Green or Greenwashing?
How Manager and Investor Preferences Shape Firm Strategy.

Nathan Barrymore
Robert H. Smith School of Business, University of Maryland
College Park, Maryland, U.S.A.

This Version: December 6, 2021

Abstract

When faced with pressures from stakeholders, firms may respond by substantively changing their products and processes, or by symbolically responding without substantive change. In this paper, I theorize and empirically study how managers’ and investors’ preferences with regards to one such pressure – that for environmental and social (ESG) responsibility – causes firms to either make substantive changes that result in improved outcomes or to greenwash: adopt symbolic policies. Empirically, I find that managers’ ESG preferences, as proxied using their language on earnings calls, correlate with both ESG policies and outcomes. However, investors’ ESG preferences correlate with only policies and not outcomes, suggestive of greenwashing. These results are consistent with a principal-agent model in which managers greenwash due to information asymmetry. Finally, I show that my measure of greenwashing correlates with ESG ratings disagreement, providing practical insight for managers and investors. These results have strategic implications in situations in which investors must rely on third party enforcement to verify firms’ claims, as well as for the current effectiveness of investor-led stakeholder capitalism. Furthermore, this paper shows how symbolic management and decoupling can be caused by individual stakeholders’ preferences.

Keywords: Greenwashing; Symbolic Management; ESG; CSR; Top Management
INTRODUCTION

Firms can respond to normative pressures by substantively enacting change within the organization, or by symbolically adopting policies that do not result in tangible outcomes (Durand, Hawn, and Ioannou, 2019). Because substantive change can be difficult and costly, firms may symbolically adopt corporate governance structures, product portfolios, and ethical codes in response to pressures from customers, investors, regulatory bodies, social activists, and external evaluators (Fiss and Zajac, 2004; Kim and Jensen, 2011; Stevens et al., 2005). Even CEO dismissals can be symbolic responses that aren’t intended to substantively change firm outcomes (Gentry et al., 2021). Research has identified how institutional environments, activist pressure, firm capabilities, and issue salience affect a firm’s decision to substantively or symbolically respond to normative pressures (Bundy, Shropshire, and Buchholtz, 2013; Durand et al., 2019; Greenwood et al., 2010; McDonnell and King, 2013). I extend this work by turning attention toward how preferences of specific actors within firms – top managers and investors – shape whether firms respond substantively or symbolically to normative pressures.

I study the question of symbolic or substantive responses in the context of firms’ environmental and social actions. Customers, employees, retail investors and activists pressure firms to improve their environmental and social (ESG) impacts (Burbano, 2016; Dyck et al., 2019; Hartzmark and Sussman, 2019; McDonnell, King, and Soule, 2015). Managers and institutional investors have responded by stating their commitment to improve the ESG outcomes produced by the firms they manage and own (Chatterji and Toffel, 2018; Gond and Piani, 2013). However, assessing firms’ ESG outcomes, and thus whether firms are making substantive changes or simply taking symbolic stances, remains a challenge because of the potential for firms to greenwash, that is, symbolically adhere to normative pressures by presenting a
deceptively positive impression of their environmental and social outcomes (Allen, 2018; Berg, Kolbel, and Rigobon, 2020; Chatterji et al., 2016; Kim and Lyon, 2015; Lyon and Maxwell, 2011; Marquis and Qian, 2014; Terrachoice, 2010). This paper asks the question: which firms greenwash? Specifically, I how do managers’ and investors’ ESG preferences correlate with greenwashing, as measured by the gap between a firm’s ESG policies and its ESG outcomes.

This investigation into greenwashing is motivated by the claims and language of stakeholder capitalism emanating from CEOs and investment managers. One third of assets under management in the U.S. are invested in funds which state that they consider ESG information in their investment decisions, a dramatic increase over the past decade (Nason, 2020; Principles for Responsible Investment, 2019). Meanwhile, several hundred CEOs of the largest U.S. corporations have publicly stated that their firms have responsibilities to non-shareholding stakeholders, and these same CEOs are increasingly taking public stances on social issues (Business Roundtable, 2019; Chatterji and Toffel, 2018).

While firms claim they are doing good by society, environmental activists, employees and customers have accused firms of greenwashing (Lyon and Montgomery, 2015). Unlike financial accounting, ESG reporting over the past decade has largely been voluntary, non-standardized and unaudited. Though ESG ratings agencies and other organizations aim to provide unbiased assessments of firm’s overall ESG outcomes, their ratings are, out of necessity, largely based on these unaudited and self-reported data, as this self-reported data has historically been the bulk of the data available to investors (Peirce, 2019). As a result, ESG ratings are both noisy and potentially biased (Berg et al., 2021, 2020). This fertile ground for greenwashing prohibits investors from more accurately allocating capital according to their ESG preferences, provides poor signals to managers as to their actual ESG outcomes, and limits firms’ ability to choose
exchange partners (e.g., suppliers) that match their social or environmental preferences (Dai, Liang, and Ng, 2020). Furthermore, greenwashing threatens to impact any link present between ESG outcomes and financial performance because it may prohibit firms from credibly signaling their ESG outcomes to interested stakeholders. Firms with higher ESG outcomes have lower cost of capital (Henisz and McGlinch, 2019) and firms can reduce worker wage requirements through socially conscious messaging (Burbano, 2016). However, in situations where greenwashing is prevalent, creditors, employees, and other stakeholders will not be able to distinguish firms with positive outcomes from those which are simply greenwashing.

Thus, stakeholders are left with a series of questions around the impact of the ESG movement. Which firms are greenwashing? How do CEOs and top managers influence a firm’s greenwashing decision? Do investment companies claiming to have sustainable investments invest in firms with better ESG outcomes? I address these questions using formal theory and a correlational study.

Empirically, in a sample of the 1,500 largest public firms in the U.S. from 2008 through 2019, I find that greenwashing is associated with investor, and not manager, characteristics. In my sample, CEO preferences, as proxied by the ESG language they use on earnings calls, correlate positively with both ESG policies and ESG outcomes across firms. This is consistent with the idea that when CEOs have a preference for better ESG outcomes, their firms put in place substantive ESG policies that result in positive ESG outcomes. However, investors’ preferences, as proxied by both sustainable fund volumes and time horizons, correlate with firms’ ESG policies, but not with firms’ ESG outcomes. This ESG policies-outcomes gap is suggestive of greenwashing, consistent with the idea that firms with ESG-oriented investors
greenwash by putting in place symbolic ESG policies that do not result in improved ESG outcomes.

These results are consistent with a principal-agent model in which investors (the principals) are unable to align the incentives of firm managers (the agents) with their own preferences. Using a principal-agent framework, I model greenwashing and brownwashing1 as the misalignment between a firm’s ESG policies (what they say they do) and their ESG outcomes (what they actually do). The model assumes that managers have access to ESG outcome information but present potentially greenwashed ESG information to investors in the form of ESG policies on the books. Managers are constrained, however, by environmental activists that punish greenwashing. Under these conditions, managers increasingly greenwash as investors become more ESG-oriented, as I find in the empirical study.

Finally, my empirical measure of greenwashing correlates with ESG ratings disagreement (Berg et al., 2020). This finding both adds credence to the use of the ESG policies-outcomes gap as a measure of greenwashing and suggests that investors concerned about greenwashing may want to commit additional resources to scrutinize the ESG outcomes of firms with noisy ESG ratings before adding these firms to their ESG investments products.

Theoretically, these results add to the literature on symbolic management and decoupling (Meyer and Rowan, 1977) by identifying firm-specific sources of symbolic management: the normative preferences of managers and investors. While past research has identified how institutional and firm-level characteristics are correlated with firms’ decisions to adopt symbolic policies, the domain of environmental and social performance highlights the possibility that the

---

1 Firms are brownwashing when they present an unrepresentatively negative view of their ESG outcomes (Kim and Lyon, 2015). Firms may brownwash to, for example, avoid being perceived as hypocritical (Carlos and Lewis, 2018) or because investors perceive ESG outcomes as substitutes for financial returns (Kim and Lyon, 2015).
normative preferences of firm insiders such as managers may determine whether a firm substantively or symbolically responds to normative pressures from outside stakeholders.

This study also has implications for managers, investors, and civil society actors concerned with how stakeholder capitalism manifests in firm strategic decisions around ESG issues. For investors, these results point to potential biases in ESG data used over the past decade. While investment has flowed into firms with high ESG ratings by third party ratings agencies, these ratings have not correlated with ESG outcomes reported in the news. This suggests that investors must both consider new independently reported data when assessing firms and be weary of the potential for firms to greenwash their ESG policies.

For managers, these results are consistent with the idea that greenwashing has been a common strategy over the past decade. This suggests that, at a minimum, the average firm genuinely committed to ESG outcomes has not been able to credibly differentiate themselves from their greenwashing peers. Thus, managers wishing to differentiate their firms based on their ESG outcomes will need to find novel mechanisms to do so that are costly for greenwashing firms to imitate.

Most broadly, these results have implications for whether outside pressures lead firms to implement strategic change, particularly in the context of environmental and social concerns. If firms can greenwash and investors are unable distinguish the greenwashing firms from those with positive ESG outcomes, then pressure from investors’ capital allocation decisions alone cannot produce positive ESG outcomes for society. This study suggests that managers may be able to do so, but it is doubtful that stakeholders ought to rely solely on managers’ consciences to produce positive ESG outcomes for society. Thus, the results in the paper are consistent with the idea that alternative solutions, such as investors’ use of voice through shareholder proposals and
proxy votes to change management, or, more broadly, governmental action, may be more effective levers of change (Fancy, 2021; Flammer, 2015).

This paper proceeds as follows: I first synthesize existing theory on the drivers of greenwashing and how managers and investors impact ESG performance. I then present the formal model grounded in this existing theory. After presenting the model, I introduce my empirical setting, outline my measures, and show my results. The paper ends with a discussion.

THEORY

Firms can greenwash in a variety of ways, each stymieing investors and other stakeholders from accurately evaluating firms’ ESG outcomes. I define greenwashing as actions taken by managers in organizations that result in a view of the organization as being more environmental or socially responsible than justified given the environmental or social outcomes the firm produces. This definition of greenwashing follows recent calls to expand the study of greenwashing beyond only disclosure decisions (Bowen, 2014; Delmas and Burbano, 2011; Lyon and Montgomery, 2015). This definition also encompasses social as well as environmental outcomes – focusing more broadly than the term’s origin in environmental performance. Importantly, this definition of greenwashing can encompass intentional and unintentional actions by management. For example, well-intentioned ESG policies may which do not produce positive environmental or social outcomes would be considered greenwashing. As such, greenwashing can be a strategic choice, a failure of execution, or simply the result of circumstances outside of management’s control.

There are numerous examples of firms greenwashing, or implementing ESG policies which are not supported by or followed up with action. Delmas and Montes-Sancho (2010) find that late entrants to the U.S. department of energy’s Climate Change Program were likely to
symbolically cooperate in the program, but not reduce their carbon emissions. Ramus and Montiel (2005) find that companies adopt environmental policies due to isomorphic pressures, but firms in certain industries choose not to adopt the most impactful policies because they would have the greatest negative financial impacts. Environmental organizations point out that most environmental claims on products are in some way misleading (Terrachoice, 2010). For example, BP was sued for advertising environmentally friendly credentials, despite their core business remaining in oil extraction (Dempsey, Raval, and Times, 2019).

Past research has looked at institutional drivers (Marquis, Toffel, and Zhou, 2016), industry factors (Ramus and Montiel, 2005), and firm-specific (Kim and Lyon, 2015) drivers of greenwashing. I extend this work by considering individual actors’ preferences. The public stances by executives and investors, as well as research into how these actors shape ESG outcomes suggest that these actors may also shape greenwashing behavior. Thus, this study moves to a more micro level than past work and examines the effect of actors specific to the firm – investors and executives – on greenwashing.

For context, I outline the existing research on the institutional and industry level drivers of greenwashing below. Firms greenwash due to market, nonmarket, organizational, and individual psychological drivers (Delmas and Burbano, 2011). At the institutional level, the combination of stakeholder pressure and a weak regulatory environment encourages greenwashing (Delmas and Burbano, 2011; Delmas and Montes-Sancho, 2010). Scrutiny by governments and social norms around environmental outcomes both decrease greenwashing behavior (Marquis et al., 2016; Ramus and Montiel, 2005). Regulation, however, can also be insufficient. Kim and Lyon (2011) find that governmental threat of regulation, without government oversight, drives firms to greenwash as a tactic to preempt the threatened regulation.
Firm-level drivers of greenwashing include industry, organization size, (low) visibility, and growth. Kim & Lyon (2015) find that growing investor-owned utilities greenwash to gain support from stakeholders necessary for expansion, while firms with poor financial performance brownwash, possibly to avoid giving investors the impression that they are sacrificing financial value for environmental outcomes. Ramus and Montiel (2005) found that implementation of environmental policies was weaker for firms in the service sector, relative to manufacturing firms. Kim and Lyon (2011) found that large firms were more likely to greenwash. While this paper focuses on managers’ and investors’ impacts on greenwashing, informed by these results, I include financial controls such as industry and firm size in my empirical models.

**Managers and ESG Performance**

Research has found that managers personal ethical beliefs, personality characteristics and political affiliations all influence or correlate with ESG ratings, which are meant to capture ESG outcomes. Chin, Hambrick, and Treviño (2013) find that firms with more liberal CEOs are more likely to receive higher ESG from third parties. A similar stream of research finds that personality characteristics – narcissism and hubris – also predict a firm’s ESG rating (Petrenko et al., 2016; Tang, Mack, and Chen, 2018). More directly, Muller and Kolk (2010) find that, in a sample of auto part manufacturing firms in Mexico, managers’ personal commitment to ethics is the main driver of their firms’ social outcomes. While these studies all point toward managers’ personalities and ethical commitments driving ESG outcomes, they point to different motivations – ethical commitment, liberalism, narcissism, and hubris – for why some managers may pursue a greenwashing strategy (implementing symbolic policies), while others may be committed to improving ESG outcomes.
This research has largely theorized ESG performance as a single construct, which combines both ESG outcomes and ESG policies. Furthermore, these studies largely measure this construct through ESG ratings. For the past few decades, ESG ratings have been the best source of aggregate ESG performance measures, but recent work has found that ESG ratings have high disagreement across ratings agencies and contain significant biases (Berg et al., 2020; Chatterji et al., 2016). To approach greenwashing, I disaggregate my ESG construct into ESG outcomes and ESG policies and ask how managers influence each. Thus, the first questions this paper asks are:

*Question 1a: How do managers’ ESG preferences affect ESG policies?*

*Question 1b: How do managers’ ESG preferences affect ESG outcomes?*

**Investors and ESG Performance**

Examining investor heterogeneity is essential for understanding greenwashing. ESG investing has become a mainstream investing strategy (Principles for Responsible Investment, 2019), and there is evidence that the average investor prefers holding stocks that meet some level of social and environmental outcomes. Hartzmark and Sussman (2019) find that when Morningstar made their fund ESG ratings more visible, investment flowed into high-sustainability funds and out of low-sustainability funds. In the early 2000s, sin stocks, that is, companies in industries such as tobacco or oil, financially outperformed the market (Hong and Kacperczyk, 2009). These stocks are often excluded from ESG investments, suggesting that investors were willing to pay a price to not invest in negative ESG performers. Shin, Lee, and Bansal (2021) find that, on average, investors rewarded CEOs for poor ESG outcomes (as proxied by company-level ESG ratings) in the 1980s and 1990s, but that this relationship
reversed in the 2000s. Thus, there is evidence that investors on average have changed their view on how to evaluate ESG outcomes.

Moving beyond average effects, we have additional evidence that different investors have heterogeneous preferences for ESG outcomes. Institutional investors (Chen, Dong, and Lin, 2020), and specifically European institutional investors (Dyck et al., 2019) appear to value firms’ ESG outcomes. Furthermore, they impact firms’ ESG ratings through both choice (ownership) and voice (shareholder proposals and contact with management) (Chen et al., 2020). Long term investors also own firms with higher average ESG ratings (Starks, Venkat, and Zhu, 2017), and push firms to use integrated reporting, a type of reporting consistent with a commitment to ESG outcomes (Serafeim, 2015). However, U.S. hedge funds appear to view ESG investments as a wasteful expenditure which sacrifices firm financial value. Hedge funds target firms with higher ESG scores (DesJardine and Durand, 2020), and then reduce the ESG ratings of firms they acquire (DesJardine, Marti, and Durand, 2020). Consequently, firms targeted by activist funds have increases in employee injury rates and sickness after they have been targeted (Xia and Meyer-Doyle, 2021). Given this evidence, it is essential to discriminate a firms’ investors by their ESG preferences to determine their impact on the firms’ ESG investments and greenwashing.

We thus have evidence that (a) investors on average have preferences for holding firms with positive ESG outcomes, but that (b) ESG-preferences are heterogeneous across investors. However, distinguishing between investors’ effect on outcomes and policies has been limited due to both data availability (i.e., ESG ratings as discussed above), as well as the recency of the mass adoption of ESG investing. Thus, we do not have an integrated theory as to how the ESG
preferences or the ESG commitments of investors will lead to greenwashing behavior. This leads to my second set of questions:

*Question 2a*: How do investors’ ESG preferences affect ESG policies?

*Question 2b*: How do investors’ ESG preferences affect ESG outcomes?

**Environmental and Social Activists**

Environmental and social activists provide a limit to the degree to which firms can greenwash their ESG outcomes. These activists can damage a firm’s reputation and impact a firm’s financial position if they perceive the firm to be environmentally or socially irresponsible (Kim and Lyon, 2011; McDonnell and King, 2013; McDonnell et al., 2015). This activist-imposed cost to the firm may take the form of a boycott, leading to loss in sales or lower employee engagement. Baron and Diermeier (2007) develop a theoretical model in which activists threaten firms with high brand value and low ESG outcomes, and firms increase their ESG outcomes when threatened. Empirically, King and McDonnell (2012) find that activists target firms with high advertised ESG but low ESG outcomes. These studies all point to activists sanctioning firms based on the gap between their ESG policies and their ESG outcomes. Thus, my model of greenwashing should include an activist penalty increasing based on this ESG policies – ESG outcomes gap.

**Principal-Agent Models and Greenwashing**

To theorize as to how to investors and managers shape greenwashing, we must consider whether the agency problem at hand is one of managers overinvesting in ESG or underinvesting in ESG outcomes, relative to the investment the firms’ shareholders would prefer.² Benabou &

---

² I do not consider the important issue of social optimality in this paper. ESG performance may have positive externalities through, for example, the reduction in greenhouse gas emissions. If this is the case, it may be that all firms are underinvesting in ESG relative to the social optimum. However, if a firm’s investors perceive a tradeoff between personal financial returns in stock price and the firm making the socially optimal reduction in carbon
Tirole (2010) outline theoretical arguments in both directions. There is the potential that ESG investment are win-win scenarios which managers may not recognize unless pushed by investors (Flammer, Hong, and Minor, 2019) as well as the potential that managers invest in ESG programs to gain private nonpecuniary benefits, against the preferences of shareholders (Cheng, Hong, and Shue, 2014). Empirical findings also identify both effects.

Studies which find that firms overinvest in ESG generally conclude that this overinvestment is due to managers reaping private benefits at the expense of investor returns. Werbel and Carter (2002) find that a firm’s charity foundation giving is consistent with nonpecuniary benefits for the CEO, and not to the firm. Cheng et al (2013) find that increasing managerial ownership and financial alignment with the firm decreases firm CSR, suggesting managers overinvest in CSR in relation to what shareholders would prefer. Personality characteristics often characterized as negative also appear to contribute to ESG activities. Petrenko et al (2016) find that narcissistic CEOs invest more in CSR, a signal that CSR is a wasteful nonpecuniary CEO benefit for some firms. Tang et al (2020) finds that CEO narcissism and comparison to peer firms influences a CEO’s investment in ESG, suggesting that ESG outcomes provides nonpecuniary benefits for CEOs.

However, as Hart and Zingales (2017) argue, there are good reasons to believe that ESG outcomes are of interest to shareholders, and that managers should maximize shareholder welfare, not wealth. Maximizing shareholder welfare, they argue, means managers considering shareholders’ preferences over environmental and social outcomes. Following this logic, corporations have recently begun to contract around ESG outcomes. Flammer et al., (2019) find that adding CSR criteria in executive compensation increases firms’ long-term orientation and emissions reduction by the firm, they may prefer the former. Thus, when I discuss overinvesting (greenwashing) or underinvesting (brownwashing) in ESG, the reference point is always to investors’ preferences.
value, consistent with the idea that CSR contracting focuses executives’ attention in areas that provide value but are not executives’ default focus. Furthermore, this suggests that when investors can measure ESG outcomes well, firm financial value can be increased through incentivizing managerial attention to these issues. Detemple and Xing (2020) develop a formal model in which ESG effort by managers is inversely related to pay-for-performance incentives popularized in the 21st century. Thus, these pay-for-performance contracts may be creating their own agency problems if shareholders prefer that executives pursue some level of ESG outcomes, as the studies previously outlined (e.g., Hartzmark and Sussman, 2019) have shown. In fact, including ESG related targets has become more popular in the past 15 years, with around a third of S&P 500 firms using such incentives by 2012 (Flammer et al., 2019; Maas, 2018; Maas and Rosendaal, 2016). These contracts suggest that some investors perceive agency issues related to too little investment in ESG outcomes.

The agency problem in ESG outcomes is thus not unidirectional. Investors are heterogeneous in their preferences for ESG outcomes, as are executives. Research has identified agency issues in managers overinvesting in ESG projects relative to investor preferences, as well as underinvesting. Synthesizing these findings on managers, investors, and agency issues, an appropriate model of greenwashing will include (a) heterogenous preferences for ESG outcomes in both investors and managers, (b) ESG communications by the firm, which may or may not be accurate, and which are unverifiable to investors, and (c) the threat of attack by social or environmental activists when firms greenwash. Finally, the model must allow for the possibility of greenwashing and brownwashing (Kim and Lyon, 2015). These components lead me to the development of the following model.
THE MODEL

Consider a firm’s decision to self-report their ESG policies and invest in costly ESG outcomes (which may be difficult or impossible to verify). There are three actors: a manager, a shareholder, and an environmental activist\(^3\) (e.g., The Sierra Club). The manager invests \(c(x)\) into ESG programs to achieve ESG outcomes of level \(x\).\(^4\) The manager self-reports the firm’s ESG policies at level \(r\), which, for now, we will assume is also set by the manager (the agent for the firm). The manager’s preference for ESG is defined as \(\alpha\), and could be determined by the manager’s personal preferences for ESG investment (e.g., Chin et al., 2013) or by financial incentives to achieve ESG goals (e.g., Flammer et al., 2019). The shareholder’s preference for ESG is defined as \(\beta\), with, for example, European Institutional investors (Dyck et al., 2019) having a greater preference for ESG outcomes than U.S. hedge funds (DesJardine et al., 2020). Higher values of \(\alpha\) and \(\beta\) indicate a greater preference for ESG outcomes. The shareholder influences the manager’s choice of ESG policies and investment in outcomes through the use of voice (i.e., calling the manager), choice (capital allocation) and contracting around ESG issues in CEO compensation packages (Amel-Zadeh and Serafeim, 2018; Flammer et al., 2019). The environmental activist punishes the firm for self-reporting high levels of ESG policies while having low ESG outcomes, indicated by the cost function \(f(r - x)\) (Baron and Diermeier, 2007; King and McDonnell, 2012). This activist penalty increases with the concentration of activist oversight of the firm, indicated by \(\delta\).

The manager determines the firm’s ESG policies and outcomes considering their own preferences as well as the preferences of the shareholder and activist. The manager maximizes

---

\(^3\) This could be an environmental or social activist. I use “environmental activist,” or “activist” for brevity.

\(^4\) Greenwashing may also be unintentional due to an ESG policy producing a worse than expected outcome. For simplicity, I assume ESG investments produce known ESG outcome \(x\), rather than noisy ESG outcome \(x + \varepsilon\).
the following expression, which I call the manager’s utility ($u^F: u^F = \alpha x + \beta r - c(x) - \delta f(r - x)$)

The terms in the utility function are defined in Table 1, along with a column listing an example application of a utility firm.

[Insert Table 1 here]

**Results when Investors Trust Self-Reported ESG Policies**

I will always assume that the manager selects the firm’s ESG investment, and thus determines the firm’s ESG outcomes, $x$. For now, suppose the shareholder is trusting such that the manager selects the firm’s ESG policies, and the shareholder believes the firm’s level of ESG outcomes are equal to their level of self-reported ESG policies. Specifically, they believe (potentially incorrectly) that $x = r$. The game proceeds as follows.

1. The manager, $F$, observes her type $\alpha$ (the shareholder’s type, $\beta$, is public knowledge).

   The manager then selects a level of ESG investment which produces ESG outcomes $x$, and selects a level of self-reported ESG policies, $r$.

2. The firm’s utility is determined based on $\{x, r\}$.

   Under this initial assumption the manager sets the value of both $x$ and $r$, the manager is the only actor, and all given values and functions are common knowledge: $\{\alpha, \beta, \delta, c(\cdot), f(\cdot)\}$. I assume the cost and penalty functions are both increasing and convex in their arguments. The equilibrium under these assumptions is simply the solution to the following maximization problem:

   $$\max_{x,r} u^F = \max_{x,r} [\alpha x + \beta r - c(x) - \delta f(r - x)]$$

   The complete proofs for all equilibria (along with all function and parameter assumptions) are in the Appendix. For now, I simply present an overview of the results of the model. The manager’s optimal choice of ESG policies, $r^*$, and ESG outcomes, $x^*$, are as follows:
\[ x^* = c_x^{-1}(\alpha + \beta) \]
\[ r^* = x^* + f_g^{-1}(\beta) \]

The manager faces incentives to increase ESG outcomes up to the point where the marginal benefits (from both the \( \alpha x \) term and a reduction in the penalty function) are equal to the marginal costs (from the cost function). They also have the incentive to set ESG policies at the point where the marginal benefits from ESG policies (from the \( \beta r \) term) are equal to the marginal costs of ESG policies (from the increased activist penalty). Depending on the relative importance of the manager’s own preferences, the shareholder’s preferences, and the activist’s penalty, the manager will want to understate, accurately report, or overstate the firm’s ESG. If the penalty from activists is high relative to the benefits of self-reporting a high level of ESG policies, a firm will brownwash: underreport their ESG (see Carlos and Lewis, 2018 or Kim and Lyon, 2015 for examples). Alternatively, if benefits to self-reporting a high level of ESG policies are high relative to the penalty from activists, the firm will greenwash.

Under this trusting shareholder assumption, as the manager’s and investor’s ESG preferences change, the firm’s ESG policies and ESG outcomes change in the following ways:

<table>
<thead>
<tr>
<th>Increasing Manager’s ESG Preferences ...</th>
<th>Increasing Shareholder’s ESG Preferences ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \frac{\partial x^*}{\partial \alpha} = \frac{\partial [c_x^{-1}(x)]}{\partial \alpha} &gt; 0 ]</td>
<td>[ \frac{\partial x^<em>}{\partial \beta} = \frac{\partial [c_x^{-1}(x)]}{\partial \beta} = \frac{\partial [c_x^{-1}(x)]}{\partial \alpha} = \frac{\partial x^</em>}{\partial \alpha} &gt; 0 ]</td>
</tr>
<tr>
<td>[ \frac{\partial r^<em>}{\partial \alpha} = \frac{\partial x^</em>}{\partial \alpha} &gt; 0 ]</td>
<td>[ \frac{\partial r^<em>}{\partial \beta} = \frac{\partial x^</em>}{\partial \beta} + \frac{\partial [f_g^{-1}(\beta)]}{\partial \beta} &gt; \frac{\partial x^*}{\partial \beta} &gt; 0 ]</td>
</tr>
</tbody>
</table>

These results mean that the higher the manager’s preferences for ESG (\( \alpha \)), the higher the firm’s ESG policies (\( r \)) and ESG outcomes (\( x \)). In fact, as the manager’s ESG preferences change, the firm’s ESG outcomes change in the following ways:

\[ \frac{\partial x^*}{\partial \alpha} = \frac{\partial [c_x^{-1}(x)]}{\partial \alpha} > 0 \]
\[ \frac{\partial r^*}{\partial \alpha} = \frac{\partial x^*}{\partial \alpha} > 0 \]
\[ \frac{\partial r^*}{\partial \beta} = \frac{\partial x^*}{\partial \beta} + \frac{\partial [f_g^{-1}(\beta)]}{\partial \beta} > \frac{\partial x^*}{\partial \beta} > 0 \]

Here, I use overstate to mean they report higher policies than they achieve outcomes, that is choose \( r | r > x \).
increase, the firm’s ESG policies and ESG outcomes will increase at the same rate. As the shareholder’s preferences ($\beta$) for ESG increases, again both ESG policies and ESG outcomes increase. However, unlike with manager preferences, when investor ESG preferences increase, the firm increases their ESG policies ($r$) more than they increase their ESG outcomes ($x$), leading to greenwashing.

The preceding solution assumes that shareholders respond to a firm’s self-reported ESG policies as if they reflect the firm’s ESG outcomes. This could be because investors trust firms’ self-reported ESG policies, or because they face institutional pressures to do something, lack more credible information, and think their eventual customers won’t notice the difference. This latter point would suggest a different foundational source of greenwashing.

Taking an equilibrium approach would require the model to account for shareholders rationally discounting firms’ self-reported ESG policies. This equilibrium model would provide an alternative set of relationships between stakeholder preferences and ESG outcomes and self-reporting. Thus, this model framework provides two possible sets of predictions: one in which shareholders trust firms ESG policies, and another in which they rationally discount ESG policies as cheap talk.

Do investors use firms’ self-disclosed ESG policies, or do they rationally discount self-disclosed information as cheap talk? Recent academic literature and articles in the popular press suggest that investors are forced to use ratings based significantly on self-disclosed data due to high demand for such products and the lack of better options. While investors are rationally concerned about greenwashing (Ramanna, 2021), the rise in the quantification of ESG metrics has resulted in an information environment where investors are left with little option other than relying on largely voluntary, non-standardized and unaudited ESG disclosures (Peirce, 2019).
Investors rely on ESG ratings in making investment decisions (Amel-Zadeh and Serafeim, 2018), and yet they say their top concern about this data they use is greenwashing (Carlson, 2021). Furthermore, there are known issues with ESG ratings such as divergence (Berg et al., 2020; Chatterji et al., 2016) and the impact of firm resources on ESG scores (Drempetic, Klein, and Zwergel, 2020).

The non-equilibrium predictions presented above, in which shareholders trust ESG policies, are more consistent with this empirical evidence and with testimonials in the popular press. Additionally, they are consistent with the results I present below in the empirical section. As such, I present this trusting shareholder version of the model as my inference to the best explanation and relegate the equilibrium solution to the model to the appendix.

Given this inference to the best explanation, the propositions from the trusting outside stakeholder version of the model are formally stated as follows:

**Proposition 1:** As managers’ ESG preferences increase, firms’ ESG policies will increase at the same rate as their ESG outcomes.

**Proposition 2:** As shareholders’ ESG preferences increase, firms’ ESG policies will increase more quickly than their ESG outcomes.

**EMPIRICAL APPLICATION**

**Data and Sample**

I examine the question of how top management and investors’ ESG stances predict ESG policies and ESG outcomes using a large sample of public U.S. firms. I gather data on the 1,500 largest U.S. public firms from 2008 to 2019. I source ESG data from TruValue (now owned by FactSet), Thomson Reuters, MSCI, Sustainalytics and RobecoSAM, earnings call transcripts
from Capital IQ, investors holdings from Morningstar and 13F filings accessed through WRDS, and firm financial and industry membership data from Compustat.

**Dependent Variables**

My outcome variables are firms’ ESG policies and firms’ ESG outcomes. To measure ESG policies, I use individual datapoints from the Thomson Reuters ESG dataset (formerly ASSET4). This dataset includes the individual metrics used in developing firms’ overall ESG scores, which are unavailable from other ESG data providers. These policies include, for example, “Water Reduction Program” and are rated as either 0 (no policy) or 1 (firm has a policy). I generate a total policies variable as the total count of policies.

I measure ESG outcomes from news media coverage gathered from TruValue Labs, an ESG data provider. TruValue Labs machine reads articles from tens of thousands of news media sources daily. Their algorithm then uses natural language processing to identify articles covering firms, categorizes these articles by ESG issue and scores the articles according to their sentiment on a 0-100 scale. For example, an article reporting on a company being accused of manipulating test results to hide dangerous emissions from their plant will be associated with the firm, categorized as having negative sentiment and categorized under “Waste & Hazardous Material Management.” Alternatively, an article describing how a firm has announced that they will produce only LEED certified building will be rated as having positive sentiment and categorized under “Energy Management.”

TruValue generates firm-day level observations based on each article. To capture a firm’s overall media coverage at a point in time, these firm-day level observations are aggregated into a firm “insight score” using an exponentially weighted moving average of all articles about the firm. I use this “insight score” data point as the TruValue ESG Score for a firm. I then match the
TruValue ESG Score to financial and other ratings data by choosing the last day of the firm’s fiscal year, rounded to the nearest month end. The final TruValue measure captures ESG events associated with the firm which are independently reported by the news media. Higher values of this measure indicate more positive coverage, while lower values of the TruValue measure indicate more negative coverage.

**Independent Variables**

To find an empirical proxy for investors and managers’ ESG preferences, I turn to their public ESG stances. For investors I use the descriptions of their investment fund offerings, and for managers I use their statements to investors in earnings calls.

*Investors’ Stated ESG Preferences*

To measure a firm’s investors’ stated ESG preferences, I measure the proportion of their institutional shareholders’ funds labeled as sustainable. I first develop an ESG measure at the institutional investor level. To do so, I gather all funds listed as sustainable according to their prospectus from Morningstar. I merge this data with 13f filings to calculate an institutional investor-year level measure of total assets under management (AUM) and sustainable AUM. I then calculate, at the institutional investor-year level, a measure of ESG orientation by dividing sustainable AUM by total AUM. Using 13f filings to identify each firm’s investors, I then calculate, at the firm-year level, a weighted average of all the firm’s investors’ ESG orientations.

Previous research has found that long term investors have higher ESG preferences than short term investors (Oikonomou, Yin, and Zhao, 2020). As a robustness check, I calculate a measure of investor time horizon as an alternative measure of a firm’s investors’ ESG preferences. I follow past research and calculate a firm’s investors’ time horizon as the proportion of long term

---

6 All institutional investors with over $100M in assets must file 13f reports indicating how their assets are invested.
(dedicated) investors minus short term (transient) investors using the classification scheme from Bushee (2001).

*Top Management’s ESG Orientation*

To develop an empirical proxy for the CEO and top management team’s ESG preferences, I examine the language they use on earnings conference calls, following past work in finance (Brochet, Loumioti, and Serafeim, 2015; Loughran and McDonald, 2010; Yang, UY, and Huang, 2020) and emerging work around ESG issues (Smeuninx, De Clerck, and Aerts, 2020).

I measure the volume and sentiment of ESG-related language managers use on earnings calls and use this as a proxy for their ESG preferences. I identify ESG-related text in earnings calls using two methods: semantic distance and word counts. My main measure of ESG language is a novel measure of the semantic distance between executives’ language and SASB sustainability standards. I calculate this semantic distance by converting earnings call paragraphs into vectors of real numbers using MPNet and Sentence Transformers (Song et al., 2020), and then calculating the cosine similarity between each paragraph of earnings call text’s vector representation and each of the 26 SASB standards’ vector representations. This method identifies paragraphs that are, for example, close to the phrase “Communities in which we operate and impacts on core human rights and the treatment of indigenous peoples” or “Direct (Scope 1) greenhouse gas (GHG) emissions that a company generates.” By using this large scale pre-trained model and sentence embeddings (vector representations), the algorithm takes advantage of the dense information in natural language, including both the semantic meaning of individual words as well as their position in sentences and paragraphs. This allows the algorithm to identify phrases and paragraphs with similar meanings, rather than finding only matching words or synonyms.
As a robustness check I use a word count measure, which comes with the advantage of being easily interpretable, but the disadvantage of being a rough measure that doesn’t take advantage of the nuance or information present in language. I manually develop a list of ESG-related words by reading dozens of ESG reports. I then calculate, at the paragraph level, the proportion of ESG-words in the paragraph. This proportion, aggregated to the call level, is the “ESG volume” measure. I then, again at the paragraph level, calculate a sentiment score using TextBlob, a widely used natural language processing package. To calculate an ESG-sentiment score, I multiply, at the paragraph level, the ESG volume measure and the sentiment measures and then sum this score across all paragraphs spoken by executives. For all ESG language measures, I generate separate measures for the presentation and Q&A sections.

Control Variables

I control for firm and industry characteristics using financial variables from Compustat. I control for the number of employees, total revenue, and total assets in regression models. I also control for sector effects by using 2-digit SIC sector values. Finally, I use year fixed effects to control for the increasing salience of ESG issues over my sample timeframe.

Descriptive Statistics

I start with the sample of the largest 1,500 firms in the U.S. from 2008 to 2019. My sample is then restricted to firms covered in the TruValue dataset and firms for which I can gather earnings calls. Investor data from Morningstar and 13f filings does not significantly reduce my sample size. Merging in Thomson Reuters ESG data restricts the sample to 5,486 firm-year observations and skews the data toward recent years. Long term shareholding data, which I use in a robustness check, drops a few additional firm-year observations. Given this year skew of the sample, my results are thus largely driven by recent years (2015-2019). The sample is outlined in Table 2.
Summary statistics for my key variables are outlined in Table 3. Firms in the sample have between 0 and 67 ESG related policies, with an average of 21 and a median of 16 policies. My measure of outcomes, news sentiment from TruValue, has a mean value of 61.6 and a median value of 63.3, suggesting that the average ESG related news article in the sample has a slight positive sentiment (as a score of 50 represents neutral sentiment). Earnings call ESG focus measures have a slight positive skew (and are normed in the empirical models). Firms’ investors have, on average, around 1% of their AUM labeled as sustainable.7 In unreported analysis, I find that both ESG mentions on earnings calls and investor ESG percentages are increasing greatly in later years of the sample. As a robustness check, I use shareholder time horizon. Long Term Shareholders is defined as the difference between the percentage of a firms shares held by dedicated shareholders and transient shareholders. On average, firms have more transient shareholders than dedicated shareholders.

Environmental and social issues vary significantly across sectors and have become more salient for investors and managers over the past decade. To identify the correlations between my ESG variables of interest, I normalize all ESG variables within sector-year for my empirical analysis. In the sections that follow, all key ESG variables (policies, outcomes, language on earnings calls, and investor ESG orientation and time horizon) are all normalized at the sector-year level. For example, the normalized ESG policy measure is calculated as

\[
\text{Normalized Policies} = \frac{\text{Policies} - \text{Policies}_{jt}}{\sigma_{\text{Policies}_{jt}}}, \text{ where } \text{Policies}_{jt} \text{ is the average number of policies in }
\]

7 While one third of assets under management in the United States claim to consider ESG information when making asset allocation decisions, funds explicitly labeled as sustainable in their prospectus make up a much smaller portion of total assets under management.
the focal sector \((j)\) and year \((t)\), and \(\sigma_{\text{policies}_{jt}}\) is the standard deviation of the number of policies firms have in the focal sector-year. Summary statistics for these normalized variables are in the appendix, and, by construction, all resemble a normal distribution.

The earnings calls ESG measures used here are novel. In summary, managers do talk about ESG on earnings calls, examples are provided in the appendix. ESG discussions experienced a short peak in 2008 during the financial crisis, and, after falling quickly, have been increasing over the 2011-2019 period, especially since 2018. When, and how much management discusses ESG issues appears to be reflective of a firm’s ESG reputation, at least among firms with strong ESG reputations (both positive and negative). Unilever, a firm with a very high ESG reputation, and Exxon Mobile, a firm with a very low ESG reputation, provide a key contrast. In their Q3 2020 earnings call, the Unilever CEO explicitly states their stakeholder-oriented framework in the opening paragraphs of the call. Alternatively, Exxon Mobile executives do not discuss ESG issues at all in their presentation, and only touch on renewable energy sources when asked by an analyst in the Q&A section in their Q3 2020 call. These differences are not just driven by industry. In BP’s Q3 2020 earnings call, executives discuss stakeholders and renewable resources over multiple paragraphs of the presentation section of the call. For further summary statistics, see the appendix.

**RESULTS**

Correlations between all key variables are presented in Table 4. Of note, the number of ESG policies a firm has correlates positively, but very weakly, with ESG outcomes in the news. The raw correlation between the two variables is 0.02. ESG language on earnings calls correlates with both ESG policies and ESG outcomes. Investor ESG orientation, on the other hand, correlates positively with ESG policies but have a very small and negative correlation with ESG
outcomes. These correlations summarize the results told more precisely through regression analysis and visually through binned scatterplots in the remainder of the results section.

[Insert Table 4 here]

The main regression results showing the correlation between managers’ and investors’ ESG preferences, controlling for financial characteristics, are found in Table 5 through Table 7. I first discuss the results related to managers’ ESG language, and then turn to investor characteristics.

**Managers**

The relationship (with no controls) between managers’ ESG language and ESG policies and outcomes is shown visually in the binned scatterplot in Figure 1. From this raw correlation, we can see that firm-years with greater discussion of ESG issues on earnings calls have more ESG policies on the books, and better ESG outcomes as reported by news media.

[Insert Figure 1 here]

To examine whether these relationships are accounted for by alternative drivers such as firm financial characteristics or yearly trends, Table 5 regresses ESG policies and outcomes on both managers’ ESG language and investors’ ESG investments. Managers’ ESG language is measured by calculating the semantic distance between managers’ language and SASB sustainability standards. In models (1) and (2) of Table 5, the dependent variable is ESG policies on the books, while in models (3) and (4) the dependent variable is ESG outcomes. Models (1) and (3) include sector and year fixed effects, while models (2) and (4) include firm and year fixed effects. The independent variables and control variables are lagged by one year in all models.\(^8\)

The coefficient on Managers’ ESG Language is positive in all models, but not always statistically significant. The coefficient on Managers’ ESG Language has t-statistics of 2.66 and

---

\(^8\) In unreported results, I find that all results are substantively the same when using concurrent independent variables, or concurrent independent and control variables.
3.02 in models (1) and (3), respectively. This means that across firms in the same industry, managers’ ESG language is predictive of more ESG policies and better ESG outcomes. All key variables are normalized thus these results indicate that, across firms in the same sector, an increase of ESG language on earnings calls by standard deviation is associated with a 0.03 standard deviation increase in policies on the books and a 0.05 standard deviation increase in ESG outcomes.

In firm fixed-effects models, results for managers’ language are consistent directionally, but the point estimates in these models are around half to one third the size of the sector-fixed effects models and have confidence intervals which overlap with zero. Thus, these results suggest that managers’ ESG language is predictive of ESG outcomes in the news across firms within the same industry but does not provide evidence that these effects are present within firms over time.

[Insert Table 5 Here]

To examine the robustness of measuring managers’ language using a sentence embedding model, and for greater interpretability, I examine the results using an ESG word frequency measure. These results are displayed in Table 6. As in table 5, the dependent variable in the first two columns is ESG polices, and in the last two columns is outcomes. Sector fixed effects are used in columns (1) and (3), and firm fixed effects are used in columns (2) and (4). The ESG word count measure is calculated as the proportion of ESG related words in earnings calls and replace the semantic distance measure from Table 5. The results from this model are consistent with the results using the SASB semantic distance measure, and slightly stronger. In the sector fixed effects models, a one standard deviation increase in managers’ use of ESG words is associated with a 0.04 standard deviation increase in ESG policies on the books and a 0.08
standard deviation increase in ESG outcomes the following year. In the firm fixed effects model, the coefficient on managers’ ESG language is positive and not statistically significant.

[Insert Table 6 Here]

**Shareholders**

Turning to shareholders, these results show a clear distinction between the correlation between shareholder ESG preferences and ESG policies, and between shareholders and ESG outcomes. These relationships are presented visually in the binned scatterplot in Figure 2. The binned scatterplot shows that, across firms, those firms with investors who have more assets under management in sustainable-labeled fund have more ESG policies on the books, but there does not appear to be a clear relationship between ESG funds and ESG outcomes.

[Insert Figure 2 here]

Turning to regression results in Table 5, the coefficient on Investors’ ESG Orientation in columns (1) and (2) of are both positive, with t-statistics of 3.45 and 2.48, respectively. These coefficients indicate that, across firms in the same industry, a one standard deviation increase in a firms’ investors’ ESG orientation is associated with a 0.035 standard deviation increase in their ESG policies on the books. Within firms over time, a one standard deviation increase in investor ESG orientation is associated with a 0.02 standard deviation increase in ESG policies.

The correlation between investor ESG orientation and ESG outcomes is both smaller in economic terms, relative to the effect on policies, and not statistically different than zero. The coefficients on Investors’ ESG Orientation in columns (3) and (4) have t-statistics of 1.32 and 0.25, respectively. These results provide evidence that increases in a firm’s investors’ ESG preference is correlated with an increase in ESG policies, but does not provide evidence that ESG outcomes change with investors’ ESG preference.
Previous research has found that long term investors prefer firms with high ESG ratings, and push firms to improve their ESG ratings. Following this research, I use ESG time horizon (the proportion of dedicated investors minus the proportion of transient investors according to the classification system developed in Bushee, 2001) as a robustness check of the results from my measure of investor ESG orientation in Table 7. This alternate measure has the advantage of being established in the literature, but the disadvantage of being a less direct measure of ESG orientation. Table 7 is set up just as Table 5, but uses investor time horizon rather than investors’ ESG funds as the key independent variable for investors. Column (1) shows that a one standard deviation increase in long term investors is correlated with a 0.036 standard deviation increase in ESG policies on the books in the sector fixed effects model. Within firms over time, long term shareholders are associated with lower ESG policies, but the coefficient has a 95% confidence interval overlapping with zero (t-stat of -1.69). With regards to outcomes, a one standard deviation increase in long term investors is correlated with a 0.0385 standard deviation decrease in ESG outcomes in the sector fixed effects model (t-stat of -2.55) and with a 0.0344 decrease in ESG outcomes in the firm fixed effects model (t-stat of -1.84).

In aggregate, these results suggest that the ESG language managers use in earnings calls correlates with both increased ESG policies and more positive ESG outcomes across firms, but not significantly within firms. However, investor ESG orientation is associated with increased ESG policies, but appears to have little or no association with ESG outcomes, and long-term shareholders may even have a negative association with ESG outcomes.

Ratings Disagreement
A practical implication of greenwashing is that it increases the cost for investors and stakeholders to accurately assess firms’ environmental and social outcomes, perhaps prohibitively so. This could be reflected in the fact that there is high disagreement between ESG ratings. Chatterji et al., (2016) found that as of the mid 2010s, individual companies’ ESG ratings are highly variable across ratings agencies, with inter-rater correlations of around 0.5, in comparison to financial ratings inter-rater correlations of around 0.97. Berg et al., (2020) find the same pattern still exists through 2019, with the source of divergence split between scope disagreement (i.e., what behaviors and outcomes are ESG ratings meant to capture) and measurement error. Practitioners have also questioned our current ability to assess firms’ ESG outcomes, citing both the disagreement over methodology and data (Mackintosh, 2018; Peirce, 2019). Researchers and practitioners are thus left with the question of what ESG information can be valuable, and what drives firms to potentially misrepresent their ESG outcomes.

As both a robustness check of the validity of using ESG outcomes reported in the news and ESG policies to assess greenwashing behavior, and as a practical insight for researchers and practitioners, I next examine the relationship between the ESG policy-outcome gap and ESG ratings disagreement. I define greenwashing empirically in this section as the number of ESG policies a firm has, normed at the sector-year level, minus the TruValue measure of ESG outcomes, again normed at the sector-year level for comparability. I calculate ratings disagreement as the variance between three large ESG ratings providers: MSCI, Sustainalytics and RobecoSAM. Due to data access, I have ESG ratings disagreement for only two years: 2018 and 2019, thus I present results only with sector fixed effects, as firm fixed effects would only capture one year of change. Results are presented visually in Figure 3. There is a positive

---

9 I exclude Thomson Reuters because I gather my policies measure using Thomson Reuters data, but these results all hold if the Thomson Reuters ESG score is included in the ratings variance measure.
relationship between greenwashing and ratings disagreement. Regression analysis is presented in Table 8, and confirms that this relationship is present across firms in the same industry.

[Insert Figure 3 and Table 8 here]

**DISCUSSION**

This paper presents a theory of how a firm’s stakeholders shape the firm’s use of symbolic policies. Specifically, in the context of social and environmental performance, I examine how managers and investors shape greenwashing. Using a principal-agent framework, I hypothesize that firms greenwash in response to investor pressure around ESG issues when investors do not have sufficient information to verify managers ESG claims, and when managers are not ESG-oriented.

Empirically, I examine whether the ESG preferences of management and investors predict ESG outcomes and policies and find correlational evidence consistent with the principal-agent framework. Over the past decade, managers’ ESG preferences, as proxied by their ESG language in earnings calls, correlates with ESG policies and ESG outcomes. However, a firm’s investors’ ESG preferences correlate strongly with ESG policies and the books, but not with outcomes, suggestive of greenwashing. These results are robust to alternative measures of managers and investor preferences, including investor time horizons and multiple methods of measuring managers’ language.

This study has theoretical implications to multiple streams of research. First, this study suggests that individual and organizational preferences, especially those of top management, around ESG issues are driving factors in firms’ ESG strategies. This promotes that idea that looking inside the firm can provide insight into how firms can benefit society. However, this study also provides a note of caution for ESG research. Significant ESG research has focused on
the relationship between firms’ ESG (or CSR) performance and financial performance. This study suggests that if the measures the research community uses to proxy for ESG performance are at best noisy and at worst biased due to greenwashing, the true relationship between social and financial performance may be obscured when using historical ESG ratings.

More broadly, this paper contributes to sociological research on decoupling (DiMaggio and Powell, 1983; Meyer and Rowan, 1977). The empirical patterns observed here are consistent with organizations decoupling policy from practice to maintain external legitimacy while maintaining internal efficiency. Adding to past research, this paper shows that individual preferences, beyond industry and financial factors, may also affect the decoupling decision. Lastly, this paper suggests that agency problems around firms’ ESG strategy could result in either overinvestment or underinvestment in ESG, relative to investors’ preferences. Research has presented conflicting views of the directionality of the agency problem in ESG (Cheng et al., 2014; Flammer et al., 2019). This paper provides a model in which both are possible, depending on the relative ESG preferences of managers and investors.

This study contributes novel data measures and techniques to ESG research. I machine read earnings call transcripts and semantic distance and ESG word frequency to measure managers’ ESG orientation. Semantic distance specifically can be used to generate insight across a broad array of strategic management questions, by providing an automated method for identifying what managers are talking about that can use the dense meaning of text across large corpuses. I also develop a new measure of greenwashing by comparing ESG policies to ESG outcomes reported in the news. This new method covers a larger set of firms and a wider set of ESG concerns than past research into greenwashing.
The sample, the correlational results, and the empirical proxies used in this study provide reasons to extend this work to identify causality and to consider the boundary conditions in interpreting this study. Greenwashing and the investor ESG movement are currently changing significantly in terms of scope and regulatory environment. Thus, it is important to note that this study examines the behavior of firms in the United States between 2008 and 2019. The results presented here provide insight into this time and place and could change dramatically across jurisdictions and timeframes in which the reporting standards and culture around ESG issues vary. For example, ESG reporting in the United States is currently moving toward standardized frameworks, led by organizations such as SASB and the Global Reporting Initiative (GRI), and accounting for ESG reporting is moving toward assured, if not audited, data. These changes may help attenuate greenwashing.

The empirical results presented in this paper are correlational, and thus warrant additional investigation. While the results are consistent with a causal story and with the principal-agent theory, they do not provide causal evidence. Managers (and thus their ESG preferences) are not randomly assigned to firms, nor, likewise, are investors assigned randomly. Further research could examine exogenous changes to a firms’ management and investor base or their preferences, and how greenwashing behavior changes as a result. Additionally, research could investigate whether greenwashing firms match with certain types of managers or investors.

The empirical proxies used in this study could introduce bias to the empirical analysis. ESG language may reflect managers’ perceptions of demands from stakeholders, in addition to or rather than a reflection of their own preferences. Without dismissing this possibility, at a minimum this study’s results indicate that this language is a signal that firms’ ESG outcomes are judged to be more positive (and less harmful) by a broad set of societal stakeholders.
measuring ESG outcomes through media coverage may introduce bias. While individual media outlets undoubtedly come at reporting issues from a specific perspective, the breadth and specificity of the data sources gathered by TruValue, along with the elimination of press releases in their algorithms, provide some level of comfort that the events reported reflect actual outcomes. The use of these proxies is motivation for further research using alternative measures of ESG outcomes. Recent research looking at only managers’ language using alternative measures of ESG performance have found consistent results, suggesting that news media is in line with direct and more specific measures (Chava, Du, and Malakar, 2021).

Beyond theory, these results have significant implications for managers, investors, and regulators. The surface level takeaway is that over the past decade, greenwashing has been a feasible strategy for managers who are not committed to ESG issues. However, this result also indicates that for managers who want to pursue substantive ESG strategies, they will have to establish more robust methods to credibly signal their ESG commitments. Following the signaling perspective taken in this paper, this means that ESG-oriented managers must find practices that are sufficiently costly for low-ESG firms to imitate such that stakeholder will be able to differentiate high and low performing firms.

For investors, these results suggest that over the past decade, many firms have put in place symbolic ESG policies as a result of investor pressure, and these policies have not, on average, resulted in improved outcomes. Again, these results could have two interpretations. If institutional investors do not care about ESG outcomes, and only want to attract retail investors attracted to ESG-labeled products, these results indicate that institutional investors themselves can successfully greenwash their products. However, if institutional investors do want to push firms toward better ESG outcomes (as their retail clients appear to want based on their
investment choices), they will need more reliable ESG signals and data from firms. This could mean demanding standardized, mandatory, and/or audited or assured ESG data. In terms of immediate practice, the finding that greenwashing correlates with ratings disagreement suggests that investors interested in measuring firms’ ESG performance ought to consider using data reliant only on independently reported ESG outcomes in addition to traditional ESG ratings.

Governments and regulators both in the US and across the world are currently debating how to standardize ESG reporting. These results suggest that these steps will be necessary to turn the goals of the ESG and stakeholder capitalism movement into improved firm outcomes and impacts. If investors do not have reliable signals of ESG outcomes, firms can greenwash, and investor led stakeholder capitalism will not produce the substantive change it claims to pursue.

References


Nason D. 2020, December 21. ‘Sustainable investing’ is surging, accounting for 33% of total U.S. assets under management. *CNBC*. Available at: https://www.cnbc.com/2020/12/21/sustainable-investing-accounts-for-33percent-of-total-us-assets-under-management.html [1 February 2021].


Ramanna K. 2021, January 17. ESG accounting needs to cut through the greenwash. Financial Times. Available at: https://www.ft.com/content/99b7a241-dfd4-4770-82b6-13de5a2804bf.


### Table 1: Summary of Model Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Investor-Owned Utilities Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Manager’s preference for ESG</td>
<td>The top management team’s perception that lower carbon emissions will lead to higher profits via, for example, lower regulatory burden or higher consumer loyalty, and/or their perceptions that the firm has a social responsibility to reduce their carbon emissions.</td>
</tr>
<tr>
<td>$x$</td>
<td>ESG outcomes</td>
<td>Firm’s overall carbon emissions.</td>
</tr>
<tr>
<td>$c(x)$</td>
<td>Cost of ESG outcomes</td>
<td>The cost of new solar power plant or wind farm installations and/or the costs of modifying existing plants to reduce or capture emissions.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Shareholder’s preference for ESG</td>
<td>Investors’ demand for lowering carbon emissions. This could be based on the belief that reducing carbon emissions provides financial benefits through reducing regulatory costs, or from a belief that the firm has a moral obligation to reduce carbon emissions to lessen climate change.</td>
</tr>
<tr>
<td>$r$</td>
<td>ESG policies (self-reported)</td>
<td>A firm’s reported carbon reductions (which could be from selected projects, and not representative of the overall firm’s emissions), or their self-reported carbon-reduction policies.</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Level of activist oversight</td>
<td>Oversight of the utility’s ESG activities by environmental activists. This could be, for example, the membership and activity level of the local Sierra Club.</td>
</tr>
<tr>
<td>$f(r - x)$</td>
<td>Penalty activist imposes</td>
<td>The expected cost of an activist campaign to the firm, and by extension the manager. For example, the Sierra Club could organize a campaign against the utility for exaggerating their carbon reductions.</td>
</tr>
</tbody>
</table>

### Table 2: Sample

<table>
<thead>
<tr>
<th>Year</th>
<th>TruValue, Compustat, Earnings Calls, Investor data</th>
<th>Previous variable and Thomson Reuters ESG</th>
<th>All Variables and Robustness Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>349</td>
<td>102</td>
<td>96</td>
</tr>
<tr>
<td>2009</td>
<td>417</td>
<td>125</td>
<td>118</td>
</tr>
<tr>
<td>2010</td>
<td>499</td>
<td>153</td>
<td>147</td>
</tr>
<tr>
<td>2011</td>
<td>575</td>
<td>179</td>
<td>172</td>
</tr>
<tr>
<td>2012</td>
<td>654</td>
<td>195</td>
<td>186</td>
</tr>
<tr>
<td>2013</td>
<td>731</td>
<td>219</td>
<td>211</td>
</tr>
<tr>
<td>2014</td>
<td>838</td>
<td>229</td>
<td>223</td>
</tr>
<tr>
<td>2015</td>
<td>972</td>
<td>471</td>
<td>461</td>
</tr>
<tr>
<td>2016</td>
<td>1,097</td>
<td>762</td>
<td>733</td>
</tr>
<tr>
<td>2017</td>
<td>1,173</td>
<td>905</td>
<td>842</td>
</tr>
<tr>
<td>2018</td>
<td>1,203</td>
<td>996</td>
<td>901</td>
</tr>
<tr>
<td>2019</td>
<td>1,169</td>
<td>1,150</td>
<td>931</td>
</tr>
<tr>
<td>Total</td>
<td>9,677</td>
<td>5,486</td>
<td>5,021</td>
</tr>
</tbody>
</table>

Counts refer to the number of firm-level observations in each year.
Table 3: Summary Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>ESG Policies</th>
<th>ESG Policies</th>
<th>Earnings Call ESG (SASB)</th>
<th>Earnings Call ESG (Word Count)</th>
<th>Investor ESG Orientation</th>
<th>Long Term Shareholders</th>
<th>Employees, log</th>
<th>Revenue, log</th>
<th>Assets, log</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>4,990</td>
<td>4,990</td>
<td>4,990</td>
<td>4,990</td>
<td>4,990</td>
<td>4,986</td>
<td>4,990</td>
<td>4,990</td>
<td>4,990</td>
</tr>
<tr>
<td>Mean</td>
<td>21.3</td>
<td>61.7</td>
<td>0.002</td>
<td>0.552</td>
<td>0.009</td>
<td>-0.134</td>
<td>2.05</td>
<td>7.61</td>
<td>8.28</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15.0</td>
<td>15.9</td>
<td>0.002</td>
<td>0.357</td>
<td>0.011</td>
<td>0.132</td>
<td>1.46</td>
<td>2.02</td>
<td>1.80</td>
</tr>
<tr>
<td>Median</td>
<td>16.0</td>
<td>63.3</td>
<td>0.001</td>
<td>0.463</td>
<td>0.006</td>
<td>-0.134</td>
<td>1.84</td>
<td>7.66</td>
<td>8.16</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.0</td>
<td>1.0</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.633</td>
<td>0.00</td>
<td>0.00</td>
<td>3.86</td>
</tr>
<tr>
<td>Maximum</td>
<td>67.0</td>
<td>99.8</td>
<td>0.032</td>
<td>2.650</td>
<td>0.100</td>
<td>0.782</td>
<td>7.70</td>
<td>13.20</td>
<td>14.80</td>
</tr>
</tbody>
</table>

Table 4: Correlation Table (4,032 Observations)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) ESG Policies</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) ESG Outcomes</td>
<td>0.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Earnings Call ESG (SASB)</td>
<td>0.08</td>
<td>0.06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Earnings Call ESG (Word Count)</td>
<td>0.11</td>
<td>0.09</td>
<td>0.26</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) Investor ESG Orientation</td>
<td>0.12</td>
<td>-0.01</td>
<td>-0.00</td>
<td>0.03</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Long Term Shareholders</td>
<td>0.15</td>
<td>-0.03</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.07</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) Employees, log</td>
<td>0.58</td>
<td>-0.07</td>
<td>0.03</td>
<td>0.06</td>
<td>0.11</td>
<td>0.10</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>(8) Revenue, log</td>
<td>0.57</td>
<td>-0.10</td>
<td>0.04</td>
<td>0.12</td>
<td>0.11</td>
<td>0.09</td>
<td>0.81</td>
<td>1.00</td>
</tr>
<tr>
<td>(9) Assets, log</td>
<td>0.60</td>
<td>-0.12</td>
<td>0.01</td>
<td>0.06</td>
<td>0.15</td>
<td>0.14</td>
<td>0.68</td>
<td>0.85</td>
</tr>
</tbody>
</table>

All dependent and independent variables (1 through 7) are normalized at the sector-year level. Control variables (8, 9 and 10) are logged. All values above 0.03 (or below -0.03) have 95% confidence intervals that do not overlap with zero. All values above 0.04 have 99% confidence intervals that do not overlap with zero.
Table 5: Managers’ and Investors’ ESG Stated Preferences, and ESG Policies & Outcomes

<table>
<thead>
<tr>
<th></th>
<th>(1) Policies</th>
<th>(2) Policies</th>
<th>(3) Outcomes</th>
<th>(4) Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers’ ESG Language</td>
<td>0.0295**</td>
<td>0.0142</td>
<td>0.0474**</td>
<td>0.0152</td>
</tr>
<tr>
<td>(SASB Semantic Distance)</td>
<td>[2.66]</td>
<td>[1.48]</td>
<td>[3.02]</td>
<td>[0.95]</td>
</tr>
<tr>
<td>Investors’ ESG Orientation</td>
<td>0.0345***</td>
<td>0.0201*</td>
<td>0.0187</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>[3.45]</td>
<td>[2.48]</td>
<td>[1.32]</td>
<td>[0.25]</td>
</tr>
<tr>
<td>Employees, log</td>
<td>0.315***</td>
<td>0.214***</td>
<td>0.0342</td>
<td>-0.131</td>
</tr>
<tr>
<td></td>
<td>[19.62]</td>
<td>[5.16]</td>
<td>[1.51]</td>
<td>[-1.88]</td>
</tr>
<tr>
<td>Assets, log</td>
<td>0.169***</td>
<td>0.0918**</td>
<td>-0.0990***</td>
<td>0.155**</td>
</tr>
<tr>
<td></td>
<td>[12.25]</td>
<td>[2.95]</td>
<td>[-5.07]</td>
<td>[2.99]</td>
</tr>
<tr>
<td>Revenue, log</td>
<td>0.0525***</td>
<td>-0.0219</td>
<td>0.0334</td>
<td>-0.0571</td>
</tr>
<tr>
<td></td>
<td>[3.91]</td>
<td>[-1.06]</td>
<td>[1.76]</td>
<td>[-1.65]</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.852***</td>
<td>-1.639***</td>
<td>-0.723*</td>
<td>-0.494</td>
</tr>
<tr>
<td></td>
<td>[-7.96]</td>
<td>[-8.54]</td>
<td>[-2.20]</td>
<td>[-1.54]</td>
</tr>
</tbody>
</table>

Fixed Effects: Sector x, Firm x, Year x

Observations 4,978 4,978 4,978 4,978
R-squared 0.56 0.40 0.12 0.01

Managers’ ESG Language is a measure of the semantic distance between their language on earnings calls and sustainability standards (from SASB).

Investors’ ESG Orientation is a measure of the proportion of a firms’ investors’ assets under management in funds labeled as sustainable in their prospectus.

Investor ESG Orientation, ESG Policies, and ESG Outcomes are all normed at the sector-year level.

All independent and control variables lagged by 1 year.

T-statistics in brackets. * p<0.05, ** p<0.01, *** p<0.001.
<table>
<thead>
<tr>
<th></th>
<th>(1) Policies</th>
<th>(2) Policies</th>
<th>(3) Outcomes</th>
<th>(4) Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers’ ESG Language</td>
<td>0.0396***</td>
<td>0.0125</td>
<td>0.0781***</td>
<td>0.0299</td>
</tr>
<tr>
<td>(Word Count Measure)</td>
<td>[3.64]</td>
<td>[0.94]</td>
<td>[5.09]</td>
<td>[1.35]</td>
</tr>
<tr>
<td>Investors’ ESG Orientation</td>
<td>0.0334***</td>
<td>0.0199*</td>
<td>0.017</td>
<td>0.00296</td>
</tr>
<tr>
<td></td>
<td>[3.33]</td>
<td>[2.46]</td>
<td>[1.2]</td>
<td>[0.22]</td>
</tr>
<tr>
<td>Employees, log</td>
<td>0.317***</td>
<td>0.215***</td>
<td>0.0344</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>[19.75]</td>
<td>[5.18]</td>
<td>[1.52]</td>
<td>[-1.88]</td>
</tr>
<tr>
<td>Assets, log</td>
<td>0.172***</td>
<td>0.0918**</td>
<td>-0.0925***</td>
<td>0.154**</td>
</tr>
<tr>
<td></td>
<td>[12.37]</td>
<td>[2.95]</td>
<td>[-4.72]</td>
<td>[2.97]</td>
</tr>
<tr>
<td>Revenue, log</td>
<td>0.0481***</td>
<td>-0.0224</td>
<td>0.0245</td>
<td>-0.0573</td>
</tr>
<tr>
<td></td>
<td>[3.57]</td>
<td>[-1.08]</td>
<td>[1.28]</td>
<td>[-1.66]</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.802***</td>
<td>-1.638***</td>
<td>-0.633</td>
<td>-0.486</td>
</tr>
<tr>
<td></td>
<td>[-7.75]</td>
<td>[-8.53]</td>
<td>[-1.93]</td>
<td>[-1.52]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Sector</th>
<th>x</th>
<th>x</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Observations: 4,606
R-squared: 0.57

Managers’ ESG Language is a measure of the proportion of ESG-related words managers use on earnings calls.
Investors’ ESG Orientation is a measure of the proportion of a firm’s investors’ assets under management in funds labeled as sustainable in their prospectus.
Investor ESG Orientation, ESG Policies, and ESG Outcomes are all normed at the sector-year level.
All independent and control variables lagged by 1 year.
T-statistics in brackets. * p<0.05, ** p<0.01, *** p<0.001.
Table 7: ESG Stated Preferences, and ESG Policies & Outcomes: Using Investor Time Horizons

<table>
<thead>
<tr>
<th></th>
<th>(1) Policies</th>
<th>(2) Policies</th>
<th>(3) Outcomes</th>
<th>(4) Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers’ ESG Language (SASB Semantic Distance)</td>
<td>0.0300**</td>
<td>0.0142</td>
<td>0.0470**</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>[2.70]</td>
<td>[1.48]</td>
<td>[3.00]</td>
<td>[0.94]</td>
</tr>
<tr>
<td>Investors’ Time Horizon (Long Term Orientation)</td>
<td>0.0356***</td>
<td>-0.0184</td>
<td>-0.0385*</td>
<td>-0.0334</td>
</tr>
<tr>
<td></td>
<td>[3.33]</td>
<td>[-1.69]</td>
<td>[-2.55]</td>
<td>[-1.84]</td>
</tr>
<tr>
<td>Employees, log</td>
<td>0.312***</td>
<td>0.222***</td>
<td>0.0402</td>
<td>-0.122</td>
</tr>
<tr>
<td></td>
<td>[19.35]</td>
<td>[5.32]</td>
<td>[1.77]</td>
<td>[-1.76]</td>
</tr>
<tr>
<td>Assets, log</td>
<td>0.169***</td>
<td>0.0933**</td>
<td>-0.0919***</td>
<td>0.157**</td>
</tr>
<tr>
<td></td>
<td>[12.2]</td>
<td>[2.99]</td>
<td>[-4.70]</td>
<td>[3.02]</td>
</tr>
<tr>
<td>Revenue, log</td>
<td>0.0552***</td>
<td>-0.024</td>
<td>0.0269</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>[4.09]</td>
<td>[-1.16]</td>
<td>[1.41]</td>
<td>[-1.74]</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.895***</td>
<td>-1.650***</td>
<td>-0.738*</td>
<td>-0.51</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Firm</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Observations</td>
<td>4,610</td>
<td>4,610</td>
<td>4,610</td>
<td>4,610</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.57</td>
<td>0.40</td>
<td>0.12</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Managers’ ESG Language is a measure of the semantic distance between their language on earnings calls and sustainability standards (from SASB).
Investors’ ESG Orientation is a measure of the proportion of a firm’s shares held by dedicated shareholders minus the proportion held by transient investors, using the Bushee (2001) classification.
Investor ESG Orientation, ESG Policies, and ESG Outcomes are all normed at the sector-year level.
All independent and control variables lagged by 1 year.
T-statistics in brackets. * p<0.05, ** p<0.01, *** p<0.001.
Table 8: Ratings Variance and Greenwashing

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenwashing</td>
<td>0.0345*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[2.13]</td>
<td></td>
</tr>
<tr>
<td>log Employees</td>
<td>-0.00543 [-0.19]</td>
<td>-0.00683 [-0.24]</td>
</tr>
<tr>
<td>log Assets</td>
<td>0.00885 [0.3]</td>
<td>0.00331 [0.11]</td>
</tr>
<tr>
<td>log Revenue</td>
<td>0.0695 [1.78]</td>
<td>0.0548 [1.39]</td>
</tr>
<tr>
<td>Constant</td>
<td>0.24 [0.8]</td>
<td>0.388 [1.27]</td>
</tr>
</tbody>
</table>

Observations 607 607

Greenwashing is defined as ESG policies (from Thomson Reuters) minus ESG outcomes (independently reported from TruValue, both variables are normed at the sector-year level before differenced). Ratings disagreement is the difference between (year normed) MSCI and Thomson Reuters ESG ratings. Year and sector fixed effects are included in both models. 

$t$ statistics in brackets. * $p<0.05$, ** $p<0.01$, *** $p<0.001$
Figures

Figure 1: Management’s ESG Language on Earnings Calls (SASB)

Figure 2: Shareholder’s ESG Orientation (Sustainable AUM Percentage)
Figure 3: Ratings Disagreement and Greenwashing

Ratings Variance and Greenwashing
Large Public US Firms, 2018-2019

Sample is 1,104 firm-year observations.
Greenwashing is defined as TruValue Insight minus (normalized) policies from Thomson Reuters.
Ratings variance is calculated using MSCI, Sustainalytics and RobecoSAM.
All ESG data was normalized at the industry-year level.
Appendix A: Proofs for optimal firm behavior in non-equilibrium model (outside stakeholders trust self-reported ESG)

Assume there is a firm $F$ with type $\alpha \in (\underline{\alpha}, \overline{\alpha})$. The firm selects a level of costly performance, $x$ with cost $c(\cdot)$, and reported performance, $r$. The only cost from reported performance coming from a penalty function, $\delta f(\cdot)$, increasing in $(r - x)$. The one-sided equilibrium game proceeds as follows:

1. The firm, $F$, observes its type $\alpha$. It then selects a level of performance $x$, and a reported level of performance, $r$.
2. Utilities are determined based on $\{x, r\}$.

Under this initial assumption that the informed agent sets the value of both $x$ and $r$. The informed agent is the only actor, all given values and functions are common knowledge: \{$\alpha, \beta, \delta, c(\cdot), f(\cdot)$\}. The cost and penalty functions are both increasing and convex in their arguments. The equilibrium under these assumptions is simply the following maximization problem.

$$\max_{x, r} u^F = \max_{x, r} [\alpha x + \beta r - c(x) - \delta f(r - x)]$$

Benefits to actual performance are $\alpha x$, where $\alpha > 0$. Benefits to reported performance are $\beta r$, where $\beta > 0$. Define the costs to actual performance and associated partial derivatives (indicated using subscripts) as $c(x)$, $\frac{\partial c(x)}{\partial x} = c_x(x)$ and $\frac{\partial^2 c(x)}{\partial x^2} = c_{xx}(x)$. Further, assume that costs as a function of actual performance are continuous, twice differentiable and convex: $c_x(0) = 0, c_x(x) \geq 0$ and $c_{xx}(x) > 0$. These assumptions ensure the minimum a firm will spend on performance is greater than 0.

There are no direct costs of broadcasting performance (as in cheap talk models, i.e. Crawford & Sobel, 1982). However, I assume firms face a penalty for misrepresenting their performance. Call this gap between reported and actual performance $g \equiv r - x$ and the cost of
misrepresenting performance $f(g)$. Assume the cost to firms for misrepresenting performance is continuous, twice differentiable, and convex: using subscripts to denote partial derivatives, $f_g > 0$ and $f_{gg} > 0$. This formulation captures a variety of possible cost structures for misrepresentation. Firms may face a stochastic possibility of being caught misrepresenting performance, such as through investigative reporting or regulatory oversight, in which case $f(g)$ represents the expected value of the cost. Alternatively, firms may face a consistent penalty from a reported/actual performance gap. For example, a set of discriminating consumers may be able to identify firms with gaps between reported and actual performance and choose not to purchase goods from those firms.

Because firms know they will face this broadcast decision when investing in performance, firms will make their investment and reporting decisions simultaneously. Formally, the firm’s problem is to maximize the following utility function:

$$u_i = \alpha x + \beta r - c(x) - f(g)$$

$$\max_{x,r} \alpha x + \beta r - c(x) - f(g)$$

Where $c_x(x) \geq 0$, $c_{xx}(x) > 0$, $g \equiv r - x$, $f_g > 0$ and $f_{gg} > 0$

To differentiate between indicate inverse functions and the multiplicative inverse, I use notation as follows:

Inverse functions are indicated by a superscript. If $f(g) = \beta$, then $g = f^{-1}(\beta)$.

Multiplicative inverses are indicated by a fraction. If $f(g) = \frac{1}{\beta}$, then $\beta = \frac{1}{f(g)}$

I know solve for the equilibrium:

$$\max_{x,r} u^F = \max_{x,r} \alpha x + \beta r - c(x) - f(g), such that g = r - x$$
The partial derivatives of the utility with respect to the choice variables must be equal to zero.

\[
\frac{\partial u}{\partial x} = \alpha - c_x(x) - f_x(g) = \alpha - c_x(x) - f_g(g)g_x = \alpha - c_x(x) + f_g(g) = 0
\]

\[
\frac{\partial u}{\partial r} = \beta - f_g(g)g_r = \beta - f_g(g) = 0
\]

Solving for \(x^*\) and \(r^*\) using the systems of equations:

\[
\alpha - c_x(x) + f_g(g) = \beta - f_g(g)
\]

\[
2f_g(g) = c_x(x) + \beta - \alpha
\]

\[
f_g(g) = \frac{c_x(x) + \beta - \alpha}{2}
\]

Plugging back into the equation for \(\frac{\partial u}{\partial x}\)

\[
\alpha - c_x(x) + f_g(g) = 0
\]

\[
\alpha - c_x(x) + \frac{c_x(x) + \beta - \alpha}{2} = 0
\]

\[
2\alpha - 2c_x(x) + c_x(x) + \beta - \alpha = 0
\]

\[
c_x(x) = \alpha + \beta
\]

\[
x^* = c_x^{-1}(\alpha + \beta)
\]

Plugging back into the equation for \(\frac{\partial u}{\partial r}\)

\[
\beta - f_g(g) = 0
\]

\[
f_g(r - x^*) = \beta
\]

\[
r - x^* = f_g^{-1}(\beta)
\]

\[
r^* = x^* + f_g^{-1}(\beta)
\]

Therefore, the firm’s optimal behavior (indicated by an asterisk), is:
\[ x^* = c_x^{-1}(\alpha_i + \beta) \]

\[ r^* = x^* + f_{\beta}^{-1}(\beta) \]

Because \( c_x(0) = 0, c_x > 0, c_{xx} > 0 \) and \( \alpha + \beta > 0 \), we know that \( c_x^{-1}(\alpha + \beta) > 0 \).

Therefore, a firm will always spend a positive amount on actual performance, ensuring \( x_i^* > 0 \).

A firm will not always report positive performance. As \( \beta \to 0 \), \( f_{\beta}^{-1}(\beta) \to -\infty \). As \( \beta \to \infty \), \( f_{\beta}^{-1}(\beta) \to \infty \). Therefore, \( f_{\beta}^{-1}(\beta) \) may be negative or positive, and firms may brownwash (underreport their performance) or greenwash (overreport their performance) based on the benefits they receive from reporting positive performance.

How do changes in \( \alpha_i \), the benefits to actual performance, and \( \beta_i \), the benefits of reported performance, shape a firm’s choice over actual and reported performance? To examine these effects, we must sign the following partial derivatives.

\[
\frac{\partial x^*}{\partial \alpha} = \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial \alpha}, \quad \frac{\partial x^*}{\partial \beta} = \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial \beta}, \quad \frac{\partial r^*}{\partial \alpha} = \frac{\partial x^*}{\partial \alpha} + \frac{\partial [f_{\beta}^{-1}(\beta)]}{\partial \alpha}, \quad \frac{\partial r^*}{\partial \beta} = \frac{\partial x^*}{\partial \beta} + \frac{\partial [f_{\beta}^{-1}(\beta)]}{\partial \beta}
\]

We are interested in the sign of \( \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial \alpha} \) and \( \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial \beta} \) over the domain \( x_i \in (0, \infty), \alpha_i \in (0, \infty), \beta \in (0, \infty) \). We know that \( c(\cdot) \) is a convex function where \( c_x(x) > 0 \) and \( c_{xx}(xx) > 0 \).

Therefore, \( c_x(x) \) is a strictly increasing monotone function. The inverse of any strictly increasing monotone function is also strictly increasing and monotone. Therefore \( c_x^{-1}(x) \) is a strictly increasing monotonic function. Because this function is increasing monotonically, taking the partial derivative of this function with respect to \( \alpha_i \) and \( \beta_i \) will always result in a positive value.

Solving, we find

\[
\frac{\partial x^*}{\partial \alpha} = \frac{\partial [c_x^{-1}(\alpha_i + \beta)]}{\partial \alpha} = \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial (\alpha + \beta)} \cdot \frac{\partial (\alpha + \beta)}{\partial \alpha} = \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial (\alpha + \beta)} > 0
\]
\[
\frac{\partial x^*}{\partial \beta} = \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial \beta} = \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial (\alpha + \beta)} \cdot \frac{\partial (\alpha + \beta)}{\partial \beta} = \frac{\partial [c_x^{-1}(\alpha + \beta)]}{\partial (\alpha + \beta)} > 0
\]

We are also interested in \(\frac{\partial [f^{-1}_\beta(\beta)]}{\beta}\). By assumption, \(f(\cdot)\) is a monotonically increasing convex function, so \(f_\beta(\beta)\) is a strictly increasing monotonic function. Therefore, \(f^{-1}_\beta(\beta)\) is also a strictly increasing monotonic function. Therefore \(\frac{\partial [f^{-1}_\beta(\beta)]}{\beta} > 0\). We will now use the signs of these partial derivatives to determine the direction and relative magnitude of a firm’s optimal behavior in response to changes in benefits to actual or reported performance.

Examining the partial derivatives with respect to \(\alpha\), we see that

\[
\frac{\partial x^*}{\partial \alpha} = \frac{\partial [c_x^{-1}(x_i)]}{\partial \alpha} > 0
\]

\[
\frac{\partial r^*}{\partial \alpha} = \frac{\partial x^*}{\partial \alpha} > 0
\]

This optimal behavior suggests that increasing \(\alpha\), the pressure for real performance, increases actual performance (through higher investment in performance by the firm) and higher broadcast performance. For example, if solar power becomes more efficient and provides increased operational benefits (\(\alpha_i\ increases\)), then the firm will invest more in solar panels, increasing actual environmental performance (\(x_i\)) and will increase broadcast behavior (\(r_i\)) by the same amount.

The partial derivatives with respect to \(\beta_i\) are as follows:

\[
\frac{\partial x^*}{\partial \beta_i} = \frac{\partial [c_x^{-1}(x)]}{\partial \beta_i} = \frac{\partial [c_x^{-1}(x)]}{\partial x} = \frac{\partial x^*}{\partial \alpha} > 0
\]

\[
\frac{\partial r^*}{\partial \beta_i} = \frac{\partial x^*}{\partial \beta_i} + \frac{\partial [f^{-1}_\beta(\beta)]}{\beta} > \frac{\partial x^*}{\partial \beta_i} > 0
\]
These partial derivatives show that increasing $\beta$, pressures for broadcast performance, also increases both actual and broadcast performance. However, broadcast performance increases more than actual performance.

For example, if institutional investors demand higher reported environmental performance from a firm, the firm faces two countervailing forces. The firm is rewarded for increasing their reported performance but increasing their reported performance without increases their actual performance increases the penalty they pay for misrepresenting performance. As a result, the firm will increase its actual performance, but not as much as they increase their reported performance due to the convexity of real performance costs. Thus, the gap between the reported and actual performance increases.
Appendix B: Proofs from equilibrium model (outside stakeholders rationally discount self-reported ESG)

Suppose now that there is an audience which rationally sets the firm’s ESG rating, \( r \), with the goal of setting the rating as close to the actual performance as possible. Specifically, the audience’s rule is to set \( E[x|r] = r \). The only communication between the inside stakeholder and audience is a message sent by the inside stakeholder to the audience, which can be thought of as a noisy signal of the firm’s actual ESG performance, \( x \). Under this assumption, the model follows the structure of the cheap talk model introduced by Crawford and Sobel (1982). The formal solutions are presented below, but first I present a verbal summary of the equilibrium:

The solution concept for the model is a partition equilibrium. In a partition equilibrium, firms are sorted into one or more groups corresponding to ranges of performance, and the firm’s rating, \( r \), is equal to the average performance of all firm types in the range under which their performance falls. As the average (absolute value of the) bias of firms increases, the outside stakeholder can less accurately sort firms. As the bias becomes large, eventually there is a point at which the only solution is a 1-partition equilibrium, in which all firm types would be given a single rating equal to the average performance of all firm types.

For example, say the parameters of the model are such that at most a two-partition equilibrium is possible (a one-partition equilibrium is always possible, but higher partition equilibria always produce greater utility when they are possible). In a two-partition equilibrium, two ratings are possible: \( r' \) and \( r \), such that \( r' > r \). If the firm is type \( \alpha > \bar{\alpha} \), then the firm will choose a performance level and send a message such that they receive the rating \( r' \). If the firm is of the type \( \alpha < \bar{\alpha} \), they will choose a performance level and send a message such that they receive the rating \( r \). For such an equilibrium to hold, the average performance of all firms with type \( \alpha > \bar{\alpha} \) must be \( r' \), given that these firms know they will receive the rating \( r' \). The equivalent must hold for all firms of type \( \alpha < \bar{\alpha} \). As a result, firms are accurately, but imprecisely, sorted into “high” or “low” performers.

The intuitive tension in the model is that if the firm has a bias to have a rating that is significantly different than their actual performance, the audience will discount any message they send. The audience treats the message as cheap talk, and will assign a conservative (and, on average, less accurate) rating to the firm. As the audience’s expectation of the firm’s bias moves closer to zero, the audience will be better able to distinguish between different firms based on the messages they send, and average ratings accuracy will increase. The interesting component of the model is that the firm’s bias can be negative (underreporting ESG performance, \( r < x \)), or positive (greenwashing, \( r > x \)), and that this bias is a function of the inside and outside stakeholders’ preferences and activist presence. If outside stakeholders have low preferences for ESG performance relative to inside stakeholders’ preferences and the level of activist presence, firms will underreport (i.e., not report) their ESG performance so as to not draw activist attention. As a result, third party assessments of ESG performance reliant on firms’ self-reported data would be imprecise and inconsistent across raters. The lack of attention to ESG issues from investors in fossil fuels relative to the presence of environmental activists in the 1990s would fit this scenario.
Alternatively, if outside stakeholders have a high demand for ESG performance, while inside stakeholders do not, firms will greenwash their ESG behavior, also leading to imprecise and inconsistent third party ESG assessments. This may presently be the case in some firms facing high investor demands for ESG performance, but with top management doubtful of climate science or believing there is responsibility for business to reduce their harmful externalities. I expect this greenwashing effect to dominate, given the recent rapid rise in ESG investing.

From this two-sided equilibrium, I develop the following proposition:

As the outside stakeholder’s preference for ESG performance increases (i.e. investors’ increased commitment to ESG ratings over the past decade) relative to inside stakeholder’s ESG preferences, the accuracy of these ratings will increase up to a point, and then decrease. The disagreement between ratings will be directly inversely correlated with ratings accuracy, so stated using ratings disagreement, this formulation of the model leads to the following propositions:

**Proposition 3:** As outside stakeholder preferences for ESG performance \( (\beta) \) increases relative to that of the inside stakeholder \( (\alpha) \), ratings disagreement will fall and then rise, following a U shape.

**Complete Proof:**

Begin with the setup of the one-sided model, but now assume that the uninformed stakeholder rationally sets a firm’s reported performance, \( r \), as the expected value of their performance given that they receive the rating.

The game proceeds in three stages:

1. The firm, \( F \), observes its type \( \alpha \), selects a level of performance \( x \), and sends a message, \( n \), to the audience, \( A \). This signal can be random and can be viewed as a noisy estimate of the firm’s actual performance \( x \).
2. The audience sets a rating for the firm, \( r \).
3. Utilities are established.

The parameters of the model are defined as follows.

**Parameters**

Assume there is a distribution of firms defined by \( \alpha \in (\alpha, \bar{\alpha}) \) with a cumulative distribution function \( G(\alpha) \) with density \( g(\alpha) \) such that \( G(\alpha) = 0 \) and \( G(\bar{\alpha}) = 1 \).
The benefits to reported performance are defined by the parameter $\beta$ (as in the one-sided model). The rating is assigned by the audience (unlike in the one-sided model, where the rating is set by the firm) and is defined as $r \in \mathbb{R}$.

**Utility Functions**

The utility of the firm is defined as $u^F(x, r, \beta)$, a continuous and twice differentiable function. The utility of the audience is defined as $u^A(r, x)$, a continuous and twice differentiable function.

**Partial Derivative Assumptions**

I use the notation $\frac{\partial u^F(x, r, \beta)}{\partial x} \equiv u^F_x(x, r, \beta) \equiv u^F_1(x, r, \beta)$ and $\frac{\partial^2 u(x, r, \beta)}{\partial x^2} \equiv u^F_{xx}(x, r, \beta) \equiv u^F_{11}(x, r, \beta)$. Further, I use $u^F(\cdot)$ to indicate the utility function with the arguments in any order, so a statement about $u^F_1(\cdot)$ is a statement about the partial derivative with respect to any of the arguments.

Assume, for $i \in \{F, A\}$, that for some $r$, $u^i_{rr}(r, x) < 0$, meaning that $u^i$ has a unique maximum $r$ for a given $x$. Further, assume that $u^F_{rr}(r, x, \beta) = 0$ for some $r$ and $u^F_{rr}(r, x, \beta) < 0$, so $u^F$ has a unique maximum $r$ for a given $(x, r, \beta)$. Finally, assume that $u^i(r, x) > 0$. This last assumption is a sorting requirement meaning that the best value of $r$ for a fully informed agent is increasing in $x$.

**Equilibrium**

I use a Bayesian Nash Equilibrium to define the solution. While the firm’s choice of performance and message precede the audience’s ratings assignment in time, the audience sees only the message, and therefore all these actions are strategically simultaneous.

The equilibrium consists of a signaling rule for the firm defined as, $q(n|x)$, and an action rule for the audience defined as $r(n)$, such that
1) The firm’s signaling rule maximizes expected utility for a given type, defined by \( \alpha \), taking the audience’s ratings rule as a given. Specifically, for each \( \alpha \in (\alpha, \bar{\alpha}) \), \( \int_N q(n|x)dn = 1 \), where the Borel set \( N \) is the set of feasible signals, and if \( n^* \) is in the support of \( q(n|x) \), then \( n^* \) solves \( \max_{n \in N} u^F(r(n), x, \beta) \).

2) The audience responds optimally to each signal, using Bayes’ rule to update their prior, taking into account \( G(\alpha) \), the firm’s signaling strategy, and the signal received. Specifically, for each \( n \), \( y(n) \) solves \( \max_r \int q^\alpha(r, x) p(x|\alpha) d\alpha, \) where \( p(\alpha|n) \equiv \frac{q(n|\alpha)g(\alpha)}{\int_\alpha q(n, \alpha)f(\alpha)d\alpha} \).

**Modifications from the classical Crawford and Sobel (1982) model.**

Assume the firm’s utility takes the functional form from the one-sided model:

\[
u^F = \alpha x + \beta r - c(x) - f(r - x)\]

Where \( x \geq 0, c(x) \) is defined over the domain \( x \in [0, \infty), c(0) = 0, c_x(x) \geq 0, c_{xx}(x) > 0, \) \( g_i \equiv r - x, f_g > 0 \) and \( f_{gg} > 0 \).

The one-sided equilibrium is solved in the main paper. Here I restrict my attention to the two-sided equilibrium. In this equilibrium, the firm takes \( r_i \) as a given and maximizes their utility by choosing \( x_i \):

\[
u_i = \alpha_i x_i + \beta_i r_i - c(x_i) - f(r_i - x_i)\]

\[
\frac{\partial u_i}{\partial x_i} = \alpha_i - c_x(x_i) + f_x(r_i - x_i) = 0
\]

\[
x_i^* = \{x_i | \alpha_i - c_x(x_i) + f_x(r_i - x_i) = 0\} \tag{1}
\]

We are interested in the point where \( \alpha_i + f_x(r_i - x_i) = c_x(x_i) \)

**Lemma 1:** The point \( x^* \) is unique for a given \((\alpha, r)\).

**Proof:** The function \( c_x(x) \) is monotonically increasing and convex, and the function \( \alpha + f_x(r - x) \) is convex and monotonically decreasing in \( x_i \). Therefore, the conditions for \( x_i^* \) satisfy
the single crossing property, and there will be only one point \( x^* \) at which \( \alpha_i + f_x(r - x) = c_x(x) \). Therefore, \( x^* \mid r \) exists and is unique for each \( \alpha > 0 \).

**Lemma 2:** \( x^* \) is strictly increasing in \( \alpha \).

**Proof:** From Equation (1), \( x^* = \{ x \mid \alpha - c_x(x) + f_x(r - x) = 0 \} \). By assumption, \( \alpha_i > 0 \) and \( c_x(x) > 0 \). Since \( \alpha - c_x(x) + f_x(r - x) = 0 \), we know \( c_x(x) > f_x(r - x) \). If we increment \( \alpha \), then \( c_x(x) - f_x(r - x) \) increases. Suppose \( x \) decreased (or remained constant) as a result of incrementing \( \alpha \). Then, \( c_x(x) \) would decrease, at a greater rate than \( f_x(r - x) \) would increase (or both would remain constant if \( \alpha \) remained constant). Then \( c_x(x) - f_x(r - x) \) would decrease (remain constant), and we have a contradiction.

**Lemma 3:** If there are two possible ratings to select into, \( r' > r \), there is a specific point \( \bar{\alpha} \) at which a firm will be indifferent between rating \( r \) and rating \( r' \). Furthermore, all firms with \( \alpha > \bar{\alpha} \) will choose the \( r' \) and all firms for which \( \alpha < \bar{\alpha} \) will select into \( r \).

**Proof:** Assume a firm can select to send a message such that they will be assigned either the rating \( r \) or the rating \( r' > r \). A firm will be indifferent when \( u(x^* \mid \alpha, r) = u(x^* \mid \alpha, r') \).

From the previous proof, \( \frac{\partial x^*}{\partial \alpha} > 0 \). Similarly, \( \frac{\partial x^*}{\partial r} > 0 \).
Then the utility functions satisfy the single crossing property, and therefore there exists some $\alpha : u(x^*|\alpha, r) = u(x^*|\alpha, r')$. Therefore, firms with $\alpha > \bar{\alpha}$ will choose the $r'$ and all firms for which $\alpha < \bar{\alpha}$ will select into $r$. A visual example is shown below:

![Graph showing firm utility vs. benefits from performance](image)

**Return to (Crawford and Sobel (1982) proof):**

The previous section has assumed or established the following:

1) $\alpha \in (\alpha, \bar{\alpha})$
2) $x^*_\alpha(\alpha, r, \beta) > 0$ and $x^*_{\alpha\alpha}(\alpha, r, \beta) < 0$. Therefore $x(\alpha)$ is a monotonic transformation.

Define $\overline{m} \equiv \max\{x^*|r\}$, $\underline{m} \equiv \min\{x^*|r\}$, and define the function $m \equiv \frac{\text{argmax} u^r|\alpha \beta - \overline{m}}{\overline{m} - \underline{m}}$.

Now, for a given set of ratings, $m$ fits the required firm definition to fit into the Crawford and Sobel (1982) model of cheap talk.

**Theorem 1:**

Crawford and Sobel (1982) version: The number of partitions possible in equilibrium is increasing in the misalignment between the sender and receiver.

My model: The number of partitions possible is increasing, then decreasing in $\beta$.

Specifically, $\frac{\partial^2 N(\beta)}{\partial \beta^2} < 0$.  

58
Proof: The bias term is decreasing, then increasing in $\beta$. At the point $\beta = \beta^*$ where $0 < \beta^* < \infty$, the bias term equals zero. Therefore, the number of partitions possible is increasing, then decreasing in $\beta$.

Conclusions

The point $x^*$ is unique for a given $(\alpha, r)$, and further, $x^*$ is strictly increasing in $\alpha$ and $r$ (Lemmas 1, 2 and 3). The structure of the firm’s utility is such that there is a single value of $x$ which maximizes its utility if it takes its type $(\alpha)$ and ratings $(r)$ as a given. This is due to the convexity of costs to actual performance $(x)$, concave benefits to reducing the reported/actual gap $(r - x)$, and linear benefits to performance $(x)$. Partial derivatives of the firm’s utility function reveal that $x^*$ is increasing in both $\alpha$ and $r$.

The main conclusion from the two-sided model relates to ratings accuracy. If reported performance gives firms zero benefit, but there are penalties increasing in $(r - x)$, firms will be biased to report the lowest possible value they can in order to minimize the penalty $f(r - x)$. As $\beta$ approaches $\infty$, the firm will garner such benefits from reported performance that every firm, regardless of their ability or performance, will prefer to garner the highest rating, thus eliminating the ability for the audience to distinguish between high and low type firms. In between the extremes of $\beta = 0$ and $\beta = \infty$, there is a range of $\beta$ over which the firm’s preferred rating coincides sufficiently closely with their actual performance. In these cases, firms prefer to be rated accurately, and signal as such. Resultingly, Theorem 1 states that the number of partitions possible is increasing, then decreasing in $\beta$.

This model makes the following additions to the model in Crawford and Sobel (1982). In this paper’s model, firms select their level of performance, whereas in Crawford and Sobel (1982) the firm (which they call the sender) has an assigned value which they cannot modify. Further, the
bias term may be different signed for different firms in the distribution of firms being rated, unlike in Crawford and Sobel (1982). Finally, these two changes are made explicit by the costs and benefits to firm performance and the firm’s rating. The result of all these changes is the novel Theorem 1, as well as secondary propositions, all centered upon endogenously chosen performance levels.
# Appendix C: Examples of Text from ESG Articles and Earnings Calls

## Examples of ESG Related Articles from TruValue

### MT / ArcelorMittal

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Waste &amp; Hazardous Materials Management</td>
<td>ArcelorMittal has been accused of manipulating test results to conceal the amount of ammonia and cyanide that is being discharged from its plant in Indiana.</td>
</tr>
</tbody>
</table>


### AMZN / Amazon.com

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Labor Practices</td>
<td>Amazon has allegedly threatened to fire employees for criticizing the company's environmental policies.</td>
</tr>
</tbody>
</table>


### JNJ / Johnson & Johnson

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Product Quality &amp; Safety</td>
<td>Washington has sued Johnson &amp; Johnson for allegedly deceiving patients and doctors about the addictive nature of the opioid pain killers it developed and marketed.</td>
</tr>
</tbody>
</table>


### ORSTED / Orsted A/S

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Systemic Risk Management</td>
<td>Orsted will pay £4.5 million because its power station failed to remain connected after a lightning strike, causing a power outage in the UK that affected more than one million customers.</td>
</tr>
</tbody>
</table>


### GE / General Electric Co

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Product Design &amp; Lifecycle Management</td>
<td>GE said it won 1,215 MW of wind turbine orders or commitments from customers in China in 2019.</td>
</tr>
</tbody>
</table>


### TXT / Textron Inc

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Critical Incident Risk Management</td>
<td>Several employees of Textron Inc were reported injured following an explosion at a plant in Wichita, Kansas.</td>
</tr>
</tbody>
</table>


### ARCO / Arcos Dorados Holdings

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Employee Health &amp; Safety</td>
<td>Peruvian authorities say that McDonald's Latin America franchisee Arcos Dorados violated the law with its working conditions after two employees were electrocuted and killed.</td>
</tr>
</tbody>
</table>

Source: Haaretz: haaretz.com/world-news/wires/1.8321179

### DRE / Duke Realty

<table>
<thead>
<tr>
<th>ESG Effect</th>
<th>Category</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Energy Management</td>
<td>Duke Realty has announced that, starting this year, it will develop only LEED-certified buildings.</td>
</tr>
</tbody>
</table>

### Sample of ESG Related Quotes from Earnings Calls

Some companies explicitly discuss their stakeholder orientation or strategy

<table>
<thead>
<tr>
<th>Company</th>
<th>Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>“Now, from the start of this COVID-19 crisis, we’ve been guided by a clear set of priorities that are very much in line with our multi-stakeholder business model to protect our people, safeguard supply, respond to new patterns of demand, obviously support our communities and preserve cash and balance sheet strength.” -BP, 2020 Q3</td>
</tr>
<tr>
<td>Unilever</td>
<td>“It’s never been more relevant for brands to demonstrate their positive contribution to society and address the issues that our consumers care about in an authentic way. So we’re investing more of our marketing spend on communication, which is explicitly purposeful,” Unilever, 2020 Q3</td>
</tr>
<tr>
<td>Unilever</td>
<td>“In addition, while the world continues to grapple with the devastating effects of COVID-19 and the issues of escalating inequality, it’s more important than ever that we don’t lose sight of the climate crisis and the very real and serious threats that it creates for all of us. Climate change, nature degradation, biodiversity decline, water scarcity, they’re deeply interconnected problems and we are committed to find ways to tackle them simultaneously. So in June, we announced a wide-ranging set of commitments . . .” Unilever, 2020 Q3</td>
</tr>
</tbody>
</table>

Discussion of ESG only in the Q&A section may suggest ESG issues are a low priority

<table>
<thead>
<tr>
<th>Company</th>
<th>Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exxon Mobile</td>
<td>“I guess the first question, there’s been a lot certainly that you guys have talked about and that’s been written about, carbon, and a lot of your peers who have come out with explicit carbon targets, I just wanted to know where Exxon is in that journey in terms of coming out with carbon targets and how are you thinking about setting them to the extent that’s the path you choose to go down?” [Answer by Jack Williams, Senior VP, Exxon Mobile]</td>
</tr>
<tr>
<td></td>
<td>“Yes. Thanks, Niel. Let me just take a step back a little bit on how we see things and what we’re doing. Basically, we see a world that’s going to need more energy going forward; through population, through GDP growth, and a lot of that goes through the non-OECD countries. Energy consumption is tied to the population and GDP. And we don’t think the current solutions set is really complete.” Exxon Mobile, 2020 Q3</td>
</tr>
</tbody>
</table>
Appendix D

Summary Statistics, Normalized Variables

<table>
<thead>
<tr>
<th>Statistic</th>
<th>ESG Policies</th>
<th>ESG Outcomes</th>
<th>Negative ESG Outcomes</th>
<th>Earnings Call ESG (SASB)</th>
<th>Earnings Call ESG (Word count)</th>
<th>Investor ESG Orientation</th>
<th>Long Term Shareholders</th>
<th>Employees, log</th>
<th>Revenue, log</th>
<th>Assets, log</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>4,982</td>
<td>4,982</td>
<td>4,914</td>
<td>4,983</td>
<td>4,983</td>
<td>4,056</td>
<td>4,986</td>
<td>4,990</td>
<td>4,990</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0</td>
<td>0.0</td>
<td>0.04</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.05</td>
<td>-0.08</td>
<td>2.05</td>
<td>7.61</td>
<td>8.28</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1.0</td>
<td>1.0</td>
<td>1.12</td>
<td>0.94</td>
<td>0.99</td>
<td>1.00</td>
<td>0.95</td>
<td>1.46</td>
<td>2.02</td>
<td>1.80</td>
</tr>
<tr>
<td>Median</td>
<td>-0.3</td>
<td>0.1</td>
<td>-0.16</td>
<td>-0.26</td>
<td>-0.21</td>
<td>-0.11</td>
<td>-0.09</td>
<td>1.84</td>
<td>7.66</td>
<td>8.16</td>
</tr>
<tr>
<td>Minimum</td>
<td>-2.1</td>
<td>-4.4</td>
<td>-0.33</td>
<td>-1.38</td>
<td>-2.01</td>
<td>-1.43</td>
<td>-3.97</td>
<td>0.00</td>
<td>0.00</td>
<td>3.86</td>
</tr>
<tr>
<td>Maximum</td>
<td>3.8</td>
<td>2.7</td>
<td>14.40</td>
<td>8.32</td>
<td>4.90</td>
<td>15.50</td>
<td>7.20</td>
<td>7.70</td>
<td>13.20</td>
<td>14.80</td>
</tr>
</tbody>
</table>