Mandatory vs. voluntary ESG disclosure, efficiency, and real effects*

July 30, 2022

Abstract

In this study, we examine the equilibrium effects of ESG quality disclosure in both voluntary and mandatory regimes. A firm manager makes a private investment decision in an environmentally friendly or unfriendly project that affects future cash flows and the social externalities produced by the firm. We build from Shin (2003) and allow an informed manager to make potentially disparate disclosure decisions on multiple interdependent outcomes—future financial performance and ESG quality. We find that mandating ESG quality disclosure results in *over-investment* in the sustainable technology. That is, the manager often implements sustainable investment even though this is overall less preferred by shareholders. Moreover, a voluntary disclosure regime can be more efficient for investment than a mandatory regime, from the perspective of shareholders. The results also show that mandating ESG disclosure leads to a greater prevalence of sustainable investing. The results provide insights that can be relevant for public policy considerations regarding mandatory ESG disclosure as well as implications that can help to guide empirical research.

Keywords: ESG disclosure, voluntary disclosure, mandatory disclosure, ESG score, project choice, investment, real effects, efficiency.
JEL classification: C72, D82, G11, G23, M41.

1 Introduction

Environmental, social, and governance (ESG) considerations have become prevalent for institutional and retail investors in recent years.[1] The rise in demand for ESG performance has led some firms to voluntarily release ESG information. For example, in the U.S., the majority of firms in the S&P 500 issue sustainability or corporate responsibility reports (Christensen et al. (2021)). Recently, the Securities and Exchange Commission (SEC) has approved a proposal for extensive mandatory climate-related disclosure requirements, which "could mark the most sweeping overhaul of corporate disclosure rules in more than a decade" (*The Washington Post*, March 15, 2022).[2] Moreover, the European Union and at least twenty-five countries to date have imposed mandatory ESG disclosure requirements in some form on publicly traded firms (Krueger et al. (2021)).[3] An important question is to what extent disclosure requirements influence ESG activities and investment decisions, and whether such effects are efficiency enhancing.[4] Despite the rising prevalence of ESG reporting and mandatory ESG disclosure requirements, as well as the large empirical literature which has emerged, there is relatively little theoretical investigation of ESG disclosure and investment. The goal of this paper is to provide theoretical guidance on the role of ESG disclosure in firm investment, in both mandatory and voluntary regimes, and its implications for efficiency and investor welfare.

We consider the equilibrium effects of mandatory versus voluntary ESG disclosure in a parsimonious model. Our setting is one where a firm manager privately selects a project (or an investment or production technology). The project varies along two dimensions in terms

[1] For example, sustainable investments made up 33% ($17 trillion) of assets professionally managed in the U.S. in 2020 (US SIF (2020), Christensen et al. (2021)). Moreover, net flows to 300 mutual funds with ESG mandates quadrupled in 2019 (Gillan et al. (2021)). Global investment in sustainable investments or with explicit ESG goals exceeded $30 trillion in 2018, according to the 2018 Global Sustainable Investment Review. For empirical evidence on the demand for ESG investing and disclosure, see, e.g., Dyck et al. (2019), Hartzmark and Sussman (2019), Krueger et al. (2020), Ilhan et al. (2020), Bauer et al. (2021), and Humphrey et al. (2021).

[2] The proposal was approved by the SEC on March 21, 2022. The proposal is currently undergoing the 60-day public comment process before the agency's four commissioners vote on the final rule. See the SEC press release for more information: https://www.sec.gov/news/press-release/2022-46.

[3] The European Union requires large public companies in EU member states to disclose information through the Non-Financial Reporting Directive 2014/95/EU, which went into effect in 2018. Similar legislative attempts have been made in the U.S. Congress; on June 16, 2021, the U.S. House of Representatives passed the Corporate Governance Improvement and Investor Protection Act (HR 1187), which would require public companies to disclose ESG metrics and allow enforcement by the SEC. See https://www.congress.gov/bill/117th-congress/house-bill/1187.

[4] As noted in the recent review by Christensen et al. (2021), "It is very difficult to predict whether the described firm responses are net positive or negative from the perspective of investors, other stakeholders, or society. [...] We need more research to better understand these tradeoffs as well as how and why firms respond to specific reporting requirements" (p. 1232).

of future expected performance—ESG quality and cash flows. In particular, an investment in the *sustainable* project is more likely to produce a low negative social externality (i.e., high ESG quality) but also relatively lower future cash flows. Conversely, investing in the *non-renewable* technology results in both higher expected future cash flows and social externality (i.e., low ESG quality). This structure captures the basic trade-off in green investment: firms can pursue environmentally friendly projects, however the green technology is more costly to implement or requires greater knowledge investment (R&D), resulting in lower expected profitability.

Following the investment decision, the manager potentially receives private information regarding future financial performance and ESG quality. We consider two disclosure regimes—voluntary and mandatory. In the voluntary regime, an informed manager can decide what to disclose or withhold. In the mandatory regime, the manager must disclose ESG quality but she continues to have discretion over disclosure of future expected financial performance, if informed. Investors have heterogeneous preferences over the ESG quality of the firm, consistent with recent survey and empirical evidence documenting ESG concerns in investor preferences (e.g., Riedl and Smeets (2017), Hartzmark and Sussman (2019), Krueger et al. (2020), Ilhan et al. (2020), Bauer et al. (2021)). In particular, investors are composed of two types: (*i*) financial, or traditional, investors who are concerned only with the firm's financial performance, and (*ii*) social investors, who care about both financial performance and ESG quality. The manager is concerned over each class of investors' preferences proportional to their mass of the shareholder base.

We begin our analysis by establishing the first-best benchmark, where the market always observes the realizations of both dimensions. In this case, the manager chooses the sustainable (or "clean") project when the fraction of social investors is sufficiently high, and the non-renewable project otherwise. The first-best benchmark is efficient in the sense that the manager fully internalizes aggregate shareholder preferences and maximizes aggregate shareholder welfare.

We then turn to our main analysis where the manager has discretion over disclosure in the voluntary and mandatory regimes, examining first the voluntary regime. Due to the presence of multiple project outcomes, we follow the innovative work of Shin (2003) and consider equilibria in the disclosure subgame which satisfy the intuitively appealing and simple *sanitation* strategy, whereby an informed manager discloses positive signals and withholds negative ones. Moreover, since the outcomes of both dimensions are affected by the underlying project choice, outcomes are *interdependent* in our setting; disclosure or non-disclosure along one dimension influences beliefs over the other dimension as well. We characterize market beliefs and show that sanitation forms an equilibrium disclosure strategy.

In terms of investment, we find that the manager chooses the sustainable (resp. non-renewable) project with probability one when the share of investors with concerns over ESG quality is sufficiently high (resp. low). Interestingly, however, we find that the manager mixes over project choices when there is an intermediate fraction of social or financial investors, in contrast to the first-best benchmark. This occurs as the manager can benefit by privately deviating if the market conjectures a pure strategy of investment when no investor group is sufficiently dominant. For example, if the market expects the clean project to be chosen, the manager can instead privately deviate to selecting the non-renewable investment. By doing so, the manager can increase the chance of realizing the high financial performance outcome, thereby raising market beliefs after positive disclosure on this dimension. Of course, the private deviation also lowers the likelihood that high ESG quality is realized; however, since the market believes that the manager selected the clean investment, beliefs over the ESG dimension in the event of non-disclosure are inflated upwards. In this way, the manager can win the best of both worlds by attempting to satisfy both classes of investors. Since the market rationally anticipates this deviation, the equilibrium investment strategy is mixed when no investor group is dominant.

We then proceed to characterizing the equilibrium of the mandatory regime and providing efficiency and welfare analyses. Our first key efficiency result is that *mandatory disclosure of ESG quality results in over-investment of the clean technology relative to the first-best level.* The reason for this relates to the discussion above; under voluntary disclosure, the manager can benefit through private deviations. However, when disclosure is mandatory, the manager can no longer strategically withhold ESG information, as the market observes the signal realization of ESG quality. Consequently, if the manager privately deviates from the clean to the non-renewable technology, she is unable to benefit from inflated market beliefs along the ESG dimension following non-disclosure. As a result, the manager faces little incentive to privately deviate when the market expects her to invest in the sustainable technology. However, the manager continues to have an incentive to privately deviate when the market expects the non-renewable technology, as she still has discretion over disclosing future expected financial performance. The net result is under-adoption of the non-renewable technology and over-adoption of the clean technology, relative to the first-best level.

In light of the discussion above, one might expect that mandatory disclosure is nevertheless efficiency enhancing over voluntary disclosure, as one valve of the manager's private deviation incentive is shut off under the mandatory regime. However, we find that this is not always the case. Indeed, voluntary disclosure can result in investment that is closer to the first-best level than mandatory disclosure under certain conditions. In particular, when the fraction of social shareholders is not sufficiently high, mandatory disclosure results in lower aggregate shareholder welfare relative to voluntary disclosure. Two countervailing effects are present. First, as noted above, the manager has less incentive to privately deviate when the market expects the clean investment under mandatory disclosure. This effect brings the project selection closer to the first-best implementation relative to voluntary disclosure.

The second effect is that the manager has a *heightened* incentive to privately deviate from the non-renewable investment when the market expects that project. Because the market observes the signal realization of ESG quality in the mandatory regime, the manager can no longer hide poor ESG signal realizations by mimicking uninformedness. Consequently, she is "punished" relatively more severely following bad ESG outcomes, even when this outcome was expected by the market. Moreover, the manager continues to have discretion over disclosure of the signal of future financial performance. As such, she is more willing to privately deviate when the market expects the non-renewable project, even if this eventually results in a lower belief along the financial dimension (due to non-disclosure). Put differently, the manager's lower payoff from poor ESG outcomes under the mandatory regime *intensifies* her inclination towards private deviation. This effect lowers investment efficiency and results in lower aggregate shareholder welfare relative to the voluntary regime. We find that the first effect dominates when the proportion of social investors is sufficiently high, while the second effect dominates otherwise. Hence, our second main efficiency result is that *voluntary disclosure can be efficiency enhancing relative to mandatory ESG disclosure.*

Our results provide a number of empirical implications with respect to the likelihood of ESG disclosure, the prevalence of sustainable investments, and the level of real investment efficiency. The equilibrium characterization of the voluntary regime implies that firms with a greater share of investors who have preferences over the social outcomes of the firm are more likely to issue ESG disclosures, consistent with the findings of Ilhan et al. (2020). These firms are also more likely to make sustainable investments, consistent with Dyck et al. (2019) and Chen et al. (2020). Moreover, we expect to see greater ESG disclosure and sustainable investment in firms or industries which have greater uncertainty over future cash flows, such as growth industries or industries with rapidly evolving product markets, and less ESG disclosure in stable industries.

Furthermore, in terms of within-region implications, we expect a greater prevalence of sustainable investing when shifting from a voluntary to mandatory ESG disclosure regime. This helps to explain the results of Chen et al. (2018), Downar et al. (2021), and Jouvenot and Krueger (2021), who find that a country's shift from voluntary to mandatory environmental disclosure was followed by greater adoption of ESG activities and a decrease in negative externalities produced by firms, on average. As noted above, a shift in the disclosure regime can also be met with lower efficiency for shareholders. Our results predict that mandatory disclosure is efficiency-increasing for firms that have a larger proportion of social investors and efficiency-decreasing for firms with a lower proportion of these investors, pointing to an overall non-monotone effect on efficiency. This helps to explain the findings of Grewal et al. (2019), who document a non-monotone reaction to the European Union's ESG disclosure mandates. These predictions, as well as others, are thoroughly discussed in Section 6.

1.1 Related Literature

This study connects to a number of different literatures. As noted above, our model incorporates voluntary disclosure of private information. Grossman and Hart (1980), Grossman (1981), and Milgrom (1981) first investigate voluntary disclosure and show the influential unraveling principle, whereby the sender reveals her private information in the absence of disclosure frictions. We follow Dye (1985) and Jung and Kwon (1988) and assume uncertainty over the manager's endowment of information as the basic disclosure friction.

A central feature of our setting is the presence of multiple disclosure outcomes. As such, our model builds from the pioneering work of Shin (2003). As in Shin (2003), we assume that the manager can disclose multiple binomial outcomes; we therefore incorporate Shin (2003)'s elegant "sanitation" strategy, which narrows the range of disclosure strategies to one that satisfies intuitive and natural properties. Our work varies from Shin (2003) in two ways. First, the outcomes in our disclosure subgame are interdependent. Specifically, outcomes are connected through the underlying project choice in our setting, whereas outcomes are unconditionally independent in Shin (2003). Second, we endogenize project choice, thus allowing the manager to influence project outcomes, whereas the underlying distribution that determines outcomes is assumed to be exogenous in Shin (2003).

Ben-Porath et al. (2018) and Guttman and Meng (2021) similarly examine voluntary disclosure and investment decisions jointly. These studies find that the presence of voluntary disclosure leads to riskier project selection and can amplify information acquisition, respectively. Our study varies as we analyze project choice when the manager is potentially informed over multiple disparate outcomes (and thus communication is multidimensional), we examine the interaction of mandatory and voluntary disclosure, and we allow investors to have heterogeneous preferences. We consequently find that voluntary disclosure following project selection can be efficiency enhancing relative to partial mandatory disclosure.[5]

Our paper is also related to the small but growing theoretical literature on the role of ESG concerns in capital markets. As in the present study, this literature generally features a class of investors who have preferences for high ESG quality. Heinkel et al. (2001), Luo and Balvers (2017), Baker et al. (2018), Baker et al. (2020), Zerbib (2020), Pástor et al. (2021), and Pedersen et al. (2021) consider the asset pricing implications of ESG quality in portfolio choice models in the spirit of Fama and French (2007). These studies generally show that "green" firms with high ESG quality have lower expected returns relative to firms with low ESG quality. This literature assumes that firm ESG quality is commonly known to investors and is determined exogenously, in contrast to our setting. Uncertainty over ESG quality is also studied in the models of Friedman and Heinle (2016), Avramov et al. (2021), and Goldstein et al. (2022), but ESG quality continues to be exogenously determined and strategic ESG disclosure is not considered in these settings.

Lyon and Maxwell (2011) examine a voluntary disclosure model (à la Shin (2003)) where firms have private information over multiple environmental activities. Lyon and Maxwell (2011) assume that an activist auditor can investigate the firm in the event of non-disclosure and impose monetary punishment if the firm is found to have withheld negative information (akin to litigation risk of non-disclosure, e.g., Marinovic and Varas (2016)). Among other differences, in our study we endogenize the distribution of outcomes, embed both financial and ESG dimensions in disclosure, incorporate interdependent outcomes, and we do not consider audit/litigation costs of non-disclosure. Friedman et al. (2021a) consider strategic misreporting of ESG activities, where investors have uncertainty over the manager's objective. The present model varies as we examine both mandatory and voluntary disclosure regimes of ESG quality in the face of an investment decision.

A few papers consider the role of activist investors in impacting firm investment decisions towards socially responsible projects, such as Gollier and Pouget (2014), Chowdhry et al. (2019), Landier and Lovo (2020), Oehmke and Opp (2020), Friedman and Heinle (2021), Green and Roth (2021), and Gupta et al. (2021). Our study varies as we consider the role of disclosure in affecting project decisions when a portion of atomistic shareholders have ESG

[5] Other papers that examine voluntary disclosure and investment include Kumar et al. (2012) and Wen (2013). Kumar et al. (2012) models voluntary disclosure to an activist shareholder who controls investment, and Wen (2013) examines strategic disclosure conditional on whether the manager has invested in an exogenous project opportunity.

preferences.

Our study also relates to the stream of literature that explores the real effects of disclosure (e.g., Kanodia and Lee (1998), Edmans et al. (2016), Goldstein and Yang (2019); see Kanodia and Sapra (2016) for a review). These studies analyze the efficiency effects of the precision or frequency of reporting. Our study adds to this literature as we consider real effects arising from the interplay of voluntary and mandatory disclosures, while the prior studies largely consider mandatory disclosure. Relatedly, our study connects to the literature which examines the interaction between mandatory and voluntary disclosure, such as Gigler and Hemmer (1998), Einhorn (2005), Bertomeu et al. (2021), and Friedman et al. (2021b). We add to this literature by investigating how the interaction between mandatory and voluntary disclosure impacts real investment decisions.

Pae (2005) and Bertomeu and Marinovic (2016) investigate settings with multidimensional communication from a privately informed manager. Pae (2005) extends the Dye (1985) framework to multiple signals and disclosures, while Bertomeu and Marinovic (2016) considers a setting where the manager can disclose a verifiable message along with a manipulable message. Our setting similarly includes multidimensional disclosure, however our model varies as we endogenize the underlying distribution of private information through an investment decision.

The paper proceeds as follows. We present the model in the next section and establish the benchmark case of first-best project implementation in Section 3. In Section 4 we characterize the equilibrium of the voluntary regime, while the mandatory regime and efficiency results are presented in Section 5. In Section 6 we examine comparative statics and discuss empirical predictions. The final section concludes. All proofs are relegated to the Appendix.

2 Model

We assume that a risk-neutral firm manager has access to two different investment technologies or projects, denoted by $\tau \in \{C, D\}$. The first, $\tau = C$, is a sustainable, or "clean," investment, while the second, $\tau = D$, represents a non-renewable, or traditional, investment. Each project τ affects the firm's future performance in two dimensions: financial and social. The future *financial* performance of the project is denoted as $a \in \{a_h, a_l\}$. This can be thought of as the future financial impact of the project for the firm, such as cash flows or net return from the investment (we often use cash flows for expositional ease). The high cash flow a_h is realized with probability ϕ_τ, and the low cash flow a_l is realized with probability $1 - \phi_\tau$. Each project similarly generates a non-pecuniary negative social externality, $x \in \{x_h, x_l\}$, where the high externality x_h is realized with probability $1 - p_\tau$ and the low externality x_l with probability p_τ. The negative externality can capture, for example, carbon emissions or other social considerations. We assume that the sustainable technology is less effective than the non-renewable technology in generating high future cash flows for the firm but is also less likely to impose a high negative social externality. In particular, we assume that ϕ_τ and p_τ have the following properties:

$$0 < \phi_C < \phi_D < 1,$$
$$0 < p_D < p_C < 1.$$

The stochastic nature of the outcomes is meant to capture uncertainty regarding the eventual efficacy or success of the project/technology along both the financial and social dimensions. For example, with respect to the clean investment, firms must innovate in potentially uncharted territory when shifting to projects that are more sustainable and therefore cannot guarantee the technology's social efficacy. Moreover, the advantage in sustainability may raise the cost of the product or limit the rate of production, potentially resulting in lower generated cash flows (i.e., uncertainty over a). Likewise, financial and social outcomes associated with the non-renewable investment, $\tau = D$, are also stochastic. This is meant to capture, for instance, residual uncertainty over demand for the product (uncertainty over a), or exogenous innovations in the supply chain or other technological advances which allow the firm to claim sustainable production (uncertainty over x).

Conditional on the investment choice, the realization of the signal for future cash flows is independent of the realization of the signal for the negative social externality. The manager privately chooses the investment τ at the beginning of the game. We allow for mixed strategies in the investment decision; we denote by $\sigma \in [0, 1]$ the probability that the manager chooses $\tau = C$.

We assume that there is a continuum of risk-neutral investors of mass one composed of two types—financial and social investors. Financial investors, who are of mass $1 - \omega$, are concerned only with the financial performance of the firm, i.e., the *financial value* of the firm, which is given as

$$a = a_l + \alpha f, \tag{1}$$

where $\alpha = a_h - a_l$, and f is equal to one with probability ϕ_τ and zero otherwise. The remaining mass $\omega \in [0, 1]$ of investors are social investors, who care about both future cash flows and the negative social externality produced by the firm. This *social value* of the firm is denoted by v and defined as

$$v \equiv a - x = a_l + \alpha f - (x_h - \beta g) = a_l - x_h + \alpha f + \beta g, \tag{2}$$

where $\beta = x_h - x_l$, and g is one with probability p_τ and zero otherwise.[6,7]

We see that the non-renewable technology has a higher expected financial value than the clean investment, $E_C[a] < E_D[a]$. We also assume that the clean investment generates a higher expected social value: $E_C[v] > E_D[v]$.[8] To better focus the exposition and more clearly illustrate the economic forces of the model, we impose the following relationship between the two investments.

Assumption 1. *$\phi_D p_D = \phi_C p_C$, which implies*

$$\begin{aligned} \frac{\triangle_p}{p_C} &= \frac{\triangle_\phi}{\phi_D} \\ \frac{\alpha}{\beta} &< \frac{p_C}{\phi_D}, \end{aligned}$$

where $\triangle_\phi = \phi_D - \phi_C$ and $\triangle_p = p_C - p_D$.

Assumption 1 implies that the loss in cash flow from the clean investment is proportionate to the increase in negative social externality from the non-renewable investment. This assumption essentially rules out one investment being dominant to the other and captures the trade-off we seek to model between the two technologies. We note that the results are not qualitatively sensitive to this assumption and continue to hold insofar as one investment type does not become superior relative to the other.

Following the investment choice, the manager probabilistically receives signals regarding future cash flows and the social externality. In particular, the manager observes a perfect signal of future cash flows, which can be represented with parameter f, with probability $q_f \in (0, 1)$, and a perfect signal of the social externality, g, with probability $q_g \in (0, 1)$.[9] For expositional ease, we often refer to the outcome of $g = 1$ as high ESG quality (or ESG

[6] An alternative interpretation of this structure is that investors are homogeneous and place a weight of $\hat{\omega}$ on financial value and $1 - \hat{\omega}$ on social value.

[7] Riedl and Smeets (2017), Krueger et al. (2020), Ilhan et al. (2020), and Bauer et al. (2021) provide survey evidence of investor preferences for social value, while Hartzmark and Sussman (2019) provides evidence from observational data. In particular, Bauer et al. (2021) finds that "social preferences rather than financial beliefs or confusion drive the choice for more sustainability. [...] Even among those who expect lower financial returns, the majority of 58% choose more sustainable investments" (p. 3979).

[8] This assumption implies that $\alpha \triangle_\phi < \beta \triangle_p$.

[9] Allowing the signals to be imperfect would not qualitatively change the results.

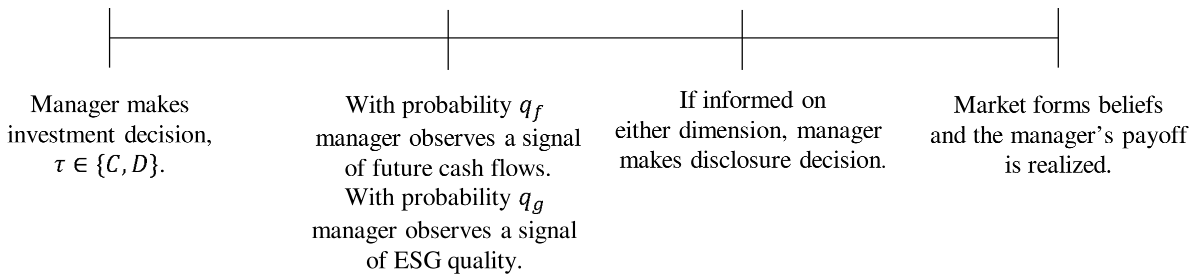

Figure 1: Timeline.

score), and $g = 0$ as low ESG quality. For example, the manager observes preliminary results regarding the project's status, such as whether it was successful or the efficacy of the technology, and can accurately gauge the expected future cash flows and social externality that the project will generate. The manager's likelihood of observing each event is independent of one another, the realized (f, g), and the investment decision. The manager's information endowment is summarized as

$$s = \begin{cases} (\emptyset, \emptyset) & \text{with probability } (1 - q_f)(1 - q_g), \\ (\emptyset, g) & \text{with probability } (1 - q_f)q_g, \\ (f, \emptyset) & \text{with probability } q_f(1 - q_g), \\ (f, g) & \text{with probability } q_f q_g. \end{cases}$$

The manager then makes a disclosure decision of f, if she is informed, and of g, again if she informed on this dimension. A manager who is uninformed with respect to a dimension (cash flows or social externality) cannot disclose that dimension and, as in Dye (1985), is unable to convey to the market that she is uninformed. A manager who is informed on both dimensions can choose to disclose both signals, one signal, or neither signal. We also consider a mandatory disclosure regime for comparison where the manager, if informed, only has discretion on the disclosure of future financial performance, f, while the market always observes ESG quality g.

The manager weighs the preferences of both investors according to their mass. The manager's payoff is given by

$$U(s, d, \sigma) = \omega E[v|d(s), \sigma] + (1 - \omega)E[a|d(s), \sigma],$$

where $d(s)$ is the manager's disclosure decision, and σ represents the manager's investment strategy. The manager thus aims to maximize the aggregate posterior beliefs of the two types of investors, scaled by their relative mass, through her investment and disclosure decisions. This payoff function can be microfounded by, for example, the manager's interest in raising capital from investors.[10] Following the disclosure decision, investors form beliefs and the manager's payoff is realized. The sequence of the model is summarized in Figure 1. The equilibrium concept we employ is perfect Bayesian equilibrium.

3 First-best Benchmark

We begin with a benchmark case that will be a helpful point of comparison for the ensuing analysis. In particular, we consider the first-best investment decision under the assumption that the manager is always informed (i.e., $q_f = q_g = 1$), or equivalently, the market always observes future cash flows f and the social externality g following the investment decision.[11]

The manager's payoff from following investment strategy σ and disclosure strategy $d(s)$ is

$$U(s, d(s), \sigma) = \omega(a_l - x_h + \alpha E[f|d(s), \sigma] + \beta E[g|d(s), \sigma]) + (1-\omega)(a_l + \alpha E[f|d(s), \sigma]).$$

(Recall that $\alpha = a_h - a_l$ and $\beta = x_h - x_l$.) The manager's payoff can be expressed more simply, as the market's inference of f and g is always perfect. We let $k \equiv \omega(a_l - x_h) + (1-\omega)a_l$ denote manager's baseline payoff and determine the manager's expected utility as

$$\begin{aligned} E_\tau[U(s, d(s), \sigma)] &= k + \omega(\alpha\phi_\tau + \beta p_\tau) + (1-\omega)\alpha\phi_\tau \qquad (3) \\ &= k + \alpha\phi_\tau + \omega\beta p_\tau. \end{aligned}$$

By maximizing equation (3), we derive the manager's optimal investment strategy in this first-best case.

Proposition 1. *Under the first-best benchmark, there exists a unique equilibrium where the*

[10] More specifically, the manager derives utility proportional to the total amount of raised capital from the two types of investors. If investors incur a private cost of investment, given as

$$i = \arg\max_i E[v|d, \sigma]i - \frac{i^2}{2} = E[v|d, \sigma],$$

then the capital raised from social investors will be $\omega E[v|d, \sigma]$. Similarly, the capital raised from financial investors will be $(1-\omega)E[a|d, \sigma]$, so that the total capital is given by $\omega E[v|d, \sigma] + (1-\omega)E[a|d, \sigma]$.

[11] By a standard unraveling argument, we show in the Appendix that non-disclosure of a dimension results in the lowest possible market beliefs of that dimension.

manager selects the clean investment with probability one (i.e., $\sigma = 1$) when

$$\omega > \omega_{FB} \equiv \frac{\alpha \triangle_\phi}{\beta \triangle_p}. \tag{4}$$

Otherwise, the manager selects the non-renewable investment ($\sigma = 0$) in the unique equilibrium.

In Proposition 1, we see that the manager chooses the clean investment when the fraction of social investors is sufficiently high. By selecting $\tau = C$, the manager is willing to sacrifice a lower Bayesian belief of the financial value for a higher expected belief among social investors. Moreover, the improvement in social investor beliefs in terms of social value must be sufficiently high to overcome the decline in both the financial investors' beliefs as well as the financial value component of social investors' beliefs. Hence, when the expected drop in financial value from selecting the clean investment, captured in part by the numerator $\alpha\triangle_\phi$ in the right-hand side of condition (4), is very large, a greater mass of social investors is necessary for the clean investment to be worthwhile.

Proposition 1 also implies that the manager fully internalizes shareholder welfare in the investment decision in this case. In particular, the first-best level maximizes aggregate shareholder utility. We later use this first-best level to measure distortions in the investment decision and reductions in aggregate shareholder welfare.

4 Equilibrium

We now solve for the equilibrium of our baseline setting where the manager is probabilistically informed regarding each dimension and can strategically withhold information, i.e., the *voluntary regime*. Due to the presence of multiple disclosure outcomes, we build from the work of Shin (2003, 2006) and conjecture that the manager follows a *sanitation disclosure strategy* in equilibrium. In particular, under this conjectured strategy, the manager discloses only positive outcomes and withholds negative ones, when informed. As noted by Shin (2003), this disclosure strategy is appealing due to its simplicity and since it embeds desirable properties. In particular, sanitation is optimal when the manager's utility is monotone in aggregate shareholder beliefs, and thus when aggregate shareholder beliefs are monotone.[12]

[12]Stated differently, for $s = (f, g)$ and $s' = (f', g')$, if $f + g \geq f' + g'$ then $U(s, d(s), \sigma) \geq U(s', d(s'), \sigma)$. This implies that the manager's payoff is lowest when no outcomes are disclosed and highest when both outcomes are disclosed.

Recall that the manager learns two outcomes—future cash flows, denoted by $f \in \{0, 1\}$, and the social externality, or ESG score, denoted by $g \in \{0, 1\}$, with probability q_f and q_g, respectively. We simplify notation by representing the manager's signals (and thus her information set) with the pair $s = (f, g)$. We use the notation $\emptyset$ to denote that the manager is uninformed on a particular dimension or if the market does not observe a dimension. Formally, the conjectured sanitation strategy maps the manager's disclosure decision as a function of her signals:

$$d(s) = \begin{cases} (\emptyset, \emptyset) & \text{if } s \in \{(\emptyset, \emptyset), (\emptyset, 0), (0, \emptyset), (0, 0)\} \\ (\emptyset, 1) & \text{if } s \in \{(\emptyset, 1), (0, 1)\} \\ (1, \emptyset) & \text{if } s \in \{(1, \emptyset), (1, 0)\} \\ (1, 1) & \text{if } s = (1, 1). \end{cases}$$

We first analyze the effects of disclosure on market beliefs regarding future cash flows f and the social externality g. We then show that the sanitation strategy is supported in equilibrium, which we refer to as the *sanitation equilibrium*. Finally, we solve for the optimal investment strategy.

Voluntary disclosure and market beliefs

In deriving the market beliefs, we consider both pure and mixed investment strategies, i.e., $\sigma \in [0, 1]$, where σ is the equilibrium probability that the manager selects the clean investment. We begin by taking the manager's investment strategy σ as given. When forming beliefs, the market considers the manager's investment strategy, disclosure strategy, as well as the disclosure decisions. An important point to note is that disclosure along one of the dimensions can influence the beliefs of both dimensions. This is due to the fact that outcomes (cash flows and ESG score) are interdependent, as they arise from the same underlying project choice.[13] In particular, the market updates its beliefs regarding g taking into account the disclosure decisions of both f and g. For example, if the manager discloses f, but g is not disclosed, then the market revises its beliefs about g using both f and the fact that g was not disclosed (as well as the investment strategy).

To better understand and disentangle the different sources of market updating, we present

[13]In terms of the disclosure subgame, this interdependency in outcomes is the main departure from Shin (2003, 2006). The second major departure is that these outcomes are endogenized in our setting, considered in the following section.

the belief revision as occurring sequentially in two steps.[14] Continuing with the aforementioned example, following disclosure of f and non-disclosure of g, we first consider the impact of disclosure of f on market beliefs regarding g, and then, after this first update, we examine how beliefs over g are updated after taking into account that g is not disclosed.

We denote the disclosure decision regarding future cash flows f as d_f. Given the sanitation disclosure strategy, the market observes either disclosure of $f = 1$ or non-disclosure of this dimension, i.e., $d_f \in \{\emptyset, 1\}$. Proceeding with the two-step belief revision process described above, we first examine the market inference that comes just from the disclosure decision about f. In other words, we first consider the posterior probability that the manager selected the clean investment in the first stage conditional *only* on the disclosure decision and strategy of f (and of the investment strategy). Conditional on non-disclosure of f, the posterior probability that the manager has chosen the clean investment, denoted by $\delta_\emptyset$, is given by

$$\begin{aligned}
\delta_\emptyset &= \Pr(\tau = C | d_f = \emptyset, \sigma) \\
&= \frac{\sigma \Pr(d_f = \emptyset | \tau = C)}{\sigma \Pr(d_f = \emptyset | \tau = C) + (1 - \sigma) \Pr(d_f = \emptyset | \tau = D)} \\
&= \frac{\sigma \left\{1 - q_f + q_f(1 - \phi_C)\right\}}{\sigma \left\{1 - q_f + q_f(1 - \phi_C)\right\} + (1 - \sigma) \left\{1 - q_f + q_f(1 - \phi_D)\right\}} \\
&= \frac{\sigma(1 - q_f \phi_C)}{1 - q_f \phi_\sigma}, \qquad (5)
\end{aligned}$$

where $\phi_\sigma \equiv \phi_D + \sigma(\phi_C - \phi_D)$. Likewise, the posterior that $\tau = C$ conditional on disclosure of f, denoted by δ_1, is

$$\delta_1 = \Pr(\tau = C | d_f = 1, \sigma) = \frac{\sigma \phi_C}{\phi_\sigma}. \qquad (6)$$

We observe two properties from these Bayesian updates. First, when the manager follows a pure strategy for investment, $\sigma \in \{0, 1\}$, the posterior of the clean investment given the disclosure decision d_f is the same in both cases, i.e., $\delta_\emptyset = \sigma = \delta_1$. This is natural, as the market should not have any residual uncertainty regarding the investment decision when the manager chooses a particular project with probability one.

Second, when the manager uses a mixed strategy, $\sigma \in (0, 1)$, the market updates on the likelihood of $\tau = C$ (and thus on g) given the disclosure decision of f, as there continues to

[14] We note that, due to the simultaneous nature of disclosure on both dimensions, this belief revision actually occurs simultaneously. We present updating as sequential here to better illustrate the influence of disclosure on market beliefs.

be uncertainty following the project decision. In particular, we have $\delta_\emptyset \geq \sigma \geq \delta_1$. That is, disclosing high future cash flows $f = 1$ lowers the market's belief that the clean project was chosen. This is due to the fact that the non-renewable investment has a higher likelihood of generating high future cash flows, and this outcome is less likely if the clean project is chosen. Disclosure of $f = 1$ therefore leads the market to place a lower weight on $\tau = C$. Similarly, upon non-disclosure of f, the market updates its belief *upward* that the clean investment was chosen. Because non-disclosure pools the low signal realization of future cash flows with uninformedness, the market places a higher weight on $\tau = C$ following non-disclosure, as the low cash flow signal is more likely to arise under this project.

Given the posterior probabilities of the clean investment above, we next determine the conditional expectation of g, following the disclosure decision of f, $d_f \in \{\emptyset, 1\}$. Continuing with our "sequential" market belief revision, we proceed to condition this belief update *only* on the disclosure decision of f and the investment strategy σ, and *not* on the disclosure decision of g. The market's conditional expectation of g in this case is given by

$$p_{d_f} \equiv E[g|d_f, \sigma] = \delta_{d_f} p_C + (1 - \delta_{d_f}) p_D. \tag{7}$$

Equations (5) and (6) imply that the market assesses a higher probability of the low social externality upon observing non-disclosure of f, relative to the prior, while this posterior is lowest when f is disclosed, i.e., $p_\emptyset \geq p_\sigma \geq p_1$. The reasoning is similar to that discussed earlier—high cash flows are less likely to arise under the clean investment, resulting in a downward belief revision of g when the market observes $f = 1$.

We now proceed to the second step of the market's Bayesian updating of g and factor in the manager's disclosure decision and strategy of g. Of course, if g is disclosed, the market's posterior of g is simply the disclosed value (under the sanitation strategy, this only occurs when $g = 1$). Following non-disclosure, when updating beliefs the market takes into account the fact that g can be strategically withheld following a poor realization. Let $\hat{g}_{d_f}$ denote the market belief conditional on the disclosure decision of f, $d_f \in \{\emptyset, 1\}$, non-disclosure of g, and the manager's disclosure and investment strategy: $\hat{g}_{d_f} \equiv E[g|d = (d_f, \emptyset), \sigma]$.

To derive this posterior belief $\hat{g}_{d_f}$ and to better illustrate the role of non-disclosure of g in belief revision, we expand p_{d_f} from equation (7) using the law of iterated expectations:

$$\begin{aligned} p_{d_f} &\equiv E[g|d_f, \sigma] = E[E[g|d = (d_f, d_g), \sigma]|d_f, \sigma] \\ &= \Pr(d_g = \emptyset|d_f, \sigma) E[g|d = (d_f, \emptyset), \sigma] + \Pr(d_g = 1|d_f, \sigma) E[g|d = (d_f, 1), \sigma] \\ &= (1 - q_g + q_g(1 - p_{d_f}))\hat{g}_{d_f} + q_g p_{d_f}. \end{aligned} \tag{8}$$

The first term in the right-hand side of equation (8) corresponds to the event that the market does not observe g, either due to strategic non-disclosure (probability $q_g(1-p_{d_f})$) or because the manager is uninformed on this dimension (probability $1-q_g$). The market's belief of g in this case is $\hat{g}_{d_f}$. The second term captures the event that the manager is informed and $g=1$ (probability $q_g p_{d_f}$), in which case the market's belief of g is equal to the disclosed value.

By solving for $\hat{g}_{d_f}$, we have

$$\hat{g}_{d_f} = \frac{1-q_g}{1-q_g p_{d_f}} p_{d_f}. \tag{9}$$

The posterior belief $\hat{g}_{d_f}$ reflects the interdependency between outcomes. As future cash flows and ESG performance are both influenced by the same underlying project choice, the market uses the disclosure behavior of both outcomes to draw inferences on a single dimension. Moreover, $\hat{g}_{d_f}$ exhibits the same negative complementarity as demonstrated above, where beliefs over g are decreasing in disclosure of f, i.e., $\hat{g}_\emptyset \geq \hat{g}_\sigma \geq \hat{g}_1$.

Thus far, we have only focused on the belief characterization of the ESG score g. By a similar argument, all of the above properties hold analogously for market beliefs regarding future cash flows following the disclosure decision. In particular, we can derive the market beliefs of f following non-disclosure of this dimension, denoted by $\hat{f}_{d_g}$, as

$$\hat{f}_{d_g} \equiv E[g|d=(\emptyset, d_g), \sigma] = \frac{1-q_f}{1-q_f \phi_{d_g}} \phi_{d_g}, \tag{10}$$

where ϕ_{d_g} is the market's posterior of high future cash flows conditional on the disclosure decision of g.[15] We analogously have that $\hat{f}_\emptyset \geq \hat{f}_\sigma \geq \hat{f}_1$. Having characterized market beliefs, the following proposition establishes that the sanitation strategy forms an equilibrium of the disclosure subgame.

Proposition 2. *Consider a fixed $\sigma \in [0,1]$. There exists a perfect Bayesian equilibrium in the disclosure subgame in which the manager's optimal disclosure strategy is the sanitation*

[15] Specifically, ϕ_{d_g} is determined as

$$\phi_{d_g} = \kappa_{d_g}\phi_C + (1-\kappa_{d_g})\phi_D,$$

where κ_{d_g} is the posterior of the clean technology conditional on the disclosure decision of g:

$$\begin{aligned} \kappa_\emptyset &= \frac{\sigma(1-q_g p_C)}{1-q_g p_\sigma}, \\ \kappa_1 &= \frac{\sigma p_C}{p_\sigma}. \end{aligned}$$

strategy, i.e., a sanitation equilibrium, if we have

$$\frac{\alpha(\hat{f}_\emptyset - \hat{f}_1)}{\beta(1-\hat{g}_\emptyset)} \leq \omega \leq \frac{\alpha(1-\hat{f}_\emptyset)}{\beta(\hat{g}_\emptyset - \hat{g}_1)} \tag{11}$$

In the event of non-disclosure on either dimension, market beliefs are given by equations (9) and (10) in any sanitation equilibrium.

Due to the nature of interdependence between future cash flows and ESG quality, whether the sanitation strategy forms an equilibrium under a mixed investment strategy $\sigma \in (0,1)$ is not straightforward. Condition (11) of Proposition 2 ensures that market beliefs are well-behaved when $\sigma \in (0,1)$. As noted by Shin (2003), the sanitation equilibrium is not unique in the disclosure game. Due to the presence of out-of-equilibrium beliefs, we may construct other, perhaps less plausible, equilibria (see, e.g., Appendix A of Shin (2003)), a point not uncommon in sequential games of incomplete information. However, sanitation appears to be the most intuitive and natural disclosure strategy, as it embeds reasonable properties such as good news disclosure and monotonicity in the manager's payoff and aggregate shareholder beliefs. Proposition 2 takes the investment strategy σ as exogenously given; we endogenize this choice in the following section.

Optimal project choice

Given the equilibrium in the disclosure subgame established in Proposition 2, we now investigate the manager's optimal project choice. The equilibrium project choice therefore takes into account market beliefs in the disclosure subgame.

The manager's expected utility from choosing project τ is given by

$$\begin{aligned} E_\tau[U(s,d(s),\sigma)] \;=\; & \underbrace{k}_{\substack{\text{Manager's} \\ \text{"baseline" payoff}}} + \underbrace{(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset)(1-q_g p_\tau)(1-q_f\phi_\tau)}_{\substack{\text{Payoff upon non-disclosure} \\ \text{of } f \text{ and } g}} \\ & + \underbrace{(\alpha\hat{f}_1 + \omega\beta)q_g p_\tau(1-q_f\phi_\tau)}_{\substack{\text{Payoff upon disclosure of } g \\ \text{and non-disclosure of } f}} + \underbrace{(\alpha + \omega\beta\hat{g}_1)q_f\phi_\tau(1-q_g p_\tau)}_{\substack{\text{Payoff upon disclosure of } f \\ \text{and non-disclosure of } g}} \\ & + \underbrace{(\alpha+\omega\beta)q_f\phi_\tau q_g p_\tau}_{\text{Payoff upon disclosure of } f \text{ and } g} \end{aligned}$$

(Recall that $\alpha = a_h - a_l$ and $\beta = x_h - x_l$.) The manager's project choice affects both the underlying outcomes and the manager's likelihood of disclosure. Moreover, in equilibrium, the market's conjectured investment strategy must coincide with the manager's desired strategy. The investment therefore determines beliefs in the disclosure subgame.

We first present the full characterization of the sanitation equilibrium. We then discuss the economic forces underlying this result.

Theorem 1. *The manager's investment strategy in the sanitation equilibrium is characterized as follows.*

(i) *The manager selects the clean investment with probability one (i.e., $\sigma = 1$) if and only if*

$$\omega \geq \overline{\omega} \equiv \omega_{FB} \frac{q_f(1-\phi_C)}{1-q_f\phi_C} \frac{1-q_g p_C}{q_g(1-p_C)}. \tag{12}$$

(ii) *The manager selects the non-renewable investment with probability one (i.e., $\sigma = 0$) if and only if*

$$\omega \leq \underline{\omega} \equiv \omega_{FB} \frac{q_f(1-\phi_D)}{1-q_f\phi_D} \frac{1-q_g p_D}{q_g(1-p_D)}, \tag{13}$$

where $\underline{\omega} < \overline{\omega}$.

(iii) *For $\omega \in (\underline{\omega}, \overline{\omega})$, there exists a unique $\sigma(\omega)$ in the sanitation equilibrium.*

Following the investment choice, the manager uses the sanitation disclosure strategy and the equilibrium of the disclosure subgame is characterized in Proposition 2.

Parts (i) and (ii) of Theorem 1 establish that the manager uses a pure strategy for the project choice, $\sigma \in \{0, 1\}$, given certain conditions on the fraction of financial and social investors. Similar to Proposition 1, the manager chooses the sustainable investment when the mass of social investors is sufficiently high, $\omega \geq \overline{\omega}$, and the non-renewable investment when the mass of social investors is sufficiently low, $\omega \leq \underline{\omega}$. These conditions are determined such that the manager cannot benefit through a private deviation from the pure strategy. For example, for $\tau = C$ to be a pure strategy in equilibrium, we must have $E_C[U(s, d(s), \sigma = 1)] \geq E_D[U(s, d(s), \sigma = 1)]$.

Furthermore, we see in part (iii) of Theorem 1 that there is a region under which no pure investment strategy equilibrium exists. That is, when $\omega \in (\underline{\omega}, \overline{\omega})$, the only possible equilibrium is in mixed strategies for the investment choice. The manager faces the following trade-off in her project selection. By choosing the clean (non-renewable) investment, the manager can maximize beliefs of social (financial) investors, but she generates a lower expected belief among financial (social) investors. In the range $\omega \in (\underline{\omega}, \overline{\omega})$, the manager attempts to capture the best of both worlds and inflate beliefs of both types of investors.

To see this, suppose that $\omega \in (\underline{\omega}, \overline{\omega})$, but the manager uses a pure strategy of selecting the clean investment. Under the conjectured strategy, upon non-disclosure of g, the market continues to assess a high likelihood that the high ESG outcome will be realized, as investors believe the clean project was chosen with probability one. The manager can therefore deviate to choosing the non-renewable technology to increase the chance that the high future cash flow outcome is realized. In this way, upon non-disclosure of g, the manager is not penalized heavily since the market places a higher probability that the manager is uninformed on this dimension and, moreover, by privately deviating the manager has a higher likelihood of disclosing $f = 1$. Hence, under any conjectured pure strategy when $\omega \in (\underline{\omega}, \overline{\omega})$, the manager can strictly improve by privately deviating and selecting the other project.

In contrast, when $\omega \geq \overline{\omega}$ or $\omega \leq \underline{\omega}$, the manager no longer finds the private deviation to be worthwhile and the pure investment strategy is supported in equilibrium. This occurs because non-disclosure on the investors' preferred dimension becomes too costly for the manager when the fraction of financial or social investors is sufficiently high. By privately deviating from, for example, $\tau = C$ to $\tau = D$, the manager increases the chance of non-disclosure of g. Even if investors believe that the clean investment was chosen with certainty, the posterior belief of $\hat{g}_{d_f}$ after non-disclosure is still lower than the belief following disclosure. When the fraction of social investors is sufficiently high, this downward revision from non-disclosure becomes too costly for the manager and she therefore has no incentive to deviate from the pure strategy of $\sigma = 1$.

The incentive to privately deviate is also reflected in conditions (12) and (13) that characterize the respective boundaries $\underline{\omega}$ and $\overline{\omega}$. For example, in condition (12), the degree of distortion relative to the first-best benchmark w_{FB} is

$$\frac{q_f(1-\phi_C)}{1-q_f\phi_C}\frac{1-q_g p_C}{q_g(1-p_C)} = \frac{\Pr(s_f=0|d_f=\emptyset,\tau=C)}{\Pr(s_g=0|d_g=\emptyset,\tau=C)}. \tag{14}$$

As shown above, this distortion term can be represented as the likelihood ratio of low cash flows over low ESG quality, both conditional on non-disclosure.[16] In other words, (14) is the manager's "punishment" from strategic non-disclosure of cash flows relative to non-disclosure of ESG quality. As the market's downward belief revision following non-disclosure of f becomes more severe (i.e., when the numerator increases), the manager is more inclined to privately deviate when the market expects the clean investment, which pushes $\overline{\omega}$ further

[16]The condition can be rewritten as

$$\frac{q_f(1-\phi_C)}{1-q_f\phi_C}\frac{1-q_g p_C}{q_g(1-p_C)} = \frac{q_f(1-\phi_C)}{1-q_f\phi_C} \Big/ \frac{q_g(1-p_C)}{1-q_g p_C},$$

where the numerator of the right-hand side is $\Pr(s_f = 0|d_f = \emptyset, \tau = C)$ and the denominator is $\Pr(s_g = 0|d_g = \emptyset, \tau = C)$.

from ω_{FB}. The opposite effect occurs as the market becomes more pessimistic following non-disclosure of g (i.e., when the denominator increases).

Finally, to complete the equilibrium characterization, we show that a unique mixed strategy $\sigma \in (0,1)$ of the project selection exists in the sanitation equilibrium. As noted above, this occurs under an intermediate fraction of social investors, $\omega \in (\underline{\omega}, \overline{\omega})$. For mixing to be sustained in equilibrium, the manager must be indifferent between selecting the clean and non-renewable technology:

$$E_C[U(s, d(s), \sigma)] = E_D[U(s, d(s), \sigma)]. \tag{15}$$

We later determine comparative statics for the mixed strategy distribution, as well for the bounds $\underline{\omega}$ and $\overline{\omega}$, in Section 6.

Theorem 1 shows that the presence of discretionary disclosure with multiple interdependent outcomes can distort investment decisions. Moreover, the departure from the first-best level naturally implies inefficiency in the investment choice. In the next section, we examine inefficiency and real effects of both the voluntary and mandatory disclosure regimes.

5 Efficiency and real effects: Comparison between mandatory and voluntary disclosure of ESG quality

In the previous section, we examine investment and disclosure behavior when the manager has discretion over the release of both future financial performance and ESG quality. While voluntary ESG quality disclosure is the prevailing convention in the U.S., at least twenty-five countries and the European Union have adopted mandatory ESG disclosure laws or directives (Krueger et al. (2021)). A pertinent question therefore is to what extent mandatory ESG disclosure affects investment decisions and which disclosure regime induces less investment distortion. We begin our analysis in this section by discussing the real investment effects from mandatory ESG disclosure. We then compare the differences in efficiency and real effects on investment between the two disclosure regimes.

To convey the economic forces driving the results, in the ensuing analysis we often consider the manager's incentive to privately deviate from the "market's conjecture" of a pure strategy. By this, we mean that we first consider a situation where the market's conjecture of the manager's project strategy σ coincides with the first-best implementation characterized in Proposition 1. We then consider the strength of the manager's incentive to depart from this first-best level in light of the disclosure regime. (We also use this heuristic approach in conveying intuition for the comparative statics analyses in Section 6.)

Mandatory ESG quality disclosure

We first examine the case of mandatory disclosure of ESG performance, i.e., the *mandatory regime*. We assume that, under the mandatory ESG disclosure regime, the manager must disclose g to the market. This can be thought of as regulation which requires firms to convey their ESG performance and imposes fines or penalties in the event noncompliance. Due to the disclosure regulation, managers can, for example, expend additional resources on information acquisition to obtain a signal of g. As such, we assume in this mandatory regime that the manager is always informed about g, i.e., $q_g = 1$ We also assume that the disclosed value g cannot be manipulated so that we can make a more direct comparison with the voluntary regime. Disclosure of future financial performance, f, continues to be voluntary and $q_f \in (0, 1)$.

We solve for the cutoffs for clean and non-renewable investment as in Proposition 1. By substituting $q_g = 1$ into conditions (12) and (13), the upper and lower thresholds for a pure investment strategy under mandatory ESG disclosure are respectively given by

$$\begin{aligned} \overline{\omega}_m &\equiv \omega_{FB} \frac{q_f(1-\phi_C)}{1-q_f\phi_C}, \\ \underline{\omega}_m &\equiv \omega_{FB} \frac{q_f(1-\phi_D)}{1-q_f\phi_D}. \end{aligned}$$

An important property of these bounds is that both are lower than the first-best benchmark,

$$\underline{\omega}_m < \overline{\omega}_m < \omega_{FB}, \tag{16}$$

which implies that mandatory ESG disclosure induces *over-investment* in the clean technology relative to the first-best benchmark for $\omega \in (\underline{\omega}_m, \omega_{FB})$. Moreover, under-investment (over-investment) of the clean (non-renewable) technology does not occur for any ω, and thus, in expectation, over-investment of the clean technology is always positive in the mandatory regime.[17] We note that this property emerges without additional conditions on the exogenous parameters.

[17] While we have not specified a distribution for ω, this implication holds as long as there is positive density over $\omega \in (\underline{\omega}_m, \omega_{FB})$ in the distribution of ω.

Proposition 3. *Under mandatory disclosure of ESG quality, over-investment in the clean technology relative to the first-best benchmark occurs with probability one for* $\omega \in (\overline{\omega}_m, \omega_{FB})$ *and with positive probability for* $\omega \in (\underline{\omega}_m, \overline{\omega}_m)$. *Under-investment in the clean technology does not occur for any* ω.

This result is perhaps surprising, as intuition would suggest that the manager has little incentive to deviate from the first-best level and excessively adopt the clean investment when she does not have private information along the ESG dimension. To better understand Proposition 3, consider the situation where the manager has discretion over both dimensions, if informed. As discussed in Section 4, for intermediate levels of ω, the manager has an incentive to privately deviate from a pure strategy. For example, if the market expects the clean investment (i.e., $\tau = C$ or $\sigma = 1$), then by privately deviating to $\tau = D$, the manager can hide negative news under inflated market beliefs over ESG quality, in an attempt to attain the high outcome on future cash flows. In contrast, when the market always observes the ESG outcome, as in the mandatory regime, the manager has little incentive to deviate from the market's conjecture of $\tau = C$. By deviating from the pure strategy of $\tau = C$, the manager can no longer hide poor ESG realizations nor capitalize on inflated market beliefs of ESG quality following non-disclosure. Hence, following a private deviation from $\tau = C$ to $\tau = D$, the outcome is more likely to be $g = 0$, which will be accordingly factored into market beliefs upon revelation.

The above argument would suggest *more* efficient investment of the clean technology, as the manager has less incentive to deviate from the conjectured pure strategy of $\tau = C$. However, in the mandatory regime, the manager continues to have discretion over the financial dimension, f. She therefore continues to benefit from privately deviating from the market's conjecture of the pure strategy of selecting the non-renewable technology $\tau = D$ (under intermediate ω). Consequently, since the manager's benefit from private deviation is preserved when the market expects $\tau = D$ (under intermediate ω), but is shut off when the market expects $\tau = C$, over-investment in the clean technology is always induced under the mandatory regime. Figure 2 illustrates the departure from first-best graphically, where the degree of inefficiency relative to the first-best level is the green shaded region.

Investment distortion in the voluntary regime

The equilibrium investment strategy under the voluntary regime also differs from the first-best implementation. This is exemplified in Theorem 1, where the manager uses a mixed strategy in the region $\omega \in (\underline{\omega}, \overline{\omega})$. Moreover, the bounds of the pure investment strategy

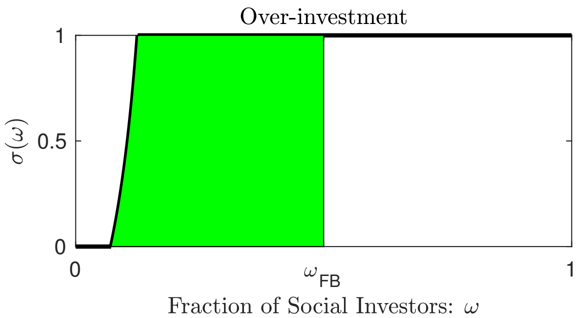

Figure 2: **Over-investment in the mandatory regime**
This figure plots the equilibrium technology choice for the mandatory regime: $q_g = 100\%$. The baseline parameters are $p_C = 30\%$, $p_D = 20\%$, $\phi_C = 50\%$, $\phi_D = 75\%$, $\alpha = 0.2$, $\beta = 1$, and $q_f = 40\%$. The first-best threshold is $\omega_{FB} = 50\%$. The shaded area represents inefficiency relative to the first-best implementation.

under voluntary disclosure can also vary from the first-best level. The following corollary establishes that over- or under-investment in the clean technology can occur in the equilibrium of the voluntary regime.

Corollary 1. *Under the sanitation equilibrium in the voluntary disclosure regime, under-investment in the clean technology relative to the first-best benchmark occurs if we have* $\underline{\omega} \geq \omega_{FB}$*, while over-investment in the clean technology occurs if we have* $\overline{\omega} \leq \omega_{FB}$*.*

A perhaps unexpected finding from Corollary 1 is that voluntary disclosure does not necessarily lead to under-investment in the clean technology relative to the first-best benchmark. Suppose that we have $\underline{\omega} \geq \omega_{FB}$. This implies that for $\omega \in (\omega_{FB}, \overline{\omega})$, the manager optimally chooses the non-renewable technology with positive probability, even though she chooses the clean investment with probability one in the first-best benchmark. The condition that $\underline{\omega} \geq \omega_{FB}$ is equivalent to

$$\Pr(s_f = 0 | d_f = \emptyset, \tau = D) \geq \Pr(s_g = 0 | d_g = \emptyset, \tau = D), \quad (17)$$

which states that, conditional on $\tau = D$, the likelihood that the manager observed a low signal of future cash flows upon non-disclosure of f is at least as high as the analogous likelihood of the low signal of ESG quality upon non-disclosure of g. Condition (17) implies that non-disclosure of f is more punitive for the manager, which limits her incentive to privately deviate from the non-renewable technology if $\tau = D$ is the conjectured pure strategy. This therefore leads to over-adoption (under-adoption) of the non-renewable (clean) project.

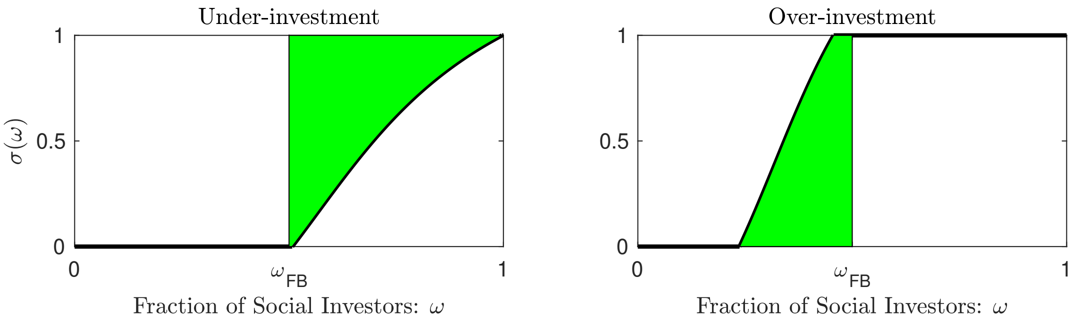

Figure 3: **Under- and Over-investment in the clean technology**
This figure plots the equilibrium technology choice for two cases: $q_g = 17\%$ (left panel) and 35% (right panel). The baseline parameters are $p_C = 30\%$, $p_D = 20\%$, $\phi_C = 50\%$, $\phi_D = 75\%$, $\alpha = 0.2$, $\beta = 1$, and $q_f = 40\%$. The first-best threshold is $\omega_{FB} = 50\%$. The shaded area represents inefficiency relative to the first-best.

Condition (17) is satisfied, for example, when q_f is high relative to q_g.[18]

Likewise, the condition $\overline{\omega} \leq \omega_{FB}$ implies that

$$\Pr(s_f = 0 | d_f = \emptyset, \tau = C) \leq \Pr(s_g = 0 | d_g = \emptyset, \tau = C). \tag{19}$$

By a similar reasoning as above, the manager is less inclined to privately deviate from the market's conjecture of the pure strategy $\tau = C$ in this case, which implies over-investment in the clean project. Figure 3 portrays the dual inefficiencies in the voluntary regime.

Mandatory vs. voluntary ESG disclosure

In the preceding analysis, we see that both disclosure regimes induce investment distortions. We now compare the two regimes in terms of their investment efficiency. To measure investment inefficiency, we consider the deviation from the first-best benchmark. Under the first-best implementation, the manager seeks to maximize aggregate shareholder welfare according to investor preferences. Any deviation from the first-best benchmark therefore implies a net loss for shareholders in the aggregate.

We see that the manager's ex ante payoff, and thus the shareholders' ex ante aggregate payoff, is given by

$$V(\sigma) = k + \sigma E_C[U(s, d(s), \sigma)] + (1 - \sigma) E_D[U(s, d(s), \sigma].$$

[18]Equation (17) can written as

$$\Pr(s_f = 0 | d_f = \emptyset, \tau = D) = \frac{q_f(1 - \phi_D)}{1 - q_f \phi_D} \geq \frac{q_g(1 - p_D)}{1 - q_g p_D} = \Pr(s_g = 0 | d_g = \emptyset, \tau = D). \tag{18}$$

After some calculations, we can express this payoff as

$$V(\sigma) = k + \alpha\phi_\sigma + \omega\beta p_\sigma,$$

where σ is the optimal investment choice under either the voluntary or mandatory regime. Our specific measure of real efficiency is therefore defined as the ratio between the expected aggregate shareholder welfare in equilibrium and the analogous welfare under the first-best implementation:

$$RE = \frac{k + \alpha\phi_\sigma + \omega\beta p_\sigma}{k + \alpha\phi_{FB} + \omega\beta p_{FB}}, \tag{20}$$

where $\phi_{FB} = \phi_C$ and $p_{FB} = p_C$ if $\omega > \omega_{FB}$, while $\phi_{FB} = \phi_D$ and $p_{FB} = p_D$ if $\omega < \omega_{FB}$.

The prior analysis in this section suggests that efficiency comparisons between the two regimes are not straightforward, as the voluntary regime includes two kinds of investment distortion which can occur under different conditions. The following theorem provides an efficiency ranking of the two regimes based on the composition of shareholders.

Theorem 2. *Real efficiency (RE) in the voluntary regime is always weakly greater than RE in the mandatory regime for $\omega < \omega_{FB}$. RE in the voluntary regime is always weakly lower than in the mandatory regime for $\omega > \omega_{FB}$.*

Theorem 2 establishes that the voluntary regime can be efficiency-enhancing relative to mandatory disclosure when the fraction of social investors is not sufficiently high. This is perhaps surprising, as the prior analysis shows that the manager has only one deviation channel in the mandatory regime, whereas she has two incentives for private deviation under the voluntary regime. Hence, one might expect that the mandatory regime should always be more efficient than the voluntary regime, since the manager's incentive to privately deviate ultimately gives rise to inefficiency. However, two effects are at play in the mandatory regime. The first effect is the one just mentioned—the manager has little incentive to privately deviate when the market expects the clean technology, $\tau = C$. This effect raises real efficiency relative to the voluntary regime and brings project selection closer to the first-best level.

The second effect is that the presence of mandatory disclosure of ESG quality *intensifies* the manager's incentive to privately deviate when the market expects the non-renewable investment, $\tau = D$. In contrast to the voluntary regime, the manager cannot hide poor signal realizations of ESG quality by mimicking uninformedness when ESG quality is mandatorily disclosed. Consequently, the manager endures a greater "punishment" from poor ESG outcomes in the mandatory regime relative to voluntary disclosure—even if the market expects the non-renewable investment and thus the poor ESG outcome. Moreover, the manager continues to maintain the benefit of hiding poor signal realizations of future financial performance in the mandatory regime. The manager is therefore more often willing to accept the non-disclosure belief over financial performance ($\hat{f}_{d_g}$) in the mandatory regime in an attempt to raise beliefs about ESG quality through private deviation. In other words, the lower expected payoff following realizations of $g = 0$ in the mandatory regime intensifies the manager's incentive to privately deviate when the market expects the non-renewable investment. This second effect *lowers* real efficiency relative to the voluntary regime and drives project selection further from the first-best implementation.

Theorem 2 establishes that the first effect dominates when the fraction of social investors is sufficiently high, $\omega > \omega_{FB}$, resulting in greater efficiency under the mandatory regime. When $\omega > \omega_{FB}$, the first-best project selection is the clean investment (Proposition 1). The manager has little incentive to privately deviate from the market's conjecture of $\tau = C$ under mandatory disclose (Proposition 3), but she can continue to have this incentive under the voluntary regime (Theorem 1 and Corollary 1). Consequently, the manager's project decision is more aligned with aggregate shareholder preferences in the mandatory regime when $\omega > \omega_{FB}$.

Relatedly, the aforementioned second effect dominates when the fraction of social investors is sufficiently low, $\omega < \omega_{FB}$, which implies that the voluntary regime is more efficient. In this case, the manager selects the non-renewable project $\tau = D$ in the first-best benchmark. The manager departs from this selection in both the mandatory and voluntary regimes, but she departs *more* under the mandatory regime (i.e., the manager more often selects $\tau = C$), due to her greater incentive to privately deviate from the market's conjecture of $\tau = D$. As a result, aggregate shareholder welfare, and thus real efficiency, declines more under mandatory disclosure when $\omega < \omega_{FB}$.

Figure 4 illustrates the optimal project choice and the corresponding real efficiency levels for a specific parameterization. We see in the left panel that the manager begins to select the clean project with positive probability at a lower level of ω (and below ω_{FB}), as compared to the voluntary regime, consistent with Proposition 3.

An interesting implication of Theorem 2 is that the manager more often invests in the clean project under mandatory disclosure than she does in the voluntary regime. That is, under the mandatory regime, the manager's threshold for implementing the clean project with probability one is lower and a greater probability mass is placed on $\tau = C$ in the mixed region. This implies that shifting from a voluntary to mandatory ESG disclosure regime weakly increases the prevalence that the clean project is chosen.

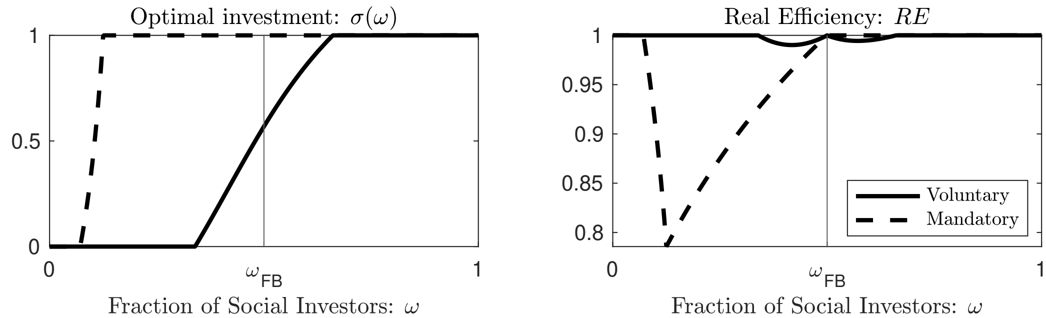

Figure 4: **Mandatory vs. voluntary ESG disclosure**
This figure plots the equilibrium project choice (left panel) and real efficiency (right panel) under the voluntary and mandatory disclosure regimes. The baseline parameters are $p_C = 30\%$, $p_D = 20\%$, $\phi_C = 50\%$, $\phi_D = 75\%$, $\alpha = 0.2$, $\beta = 1$, $q_g = 25\%$, and $q_f = 40\%$. The first-best threshold is $\omega_{FB} = 50\%$.

Corollary 2. *The probability that the clean project is chosen is weakly higher in the mandatory regime than in the voluntary regime.*

6 Comparative statics and empirical predictions

In this section, we discuss empirical predictions that arise from the model. These implications are with respect to the likelihood of ESG quality disclosure, prevalence of clean technology adoption, and level of real investment efficiency. Moreover, our comparative statics analysis provides predictions for cross-industry (or cross-firms) variation in these outcomes, which may be helpful in guiding future empirical research. While several of our implications have not yet been explored empirically, we make connections to the empirical literature when possible.

The equilibrium characterization in Theorem 1 implies greater adoption of clean projects and more ESG quality disclosure when there is a greater share of investors who have preferences regarding the non-financial performance of the firm. This aligns with recent empirical evidence that firms with greater social investor ownership provide more ESG disclosures (Dhaliwal et al. (2011), Ilhan et al. (2020), Pawliczek et al. (2021)) and invest in more sustainable projects (Dyck et al. (2019), Chen et al. (2020)). Moreover, Proposition 3 suggests that we should observe over-investment in clean projects in mandatory ESG disclosure regimes, on average. Similarly, sustainable investments should be more prevalent among firms in mandatory ESG disclosure regimes relative to firms in voluntary regimes, on average. Likewise, as implied by Corollary 2, we expect to see a greater prevalence of clean project implementation upon shifting from a voluntary to a mandatory ESG disclosure regime within a country or region. This helps to explain the findings of Chen et al. (2018), Downar et al.

(2021), and Jouvenot and Krueger (2021) that a within-country shift to mandatory environmental disclosure led to, on average, greater adoption of ESG activities and lower negative social externalities produced by firms.

In addition, Corollary 1 implies variation within voluntary disclosure regimes regarding over- or under-investment of clean projects. Our comparative statics analysis provides insights regarding variation in investment and disclosure across firms or industries within a disclosure regime:

Proposition 4. *In the voluntary regime, the upper and lower boundaries of the mixed strategy region are decreasing in the probability of being informed about ESG quality, q_g. In the mandatory and voluntary regimes, the upper and lower boundaries are increasing in the probability of being informed about future financial performance, q_f, and α/β. Moreover, the mixed strategy region is narrower if q_g is higher and wider if α/β is greater.*

Proposition 4 states that, under the voluntary regime, the boundaries $\underline{\omega}$ and $\overline{\omega}$ are both decreasing as the firm is more likely to be informed regarding ESG quality. The reason for this is that market beliefs following non-disclosure of g decline as q_g increases, as it becomes more likely that the manager is mimicking uninformedness after observing a poor signal realization. In turn, the manager has less incentive to privately deviate when the market expects $\tau = C$, leading to a decrease in $\overline{\omega}$, and more inclined to privately deviate when the market expects $\tau = D$, resulting in a decrease of $\underline{\omega}$. Moreover, the decreased bounds $\underline{\omega}$ and $\overline{\omega}$ mean that the firm is more often adopting the clean technology. Since clean investment leads to a higher likelihood of positive ESG quality signal realizations, the lower bounds similarly imply a higher likelihood of ESG disclosure. This implies that, we should expect greater adoption of sustainable projects/investments and greater ESG disclosure among firms or industries that have less uncertainty or better information over the ESG quality of their projects/investments.

Relatedly, in both disclosure regimes, the upper and lower boundaries of the mixed strategy region are increasing in the likelihood that the manager is informed about future financial performance, q_f. The reasoning is analogous to that of q_g above; due to lower market beliefs following non-disclosure of f, the manager has a lower (higher) incentive to privately deviate when the market expects the non-renewable (clean) investment. The parameter q_f can correspond to the firm's uncertainty over its future financial performance. For example, a low q_f aligns with firms or industries with substantial performance uncertainty, such as growth industries or industries with rapidly evolving product markets, whereas a high q_g can correspond to firms in more stable industries. As the manager adopts the non-renewable technology more often as q_f increases, we expect less clean investment and less ESG quality disclosure in stable industries and more clean adoption and ESG disclosure in growth industries.

In terms of the relative project realization values, α/β, the manager is more inclined to choose the non-renewable investment as this project becomes more appealing relative to the clean investment (recall that $\alpha = a_h - a_l$ and $\beta = x_h - x_l$). This is natural, as the incentive to privately deviate from $\tau = D$ becomes weaker as α increases or β decreases. Accordingly, we expect less ESG disclosure as the upside value of the non-renewable project increases.

As discussed previously, the manager does not use a pure investment strategy in the mixed strategy region $(\underline{\omega}, \overline{\omega})$ or $(\underline{\omega}_m, \overline{\omega}_m)$. Consequently, investors are uncertain as to the manager's project selection in this region. We therefore expect greater information asymmetry among firms or industries that have an intermediate level of investors with social preferences relative to firms/industries with a very high or low proportion of these investors. We additionally see in Proposition 4 that the size of the mixed strategy region expands as α/β increases. This implies that the upper threshold, $\overline{\omega}$ or $\overline{\omega}_m$, increases at a faster rate than $\underline{\omega}$ or $\underline{\omega}_m$, respectively, as α/β increases. Similarly, the length of the mixed region truncates as q_g increases.

While the above analysis examines the boundaries of the mixed strategy region, we find that the manager's incentives behave similarly within this region:

Proposition 5. *The equilibrium probability for the manager to choose the sustainable project, σ, is weakly increasing in ω, β, q_g, and weakly decreasing in α and q_f.*

Recall that the mixed strategy distribution $\sigma \in (0, 1)$ is determined such that the manager's indifference condition, $E_C[U(s, d(s), \sigma)] = E_D[U(s, d(s), \sigma)]$, is satisfied. Hence, a change in one parameter requires σ to be adjusted so that indifference continues to hold. For example, the manager's likelihood of selecting the clean technology is increasing in the fraction of social investors, ω. This occurs because the manager suffers a greater loss following non-disclosure of g (relative to disclosure of $g = 1$) when the fraction of social investors is higher. Accordingly, a greater weight must be placed on the sustainable project when ω is higher to satisfy the indifference condition. The other relations established in Proposition 5 are in line with those in Proposition 4. In particular, σ is increasing in q_g and β, as the manager's incentive to select the clean project increases. Likewise, the manager is more inclined to the invest in the non-renewable project (i.e., σ decreases) as α or q_f increases.

The model also provides implications regarding investment efficiency, defined by equation (20). Theorem 2 implies that real efficiency should be higher in voluntary ESG disclosure regimes relative to mandatory regimes for firms or industries with a low proportion of social investors, and vice versa when the share of social investors is high. In terms of within-region implications, as noted above, a within-region shift from voluntary to mandatory disclosure results in greater sustainable investment (Corollary 2), but such a shift may not be efficient. Our results predict that a shift to mandatory disclosure within a region is efficiency-increasing for firms that have a larger proportion of social investors and efficiency-decreasing for firms with a lower proportion of these investors, which implies an overall non-monotone effect on efficiency.

Grewal et al. (2019) finds a non-monotone effect on stock returns following the EU's adoption of ESG disclosure mandates. In particular, Grewal et al. (2019) documents that firms with low ESG disclosure and activity prior to the reform experienced an average negative return, while the return was positive for firms with high ESG disclosure in the pre-period. Our framework can help to explain these results: Low ESG disclosure firms in our model arise from a higher concentration of financial investors (i.e., low or intermediate ω), and these sophisticated investors consequently anticipated inefficient over-investment in ESG projects following the reform. To the extent that returns reflected the change in efficiency, our results help to explain this non-monotone finding.

The following proposition establishes further properties of real efficiency:

Proposition 6. *For $\omega < \omega_{FB}$, real efficiency is weakly decreasing in q_g and weakly increasing in q_f. For $\omega > \omega_{FB}$, the opposite holds.*

A property we observe from equation (20) is that, when $\omega < \omega_{FB}$, RE is decreasing in σ. In the region $\omega \in [0, \omega_{FB}]$, the manager can maximize aggregate shareholder welfare by selecting the non-renewable project with probability one. As greater weight is placed on the clean project in this region ($\sigma > 0$), the manager departs more from the first-best level and thus real efficiency declines. Likewise, RE is increasing in σ when $\omega > \omega_{FB}$ by a similar reasoning.

As established in Proposition 4, the boundaries $\underline{\omega}$ and $\overline{\omega}$ are decreasing in q_g, while Proposition 5 shows that σ is increasing in q_g. This implies that the manager more often chooses the clean project as q_g increases, which results in a weakly lower RE when $\omega < \omega_{FB}$ and a weakly higher efficiency when $\omega > \omega_{FB}$. Hence, real efficiency is non-monotone and U-shaped in the uncertainty firms have regarding their ESG quality. An analogous reasoning applies to q_f and thus RE is inverse U-shaped in the strength of firm information over future financial performance.

7 Concluding remarks

Environmental and social concerns have become prominent among investors in recent years. Voluntary disclosure of ESG performance during the same time has also risen tremendously. Meanwhile, numerous countries have implemented regulations requiring firms to disclose ESG information. We formally investigate the real investment effects of mandatory and voluntary ESG disclosure in a parsimonious model with multidimensional private information.

Our results show that, in the voluntary regime, the manager withholds bad signal realizations and releases good ones, consistent with the sanitation strategy of Shin (2003). Optimal investment in both regimes is distorted from the first-best level, as the manager can privately deviate in an attempt to manipulate market beliefs. We find that the mandatory regime results in over-investment in the clean technology, which suggests that mandatory ESG regulation can have unintended effects due to the presence of voluntary disclosure in other areas. In particular, the manager continues to have discretion to disclose information related to future financial performance. Consequently, the manager has a heightened incentive to privately deviate—relative to the voluntary regime—from the non-renewable to the sustainable project.

We additionally characterize conditions under which voluntary or mandatory disclosure is more efficient for investors. In particular, when the fraction of shareholders who care about ESG quality is not sufficiently high, the voluntary regime is more efficient and improves aggregate shareholder welfare relative to the mandatory regime. This result is perhaps surprising, as the manager has less incentive to privately deviate from the clean technology when she does not have discretion on this dimension. However, the fact that ESG quality is always disclosed under the mandatory regime intensifies the manager's incentive to privately deviate from the non-renewable project when the market expects this investment. This overreaction can consequently result in a lower real efficiency and a greater departure from the first-best level. In contrast, when the share of social investors is high, the manager's decreased incentive to privately deviate from the clean technology dominates, and voluntary disclosure is more efficient for shareholders.

The implications outlined in Section 6 include both cross-region and within-region predictions, which can be applicable to areas that have seen shifts in their ESG disclosure regimes. These include, cross-industry variation in the level of ESG disclosure or activity; for example, high growth industries should exhibit greater ESG disclosure than low growth or stable industries. The results also provide a novel prediction on efficiency and real effects—investment efficiency should be non-monotone following a shift in the disclosure regime. Nevertheless, we expect an increase, on average, in clean investment following a shift from voluntary to mandatory ESG reporting. Our results thus offer a number of avenues for future research and help to inform the policy debate on ESG disclosure.

References

Avramov, D., S. Cheng, A. Lioui, and A. Tarelli (2021): "Sustainable investing with ESG rating uncertainty," *Journal of Financial Economics*.

Baker, M., D. Bergstresser, G. Serafeim, and J. Wurgler (2018): "Financing the response to climate change: The pricing and ownership of US green bonds," Tech. rep., National Bureau of Economic Research.

Baker, S. D., B. Hollifield, and E. Osambela (2020): "Asset prices and portfolios with externalities," *Available at SSRN 3344940*.

Bauer, R., T. Ruof, and P. Smeets (2021): "Get real! Individuals prefer more sustainable investments," *The Review of Financial Studies*, 34, 3976–4043.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2018): "Disclosure and choice," *The Review of Economic Studies*, 85, 1471–1501.

Bertomeu, J. and I. Marinovic (2016): "A theory of hard and soft information," *The Accounting Review*, 91, 1–20.

Bertomeu, J., I. Vaysman, and W. Xue (2021): "Voluntary versus mandatory disclosure," *Review of Accounting Studies*, 26, 658–692.

Chen, T., H. Dong, and C. Lin (2020): "Institutional shareholders and corporate social responsibility," *Journal of Financial Economics*, 135, 483–504.

Chen, Y.-C., M. Hung, and Y. Wang (2018): "The effect of mandatory CSR disclosure on firm profitability and social externalities: Evidence from China," *Journal of accounting and economics*, 65, 169–190.

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

Christensen, H. B., L. Hail, and C. Leuz (2021): "Mandatory CSR and sustainability reporting: economic analysis and literature review," *Review of Accounting Studies*, 26, 1176–1248.

Dhaliwal, D. S., O. Z. Li, A. Tsang, and Y. G. Yang (2011): "Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting," *The accounting review*, 86, 59–100.

Downar, B., J. Ernstberger, S. Reichelstein, S. Schwenen, and A. Zaklan (2021): "The impact of carbon disclosure mandates on emissions and financial operating performance," *Review of Accounting Studies*, 26, 1137–1175.

Dyck, A., K. V. Lins, L. Roth, and H. F. Wagner (2019): "Do institutional investors drive corporate social responsibility? International evidence," *Journal of Financial Eco-*

nomics, 131, 693–714.

Dye, R. A. (1985): "Disclosure of nonproprietary information," *Journal of accounting research*, 123–145.

Edmans, A., M. S. Heinle, and C. Huang (2016): "The real costs of financial efficiency when some information is soft," *Review of Finance*, 20, 2151–2182.

Einhorn, E. (2005): "The nature of the interaction between mandatory and voluntary disclosures," *Journal of Accounting Research*, 43, 593–621.

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

Friedman, H. L. and M. S. Heinle (2016): "Taste, information, and asset prices: Implications for the valuation of CSR," *Review of Accounting Studies*, 21, 740–767.

——— (2021): "Interested investors and intermediaries: When do esg concerns lead to esg performance?" *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Friedman, H. L., M. S. Heinle, and I. M. Luneva (2021a): "A Theoretical Framework for ESG Reporting to Investors," *Available at SSRN*.

Friedman, H. L., J. S. Hughes, and B. Michaeli (2021b): "A rationale for imperfect reporting standards," *Management Science*.

Gigler, F. and T. Hemmer (1998): "On the frequency, quality, and informational role of mandatory financial reports," *Journal of Accounting Research*, 36, 117–147.

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*, 101889.

Goldstein, I., A. Kopytov, L. Shen, and H. Xiang (2022): "On esg investing: Heterogeneous preferences, information, and asset prices," Tech. rep., National Bureau of Economic Research.

Goldstein, I. and L. Yang (2019): "Good disclosure, bad disclosure," *Journal of Financial Economics*, 131, 118–138.

Gollier, C. and S. Pouget (2014): "The" washing machine": Investment strategies and corporate behavior with socially responsible investors," *TSE Working Paper*.

Green, D. and B. Roth (2021): "The allocation of socially responsible capital," *Available at SSRN 3737772*.

Grewal, J., E. J. Riedl, and G. Serafeim (2019): "Market reaction to mandatory nonfinancial disclosure," *Management Science*, 65, 3061–3084.

Grossman, S. J. (1981): "The informational role of warranties and private disclosure about product quality," *The Journal of Law and Economics*, 24, 461–483.

Grossman, S. J. and O. D. Hart (1980): "Disclosure laws and takeover bids," *The Journal of Finance*, 35, 323–334.

Gupta, D., A. Kopytov, and J. Starmans (2021): "The Pace of Change: Socially Responsible Investing in Private Markets," *Available at SSRN 3896511*.

Guttman, I. and X. Meng (2021): "The effect of voluntary disclosure on investment inefficiency," *The Accounting Review*, 96, 199–223.

Hartzmark, S. M. and A. B. Sussman (2019): "Do investors value sustainability? A natural experiment examining ranking and fund flows," *The Journal of Finance*, 74, 2789–2837.

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

Humphrey, J., S. Kogan, J. Sagi, and L. Starks (2021): "The asymmetry in responsible investing preferences," Tech. rep., National Bureau of Economic Research.

Ilhan, E., P. Krueger, Z. Sautner, and L. T. Starks (2020): "Climate risk disclosure and institutional investors," *Swiss Finance Institute Research Paper*.

Jouvenot, V. and P. Krueger (2021): "Mandatory corporate carbon disclosure: Evidence from a natural experiment," *Available at SSRN 3434490*.

Jung, W.-O. and Y. K. Kwon (1988): "Disclosure when the market is unsure of information endowment of managers," *Journal of Accounting research*, 146–153.

Kanodia, C. and D. Lee (1998): "Investment and disclosure: The disciplinary role of periodic performance reports," *Journal of accounting research*, 36, 33–55.

Kanodia, C. and H. Sapra (2016): "A real effects perspective to accounting measurement and disclosure: Implications and insights for future research," *Journal of Accounting Research*, 54, 623–676.

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

Krueger, P., Z. Sautner, D. Y. Tang, and R. Zhong (2021): "The Effects of Mandatory ESG Disclosure around the World," *Available at SSRN 3832745*.

Kumar, P., N. Langberg, and K. Sivaramakrishnan (2012): "Voluntary disclosures, corporate control, and investment," *Journal of Accounting Research*, 50, 1041–1076.

Landier, A. and S. Lovo (2020): "ESG Investing: How to Optimize Impact?" *HEC Paris Research Paper No. FIN-2020-1363*.

Luo, H. A. and R. J. Balvers (2017): "Social screens and systematic investor boycott risk," *Journal of Financial and Quantitative Analysis*, 52, 365–399.

Lyon, T. P. and J. W. Maxwell (2011): "Greenwash: Corporate environmental disclosure under threat of audit," *Journal of Economics & Management Strategy*, 20, 3–41.

MacMillan, D. and M. Joselow (2022): "SEC plans to force public companies to disclose greenhouse gas emissions," *The Washington Post*, March 15, 2022.

Marinovic, I. and F. Varas (2016): "No news is good news: Voluntary disclosure in the face of litigation," *The RAND Journal of Economics*, 47, 822–856.

Milgrom, P. R. (1981): "Good news and bad news: Representation theorems and applications," *The Bell Journal of Economics*, 380–391.

Oehmke, M. and M. M. Opp (2020): "A theory of socially responsible investment," .

Pae, S. (2005): "Selective disclosures in the presence of uncertainty about information endowment," *Journal of Accounting and Economics*, 39, 383–409.

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

Pawliczek, A., A. N. Skinner, and L. Wellman (2021): "A New Take on Voice: The Influence of BlackRock's 'Dear CEO' Letters," *Available at SSRN 3763042*.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski (2021): "Responsible investing: The ESG-efficient frontier," *Journal of Financial Economics*, 142, 572–597.

Riedl, A. and P. Smeets (2017): "Why do investors hold socially responsible mutual funds?" *The Journal of Finance*, 72, 2505–2550.

Shin, H. S. (2003): "Disclosures and asset returns," *Econometrica*, 71, 105–133.

——— (2006): "Disclosure risk and price drift," *Journal of Accounting Research*, 44, 351–379.

US SIF (2020): "Report on US sustainable and impact investing trends," https://www.ussif.org/trends.

Wen, X. (2013): "Voluntary disclosure and investment," *Contemporary Accounting Research*, 30, 677–696.

Zerbib, O. D. (2020): "A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from green investing and sin stock exclusion," *Available at SSRN 3455090*.

Appendix

Proof of Proposition 1

Suppose that the manager is always informed of the firm's total value, $q_f = q_g = 1$. The sanitation disclosure strategy forms an equilibrium if we have

$$U(s, d(s), \sigma) \geq U(s, d'(s), \sigma)$$

By inspecting all possible alternative disclosure choices for $s \in \{(0,0),(0,1),(1,0),(1,1)\}$, we obtain the following four conditions:

$$\begin{aligned}
\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset &\geq \alpha\hat{f}_0 && (21)\\
\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset &\geq \omega\beta\hat{g}_0 && (22)\\
\alpha\hat{f}_1 + \omega\beta &\geq \alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset && (23)\\
\alpha + \omega\beta\hat{g}_1 &\geq \alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset, && (24)
\end{aligned}$$

where $\hat{f}_0$ is off-equilibrium belief about f upon observing nondisclosure of f and disclosure of $g = 0$. Similarly, $\hat{g}_0$ is off-equilibrium belief about g upon observing disclosure of $f = 0$ and nondisclosure of g.

We can easily see that $\hat{g}_\emptyset = 0$ and $\hat{g}_1 = 0$ since $d(s) = (\emptyset,\emptyset),(1,\emptyset)$ correspond to $s = (0,0),(1,0)$, respectively. Similarly, we can see that $\hat{f}_\emptyset = \hat{f}_1 = 0$. We can also determine off-equilibrium belief $\hat{g}_0$ as follows. The expectation of g conditional on $f = 0$ is given by

$$p_0 \equiv \delta_0 E_C[g] + (1-\delta_0)E_D[g] = \delta_0 p_C + (1-\delta_0)p_D$$

where δ_0 is the posterior probability that the manager has chosen the technology C conditional on $f = 0$:

$$\begin{aligned}
\delta_0 &= \Pr(\tau = C | f = 0, \sigma)\\
&= \frac{\sigma \Pr(f = 0|\tau = C)}{\sigma \Pr(f = 0|\tau = C) + (1-\sigma)\Pr(f = 0|\tau = D)}\\
&= \frac{\sigma(1-\phi_C)}{1-\phi_\sigma} > \sigma
\end{aligned}$$

This implies that $p_0 > p_\sigma$. We should have

$$\begin{aligned}
p_0 &= \delta_0(p_C + (1-p_C)\hat{g}_0) + (1-\delta_0)(p_D + (1-p_D)\hat{g}_0)\\
&= p_0 + (1-p_0)\hat{g}_0
\end{aligned}$$

Thus, we have $\hat{g}_0 = 0$. Similarly, we can see that $\hat{f}_0 = 0$. It is obvious that market's beliefs satisfy the conditions (21) – (24) so that the sanitation disclosure strategy forms an equilibrium.

The manager selects the clean investment with probability one if

$$\alpha\phi_C + \omega\beta p_C > \alpha\phi_D + \omega\beta p_D. \leftrightarrow \omega > \omega_{FB}$$

Otherwise, the manager chooses the non-renewable investment.

Proof of Proposition 2

Our conjectured disclosure strategy forms an equilibrium if we have

$$U(s, d(s), \sigma) \geq U(s, d'(s), \sigma)$$

By considering each of the manager's signals and any deviation from the sanitation strategy, it is easy to check that the sanitation strategy forms an equilibrium if the conditions (21) – (24) hold.

To complete the construction of the sanitation equilibrium, we must specify market beliefs following events that are not reached in equilibrium. In particular, upon disclosure of the low signal in either dimension and non-disclosure in the other dimension (e.g., consider a deviant disclosure of $d = (\emptyset, 0)$), the market must have some belief regarding f in this off-path event. Rather than deriving off-equilibrium beliefs directly as in the first-best benchmark, we follow Shin (2003) and specify out-of-equilibrium beliefs to be that, in the off-path event that the manager discloses a poor outcome in one dimension, the market believes that the manager is fully informed and the remaining dimension (if undisclosed) resulted in a poor outcome as well. (see p. 115 of Shin (2003)). Hence, by specifying the out-of-equilibrium beliefs $\hat{f}_0$ and $\hat{g}_0$ to be zero, the conditions (21) and (22) are always satisfied. The conditions (23) and (24) are equivalent to (11).

Proof of Theorem 1

Suppose that the manager uses a pure strategy $\sigma \in \{0, 1\}$. Then, we have

$$\hat{f}_\emptyset = \hat{f}_1 = \frac{1 - q_f}{1 - q_f \phi_\sigma} \phi_\sigma$$
$$\hat{g}_\emptyset = \hat{g}_1 = \frac{1 - q_g}{1 - q_g p_\sigma} p_\sigma$$

Next, we can find $\overline{\omega}$ by the indifference condition, $E_C[U(s, d(s), 1)] = E_D[U(s, d(s), 1)]$, which is equivalent to

$$\alpha \triangle_\phi \frac{q_f(1 - \phi_C)}{1 - q_f \phi_C} = \omega \beta \triangle_p \frac{q_g(1 - p_C)}{1 - q_g p_C}$$

Solving for ω gives us $\overline{\omega}$. Similarly, we can find $\underline{\omega}$. We have $\underline{\omega} < \overline{\omega}$ since $\frac{1-\phi}{1-q_f\phi}$ is decreasing in ϕ and $\frac{1-q_g p}{1-p}$ is increasing in p.

If $\omega \in (\underline{\omega}, \overline{\omega})$, we need to show there exists a mixed strategy $\sigma \in (0, 1)$ solving (15), which is equivalent to $F(\sigma, \omega) = 0$, where a function $F(\sigma, \omega)$ is defined as

$$F(\sigma, \omega) = (q_g \triangle_p - q_f \triangle_\phi)(\alpha \hat{f}_\emptyset + \omega \beta \hat{g}_\emptyset) + q_f \triangle_\phi(\alpha + \omega \beta \hat{g}_1) - q_g \triangle_p(\alpha \hat{f}_1 + \omega \beta) \tag{25}$$

We need to show that there exists σ satisfying (15). Notice that

$$\begin{aligned} F(0,\omega) &= q_f \triangle_\phi \alpha \frac{1-\phi_D}{1-q_f\phi_D} - q_g \triangle_p \omega\beta \frac{1-p_D}{1-q_g p_D} < 0 \\ F(1,\omega) &= q_f \triangle_\phi \alpha \frac{1-\phi_C}{1-q_f\phi_C} - q_g \triangle_p \omega\beta \frac{1-p_C}{1-q_g p_C} > 0 \end{aligned}$$

and F is continuous in σ. Thus, there exists a $\sigma \in (0,1)$ solving (15). Since market beliefs are monotone in σ and since beliefs at the bounds $\underline{\omega}$ and $\overline{\omega}$ are uniquely pinned down by the pure strategy, continuity of F implies that σ satisfying (15) must be unique.

Finally, to show that σ solving $F(\sigma,\omega) = 0$ satisfies (11) and thus forms the sanitation equilibrium, notice that $F(\sigma,\omega) = 0$ can be re-written as

$$q_f \triangle_\phi(\alpha + \omega\beta\hat{g}_1 - \alpha\hat{f}_\emptyset - \omega\beta\hat{g}_\emptyset) = q_g \triangle_p(\alpha\hat{f}_1 + \omega\beta - \alpha\hat{f}_\emptyset - \omega\beta\hat{g}_\emptyset)$$

Thus, if a mixed strategy σ solving $F(\sigma,\omega) = 0$ does not satisfy (11), it should be the case that the following two inequalities hold simultaneously:

$$\begin{aligned} \omega &< \frac{\alpha}{\beta}\frac{\hat{f}_\emptyset - \hat{f}_1}{1-\hat{g}_\emptyset} \\ \omega &> \frac{\alpha}{\beta}\frac{1-\hat{f}_\emptyset}{\hat{g}_\emptyset - \hat{g}_1} \end{aligned}$$

which contradicts the sanitation condition.

Proof of Proposition 3

It directly follows from (16). For $\omega \in (\underline{\omega}_m, \omega_{FB})$, a mixed strategy $\sigma \in (0,1]$ is optimal, which is clearly higher than the first-best choice $\sigma_{FB} = 0$.

Proof of Corollary 1

It directly follows from (12) and (13). If $\underline{\omega} \geq \omega_{FB}$, for $\omega \in (\omega_{FB}, \overline{\omega})$, a mixed strategy $\sigma \in [0,1)$ is optimal, which is clearly lower than the first-best choice $\sigma_{FB} = 1$. Similarly, if $\overline{\omega} \leq \omega_{FB}$, for $\omega \in (\underline{\omega}, \omega_{FB})$, a mixed strategy $\sigma \in (0,1]$ is optimal, which is clearly higher than the first-best choice $\sigma_{FB} = 0$.

Proof of Theorem 2 and Corollary 2

From Proposition 5, we have $\frac{\partial\sigma}{\partial q_g} \geq 0$, i.e., $\sigma_m \geq \sigma$ for a given ω, which proves Corollary 2. Also, notice that when $\omega < \omega_{FB}$, RE is decreasing in σ:

$$\alpha(\phi_C - \phi_D) + \omega\beta(p_C - p_D) = \omega\beta\triangle_p - \alpha\triangle_\phi < 0$$

Thus, RE under the voluntary regime is always weakly greater than RE in the mandatory regime. Similarly, when $\omega > \omega_{FB}$, RE is increasing in σ and thus we have the opposite result.

Proof of Proposition 4

We first derive the comparative statics of the upper boundary:

$$\begin{aligned}
\frac{\partial\overline{\omega}}{\partial q_g} &= -\omega_{FB}\frac{q_f(1-\phi_C)}{1-q_f\phi_C}\frac{1}{q_g^2(1-p_C)} < 0 \\
\frac{\partial\overline{\omega}}{\partial q_f} &= \omega_{FB}\frac{1-\phi_C}{(1-q_f\phi_C)^2}\frac{1-q_g p_C}{q_g(1-p_C)} > 0 \\
\frac{\partial\overline{\omega}}{\partial(\alpha/\beta)} &= \frac{\triangle_\phi}{\triangle_p}\frac{q_f(1-\phi_C)}{1-q_f\phi_C}\frac{1-q_g p_C}{q_g(1-p_C)} > 0
\end{aligned}$$

Similarly, we can find that $\frac{\partial\underline{\omega}}{\partial q_g} < 0$, $\frac{\partial\underline{\omega}}{\partial q_f} > 0$, and $\frac{\partial\underline{\omega}}{\partial(\alpha/\beta)} > 0$. The mandatory regime is a special case with $q_g = 1$.

Define $x \equiv \overline{\omega} - \underline{\omega}$. Then, we have

$$\begin{aligned}
\frac{\partial x}{\partial q_g} &= \frac{\partial\overline{\omega}}{\partial q_g} - \frac{\partial\underline{\omega}}{\partial q_g} \\
&= \frac{\omega_{FB}q_f}{q_g^2}\left\{\frac{1-\phi_D}{1-q_f\phi_D}\frac{1}{1-p_D} - \frac{1-\phi_C}{1-q_f\phi_C}\frac{1}{1-p_C}\right\} < 0 \\
\frac{\partial x}{\partial(\alpha/\beta)} &= \frac{\phi_D q_f}{p_C q_g}\left\{\frac{1-\phi_C}{1-q_f\phi_C}\frac{1-q_g p_C}{1-p_C} - \frac{1-\phi_C}{1-q_f\phi_C}\frac{1-q_g p_C}{1-p_C}\right\} > 0
\end{aligned}$$

since we have $\frac{1-\phi}{1-q_f\phi}$ is decreasing in ϕ, $\phi_C < \phi_D$, and $p_C > p_D$.

Proof of Proposition 5

A mixed strategy σ solves $F(\sigma) = 0$, where F is defined in (25). Thus, any comparative statics can be derived as

$$\frac{\partial\sigma}{\partial\epsilon} = -\frac{F_\epsilon}{F_\sigma}, \tag{26}$$

where $\epsilon \in \{\omega, \alpha, \beta, q_g, q_f\}$ is a variable of interest. Since we know that $F_\sigma > 0$ at σ solving $F(\sigma) = 0$ from Theorem 1, the sign of (26) is equal to $-F_\epsilon$. First, we can see that

$$\begin{aligned}
F_\omega &= (q_g\triangle_p - q_f\triangle_\phi)\beta\hat{g}_\emptyset + q_f\triangle_\phi\beta\hat{g}_1 - q_g\triangle_p\beta \\
&= -q_g\triangle_p\beta(1-\hat{g}_\emptyset) - q_f\triangle_\phi\beta(\hat{g}_\emptyset - \hat{g}_1) \\
&< 0
\end{aligned}$$

Thus, σ is increasing in ω in the mixed region. Second, we have

$$\begin{aligned}
F_\alpha &= q_f\triangle_\phi(1-\hat{f}_\emptyset) + q_g\triangle_p(\hat{f}_\emptyset - \hat{f}_1) > 0 \\
F_\beta &= -\omega\left\{q_g\triangle_p(1-\hat{g}_\emptyset) + q_f\triangle_\phi(\hat{g}_\emptyset - \hat{g}_1)\right\} < 0,
\end{aligned}$$

Thus, σ is decreasing (increasing) in α (β).

Third, to prove $F_{q_g} < 0$ in the mixed region so that σ is increasing in q_g, we fix ω here and assume that

$\omega > \underline{\omega}_m$. Otherwise, the optimal project choice is always the non-renewable one. We have

$$\begin{aligned}
F(1,q_g) &= q_f\triangle_\phi\alpha(1-\hat{f}) - q_g\triangle_p\omega\beta(1-\hat{g}) \\
&= q_f\triangle_\phi\alpha\frac{1-\phi_C}{1-q_f\phi_C} - q_g\triangle_p\omega\beta\frac{1-p_C}{1-q_g p_C} \\
F_{q_g}(1,q_g) &\propto -\frac{1}{(1-q_g p_C)^2} < 0
\end{aligned}$$

Thus, there exists a unique $\overline{q}_g$ satisfying $F(1,\overline{q}_g) = 0$

$$\begin{aligned}
\frac{q_g}{1-q_g p_C} &= \frac{\triangle_\phi\alpha}{\triangle_p\omega\beta}\frac{q_f}{1-p_C}\frac{1-\phi_C}{1-q_f\phi_C} \\
\overline{q}_g &= \left(p_C + \frac{\triangle_p\omega\beta(1-p_C)(1-q_f\phi_C)}{\triangle_\phi\alpha q_f(1-\phi_C)}\right)^{-1}
\end{aligned}$$

For $q_g \geq \overline{q}_g$, $\sigma = 1$ is the optimal choice. Similarly, for $q_g \leq \underline{q}_g$, $\sigma = 0$ is the optimal choice, where $\underline{q}_g$ satisfies $F(0,\underline{q}_g) = 0$:

$$\underline{q}_g = \left(p_D + \frac{\triangle_p\omega\beta(1-p_D)(1-q_f\phi_D)}{\triangle_\phi\alpha q_f(1-\phi_D)}\right)^{-1}$$

We can show that $\underline{q}_g < \overline{q}_g$. Suppose that it's not by contradiction, which implies

$$\begin{aligned}
p_D + \frac{\triangle_p\omega\beta(1-p_D)(1-q_f\phi_D)}{\triangle_\phi\alpha q_f(1-\phi_D)} &\leq p_C + \frac{\triangle_p\omega\beta(1-p_C)(1-q_f\phi_C)}{\triangle_\phi\alpha q_f(1-\phi_C)} \\
\frac{\omega\beta(1-p_D)(1-q_f\phi_D)}{\triangle_\phi\alpha q_f(1-\phi_D)} &\leq 1 + \frac{\omega\beta(1-p_C)(1-q_f\phi_C)}{\triangle_\phi\alpha q_f(1-\phi_C)} \\
\omega\beta(1-p_D)(1-\phi_C)(1-q_f\phi_D) &\leq \triangle_\phi\alpha q_f(1-\phi_D)(1-\phi_C) + \omega\beta(1-p_C)(1-\phi_D)(1-q_f\phi_C) \\
&< \triangle_\phi\alpha q_f(1-\phi_D)(1-\phi_C) + \omega\beta(1-p_C)(1-\phi_C)(1-q_f\phi_D)
\end{aligned}$$

since we have $(1-\phi_D)(1-q_f\phi_C) < (1-\phi_C)(1-q_f\phi_D)$. This implies that

$$\begin{aligned}
\omega\beta\triangle_p(1-q_f\phi_D) &< \triangle_\phi\alpha q_f(1-\phi_D) \\
\omega &< \omega_{FB}\frac{q_f(1-\phi_D)}{1-q_f\phi_D} = \underline{\omega}_m
\end{aligned}$$

which contradicts the assumption of $\omega > \underline{\omega}_m$.

Next, notice that

$$F(\sigma,\underline{q}_g) > F(0,\underline{q}_g) = 0$$

since F is increasing in σ. Similarly, we have

$$F(\sigma,\overline{q}_g) < F(1,\overline{q}_g) = 0$$

Thus, there should be at least one $q_g \in (\underline{q}_g, \overline{q}_g)$ satisfying $F(\sigma,q_g) = 0$. We claim that there is a unique q_g, which implies that F_{q_g} should be non-positive at q_g solving $F(\sigma,q_g) = 0$. To show this, consider $q_g^{'} = \inf\{q_g : F(\sigma,q_g) = 0\}$. It must be the case that $F_{q_g}(\sigma,q_g^{'}) \leq 0$. Take the first derivative of F with respect to q_g:

$$\begin{aligned} F_{q_g} &= \triangle_p(\alpha\hat{f}_\emptyset+\omega\beta\hat{g}_\emptyset)+(q_g\triangle_p-q_f\triangle_\phi)\frac{\partial}{\partial q_g}(\alpha\hat{f}_\emptyset+\omega\beta\hat{g}_\emptyset) \\ &\quad +q_f\triangle_\phi\omega\beta\frac{\partial\hat{g}_1}{\partial q_g}-\triangle_p(\alpha\hat{f}_1+\omega\beta)-q_g\triangle_p\alpha\frac{\partial\hat{f}_1}{\partial q_g} \end{aligned}$$

where the first derivatives are obtained after some calculations:

$$\begin{aligned} \frac{\partial\hat{f}_\emptyset}{\partial q_g} &= \frac{(1-q_f)\sigma(1-\sigma)\triangle_p\triangle_\phi}{(1-q_f\phi_\sigma-q_gp_\sigma+q_fq_g\phi_Cp_C)^2}>0 \\ \frac{\partial\hat{f}_1}{\partial q_g} &= 0 \\ \frac{\partial\hat{g}_\emptyset}{\partial q_g} &= -\frac{(p_\sigma-q_f\phi_Cp_C)(1-q_f\phi_\sigma-p_\sigma+q_f\phi_Cp_C)}{(1-q_f\phi_\sigma-q_gp_\sigma+q_fq_g\phi_Cp_C)^2}<0 \\ \frac{\partial\hat{g}_1}{\partial q_g} &= -\frac{\phi_Cp_C(\phi_\sigma-\phi_Cp_C)}{(\phi_\sigma-q_g\phi_Cp_C)^2}<0, \end{aligned}$$

since we know that $(1-p_\emptyset)(1-q_f\phi_\sigma)=1-q_f\phi_\sigma-p_\sigma+q_f\phi_Cp_C>0$. Take the second derivative of F with respect to q_g:

$$F_{q_gq_g}=2\triangle_p(\alpha\frac{\partial\hat{f}_\emptyset}{\partial q_g}+\omega\beta\frac{\partial\hat{g}_\emptyset}{\partial q_g})+(q_g\triangle_p-q_f\triangle_\phi)(\alpha\frac{\partial^2\hat{f}_\emptyset}{\partial q_g^2}+\omega\beta\frac{\partial^2\hat{g}_\emptyset}{\partial q_g^2})+q_f\triangle_\phi\omega\beta\frac{\partial^2\hat{g}_1}{\partial q_g^2}$$

where the second derivatives are given by

$$\begin{aligned} \frac{\partial^2\hat{f}_\emptyset}{\partial q_g^2} &= 2\frac{\partial\hat{f}_\emptyset}{\partial q_g}\frac{p_\sigma-q_f\phi_Cp_C}{1-q_f\phi_\sigma-q_gp_\sigma+q_fq_g\phi_Cp_C}>0 \\ \frac{\partial^2\hat{g}_\emptyset}{\partial q_g^2} &= 2\frac{\partial\hat{g}_\emptyset}{\partial q_g}\frac{p_\sigma-q_f\phi_Cp_C}{1-q_f\phi_\sigma-q_gp_\sigma+q_fq_g\phi_Cp_C}<0 \\ \frac{\partial^2\hat{g}_1}{\partial q_g^2} &= 2\frac{\partial\hat{g}_1}{\partial q_g}\frac{\phi_Cp_C}{\phi_\sigma-q_g\phi_Cp_C}<0 \end{aligned}$$

Substituting the second derivatives into $F_{q_gq_g}$ yields

$$F_{q_gq_g}=2\left(\alpha\frac{\partial\hat{f}_\emptyset}{\partial q_g}+\omega\beta\frac{\partial\hat{g}_\emptyset}{\partial q_g}\right)\left(\frac{\triangle_p-q_f\phi_\sigma\triangle_p-q_f\triangle_\phi(p_\sigma-q_f\phi_Cp_C)}{1-q_f\phi_\sigma-q_gp_\sigma+q_fq_g\phi_Cp_C}\right)+2q_f\triangle_\phi\omega\beta\frac{\partial\hat{g}_1}{\partial q_g}\frac{\phi_Cp_C}{\phi_\sigma-q_g\phi_Cp_C}$$

The numerator of the coefficient on $\alpha\frac{\partial\hat{f}_\emptyset}{\partial q_g}+\omega\beta\frac{\partial\hat{g}_\emptyset}{\partial q_g}$ is independent of σ and when $\sigma=1$, and we have

$$\begin{aligned} \triangle_p(1-q_f\phi_C)-q_f\triangle_\phi p_C(1-q_f\phi_C) &= (1-q_f\phi_C)(\triangle_p-q_f\triangle_\phi p_C) \\ &= \triangle_p(1-q_f\phi_C)(1-q_f\frac{\triangle_\phi}{\triangle_p}p_C) \\ &= \triangle_p(1-q_f\phi_C)(1-q_f\phi_D)>0 \end{aligned}$$

Thus, the coefficient is positive. We can see that if $\alpha\frac{\partial\hat{f}_\emptyset}{\partial q_g}+\omega\beta\frac{\partial\hat{g}_\emptyset}{\partial q_g}\leq 0$, the second derivative is also negative. We claim that this is the case at $q_g=q_g^{'}$ so that $F(\sigma,q_g)$ is concave in q_g and there is a unique

q_g solving $F(\sigma, q_g) = 0$. Suppose that $\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_g} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_g} > 0$ at $q_g = q_g^{'}$ by contradiction. Then, there might be another $q_g^{''} > q_g^{'}$ solving $F(\sigma, q_g^{''}) = 0$ with $F_{q_g}(\sigma, q_g^{''}) \geq 0$. Under the sanitation equilibrium, we have $\alpha\hat{f}_1 + \omega\beta \geq \alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset$. Thus, the fact that $F_{q_g}(\sigma, q_g^{''}) \geq 0$ implies that

$$\begin{aligned}
(q_g\triangle_p - q_f\triangle_\phi)\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_g} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_g}\right) + q_f\triangle_\phi\omega\beta\frac{\partial \hat{g}_1}{\partial q_g} &> 0 \\
(q_g\triangle_p - q_f\triangle_\phi)\frac{\partial}{\partial q_g}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) &> 0
\end{aligned}$$

at $q_g = q_g^{''}$. It must be the case that $q_g\triangle_p > q_f\triangle_\phi$ and $\frac{\partial}{\partial q_g}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) > 0$ since $\frac{\partial}{\partial q_g}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) > 0$ at $q_g = q_g^{'}$ and it is increasing in q_g:

$$\frac{\partial^2}{\partial q_g^2}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) = \frac{2(p_\sigma - q_f\phi_C p_C)}{1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C}\frac{\partial}{\partial q_g}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) > 0$$

We can re-express $F_{q_g q_g}$ as

$$\begin{aligned}
F_{q_g q_g} \quad &\propto \quad \left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_g} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_g}\right)\frac{\triangle_p - q_f\phi_\sigma\triangle_p - q_f\triangle_\phi(p_\sigma - q_f\phi_C p_C)}{1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C} \\
&\quad + q_f\triangle_\phi\omega\beta\frac{\partial \hat{g}_1}{\partial q_g}\frac{\phi_C p_C}{\phi_\sigma - q_g\phi_C p_C} \\
&> \quad \left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_g} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_g}\right)\left\{\triangle_p(1 - q_f\phi_\sigma) - q_f\triangle_\phi(p_\sigma - q_f\phi_C p_C)\right\} \\
&\quad + q_f\triangle_\phi\omega\beta\frac{\partial \hat{g}_1}{\partial q_g}(p_\sigma - q_f\phi_C p_C) \\
&= \quad \left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_g} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_g}\right)\triangle_p(1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C) \\
&\quad + (p_\sigma - q_f\phi_C p_C)\left\{(q_g\triangle_p - q_f\triangle_\phi)\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_g} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_g}\right) + q_f\triangle_\phi\omega\beta\frac{\partial \hat{g}_1}{\partial q_g}\right\}
\end{aligned}$$

since $\hat{g}_\emptyset \geq \hat{g}_1$, which implies

$$\frac{p_\sigma - q_f\phi_C p_C}{1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C} \geq \frac{\phi_C p_C}{\phi_\sigma - q_g\phi_C p_C}$$

Thus, the second derivative at $q_g = q_g^{''}$ is positive so that F_{q_g} continues to be positive. This implies that $F(\sigma, q_g) \geq 0$ as $q_g \to \overline{q}_g$, which contradicts $F(\sigma, \overline{q}_g) < 0$.

We can use the same technique to prove $F_{q_f} > 0$ so that σ is decreasing in q_f. We fix ω here and assume that $\omega < \overline{\omega}_f \equiv \frac{\alpha\triangle_\phi}{\beta\triangle_p}\frac{1 - q_g p_C}{q_g(1 - p_C)}$. Otherwise, the optimal project choice is always the sustainable one. We can show that there exists a range of $(\underline{q}_f, \overline{q}_f)$ in which a mixed strategy is optimal:

$$\begin{aligned}\underline{q}_f &= \left(\phi_C + \frac{\triangle_\phi \alpha(1-\phi_C)(1-q_g p_C)}{\triangle_p \omega\beta q_g(1-p_C)}\right)^{-1} \\ \overline{q}_f &= \left(\phi_D + \frac{\triangle_\phi \alpha(1-\phi_D)(1-q_g p_D)}{\triangle_p \omega\beta q_g(1-p_D)}\right)^{-1}\end{aligned}$$

For $q_f \le \underline{q}_f$, $\sigma = 1$ is the optimal choice, and for $q_f \ge \overline{q}_f$, $\sigma = 0$ is the optimal choice.

It is easy to find that for $\sigma \in (0,1)$, we have

$$\begin{aligned}F(\sigma, \underline{q}_f) &< F(1, \underline{q}_f) = 0 \\ F(\sigma, \overline{q}_f) &< F(0, \overline{q}_f) = 0\end{aligned}$$

since F is increasing in σ. Thus, there should be at least one $q_f \in (\underline{q}_f, \overline{q}_f)$ satisfying $F(\sigma, q_f) = 0$. To prove uniqueness, consider $q_f^{'} = \inf\{q_f : F(\sigma, q_f) = 0\}$. It must be the case that $F_{q_f}(\sigma, q_f^{'}) \ge 0$. Take the first derivative of F with respect to q_f:

$$\begin{aligned}F_{q_f} &= -\triangle_\phi(\alpha \hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) + (q_g\triangle_p - q_f\triangle_\phi)\frac{\partial}{\partial q_f}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) \\ &\quad + q_f\triangle_\phi\omega\beta\frac{\partial \hat{g}_1}{\partial q_f} + \triangle_\phi(\alpha + \omega\beta\hat{g}_1) - q_g\triangle_p\alpha\frac{\partial \hat{f}_1}{\partial q_f}\end{aligned}$$

where the first derivatives are given by

$$\begin{aligned}\frac{\partial \hat{g}_\emptyset}{\partial q_f} &= \frac{(1-q_g)\sigma(1-\sigma)\triangle_p\triangle_\phi}{(1-q_f\phi_\sigma - q_g p_\sigma + q_f q_g \phi_C p_C)^2} \\ \frac{\partial \hat{g}_1}{\partial q_f} &= 0 \\ \frac{\partial \hat{f}_\emptyset}{\partial q_f} &= -\frac{(\phi_\sigma - q_g\phi_C p_C)(1 - q_g p_\sigma - \phi_\sigma + q_g\phi_C p_C)}{(1-q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C)^2} \\ \frac{\partial \hat{f}_1}{\partial q_f} &= -\frac{\phi_C p_C(p_\sigma - \phi_C p_C)}{(p_\sigma - q_f\phi_C p_C)^2}\end{aligned}$$

Take the second derivative of F with respect to q_f:

$$F_{q_f q_f} = -2\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f}\right)\left(\frac{\triangle_\phi - q_g p_\sigma\triangle_\phi - q_g\triangle_p(\phi_\sigma - q_g\phi_C p_C)}{1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C}\right) - 2q_g\triangle_p\alpha\frac{\partial \hat{f}_1}{\partial q_f}\frac{\phi_C p_C}{p_\sigma - q_f\phi_C p_C}.$$

Again, we can show that the coefficient on $\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_g} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_g}$ is positive as we did for q_g. Thus, if we have $\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f} \le 0$, the second derivative is positive. We claim that this is the case at $q_f = q_f^{'}$ so that $F(\sigma, q_f)$ is convex in q_f so that there is a unique q_f solving $F(\sigma, q_f) = 0$. Suppose that $\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f} > 0$ at $q_f = q_f^{'}$ by contradiction. Then, there might be another $q_f^{''} > q_f^{'}$ solving $F(\sigma, q_f^{''}) = 0$ with $F_{q_f}(\sigma, q_f^{''}) \le 0$. Under the sanitation equilibrium, we have $\alpha + \omega\beta\hat{g}_1 \ge \alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset$. Thus, the fact that $F_{q_f}(\sigma, q_f^{''}) \le 0$ implies that

$$
\begin{aligned}
(q_g\triangle_p - q_f\triangle_\phi)\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f}\right) - q_g\triangle_p\alpha\frac{\partial \hat{f}_1}{\partial q_f} &< 0 \\
(q_g\triangle_p - q_f\triangle_\phi)\frac{\partial}{\partial q_f}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) &< 0
\end{aligned}
$$

at $q_f = q_f^{''}$. It must be the case that $q_g\triangle_p < q_f\triangle_\phi$ and $\frac{\partial}{\partial q_f}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) > 0$ since $\frac{\partial}{\partial q_f}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) > 0$ at $q_f = q_f^{'}$ and it is increasing in q_f:

$$
\frac{\partial^2}{\partial q_f^2}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) = \frac{2(\phi_\sigma - q_g\phi_C p_C)}{1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C}\frac{\partial}{\partial q_f}(\alpha\hat{f}_\emptyset + \omega\beta\hat{g}_\emptyset) > 0
$$

We can re-express the second derivative as

$$
\begin{aligned}
F_{q_f q_f} \quad &\propto \quad -\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f}\right)\frac{\triangle_\phi - q_g p_\sigma\triangle_\phi - q_g\triangle_p(\phi_\sigma - q_g\phi_C p_C)}{1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C} \\
&\qquad -q_g\triangle_p\alpha\frac{\partial \hat{f}_1}{\partial q_f}\frac{\phi_C p_C}{p_\sigma - q_f\phi_C p_C} \\
&< \quad -\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f}\right)\left\{\triangle_\phi(1 - q_g p_\sigma) - q_g\triangle_p(\phi_\sigma - q_g\phi_C p_C)\right\} \\
&\qquad -q_g\triangle_p\alpha\frac{\partial \hat{f}_1}{\partial q_f}(\phi_\sigma - q_g p_C\phi_C) \\
&= \quad -\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f}\right)\triangle_\phi(1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C) \\
&\qquad +(\phi_\sigma - q_g p_C\phi_C)\left\{(q_g\triangle_p - q_f\triangle_\phi)\left(\alpha\frac{\partial \hat{f}_\emptyset}{\partial q_f} + \omega\beta\frac{\partial \hat{g}_\emptyset}{\partial q_f}\right) - q_g\triangle_p\alpha\frac{\partial \hat{f}_1}{\partial q_f}\right\}
\end{aligned}
$$

since $\hat{f}_\emptyset \geq \hat{f}_1$, which implies

$$
\frac{\phi_\sigma - q_g p_C\phi_C}{1 - q_f\phi_\sigma - q_g p_\sigma + q_f q_g\phi_C p_C} \geq \frac{p_C\phi_C}{p_\sigma - q_f\phi_C p_C}
$$

Thus, the second derivative at $q_g = q_g^{''}$ is negative so that F_{q_f} continues to be negative. This implies that $F(\sigma, q_f) \leq 0$ as $q_f \to \overline{q}_f$, which contradicts $F(\sigma, \overline{q}_f) > 0$.

Proof of Proposition 6

As we show in the proof of Theorem 1, RE is decreasing (increasing) in σ when $\omega < \omega_{FB}$ ($\omega > \omega_{FB}$). Since σ is weakly increasing (decreasing) in q_g (q_f), RE is weakly decreasing (increasing) in q_g and weakly increasing (decreasing) in q_f when $\omega < \omega_{FB}$ ($\omega > \omega_{FB}$).