^{***}		0.96^{***}	0.96^{***}
		(0.003)	(0.003)		(0.001)	(0.001)
Constant	15.33^{***}	0.06	0.15	-0.27	0.15^{***}	0.03
	(1.56)	(0.05)	(0.23)	(0.21)	(0.02)	(0.07)
Residual Standard Error (M Tons CO2e)	16.53	2.27	2.27	14.95	4.28	4.28
Marginal reduction in RSE		86%	0%		71%	0%
Degrees of freedom	2,563	2,332	2,331	51,328	45,269	45,237
Note:				*p<0.1; **p<0.05; ***p<0.01		

Appendix

A Proofs for Principal-Agent Model (Section 2.4)

A.1 Derivation of Optimal Incentive Weights

Given the linear contract $w = \alpha + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_3$, the agent's certainty equivalent is:

$$\begin{aligned}\mathrm{CE}_A = \alpha + (\beta_1 + \beta_2 + \beta_3)e + \beta_2 d - \frac{1}{2}k_e e^2 - \frac{1}{2}k_d d^2 \\ - \frac{a}{2}\left(\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2 + \beta_3^2\sigma_3^2\right).\end{aligned}$$

The agent's first-order conditions with respect to e and d yield:

$$e^*(\beta) = \frac{\beta_1 + \beta_2 + \beta_3}{k_e}, \qquad d^*(\beta) = \frac{\beta_2}{k_d}. \tag{33}$$

Using the binding participation constraint, the principal's problem becomes:

$$\max_{\beta} \quad ve^*(\beta) - \frac{1}{2}k_e e^{*2}(\beta) - \frac{1}{2}k_d d^{*2}(\beta) - \frac{a}{2}\sum_{i=1}^{3}\beta_i^2\sigma_i^2. \tag{34}$$

Substituting the agent's optimal responses and defining precisions $p_i = \sigma_i^{-2}$, the first-order conditions are:

$$\frac{\partial}{\partial \beta_1}: \quad \frac{v}{k_e} - e^* - \frac{a\beta_1}{p_1} = 0, \tag{35}$$

$$\frac{\partial}{\partial \beta_2}: \quad \frac{v}{k_e} - e^* - \frac{\beta_2}{k_d} - \frac{a\beta_2}{p_2} = 0, \tag{36}$$

$$\frac{\partial}{\partial \beta_3}: \quad \frac{v}{k_e} - e^* - \frac{a\beta_3}{p_3} = 0. \tag{37}$$

Let $S = p_1 + p_2 + p_3$ and $S_{13} = p_1 + p_3$. Solving this system of equations yields:

$$\beta_1^* = \frac{v}{a}\frac{p_1\left(k_d + \frac{p_2}{a}\right)}{\Delta}, \tag{38}$$

$$\beta_2^* = \frac{v}{a}\frac{p_2 k_d}{\Delta}, \tag{39}$$

$$\beta_3^* = \frac{v}{a}\frac{p_3\left(k_d + \frac{p_2}{a}\right)}{\Delta}, \tag{40}$$

where

$$\Delta = \left(k_e + \frac{S}{a}\right)\left(k_d + \frac{p_2}{a}\right) - \left(\frac{p_2}{a}\right)^2. \tag{41}$$

Equilibrium effort and greenwashing are:

$$e^* = \frac{v}{ak_e}\frac{k_d S + \frac{p_2}{a}S_{13}}{\Delta}, \tag{42}$$

$$d^* = \frac{v}{a}\frac{p_2}{\Delta}. \tag{43}$$

A.2 Proof of Proposition 1

Part 1: Comparative statics with respect to k_d.

Taking the derivative of β_2^* with respect to k_d:

$$\frac{\partial \beta_2^*}{\partial k_d} = \frac{v}{a}\frac{p_2}{\Delta^2}\left[\frac{\partial \Delta}{\partial k_d}\right], \tag{44}$$

where

$$\frac{\partial \Delta}{\partial k_d} = k_e + \frac{S}{a} > 0. \tag{45}$$

Since the numerator of β_2^* grows linearly in k_d while the denominator Δ grows faster (quadratically through the cross-term), we have:

$$\frac{\partial \beta_2^*}{\partial k_d} = \frac{vp_2}{a}\frac{\left(k_e + \frac{S}{a}\right)\left(k_d + \frac{p_2}{a}\right) - 2\left(\frac{p_2}{a}\right)^2 - \left(k_e + \frac{S}{a}\right)k_d}{\Delta^2}. \tag{46}$$

Simplifying the numerator:

$$\left(k_e + \frac{S}{a}\right)\frac{p_2}{a} - 2\left(\frac{p_2}{a}\right)^2 < 0 \tag{47}$$

for sufficiently small p_2 (noisy manipulable signal). Therefore, $\frac{\partial \beta_2^*}{\partial k_d} < 0$: cheaper greenwashing leads the principal to reduce weight on the manipulable signal.

By symmetry and the budget constraint on total incentive power, β_1^* and β_3^* must increase as β_2^* decreases.

Part 2: Effort depression.

When greenwashing is impossible ($k_d \to \infty$), the standard Holmström-Milgrom result obtains:

$$e^*_{\text{no-gw}} = \frac{v}{k_e} \cdot \frac{S}{ak_e + S}. \tag{48}$$

For finite k_d, equilibrium effort is:

$$e^* = \frac{v}{ak_e} \frac{k_d S + \frac{p_2}{a} S_{13}}{\Delta}. \tag{49}$$

We need to show $e^* < e^*_{\text{no-gw}}$. This is equivalent to:

$$\frac{k_d S + \frac{p_2}{a} S_{13}}{\Delta} < \frac{S}{ak_e + S}. \tag{50}$$

Substituting the definition of Δ and simplifying:

$$\frac{k_d S + \frac{p_2}{a} S_{13}}{\left(k_e + \frac{S}{a}\right)\left(k_d + \frac{p_2}{a}\right) - \left(\frac{p_2}{a}\right)^2} < \frac{S}{ak_e + S} \tag{51}$$

$$\iff (k_d S + \frac{p_2}{a} S_{13})(ak_e + S) < S\left[\left(k_e + \frac{S}{a}\right)\left(k_d + \frac{p_2}{a}\right) - \left(\frac{p_2}{a}\right)^2\right]. \tag{52}$$

Expanding both sides and canceling common terms, the inequality reduces to:

$$0 < k_d k_e \frac{p_2}{a}(S - S_{13}) = k_d k_e \frac{p_2^2}{a} > 0, \tag{53}$$

which holds for all finite $k_d > 0$. Therefore, $e^* < e^*_{\text{no-gw}}$ whenever greenwashing is possible. ■

A.3 Proof of Proposition 2

We prove the existence of parameter regions satisfying conditions (i)-(iv).

Condition (i): $\beta_2^*(v)$ increasing in v.

From the closed-form solution:

$$\beta_2^*(v) = \frac{v}{a}\frac{p_2 k_d}{\Delta}. \tag{54}$$

Since Δ does not depend on v (it depends only on cost and precision parameters), we have:

$$\frac{\partial \beta_2^*}{\partial v} = \frac{p_2 k_d}{a\Delta} > 0. \tag{55}$$

This holds for all parameter values. ✓

Condition (ii): Type L responds more to increases in v.

For each type $\theta \in \{L, H\}$, greenwashing is:

$$d^*(\theta; v) = \frac{\beta_2^*(v)}{k_d(\theta)} = \frac{v}{a}\frac{p_2 k_d}{\Delta k_d(\theta)}. \tag{56}$$

Taking derivatives:

$$\frac{\partial d^*(L; v)}{\partial v} = \frac{p_2 k_d}{a\Delta k_d^L}, \qquad \frac{\partial d^*(H; v)}{\partial v} = \frac{p_2 k_d}{a\Delta k_d^H}. \tag{57}$$

Since $k_d^L < k_d^H$, we have:

$$\frac{\partial d^*(L; v)}{\partial v} = \frac{p_2 k_d}{a\Delta k_d^L} > \frac{p_2 k_d}{a\Delta k_d^H} = \frac{\partial d^*(H; v)}{\partial v}. \tag{58}$$

This holds for all $k_d^L < k_d^H$. ✓

Condition (iii): Type uncertainty increases dispersion.

True impact for each type is $I(\theta) = e^* - \gamma d^*(\theta)$. Since effort e^* is type-independent, the variance of impact across types is:

$$\mathrm{Var}(I(\theta)) = \gamma^2 \mathrm{Var}(d^*(\theta)) = \gamma^2 \pi(1-\pi)[d^*(L) - d^*(H)]^2. \tag{59}$$

The difference in greenwashing between types is:

$$d^*(L) - d^*(H) = \frac{v}{a}\frac{p_2 k_d}{\Delta}\left(\frac{1}{k_d^L} - \frac{1}{k_d^H}\right) = \frac{v p_2 k_d}{a\Delta}\frac{k_d^H - k_d^L}{k_d^L k_d^H}. \tag{60}$$

Therefore:

$$\mathrm{Var}(I(\theta; v)) = \gamma^2 \pi(1-\pi)\left[\frac{v p_2 k_d}{a\Delta}\frac{k_d^H - k_d^L}{k_d^L k_d^H}\right]^2. \tag{61}$$

Taking the derivative with respect to v:

$$\frac{d}{dv}\mathrm{Var}(I(\theta; v)) = 2\gamma^2 \pi(1-\pi)\left[\frac{p_2 k_d}{a\Delta}\frac{k_d^H - k_d^L}{k_d^L k_d^H}\right]^2 v > 0. \tag{62}$$

This holds for all $v > 0$ when $k_d^L \neq k_d^H$. ✓

Condition (iv): Expected impact decreasing in v for some interval.

Expected true impact is:

$$\mathbb{E}[I(v)] = e^*(v) - \gamma \mathbb{E}[d^*(\theta; v)] = e^*(v) - \gamma[\pi d^*(L; v) + (1-\pi) d^*(H; v)]. \tag{63}$$

From the equilibrium characterization:

$$e^*(v) = \frac{v}{a k_e}\frac{k_d S + \frac{p_2}{a} S_{13}}{\Delta}, \tag{64}$$

$$\mathbb{E}[d^*(\theta; v)] = \frac{v}{a}\frac{p_2 k_d}{\Delta}\left[\frac{\pi}{k_d^L} + \frac{1-\pi}{k_d^H}\right] = \frac{v}{a}\frac{p_2 k_d}{\Delta \bar{k}_d}, \tag{65}$$

where $\bar{k}_d^{-1} = \pi k_d^{L-1} + (1-\pi) k_d^{H-1}$ is the harmonic mean of greenwashing costs.

Therefore:

$$\mathbb{E}[I(v)] = \frac{v}{a\Delta}\left[\frac{k_d S + \frac{p_2}{a} S_{13}}{k_e} - \gamma \frac{p_2 k_d}{\bar{k}_d}\right]. \tag{66}$$

Taking the derivative:

$$\frac{d\mathbb{E}[I(v)]}{dv} = \frac{1}{a\Delta}\left[\frac{k_d S + \frac{p_2}{a} S_{13}}{k_e} - \gamma \frac{p_2 k_d}{\bar{k}_d}\right]. \tag{67}$$

Expected impact is *decreasing* in v when:

$$\gamma > \gamma^* \equiv \frac{\bar{k}_d(k_d S + \frac{p_2}{a} S_{13})}{p_2 k_d k_e}. \tag{68}$$

For $\gamma > \gamma^*$, we have $\frac{d\mathbb{E}[I(v)]}{dv} < 0$ for all $v > 0$.

Existence of parameter region.

Choose parameters:

- $k_d^L = 1$, $k_d^H = 10$ (large heterogeneity in greenwashing costs)
- $\pi = 0.5$ (equal type probabilities)
- p_2 large relative to p_1, p_3 (manipulable signal is precise)
- γ large (greenwashing strongly contaminates impact measurement)
- k_e moderate, a moderate

With these parameters, conditions (i)-(iii) hold by construction from the above derivations. Condition (iv) holds when γ exceeds the threshold γ^*. Since γ^* is finite and depends continuously on the other parameters, there exists a non-empty open set of parameters satisfying all four conditions simultaneously.

Moreover, the negative spiral is most pronounced when:

1. Type heterogeneity is large ($k_d^L \ll k_d^H$)
2. The manipulable signal is relatively precise (p_2 large)
3. Greenwashing strongly distorts impact (γ large)
4. The prior probability of low-cost types is intermediate ($\pi \approx 0.5$)

Under these conditions, increasing the principal's valuation of impact v creates strong incentives on the manipulable signal, which low-cost types exploit through greenwashing, raising type-induced dispersion and lowering expected true impact. ■

A.4 Benchmark: No Greenwashing

When manipulation is impossible ($k_d \to \infty$), the solution converges to the standard CARA-Normal agency model with multiple signals. The optimal weights become:

$$\lim_{k_d \to \infty} \beta_i^* = \frac{v p_i}{a(k_e + \frac{S}{a})}, \quad i = 1, 2, 3, \tag{69}$$

which are proportional to signal precisions (Holmström-Milgrom aggregation result). Equilibrium effort becomes:

$$\lim_{k_d \to \infty} e^* = \frac{v}{k_e} \cdot \frac{S}{a k_e + S} = \frac{e_{\text{FB}}}{1 + \frac{a k_e}{S}}, \tag{70}$$

where $e_{\text{FB}} = v/k_e$ is the first-best effort level. As total signal precision $S \to \infty$, effort approaches first-best.

A.5 Additional Comparative Statics

Precise unbiased signals restore first-best. If $p_1 \to \infty$ or $p_3 \to \infty$, then $S \to \infty$ and:

$$\lim_{S \to \infty} e^* = \lim_{S \to \infty} \frac{v}{a k_e} \frac{k_d S + \frac{p_2}{a} S_{13}}{\Delta} = \frac{v}{k_e} = e_{\text{FB}}. \tag{71}$$

Cheap manipulation eliminates the manipulable signal. If $k_d \to 0$:

$$\lim_{k_d \to 0} \beta_2^* = 0, \tag{72}$$

$$\lim_{k_d \to 0} e^* = \frac{v}{k_e} \cdot \frac{S_{13}}{a k_e + S_{13}}, \tag{73}$$

which is the effort level using only the unbiased signals s_1 and s_3. The manipulable signal s_2 is completely discarded.

A.6 Extension to Continuous Type Distributions

The negative spiral result in Proposition 2 is robust to continuous type distributions. We show this by replacing the discrete types $\{L, H\}$ with a log-normal distribution of greenwashing costs.

Continuous Type Specification. Let greenwashing costs be distributed according to:

$$\ln(k_d(\theta)) \sim N(\mu_{\ln k}, \sigma^2_{\ln k}), \qquad k_d(\theta) > 0, \tag{74}$$

where $\mu_{\ln k}$ determines the mean greenwashing cost and $\sigma_{\ln k}$ captures the degree of heterogeneity across firms. This specification ensures strictly positive costs while allowing for substantial variation. Some firms have low greenwashing costs (high $\ln(k_d)$ realizations), while others find manipulation prohibitively expensive (low $\ln(k_d)$ realizations).

The log-normal distribution has desirable properties for our application:

1. Natural lower bound at zero ensures $k_d(\theta) > 0$ for all types
2. Right-skewed: captures the idea that most firms have moderate costs with a long tail of high-cost types
3. Tractable moments: allows closed-form expressions for key statistics

Expected Greenwashing and Variance. Given contract (α, β), each type θ chooses greenwashing $d^*(\theta) = \beta_2/k_d(\theta)$. Using properties of the log-normal distribution, the expected greenwashing across types is:

$$\mathbb{E}[d^*(\theta)] = \beta_2 \mathbb{E}\left[\frac{1}{k_d(\theta)}\right] = \beta_2 \exp\left(-\mu_{\ln k} + \frac{\sigma^2_{\ln k}}{2}\right). \tag{75}$$

The variance of greenwashing responses is:

$$\mathrm{Var}(d^*(\theta)) = \beta_2^2 \exp(-2\mu_{\ln k} + \sigma^2_{\ln k}) \left[\exp(\sigma^2_{\ln k}) - 1\right]. \tag{76}$$

Crucially, $\mathrm{Var}(d^*(\theta))$ is increasing in both β_2^2 and $\sigma^2_{\ln k}$. Greater heterogeneity in greenwashing costs amplifies the dispersion of responses to incentives.

Expected Impact. True impact is $I(\theta) = e^* - \gamma d^*(\theta)$. Since effort e^* is type-independent, expected impact is:

$$\mathbb{E}[I(v)] = e^*(v) - \gamma \mathbb{E}[d^*(\theta; v)] = \frac{v}{ak_e} \frac{k_d S + \frac{p_2}{a} S_{13}}{\Delta} - \gamma \beta_2(v) \exp\left(-\mu_{\ln k} + \frac{\sigma^2_{\ln k}}{2}\right). \tag{77}$$

The Negative Spiral with Continuous Types. Taking the derivative with respect to v:

$$\frac{d\mathbb{E}[I(v)]}{dv} = \frac{1}{a\Delta}\left[\frac{k_d S + \frac{p_2}{a}S_{13}}{k_e} - \gamma\frac{p_2}{a}\exp\left(-\mu_{\ln k} + \frac{\sigma^2_{\ln k}}{2}\right)\right]. \tag{78}$$

Expected impact is *decreasing* in v when:

$$\gamma > \gamma^* \equiv \frac{k_d S + \frac{p_2}{a}S_{13}}{p_2 k_e}\cdot\left[\exp\left(-\mu_{\ln k} + \frac{\sigma^2_{\ln k}}{2}\right)\right]^{-1}. \tag{79}$$

For $\gamma > \gamma^*$, we have $\frac{d\mathbb{E}[I(v)]}{dv} < 0$: higher valuation of impact leads to lower expected true impact.

Comparison to Discrete Types. The continuous specification delivers the same qualitative result as the discrete two-type model but with important extensions:

1. **Smooth comparative statics**: The effect of heterogeneity on the negative spiral can be analyzed continuously by varying $\sigma_{\ln k}$, rather than jumping between two discrete cases.
2. **Richer dispersion effects**: The variance of impact is:

$$\mathrm{Var}(I(\theta; v)) = \gamma^2\beta_2^2(v)\exp(-2\mu_{\ln k} + \sigma^2_{\ln k})\left[\exp(\sigma^2_{\ln k}) - 1\right], \tag{80}$$

   which shows explicitly how heterogeneity $\sigma^2_{\ln k}$ amplifies the type-induced noise in impact measurement. As $\sigma_{\ln k} \to 0$, we recover the no-heterogeneity case with $\mathrm{Var}(I) = 0$.
3. **Threshold interpretation**: The critical value γ^* now depends on the full distribution of types through the moment generating function, not just on two specific values k_d^L and k_d^H.
4. **Empirical implementation**: The log-normal specification can be estimated from data on ESG score dispersion or greenwashing behavior, providing a direct link to empirical work.

Robustness. The negative spiral mechanism is not an artifact of the two-type specification. It arises from the fundamental tension between three forces:

1. Higher v induces the principal to rely more on the manipulable signal (β_2 in-

creases)

2. Type heterogeneity creates dispersion in greenwashing responses to incentives
3. This dispersion contaminates the information environment, reducing the effectiveness of performance measurement

This mechanism operates under any distribution with sufficient heterogeneity, including log-normal, truncated normal, and other continuous specifications. The two-type model in the main text isolates this mechanism in its simplest form.

B Alternative Measures

B.1 Market Adjusted E Measure

An environmental score adjusted for the correlation with the market outcomes is given as

$$\tilde{\mathcal{E}}_{i,t,n} \quad \equiv \quad -\mathcal{P}_{i,t,n} exp(n\theta_{i,t,n}).$$

B.2 Input Adjusted E Measure (Scope 3 Equivalent)

We can construct an E measure which takes into account whether a firm uses inputs, which have lead to emissions in their production. We call these input adjusted E measures, which are in line with the goal of Scope 3 emission measures. They are given as

$$\mathcal{E}^{③}_{i,t,n} \quad \equiv \quad [\mathbb{1} - \Sigma]^{-1} \, \mathcal{E}_{i,t,n},$$

where $\mathbb{1}$ is the n-dimensional identity matrix, and Σ is a n-by-n dimensional matrix which for row i and column j shows the fraction of firm j's output used in firm i's production.

To see this, start with the definition of $\mathcal{E}^{③}_{i,t,n}$ being the sum of its own standard emissions and the input-adjusted emissions of its inputs, $\mathcal{E}_{i,t,n} + \Sigma\mathcal{E}^{③}_{j,t,n}$, and solve for input-adjusted emissions using the fact that i and j are the same firms (and hence $\mathcal{E}^{③}_{i,t,n} = \mathcal{E}^{③}_{j,t,n}$).

B.3 Market Adjusted Green Impact

The market corrected impact score is given as

$$\tilde{I}_{i,t,n} \equiv \mathcal{F}_{i,t-1,n} \exp(-\theta^e_{i,t-1,n}) - \mathcal{F}_{i,t,n-1}.$$

B.4 Green Impact per Dollar Measure

A green impact investor may be interested how to achieve the highest impact per dollar, in which case it would be

$$\mathcal{G}_{i,t,n} \equiv \frac{R^{\$}_{i,t,n}}{P_{i,t-1}}.$$

B.5 Emission Reduction Measures

A set of measures that capture expected CO_2 reductions are given by:

$$\begin{aligned} \mathcal{R}_{i,t,n} &\equiv y^f_{i,t,n}, \\ \tilde{\mathcal{R}}_{i,t,n} &\equiv -g_{i,t,n}. \end{aligned}$$

where the tilde ($\sim$) denotes a measure adjusted for correlation with the market.

C Derivations

C.1 Derivation of Optimal ESG Score

We have the following cash flow equation

$$D^c_{i,t}(E_{i,t}) = D^{c,0}_{i,t} \exp\left(-B_{i,t}\left(E_{i,t}\right)\right) = D^{c,0}_{i,t} \exp\left(-k_i E^{1/\eta_i}_{i,t}\right),$$

And cost of equity equation

$$r^c_{i,t} = \beta^M_i r^M_t - s \ln(E_{i,t}),$$

and optimality condition

$$D_{i,t}^{c'}(E_{i,t}) = \mathbb{E}_t \left[-r_{i,t+1}^{C'}(E_{i,t}) \frac{D_{i,t+1}^{c}(E_{i,t+1})}{\exp\left(r_{i,t+1}^{c}(E_{i,t})\right)} \right].$$

Then start from Equation 13 repeated below for which it follows that

$$\underbrace{\frac{k_i E^{\frac{1-\eta_i}{\eta_i}} D_{i,t}^{c,0}}{\exp\left(k_i E_{i,t}^{1/\eta_i}\right)}}_{\text{Marginal cost}} = \underbrace{s E_i^{-1} \mathbb{E}_t \left[\frac{D_{i,t+1}^{c,0} \exp\left(-k_i E_i^{1/\eta_i}\right)}{\exp\left(\beta_i^M r_{t+1}^M - s\ln(E_i)\right)} \right]}_{\text{Marginal benefit}},$$

$$\begin{aligned} k_i E^{\frac{1-\eta_i}{\eta_i}} &= s E_i^{-1} \mathbb{E}_t \left[\exp\left(s\ln(E_i) - \beta_i^M r_{t+1}^M\right) \right], \\ k_i E^{\frac{1}{\eta_i}} &= s \, \mathbb{E}_t \left[\exp\left(s\ln(E_i) - \beta_i^M r_{t+1}^M\right) \right] \\ &= s E_i^s \, \mathbb{E}_t \left[\exp\left(-\beta_i^M r_{t+1}^M\right) \right]. \end{aligned}$$

Switch sides and further simplify to find

$$E_i^{s-\frac{1}{\eta_i}} = \mathbb{E}_t \left[\exp\left(\beta_i^M r_{t+1}^M\right) \right] \frac{k_i}{s},$$

$$E_i^* = \left(\mathbb{E}_t \left[\exp\left(\beta_i^M r_{t+1}^M\right) \right] \frac{k_i}{s} \right)^{1/(s-\frac{1}{\eta_i})},$$

$$E_i^* = \left(\frac{k_i}{s} \exp\left(\beta_i^M \mu_t^M - \frac{1}{2} {\beta_i^M}^2 {\sigma_t^M}^2 \right) \right)^{\frac{-1}{\frac{1}{\eta_i}-s}}.$$

$$E_i^* = \left(\frac{s}{k_i} \exp\left(-\beta_i^M \mu_t^M + \frac{1}{2} {\beta_i^M}^2 {\sigma_t^M}^2 \right) \right)^{\frac{1}{\eta_i}-s}.$$

As $\eta \in (0,1)$, k is strictly positive, and reasonable values for sentiment lie within $s \in (0,1)$ it means that

$$E_i^* = \psi\left(s, k_i, \beta_i^M, \mu_t^M, \sigma_t^M\right)^{\phi(\eta_i, s)},$$

where ψ and ϕ are positive definite.

This means that E will be increasing in ψ with a strength determined by ϕ. Hence E_i^* is increasing in s and σ_t^M and decreasing in k_i and μ_t^M. The market exposure β_i could be either. The strength of this effect is increasing in both s and η.

C.2 Derivation of Noisy Ratings Impact Decrease

Let the realised E_i^R score be the target score E_i, plus a realisation of a log-normally distributed noise term σ^R. The marginal cost is the same as it is based on the targeted score, however the marginal benefit will change to:

$$sE_i^{-1}\,\mathbb{E}_t\left[\frac{D_{i,t,1}^{c,0}\exp\left(-k_iE_i^{1/\eta_i}\right)}{\exp\left(\beta_i^M r_{t+1}^M - s\ln(E_i^R)\right)}\right] \tag{81}$$

$$= sE_i^{-1}\frac{D_{i,t,1}^{c,0}\exp\left(-k_iE_i^{1/\eta_i}\right)}{\mathbb{E}_t\left[\exp\left(\beta_i^M r_{t+1}^M - s\ln(E)+\frac{1}{2}\sigma^{R^2}\right)\right]} \tag{82}$$

$$= sE_i^{s-1}\frac{D_{i,t,1}^{c,0}\exp\left(-k_iE_i^{1/\eta_i}\right)}{\mathbb{E}_t\left[\exp\left(\beta_i^M r_{t+1}^M +\frac{1}{2}\sigma^{R^2}\right)\right]} \tag{83}$$

$$= MB_i^M\exp\left(-\frac{1}{2}\sigma^{R^2}\right), \tag{84}$$

where MB_i^M is the marginal benefit without noisy ESG ratings.

This means the new equilibrium E score will be

$$\underbrace{\frac{k_iE^{\frac{1-\eta_i}{\eta_i}}D_{i,t}^{c,0}}{\exp\left(k_iE_{i,t}^{1/\eta_i}\right)}}_{\text{Marginal cost}} = \underbrace{sE_i^{-1}\,\mathbb{E}_t\left[\frac{D_{i,t+1}^{c,0}\exp\left(-k_iE_i^{1/\eta_i}\right)}{\exp\left(\beta_i^M r_{t+1}^M - s\ln(E_i)\right)}\right]\exp\left(-\frac{1}{2}\sigma^{R^2}\right)}_{\text{Marginal benefit}},$$

$$k_iE^{\frac{1-\eta_i}{\eta_i}} = sE_i^{-1}\,\mathbb{E}_t\left[\exp\left(s\ln(E_i)-\beta_i^M r_{t+1}^M\right)\right]\exp\left(-\frac{1}{2}\sigma^{R^2}\right),$$

$$k_iE^{\frac{1}{\eta_i}} = s\,\mathbb{E}_t\left[\exp\left(s\ln(E_i)-\beta_i^M r_{t+1}^M\right)\right]\exp\left(-\frac{1}{2}\sigma^{R^2}\right)$$

$$= sE_i^s\,\mathbb{E}_t\left[\exp\left(-\beta_i^M r_{t+1}^M\right)\right]\exp\left(-\frac{1}{2}\sigma^{R^2}\right).$$

Switch sides and further simplify to find

$$E_i^{s-\frac{1}{\eta_i}} = \mathbb{E}_t\left[\exp\left(\beta_i^M r_{t+1}^M\right)\right]\frac{k_i}{s}\exp\left(\frac{1}{2}\sigma^{R^2}\right),$$

$$E_i^* = \left(\mathbb{E}_t\left[\exp\left(\beta_i^M r_{t+1}^M\right)\right]\frac{k_i}{s}\exp\left(\frac{1}{2}\sigma^{R^2}\right)\right)^{1/(s-\frac{1}{\eta_i})},$$

$$E_i^* = \left(\frac{k_i}{s}\exp\left(\beta_i^M \mu_t^M - \frac{1}{2}\beta_i^{M^2}\sigma_t^{M^2}\right)\right)^{\frac{-1}{\frac{1}{\eta_i}-s}}\exp\left(\frac{1}{2}\sigma^{R^2}\right)^{\frac{-1}{\frac{1}{\eta_i}-s}},$$

$$E_i^{*,R} = E_i^{*,M}\exp\left(\frac{1}{2}-\sigma^{R^2}\right)^{\frac{1}{\eta_i}-s},$$

where $E_i^{*,R}$ is the equilibrium E score without noisy ratings and $E_i^{*,M}$ without.

C.3 Derivation of Cheap Talk Effects

Let E be governed by the Cobb-Douglas CES function:

$$E_{i,t} = \frac{1}{k_i}(B^{\theta}_{i,t}T^{1-\theta}_{i,t})^{\eta}_{i}, \tag{85}$$

where θ signifies the importance of that factor in the production function.

We start by finding the optimal mix of impact and cheap talk to achieve ESG score E. For B and T to be in optimal proportions the marginal rate of substitution between B and T must equal the relative cost of B and T:

$$MRS = \frac{p_I}{p_T}. \tag{86}$$

We can find the MRS defined as $\partial T/\partial B$ by taking the ratio of $\partial E/\partial B$ and $\partial E/\partial T$ as

$$\begin{aligned} \frac{\partial E}{\partial B} &= \frac{\theta\eta}{B}(B^{\theta}T^{1-\theta})^{\eta} = \frac{\theta\eta}{B}E, \\ \frac{\partial E}{\partial T} &= \frac{(1-\theta)\eta}{T}(B^{\theta}T^{1-\theta})^{\eta} = \frac{(1-\theta)\eta}{T}BE, \end{aligned} \tag{87}$$

so

$$\frac{\partial T}{\partial B} = \frac{\theta}{1-\theta}\frac{T}{B} = MRS. \tag{88}$$

In optimum you cannot be better off from substituting between the two inputs compared to their input prices and hence in equilibrium

$$MRS = \frac{\theta}{1-\theta}\frac{T^*}{B^*} = \frac{p_I}{p_T} = \text{relative price}. \tag{89}$$

Solving for T^* gives

$$T^* = \frac{p_I}{p_T}\frac{1-\theta}{\theta}B^*. \tag{90}$$

The budget B' is defined as the sum of expenditures on impact and cheap talk:

$$B' = p_I B + p_T T \tag{91}$$

Substituting T^* into the budget equation, and solving for B, we get the optimal impact as

$$B^* = \frac{B'}{p_I}\theta. \tag{92}$$

Which we can rewrite to get the share of the budget spend on impact ($B^* p_I / B'$) as:

$$\frac{B^* p_I}{B'} = \theta. \tag{93}$$

Similarly we get

$$T^* = \frac{B'}{p_T}(1-\theta). \tag{94}$$

Hence the ratio is given by

$$\frac{B^*}{T^*} = \frac{\theta}{1-\theta}\frac{p_T}{p_I}. \tag{95}$$

We get the budget needed to achieve a score E by taking the ESG production function (Equation 85) and substituting in the equations for optimal impact and cheap talk (Equations 92 and 94), and solving for B':

$$B' = k^{1/\eta} E^{1/\eta} \left(\frac{p_I}{\theta}\right)^{\theta} \left(\frac{p_t}{1-\theta}\right)^{1-\theta}. \tag{96}$$

Wlog we can normalise $p_{I,i,t+n} = 1$. Hence they do more impact, the higher θ is and the higher $p_{T,i,t+n}$ is (where the price of cheap talk now is relative to the price of impact), which may be firm specific. As a simplifying example consider if $p_{T,i,t+n} = 1$, here firms do most impact and less cheap talk if θ is above 0.5.

This means the cash dividends become

$$D^c_{i,t,0}(E_{i,t}) = D^{c,0}_{i,t,0} \exp\left(-B'_{i,t}\left(E_{i,t}\right)\right) = D^{c,0}_{i,t,0} \exp\left(-k'_i E^{1/\eta_i}_{i,t}\right),$$

where k'_i is the constant $k_i \left(\frac{1}{\theta}\right)^{\theta} \left(\frac{p_T}{1-\theta}\right)^{1-\theta}$.

It is useful to note that we can recover the version without cheap talk versus impact by setting $k'_i = k_i$.

Credability of Cheap Talk. Next we turn to the consequence of firms needing to be credible to ensure their ESG discount. As the investor is not interested in funding cheap talk the investors strategy is that they only give discount if the firm does not do cheap talk. They infer the firm does cheap talk if they are more than 95% sure their

ESG score is from cheap talk. Both the firms cheap talk and impact is a signal.

Corollary 1 (Cheap Talk to Impact Ratio). *The largest cheap talk to impact ratio that they can have before losing benefit is*

$$\frac{T}{B} = 1.96\sqrt{\sigma_T^2 + \sigma_B^2}. \tag{97}$$

Proof. The signal the investor checks is larger than zero is

$$\frac{T}{B} = \frac{T}{E - T}. \tag{98}$$

And the noise, when uncorrelated, is

$$\sqrt{\sigma_T^2 + \sigma_B^2}. \tag{99}$$

So test is on

$$\frac{T/B}{\sqrt{\sigma_T^2 + \sigma_B^2}}. \tag{100}$$

Hence, the investor does not think the firm does cheap talk when [17]

$$\frac{T/B}{\sqrt{\sigma_T^2 + \sigma_B^2}} \leq 1.96. \tag{101}$$

So the largest fraction of cheap talk to impact that the firm can have before losing their benefit is

$$\frac{T}{B} = 1.96\sqrt{\sigma_T^2 + \sigma_B^2}. \tag{102}$$

■

C.4 Derivation of Sustainability-based Pay Impact Increase

Let the negative of the external value of the firm be the price of an emission future $\mathcal{P}_{i,t}$. Then his income Y is:

$$Y_{m,t} = S + k^I P_{i,t} - \mathcal{P}_{i,t} k^E P_{i,t}.$$

[17] Assuming a large number of observations. Which implies that the threshold is larger for younger firms.

Divide through by k^I and introduce $\kappa = k^E/k^I$

$$Y_{m,t}/k^I = S/k^I + P_{i,t} - \mathcal{P}_{i,t}\kappa P_{i,t}.$$

The manager's objective is to maximise his income by choosing which ESG score $E_{i,t}$ the firm should aim for:

$$\max_{E_{i,t}} Y_{m,t} = \max_{E_{i,t}} Y_{m,t}/k^I.$$

Then by comparison to Equation 13 his costs versus benefits optimality condition becomes:

$$\frac{k_i E^{\frac{1-\eta_i}{\eta_i}} D^{c,0}_{i,t}}{\exp(k_i E_i^{1/\eta_i})} - \frac{\kappa D^{c,0}_{i,t}}{\exp(k_i E^{1/\eta_i})} = sE_i^{-1}\,\mathbb{E}_t\left[\frac{D^{c,0}_{i,t+1}(1+\kappa E_i)\exp\left(-k_i E_i^{1/\eta_i}\right)}{\exp\left(\beta_i^M r^M_{t+1} - s\ln(E_i)\right)}\right].$$

$$k_i E^{\frac{1-\eta_i}{\eta_i}} - \kappa = sE_i^{-1}\,\mathbb{E}_t\left[\frac{1+\kappa E_i}{\exp\left(\beta_i^M r^M_{t+1}\right)}\right]E_i^s.$$

$$k_i E^{\frac{1-\eta_i}{\eta_i}} - \kappa = sE_i^{s-1}\frac{1+\kappa E_i}{\mathbb{E}_t\left[\exp\left(\beta_i^M r^M_{t+1}\right)\right]}.$$

An extreme case would be for the situation with no extra help from sustainability sentiment, $s = 0$, in that case it will be that

$$k_i E^{\frac{1-\eta_i}{\eta_i}} = \kappa,$$

and optimal E rating is

$$E_i^* = \left(\frac{\kappa}{k_i}\right)^{\frac{\eta_i - 1}{\eta_i}}.$$

Which means that the increase in impact from sustainability linked pay is

$$I^{R,M}_{i,t,0} = \left(\frac{\kappa}{k_i}\right)^{\frac{\eta_i - 1}{\eta_i}}.$$

D Additional Results and Robustness Tests

Figure D.1: Carbon Emission Intensity Across the World

This figure plots the carbon dioxide (CO_2) emissions intensity of the United States and the World over time. CO_2 intensity measured in kilograms of CO_2 per $ of GDP (measured in international-$ in 2011 prices). Data from Global Carbon Project (Andrew and Peters, 2021) and Maddison Project Database 2020 (Bolt and van Zanden, 2020).

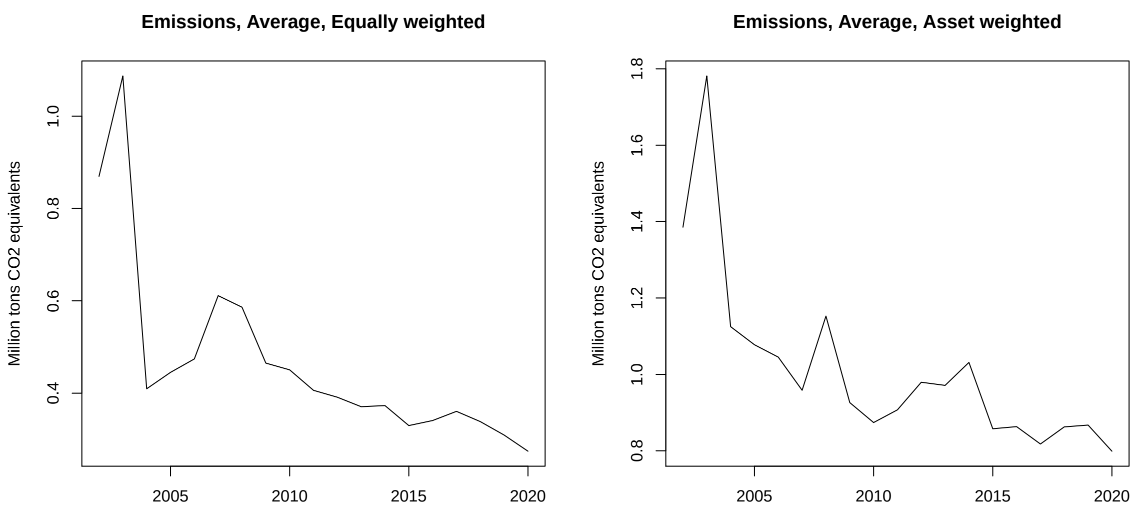

Figure D.2: Emissions over time

These figures plot the average carbon dioxide (CO_2) emissions across US firms over time. The left hand figure's average is computed using equal weights across firms, where as the right hand figure is computed using weights proportional to the assets of the firm.

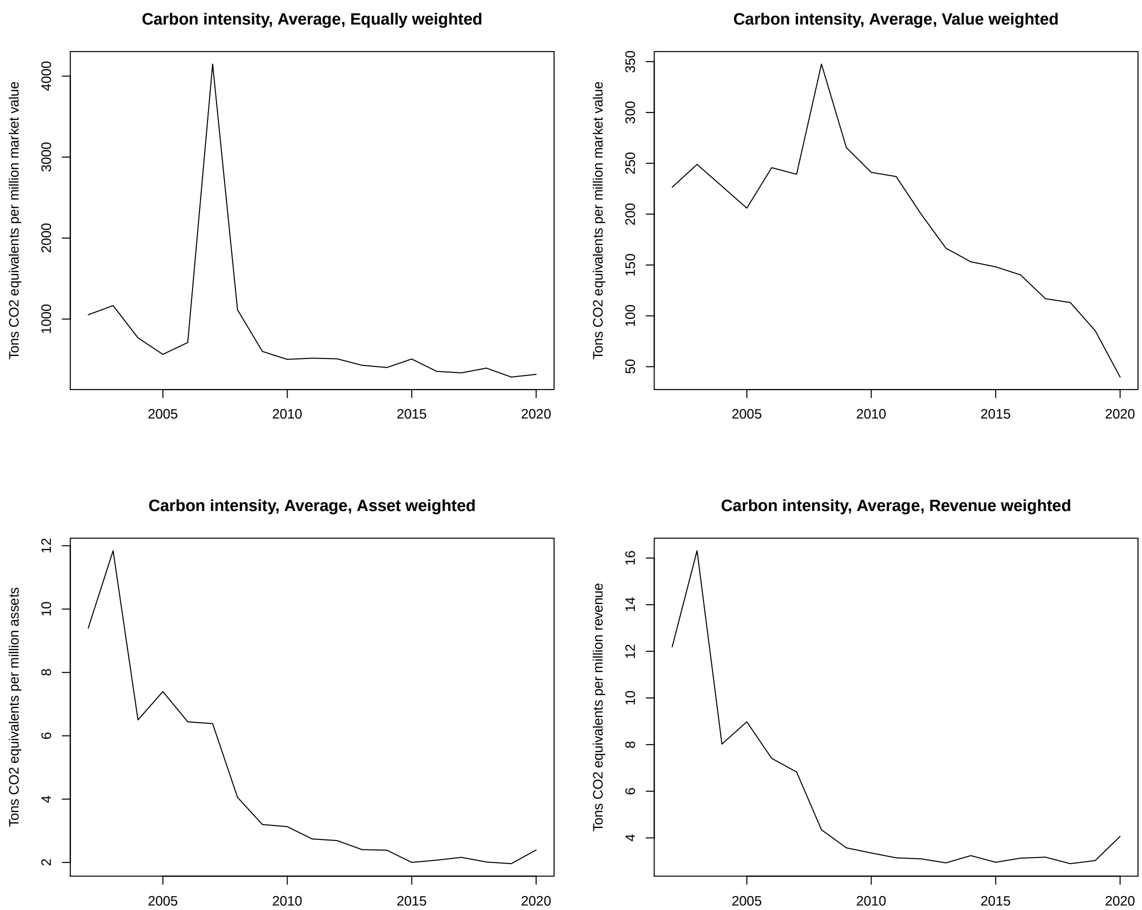

Figure D.3: Emission intensity over time

These figures plot the average carbon dioxide (CO_2) emissions intensity across american firms over time. The top left hand figure's average is computed using equal weights across firms, where as the top right hand figure is computed using weights proportional to the market value of the firm. The bottom left hand figure's average is computed using weights proportional to the asset size of the firm, and the bottom right hand figure uses weights proportional to the the firm's revenue.

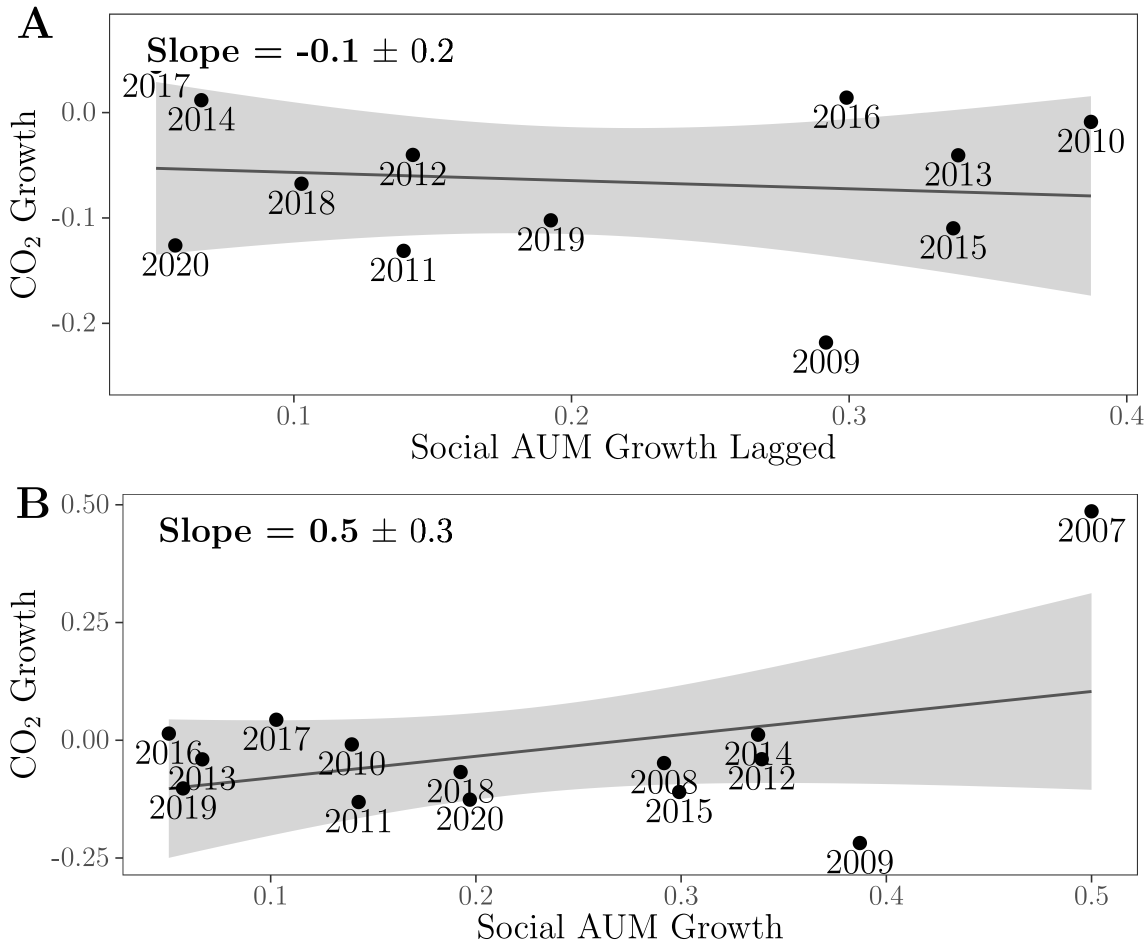

Figure D.4: CO_2 Emissions versus Socially Invested Capital

This figure plots the average emissions of firms over time as well as the amount of socially invested capital as reported by United Nations Principles of Responsible Investment. Panel A is for lagged social investment growth, and Panel B is for simultaneous growth. The CO2 emissions include direct CO_2 emissions. R^2 of Panel A is 0.02 and R^2 of Panel B is 0.15.

Figure D.5: Inclusion to Index Effect

This figure shows the inclusion to index effect in terms of changes in their emissions. Emissions change is compared to cross-sectional average by including time fixed effects. Band is a standard deviations away from the estimate. Standard errors clustered at firm-month.

Figure D.6: Inclusion to Index Effect: Saints versus Sinners

This figure shows the emissions of saint and sinner firms over time. Specifically, Panel A shows the average emissions per group per period. Panel B shows the period-by-period change in percentages for each group. Panel C shows the cumulative effect of emission changes in ratios. Band is 3 standard deviations away from the period estimate.

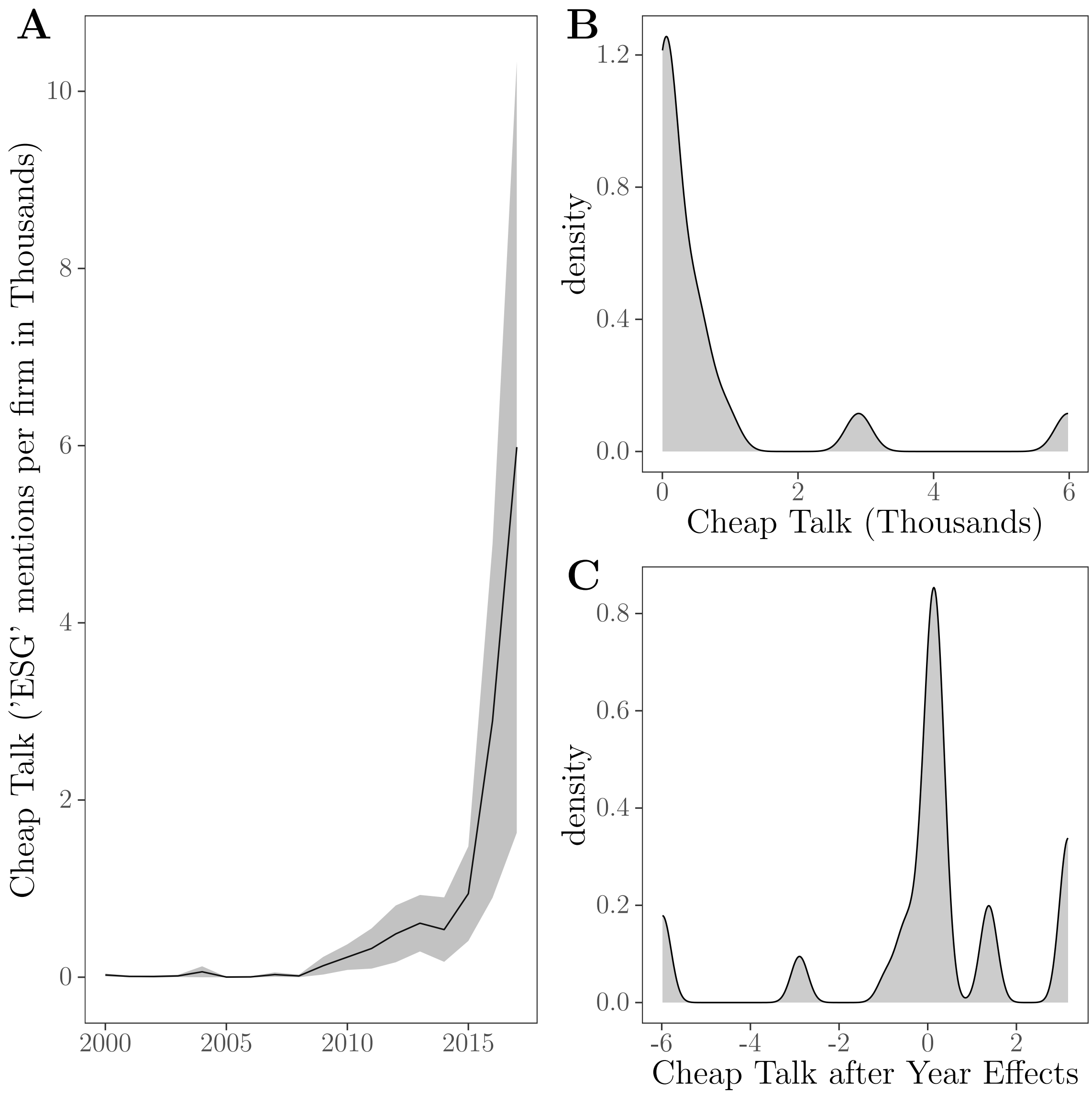

Figure D.7: Cheap Talk in Time Series and Cross-section

Cheap talk is measured as 'ESG' mentions per firm in Thousands. Data from their public reports to the securities exchange commission. In Panel A grey area represents two standard deviations variation.

Table D.1: Effect of being in index, Robustness with different specifications and models

Table shows how firms in the social index compares to firms outside of the index in terms of their emissions for different models. Column (1) and (2) repeats Column (3) and (4) in Table 3. The columns following add controls to Column (2). Specifically, Column (3) adds firm and time fixed effects, Columns (4), (5), and (6) add a dummy for the date of firm inclusion to the social index. Additionally, Column (5) adds time fixed effects and Column (6) has firm and time fixed effects. Emission changes are 1-year changes in CO_2 equivalents.

	Δ $CO_{2,t,t+1}$ (T Tons)	$\Delta \log CO_{2,t,t+1}$				
	(1)	(2)	(3)	(4)	(5)	(6)
Firm in Index	69**	2.3%**	1%**	2%**	3%′	1%**
	(3.60)	(6.50)	(2.72)	(6.39)	(0.02)	(0.01)
Firm just included				3%′	138%*	3%
				(1.33)	(0.77)	(0.02)
Constant	−86**	−2.93%**		−3%**	−4%**	
	(−6.25)	(−11.69)		(−11.69)	(0.01)	
FE	None	None	Firm+Time	None	Time	Firm+Time
Observations	37,056	37,056	37,056	37,056	352	37,056
R^2	0.0003	0.001	0.0002	0.001	0.01	0.0002
Note:						*p<0.2; **p<0.1; ***p<0.05

Table D.2: Effect of being in Index, Robustness to 1 to 3 years ahead

Table shows how firms in the social index compares to firms outside of the index in terms of their emissions across different horizons. Columns (1) to (6) are log changes and Columns (7) to (12) are absolute changes. For each group of two the first column is for a horizon of one year, second two years, and third three years. Within each group there is the first the ordinary result and then controlling for inclusions. Emission changes are 1-year changes in CO_2 equivalents.

	$\Delta \log CO_{2,t,t+1}$		$\Delta \log CO_{2,t,t+2}$		$\Delta \log CO_{2,t,t+3}$		$CO_{2,t,t+1}$ (T Tons)		$CO_{2,t,t+2}$ (T Tons)		$CO_{2,t,t+3}$ (T Tons)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Firm in Index	0.02**	0.02**	0.04**	0.04**	0.06**	0.06**	69**	70**	193**	195**	336**	335**
	(6.50)	(6.39)	(8.43)	(8.36)	(10.23)	(10.25)	(3.60)	(3.62)	(6.60)	(6.65)	(8.68)	(8.64)
Just Included		0.03′		0.02		−0.03		−61		−204		94
		(1.33)		(0.58)		(−0.68)		(−0.45)		(−0.97)		(0.30)
Constant	−0.03**	−0.03**	−0.06**	−0.06**	−0.09**	−0.09**	−86**	−86**	−218**	−218**	−352**	−352**
	(−11.69)	(−11.69)	(−17.11)	(−17.11)	(−21.63)	(−21.63)	(−6.25)	(−6.25)	(−10.63)	(−10.63)	(−13.23)	(−13.23)
Observations	37,056	37,056	32,832	32,832	28,812	28,812	37,056	37,056	32,832	32,832	28,812	28,812
R^2	0.001	0.001	0.002	0.002	0.004	0.004	0.0003	0.0004	0.001	0.001	0.003	0.003
Note:												*p<0.2; **p<0.1; ***p<0.05

Table D.3: Effects of Cheap Talk on ESG Scores, Granger Causality, Group of Words

Table considers how cheap talk by firms affects future ESG scores and impact across cheap talk as measured on different words. The cheap talk measure is frequency of the given word. Analyses take a Granger causality approach meaning they see if the dependent variable is predicted by lagged realisations of an independent variable in excess of its own lagged realisations.

	ESG Score Next Year			
	(1)	(2)	(3)	(4)
SUSTAINABILITY	0.016***	0.014**	0.093***	0.014**
	(0.005)	(0.005)	(0.011)	(0.005)
SUSTAINABLE	0.021***	0.010	0.063***	0.010
	(0.006)	(0.007)	(0.013)	(0.007)
ESG	0.004	−0.004	−0.023	−0.004
	(0.009)	(0.013)	(0.026)	(0.013)
ENVIRONMENT	0.004***	0.001	−0.004	0.001
	(0.001)	(0.001)	(0.003)	(0.001)
SOCIAL	0.005***	0.002	0.014***	0.001
	(0.001)	(0.002)	(0.003)	(0.002)
GOVERNANCE	0.001***	0.0005	0.002*	0.0004
	(0.0004)	(0.0005)	(0.001)	(0.0005)
CLIMATE	0.009***	0.002	0.011	0.004
	(0.003)	(0.003)	(0.007)	(0.004)
ESG Score This Year	0.889***	0.821***		0.821***
	(0.003)	(0.004)		(0.004)
CO_2/Assets			−0.001***	−0.0003***
			(0.0001)	(0.00005)
Constant	✓	✓	✓	✓
Observations	23,297	12,598	12,598	12,598
R^2	0.827	0.764	0.012	0.764
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table D.4: Elasticities of Cheap Talk and Impact

Table considers how firms' cheap talk and real impact affects their one year ahead ESG or E scores. Cheap talk is the word frequency of 'Sustainability'.

	Dependent variable:				
	log ESG_{t+1}	log E_{t+1}	E_{t+1}		ESG_{t+1}
	(1)	(2)	(3)	(4)	(5)
$\log(CO_2/\text{Assets})_t$	0.003**	−0.017***			
	(0.001)	(0.002)			
$\log(\text{Cheap Talk})_t$	0.024***	0.029***			
	(0.003)	(0.004)			
$CO_{2,t}$			−0.00000***		
			(0.00000)		
CO_2/Assets_t				−0.003***	−0.001***
				(0.0004)	(0.0002)
Cheap Talk_t			0.073***	0.072***	0.084***
			(0.019)	(0.019)	(0.012)
Constant	4.090***	4.136***	62.567***	62.473***	62.516***
	(0.007)	(0.011)	(0.361)	(0.358)	(0.233)
Observations	1,505	1,505	1,509	1,505	1,505
R^2	0.056	0.062	0.045	0.046	0.036
Note:			*p<0.1; **p<0.05; ***p<0.01		