Talk or Walk the Talk?
The Real Impact of ESG Investing*

Huiyao Chen†

September 29, 2023

please click here for the latest version

Abstract

I propose a model to study how environmental, social, and governance (ESG) investors influence firms' ESG-related investments and disclosures. Paradoxically, when the firm manager can easily manipulate ESG disclosures, stronger investor ESG preference can decrease green investment: though investors attach a higher value to green outcomes, more greenwashing is induced, making ESG disclosures less reliable. Investors therefore give less reward to firms that claim to be green. Moreover, firms with poor business performance are particularly likely to greenwash and reap benefits from investors. My analysis raises concerns that the rise of ESG investing may have unintended consequences, especially when ESG-disclosure regulations are weak.

Keywords: Greenwashing, ESG, real effects, socially responsible investing, ESG reporting, externality
JEL Classification: G11, G23, G32, M14

*I am extremely grateful to my advisors Vincent Glode, Itay Goldstein, Luke Taylor, and Yao Zeng for their invaluable guidance and support. I thank Mirko Heinle, Xiao Lin, and seminar participants at the Wharton School and the 2023 FTG Summer School for their insightful comments.

†The Wharton School, University of Pennsylvania: chenhy@wharton.upenn.edu

1 Introduction

The dramatic growth of environmental, social, and governance (ESG) investing reflects investors' non-pecuniary preferences and their desire to influence the ESG practices of companies.[1] However, the empirical evidence on the real impact of ESG investing is mixed, while the prevalence of greenwashing has been widely documented: Specifically, as investors increasingly prioritize ESG outcomes, firms are more inclined to manipulate their ESG disclosures to cater to investors, which can result in significant discrepancies between their claims and their actual practices. In this paper, I propose a model to analyze this controversial question: as billions of dollars are poured into ESG assets, does it incentivize more real ESG activities, or just induce more greenwashing?

There is widespread concern that firms are engaging in greenwashing or diversity-washing and reaping benefits from ESG investors. For example, Baker et al. (2023) finds that firms may make commitments to diversity, equity, and inclusion (DEI) in their disclosures, yet have poor hiring practices that result in less workplace diversity and future outflows of diverse employees. They also highlight that those diversity-washing firms get superior ESG scores and attract investment from ESG funds. As the industry observations suggest, "corporate leaders who talk the most about diversity may benefit from greater investment in their companies by socially conscious funds, even if hiring and promotion efforts are lackluster."[2] Exacerbating the issue is the fact that investors lack reliable information to monitor ESG performance and

[1]Most institutional investors, especially those large companies like BlackRock, have made commitments to support ESG activities. For example, The collective AUM represented by all 3826 PRI signatories (3404 investors and 422 service providers) has reached $121 trillion as of March 31, 2021 (`https://www.unpri.org/annual-report-2021/how-we-work/building-our-effectiveness/enhance-our-global-footprint`); the collective AUM represented by all Net Zero Asset Managers Initiative signatories has reached $59 trillion as of December 31, 2022 (`https://www.netzeroassetmanagers.org/signatories/`). Survey studies also show that investors are willing to support ESG activities. For example, the 2022 Survey of Investors, Retirement Savings, and ESG by Stanford Rock Center for Corporate Governance shows that young investors claim to be willing to lose between 6 and 10 percent of their retirement savings to support ESG causes (`https://www.gsb.stanford.edu/sites/default/files/publication/pdfs/survey-investors-retirement-savings-esg.pdf`).

[2]CEOs Who Are All Talk and No Action on Inclusion Still Benefit (Bloomberg): `https://www.bloomberg.com/news/articles/2023-01-19/-diversity-washing-funds-can-aid-companies-even-if-they-don-t-improve-hiring#xj4y7vzkg`

identify greenwashing. In reality, ESG information is usually subjective, multi-dimensional, and lacks a definite realization to discipline ex-ante evaluations. Even ESG rating agencies, which specialize in assessing ESG information and providing aggregated scores, might have substantial disagreements over the ESG performance of firms (Berg et al., 2022).[3] Thus, it is challenging for investors to distinguish actual green firms from greenwashing firms.

In this paper, I propose a model to analyze how firms' incentives for real ESG activities and greenwashing are jointly affected by ESG investors.[4] Surprisingly, my model shows that when the cost of misreporting ESG outcomes is low, higher intensity of investors' ESG preference (i.e., investors attach a higher value to ESG outcomes) might induce more greenwashing and reduce real ESG activities, thus harming real efficiency. A novel channel driving this result is highlighted in the analysis: as investors increasingly value ESG outcomes, more greenwashing is induced since brown firms[5] have larger incentives to misreport and mimic green firms, and this effect intensifies especially when ESG information discipline is weak. Thus, ESG information becomes less reliable: among those firms that claim to be green, there are more greenwashing firms relative to actual green firms. Consequently, investors rationally give less reward to firms claiming to be green, which reduces the incentive of real green investment in the first place.

My model highlights the intrinsic paradox of incentivizing ESG activities through financial markets. On the one hand, we want investors to focus more on ESG outcomes (e.g., carbon emissions, gender equality, etc.) relative to traditional business performance, such

[3]An example of poor incentives to take ESG practices due to disagreement (Shareholders Push an Array of ESG Proposals: `https://www.wsj.com/articles/shareholders-push-array-of-esg-proposals-11651004156`): in Apple's annual meeting in March 2022, a majority of Apple investors supported a resolution requiring the board to hire a third party to undertake a "civil rights audit" of issues at the company including pay equity, leadership diversity and others. In opposing the proposal, Apple said in its proxy statement that it "already fulfills the objectives of the proposal in several ways, including through impact and risk assessments, active governance and board oversight, engagement with our communities and key stakeholders, and regular, transparent public reporting."

[4]Throughout this paper, I refer to the investors who value ESG outcomes and desire to incentivize ESG activities through investment, and I use the word ESG investors, impact investors, and socially responsible investors interchangeably.

[5]Note that for concreteness, I use the words "brown firm" and "green firm" to refer to firms with bad and good ESG outcomes respectively. Generally, the same analysis can be applied to other ESG issues.

that firms are willing to allocate resources to ESG activities rather than just maximizing profits. On the other hand, it also means that investors increasingly value the outcomes that are subject to more manipulation, which naturally induces more greenwashing and thus less incentive for actual ESG activities.

My model naturally fits a wide range of empirical contexts in which principals with ESG preference incentivize agents to undertake ESG activities. For example, the model can be applied to analyze the greenwashing of ESG funds: the customers of ESG funds want to invest in green stocks, but fund managers might just want to maximize their own payoff and they might label themselves as ESG funds in order to get a higher management fee than traditional funds. The key insight of the model rationalizes the empirical findings that ESG funds hold stocks with more voluntary ESG disclosure but worse actual ESG performance (e.g., Raghunandan and Rajgopal, 2022). In addition, my model can be applied to other empirical contexts such as green bond issuance, ESG-focused venture capital, etc.

The model has three stages: real investment stage, disclosure stage, and trading stage. To fix ideas, we consider the case of green versus brown firms:

1. At the real investment stage, firms have no project in place initially. Each firm can get either a green or brown investment opportunity, and the firm manager must decide whether to take the investment or not. The real investment has both externality value and financial value: specifically, the green (brown) investment generates positive (negative) externality, with a negative (positive) NPV. For example, we can think of the green investment as the adoption of new clean energy, which reduces carbon emissions but the transition is costly and requires substantial upfront expenditure, and the brown investment as the expansion of production using traditional energy, which increases carbon emissions but generates high profits.

2. At the disclosure stage, firms that make the investment must disclose the externality value of the project. Each firm can either truthfully report with no cost, or misreport with an information manipulation cost. Particularly, we say that a brown firm engages

in greenwashing if it claims to generate positive externality, while actually having a brown project in place and generating negative externality.

3. At the trading stage, a competitive financial market opens, and ESG investors trade the stock of the firm conditional on the disclosure of the firm, which generates a market-clearing price.[6] Note that I introduce a key parameter at this stage: the intensity of investors' ESG preferences. As this intensity of ESG preference becomes larger, investors attach a higher value to each unit of externality created by the firm (or we can say investors internalize a larger proportion of the externality).

Note that there are conflicts of interest between the firm manager and ESG investors. Specifically, ESG investors care about externality, and they attach a higher value to externality as their intensity of ESG preference increases. On the other hand, firm managers do not value externality directly. They determine their real investment and disclosure strategies to maximize their compensation, which is determined by the stock price and NPV of the investment, minus any information manipulation cost.

I show that a higher intensity of investors' ESG preference has two countervailing effects on compensation to green investment. On the one hand, firms get larger compensation for green investments if the information quality at the disclosure stage is fixed, as investors internalize more externality value and bid up the stock price of green firms. On the other hand, more greenwashing is induced in equilibrium at the disclosure stage, worsening the information quality endogenously, since brown firms get larger penalties by truthfully reporting and larger rewards by pooling with green firms. Since investors have rational expectations and update their beliefs upon receiving green disclosure, they decrease compensation to firms that claim to be green as greenwashing becomes more severe. Thus, increasing the intensity of ESG preference has non-monotonic effects on the compensation for externality value: when

[6]The key of the last stage is that investors reward or punish the manager depending on the perceived externality value according to their beliefs. We can also think of investors' actions as activism and engagement, i.e., there is a representative ESG investor issuing compensation to the manager contingent on the disclosure.

the intensity of investors' ESG preference is small, the benefit of greenwashing is smaller than the manipulation cost, so firms truthfully disclose externality value and compensation for green investment is increasing in ESG preference. As the intensity of investors' ESG preference becomes large enough (such that the benefit of greenwashing can cover information manipulation costs), brown firms start to engage in greenwashing. Moreover, the share of greenwashing firms increases with the intensity of investors' ESG preference, which decreases compensation for green investment. Since the green project is invested if the compensation for investing is larger than the cost, an increase in investors' ESG preference might have negative effects on real efficiency when greenwashing is severe.

Based on this novel mechanism, I discuss three regimes of information discipline (which represent different levels of misreporting cost) and the corresponding equilibrium. The most interesting case is when there is an intermediate level of information discipline (i.e., the cost of misreporting is in the intermediate region): to incentivize green investment in this case, the intensity of investors' ESG preference should be large enough such that the compensation to green investment can cover the financial cost of investment, but it should not be too large such that too much greenwashing undermines the information quality substantially. On the other hand, if information discipline is very weak (i.e., the cost of misreporting is low), green investment cannot be incentivized by ESG investors no matter how large the intensity of their ESG preference is, because even slight rewards for taking green projects can induce very severe greenwashing; if information discipline is very strong (i.e., the cost of misreporting is high), green investment can be easily incentivized as firms tend to truthfully report ESG outcomes.

These results have very important policy implications: for those ESG outcomes that are hard to measure or have substantial disagreement (e.g., gender diversity), regulation on information disclosure is more crucial than increasing incentivization from ESG investors; for those ESG outcomes that are relatively more measurable but still subject to greenwashing concerns (e.g., long-term carbon emission objectives), we should be careful as more ESG

investors can backfire and reduce real ESG activities; for those ESG outcomes which are easy to measure (e.g., corporate governance issues), more ESG investors can always play a positive role.

Then I analyze the extension in which there is exogenous disclosure of financial information that precisely reveals the NPV of the real investment, and investors could infer the actual ESG performance from financial information. This setup closely mirrors the fundamental disparity between ESG and financial information in the real world: While financial data, such as earnings announcements, are subjected to rigorous auditing and regulations, ESG information, particularly voluntary ESG disclosures, often remains subjective and lacks the same level of discipline. This crucial distinction underpins the key result of this extension: when earnings announcements suggest a negative NPV for a firm's investments, this could stem from two possibilities. It might be due to the firm making poor investment choices, or it could be a result of the company diverting resources towards ESG activities. In situations where information discipline is limited, firms often tend to claim the latter, seeking compensation from ESG investors. Consequently, brown firms with a negative NPV are more likely to engage in greenwashing and reap substantial benefits. This result rationalizes empirical observations that many companies publicly embrace ESG initiatives as a cover for poor business performance (e.g., Flugum and Souther, 2022; Baker et al., 2023).

To show that the key insight from the baseline model is robust, I consider two more general model setups. First, I consider an extension in which agents have heterogeneous private information manipulation costs. I derive a threshold equilibrium: the firm manager engages in greenwashing if his cost of information manipulation is lower than a threshold instead of randomizing between engaging in greenwashing or not at the indifference point. Second, I consider an extension in which investment opportunities are endogenized, i.e., brown firms have the option of adopting green technology and becoming green firms. In equilibrium, high-pollution brown firms with a high cost of adopting green technology choose to engage in greenwashing, which undermines information quality as well as the incentive of

low-pollution firms to take green investments. It illustrates the point that the main result in the baseline model is not driven by the exogeneity of investment opportunities. Instead, the discipline on ESG information disclosure is the key factor that determines real efficiency.

Next, I explore two additional extensions related to policy interventions. First, I examine how uncertainty regarding investors' ESG preferences influences the equilibrium. In reality, such uncertainty may arise from factors like anti-ESG policies, such as the Texas anti-ESG laws introduced in 2021, which might flip market participants' preferences, making ESG companies unfavorable to investors. It might seem intuitive to think that it would diminish the incentive for undertaking ESG activities, as preference uncertainty naturally reduces expected compensation from ESG investors. However, my model suggests that in situations where greenwashing is prevalent, a slight degree of preference uncertainty could benefit actual green firms because it could reduce greenwashing. Second, I discuss the impact of direct incentivization, i.e., a portion of managers' compensation is tied to actual ESG performance. As an example, one can think of this as integrating ESG criteria into executive compensation through clawback policies, thereby linking executive compensation to the achievement of long-term ESG goals. I show that direct incentivization not only directly affects the manager's real decision but also complements market discipline by reducing greenwashing motives.

Related Literature My paper adds to the growing literature on ESG investing and its real impact. On the empirical side, many studies report contradictory findings. De Angelis et al. (2022), Liang et al. (2022), Gantchev et al. (2022), and Flammer (2021) show that ESG investing generates substantial real impact and motivates firms to adopt ESG practices, while Berg et al. (2022) and Duchin et al. (2022) show that ESG investing does not have a significant effect on disciplining firms. On the theoretical side, most existing models (e.g., Heinkel et al., 2001; Chowdhry et al., 2019; Pástor et al., 2021; Oehmke and Opp, 2022; Edmans et al., 2022, etc.) show that when investors internalize more social benefits and costs (or the share of ESG investors increases), more efficient social outcomes can be

induced. However, my model emphasizes the unintended consequence of ESG investing: higher intensity of ESG preference might backfire and harm real efficiency when information discipline is weak. Moreover, in their models, the information quality about ESG practices is either irrelevant or taken as exogenous, but in my model, it is endogenous to the extent of greenwashing by brown firms, which drives the key result.

My paper also closely relates to the growing literature on ESG information disclosure. On the one hand, empirical evidence suggests that ESG investors induce more voluntary disclosure related to ESG (Ilhan et al., 2023; Flammer et al., 2021). On the other hand, empirical studies show that the quality of ESG disclosure is questionable and greenwashing is prevalent (Baker et al., 2023; Bailey et al., 2022; Liang et al., 2022).[7] Specifically, there are significant discrepancies between firms' disclosure and actual ESG practices, and firms may selectively release favorable information due to the absence of standards and frameworks for ESG disclosure. Another important source of ESG information is ESG rating. Berg et al. (2022), Berg et al. (2021), Serafeim and Yoon (2022), and Christensen et al. (2022) analyze the disagreement among ESG ratings and how it affects asset prices and predicts future news. Avramov et al. (2022) theoretically analyzes how ESG rating uncertainty affects market risk premium and alpha in an asset pricing framework. Distinct from existing literature that either considers ESG real effects or information disclosure only, my paper jointly considers incentives for real ESG activities and ESG disclosure, and my model emphasizes the interaction between greenwashing and real efficiency.

From a theoretical perspective, my paper relates to the broad literature on information manipulation and earnings management. Particularly, Goldman and Slezak (2006) analyzes how stock-based compensation induces managers to exert productive effort but also to misreport performance. One key difference between my model and their model is that in their model market participants can perfectly predict the agent's equilibrium choices and they

[7]My model can also be applied to analyze the greenwashing of ESG funds, which is highlighted by many empirical papers (e.g., Kim and Yoon, 2023; Raghunandan and Rajgopal, 2022; Gibson Brandon et al., 2022; Liang et al., 2022).

derive a signal-jamming equilibrium (e.g., Fudenberg and Tirole, 1986; Stein, 1989, etc.), while in my model there is uncertainty about the agent's type (green vs. brown). In this sense, my model is more closely connected to the earnings management literature in which the equilibrium choice is not perfectly predictable (e.g., Dye, 1988; Fischer and Verrecchia, 2000, etc.). More recently, Beyer and Guttman (2012) also considers the joint decision of real investment and voluntary disclosure, but the manager can only choose either no disclosure or truthful disclosure. Instead, in my model, the key channel is that brown firms misreport and pool with green firms, which decreases incentives for green investment. More importantly, my model features a manipulable externality fundamental versus a non-manipulable financial fundamental. The dilemma is that when market participants value the externality fundamental more in order to incentivize socially optimal real decisions, it inevitably induces more information manipulation, which may undermine real efficiency unexpectedly.

In addition, my paper can connect to the literature that explores the relationship between ESG criteria in CEO compensation and firms' ESG performance. Gillan et al. (2021) and Bebchuk and Tallarita (2022) provide a comprehensive overview, suggesting that while incorporating ESG metrics in compensation could incentivize ESG activities, it may also have detrimental effects on welfare by exacerbating severe agency problems. A growing body of empirical evidence, including studies by Cohen et al. (2022) and Berrone and Gomez-Mejia (2009), indicates that environmental governance mechanisms and ESG metrics in executive compensation contracts can enhance firms' ESG performance.

2 Model

The model has three dates, $t \in \{1, 2, 3\}$. Figure 1 describes the timing of actions and events in the model. Firms have no project in place ex-ante. At $t = 1$, there are two types of firms, and the firm type is denoted by $\theta \in \Theta := \{G, B\}$, with a share of π and $1 - \pi$ respectively: $\theta = G$ represents that the firm gets a green investment opportunity, and $\theta = B$ represents

that the firm gets a brown investment opportunity. The firm manager must decide whether to take the investment opportunity or not, denoted by $I \in \{0, 1\}$: if the manager invests in the project ($I = 1$), it generates externality value e_θ with an NPV $-k_\theta e_\theta$; if the manager keeps the status quo ($I = 0$), both the externality value and NPV is 0. Specifically, I define $v_e(\theta, I)$ as the externality value of the project as follows:

$$v_e(\theta, I) = \begin{cases} e_G & \text{if } \theta = G, I = 1 \\ e_B & \text{if } \theta = B, I = 1, \\ 0 & \text{if } I = 0, \end{cases} \tag{1}$$

and the NPV of the project is defined as $v(\theta, I) = -k_\theta v_e(\theta, I)$. In addition, I assume $k_\theta \in [0, 1]$, so the green investment has a positive social value, and the brown investment has a negative social value: for a manager who only cares about financial value, it is optimal to forgo the green investment and take the brown investment; for a social planner who maximizes the social value, it is efficient to take the green investment and forgo the brown investment.

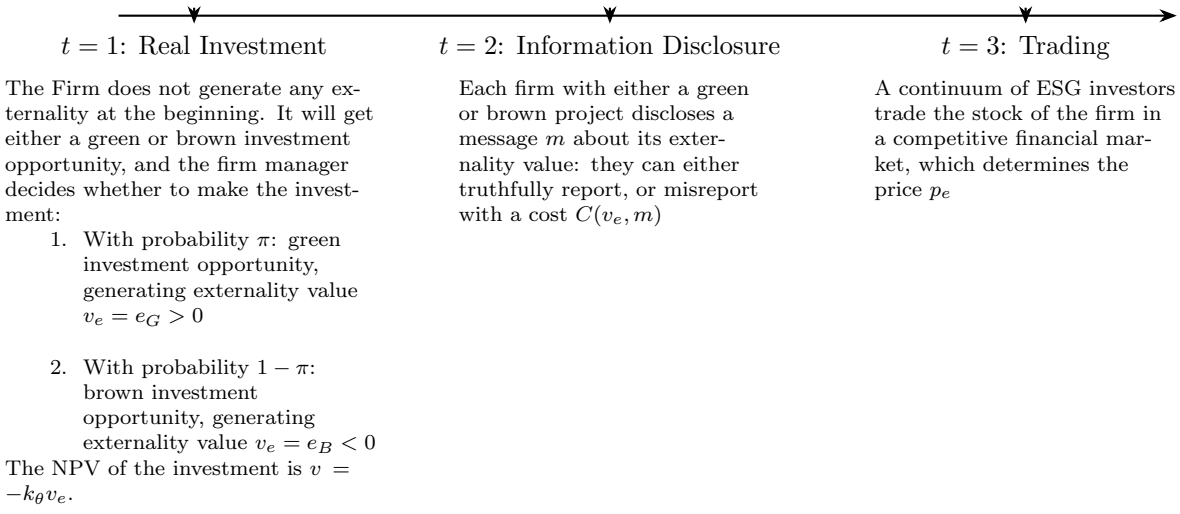

Figure 1: Timeline

Next, I specify the real investment problem of the manager. The manager does not value

the externality value directly. Instead, he cares about his compensation from creating the externality value, which is determined by the stock price realized at the end of $t = 3$ as well as the NPV of the investment and any disclosure costs. Thus, the manager's problem at $t = 1$ is as follows:

$$\max_{I\in\{0,1\}} U(\theta, I, m) = p_e - C(v_e(\theta, I), m) + v(\theta, I), \tag{2}$$

where p_e is the stock price realized at the end of $t = 3$, and $C(v_e, m)$ is the disclosure cost incurred at $t = 2$. I call $t = 1$ the real investment stage. After the real investment is made, investors can observe whether the firm has a project in place, but they do not know the type of the project.

At $t = 2$, if the firm has either a green or brown project, then the manager can disclose how much externality is created by sending a message $m(v_e) \in M = \{e_G, e_B\}$ to investors. For example, the manager can issue a corporate social responsibility (CSR) report to claim how well the firm achieves social goals. However, the manager is able to manipulate the information in CSR reports and overstate the positive externality, which is known as "greenwashing". Specifically, I assume the manager can truthfully disclose the externality value at no cost, or manipulate the message at a cost c, i.e., the disclosure cost function is $C(v_e, m) = c \cdot \mathbb{1}_{\{m\neq v_e\}}$. We can think about c as the ex-post penalty for manipulating a CSR report: for example, with probability q, it can be verified ex-post that the firm misled investors in its CSR report, which incurs a fixed cost D (e.g., penalties from regulators, downward stock price reactions, etc.), so we have manipulation cost $c = qD$. At this stage, the manager's disclosure problem is:

$$\max_{m(v_e)} U_2(v_e, m(v_e)) = \max_{m(v_e)} \{p_e - C(v_e, m(v_e))\}. \tag{3}$$

I call $t = 2$ the investment stage. I use $\widehat{U}_2(v_e^*) = a_e^* - C(v_e^*, m^*(v_e))$ to denote the utility (at the disclosure stage) of creating externality v_e^* in equilibrium.

At $t = 3$, a competitive financial market opens and a continuum of ESG investors trade the stock of the firm conditional on the message disclosed at $t = 2$. ESG investors derive utility from holding green stocks and disutility from holding brown stocks following the literature on ESG investing (e.g., Pástor et al., 2021). The utility function of each ESG investor[8] is

$$u_i(q_i) = (\beta v_e - p_e)q_i,$$

where β is the intensity of ESG preference, v_e is the externality value of the project, p_e is the market-clearing price, q_i is the position he takes in the market. Note that the intensity of ESG preference β measures how much value the ESG investor assigns to each unit of externality the firm created (in reality, β can be thought of as the average ESG preference of investors, the total share of ESG investors, etc.). Alternatively, in the context of active engagement and governance, p_e can be thought of as ESG shareholders setting monetary compensation for the manager for achieving ESG goals: for each unit of positive (negative) externality value perceived by the investors, the manager gets β as the monetary payoff.

The investor's expected utility upon receiving the firm's disclosure m is

$$\mathbb{E}[u_i|m] = (\beta\mathbb{E}[v_e|m] - p_e)q_i.$$

Thus, the market-clearing price is

$$p_e = \beta\mathbb{E}[v_e|m] \tag{4}$$

The payoff to the manager is realized at the end of this period, and I call $t = 3$ the trading stage.

The solution concept of the model is a Perfect Bayesian Equilibrium.

[8]In the baseline model, I assume ESG investors only care about externality value to shut down the channel that they might infer the externality value from the information on financial value. In Section 4, I analyze the extension in which ESG investors care about both externality value and financial value.

Definition 1 *The ESG investors' trading strategy, the firms' investment strategy, and the firms' disclosure strategy constitute an equilibrium if*

(1) Given the firms' investment and disclosure strategy, the ESG investors trade the stock at a price such that they are break-even conditional on the message disclosed by the firms (which generates a competitive market pricing function).

(2) Given the competitive pricing function of the market, firm managers (with either a green or brown investment opportunity) choose the disclosure and investment strategy to maximize their compensation.

(3) The ESG investors update their beliefs according to Bayes' rule whenever possible.

3 Equilibrium

3.1 Disclosure Stage

In order to solve for the Perfect Bayesian Equilibrium, I first guess an equilibrium and then verify it. Suppose that at the beginning of the disclosure stage, there is a share α of green firms (i.e., firms that make the green investment), and a share $1-\alpha$ of brown firms (i.e., firms that make the brown investment). Note that in this section, I analyze the case where there are both green and brown firms, i.e. $\alpha \in (0,1)$. The reason is that the equilibrium in which $\alpha \in \{0,1\}$ is trivial as the ESG investors know the disclosing firm's type with certainty and thus there is no space for information manipulation.

First, note that in any equilibrium the manager always truthfully reports if the firm generates positive externality e_G, as the manager will not pay the manipulation cost to claim to be brown. Thus, there are three possible equilibria: "full disclosure", "greenwashing" (the brown firm always claims to be green), and "partial greenwashing" (the brown firm claims to be green with some probability). I define q as the probability that a brown firm engages

in greenwashing. Note that q can also be interpreted as the share of "greenwashing" firms if we consider a continuum of firms instead of a single firm.

The following lemma summarizes the equilibrium in this stage:

Lemma 1 *Suppose there is a share α of green firms and a share $1-\alpha$ of brown firms at the disclosure stage, then the disclosure strategy in equilibrium is determined by the ESG preference to manipulation cost ratio $\frac{\beta}{c}$:*

- *If $\frac{\beta}{c} \leq \frac{1}{e_G - e_B}$, then the equilibrium is "full disclosure", i.e., $m(e_G) = e_G$, $m(e_B) = e_B$;*
- *If $\frac{\beta}{c} \geq \frac{1}{\alpha}\frac{1}{e_G - e_B}$, then the equilibrium is "greenwashing", i.e., $m(e_G) = m(e_B) = e_G$ (the off-path-belief and equilibrium refinement rule is specified in the proof);*
- *if $\frac{\beta}{c} \in (\frac{1}{e_G - e_B}, \frac{1}{\alpha}\frac{1}{e_G - e_B})$, then the equilibrium is "partial greenwashing", i.e., $m(e_G) = e_G$, $m(e_B) = e_B$ with probability $1-q$ and $m(e_B) = e_G$ with probability q (q is specified in the proof).*

Figure 2 depicts the full picture of equilibrium for any value of $\frac{\beta}{c}$. The equilibrium disclosure strategy depends on the compensation gap between reporting positive externality and reporting negative externality (which depends on the intensity of ESG preference β), relative to the manipulation cost c. If this gap is smaller than the information manipulation cost (i.e., the preference to cost ratio $\frac{\beta}{c}$ is very small), then brown firms will truthfully report, and there is no greenwashing in equilibrium. As the compensation gap becomes larger and equals the manipulation cost (i.e., the preference to cost ratio $\frac{\beta}{c}$ is in the intermediate region), then a proportion of brown firms engage in greenwashing. The probability of greenwashing q increases with $\frac{\beta}{c}$ such that brown firms are indifferent between greenwashing and truthful reporting. Note that in such partial greenwashing equilibrium, the effect of a larger share of greenwashing firms q dominates the effect of larger β, so the compensation to green firms decreases as β increases [9]. Last, if the compensation gap becomes large enough such that it

[9]Note that the effect of a larger share of greenwashing firms q on compensation does not always dominate under a more general setting. For example, in the extension with heterogeneous information manipulation cost shown in Section 5.1, this effect is small when the intensity of ESG preference β is small.

exceeds the manipulation cost even when all brown firms engaged in greenwashing (i.e., the preference to cost ratio $\frac{\beta}{c}$ is very large), then all the brown firm engages in "greenwashing", and thus two types of firms pool together and the market reacts with a pooling price of $\beta\bar{e}$ [10].

Corollary 1

The probability of greenwashing q is (weakly) increasing in the ESG preference to manipulation cost ratio $\frac{\beta}{c}$.

This corollary summarized the property about the share of greenwashing firms q in the analysis above. It implies the concerns about greenwashing when investors increasingly value ESG outcomes (β increases), while the discipline on disclosing such outcomes is still weak (c remains small).

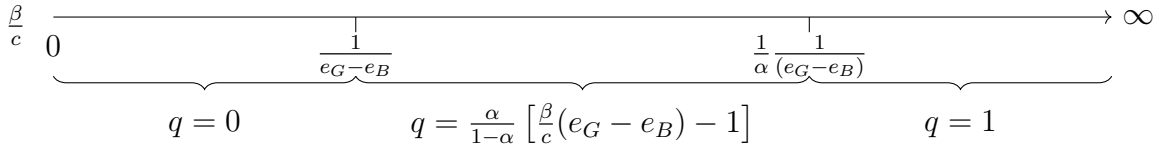

Figure 2: Disclosure Equilibrium

Next, I derive the property of the equilibrium payoff at the disclosure stage, when the firm has already invested in a green or a brown project at the beginning of this stage. Define $\widehat{U}_2(e_G)$ as the compensation for producing positive externality at the disclosure stage, $\widehat{U}_2(e_B)$ as the penalty for producing negative externality at the disclosure stage, and define $\bar{e} = \pi e_G + (1-\pi)e_B$ as the average externality value when all brown firms and green firms pool together. If $\frac{\beta}{c} \leq \frac{1}{e_G-e_B}$, since all firms truthfully report, $\widehat{U}_2(e_G) = \beta e_G$, $\widehat{U}_2(e_B) = \beta e_B$. If $\frac{\beta}{c} \geq \frac{1}{\alpha}\frac{1}{e_G-e_B}$, since all brown firms engage in greenwashing (i.e., all firms claim to be green), the market gives an average price $\bar{e}$ to firms claim to be green. Thus, $\widehat{U}_2(e_G) = \beta\bar{e}$, $\widehat{U}_2(e_B) = \beta\bar{e} - c$. In the intermediate region where $\frac{\beta}{c} \in (\frac{1}{e_G-e_B}, \frac{1}{\alpha}\frac{1}{e_G-e_B})$, a proportion

[10]Note that in this pooling equilibrium, I specify the off-equilibrium-path belief such that ESG investors assign probability 1 to brown firms if they receive message $m = e_B$. See the proof of Lemma 1 for more discussions on the choice of off-equilibrium-path beliefs.

$q = \frac{\alpha}{1-\alpha}\left[\frac{\beta}{c}(e_G - e_B) - 1\right]$ of brown firms engage in greenwashing. Since brown firms are indifferent between truthfully reporting and misreporting, $\widehat{U}_2(e_B) = \beta e_B$. The green firms report green but do not incur the information manipulation cost, so their compensation is $\widehat{U}_2(e_G) = \widehat{U}_2(e_B) + c = \beta e_B + c$. Formally, the function $\widehat{U}_2(e_G)$ and $\widehat{U}_2(e_B)$ are shown below: the equilibrium penalty for negative externality is

$$\widehat{U}_2(e_B) = \begin{cases} \beta e_B & \text{if } \frac{\beta}{c} \leq \frac{1}{\alpha}\frac{1}{e_G - e_B}, \\ \beta \overline{e} - c & \text{if } \frac{\beta}{c} > \frac{1}{\alpha}\frac{1}{e_G - e_B}, \end{cases} \tag{5}$$

and the equilibrium compensation for positive externality is

$$\widehat{U}_2(e_G) = \begin{cases} \beta e_G & \text{if } \beta \leq \frac{c}{e_G - e_B}, \\ \beta e_B + c & \text{if } \beta \in (\frac{c}{e_G - e_B}, \frac{c}{\alpha(e_G - e_B)}), \\ \beta \overline{e} & \text{if } \beta \geq \frac{c}{\alpha(e_G - e_B)}. \end{cases} \tag{6}$$

Corollary 2 *The compensation for positive externality $\widehat{U}_2(e_G)$ and penalty for negative externality $\widehat{U}_2(e_B)$ can be non-monotonic in the intensity of ESG preference β:*

1. *If $\beta \leq \frac{c}{e_G - e_B}$, $\widehat{U}_2(e_G)$ is increasing in β and $\widehat{U}_2(e_B)$ is decreasing in β;*
2. *If $\beta \in (\frac{c}{e_G - e_B}, \frac{c}{\pi(e_G - e_B)})$, $\widehat{U}_2(e_G)$ and $\widehat{U}_2(e_B)$ are both decreasing in β;*
3. *If $\beta \geq \frac{c}{\pi(e_G - e_B)}$, $\widehat{U}_2(e_G)$ and $\widehat{U}_2(e_B)$ are both increasing (decreasing) in β when $\overline{e} > 0$ ($\overline{e} < 0$).*

Note that the compensation for creating positive externality value (or penalty for creating negative externality value) is non-monotonic in the intensity of investors' ESG preference β. Specifically, when β is small, investors can distinguish green firms from brown firms as firms truthfully disclose, thus the compensation for green firms is increasing in β. When β becomes larger, brown firms are more likely to manipulate disclosure and claim to be green,

so the compensation for green firms is decreasing in β as greenwashing becomes more severe. Last, when β is large enough such that brown firms always engage in greenwashing, this relationship depends on the average externality value created: if $\bar{e} > 0$ ($\bar{e} < 0$), then the compensation is increasing (decreasing) in β.

3.2 Investment Stage

Given the equilibrium strategy and the resulting compensation at the disclosure stage, we can derive the equilibrium strategy at the investment stage. First of all, I impose some assumptions on the parameters such that we can focus on interesting cases:

Assumption 1 $\bar{e} = \pi e_G + (1-\pi)e_B < 0$

This assumption ensures that when all brown firms engage in greenwashing and pool with green firms, the average externality value is negative. Consequently, green investment cannot be incentivized when there is too much greenwashing. Analyzing the case in which $\bar{e} = \pi e_G + (1-\pi)e_B > 0$ does not change my key results [11].

Assumption 2 $0 \leq \beta < k_B$.

Assumption 3 $0 < k_G < k_B \leq 1$.

The two assumptions above both indicate that the positive NPV of brown investment is large enough. Specifically, Assumption 2 guarantees that in the regions of β that we analyze, the brown investment is always taken (i.e., $I(B) = 1$) [12], so we can analyze greenwashing and its interactions with green investment. Otherwise, the equilibrium becomes trivial when there is no brown investment. Assumption 3 says the ratio of negative NPV of green investment to its externality value k_G is smaller than k_B, such that we can find a non-empty interval of β in which there are both green and brown investments.

[11] A graphical illustration about intervals of green investment when $\bar{e} > 0$ is shown in Appendix B.

[12] To see this, note that the largest penalty to brown investment from the market is βe_B, while the NPV gain from brown investment is $k_B e_B$. If $\beta < k_B$, then it is always beneficial for the firm manager to take brown investments.

In this section, we focus on the equilibrium in which $I(G) \in \{0, 1\}$ and $I(B) = 1$. The reason is as follows: first, as mentioned above, Assumption 2 guarantees that the brown investment is always made in equilibrium such that there are brown firms at the disclosure stage. Second, we eliminate any equilibrium in which $I(G) \in (0, 1)$ because the mixed strategies are not stable in such games with strategic complementarity (see Crawford, 1989; Fudenberg and Tirole, 1986; Fudenberg and Kreps, 1993; Echenique and Edlin, 2004, etc.). More detailed discussions about equilibrium refinement rules and the choices of off-equilibrium-path beliefs are shown in the Appendix.

Next, I derive conditions under which the green investment equilibrium (i.e.,$I(G) = I(B) = 1$) exists. Note that the green investment is made if the compensation from the market is larger than the NPV loss, i.e., $\widehat{U}_2(e_G) \geq k_G e_G$. We can get the following proposition by solving this inequality:

Proposition 1 *The intervals of β in which green investment is made depend on the level of information discipline:*

- *Weak information discipline:*

  If $c < k_G(e_G - e_B)$, the green investment is never made.

- *Intermediate information discipline:*

  If $c \in [k_G(e_G - e_B), k_G e_G - k_B e_B]$, the green investment is made if $\beta \in [k_G, \frac{k_G e_G - c}{e_B}]$.

- *Strong information discipline:*

  If $c > k_G e_G - k_B e_B$, the green investment is made if $\beta \in [k_G, k_B)$.

I analyze the conditions for green investment under three regimes of information discipline respectively. The most interesting case is when there is an intermediate level of information discipline (i.e., $c \in [k_G(e_G - e_B), k_G e_G - k_B e_B)]$, as shown in Figure 3. When β is small, all firms truthfully report their externality, so the compensation to green investment increases with β, as ESG investors internalize more externality; When β reaches the first threshold

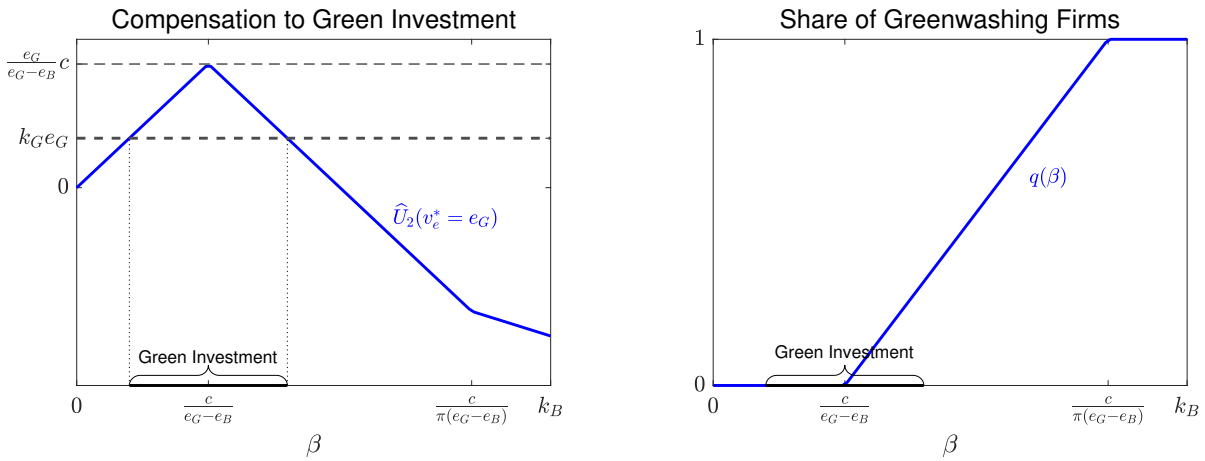

Figure 3: Green Investment When $c \in [k_G(e_G - e_B), k_G e_G - k_B e_B]$

$\frac{c}{e_G-e_B}$, brown firms start to engage in greenwashing. Moreover, as β increases, the share of greenwashing firms among all brown firms q keeps increasing, which lowers investors' expectations about the externality value of each firm that sends a green message $m = e_G$. Last, when β reaches the second threshold $\frac{1}{\pi}\frac{c}{e_G-e_B}$, all brown firms engaged in greenwashing. Because of this hump-shaped compensation function (w.r.t. β), the green investment is made only when $\beta \in [k_G, \frac{k_G e_G - c}{e_B}]$, i.e., β is large enough such that the compensation to green investment can cover the financial cost of investment, but it should not be too large such that too much greenwashing undermines the information quality.

If the information discipline is very weak (i.e., $c < k_G(e_G - e_B)$), then we can never incentivize green investment through increasing β, as shown in Figure 4. Note that the compensation to green investment is maximized at the threshold of β that the manager is about to engage in greenwashing (i.e., $\beta = \frac{c}{e_G-e_B}$), and the maximum is $\frac{e_G}{e_G-e_B}c$. Thus, the level of information discipline c determines the maximum compensation that the competitive market can achieve: if the information discipline is weak, then the role of the financial market in incentivizing green investment is highly limited. On the other hand, if the information discipline is very strong (i.e., $c > k_G e_G - k_B e_B$) as shown in Figure 5, then green investment can be easily incentivized, since the share of greenwashing firms q is low even when β becomes large.

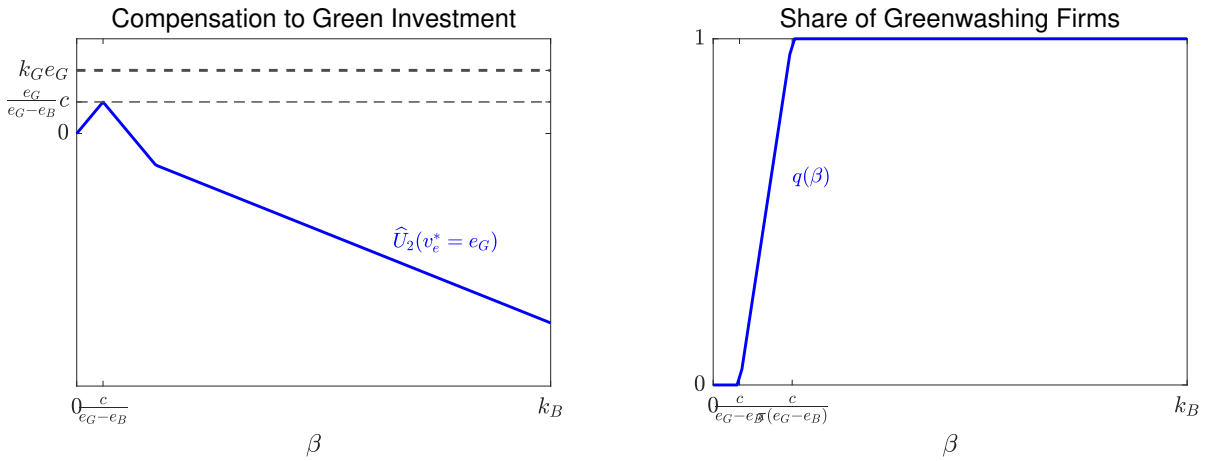

Figure 4: Green Investment When $c < k_G(e_G - e_B)$

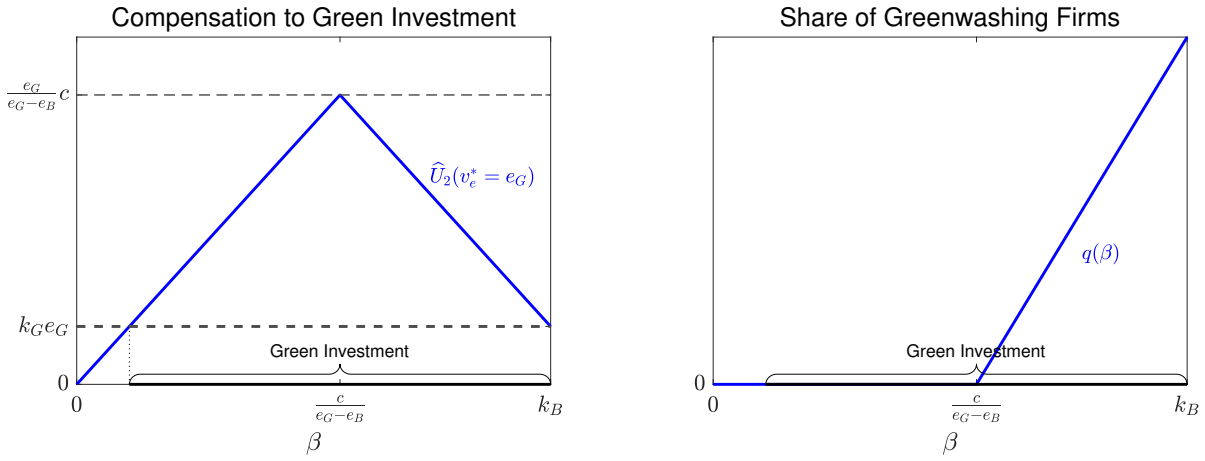

Figure 5: Green Investment When $c > k_G e_G - k_B e_B$

The results have very important policy implications: for those ESG outcomes that are hard to measure or have substantial disagreement (e.g., gender diversity), regulation on information disclosure is more crucial than increasing incentivization from ESG investors; for those ESG outcomes which are relatively more measurable but still subject to greenwashing concerns (e.g., long-term carbon emission objectives), we should be careful as more ESG investors can backfire and reduce real efficiency; for those ESG outcomes that are easy to measure (e.g., corporate governance issues), more ESG investors can always play a positive role.

4 Investors Learn from Financial Information

In reality, though ESG information is usually non-verifiable and thus subject to manipulation, financial information such as earnings announcements are much more reliable, as the fundamental is objective and verified by independent auditing agencies. This distinction between ESG information and financial information creates a new interaction between the two fundamentals in this section: a non-manipulable earnings announcement can change investors' prior beliefs about whether the company has allocated resources to ESG projects. For example, if a company does not meet earnings expectations in its announcement, this poor business performance might result from poor investment opportunities, or from engagement in ESG activities.

In this section, I will show that when ESG activities are highly likely to have a low NPV, firms might take brown projects even with negative NPV and engage in greenwashing to reap benefits from ESG investors. There is evidence that firms label themselves as ESG especially when they have poor business performance. For example, Flugum and Souther (2022) find that firm managers are more likely to claim ESG when earnings announcements fall short of market expectations. My model prediction is also consistent with some empirical findings that fund managers label themselves as ESG funds while underperforming in both actual

ESG performance and stock returns (e.g., Raghunandan and Rajgopal, 2022).

To analyze the interaction between ESG fundamentals and financial fundamentals, I impose four new setups/assumptions in this section as follows :

1. There is a public signal y revealing the realization of NPV, i.e., $y = v$. We can think about it as earnings announcements, which precisely reveals the financial fundamental and cannot be manipulated (exogenous).

2. Investors care about both NPV and ESG outcomes: ESG investor's utility function is

$$u_i(q_i) = (\beta v_e + v - p)q_i,$$

   which results in a market-clearing price:

$$p = \mathbb{E}[v|y] + \beta\mathbb{E}[v_e|y, m]$$

3. Firm managers' compensation depends only on the firm price, i.e., both NPV and ESG fundamentals enter his utility function through price: Firm manager's utility function is

$$U(\theta, I, m) = p - C(v_e(\theta, I), m)$$

4. There is a negative correlation between v_e and v, i.e., green investments are more likely to have negative NPV than brown investments:

$$v(G, 1) = \begin{cases} v_L & \text{w.p. } \rho, \\ v_H & \text{w.p. } 1 - \rho, \end{cases}$$

$$v(B, 1) = \begin{cases} v_H & \text{w.p. } \rho, \\ v_L & \text{w.p. } 1 - \rho, \end{cases}$$

where $v_H = -k_B e_B > 0 > v_L = -k_G e_G$. I define $\rho = Pr(sign(v) \neq sign(v_e)) > \frac{1}{2}$ as the correlation parameter of the ESG fundamental and financial fundamental: a large ρ means that a green (brown) project is highly likely to have a low (high) NPV.

In this case, if investors observe a firm with negative NPV ($v < 0$), they know that it is a green firm with probability $\pi\rho$ and a brown firm with probability $(1-\pi)(1-\rho)$. When all brown firms engage in greenwashing and fully pool with green firms, they get the compensation $\beta\mathbb{E}[v_e|y = v_L, m = e_G] = \beta\left[\frac{\pi\rho}{\pi\rho+(1-\pi)(1-\rho)}e_G + \frac{(1-\pi)(1-\rho)}{\pi\rho+(1-\pi)(1-\rho)}e_B\right]$. If the correlation parameter ρ is large, investors rationally infer that it is highly likely to be a green firm, so brown firms get large compensation by pooling with a large share of green firms. When the intensity of ESG preference β is also large, brown projects with negative NPV are taken if $\beta\mathbb{E}[v_e|y = v_L, m = e_G] - c + v_L > 0$. The result is summarized in the following proposition:

Proposition 2 *Suppose there is limited information discipline $c < e_G + v_L$:*

- *If $\beta = 0$, only brown investment with positive NPV is taken.*
- *If $\rho > \bar{\rho} = \frac{(1-\pi)(c-v_L-e_B)}{(1-\pi)(c-v_L-e_B)+\pi(e_G-c+v_L)}$, there exists β s.t. brown investments with negative NPV are made.*

Intuitively, if investors do not care about ESG outcomes, there is no information manipulation since the information about NPV is exogenous, so only brown investment with positive NPV is taken. However, if investors care about ESG outcomes, public disclosure (such as an earnings announcement) indicating negative NPV suggests that it is more likely to be a green investment (relative to the case in which positive NPV is disclosed). Thus, brown firms with negative NPV get larger compensation from ESG investors.

Corollary 3

- *The likelihood of greenwashing q is larger when the real investment has a negative NPV compared to when it has a positive NPV (i.e. $q_{|v<0} \geq q_{|v>0}$).*

- *The likelihood of greenwashing q is larger when the ESG fundamental and financial fundamental are more negatively correlated (i.e.,$\frac{\partial q}{\partial \rho} \geq 0$).*

Since brown firms with negative NPV are pooled with more green firms, they are also more likely to engage in greenwashing in order to get higher compensation from ESG investors, and this greenwashing incentive is strong when ESG activities are very likely to have negative NPV. This result rationalizes the empirical observations that companies publicly embrace ESG as a cover for poor business performance (e.g., Flugum and Souther, 2022).

5 Extensions: Generalized Models

In this section, I consider two extensions that generalize the assumptions of the baseline model, and I show that the key insight from the baseline model remains robust.

5.1 Heterogenous Information Manipulation Cost

In this section, I analyze the equilibrium when each manager i has a different information manipulation cost c_i, which is only known to himself. This is a more natural and realistic case than the baseline model since firm managers do not randomly choose to engage in greenwashing or not as part of a mixed-strategy equilibrium. Instead, their disclosure choice depends on their own information manipulation cost.[13] I assume the information manipulation cost c_i among firms with a brown investment opportunity has a cumulative distribution function $F(\cdot)$ (and a corresponding probability density function $f(\cdot)$). Besides, I relax Assumption 2 in the last section to analyze the equilibrium when the intensity of ESG preference β is large and brown investments are restrained.

Again, I use q to denote the share of greenwashing firms among firms with brown investment opportunities. Define $\alpha = \frac{\pi}{\pi+(1-\pi)q}$ as the share of green firms among firms taking

[13] I will focus on the pure-strategy equilibrium in this section. This extension is similar to the equilibrium choice in the baseline model, i.e., firm managers have different disclosure strategies.

investments, so $\mathbb{E}[v_e|m = e_G] = \alpha e_G + (1-\alpha)e_B$. As we have shown in the baseline model, green firms never misreport themselves as brown firms, so $\mathbb{E}[v_e|m = e_B] = e_B$.

5.1.1 Case I: Brown Investment is Always Made ($\beta < k_B$)

Note that the penalty to negative externality in any equilibrium is no more than $\beta|e_B|$ and the NPV gain from brown investment is $k_B|e_B|$, so the brown investment is always made when $\beta < k_B$.

First, I pin down the agent i^* who is indifferent between truthful reporting and misreporting,

$$\beta\mathbb{E}[v_e|m = e_G] - \beta\mathbb{E}[v_e|m = e_B] = c_{i^*}, \tag{7}$$

i.e., the difference in payoff (for a brown firm) between misreporting and truthful reporting should be equal to his information manipulation cost. It can be simplified to $c_{i^*} = \beta(e_G - e_B)\alpha$. Based on this indifference condition, for any brown firm manager i, if $c_i < c_{i^*}$, then he misreports; if $c_i > c_{i^*}$, then he truthfully reports. Thus, the share of greenwashing firms among firms with brown investment opportunities is $F(c_{i^*}) = F(\beta(e_G - e_B)\alpha)$. In equilibrium, this share should be equal to our initial guess q.

Lemma 2 *Suppose $F'(\cdot) > 0$. Given any value of β, there exists exactly one $q \in [0,1]$ as the equilibrium share of greenwashing firms (among firms with brown investment opportunities), characterized by the equation*

$$q = F(\beta(e_G - e_B)\alpha), \tag{8}$$

where $\alpha = \frac{\pi}{\pi+(1-\pi)q}$. Besides, q is strictly increasing in the intensity of ESG preference β.

Proposition 3 *Define $A = \frac{\bar{e}(\bar{\alpha})}{\pi(1-\pi)(-e_B)(e_G-e_B)}$, where $\bar{\alpha}$ and $\bar{q}$ satisfy $\bar{q} = F(k_B(e_G - e_B)\bar{\alpha})$ and $\bar{\alpha} = \frac{\pi}{\pi+(1-\pi)\bar{q}}$.*

- *There exists β_L s.t. $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} > 0$ as long as $\beta \in [0, \beta_L]$.*

- *Suppose there exists* $c < k_B(e_G - e_B)\bar{\alpha}$ *s.t.* $F'(c_i) > A$ *when* $c_i > c$. *There exists* β_H *s.t.* $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} < 0$ *as long as* $\beta \in [\beta_H, k_B]$.

The intuition is illustrated in equation (9): As investors value ESG outcomes more (β increases), it incentivizes the manager to internalize a larger proportion of externality value. However, it also induces more greenwashing meanwhile, which decreases the expected externality value of firms claiming to be green. More importantly, the positive effect is large when β is small since changing the proportion of internalized value matters more when the expected value is large; the negative effect is large when β is large since the change in expected value matters more when the manager internalizes a large proportion of it.

$$\begin{aligned} \frac{\partial \widehat{U}_2(e_G)}{\partial \beta} &= \frac{\partial\{\beta[\alpha e_G + (1-\alpha)e_B]\}}{\partial \beta} \\ &= \underbrace{[\alpha e_G + (1-\alpha)e_B]}_{>0\text{:internalized externality value} \uparrow} + \underbrace{\beta(e_G - e_B)\frac{\partial \alpha}{\partial q}\frac{\partial q}{\partial \beta}}_{<0\text{:greenwashing firms} \uparrow} \end{aligned} \tag{9}$$

Corollary 4

- *If* $F(k_G(e_G - e_B)) = 0$, *then there exists an interval* $[k_G, k_G + \delta_G]$ *s.t. the green investment is made when* $\beta \in [k_G, k_G + \delta_G]$.

- *If* $F(k_G e_G - k_B e_B) > \bar{q}$, *then there exists an interval* $[k_B - \delta_B, k_B]$ *s.t. the green investment is not made when* $\beta \in [k_B - \delta_B, k_B]$.

The results above suggest that the key insight from the baseline model can be obtained under a wide class of distributions of information manipulation cost c. Figure 6 shows the change of endogenous equilibrium variables as the intensity of ESG preference β increases when information manipulation cost c follows a uniform distribution. Specifically, more greenwashing decreases the reliability of ESG disclosure, and the equilibrium compensation to green investments is undermined especially when the intensity of ESG preference β is large.

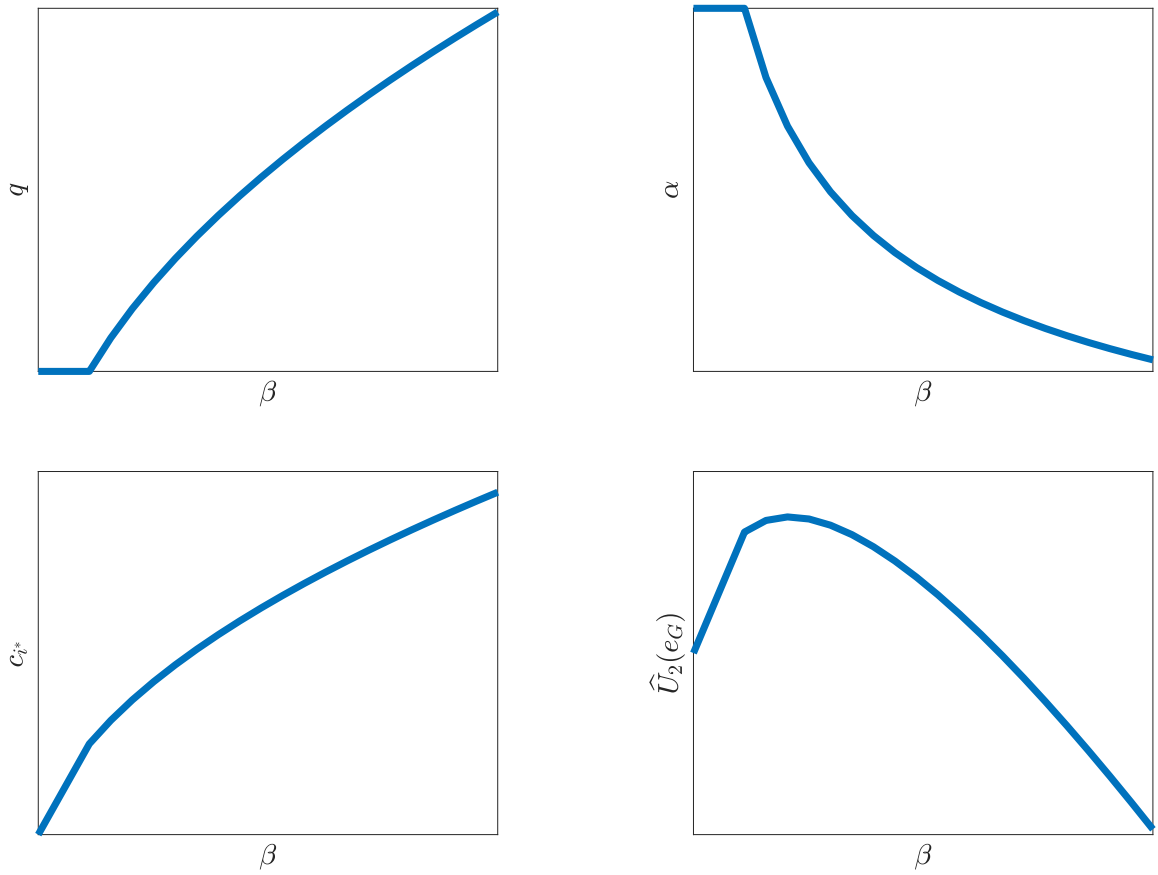

Figure 6: The figure shows the change of endogenous equilibrium variables $\{q, \alpha, c_{i^*}, \widehat{U}_2(e_G)\}$ as the intensity of ESG preference β increases. The information manipulation cost c follows a uniform distribution on $[2, 10]$, i.e., $c \sim U[2, 10]$. Other parameter values are $e_G = 10$, $e_B = -10$, $\pi = 1/3$, $k_G = 0.1$, $k_B = 0.9$.

5.1.2 Case II: Brown Investment is Restrained ($\beta > k_B$)

For completeness of the model, we derive the equilibrium when brown investment is restrained (i.e., $\beta > k_B$). Note that though the key result can also be obtained under this parameter region, the previous parameter region (in which brown investment is always taken) delivers the main insight more clearly and is a closer reflection of reality: I focus on the channel that firms with brown project in place mimic green firms and make ESG disclosure less reliable.

First, note that brown firms will never truthfully report when $\beta > k_B$ because the payoff from making brown investment and reporting truthfully is $(\beta - k_B)e_B < 0$, which is less than the payoff from keeping the status quo.

Lemma 3 *If $\beta > k_B$, brown firms never truthfully report in any equilibrium.*

Thus, firms with brown investment opportunities will either engage in greenwashing or keep the status quo, depending on their information manipulation cost. An immediate implication from Lemma 3 is that if the intensity of ESG preference β increases and exceeds the threshold k_B, there will be a sudden decrease in brown real investments.

Similarly, we pin down the firm manager i^* who is indifferent between greenwashing and keeping the status quo through the following equation:

$$\beta \mathbb{E}[v_e | m = e_G] - k_B e_B - c_{i^*} = 0, \tag{10}$$

i.e., $c_{i^*} = \beta[\alpha e_G + (1-\alpha)e_B] - k_B e_B$. Based on this indifference condition, for any brown firm manager i, if $c_i < c_{i^*}$, then he engages in greenwashing; if $c_i > c_{i^*}$, then he keeps the status quo. Thus, the share of greenwashing firms among firms with brown investment opportunities is $F(c_{i^*}) = F(\beta[\alpha e_G + (1-\alpha)e_B] - k_B e_B)$.

Proposition 4 *When $\beta > k_B$, the equilibrium share of greenwashing firms q and the equilibrium payoff to green investment $\widehat{U}_2(e_G)$ can be either increasing or decreasing in the intensity of ESG preference β, depending on the payoff to green firms at the threshold $\bar{q}$.*

1. *If* $\bar{e}(\bar{\alpha}) > 0$, *then* $\frac{\partial q}{\partial \beta} > 0$ *and* $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} > 0$ *when* $\beta \in (k_B, 1]$.

2. *If* $\bar{e}(\bar{\alpha}) < 0$, *then* $\frac{\partial q}{\partial \beta} < 0$ *and* $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} < 0$ *when* $\beta \in (k_B, 1]$.

Proposition 4 shows that if the intensity of ESG preference β further increases after the brown investment is restrained, there can be two cases depending on the average externality value of firms that claim to be green. If the average firm has a positive externality value, then more brown firms prefer to engage in greenwashing and pool together as β increases, rather than keeping the status quo. On the contrary, If the average firm has a negative externality value, then more brown firms prefer keeping the status quo instead of engaging in greenwashing as β increases.

5.2 Endogenous Investment Opportunities

Given the analysis in the baseline model, one may think that one critical assumption leading to the main result is that the types of investment opportunities are exogenous because firm managers tend to pick more green investment opportunities when investors attach higher value to green outcomes. However, in this section, I will show that this is not the case. Instead, the discipline on ESG information disclosure is still the key factor that determines real efficiency even if we consider endogenous investment opportunities.

Specifically, I consider the alternative model as follows: after the brown investment is made, brown firm managers can determine whether to adopt green technology and transfer the brown project in place into a green project. It is closely connected to reality: firms have different levels of pollution ex-ante, with the green investment cost and greenwashing cost determined by the ex-ante pollution level. In equilibrium, we can observe three kinds of firms: taking the technology revolution and becoming green firms; keeping the status quo and admitting it is still brown; keeping the status quo but claiming to be green (greenwashing). Our key insights from the baseline model still hold, i.e., more greenwashing firms undermine the information quality and decrease the incentive of becoming green firms in the first place.

5.2.1 A Simple Model: Adoption of Green Technology by Brown Firms

In this subsection, we consider the sustainable investment decision of brown firms when they cannot commit to it. Specifically, we assume that after brown investments are taken, brown firms also determine investment decisions $T(B) \in \{0,1\}$: $T(B) = 0$ represents that the brown firm keeps the brown project in place, which generates negative externality e_B with 0 private cost; $T(B) = 1$ represents that the brown firm adopts the green technology and transforms the brown projects into green projects, which generates positive externality e_G with a private cost d_B, and we assume $e_G - k_B e_B > d_B$. [14] Thus, the externality value of brown firm is

$$v_e(B,1,T) = \begin{cases} e_G & \text{if } T = 1, \\ e_B & \text{if } T = 0. \end{cases} \tag{11}$$

When brown firms do not have commitment devices for their investment strategies, we have exactly the same equilibrium outcomes as the baseline model:

Proposition 5 *If the cost of adopting green technology is larger than the information manipulation cost, i.e., $d_B > c$, then the green technology is never invested by the brown firm in equilibrium for any value of the intensity of ESG preference β.*

Note that if $\beta > \max\left\{\frac{d_B}{e_G - e_B}, \frac{d_B - c}{e_G - \bar{e}}\right\}$, the brown firm gets strictly higher payoff by adopting the green technology than keeping the status quo if investors can observe their investment strategies because

$$\underbrace{\max\{\beta e_B, \beta\bar{e} - c\}}_{\substack{\text{Investors believe that} \\ \text{the brown firm keeps the status quo}}} < \underbrace{\beta e_G - d_B}_{\substack{\text{Investors believe that} \\ \text{the brown firm adopts the green technology}}}.$$

Surprisingly, brown firms never adopt the green technology in equilibrium. The intuition is as follows: if investors believe that the green technology is adopted by the brown firm, then the firm manager can always deviate to keeping the brown project in place and then engaging

[14] Note that $d_B < e_G - k_B e_B$ ensures that the green investment of brown firms is socially efficient.

in greenwashing. Thus, investors do not believe that the firm will invest in green technology and the firm will not do so in equilibrium. Due to the lack of commitment devices, the brown firms will not adopt green technology, even though investors are willing to give very high rewards for positive externality. Thus, the first-best outcome cannot be achieved in equilibrium.

5.2.2 A General Model with Continuous Types of Brown Firms

In this subsection, I propose a more general model with continuous types of brown firms, to illustrate the point that the main result still holds when investment opportunities are endogenized.

Suppose at the beginning of $t = 1$, there is a continuum of brown firms, with type $\theta \in [e_B, 0]$ ($e_B < 0$) denoting the externality value of the brown project in place. Each firm brown decides whether to invest in green technology, which could turn its brown project into a green project with externality $e_G > 0$. Formally, each firm manager decides $I(\theta) \in \{0, 1\}$, which determines the externality value:

$$v_e(B, I) = \begin{cases} e_G & \text{if } I = 1, \\ e_B & \text{if } I = 0. \end{cases}$$

The cost (NPV) of the green investment is $C_I(\theta)$, which satisfies $C_I'(\theta) < 0$, i.e., the cost of adopting the green technology is higher for firms with more pollution ex-ante.

At $t = 2$, each firm reports the externality value of its project in place, i.e., it can send a message $m(\theta) \in \{e_G, \theta\}$. The cost of misreporting is $C_m(\theta)$, which satisfies $C_m'(\theta) < 0$, i.e., the cost of misreporting as green firms is higher for firms with more pollution ex-ante. Besides, I impose a critical single-crossing condition on $C_I(\theta)$ and $C_m(\theta)$:

Assumption 4 *Single-crossing condition: there exists $\hat{\theta} \in (e_B, 0)$ s.t. $C_I(\theta) > C_m(\theta)$ if $\theta < \hat{\theta}$ and $C_I(\theta) < C_m(\theta)$ if $\theta > \hat{\theta}$.*

It means that for firms with very high pollution ex-ante, the cost of adopting green technology and turning itself into green firms is higher than the cost of engaging in greenwashing; for firms with low pollution ex-ante, the cost of engaging in greenwashing is higher than the cost of adopting green technology and turning itself into green firms. With this assumption, only firms with $\theta > \hat{\theta}$ may take green investment, and only firms with $\theta < \hat{\theta}$ may engage in greenwashing in equilibrium.

Assumption 5 $C_I'(\theta) < -1$, $C_m'(\theta) > -\beta$.

This assumption restricts the derivatives of cost functions. It ensures that we could get a monotonic equilibrium in the sense that the equilibrium could be characterized by two thresholds (e_m, e_I) [15]: $e_m \in [e_B, \hat{\theta}]$ s.t. $I(\theta) = 0$ and $m(\theta) = e_G$ if $\theta \leq e_m$, i.e., e_m is the threshold below which greenwashing is induced; $e_I \in [\hat{\theta}, 0]$ s.t. $I(\theta) = 1$ and $m(\theta) = e_G$ if $\theta \geq e_I$, i.e., e_I is the threshold above which green real investment is taken[16]. The equilibrium is characterized by the following lemma:

Lemma 4 *The interior solution of equilibrium threshold (e_m, e_I) (i.e., $e_m \in [e_B, \hat{\theta}]$ and $e_I \in [\hat{\theta}, 0]$) is characterized by the solution to the following system of equations:*

$$\begin{cases} \beta \left\{ \alpha e_G + (1-\alpha) \int_{e_B}^{e^m} \theta \ dG(\theta) \right\} - \beta e_m = C_m(e_m), \\ \beta(e_I - e_m) = -\left[C_I(e_I) - C_m(e_m) \right], \end{cases} \tag{12}$$

where $\alpha = \frac{1-F(e_I)}{1-F(e_I)+F(e_m)}$.

Given this characterization of the equilibrium, we can derive the properties of the threshold (e_I, e_m). I show that the main result of the baseline model still holds:

[15] To see that the equilibrium is monotonic under this assumption, define $H(\theta) = C_m(\theta) + \beta\theta$, then $H'(\theta) > 0$ by assumption. The indifference condition of the equilibrium implies that $H(e_m) = \beta \mathbb{E}[v_e | m = e_G]$, so $\forall \theta < e_m$, $H(\theta) < \beta \mathbb{E}[v_e | m = e_G]$, i.e., $\beta \mathbb{E}[v_e | m = e_G] - C_m(\theta) > \beta\theta$. Thus, all firms with $\theta < e_m$ engage in greenwashing. Similar arguments can be applied to the investment threshold e_I.

[16] Note that we only consider the interior solution of (e_m, e_I) here to focus on the general property.

Proposition 6 *Suppose* (e_I^*, e_m^*) *is a solution to the system of equations* (12) *s.t.* $e_I^* \in (\hat{\theta}, 0)$ *and* $e_m^* \in (e_B, \hat{\theta})$.

1. *The share of greenwashing firms is increasing in* β *(i.e.,* $\frac{\partial e_m}{\partial \beta}_{|(e_I^*, e_m^*)} > 0$*).*
2. *The share of green firms is decreasing in* β *(i.e.,* $\frac{\partial e_m}{\partial \beta}_{|(e_I^*, e_m^*)} > 0$*) if* $|e_m^* g(e_m^*)|$ *is large enough.*

Proposition 6 shows that the key insight in the baseline model still holds even if we consider a more general setting with continuous types and endogenous investment opportunities. Given any equilibrium, increasing the intensity of ESG preference β could induce more truth-telling brown firms to engage in greenwashing. Moreover, if a large measure of brown firms are facing a tight incentive-compatible constraint of truth-telling in the equilibrium (i.e., $|e_m^* g(e_m^*)|$ is large), the negative impact of greenwashing on compensation to green investment dominates the effect of larger ESG preference, thus decreasing the share of actual green firms.

6 Extensions: Policy Interventions

6.1 Uncertainty Regarding Investors' ESG Preferences

One unique feature of ESG fundamentals is that there is substantial uncertainty regarding investors' preferences over ESG outcomes. The uncertainty usually comes from two aspects: First, there is political risk associated with anti-ESG movements and policies. For example, Texas lawmakers have already introduced anti-ESG laws in 2021, which banned banks that limit credit to the oil and gas sector from participating in public finance markets in the state. Such anti-ESG policies can flip the market participants' preference over ESG outcomes: in the previous Texas case, banks considered to be ESG became unfavorable to investors after the law was implemented. Thus, future anti-ESG policy risks create uncertainty regarding

investors' preferences. Second, disagreement about measurements and scopes of ESG outcomes could also cause such uncertainty. For example, female friendliness could be measured by the gender pay gap, the percentage of women on the board, or the percentage of women in the workforce. Firms that did well (or claimed to do well) in one measurement might fall behind in the others, and they are not sure which measurements are valued by investors.

In this section, I will analyze how such uncertainty regarding investors' preferences impacts the incentives for greenwashing as well as real investment. For concreteness, I will focus on the uncertainty of future anti-ESG policies (similar intuition can be applied to disagreement risk). Specifically, each investor's utility upon receiving the firm's disclosure

$$u_i(q_i) = (s \cdot \beta v_e - p_e)q_i,$$

where $s \in \{-1, 1\}$ is a random variable that captures policy uncertainty. I assume s satisfies

$$s = \begin{cases} -1 & w.p.\ \delta, \\ 1 & w.p.\ 1-\delta, \end{cases}$$

where $\delta < \frac{1}{2}$ [17]. With probability δ, an anti-ESG policy is implemented, so a green outcome (positive externality value v_e) becomes unfavorable to ESG investors. With probability $1-\delta$, there are no policy shocks and ESG investors value green outcomes as before. Given such policy uncertainty, the market-clearing price is

$$p_e = s \cdot \beta \mathbb{E}[v_e|m] \tag{13}$$

Note that preference uncertainty affects both the incentives for greenwashing and green investment. When there is a higher probability of implementing anti-ESG policies, both incentives decrease because green outcomes are more likely to become unfavorable to investors.

[17] The policy uncertainty should not be too large: otherwise, green firms might claim to be brown because brown outcomes are very likely to be favorable.

When greenwashing is severe (i.e., when information discipline is weak and ESG preference is strong), the effect of diminishing greenwashing could dominate, so slight preference uncertainty might induce more real ESG activities.

Proposition 7 *Suppose there is limited information discipline (i.e., $c \in [k_G(e_G - e_B), k_G e_G - k_B e_B]$) and enough preference for ESG outcomes (i.e., $\beta > \frac{c}{e_G - e_B}$). If $\delta < \frac{1 - \frac{c}{\beta(e_G - e_B)}}{2}$, the equilibrium compensation for green investment is increasing in δ (i.e., $\frac{\partial \widehat{U}_2(e_G)}{\partial \delta} > 0$).*

6.2 Direct Incentivization

In this section, we start from the baseline model in Section 2 and consider direct incentivization to induce green investment, i.e., compensation to the firm manager contingent on the realization of ESG value. Specifically, the ESG investor is able to transfer a proportion λ of the externality value through a direct compensation $w_e = \lambda \beta v_e(\theta, I)$ to the manager at $t = 3$, and the market compensation at $t = 2$ is based on the remaining proportion $1 - \lambda$ of the externality value, i.e., $a_e(m) = \mathbb{E}_I[(1 - \lambda)\beta v_e | m]$. Additionally, we assume $c > k_G(e_G - e_B)$ to simplify the analysis and focus on the main channel.

Proposition 8 *Suppose $c > k_G(e_G - e_B)$, and let $\overline{\lambda} = \frac{\pi \frac{k_G e_G}{c} - \frac{\overline{e}}{e_G - e_B}}{\pi \frac{k_G e_G}{c} + (1 - \pi)}$.*

- *If $\lambda \in [0, -\frac{\overline{e}}{e_G - \overline{e}}]$, the green investment is made if $\beta \in [k_G, \frac{c - k_G e_G}{-[\lambda e_G + (1-\lambda) e_B]}]$.*
- *If $\lambda \in (-\frac{\overline{e}}{e_G - \overline{e}}, \overline{\lambda})$, the green investment is made if $\beta \in [k_G, \frac{c - k_G e_G}{-[\lambda e_G + (1-\lambda) e_B]}] \cup [\frac{k_G e_G}{\lambda e_G + (1-\lambda)\overline{e}}, \infty)$.*
- *If $\lambda \in [\overline{\lambda}, 1]$, the green investment is made if $\beta \in [k_G, \infty)$.*

The proposition shows the equilibrium real investment when there is direct compensation for creating positive externality. Similar to the baseline model, an increase in the intensity of ESG preference β increases the externality value internalized by the investor, but it also induces more greenwashing. As more externality value enters direct incentivization, brown firms have lower greenwashing motives and green firms get larger compensation, so the regions of intensity of ESG preference β inducing green investment become larger. Particularly,

if the proportion of direct compensation λ is small, then the negative effect of greenwashing might dominate as β increases, so there is an intermediate region of β to induce green investment; If the proportion of direct compensation λ is large enough, the positive effect of internalizing externality always dominates, so real efficiency is always increasing in the intensity of ESG preference β.

Corollary 5

- *The set of the intensity of ESG preference β inducing green investment becomes larger as the proportion of direct incentivization λ becomes larger.*
- *There exists a threshold $\overline{\lambda}$ such that real efficiency is increasing in the intensity of ESG preference β if $\lambda \geq \overline{\lambda}$.*

The above results support the argument to add ESG criteria into clawback policies, which will directly affect CEO compensation if the long-term ESG objectives claimed during his tenure are not achieved and thus induce more green investment. More importantly, this direct incentivization can even make market discipline more efficient, as it decreases the greenwashing motive through long-term penalties.

7 Conclusion

As investors attach a higher value to ESG outcomes, it not only affects the incentives of real ESG activities, but also distorts the incentive for truthful reporting, i.e., more brown firms engage in greenwashing. In equilibrium, investors rationally decrease rewards for firms that claim to be green as they expect there are more greenwashing firms, which undermines the incentive for green investment in the first place. Particularly, my analysis suggests that when information discipline is limited, the ESG incentivization from market participants is weak and may even backfire. The key insight of the model still holds if heterogeneous information manipulation costs or endogenous investment opportunities are considered.

In the extensions, I first analyze the case when information about NPV indicates the realization of ESG fundamentals. The analysis rationalizes the empirical finding that firms with

poor business performance are more likely to engage in greenwashing. I also consider the preference uncertainty resulting from anti-ESG policies or disagreement over ESG measurement, and my model shows that a slight level of preference uncertainty might benefit those actual green firms if greenwashing is severe. Last, I discuss the impact of direct incentivization, and I show that it can reduce greenwashing and complement the market discipline.

My model has important policy implications for the ongoing revolutions in ESG investing. Particularly, regulation on ESG information disclosure (such as unified ESG disclosure frameworks, discipline on ESG rating agencies, etc.) is critical and should be developed in parallel with the rapid growth of ESG investing and increasing concern over ESG issues.

Appendix A Proof of Propositions

A.1 Proof of Lemma 1

First, note that in any equilibrium the manager always reports $m = e_G$ if $v_e = e_G$: otherwise, it means that $\mathbb{E}[\beta v_e | m = e_G] < \mathbb{E}[\beta v_e | m = e_B]$, so the "Brown" types must always report $m = e_B$. However, in this case, investors are certain that the firm creates $v_e = e_G$ if it reports $m = e_G$, and thus we have $\mathbb{E}[\beta v_e | m = e_G] > \mathbb{E}[\beta v_e | m = e_B]$. This is a contradiction.

Thus, the only possible equilibria are "full disclosure", "greenwashing", and "partial greenwashing". Next, we drive the conditions under which each equilibrium holds:

- "Full disclosure": $\widehat{U}_2(e_G) = \beta e_G$, $\widehat{U}_2(e_B) = \beta e_B$. It is an equilibrium if firms with $v_e = e_B$ do not deviate to reporting $m = e_G$: $\beta e_B \geq \beta e_G - c$, i.e., $\beta(e_G - e_B) \leq c$.

- "Greenwashing" (pooling at reporting $m = e_G$): $\widehat{U}_2(e_G) = \beta[\alpha e_G + (1-\alpha)e_B]$, $\widehat{U}_2(e_B) = \beta[\alpha e_G + (1-\alpha)e_B] - c$. It is an equilibrium if firms with $v_e = e_B$ do not deviate to reporting $m = e_B$ (we need to specify the off-equilibrium-path belief of investors: $\mu(v_e = e_B) = 1$ if the firm reports $m = e_B$. Note that this refinement of belief: 1. survives D1 criteria; 2. attains the largest possible regions of equilibrium): $\beta[\alpha e_G + (1-\alpha)e_B] - c \geq \beta e_B$, i.e., $\beta(e_G - e_B) \geq \frac{c}{\alpha}$.

- "Partial greenwashing" is an equilibrium if firms with $v_e = e_B$ is indifferent between reporting $m = e_B$ and $m = e_G$: suppose "Brown" types report $m = e_G$ with probability q and report $m = e_B$ with probability $1-q$, then we have $\mathbb{E}[\beta v_e | m = e_B] = \mathbb{E}[\beta v_e | m = e_G] - c$, i.e.,
$$\beta \frac{\alpha e_G + (1-\alpha) q e_B}{\alpha + (1-\alpha)q} - c = \beta e_B.$$
We can solve for $q = \frac{\alpha}{1-\alpha}\left[\frac{\beta}{c}(e_G - e_B) - 1\right]$. In addition, we need $\beta(e_G - e_B) \in (c, \frac{c}{\alpha})$ to ensure that $q \in (0,1)$. Also, note that firms with $v_e = e_G$ strictly prefer reporting $m = G$ since $\mathbb{E}[\beta v_e | m = e_G] = \mathbb{E}[\beta v_e | m = e_B] + c > \mathbb{E}[\beta v_e | m = e_B] - c$. In this case, $\widehat{U}_2(e_G) = \beta e_B + c$, $\widehat{U}_2(e_B) = \beta e_B$.

A.2 Proof of Proposition 1

First, we eliminate any equilibrium in which $I(G) \in (0,1)$ because the mixed strategies are not stable in games with strategic complementarity (for example, see Echenique and Edlin (2004) for more discussions). Note that green real investments are strategic complements: firms always get a constant payoff by keeping the status quo, while the payoff from green

investment is increasing in the share of firms taking the investment. Any perturbed beliefs about other firms' strategies can lead firms to play only one action with certainty.

Also note that the assumption $\beta < k_B$ ensures that brown investments are always made in equilibrium, i.e., $I(B) = 1$. To see this, we will verify that neither $I(B) = 0$ nor $I(B) \in (0, 1)$ can hold in equilibrium.

1. Suppose that in equilibrium $I(B) = 0$, then a necessary condition of the equilibrium is $\beta e_B - k_B e_B \geq 0$, which does not hold under the assumption $\beta < k_B$ (we need to specify off-equilibrium-path beliefs in some cases: for example, in the equilibrium where $I(G) = I(B) = 0$, we assume investors assign probability 1 to brown firms in their belief if they receive any message, which lead to the smallest lower bound of β for equilibrium).

2. Suppose that in equilibrium $I(B) \in (0, 1)$ and $I(G) = 1$ (we can eliminate the equilibrium in which $I(B) = r \in (0, 1)$ and $I(G) = 1$ using the same argument as above), then the firm manager must be indifferent between keeping the status quo and taking the brown investment. If the disclosure equilibrium is separating or partial greenwashing, then the indifference condition is $\beta e_B - k_B e_B = 0$, which only holds in knife-edge cases. If the disclosure equilibrium is full greenwashing, then the indifference condition is $\beta \frac{\pi e_G + (1-\pi) r e_B}{\pi + (1-\pi) r} - c - k_B e_B = 0$. Using the incentive-compatible constraint of greenwashing, $LHS \geq \beta e_B - k_B e_B = (k_B - \beta)(-e_B) > 0$.

Next, we derive the equilibrium in which $I(B) = 1$ and $I(G) \in \{0, 1\}$. We can write down the equilibrium payoff from green/brown investments when $I(G) = 1$, given the equilibrium in the disclosure stage.

$$U_2^*(e_G) = \begin{cases} \beta e_G & \text{if } \frac{\beta}{c} \leq \frac{1}{e_G - e_B}, \\ \beta e_B + c & \text{if } \frac{\beta}{c} \in (\frac{1}{e_G - e_B}, \frac{1}{\alpha} \frac{1}{e_G - e_B}), \\ \beta \bar{e} & \text{if } \frac{\beta}{c} \geq \frac{1}{\alpha} \frac{1}{e_G - e_B}. \end{cases} \tag{1}$$

$$U_2^*(e_B) = \begin{cases} \beta e_B & \text{if } \frac{\beta}{c} \leq \frac{1}{\alpha} \frac{1}{e_G - e_B}, \\ \beta \bar{e} - c & \text{if } \frac{\beta}{c} > \frac{1}{\alpha} \frac{1}{e_G - e_B}, \end{cases} \tag{2}$$

Thus, the equilibrium payoff from green projects $U_2^*(e_G)$ is strictly increasing in β if $\beta \leq \frac{1}{e_G - e_B} c$, strictly decreasing in β if $\beta \in (\frac{1}{e_G - e_B} c, \frac{1}{\alpha} \frac{1}{e_G - e_B} c)$, and strictly decreasing (increasing) in β if $\beta \geq \frac{1}{\alpha} \frac{1}{e_G - e_B}$ given that $\bar{e} < 0$ ($\bar{e} > 0$). The equilibrium payoff from brown projects $U_2^*(e_B)$ is strictly decreasing in β if $\beta \leq \frac{1}{\alpha} \frac{1}{e_G - e_B}$, and strictly decreasing (increasing) in β if $\beta \geq \frac{1}{\alpha} \frac{1}{e_G - e_B}$ given that $\bar{e} < 0$ ($\bar{e} > 0$).

According to previous arguments, there are two possible equilibria: either $I(G) = 0, I(B) = 1$ or $I(G) = I(B) = 1$. First, we show that the equilibrium in which $I(G) = 0$ and $I(B) = 1$ ($m(e_B) = e_B$) always exists: since only message $m = e_B$ is sent in equilibrium, we need to specify the off-equilibrium-path belief when investors receive message $m = e_G$. To derive the largest possible regions of equilibrium, we assume investors assign probability 1 to brown firms if they receive message $m = e_G$. The equilibrium exists if $\beta e_B - k_B e_B > 0$ and $\beta e_B - k_G e_G < 0$, which always hold under the previous assumptions.

Next, we focus on the equilibrium in which $I(G) = I(B) = 1$. Notice that when $\bar{e} < 0$, $|U_2^*(e_B)|$ is strictly increasing in β, while $U_2^*(e_G)$ is maximized at $\beta = \frac{c}{e_G - e_B}$ and $U_2^*(e_G)_{|\beta=\frac{c}{e_G-e_B}} = \frac{e_G}{e_G-e_B}c$. We need to verify two conditions (The whole proof can be shown graphically in the figure of $U_2^*(v_e)$).:

1. $I(B) = 1$: $U_2^*(e_B) \geq -v(B, 1)$. If $(-e_B)\frac{1}{\alpha}\frac{1}{e_G-e_B}c \geq -k_B e_B$, i.e., $c \geq k_B\alpha(e_G - e_B)$, then the condition simplifies to $\beta < k_B$; if $c < k_B\alpha(e_G - e_B)$, then the condition simplifies to $\beta < \frac{-k_B e_B - c}{-\bar{e}}$, and we can show $\frac{-k_B e_B - c}{-\bar{e}} > k_B$ when $c < k_B\alpha(e_G - e_B)$. Thus, this condition always holds under the previous assumptions.

2. $I(G) = 1$: $U_2^*(e_G) \geq -v(G, 1)$. If the maximum payoff is less than the investment cost, i.e., $\frac{e_G}{e_G-e_B}c < k_G e_G$, i.e., $c < k_G(e_G - e_B)$, then this condition never holds for any value of β; otherwise, if $c \geq k_G(e_G - e_B)$, this condition holds if $\beta \in [k_G, \min\left\{\frac{c-k_G e_G}{-e_B}, k_B\right\}]$. Thus, if $\frac{k_G e_G - c}{e_B} < k_B$, i.e., $c < k_G e_G - k_B e_B$, the condition holds when $\beta \in [k_G, \frac{k_G e_G - c}{e_B}]$; otherwise, the condition always hold given that $\beta \in [k_G, k_B)$.

A.3 Proof of Proposition 2

If $y = v < 0$, investors infer that it is a green firm with probability $\pi\rho$ and a brown firm with probability $(1-\pi)(1-\rho)$. When all brown firms engage in greenwashing and fully pool with green firms, they get the compensation $\beta\mathbb{E}[v_e|y = v_L, m = e_G] = \beta\left[\frac{\pi\rho}{\pi\rho+(1-\pi)(1-\rho)}e_G + \frac{(1-\pi)(1-\rho)}{\pi\rho+(1-\pi)(1-\rho)}e_B\right]$. Thus, brown projects with negative NPV are taken if $\beta\mathbb{E}[v_e|y = v_L, m = e_G] - c + v_L > 0$. There exists β s.t. this inequality holds if it holds when $\beta = 1$, i.e.,

$$\left[\frac{\pi\rho}{\pi\rho + (1-\pi)(1-\rho)}e_G + \frac{(1-\pi)(1-\rho)}{\pi\rho + (1-\pi)(1-\rho)}e_B\right] > c - v_L,$$

which can be simplified to $\rho > \bar{\rho} = \frac{(1-\pi)(c-v_L-e_B)}{(1-\pi)(c-v_L-e_B)+\pi(e_G-c+v_L)}$. Note that in this case, green investment is always made when this condition holds because green firms do not have the information manipulation cost and thus have a higher payoff than brown firms.

The share of greenwashing firms is $q = \frac{\pi}{1-\pi}\frac{\rho}{1-\rho}\left[\frac{\beta}{c}(e_G - e_B) - 1\right]$, which is increasing in ρ. In addition, since $\rho > \frac{1}{2}$, q is larger when the real investment has a negative NPV compared to when it has a positive NPV.

A.4 Proof of Lemma 2

As shown in Section 5.1, equation (8) is characterized by the condition that the equilibrium share of greenwashing firms equals the share of brown firms which have information cost $c_i < c_{i^*}$.

First, I prove that equation (8) pins down a unique equilibrium share q for any value of β. Note that $\alpha = \frac{\pi}{\pi+(1-\pi)q}$, so $\frac{\partial\alpha}{\partial q} = -\frac{\pi(1-\pi)}{[\pi+(1-\pi)q]^2} < 0$. Since $F(\cdot)$ is strictly increasing, $F(c_{i^*}) = F(\beta(e_G - e_B)\alpha)$ is strictly decreasing in q. Define $H(q) = F(q) - q$, which is strictly increasing in q. Since $H(0) = F(\beta(e_G - e_B)) \geq 0$ and $H(1) = F(\beta(e_G - e_B)\pi) - 1 \leq 0$, $H(q) = 0$ has exactly one solution on $[0, 1]$.

Next, I prove that the equilibrium share q is strictly increasing in the intensity of ESG preference β. First, take derivative w.r.t. β in equation (8):

$$\frac{\partial q}{\partial \beta} = F'(c_{i^*})(e_G - e_B)(\alpha + \beta\frac{\partial\alpha}{\partial q}\frac{\partial q}{\partial\beta}), \tag{3}$$

which can be simplified to

$$\frac{\partial q}{\partial \beta} = \frac{F'(c_{i^*})(e_G - e_B)\alpha}{1 - F'(c_{i^*})(e_G - e_B)\beta\frac{\partial\alpha}{\partial q}} > 0. \tag{4}$$

A.5 Proof of Proposition 3

Take the derivative of compensation to green investment $\widehat{U}_2(e_G)$ w.r.t. β:

$$\begin{aligned}
\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} &= [\alpha e_G + (1-\alpha)e_B] + \beta(e_G - e_B)\frac{\partial \alpha}{\partial \beta} \\
&= \bar{e}(\alpha) + \underbrace{\beta(e_G - e_B)\frac{\partial \alpha}{\partial q}\frac{\partial q}{\partial \beta}}_{<0} \\
&= \bar{e}(\alpha) + \beta(e_G - e_B)\frac{\partial \alpha}{\partial q}\frac{F'(c_{i^*})(e_G - e_B)\alpha}{1 - F'(c_{i^*})(e_G - e_B)\beta\frac{\partial \alpha}{\partial q}} \qquad (5) \\
&= \frac{\bar{e}(\alpha)[1 - F'(c_{i^*})(e_G - e_B)\beta\frac{\partial \alpha}{\partial q}] + F'(c_{i^*})(e_G - e_B)^2\beta\frac{\partial \alpha}{\partial q}\alpha}{1 - F'(c_{i^*})(e_G - e_B)\beta\frac{\partial \alpha}{\partial q}} \\
&= \frac{\bar{e}(\alpha) - \beta F'(c_{i^*})\frac{\partial \alpha}{\partial q}e_B(e_G - e_B)}{1 - F'(c_{i^*})(e_G - e_B)\beta\frac{\partial \alpha}{\partial q}}.
\end{aligned}$$

Since the denominator is always positive, the sign of $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta}$ is determined by the numerator.

If $\beta = 0$, we have $q = 0$ and $\alpha = 1$, so the numerator is $\bar{e}(1) = e_G > 0$. Since the derivative is continuous, there exists an interval $[0, \beta_L]$ s.t. $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} > 0$ when $\beta \in [0, \beta_L]$.

If $\beta = 1$, we have $q = \bar{q}$ and $\alpha = \bar{\alpha}$. Note that the numerator is negative if

$$\frac{\bar{e}(\alpha)}{\beta} < F'(c_{i^*})\frac{\partial \alpha}{\partial q}e_B(e_G - e_B). \qquad (6)$$

Since $|\frac{\partial \alpha}{\partial q}| \leq \pi(1-\pi)$, a sufficient condition which makes equation (6) holds is:

$$\frac{\bar{e}(\alpha)}{\beta} < \pi(1-\pi)F'(c_{i^*})(-e_B)(e_G - e_B). \qquad (7)$$

Note that we assume there exists $c < k_B(e_G - e_B)\bar{\alpha}$ s.t. $F'(c_i) > A = \frac{\bar{e}(\bar{\alpha})}{\pi(1-\pi)(-e_B)(e_G - e_B)}$ when $c_i > c$. Define the intensity of ESG preference as β_c when $c_{i^*} = c$. The LHS of equation (7) is decreasing in β, so it is minimized at $\beta = 1$ and the inequality holds at $\beta = 1$. Thus, we can find a $\beta_A < k_B$ s.t. $\frac{\bar{e}(\alpha)}{\beta} < \pi(1-\pi)A(-e_B)(e_G - e_B)$ as long as $\beta \in [\beta_A, k_B]$. Let $\beta_H = \max\{\beta_A, \beta_c\}$, then we have $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} < 0$ when $\beta \in [\beta_H, k_B]$.

A.6 Proof of Corollary 4

Since $F(k_G(e_G - e_B)) = 0$, there is no greenwashing when $\beta \leq k_G$, i.e., $q(k_G) = 0$. Thus, $\widehat{U}_2(e_G)_{|\beta \leq k_G} = \beta e_G$, and the green investment is made if $\beta = k_G$.

The interval of no green investment exists if $\widehat{U}_2(e_G)_{\beta=k_B} = k_B[\bar{\alpha}e_G + (1-\bar{\alpha})e_B] < k_G e_G$, i.e., $\bar{\alpha} < \frac{\frac{k_G}{k_B}e_G - e_B}{e_G - e_B}$. This condition is equivalent to $F(k_B(e_G - e_B)\alpha)_{|\alpha=[\frac{k_G}{k_B}e_G - e_B]/(e_G - e_B)} = F(k_G e_G - k_B e_B) > \bar{q}$.

A.7 Proof of Lemma 3

For any firm with a brown investment opportunity, the payoff from taking brown investment and engaging in greenwashing is $\beta[v_e|m = e_G] - k_B e_B - c_i$; the payoff from taking brown investment and truthfully reporting is $\beta e_B - k_B e_B$; the payoff from keep the status quo is 0. Since $\beta > k_B$, $\beta e_B - k_B e_B < 0$, so taking brown investment and truthfully reporting is strictly dominated by keeping the status quo.

A.8 Proof of Proposition 4

Similar to the Proof of Lemma 2, we can show that there is exactly one $q \in [0,1]$ as the equilibrium share of greenwashing firms based on the indifference condition $F(\beta[\alpha e_G + (1-\alpha)e_B] - k_B e_B) = q$. Take the derivative of the indifference condition w.r.t. β:

$$F'(c_{i^*})\left\{\alpha e_G + (1-\alpha)e_B + \beta(e_G - e_B)\frac{\partial \alpha}{\partial \beta}\right\} = \frac{\partial q}{\partial \beta}, \tag{8}$$

which implies

$$\frac{\partial q}{\partial \beta} = \frac{[\alpha e_G + (1-\alpha)e_B]F'(c_{i^*})}{1 - F'(c_{i^*})(e_G - e_B)\beta\frac{\partial \alpha}{\partial q}}. \tag{9}$$

Because the equilibrium q is unique, we can consider the solution as β increases from $\beta = k_B$. If $\bar{e}(\bar{\alpha}) > 0$, $\frac{\partial q}{\partial \beta}_{|\beta=k_B} > 0$, $\frac{\partial \alpha}{\partial \beta}_{|\beta=k_B} < 0$. We can show that as β increases, α never reaches the threshold in which $\bar{e}(\bar{\alpha}) = 0$. To see this, we suppose α^* is the smallest α such that $\bar{e}(\bar{\alpha}) = 0$ as β increases from k_B. Then we have $\frac{\partial q}{\partial \beta} > 0$ and $\frac{\partial \alpha}{\partial \beta} < 0$ when $\alpha \in [\bar{\alpha}, \alpha^*]$, so $F(\bar{e}(\bar{\alpha}) - k_B e_B) = \bar{q} < q_{|\alpha=\alpha^*} = F(-k_B e_B)$. However, because $\bar{e}(\bar{\alpha}) > 0$, $F(\bar{e}(\bar{\alpha}) - k_B e_B) > F(-k_B e_B)$, which is a contradiction. Thus, we must have $\frac{\partial q}{\partial \beta} > 0$ for $\beta \in [k_B, 1]$. Since $q = F(\widehat{U}_2(e_G) - k_B e_B)$, $\frac{\partial \widehat{U}_2(e_G)}{\partial \beta} > 0$ for $\beta \in [k_B, 1]$. Following the Similar arguments, we can show the results for the case where $\bar{e}(\bar{\alpha}) < 0$.

A.9 Proof of Proposition 7

The proof follows directly from the proof of Proposition 1, and now the preference parameter becomes $\beta(1-2\delta)$. The he equilibrium compensation for green investment is increasing in δ if $\beta(1-2\delta) > \frac{c}{e_G - e_B}$, i.e., $\delta < \frac{1 - \frac{c}{\beta(e_G - e_B)}}{2}$.

Appendix B Intervals of Green Investment if $\bar{e} > 0$

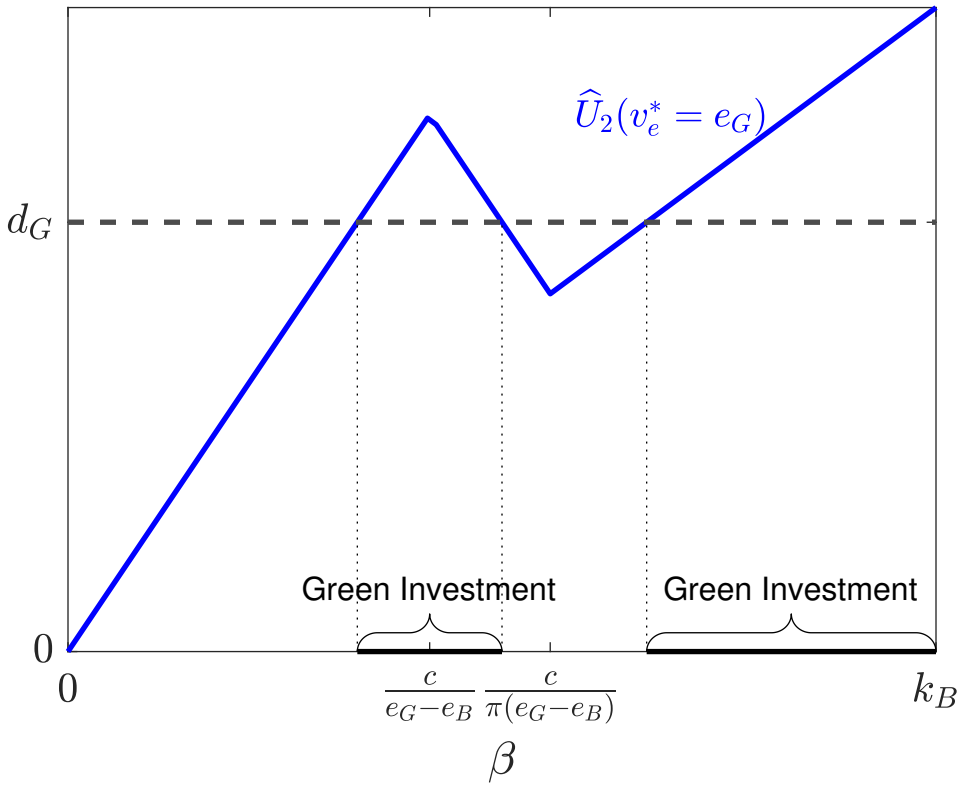

Figure 7: Intervals of green investment if $\bar{e} = \pi e_G + (1 - \pi)e_B > 0$

References

Avramov, D., S. Cheng, A. Lioui, and A. Tarelli (2022). Sustainable investing with esg rating uncertainty. *Journal of Financial Economics 145*(2), 642–664.

Bailey, M., S. Glaeser, J. D. Omartian, and A. Raghunandan (2022). Misreporting of mandatory esg disclosures: Evidence from gender pay gap information. *Available at SSRN 4192257*.

Baker, A., D. F. Larcker, C. McClure, D. Saraph, and E. M. Watts (2023). Diversity washing. *Available at SSRN 4298626*.

Bebchuk, L. A. and R. Tallarita (2022). The perils and questionable promise of esg-based compensation. *Available at SSRN 4048003*.

Berg, F., F. Heeb, and J. F. Kölbel (2022). The economic impact of esg rating changes. *Available at SSRN 4088545*.

Berg, F., J. F. Koelbel, and R. Rigobon (2022). Aggregate confusion: The divergence of esg ratings. *Review of Finance 26*(6), 1315–1344.

Berg, F., J. F. Kölbel, A. Pavlova, and R. Rigobon (2021). Esg confusion and stock returns: Tackling the problem of noise. *Available at SSRN 3941514*.

Berrone, P. and L. R. Gomez-Mejia (2009). Environmental performance and executive compensation: An integrated agency-institutional perspective. *Academy of Management Journal 52*(1), 103–126.

Beyer, A. and I. Guttman (2012). Voluntary disclosure, manipulation, and real effects. *Journal of Accounting Research 50*(5), 1141–1177.

Chowdhry, B., S. W. Davies, and B. Waters (2019). Investing for impact. *The Review of Financial Studies 32*(3), 864–904.

Christensen, D. M., G. Serafeim, and A. Sikochi (2022). Why is corporate virtue in the eye of the beholder? the case of esg ratings. *The Accounting Review 97*(1), 147–175.

Cohen, S., I. Kadach, G. Ormazabal, and S. Reichelstein (2022). Executive compensation tied to esg performance: International evidence.

Crawford, V. P. (1989). Learning and mixed-strategy equilibria in evolutionary games. *Journal of Theoretical Biology 140*(4), 537–550.

De Angelis, T., P. Tankov, and O. D. Zerbib (2022). Climate impact investing. *Management Science*.

Duchin, R., J. Gao, and Q. Xu (2022). Sustainability or greenwashing: Evidence from the asset market for industrial pollution. *Available at SSRN*.

Dye, R. A. (1988). Earnings management in an overlapping generations model. *Journal of Accounting research*, 195–235.

Echenique, F. and A. S. Edlin (2004). Mixed equilibria are unstable in games of strategic complements. *Journal of Economic Theory 118*(1), 61–79.

Edmans, A., D. Levit, and J. Schneemeier (2022). Socially responsible divestment. *European Corporate Governance Institute–Finance Working Paper* (823).

Fischer, P. E. and R. E. Verrecchia (2000). Reporting bias. *The Accounting Review 75*(2), 229–245.

Flammer, C. (2021). Corporate green bonds. *Journal of Financial Economics 142*(2), 499–516.

Flammer, C., M. W. Toffel, and K. Viswanathan (2021). Shareholder activism and firms' voluntary disclosure of climate change risks. *Strategic Management Journal 42*(10), 1850–1879.

Flugum, R. and M. E. Souther (2022). Stakeholder value: A convenient excuse for underperforming managers? *Available at SSRN 3725828*.

Fudenberg, D. and D. M. Kreps (1993). Learning mixed equilibria. *Games and economic behavior 5*(3), 320–367.

Fudenberg, D. and J. Tirole (1986). A" signal-jamming" theory of predation. *The RAND Journal of Economics*, 366–376.

Gantchev, N., M. Giannetti, and R. Li (2022). Does money talk? divestitures and corporate environmental and social policies. *Review of Finance 26*(6), 1469–1508.

Gibson Brandon, R., S. Glossner, P. Krueger, P. Matos, and T. Steffen (2022). Do responsible investors invest responsibly? *Review of Finance 26*(6), 1389–1432.

Gillan, S. L., A. Koch, and L. T. Starks (2021). Firms and social responsibility: A review of esg and csr research in corporate finance. *Journal of Corporate Finance 66*, 101889.

Goldman, E. and S. L. Slezak (2006). An equilibrium model of incentive contracts in the presence of information manipulation. *Journal of Financial Economics 80*(3), 603–626.

Heinkel, R., A. Kraus, and J. Zechner (2001). The effect of green investment on corporate behavior. *Journal of financial and quantitative analysis 36*(4), 431–449.

Ilhan, E., P. Krueger, Z. Sautner, and L. T. Starks (2023, 01). Climate Risk Disclosure and Institutional Investors. *The Review of Financial Studies*. hhad002.

Kim, S. and A. Yoon (2023). Analyzing active fund managers' commitment to esg: Evidence from the united nations principles for responsible investment. *Management Science 69*(2), 741–758.

Liang, C., B. Lourie, A. Nekrasov, and I. S. Yoo (2022). Voluntary disclosure of workforce gender diversity. *Available at SSRN 3971818*.

Liang, C. Y., Y. Qi, R. A. Zhang, and H. Zhu (2022). Does sunlight kill germs? stock market listing and workplace safety. *Journal of Financial and Quantitative Analysis*, 1–30.

Liang, H., L. Sun, and M. Teo (2022). Responsible hedge funds. *Review of Finance 26*(6), 1585–1633.

Oehmke, M. and M. M. Opp (2022). A theory of socially responsible investment. *Swedish House of Finance Research Paper* (20-2).

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

Raghunandan, A. and S. Rajgopal (2022). Do esg funds make stakeholder-friendly investments? *Review of Accounting Studies 27*(3), 822–863.

Serafeim, G. and A. Yoon (2022). Stock price reactions to esg news: The role of esg ratings and disagreement. *Review of accounting studies*, 1–31.

Stein, J. C. (1989). Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The quarterly journal of economics 104*(4), 655–669.