Executive Compensation with Social and Environmental Performance*

Pierre Chaigneau† Nicolas Sahuguet‡

Queen's University HEC Montréal

June 2024

Abstract

How to incentivize a manager to create value and be socially responsible? A manager can predict how his decisions will affect measures of social performance, and will therefore game an incentive system that relies on these measures. Still, we show that the compensation contract is based on measures of social performance when the level of social investments preferred by the board exceeds the one that maximizes the stock price. In this case, because of gaming, social investments are distorted and the sensitivity of pay to social performance is reduced. Relying on multiple measures based on different methodologies will generally mitigate inefficiencies due to gaming, i.e. harmonization of social performance measurement can backfire.

Keywords: corporate governance, corporate social responsibility, ESG measurement, ESG harmonization, executive compensation.

*For useful comments and suggestions we thank Jeremy Bertomeu, Jean de Bettignies, Milo Bianchi, Jörg Budde (discussant), Alexander Dyck, Alex Edmans, Thomas Geelen (discussant), Christian Hofmann, Nicolas Inostroza, Jan Mahrt-Smith, Lucas Mahieux, Marcus Opp (discussant), Günter Strobl, Katrin Tinn, John Van Reenen, and participants at the Accounting and Economics Society webinar, Accounting Research Workshop, Asia Meeting of the Econometric Society, CSFN Conference on Governance and Sustainability, Erasmus Corporate Governance Conference, the Québec Political Economy Conference, the SFS Cavalcade, and the University of Toronto.

†Corresponding author. Address: Smith School of Business, Goodes Hall, 143 Union Street West, K7L 2P3, Kingston, ON, Canada. Email: pierre.chaigneau@queensu.ca. Tel: 613 533 2312.

‡Address: Applied Economics Department, HEC Montréal, 3000 chemin de la côte Sainte Catherine, H3T 2A7 Montréal, QC, Canada. Email: nicolas.sahuguet@hec.ca. Tel: 514 340 6031.

There has recently been a marked increase in the propensity of firms to use social and environmental measures of performance in executive compensation (Homroy, Mavruk, and Nguyen (2023)). In particular, some executive compensation contracts rely on ESG scores and ratings (Cohen et al. (2023)) which are "third-party assessment[s] of corporations' ESG performance" (Berg, Kölbel, and Rigobon (2022)). A large majority of investors supports the inclusion of such performance metrics in incentives plans. Some institutional investors even focus their engagement on the inclusion of these metrics.[1]

Yet, prominent scholars have argued that metrics of social and environmental performance should not be used for incentive purposes (Edmans (2021), Bebchuk and Tallarita (2022)).[2] They make two main points that cannot be easily dismissed, even in well-governed firms, and which are not captured by standard models of multitasking. First, they argue that executives will tend to "hit the target but miss the point", i.e. they will improve a measure of social or environmental performance even when they are aware that it does not improve actual social or environmental outcomes. We will refer to this behavior as "gaming". Second, they argue that there is disagreement about the measurement of social and environmental performance, as documented in Berg, Kölbel, and Rigobon (2022). Not only is measurement imperfect, but there is no consensus on measurement.

Are explicit incentives based on measures of Social and Environmental Performance (SEP) truly a puzzle or can they be justified even if they lead managers to game the incentive system and performance measurement is heterogeneous? To answer this question, this paper analyzes a principal-agent model of multitasking that takes into account these factors.

A socially responsible board wants to encourage the manager to improve not just financial performance but also SEP. The manager can exert effort to improve firm value and also invest resources to improve some dimensions of the firm's SEP at a cost. SEP is not directly observed, but it is imperfectly measured along several dimensions. We assume that, given his practical

[1]The 2021 Global Benchmark Policy Survey from ISS Governance finds that 86% of investors "believe [that] incorporating non-financial Environmental, Social, and/or Governance-related metrics into executive compensation programs is an appropriate way to incentivize executives." For examples, see the ESG Engagement Campaign from Alliance Bernstein (April 2021), BlackRock's report "Our 2021 Stewardship Expectations: Global Principles and Market-level Voting Guidelines", and "ESG Performance Metrics in Executive Pay" on the Harvard Law School Forum on Corporate Governance (Jan 15 2024) for a recent summary of these practices.

[2]Even before the advent of ESG-based compensation, Shleifer and Vishny (1997) warned about the opportunities for self-dealing associated with incentive pay, and Tirole (2006) conjectured that the stakeholder society would be "best promoted through . . . a fixed wage rather than performance-based incentives."

experience running the firm and his understanding of SEP measures, the manager can anticipate how his decisions will affect its SEP measures. This is conceptually similar to the assumption in Edmans and Gabaix (2011) that the manager observes the noise in the performance measure before choosing his effort. Thus, a manager with SEP-based incentives will tend to invest more (less) in dimensions of SEP that are easy (hard) to improve, even if this leads to minimal (major) social and environmental impact. We refer to the discrepancy between the measured impact and the actual impact of some investments as the performance measure's "bias". For example, reaching the same level of carbon intensity can imply a substantial environmental impact for some firms but not for others. Since the manager knows the bias at the time of making investment decisions, SEP-based incentives will lead him to game the incentive system.

The manager's effort and investment decisions will affect the firm's stock price. Since some investors are socially responsible, we allow for a positive sensitivity of the stock price to the firm's estimated social output (with a slope which can be microfounded as investors' preference for social output). Signals of SEP are therefore already incorporated in the stock price. If the board prefers the level of social investments that maximizes the stock price, then the compensation contract only relies on the stock price.[3] Otherwise, there are two cases.

First, if the level of social investment that the board prefers is lower than the one that maximizes the stock price, then the manager's compensation should be positively related to the stock price and to the firm's profits. Intuitively, the level of social investments that maximizes the stock price exceeds the level of social investments that the board prefers, and profits-based compensation discourages such costly investment. In this case, the board can still induce the first-best level of investment – as defined as the hypothetical outcome in the absence of an agency problem.

Second, if the level of social investment that the board prefers exceeds the one that maximizes the stock price, then the manager's compensation should be positively related to the stock price and to SEP measures. These measures are used to supplement incentives for social investments already embedded in stock price-based compensation. To mitigate inefficiencies due to gaming, the sensitivity of compensation to SEP measures is decreasing in the variance of their bias.[4] For

[3]In our model, the value of additional performance metrics is not due to risk sharing or rent extraction since the manager is risk neutral and can be kept at his reservation level of utility.

[4]The low quality of these measures is not a problem per se. In Appendix B, we study a setting in which the manager does not know the measures' biases at the time of making investment decisions. We find that the first-best outcome can then always be obtained. Moreover, in this case, the sensitivity of compensation to SEP measures

these two reasons, the sensitivity of managerial compensation to SEP measures may be quite low. This is consistent with the empirical evidence that executive compensation is still "overwhelmingly based on shareholder value" in spite of the rise in SEP-based compensation (Rajan, Ramella, and Zingales (2023)).

The stock price is useful for incentive purposes because it efficiently aggregates all available information – although the weights assigned to SEP output and profits may not correspond to the board's preferences. By contrast, explicit incentives based on SEP measures are only based on the signal provided by these contractible measures. As a result, in the aforementioned second case, the board cannot offer a contract that induces the first-best level of social investments.[5] Because of less efficient incentive provision, the average level of social investment is then lower than at the first-best, in contrast to the first case discussed above.[6] This asymmetric effect is due to gaming; it would not exist under the standard assumption that performance is imperfectly measured ex-post (see Appendix B).

Next, we analyze the outcome when the board and stock market investors don't weigh all SEP measures equally. For example, the board might care more about the firm's carbon emissions but less about working conditions than stock market investors. In this case, the manager's contract can be more complex than in the baseline model. Indeed, it can simultaneously include stock price-based compensation, profits-based compensation to discourage excessive investments in working conditions (from the board's perspective), and compensation contingent on the firm's carbon intensity to encourage related investments (carbon capture, green technologies, etc.) above the level that would maximize the stock price. This will be the case when the board cares slightly more about carbon emissions than investors. By contrast, a board that cares much more about carbon emissions will be very concerned with providing efficient investment incentives on this dimension, even at the cost of not sufficiently discouraging excessive investment on other dimensions. Thus, it will not use profits-based compensation, even though its negative effect on carbon emissions could

does not depend on their quality, and it is higher than in the baseline model that features "gaming".

[5]The crucial assumption for this result is that these investments are non-contractible. The result is robust in the sense that it would still hold under more general contracts as long as these investments are not fully contractible. The important point is that the stock price aggregates information in a way that cannot be replicated by a contract. This is consistent with the notion that the stock price provides incremental information that is useful for contracting (Holmström and Tirole (1993)).

[6]Depending on the (random) bias in a SEP measure, the *realized* level of social investment can be higher or lower than the first-best level.

in principle be offset by increasing the sensitivity of pay to measures of carbon intensity. The reason is that a high sensitivity of pay to carbon intensity would substantially encourage gaming and inefficient investment decisions on this dimension, which is especially concerning for a board that cares a lot about carbon emissions. This emphasizes that gaming has nontrivial implications for the combination of performance measures in a compensation contract.

Finally, the model allows us to analyze the outcome with multiple sets of SEP measures, for example multiple ESG scores provided by multiple ESG raters. Consider two sets of SEP measures with a different quality (as measured by the variance of their "bias") and whose bias can be correlated. Even though the availability of an additional set of SEP measures can reduce investment distortions due to gaming, it can alternatively worsen them if the quality of additional SEP measures is sufficiently low. This is in contrast to models of contracting in which the value of a new performance measure is always nonnegative (e.g. Holmström (1979)). Intuitively, the stock price incorporates additional SEP measures to the extent that they are informative, regardless of the potentially detrimental incentive effects. Since they affect the stock price, which is useful for incentive provision, these measures cannot simply be "ignored" by the board. Next, when the biases in ESG scores are independent and identically distributed, we show that increasing the number of scores on a dimension of SEP always reduces investment distortions due to gaming on this dimension, and that the distortionary effect of SEP measures vanishes in the limit as the number of scores becomes very large. Intuitively, if scores are constructed differently, it is harder for a manager to game multiple scoring methodologies than to game a single methodology.

These results have normative implications for the heterogeneity of ESG scores and ratings, which is often criticized on the basis that it reflects disagreement between ESG raters. Most notably, our agency model highlights a beneficial aspect of the low correlation between ESG ratings documented by Chatterji et al. (2016), Berg, Kölbel, and Rigobon (2022), and Christensen, Serafeim, and Sikochi (2022). Indeed, while the bias of ESG scores is detrimental in the model, a low correlation between biases across ESG scores is beneficial. This has implications for the ongoing debate on the regulation and harmonization of ESG ratings, and more generally for the measurement of corporate social and environmental performance.[7]

[7]See: Regulatory Solutions: A Global Crackdown on ESG Greenwash, Harvard Law School Forum on Corporate Governance, June 23 2022; EU watchdog says ESG rating firms need rules to stop 'greenwashing', *Reuters* February 12 2020, where Steven Maijoor, chair of the European Securities and Markets Authority (ESMA) is quoted as saying that "ESG rating agencies should be regulated and supervised appropriately by public sector authorities."

Related literature

The multitasking literature analyzes incentive provision when there are more tasks than performance measures, as well as the importance of maintaining balanced incentives across tasks. Our model differs from existing multitasking models because we let the manager be privately informed about the effect of his actions on several performance measures. Indeed, we assume that the manager observes the "noise" before taking his action: he can predict the future realizations of performance measures, as opposed to their expected future realizations in other models. This introduces new possibilities for the manager to game a performance-based contract.

In Edmans and Gabaix (2011), there is only an effort problem. Thus, this timing assumption has implications for the allocation of incentives across states of the world, as determined by the curvature of the optimal contract. Intuitively, the agent must be subject to a constant incentive pressure regardless of the level of noise. In our paper, this timing assumption is instead relevant for resource allocation. It makes performance measures biased at the time when the agent makes investment decisions, and it results in lower-powered incentives.

Contracting based on ratings is studied in other papers. Rajan and Parlour (2020) show that including credit ratings in a contract may be optimal even when they are uninformative because they are contractible. By contrast, ESG ratings are informative in our model, but the stock price is useful for contracting because it is contractible. Hörner and Lambert (2021) study how to optimally design a rating for incentive purposes, whereas we take ESG scores and ratings as given.

Our paper takes an optimal contracting perspective to the provision of incentives for corporate investment in social goods. Baron (2008) analyzes the balance between "profit incentives" and "social incentives". In his model, there is no uncertainty about the firm's CSR technology, and no stock price-based incentives. Bonham and Riggs-Cragun (2022) take a broader perspective, and study the use of contracts, taxes, and disclosure regulation to encourage ESG activities which are imperfectly measured. Likewise, there is no stock price in their model. Bucourt and Inostroza (2023) study stock price-based incentives for ESG when investors have heterogeneous preferences for ESG. In their model, there is no agency problem and no contracting: the manager chooses her ESG effort to maximize shareholder value.

1 The model

1.1 Technology

Consider a firm run by a manager and controlled by a socially responsible board on behalf of shareholders. The production technology involves both a value-increasing action ("effort") and social investment decisions. At $t = 0$, a risk neutral manager chooses unobservable effort $e \in \{\underline{e}, \overline{e}\}$ at private cost $C(e)$, with $C(\underline{e}) = 0$ and $C(\overline{e}) = c_e > 0$. He also makes two observable social investment decisions, y_1 and y_2, that improve social and environmental outcomes but decrease the firm's profits. Cash flows or "profits" $\tilde{x}$ and overall "social output" $\tilde{y}$ are respectively defined as:

$$\begin{aligned} \tilde{x} &= e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x \quad \text{where } \tilde{\epsilon}_x \sim \mathcal{N}(0, \sigma_x^2) \\ \tilde{y} &= \eta_1 y_1 + \eta_2 y_2 + \tilde{\epsilon}_y \qquad \text{where } \tilde{\epsilon}_y \sim \mathcal{N}(0, \sigma_y^2) \end{aligned} \tag{1}$$

where θ_1 and θ_2 are positive constants. Cash flows are realized at $t = 2$ and paid out to shareholders. The effect of social investments on social output depends on $\{\eta_1, \eta_2\}$, which are the unobserved realizations of the random variables $\tilde{\eta}_1 \sim \mathcal{N}(\bar{\eta}, \sigma_\eta^2)$ and $\tilde{\eta}_2 \sim \mathcal{N}(\bar{\eta}, \sigma_\eta^2)$. They represent the firm's multidimensional "social productivity".[8] All random variables, $\tilde{\eta}_1$, $\tilde{\eta}_2$, $\tilde{\epsilon}_x$, $\tilde{\epsilon}_y$, are independent.

The board's objective function is:

$$\mathbb{E}\left[\tilde{x} + \alpha_B \tilde{y} - \tilde{W}\right], \tag{2}$$

where the social output $\tilde{y}$ of the firm is weighted by $\alpha_B \geq 0$, and W is the manager's contractual payment. In Appendix C, we give three possible microfoundations for this specification of the board's preferences. The notion that shareholders partly internalize the externalities generated by the firm is consistent with the empirical findings of Homroy, Mavruk, and Nguyen (2023) on the use of SEP measures in executive pay.

[8] The productivity of some social investments can be negative. This means that allocating resources to increase provision of this type of social goods in this firm would decrease its social output. This is for tractability but can sometimes be justified. For example, if gender parity is an objective and the firm currently employs more men ($\eta_1 > 0$), it means that hiring more women ($y_1 > 0$) will help achieve this objective; however, if the firm currently employs more women ($\eta_1 < 0$), it means that hiring more men ($y_1 < 0$) will help achieve this objective.

1.2 Measures of social performance and stock price

Social performance measures are realized at $t = 1$, and they imperfectly reflect the actual SEP of the firm. The measure of the firm's SEP on dimension i (for $i \in \{1, 2\}$), that we will refer to as an "ESG score" (see Appendix A) for brevity but without loss of generality, is:

$$m_i \equiv \varepsilon_i y_i \qquad \text{where } \tilde{\varepsilon}_i \sim \mathcal{N}(\eta_i, \sigma_\varepsilon^2), \tag{3}$$

where η_i is the realization of $\tilde{\eta}_i$. That is, with $\sigma_\varepsilon^2 > 0$, ε_i is a noisy measure of the firm's social productivity on dimension i. The difference between ε_i and η_i is simply the difference between the measured social impact and the actual social impact of social investments.

At $t = 0$, after contracting but before the social investment decision, the manager observes the nonverifiable variable ε_i. This timing assumption is similar to the assumption in Edmans and Gabaix (2011) that the agent chooses his action after observing the noise. In our model, it parsimoniously captures the notion that, because of his on-the-job expertise and his understanding of the SEP measures' methodology, the manager understands how investment decisions will affect SEP measures. It allows to take into account a frequent criticism leveled at SEP-based incentives (that they will encourage "gaming"). It represents a departure from standard models of contracting and multitasking with noisy ex-post performance measurement in which the principal is aware of any biases that the manager may have at the contracting phase.

A publicly observable financial report z, which is imperfectly informative about the firm's profitability, is realized at $t = 1$ such that: $\tilde{z} = e - \theta_1 y_1^2 - \theta_2 y_2^2 + \epsilon_x + \tilde{\epsilon}_z$ where $\tilde{\epsilon}_z \sim \mathcal{N}(0, \sigma_z^2)$. To account for the potential impact of corporate social performance on the stock price, we assume an exogenous stock price function that puts a weight α_I on social performance relative to profits:

$$p = \mathbb{E}\left[\tilde{x} + \alpha_I \tilde{y} \,|\, t = 1\right] \tag{4}$$

where the conditional expectation is taken with respect to the information available at $t = 1$, including social investments, the report z, and ESG scores. In Appendix E, we provide a micro-foundation based on portfolio choice by socially responsible investors.

1.3 Contracting

At $t = 0$, the board offers a compensation contract to a risk neutral manager who has an outside option worth $\bar{W} \geq 0$. As is standard, the manager is self-interested. He receives a fixed wage w, and his compensation is linear in the following performance measures: firm profits (with sensitivity β_x), stock price (with sensitivity β_p), and ESG score i (with sensitivity β_i) for $i = 1, 2$, similar to Holmström and Milgrom (1991) and Holmström and Tirole (1993).

We assume that the sensitivity of pay to profits and ESG scores must be nonnegative, i.e. $\beta_x \geq 0$ and $\beta_i \geq 0$ for $i = 1, 2$.[9] The former ($\beta_x \geq 0$) can be motivated similarly to Innes (1990): the manager would destroy output otherwise. The latter ($\beta_i \geq 0$) can be motivated by the public outcry that would likely result if a manager's pay were decreasing in measures of a firm's SEP – similar to the political constraints mentioned by Jensen and Murphy (1990).

We assume that $\bar{e} - c_e - \bar{W} > 0$, i.e. a firm which hires a manager who exerts high effort can be profitable, and that the cost of high effort for the manager, c_e, is sufficiently low that it is optimal to induce high effort in all settings considered. The discount rate is zero.

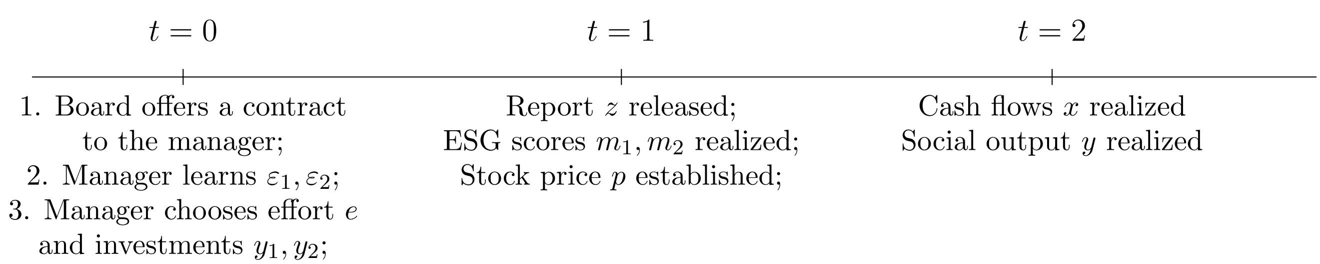

Figure 1: Timeline of the model.

2 Contracting with social performance

2.1 Baseline model

We solve the model by backward induction. Appendix E derives the stock price in several cases of interest. For publicly traded firms, the stock price is established on stock markets. For private

[9]As will be clear in Proposition 1, the sensitivity of compensation to the stock price will be positive ($\beta_p \geq 0$), although it could be negative without the constraint that $\beta_x \geq 0$.

firms, it represents the implied stock valuation established in a sale or a funding round. Lemma 1, proven in Appendix E.3, determines the $t = 1$ stock price in the baseline model.

Lemma 1 *The stock price p in equation (4) is such that:*

$$\mathbb{E}\left[\tilde{x}\,|\,t=1\right] = \hat{e} - \theta_1 y_1^2 + \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2}\left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) \quad (5)$$

$$\mathbb{E}\left[\tilde{y}\,|\,t=1\right] = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} m_1 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} y_1 \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} m_2 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} y_2 \bar{\eta} \quad (6)$$

The stock price is additive in two terms. The first is expected profits. It depends on the manager's expected effort $\hat{e}$, social investments y_1 and y_2, and the report z. The report z is informative about profits $\tilde{x}$. As a result, the report affects the stock price, which makes the stock price sensitive to the manager's actual effort (as opposed to the manager's anticipated effort $\hat{e}$, although the two coincide in equilibrium), so that the stock price provides effort incentives.

The second term is the expected social output of the firm. The perceived SEP of the firm as reflected in ESG scores affects the stock price when $\alpha_I > 0$ (see equation 4). This is consistent with the fact that investors rely on these scores and ratings for their investment decisions (Pagano et al. (2018), Berg, Kölbel, and Rigobon (2022)). Lemma 1 shows that, with $\alpha_I > 0$, there is a positive relation between ESG scores and the stock price in equilibrium, consistently with the empirical evidence (Berg et al. (2021)). Intuitively, a higher score is good news about the firm's social productivity. However, ESG scores are noisy measures and social investments are distorted accordingly, so that neither these investments nor these scores reveal the firm's actual social productivity. As a result, the expected social output of the firm ($\mathbb{E}[\tilde{y}\,|\,t = 1]$) does not only rely on its ESG scores and its social investments, but also on the prior belief $\bar{\eta}$ about the productivity of these social investments.

We now consider incentive provision and contracting. The board wants to incentivize effort and social investments. To this end, it can use three types of performance measures: the firm's stock price, its profits, and its ESG scores.

As a first preliminary step, in Appendix E.2, we consider the case without ESG scores or ratings. In this case, the stock price is informative about the levels of observable social investments ($\{y_1, y_2\}$). However, in the absence of measures of the firm's SEP, the stock price is uninformative about the productivity of these investments. As a result, we show that stock price-based compensation cannot be used to induce the manager to invest according to his signals on the productivity of the firm's technology for social output. This hypothetical case underlines the important role played by ESG scores.

As a second preliminary step, we define the first-best outcome, which provides a useful benchmark. The first-best outcome refers to the outcome without information asymmetries and without an agency problem, i.e. when there is no incentive constraint in the optimization problem. At the second-best, let $y_i(\varepsilon_i)$ be the social investment in dimension i optimally chosen by the manager given his contract and his signal ε_i, Let φ denote the conditional density function of $\tilde{\varepsilon}_i$, and ϕ denote the density function of $\tilde{\eta}_i$.

Lemma 2 *The first-best social investment is:*

$$y_i^* = \frac{\alpha_B}{2} \frac{\mathbb{E}[\tilde{\eta}_i|\varepsilon_i]}{\theta_i} \quad \textit{where} \quad \mathbb{E}[\tilde{\eta}_i|\varepsilon_i] = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta} \tag{7}$$

At the second-best, the board sets the manager's contract to induce effort and minimize the agency cost in equation (8):

$$\sum_{i=1,2} \theta_i \int_{\eta_i} \int_{\varepsilon_i} \left(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)\right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \tag{8}$$

We refer to the difference between the board's objective function at the first-best and at the second-best as the "agency cost". Given that the manager is risk neutral and there are no constraints on contracting, the agency cost is not driven by inefficient risk sharing or rent extraction.[10] Instead, the agency cost reflects the extent of the inefficiency in resource allocation. As can be seen in equation (8), it measures the deviation between social investments and their first-best levels, which are defined in equation (7).

Lemma 2 shows that minimizing the agency cost is equivalent to minimizing the sum across SEP dimensions of the expected quadratic distance between incentive-compatible social investments $y_i(\varepsilon_i)$ and first-best social investments $y_i^*(\varepsilon_i)$ multiplied by the cost parameter θ_i. This distance is a measure of the agency cost on dimension i. It measures how close the board can get to the first-best outcome described in equation (7). The agency cost is proportional to the monetary

[10]Risk sharing and rent extraction effects are arguably of second-order importance in large firms, where managerial equity holdings account for only 0.34% of firm equity for the median CEO (Edmans, Gabaix, and Jenter (2017)).

cost of social investments, as measured by θ_i. Intuitively, if social investments were costless, there would be no tradeoff between social investments and profits. On the contrary, the more costly social investments are, the more expensive is any deviation from the first-best level.

The contracting problem is a priori not simple for several reasons. First, the firm must provide incentives for effort as well as for social investment. In doing so, it faces a multitasking problem in which the sensitivity of pay to performance must be high enough (to elicit effort) and the balance of incentives matters (because of the resource allocation decisions). Second, because of the manager's knowledge of the firm's technology for social output and his understanding of the ESG ratings' methodologies, ESG ratings-based incentives will result in "gaming". Third, the levels of social investments that maximize the stock price do not necessarily maximize the board's objective function. Fourth, the sensitivity of the manager's compensation to profits and ESG ratings must be nonnegative.

As the third and last preliminary step, Lemma 3 highlights the important role played by the constraint that the sensitivity of pay to profits be non-negative.

Lemma 3 *Without a nonnegativity constraint on β_x, the board can achieve the first-best outcome ($y_i(\varepsilon_i) = y_i^*(\varepsilon_i)\ \forall \varepsilon_i$ and $e = \overline{e}$) by offering a contract such that:*

$$\beta_i = 0, \quad \beta_x = \left(\frac{\alpha_I}{\alpha_B} - 1\right)\beta_p \quad \text{and} \quad \beta_p = \frac{c_e}{\overline{e} - \underline{e}}\left(\frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} - 1\right)^{-1} \tag{9}$$

If the board's preferred level of social investments is lower than the one that maximizes the stock price ($\alpha_B < \alpha_I$), then only relying on stock price-based incentives would lead to excessive spending on social investments from the board's perspective. To counterbalance stock price-based incentives, the board then provides profits-based incentives ($\beta_x > 0$), which discourages costly social investments. On the contrary, if the board's preferred level of social investments is higher than the one that maximizes the stock price ($\alpha_B > \alpha_I$), then only relying on stock price-based incentives would lead to inadequate spending on social investments from the board's perspective. Encouraging social investments further can then be achieved by punishing the manager for achieving high profits, i.e. $\beta_x < 0$. The sensitivity of compensation to the stock price increases as needed to still provide effort incentives. In any case, the agency cost is zero: the firm's investment in dimension i of SEP, $y_i(\varepsilon_i)$, is state-by-state equal to the first-best level $y_i^*(\varepsilon_i)$ defined in equation

(7).

Proposition 1 describes the use of SEP-based compensation in the optimal linear contract with a nonnegativity constraint on β_x.

Proposition 1

- *If $\alpha_B \leq \alpha_I$, then $y_i(\varepsilon_i) = y_i^*(\varepsilon_i)\ \forall \varepsilon_i$, and the optimal linear contract is defined by:*

$$\beta_i = 0, \quad \beta_x = \left(\frac{\alpha_I}{\alpha_B} - 1\right)\frac{c_e}{\bar{e} - \underline{e}}\left(\frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} - 1\right)^{-1}, \quad \beta_p = \frac{c_e}{\bar{e} - \underline{e}}\left(\frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} - 1\right)^{-1}.$$

- *If $\alpha_B > \alpha_I$, then generically $y_i(\varepsilon_i) \neq y_i^*(\varepsilon_i)$, the expected social investment is below the first-best level:*

$$\mathbb{E}[y_i(\tilde{\varepsilon}_i)] - \mathbb{E}[y_i^*(\tilde{\varepsilon}_i)] \quad = \quad \underbrace{(\alpha_B - \alpha_I)}_{>0}\frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2 + \sigma_\eta^2 + \bar{\eta}^2} - 1\right)\bar{\eta},$$

and the optimal linear contract is defined by:

$$\beta_i = (\alpha_B - \alpha_I)\frac{c_e}{\bar{e} - \underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2 + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\right), \quad \beta_x = 0, \quad \beta_p = \frac{c_e}{\bar{e} - \underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right).$$

The structure of the optimal linear contract can be described as follows. First, the fixed wage adjusts to keep the manager at his reservation level of utility including compensation for the cost of effort. As a result, the manager does not derive rents. Second, the sensitivity of pay to the performance measures available has implications for the effort decision and for social investments. A positive sensitivity of pay to profits encourages effort and discourages social investments. A positive sensitivity of pay to ESG scores tends to encourage social investment.[11] A positive sensitivity of pay to the stock price encourages the manager to exert effort and to choose social investments that maximize the stock price.

To start, consider the optimal linear contract when the board (which represents the firm's shareholders) prefers the social investments that maximize the stock price, i.e. $\alpha_B = \alpha_I$. In this case, the manager's compensation is only contingent on the stock price: $\beta_i = 0$, $\beta_x = 0$,

[11] Because of the imperfection of ESG scores, they do not *always* encourage social investment. In the case when the sign of ε_i is opposite the sign of η_i, imperfect measurement that can be anticipated ex-ante gives rise to counterproductive incentives.

and $\beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(1+\frac{\sigma_z^2}{\sigma_x^2}\right)$. When managerial compensation is only sensitive to the stock price, a self-interested manager optimally chooses the social investments that correspond to the board's preference. The sensitivity of compensation to the stock price is then determined to provide adequate incentives for effort.

Now consider the case when the levels of social investments that the board prefers are lower than the ones that maximize the stock price ($\alpha_B < \alpha_I$). For example, as further discussed in Appendix E, this could be the case if investors enjoy sufficiently strong warm-glow utility from investing in socially responsible firms, or if there are many socially responsible funds with impact mandates (Oehmke and Opp (2024)). In this case, the manager's compensation is contingent on profits and the stock price. A positive sensitivity of compensation to profits discourages social investment relative to the level that would maximize the stock price. Since both profits-based and stock price-based compensation encourage managerial effort, a positive sensitivity of compensation to profits reduces the sensitivity of compensation to the stock price required to elicit managerial effort. In this case, the board can still induce a level of social investment y_i that corresponds to the first-best level y_i^* state-by-state (for any realization of $\tilde{\varepsilon}_i$).

Finally, consider the case when the levels of social investments that the board prefers are higher than the ones that maximize the stock price ($\alpha_B > \alpha_I$). As further discussed in Appendix C, this can arise for several reasons. In particular, this is the relevant case if investors do not intrinsically value holding equity in a socially responsible firm, so that SEP does not have a stock price impact ($\alpha_I = 0$), even though they would be better off if the firm reduced its externalities. Then a board with $\alpha_B > 0$ solves a coordination problem. More generally, whenever $\alpha_B > \alpha_I$, the manager's compensation is contingent on the stock price and ESG scores. A positive sensitivity of compensation to ESG scores encourages social investment relative to the level that would maximize the stock price. Ideally, the board would like investment in dimension i of SEP to depend on the average productivity $\bar{\eta}$ of social investments, and on the signal ε_i that the manager receives on the productivity of the firm's social investment on dimension i.

The stock market's valuation of the firm's social investments combines these two aspects (see Lemma 1). Moreover, the stock price aggregates information about the firm's social output efficiently for investment purposes, since it reflects the effect of ESG scores on beliefs about the productivity of the firm's social and environmental investments. However, the stock price's aggregation of information about the firm's social output and its profits does not correspond to the board's preference when $\alpha_B \neq \alpha_I$. When $\alpha_B > \alpha_I$, stock price-based incentives are excessively tilted toward profits maximization relative to the board's preference. This cannot be remedied by a negative sensitivity of managerial compensation to profits: the nonnegativity constraint on profits binds.

This can be partly remedied by also making managerial compensation contingent on the firm's ESG scores, so as to complement the social investment incentives already embedded in stock price-based compensation. However, by definition, ESG score m_i only depends on social investment y_i and on the signal ε_i. The latter is an imperfect signal of the firm's technology for social output (since $\sigma_\varepsilon > 0$), and it is known ex-ante by the manager. Thus, relating managerial compensation to ESG scores will lead the manager to be excessively responsive to realizations of the signal ε_i, which can be viewed as "gaming". This inefficiency reduces the sensitivity of the manager's compensation to ESG scores: this sensitivity is decreasing in the noisiness of scores, as parameterized by σ_ε. In summary, the more noisy ESG scores are, the more distorted are the manager's incentives for social investments, and the less the board encourages social investments.

ESG scores are noisy because they are imperfect indicators of social output. This imperfection combined with their predictability leads to distorted incentives whenever the scores are used for incentive purposes. For example, when ε_i is high, the manager understands that increasing y_i will have a large impact on ESG score m_i, likely over and beyond the "true impact" which is given by $\eta_i y_i$. On the contrary, when ε_i is low, the manager understands that even though increasing y_i might substantially increase social output $\eta_i y_i$, it will only have a small and possibly even negative impact on ESG score m_i.

When the board is highly socially responsible, as in the second case of Proposition 1, the firm can either underinvest or overinvest in social investments, depending on the realization of $\tilde{\varepsilon}_i$. Indeed, in this case, the optimal linear contract is such that the manager is excessively responsive to ESG scores, i.e. to the realization of $\tilde{\varepsilon}_i$. Proposition 1 still shows that, in this case, *on average* the firm underinvests in social investments from the board's perspective. Intuitively, the difficulty of aligning interests with respect to SEP reduces the second-best level of expected social investment below the first-best level.

Thus, the deviation from the first-best outcome depending on whether $\alpha_B \gtreqless \alpha_I$ is asymmetric. When $\alpha_B \leq \alpha_I$, it is possible to reach the first-best outcome using available compensation instruments – with profits-based and stock price-based compensation. On the contrary, when $\alpha_B > \alpha_I$, it is impossible to do so. In this latter case, the firm will use ESG scores-based and stock price-based compensation, and social investment will be inefficiently allocated for the reasons mentioned in preceding paragraphs. This result relies on a wedge between the observability and the contractibility of social investments, so that the stock price aggregates information in a way that could not be replicated by a contract. For simplicity, we have assumed that social investments are not contractible, but this result would still hold under the more realistic assumption that social investments are not fully contractible.

Overall, the main forces at play in the model can be summarized as follows. Equity-based and profits-based compensation are used as substitutes to elicit managerial effort. At the same time, a board that wants to encourage social investments ($\alpha_B > 0$) starts by relying on stock price-based incentives (if the stock price is sensitive to perceived SEP, i.e. $\alpha_I > 0$). The social incentives thus provided can be adjusted by changing the relative weights of stock price-based and profits-based compensation while preserving effort incentives. Only at the point when β_x cannot be decreased further (i.e. $\beta_x = 0$) does the board start using measures of SEP.[12] Indeed, since it will foster inefficient gaming, using these measures in a contract is less efficient than relying on equity-based compensation.

Proposition 1 generates empirical implications about the use of SEP measures in incentive programs. First, it suggests that many firms should not use explicit incentives based on SEP measures. Indeed, measures of SEP are only used when the board's preferred level of social investments exceeds the one that maximizes the stock price. This suggests that compensation based on SEP measures and socially responsible investors are substitutes rather than complements. For example, suppose that investors become less socially responsible because of changing investor sentiment. Then, supposing also that the degree of social responsibility preferred by boards does not change, the model predicts a rise in the use of SEP-based incentives. This can contribute to explain recent trends which might otherwise appear paradoxical.[13] Second, the intrinsic quality of

[12]Thus, the manager's compensation needs not be highly sensitive to ESG scores for social investments to be effectively incentivized. This is in contrast to the view that SEP-based compensation, which are "economically insignificant", are either inconsistent with incentive theory or suggestive of window dressing (Walker (2022)).

[13]A recent article notes that the use of ESG-based incentives is on the rise even though investors seem to be less concerned about ESG as suggested by declining inflows into ESG funds in 2023. Source: 76% of companies link

SEP measures, captured by σ_ε in the model, does not matter for the inclusion of these measures in the contract. However, it matters for the sensitivity of pay to these measures when they are used. The worse the quality of these measures (i.e. the higher σ_ε), the lower is the sensitivity of pay to these measures (see Proposition 1).

In the standard multitasking model of Holmström and Milgrom (1991), inefficiencies emanate from imperfect or inexistent ex-post performance measurement on some dimension, which prevents adequate incentive provision on other dimensions if efforts on various dimensions are complements or substitutes. In the multitasking model of Feltham and Xie (1994), inefficiencies emanate from noisy ex-post performance measurement on several dimensions, which can induce deviations from the first-best (multidimensional) action to reduce the risk borne by a risk averse agent. In our model, by contrast, the costs of various actions are independent (the cost of investment on dimension i only depends on this investment) and the manager is risk neutral. Instead, inefficiencies arise because of the manager's ex-ante awareness of the biases of each SEP measure, which results in either excessive or inadequate investment in each dimension. In Appendix B, we study the same setting with multitasking but without gaming (i.e. the manager does not observe ε_i at the time of making investment decisions). In that alternative setting, the first-best outcome can always be obtained, and the quality of SEP measures (as measured by σ_ε) does not change the sensitivity of pay to these measures.

In subsections 2.2 and 2.3, we extend the baseline model in two directions. First, we let the board have heterogeneous preferences with respect to various dimensions of SEP. Second, we study the outcome when there are several sets of ESG scores.

2.2 Heterogeneous social preferences

In this subsection, we allow the board to have a different preference with respect to each dimension i of SEP relative to profits, as measured by $\alpha_B^i \geq 0$, for $i = 1, 2$ (by contrast, in the baseline model, $\alpha_B^1 = \alpha_B^2$). We also let the stock price impact of perceived SEP, as measured by $\alpha_I^i \geq 0$, potentially be different across dimensions of SEP. For simplicity and tractability, in this subsection we assume $\sigma_y = 0$.

pay to ESG performance in rising trend: WTW, CFO Dive, Jan 24 2024.

Proposition 2

(i) If $\alpha_B^i \geq \alpha_I^i$ for $i = 1, 2$, then $\beta_x = 0$, $\beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)$, and, for $i = 1, 2$:

$$\frac{\beta_i}{\beta_p} = (\alpha_B^i - \alpha_I^i)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2 + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\right); \tag{10}$$

(ii) If $\alpha_B^i < \alpha_I^i$ for $i = 1, 2$, then $\beta_x > 0$;

(iii) If $\alpha_I^i = \alpha_I$ for $i = 1, 2$, and $\alpha_B^1 < \alpha_I \leq \alpha_B^2$ with $\frac{\alpha_I}{\alpha_B^1}$ sufficiently large and $\frac{\alpha_B^2}{\alpha_I}$ sufficiently close to 1, then $\beta_x > 0$, $\beta_p > 0$, $\beta_1 = 0$, and $\beta_2 > 0$.

(iv) If α_B^i is sufficiently large and $\alpha_I^i > 0$, then $\beta_x = 0$, $\beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)$, and $\frac{\beta_i}{\beta_p}$ is as in equation (10).

In part (i), when the board prefers a higher level of social investment than the one that maximizes the stock price on all dimensions, there is no profits-based compensation.[14] On the contrary, in part (ii), when the board prefers a lower level of social investment than the one that maximizes the stock price on all dimensions, then it uses profits-based compensation. In both cases, the intuition is as in the previous section.

In part (iii), the board prefers a much lower level of social investment than the one that maximizes the stock price on dimension 1, and a slightly higher level of social investment on dimension 2. Then, compensation is sensitive to profits to deter investment in dimension $i = 1$ of SEP, but it is also sensitive to ESG score $i = 2$ to encourage investment in dimension $i = 2$ of ESG. This is illustrated in Example 1 below. By contrast, this outcome is not possible in Proposition 1, where preferences for social output are homogeneous across dimensions of social output.

[14]More generally, we show in the proof of Proposition 2 that the firm will not use profits-based compensation ($\beta_x = 0$) if a weighted sum of $\alpha_B^i - \alpha_I^i$ is positive, as opposed to $\alpha_B - \alpha_I > 0$ in the case with homogeneous preferences for social output. In the proof of Proposition 2, we show that, letting $\beta \equiv \frac{\beta_p}{\beta_x+\beta_p}$, the optimal value of β is a solution to the following problem: $\min_\beta \sum_{i=1,2} \Gamma_i \left(\alpha_B^i - \beta\alpha_I^i\right)^2$ s.t. $\beta \leq 1$ where Γ_i is a positive constant defined in the proof of Proposition 2. That is, the optimal fraction of stock price-based incentives as a proportion of total financial incentives, $\frac{\beta_p}{\beta_x+\beta_p}$, minimizes a weighted average quadratic distance between the board's preference for dimension i of SEP and investors' preference for the same dimension of SEP times $\frac{\beta_p}{\beta_x+\beta_p}$. Accordingly, we find that this ratio is optimally either equal to 1 (when the constraint $\beta_x \geq 0$ is binding), or to $\frac{\beta_p}{\beta_x+\beta_p} = \frac{\sum_{i=1,2}\Gamma_i\alpha_I^i\alpha_B^i}{\sum_{i=1,2}\Gamma_i{\alpha_I^i}^2}$. Intuitively, the optimal ratio $\frac{\beta_p}{\beta_x+\beta_p}$ is less than one when investors tend to have a stronger preference for SEP compared to the board, which results in the board giving profits-based incentives to the manager to counterbalance stock price-based incentives.

Example 1 *Suppose that* $\alpha_B^1 = 0$, $\alpha_I^1 = 1$, $\alpha_B^2 = 1.1$, $\alpha_I^2 = 1$, $\sigma_\eta = 1$, $\sigma_\varepsilon = 1$, $\sigma_x = 1$, $\sigma_z = 1$, $\bar{\eta} = 2$, $c_e/(\overline{e} - \underline{e}) = 0.1$. *Then we have* $\beta_1 = 0$, $\beta_2 = 0.03$, $\beta_x = 0.02$, *and* $\beta_p = 0.16$.

In part (iv), when the board prefers a much higher level of social investment than the one that maximizes the stock price on dimension i, then it uses SEP-based compensation on dimension i. Moreover, there is no profits-based compensation, even if the board prefers a much lower level of social investment than the one that maximizes the stock price on the other dimension (j). This is surprising because the board could use profits-based compensation to discourage social investments on dimension j ($\beta_x > 0$), and partly offset the effect of profits-based compensation on dimension i of SEP by incentivizing investment on this dimension with the corresponding ESG score (by further increasing β_i). However, this would lead to an (already discussed) inefficiency which is especially costly for a board that cares a lot about this dimension of SEP. Thus, contract complexity, as proxied by the different types of performance measures used, does not necessarily rise when the divergence in preferences increases. This is illustrated in Example 2 below.

Example 2 *Suppose that* $\alpha_B^1 = 0$, $\alpha_I^1 = 1$, $\alpha_B^2 = 2$, $\alpha_I^2 = 1$, $\sigma_\eta = 1$, $\sigma_\varepsilon = 1$, $\sigma_x = 1$, $\sigma_z = 1$, $\bar{\eta} = 2$, $c_e/(\overline{e} - \underline{e}) = 0.1$. *Then we have* $\beta_1 = 0$, $\beta_2 = 0.17$, $\beta_x = 0$, *and* $\beta_p = 0.20$.

In Example 2, the only difference with respect to Example 1 is that the board cares even more about the second social dimension. In Example 2, a board with a strong preference for the second social dimension will not use profits-based compensation to discourage investment in the first dimension – even though it could separately encourage investment in the second dimension by further increasing β_2.

In summary, having heterogeneous preferences for social output across economic agents and across dimensions of social investments is a necessary but insufficient condition for a managerial contract to be explicitly contingent on three different types of performance measures: profits, the stock price, and some ESG scores. These results contribute to the nascent literature on the complexity of executive compensation (Murphy and Sandino (2020), Burkert et al. (2023), Albuquerque et al. (2024)).

2.3 Multiple ESG scores

We now extend the model to analyze the case with multiple ESG scores on each social dimension. In practice, several ESG rating agencies provide ESG scores and ratings. These scores are informative, in the sense that they affect firms' stock prices (Berg et al. (2021)), but there is evidence of substantial divergence across these ratings, including on the measurement of the same dimension of SEP (Berg, Kölbel, and Rigobon (2022)).

We now assume that there are N ESG raters. Each rater j provides a set of ESG scores $\{m_1^j, m_2^j\}$. ESG score on dimension i by ESG rater j is defined as $m_i^j \equiv \varepsilon_i^j y_i$. We will present results based on two different assumptions on the joint distribution of these scores.

To start, we analyze the effect of increasing the number of ESG scores with uncorrelated noise terms. Specifically, we assume that the noise terms in ESG scores are independent and identically distributed (i.i.d.), with a variance of σ_ε^2 (see section E.4 in the Appendix for additional details). This implies that ESG scores on dimension i are uncorrelated conditional on η_i, but they are unconditionally positively correlated because of their dependence on η_i. Let $\bar{\varepsilon}_i = \frac{1}{N}\sum_{j=1}^{N} \varepsilon_i^j$ be the average signal on the firm's social productivity on dimension i generated by ESG scores. This average signal is a sufficient statistic for the mean of the distribution (see section E.4).

Proposition 3 *As the number of ESG scores is increased, social investment and expected social output converge asymptotically to the first-best levels of these variables:*

$$\lim_{N\to\infty} y_i(\bar{\varepsilon}_i) = \frac{\alpha_B}{2}\frac{\bar{\varepsilon}_i}{\theta_i} \qquad \textit{and} \qquad \lim_{N\to\infty} \mathbb{E}[\tilde{\eta}_i \tilde{y}_i] = \frac{\alpha_B}{2}\frac{\bar{\eta}^2 + \sigma_\eta^2}{\theta_i}. \tag{11}$$

Proposition 3 implies that increasing the number of ESG scores allows to overcome the agency problem. The reason is as follows. In this setting, the average score on any given dimension is a sufficient statistic. Therefore, the manager considers the average score on each dimension of SEP when choosing social investments. Additional scores reduce the variance of the bias of the average score, which increases its quality, and diminishes the manager's propensity to game the ESG scoring system. Intuitively, with a single score on any given dimension of SEP, incentives are distorted in one direction only. With additional scores which are imperfectly correlated with the first score, incentives are still distorted but not necessarily in the same direction. Proposition 3 shows that, as the number of ESG scores gets very large, the incentives to game the system become negligible, and vanish in the limit. This result goes against the notion that heterogeneity in performance measurement will lead executives to game the incentive system.[15]

An important question related to ESG scores if whether adding ESG scores with a lower quality than an existing set of scores can be beneficial. For corporate governance purposes, is it better to have one high-quality set of ESG scores or is it valuable to complement these scores with additional scores which can be correlated with the former and which can also have a lower quality (higher noise)?

We now analyze the effect of adding to an existing set of scores some possibly more noisy and correlated ESG scores. Specifically, on each dimension i of SEP, there are two scores, $\tilde{\varepsilon}_i^1$ and $\tilde{\varepsilon}_i^2$, that follow a multivariate normal distribution. On any dimension i of SEP, each ESG score j is normally distributed with mean η_i and variance $\sigma_{\varepsilon_i}^{j^2}$ (or equivalently precision $1/\sigma_{\varepsilon_i}^{j^2}$); conditional on η_i, these two scores are correlated with correlation coefficient $\rho \in (-1, 1)$ (see section E.5 in the Appendix for additional details). Letting the signal-to-noise ratio of a score be $\sigma_\eta/\sigma_{\varepsilon_i}^j$, our assumptions can generate any signal-to-noise ratio for each score.

We show in Appendix E.5 that neither ε_i^1 nor ε_i^2 is a sufficient statistic for $\tilde{\eta}_i$. The additional set of ESG scores is thus informative, and a naïve application of the informativeness principle (Holmström (1979)) would conclude that it is useful for contracting. We now analyze the effect of adding ESG scores on social investment and expected social output.

Proposition 4 *With two sets of scores, social investment is equal to its first-best level if and only if $\alpha_B \leq \alpha_I$. Moreover, expected social output is higher with two sets of scores rather than one set of scores if the precision of the second scores is sufficiently high or their correlation with the first scores is sufficiently low.*

Proposition 4 shows that the result from Proposition 1 that the first-best social investment can be induced if and only if $\alpha_B \leq \alpha_I$ is robust to the addition of ESG scores.[16] Proposition 4 also shows that the addition of ESG scores can decrease the expected social output of the firm. Intuitively, additional scores which are not perfectly correlated with the first set of scores are

[15] For example, a recent press article about ESG-based executive compensation mentioned that a "lack of standardized data and disclosure may be opening the way for some to game the system." Source: "Executive compensation tied to ESG is growing but open to abuse", *Financial Post* April 17 2023.

[16] The first-best investment in dimension i of SEP with two sets of ESG scores, $y_i^*(\varepsilon_i^1, \varepsilon_i^2)$, is defined in equation (108) in the Appendix.

informative, even if their quality is low, and they consequently affect the stock price. Since the stock price is used to provide incentives, the board cannot just "ignore" additional scores.[17] The problem is that the bias of additional scores can be so strong (high $\sigma_{\varepsilon_i}^j$) and their correlation ρ with the first set of scores so high that adding these additional scores will tend to worsen the bias induced by SEP measures. Indeed, Figure 2 illustrates that the availability of another set of scores with a low quality and a high correlation with the first set of scores will decrease expected social output.

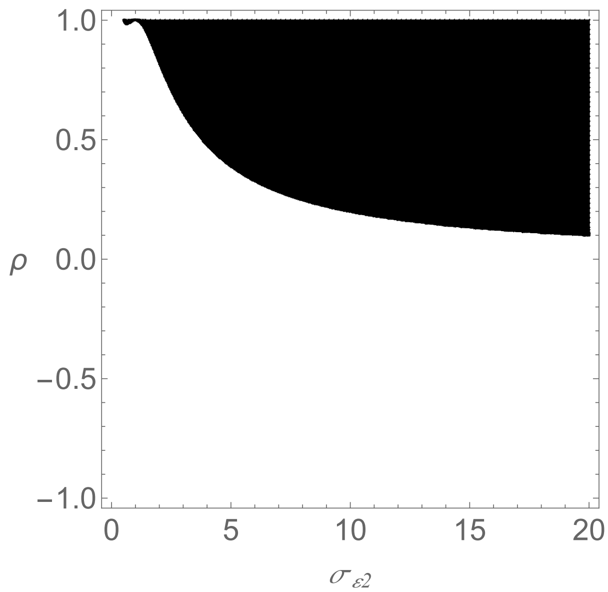

Figure 2: The black area is the subset of parameter values $\{\sigma_{\varepsilon_2}, \rho\}$ such that expected social output $\mathbb{E}[\tilde{\eta}_i \tilde{y}_i]$ on dimension i is higher with one rather than two ESG scores on this dimension. The initial score is characterized by $\sigma_{\varepsilon_1} = 1$, the additional score by σ_{ε_2}, and the correlation coefficient of the error terms in these ESG scores is ρ. We have $\bar{\eta} = 1$, and $\sigma_\eta = 1$.

Propositions 3 and 4 have important lessons for SEP measurement "convergence" or "harmonization", which is debated and frequently advocated.[18] On the one hand, Proposition 3 shows

[17]In section 2 of the Online Appendix, we show that, with $c_e \to 0$, the result that having two sets of scores instead of one can reduce the firm's expected social output does not hold in the absence of a stock price (i.e., when the firm can only contract based on profits and ESG scores).

[18]As stated by Larcker et al. (2022): "The major credit rating agencies Moody's, Standard & Poor's, and Fitch are subject to regulation by the Securities and Exchange Commission which requires covered firms to adhere to certain policies, procedures, and protections to reduce conflicts of interest and improve market confidence in their quality. Should ESG ratings be subject to similar requirements?" Accordingly, there is a regulatory push for uniform standards in SEP measurement and reporting, see: EU watchdog says ESG rating firms need rules to

that disagreement among ESG raters (i.e. a low correlation across ESG scores' noise terms) is beneficial for corporate governance purposes. If regulation or "harmonization" of ESG ratings results in a decrease in the number of available ESG ratings, our results suggest that it may result in more gaming of the remaining ESG scores and ratings by managers, for example via increased greenwashing. On the other hand, Proposition 4 shows that ESG scores with sufficiently low quality are detrimental because they affect the stock price and distort managerial incentives. Together, these results suggest that ESG raters and regulatory efforts should focus on improving ESG scores' quality rather than reducing their dispersion. Thus, the concern about the dispersion of ESG scores (low or negative correlation across scores) documented in Chatterji et al. (2016) and Berg, Kölbel, and Rigobon (2022) is not necessarily warranted.

Regulation or harmonization might admittedly improve the quality of ESG scores. Our results suggest that this increase would need to be sufficiently high to offset the increased gaming effect resulting from a decrease in the number of scores that we highlighted. In other words, an improvement in the quality of ESG scores is necessary but insufficient for regulation or harmonization to be beneficial. In the case when the noise in scores is i.i.d. as in Proposition 3, a hypothetical unique standardized score would need to have a precision which is higher than N times the precision of an individual score if it replaces N different scoring methodologies. This sets a high bar for regulation or harmonization.[19]

3 Conclusion

Criticisms levelled at ESG-based compensation cannot be answered by existing models of multitasking in which performance is measured ex-post with noise: as is well-known, imperfect measurement does not rule out performance-based compensation. To address these criticisms, we allow the manager to know the bias in performance measurement at the time of making investment decisions, which allows him to game SEP-based incentives. We also study the outcome with

stop 'greenwashing', *Reuters* February 12 2020, Regulatory Solutions: A Global Crackdown on ESG Greenwash, Harvard Law School Forum on Corporate Governance, June 23 2022. This harmonization argument is also made in influential academic research (Berg, Kölbel, and Rigobon (2022)). It is related to regulations such as government mandated CSR reporting in the European Union (Fiechter, Hitz, and Lehmann (2022)).

[19]Some of these concerns might help explain why there is only limited support in the corporate world for harmonization efforts with respect to SEP reporting. For example, despite corporate commitments to SEP, only 25% of US CFOs support the Securities and Exchange Commission's proposal to standardize climate disclosure. Source: There's an ESG backlash inside the executive ranks at top corporations, CNBC Sept 29 2022.

heterogeneous SEP measurement.

We show that SEP measures will only be used to encourage SEP when stock price-based compensation is insufficient for this purpose. Thus, SEP-based compensation will only be used when the level of social investments preferred by the board does not maximize the stock price – which contradicts the notion that SEP-based compensation will increase stock prices. Moreover, even though it is sometimes second-best optimal, the reliance on measures of SEP distorts the manager's incentives for social investments. This distortion reduces the optimal sensitivity of compensation to SEP measures. Finally, we have shown that the often criticized heterogeneity in social performance measurement can be used to mitigate gaming and improve efficiency.

The model has normative implications for the regulation of SEP measures including ESG scores and ratings. It suggests that the harmonization of SEP measures may have a counterproductive effect. Indeed, it is harder for the manager to game ESG scores and ratings when there are different ESG raters that use a variety of methodologies. By contrast, there is a recent push toward a uniform standard, which would be easier to game than a variety of methodologies. For example, the creation of the International Sustainability Standards Board (ISSB), whose objective is to develop a global standard for sustainability reporting, was announced at the 2021 United Nations Climate Change Conference. Our results contribute a new perspective to this debate.

Finally, it is worth noting that our optimal contracting perspective ignored factors such as self-dealing and rent extraction. Especially in poorly governed firms, compensation based on SEP measures might occasionally be used to undeservedly inflate managerial compensation (Bebchuk and Tallarita (2022)). By delineating the circumstances in which SEP measures-based compensation is optimal, our results can help identify instances in which this type of compensation exacerbates rather than ameliorates the agency problem.

References

Albuquerque, A.M., Carter, M.E., Guo, Z.M., Lynch, L.J., 2022. Complexity of CEO compensation packages. Working paper, Boston University.

Barbalau, A., Zeni, F., 2022. The optimal design of green securities. Working paper, University of Alberta.

Barber, B.M., Morse, A., Yasuda, A., 2021. Impact investing. *Journal of Financial Economics*, 139, 162-185.

Baron, D.P., 2008. Managerial contracting and corporate social responsibility. *Journal of Public Economics*, 92, 268-288.

Bauer, R., Ruof, T., Smeets, P., 2021. Get real! Individuals prefer more sustainable investments. *Review of Financial Studies*, 34, 3976-4043.

Bebchuk, L.A., Tallarita, R., 2022. The perils and questionable promise of ESG-based compensation. *Journal of Corporation Law*, forthcoming.

Berg, F., Kölbel, J.F., Pavlova, A., Rigobon, R., 2021. ESG confusion and stock returns: Tackling the problem of noise. Working paper, MIT.

Berg, F., Kölbel, J.F., Rigobon, R., 2022. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26, 1315-1344.

Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? *Journal of Financial Economics*, 142, 517-549.

Bonham, J., Riggs-Cragun, A., 2022. Motivating ESG activities through contracts, taxes and disclosure regulation. Working paper, University of Chicago.

Broccardo, E., Hart, O. and Zingales, L., 2022. Exit versus voice. *Journal of Political Economy*, 130, 3101-3145

Bucourt, N., Inostroza, N., 2023. ESG investing and managerial incentives. Working paper, University of Toronto.

Burkert, S., Oberpaul, T., Tichy, N., Weller, I., 2023. Executive compensation complexity and firm performance. *Academy of Management Discoveries*, forthcoming.

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

Christensen, D.M., Serafeim, G., Sikochi, A., 2022. Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review*, 97, 147-175.

Cohen, S., Kadach, I., Ormazabal, G., Reichelstein, S., 2023. Executive compensation tied to ESG performance: International evidence. *Journal of Accounting Research*, forthcoming.

Dewatripont, M., Tirole, J., 2022. The morality of markets. Working paper, ECARES.

Duchin, R., Gao, J., Xu, Q., 2022. Sustainability or greenwashing: Evidence from the asset market for industrial pollution. Working paper, Boston College.

Edmans, A., 2021, Why companies shouldn't tie CEO pay to ESG metrics. *Wall Street Journal*, June 27.

Edmans, A., Gabaix, X., 2011. Tractability in incentive contracting. *Review of Financial Studies*, 24, 2865-2894.

Edmans, A., Gabaix, X., Jenter, D., 2017. Executive compensation: A survey of theory and evidence. *The Handbook of the Economics of Corporate Governance*, 1, 383-539.

Feltham, G.A., Xie, J., 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review*, 69, 429-453.

Fiechter, P., Hitz, J.M., Lehmann, N., 2022. Real effects of a widespread CSR reporting mandate: Evidence from the European Union's CSR Directive. *Journal of Accounting Research*, 60, 1499-1549.

Friedman, H.L., Heinle, M.S., Luneva, I., 2022. A theoretical framework for ESG reporting to investors. Working paper, UCLA.

Gaynor, M., Mehta, N., Richards-Shubik, S., 2023. Optimal contracting with altruistic agents: Medicare payments for dialysis drugs. *American Economic Review*, 113, 1530-1571.

Goldstein, I., Kopytov, A., Shen, L., Xiang, H., 2022. On ESG investing: Heterogeneous preferences, information, and asset prices. Working paper, University of Pennsylvania.

Grossman, S.J., Hart, O.D., 1979. A theory of competitive equilibrium in stock market economies. *Econometrica*, 47, 293-329.

Hart, O.D., Zingales, L., 2022. The new corporate governance. *University of Chicago Business Law Review*, 1, 195-216.

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

Heeb, F., Kölbel, J.F., Paetzold, F., Zeisberger, S., 2023. Do investors care about impact? *Review of Financial Studies*, 36, 1737-1787.

Holmström, B., 1979. Moral hazard and observability. *Bell Journal of Economics*, 10, 74-91.

Holmström, B., Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics & Organization*, 7, 24-52.

Holmström, B., Tirole, J., 1993. Market liquidity and performance monitoring. *Journal of Political Economy*, 101, 678-709.

Homroy, S., Mavruk, T., Nguyen, D., 2023. ESG-Linked Compensation, CEO Skills, and Shareholders' Welfare. *Review of Corporate Finance Studies*, forthcoming.

Hörner, J., Lambert, N.S., 2021. Motivational ratings. *Review of Economic Studies*, 88, 1892-1935.

Huang, S., Hwang, B.H., Lou, D. Yin, C., 2020. Offsetting disagreement and security prices. *Management Science*, 66, 3444-3465.

Humphrey, J., Kogan, S., Sagi, J.S., Starks, L.T., 2023, The asymmetry in responsible investing preferences. NBER Working Paper 29288.

Innes, R.D., 1990. Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory*, 52, 45-67.

Jensen, M.C., Murphy, K.J., 1990. Performance pay and top-management incentives. *Journal of Political Economy*, 98, 225-264.

Larcker, D.F., Pomorski, L., Tayan, B., Watts, E.M., 2022. ESG ratings: A compass without direction. Working paper, Stanford University.

Murphy, K.J., Sandino, T., 2020. Compensation consultants and the level, composition, and complexity of CEO pay. *The Accounting Review*, 95, 311-341.

Oehmke, M., Opp, M.M., 2023. A theory of socially responsible investment. *Review of Economic Studies*, forthcoming.

Pagano, M.S., Sinclair, G., Yang, T., 2018. Understanding ESG ratings and ESG indexes. *Research Handbook of Finance and Sustainability*. Edward Elgar Publishing.

Parlour, C.A., Rajan, U., 2020. Contracting on credit ratings: Adding value to public information. *Review of Financial Studies*, 33, 1412-1444.

Pástor, L., Stambaugh, R.F., Taylor, L.A., 2021. Sustainable investing in equilibrium. *Journal of Financial Economics*, 142, 550-571.

Rajan, R., Ramella, P., Zingales, L., 2023. What purpose do corporations purport? Evidence from letters to shareholders. Working paper, University of Chicago.

Riedl, A., Smeets, P., 2017. Why do investors hold socially responsible mutual funds? *Journal of Finance*, 72, 2505-2550.

Shleifer, A., Vishny, R.W., 1997. A survey of corporate governance. *Journal of Finance*, 52, 737-783.

Tirole, J., 2006. *The Theory of Corporate Finance*. Princeton university press.

Walker, D.I., 2022. The economic (in)significance of executive pay ESG incentives. *Stanford Journal of Law, Business & Finance*, 27, 317-350

Yu, J., 2011. Disagreement and return predictability of stock portfolios. *Journal of Financial Economics*, 99, 162-183.

Appendix

A ESG ratings and scores

ESG scores and ratings, which are "third-party assessment[s] of corporations' ESG performance" (Berg, Kölbel, and Rigobon (2022)), were originally developed to allow investors to screen companies for ESG (Environmental, Social and Governance) performance.

ESG raters typically provide several different types of measures of social and environmental performance. They provide an aggregate rating for a firm, as well as separate scores that reflect its performance on various dimensions of SEP: "category scores represent a rating agency's assessment of a certain ESG category. They are based on different sets of indicators that each rely on different measurement protocols." (Berg, Kölbel, and Rigobon (2022)) These categories include greenhouse gases emissions, workplace safety, board composition, etc.

In order to be measures of SEP activities that are comparable across firms and therefore useful to investors, ESG ratings and scores are highly standardized with publicly known formulas. For example, when describing their ESG scores, S&P Global mentions: "We publish our S&P Global ESG Score methodology on our website." Likewise, Bloomberg's ESG Scores are "fully transparent including methodology & company-reported data underlying each score."[20] Other measures of SEP, such as carbon intensity or board diversity, also share this feature.

This standardization leaves them open to gaming. Indeed, it is widely acknowledged that a firm can improve its ESG ratings by engaging in actions that improve perceptions of its SEP rather than its actual SEP (Walker (2022), Duchin, Gao, and Xu (2023)). In practice, some executive compensation contracts include ESG ratings and scores as performance metrics (Cohen et al. (2023), Table 3).

B Case with no gaming

We assume that the manager does not observe ε_i at the time of making investment decisions. The manager's objective function given $\{\beta_x, \beta_p, \beta_1, \beta_2\}$, $\{\varepsilon_1, \varepsilon_2\}$, and effort e is:

$$\mathbb{E}\left[w + \beta_x\left(e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x\right) + \beta_p \tilde{p} + \beta_1 y_1 \varepsilon_1 + \beta_2 y_2 \varepsilon_2 | e\right] - C(e) \tag{12}$$

[20]Sources: Transparency and Impact: The Essential Principles of ESG, by Douglas L. Peterson, President & Chief Executive Officer of S&P Global, and Bloomberg Professional Services, www.bloomberg.com/explore/esg/.

For $\beta_x + \beta_p > 0$, the manager's objective function is concave in y_i, for $i = 1, 2$. The first-order condition (FOC) with respect to y_i is:

$$\beta_x (-2\theta_i y_i) + \beta_p (-2\theta_i y_i + \alpha_I \bar{\eta}) + \beta_i \bar{\eta} = 0 \quad \Leftrightarrow \quad y_i = \frac{\beta_i + \beta_p \alpha_I}{\beta_x + \beta_p} \frac{\bar{\eta}}{2\theta_i} \tag{13}$$

The first-best optimal value of y_i is:

$$y_i^*(\varepsilon_i) = \alpha_B \frac{\bar{\eta}}{2\theta_i} \tag{14}$$

This can be achieved by setting:

$$\frac{\beta_i + \beta_p \alpha_I}{\beta_x + \beta_p} = \alpha_B. \tag{15}$$

Similarly to the main model, the manager will optimally exert high effort ($e = \bar{e}$) if and only if:

$$\beta_x + \beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \geq \frac{c_e}{\bar{e} - \underline{e}} \tag{16}$$

Overall, the first-best can be achieved with the following contract:

$$\beta_x = \frac{c_e}{\bar{e} - \underline{e}} \tag{17}$$

$$\beta_p = 0 \tag{18}$$

$$\beta_i = \alpha_B \beta_x = \alpha_B \frac{c_e}{\bar{e} - \underline{e}} \qquad \text{for } i = 1, 2 \tag{19}$$

and w such that the manager is at his reservation level of utility given these values of $\{\beta_x, \beta_p, \beta_1, \beta_2\}$. When $\alpha_B \geq \alpha_I$, the first-best optimal outcome can also be induced with a contract that involves stock price-based compensation:

$$\beta_x = 0 \tag{20}$$

$$\beta_p = \left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right) \frac{c_e}{\bar{e} - \underline{e}} \tag{21}$$

$$\beta_i = (\alpha_B - \alpha_I) \left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right) \frac{c_e}{\bar{e} - \underline{e}} \qquad \text{for } i = 1, 2 \tag{22}$$

These results emphasize that, in the case without gaming, there is no inefficiency: the first-best outcome can always be attained. Thus, without gaming, the sensitivity of the manager's compensation to ESG measures is *not* reduced to mitigate inefficiencies.

C Discussion of modeling assumptions and preferences

There are three interpretations for the specification of the board's and investors' social and environmental preferences. In the rest of the paper, we will discuss the model using the third interpretation listed below.

The first interpretation is that the board is not intrinsically socially responsible, but it uses incentive pay to commit to an investment policy that would otherwise not be in the best interests of the firm ex-post. For example, this can be useful to raise funding from socially and environmentally responsible investors at a lower cost, or to hire employees who care about these issues. To be credible, the firm must commit to be socially responsible in the future.[21] Specifically, it should invest "as if" it were socially responsible. This can be achieved by setting a compensation contract "as if" the board had the objective function in equation (2). In this interpretation of the model, investors are socially responsible if they put a positive weight α_I on social output relative to profits. In principle, this hypothesis can explain the concomitant rise of commitments such as "sustainability pledges", and the increased reliance on social and environmental measures of firm performance in executive compensation.

The second interpretation is that the board and investors are not intrinsically socially responsible, but they are aware of political and judicial pressures emanating from activists and regulators. These third parties may punish firms that generate negative externalities, for example by requesting or mandating reparations for harm caused in the past. Even though this might not affect the firm's profitability during the manager's tenure (until $t = 2$), these actions might be costly to the firm in the distant future ($t = 3$). This heightened concern can be explained by recent shifts in public opinion. According to the US Department of Justice: "criminal prosecution acknowledges that environmental stewardship has become a mainstream value, such that most Americans recognize that polluting ... [is] repugnant." In 2023, the US Supreme Court allowed lawsuits by municipalities seeking to hold energy companies liable for harms caused by climate emissions to move forward.[22] This is related to the notion of "enlarged fiduciary duty" proposed by Tirole (2001), in which stakeholders could sue a firm whose actions did not "follow the

[21] In some instances, when green investments are well-defined, this can alternatively be induced by raising funding via green bonds (Barbalau and Zeni (2022)). In other cases, socially responsible investments are not well-defined, i.e. they cannot be described in a contract a priori, or there is not enough information at the contracting stage to determine efficient investments.

[22] Sources: https://www.justice.gov/enrd/environmental-crime-victim-assistance/prosecution-federal-pollution-crimes and https://www.nbcnews.com/politics/supreme-court/supreme-court-rejects-oil-companies-appeals-climate-change-disputes-rcna49823

mandate of the stakeholder society".

In this interpretation of the model, the random variable $\tilde{y}$ is the monetary amount in penalties imposed on the firm at $t = 3$, and α_B and α_I are respectively the board's and investors' discount factors for $t = 3$ cash flows relative to $t = 2$ cash flows. Discount factors could differ because long-term shareholders and stock market investors have different endowments and markets are incomplete (Grossman and Hart (1979)).[23] Discount factors could also differ because the firm's shareholders enjoy substantial private benefits of control (and are therefore unwilling to trade) but are more or less patient than stock market investors. For example, suppose that the firm's shareholders are more patient than the marginal stock market investor. In this case, the discount factor applied to $t = 3$ penalties relative to $t = 2$ cash flows is higher for shareholders than for the marginal stock market investor, i.e. $\alpha_B > \alpha_I$. Finally, discount factors could differ because of disagreement between the board and investors with respect to the extent of $t = 3$ penalties for social and environmental damages. In this interpretation, differences between so-called "discount factors" would reflect the different beliefs associated with $t = 3$ penalties.[24]

The third, more literal interpretation, is that shareholders (as represented by the board) and stock market investors intrinsically value the social and environmental impact of the firm. This can be justified based on the empirical evidence that investors have social and environmental concerns, and that they are willing to sacrifice financial return to this end (Riedl and Smeets (2017), Hartzmark and Sussman (2019), Barber, Morse, and Yasuda (2021), Bauer, Ruof, and Smeets (2021), Bolton and Kacperczyk (2021), Humphrey et al. (2023), Haber et al. (2022), Heeb et al. (2023)). The modelization of social preferences of the board (in equation (2)) and of investors is then similar to the one in Pástor, Stambaugh, and Taylor (2021), Broccardo, Hart, and Zingales (2022), Hart and Zingales (2022), Friedman, Heinle, and Luneva (2022), Goldstein et al. (2022), and Dewatripont and Tirole (2022). This is also similar to the modelization of altruism in Gaynor et al. (2023), and consistent with the empirical evidence on ESG-linked compensation (Homroy, Mavruk, and Nguyen (2023)).

In this interpretation of the model, $\tilde{y}$ is the amount of "social output" (including positive externalities and reductions in negative externalities) produced by the firm at $t = 3$, and α_B and α_I are respectively the board's and investors' preference for social output relative to cash flows. A discrepancy between these two preference parameters can arise because the firm's shareholders are not necessarily the same economic agents as investors who actively trade on financial markets. If investors have warm-glow preferences and

[23]Grossman and Hart (1979) note that marginal rates of substitution will then be heterogeneous across shareholders (or "investors"), i.e., each of them will have her own discount factor.

[24]It is noteworthy that disagreement across investors reduces the discount rate used for stock pricing (Yu (2011), Huang et al. (2020)). In our model, α_I is the discount factor that matters for stock pricing, i.e. disagreement across investors would increase α_I.

intrinsically value holding equity in a socially responsible firm, of if they represent fund managers with impact mandates for socially responsible investments, then their stock market investments have a stock price impact captured by $\alpha_I > 0$. On the contrary, if atomistic investors are affected by externalities, which they take as given, but do not intrinsically value socially responsible investments, then $\alpha_I = 0$ even though they would be better off if the firm reduced externalities. In this latter case, a board with $\alpha_B > 0$ solves a coordination problem by encouraging externalities mitigation above and beyond the level that would maximize the stock price (Oehmke and Opp (2024)).

D Proofs

Proof of Lemma 2:

The contract is characterized by a fixed payment w to the manager and by the sensitivity of pay to performance with respect to $\tilde{x}$, $\tilde{p}$, $\tilde{m}_1$, and $\tilde{m}_2$, which is given by $\{\beta_x, \beta_p, \beta_1, \beta_2\}$, respectively. The principal's optimization problem is:

$$\max_{e,\beta_x,\beta_p,\beta_1,\beta_2} \mathbb{E}[V(x,y,e)] \quad \text{s.t.} \quad \{e^*, y_1^*, y_2^*\} = \arg\max_{e,y_1,y_2} \mathbb{E}[u(x,y,e)], \quad \text{and} \quad \beta_x \geq 0, \beta_p \geq 0 \tag{23}$$

where:

$$\mathbb{E}[V(x,y,e)] = \mathbb{E}\left[\tilde{x} + \alpha_B \tilde{y} - \left(w + \beta_x \tilde{x} + \beta_p \tilde{p} + \sum_{i=1,2} \beta_i \tilde{m}_i\right)\right] \tag{24}$$

$$\mathbb{E}[u(x,y,e)] = \mathbb{E}\left[w + \beta_x \tilde{x} + \beta_p \tilde{p} + \sum_{i=1,2} \beta_i \tilde{m}_i\right] - C(e) \tag{25}$$

Manager's objective function given $\{\beta_x, \beta_p, \beta_1, \beta_2\}$, $\{\varepsilon_1, \varepsilon_2\}$, and effort e:

$$\mathbb{E}\left[w + \beta_x \left(e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x\right) + \beta_p \tilde{p} + \sum_{i=1,2} \beta_i y_i \varepsilon_i\right] - C(e) \tag{26}$$

where the stock price $\tilde{p}$ is as in Lemma 1. For $\beta_x + \beta_p > 0$, the manager's objective function is concave in y_i, for $i = 1, 2$. The first-order condition (FOC) with respect to y_i is:

$$\begin{aligned}
& \beta_x\left(-2\theta_i y_i\right)+\beta_p\left(-2\theta_i y_i+\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i+\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)\right)+\beta_i\varepsilon_i=0 \\
\Leftrightarrow \quad y_i(\varepsilon_i) &= \frac{\beta_i\varepsilon_i+\beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i+\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)}{\beta_x+\beta_p}\frac{1}{2\theta_i} \\
&= \frac{\left(\beta_i+\beta_p\alpha_I\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\varepsilon_i+\beta_p\alpha_I\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}}{\beta_x+\beta_p}\frac{1}{2\theta_i} \qquad (27)
\end{aligned}$$

Given his contract, the manager will optimally exert high effort ($e = \overline{e}$) if and only if:

$$\begin{aligned}
& \mathbb{E}\left[u(x,y,e)|e=\overline{e}\right]-c_e \geq \mathbb{E}\left[u(x,y,e)|e=\underline{e}\right] \\
\Leftrightarrow \quad & w+\beta_x\mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x|e=\overline{e}\right]+\beta_p\mathbb{E}\left[\mathbb{E}\left[\tilde{x}|t=1\right]+\alpha_I\mathbb{E}\left[\tilde{y}|t=1\right]\big|e=\overline{e}\right]+\beta_1 y_1\varepsilon_1+\beta_2 y_2\varepsilon_2-c_e \\
& \geq w+\beta_x\mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x|e=\underline{e}\right]+\beta_p\mathbb{E}\left[\mathbb{E}\left[\tilde{x}|t=1\right]+\alpha_I\mathbb{E}\left[\tilde{y}|t=1\right]\big|e=\underline{e}\right]+\beta_1 y_1\varepsilon_1+\beta_2 y_2\varepsilon_2 \\
\Leftrightarrow \quad & \beta_x+\beta_p\frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} \geq \frac{c_e}{\overline{e}-\underline{e}} \qquad (28)
\end{aligned}$$

The fixed component of pay, w, is set to guarantee the manager's participation for a given $e \in \{\underline{e}, \overline{e}\}$ (before the manager observes $\{\varepsilon_1, \varepsilon_2\}$):

$$\begin{aligned}
\mathbb{E}[u(x,y,e)] &= \bar{W} \\
\Leftrightarrow \quad w+\beta_x\mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x\right]+\beta_p\mathbb{E}[\tilde{p}|e]+\sum_{i=1,2}\beta_i\mathbb{E}[\tilde{m}_i] &= \bar{W}+C(e) \qquad (29)
\end{aligned}$$

Thus, equation (24) can be rewritten as:

$$\begin{aligned}
\mathbb{E}[V(x,y,e)] &= \mathbb{E}\left[(1-\beta_x)\left(e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x\right)+\alpha_B\left(\tilde{\eta}_1 y_1+\tilde{\eta}_2 y_2+\tilde{\epsilon}_y\right)-\left(w+\beta_p\tilde{p}+\sum_{i=1,2}\beta_i\tilde{m}_i\right)\right] \\
&= \mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x+\alpha_B\left(\tilde{\eta}_1 y_1+\tilde{\eta}_2 y_2+\tilde{\epsilon}_y\right)-\left(\bar{W}+C(e)\right)\right] \qquad (30)
\end{aligned}$$

We derive the "first-best" outcome without an agency problem (i.e. the incentive constraint in equation (28) can be ignored and information is symmetric):

$$\begin{aligned}
& \max_{y_1,y_2}\mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x+\alpha_B\left(\tilde{\eta}_1 y_1+\tilde{\eta}_2 y_2+\tilde{\epsilon}_y\right)-\left(\bar{W}+C(e)\right)|\varepsilon_1,\varepsilon_2\right] \\
\Leftrightarrow \quad & \max_{y_1,y_2}\left\{e-\theta_1 y_1^2-\theta_2 y_2^2+\alpha_B\left(\mathbb{E}\left[\tilde{\eta}_1|\varepsilon_1\right]y_1+\mathbb{E}\left[\tilde{\eta}_2|\varepsilon_2\right]y_2\right)-\left(\bar{W}+C(e)\right)\right\} \qquad (31)
\end{aligned}$$

where

$$\mathbb{E}\left[\tilde{\eta}_i|\varepsilon_i\right] = \mathbb{E}\left[\tilde{\eta}_i\right] + \frac{cov(\tilde{\eta}_i, \tilde{\varepsilon}_i)}{var(\tilde{\varepsilon}_i)}(\varepsilon_i - \mathbb{E}[\tilde{\varepsilon}_i]) = \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}(\varepsilon_i - \bar{\eta}) = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta} \tag{32}$$

The objective function is concave in y_i, so that the first-best optimum is given by the FOC:

$$y_i^*(\varepsilon_i) = \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta}\right) \tag{33}$$

We now consider the second-best outcome with an agency problem. The objective function of the board can be written as:

$$\mathbb{E}[V(x, y, e)] = e + \mathbb{E}\left[\sum_{i=1,2}\left(-\theta_i y_i^2 + \alpha_B \tilde{\eta}_i y_i - \left(\bar{W} + C(e)\right)\right)\right] \tag{34}$$

For given ε_i and y_i, define:

$$f(\varepsilon_i, y_i) \equiv -\theta_i y_i^2 + \alpha_B \mathbb{E}[\tilde{\eta}_i|\varepsilon_i] y_i \tag{35}$$

The first and second derivatives with respect to y_i are respectively:

$$f_y(\varepsilon_i, y_i) = -2\theta_i y_i + \alpha_B \mathbb{E}[\tilde{\eta}_i|\varepsilon_i] \tag{36}$$

$$f_{yy}(\varepsilon_i, y_i) = -2\theta_i \tag{37}$$

For a given e, maximizing the objective function of the board is equivalent to maximizing:

$$\mathbb{E}\left[\sum_{i=1,2} f(\varepsilon_i, y_i)\right] = \sum_{i=1,2}\int_{\eta_i}\int_{\varepsilon_i} f(\varepsilon_i, y_i)\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \tag{38}$$

Thus, for a given effort e and a given ε_i, the value of y_i that maximizes the board's objective function is the value of y_i that maximizes the expression in equation (38). By definition of $y_i^*(\varepsilon_i)$, for a given ε_i, the function $f(\varepsilon_i, y_i)$ is maximized by setting $y_i(\varepsilon_i) = y_i^*(\varepsilon_i)$ for $i = 1, 2$. The value of y_i that maximizes the board's objective function is the value of y_i that maximizes:

$$\max_{y_i} \mathbb{E}\left[\sum_{i=1,2}\left(f(\varepsilon_i, y_i) - f(\varepsilon_i, y_i^*(\varepsilon_i))\right)\right] = \sum_{i=1,2}\mathbb{E}\left[\left(f(\varepsilon_i, y_i) - f(\varepsilon_i, y_i^*(\varepsilon_i))\right)\right] \tag{39}$$

The function $f(\varepsilon_i, y_i)$ is quadratic in y_i. Therefore, for a given ε_i, a second-order Taylor expansion around $y_i^*(\varepsilon_i)$ is exact.

$$\begin{aligned} f(\varepsilon_i, y_i) &= f(\varepsilon_i, y_i^*(\varepsilon_i)) + f_y(\varepsilon_i, y_i^*(\varepsilon_i))(y_i - y_i^*(\varepsilon_i)) + \frac{1}{2} f_{yy}(\varepsilon_i, y_i^*(\varepsilon_i))(y_i - y_i^*(\varepsilon_i))^2 \\ &= f(\varepsilon_i, y_i^*(\varepsilon_i)) + \left(-2\theta_i y_i^*(\varepsilon_i) + \alpha_B \mathbb{E}[\tilde{\eta}_i|\varepsilon_i]\right)(y_i - y_i^*(\varepsilon_i)) - \theta_i (y_i - y_i^*(\varepsilon_i))^2 \end{aligned}$$

Thus:

$$\begin{aligned} f(\varepsilon_i, y_i) - f(\varepsilon_i, y_i^*(\varepsilon_i)) &= \left(-2\theta_i y_i^*(\varepsilon_i) + \alpha_B \left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)\right)(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)) \\ &\quad -\theta_i (y_i(\varepsilon_i) - y_i^*(\varepsilon_i))^2 \\ &= 2\theta_i \left(\left(-y_i^*(\varepsilon_i) + \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)\right)(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)) \right. \\ &\quad \left. -\frac{1}{2}(y_i(\varepsilon_i) - y_i^*(\varepsilon_i))^2\right) \\ &= 2\theta_i \left(-\frac{1}{2}(y_i(\varepsilon_i) - y_i^*(\varepsilon_i))^2\right) \\ &= -\theta_i \left(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)\right)^2 \end{aligned} \tag{40}$$

where we used equations (32) and (33) to get the first and third equalities, respectively. ■

Proof of Lemma 3 and Proposition 1:

Plugging the optimal investment y_i from equation (27) in the optimization problem using Lemma 2, gives:

$$\begin{aligned} &\min_{\beta_x,\beta_p,\beta_i} \sum_{i=1,2} \theta_i \int_{\eta_i}\int_{\varepsilon_i} \left(\frac{1}{2}\frac{1}{\theta_i}\frac{\left(\beta_i + \beta_p\alpha_I\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\varepsilon_i + \beta_p\alpha_I\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}}{\beta_x+\beta_p} - \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)\right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \\ \Leftrightarrow &\min_{\beta_x,\beta_p,\beta_i} \sum_{i=1,2} \frac{1}{\theta_i}\int_{\eta_i}\int_{\varepsilon_i}\left(\left(\frac{\beta_i}{\beta_x+\beta_p} + \left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\varepsilon_i + \left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B\right)\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \end{aligned} \tag{41}$$

Case 1. No nonnegativity constraints on contract parameters (Lemma 3).

Without nonnegativity constraints on contracting, simply set $\beta_i = 0$ and $\frac{\beta_p\alpha_I}{\beta_x+\beta_p} = \alpha_B \Leftrightarrow \beta_x = \frac{\beta_p\alpha_I}{\alpha_B} - \beta_p$, so that the expression under the integral sign in equation (41) is zero for any ε_i. Since this expression (a quadratic function) is nonnegative for any ε_i, achieving a value of zero for this expression at any ε_i maximizes the objective function of the principal for a given effort, and it implies that $y_i(\varepsilon_i) = y_i^*(\varepsilon_i)\ \forall\varepsilon_i$

(see equation (40)). To elicit high effort, use equation (28) to set:

$$\frac{\beta_p \alpha_I}{\alpha_B} - \beta_p + \beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} = \frac{c_e}{\bar{e} - \underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\bar{e} - \underline{e}} \left(\frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} - 1 \right)^{-1} \tag{42}$$

Case 2. Nonnegativity constraints on contract parameters (Proposition 1).

Conditional on η_i, the random variable $\tilde{\varepsilon}_i$, with PDF $\varphi(\varepsilon_i|\eta_i)$, is normally distributed with mean η_i and variance σ_ε^2. Thus, $\int_{\varepsilon_i} \varepsilon_i \varphi(\varepsilon_i|\eta_i) d\varepsilon_i = \eta_i$. This implies:

$$\int_{\eta_i} \int_{\varepsilon_i} \varepsilon_i \varphi(\varepsilon_i|\eta_i) \phi(\eta_i) d\varepsilon_i d\eta_i = \int_{\eta_i} \left(\int_{\varepsilon_i} \varepsilon_i \varphi(\varepsilon_i|\eta_i) d\varepsilon_i \right) \phi(\eta_i) d\eta_i = \int_{\eta_i} \eta_i \phi(\eta_i) d\eta_i = \bar{\eta} \tag{43}$$

Moreover, $\int_{\varepsilon_i} \varepsilon_i^2 \varphi(\varepsilon_i|\eta_i) d\varepsilon_i = \mathbb{E}[\tilde{\varepsilon}_i^2|\eta_i] = var(\tilde{\varepsilon}_i|\eta_i) + (\mathbb{E}[\tilde{\varepsilon}_i|\eta_i])^2 = \sigma_\varepsilon^2 + \eta_i^2$. This implies:

$$\int_{\eta_i} \int_{\varepsilon_i} \varepsilon_i^2 \varphi(\varepsilon_i|\eta_i) \phi(\eta_i) d\varepsilon_i d\eta_i = \int_{\eta_i} \left(\sigma_\varepsilon^2 + \eta_i^2 \right) \phi(\eta_i) d\eta_i = \sigma_\varepsilon^2 + \int_{\eta_i} \eta_i^2 \phi(\eta_i) d\eta_i = \sigma_\varepsilon^2 + \sigma_\eta^2 + \bar{\eta}^2 \tag{44}$$

The expression in equation (41) is globally convex with respect to β_i. The FOC w.r.t. β_i is:

$$\begin{aligned}
&\frac{1}{\theta_i} \int_{\eta_i} \int_{\varepsilon_i} \frac{2}{\beta_x + \beta_p} \varepsilon_i \Bigg(\left(\frac{\beta_i}{\beta_x + \beta_p} + \left(\frac{\beta_p \alpha_I}{\beta_x + \beta_p} - \alpha_B \right) \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} \right) \varepsilon_i \\
&+ \left(\frac{\beta_p \alpha_I}{\beta_x + \beta_p} - \alpha_B \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} \bar{\eta} \Bigg) \varphi(\varepsilon_i|\eta_i) \phi(\eta_i) d\varepsilon_i d\eta_i = 0 \\
\Leftrightarrow \quad & \left(\frac{\beta_i}{\beta_x + \beta_p} + \left(\frac{\beta_p \alpha_I}{\beta_x + \beta_p} - \alpha_B \right) \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} \right) \left(\sigma_\varepsilon^2 + \sigma_\eta^2 + \bar{\eta}^2 \right) = \left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p} \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} \bar{\eta}^2 = 0 \\
\Leftrightarrow \quad & \beta_i = \left(\alpha_B (\beta_x + \beta_p) - \beta_p \alpha_I \right) \left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} \frac{\bar{\eta}^2}{\sigma_\varepsilon^2 + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} \right),
\end{aligned} \tag{45}$$

where we used equations (43) and (44).

We establish an intermediary result: when the nonnegativity constraint on β_i does not bind, substituting for $\frac{\beta_i}{\beta_x+\beta_p}$ in equation (41):

$$\sum_{i=1,2}\frac{1}{\theta_i}\int_{\eta_i}\int_{\varepsilon_i}\Bigg(\Bigg(\left(\alpha_B-\frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\frac{\bar\eta^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar\eta^2}+\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)$$
$$+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\Bigg)\varepsilon_i+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar\eta\Bigg)^2\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i$$
$$=\left(\alpha_B-\frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)^2\frac{\sigma_\varepsilon^4}{(\sigma_\eta^2+\sigma_\varepsilon^2)^2}\sum_{i=1,2}\theta_i\int_{\eta_i}\int_{\varepsilon_i}\left(\frac{\bar\eta^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar\eta^2}\varepsilon_i-\bar\eta\right)^2\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \quad (46)$$

Start with the case $\alpha_B \leq \alpha_I$.

Supposing that the nonnegativity constraint on β_i does not bind, the expression under the integral sign in equation (46) is positive and independent from the contract. It is minimized by minimizing $\left(\alpha_B-\frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)^2$. With a nonnegativity constraint on β_x, this is achieved by setting $\beta_x=\max\{\frac{\beta_p\alpha_I}{\alpha_B}-\beta_p,0\}$. Substituting for β_x shows that the expression in equation (46) is zero as long as the nonnegativity constraint does not bind for β_x, i.e. $\frac{\beta_p\alpha_I}{\alpha_B}\geq\beta_p$, which with $\alpha_B\leq\alpha_I$ is true. This implies that $y_i(\varepsilon_i)=y_i^*(\varepsilon_i)\ \forall\varepsilon_i$ (see equation (40)).

Now supposing instead that the nonnegativity constraint on β_i binds, i.e. $\beta_i=0$, substituting for β_i in equation (41):

$$\sum_{i=1,2}\frac{1}{\theta_i}\int_{\eta_i}\int_{\varepsilon_i}\left(\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar\eta\right)^2\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \quad (47)$$

Also substituting for β_x shows that the expression under the integral sign of equation (47) is zero for any ε_i as long as the nonnegativity constraint does not bind for β_x, i.e. $\frac{\beta_p\alpha_I}{\alpha_B}\geq\beta_p$, which with $\alpha_B\leq\alpha_I$ is always true. Since the expression in equation (47) (a quadratic function) is nonnegative for any ε_i, achieving a value of zero for this expression at any ε_i maximizes the objective function of the principal for a given effort, and it implies that $y_i(\varepsilon_i)=y_i^*(\varepsilon_i)\ \forall\varepsilon_i$ (see equation (40)).

Using these results and equation (7), expected social output when $\alpha_B \leq \alpha_I$ is:

$$\begin{aligned}\mathbb{E}[\tilde{\eta}_i\tilde{y}_i] &= \mathbb{E}\left[\frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\tilde{\varepsilon}_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)\tilde{\eta}_i\right] \\ &= \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\bar{\eta}^2+\sigma_\eta^2\right) + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}^2\right) \\ &= \frac{\alpha_B}{2\theta_i}\left(\frac{\sigma_\eta^4}{\sigma_\eta^2+\sigma_\varepsilon^2}+\bar{\eta}^2\right) \qquad (48)\end{aligned}$$

Using similar arguments, it is still possible to reach the first-best outcome by setting $\beta_i = 0$ and $\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p} = 0$ if $\alpha_B \leq \alpha_I$. This upper bound was derived by setting $\beta_x = 0$ and $\beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(1+\frac{\sigma_z^2}{\sigma_x^2}\right)$ using equation (28).

Now consider the case $\alpha_B > \alpha_I$. Then the nonnegativity constraint on β_i does not bind, and the principal minimizes equation (46) while eliciting effort. This is equivalent to maximizing $\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p}$ subject to incentive compatibility, which yields $\beta_x = 0$. Thus, in this case, $\beta_i \geq 0$ in equation (45), i.e. the nonnegativity constraint on β_i does not bind. From equation (27):

$$\begin{aligned}y_i &= \frac{1}{2\theta_i}\left(\left(\frac{\beta_i}{\beta_p}+\alpha_I\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\varepsilon_i + \left(\frac{\alpha_M}{\beta_p}+\alpha_I\right)\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right) \\ &= \frac{1}{2\theta_i}\left(\left((\alpha_B-\alpha_I)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2}+\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)+\alpha_I\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\varepsilon_i + \alpha_I\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)\end{aligned}$$

So that, also using equation (7):

$$\begin{aligned}y_i(\varepsilon_i) - y_i^*(\varepsilon_i) &= \frac{1}{2\theta_i}\underbrace{(\alpha_B-\alpha_I)}_{>0}\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2}\varepsilon_i - \bar{\eta}\right) \\ \Rightarrow \quad \mathbb{E}[y_i(\tilde{\varepsilon}_i)] - \mathbb{E}[y_i^*(\tilde{\varepsilon}_i)] &= \frac{1}{2\theta_i}\underbrace{(\alpha_B-\alpha_I)}_{>0}\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2}-1\right)\bar{\eta} \qquad (49)\end{aligned}$$

■

Proof of Proposition 2: see Online Appendix.

Proof of Proposition 3:

A preliminary step is to establish that the average ESG score is a sufficient statistic: see section E.4. In particular, this implies that considering contracts based on the average score $\bar{m}_i$ on each dimension i of SEP rather than on each individual score is WLOG. The average score is defined as: $\bar{m}_i = y_i\bar{\varepsilon}_i$ with $\bar{\varepsilon}_i = \frac{1}{N}\sum_{j=1}^{N}\varepsilon_i^j$. Rewriting the manager's objective function in equation (26) accordingly gives:

$$\arg\max_{e,y_1,y_2} \mathbb{E}\left[w + \beta_x\left(e - \theta_1 y_1^2 - \theta_2 y_2^2\right) + \beta_p\tilde{p} + \sum_{i=1,2}\beta_i\bar{m}_i \middle| \bar{\varepsilon}_1, \bar{\varepsilon}_2\right] - C(e) \tag{50}$$

where the stock price $\tilde{p}$ is as in section E.4.

For $\beta_x + \beta_p > 0$, the manager's objective function is concave in y_i, for $i = 1, 2$. The FOC with respect to y_i is:

$$\beta_x\left(-2\theta_i y_i\right) + \beta_p\left(-2\theta_i y_i + \alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{\varepsilon}_i + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{\eta}\right)\right) + \beta_i\bar{\varepsilon}_i = 0$$

$$\begin{aligned}\Leftrightarrow \quad y_i \quad &= \quad \frac{\beta_i\bar{\varepsilon}_i + \beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)}{\beta_x + \beta_p}\frac{1}{2\theta_i} \\ &= \quad \frac{\left(\beta_i + \beta_p\alpha_I\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i + \beta_p\alpha_I\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}}{\beta_x + \beta_p}\frac{1}{2\theta_i}\end{aligned} \tag{51}$$

Given his contract, the manager will optimally exert high effort ($e = \overline{e}$) if and only if:

$$\begin{aligned} &\mathbb{E}\left[u(x,y,e)|e=\overline{e}\right] - c_e \geq \mathbb{E}\left[u(x,y,e)|e=\underline{e}\right] \\ \Leftrightarrow\quad & w + \beta_x\mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x|e=\overline{e}\right] + \beta_p\mathbb{E}\left[\mathbb{E}\left[\tilde{x}|t=1\right] + \alpha_I\mathbb{E}\left[\tilde{y}|t=1\right]\middle|e=\overline{e}\right] + \beta_1 y_1\bar{\varepsilon}_1 + \beta_2 y_2\bar{\varepsilon}_2 - c_e \\ & \geq w + \beta_x\mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x|e=\underline{e}\right] + \beta_p\mathbb{E}\left[\mathbb{E}\left[\tilde{x}|t=1\right] + \alpha_I\mathbb{E}\left[\tilde{y}|t=1\right]\middle|e=\underline{e}\right] + \beta_1 y_1\bar{\varepsilon}_1 + \beta_2 y_2\bar{\varepsilon}_2 \\ \Leftrightarrow\quad & \beta_x + \beta_p\frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} \geq \frac{c_e}{\overline{e}-\underline{e}} \end{aligned} \tag{52}$$

The fixed component of pay, w, is set to guarantee the manager's participation for a given $e \in \{\underline{e}, \overline{e}\}$:

$$\begin{aligned} \mathbb{E}[u(x,y,e)|e] \quad &= \quad \bar{W} \\ \Leftrightarrow \quad w + \beta_x\mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x|e\right] + \beta_p\mathbb{E}[\tilde{p}|e] + \sum_{i=1,2}\beta_i\mathbb{E}[\tilde{m}_i] \quad &= \quad \bar{W} + C(e) \end{aligned} \tag{53}$$

Thus, the board's objective function can be rewritten as:

$$\begin{aligned}\mathbb{E}[V(x,y,e)] &= \mathbb{E}\left[(1-\beta_x)\left(e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x\right)+\alpha_B\left(\tilde{\eta}_1 y_1+\tilde{\eta}_2 y_2+\tilde{\epsilon}_y\right)-\left(w+\beta_p\tilde{p}+\sum_{i=1,2}\beta_i\tilde{\bar{m}}_i\right)\right] \\ &= \mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x+\alpha_B\left(\tilde{\eta}_1 y_1+\tilde{\eta}_2 y_2+\tilde{\epsilon}_y\right)-\left(\bar{W}+C(e)\right)|e\right] \end{aligned} \quad (54)$$

Following the same steps as in the proof of Proposition 1, maximizing the board's objective function is equivalent to minimizing the expected quadratic distance between $y_i(\bar{\varepsilon}_i)$ and $y_i^*(\bar{\varepsilon}_i)$, where:

$$\mathbb{E}\left[\tilde{\eta}_i|\bar{\varepsilon}_i\right] = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta} \quad (55)$$

$$y_i^*(\bar{\varepsilon}_i) = \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right) \quad (56)$$

This problem is:

$$\begin{aligned} &\min_{\beta_x,\beta_p,\beta_i}\sum_{i=1,2}\int_{\eta_i}\int_{\bar{\varepsilon}_i}\left(\frac{1}{2}\frac{1}{\theta_i}\frac{\left(\beta_i+(\alpha_M+\beta_p\alpha_I)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i+(\alpha_M+\beta_p\alpha_I)\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}}{\beta_x+\beta_p}\right. \\ &\left.-\frac{\alpha_B+\alpha_M}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)\right)^2\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \\ \Leftrightarrow\quad &\min_{\beta_x,\beta_p,\beta_i}\sum_{i=1,2}\int_{\eta_i}\int_{\bar{\varepsilon}_i}\left(\left(\frac{\beta_i}{\beta_x+\beta_p}+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i\right. \\ &\left.+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)^2\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \end{aligned} \quad (57)$$

We have:

$$\int_{\eta_i}\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i=\int_{\eta_i}\left(\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i\varphi(\bar{\varepsilon}_i|\eta_i)d\bar{\varepsilon}_i\right)\phi(\eta_i)d\eta_i=\int_{\eta_i}\eta_i\phi(\eta_i)d\eta_i=\bar{\eta} \quad (58)$$

Moreover, $\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i^2\varphi(\bar{\varepsilon}_i|\eta_i)d\bar{\varepsilon}_i=\mathbb{E}[\tilde{\bar{\varepsilon}}_i^2|\eta_i]=var(\tilde{\bar{\varepsilon}}_i|\eta_i)+(\mathbb{E}[\tilde{\bar{\varepsilon}}_i|\eta_i])^2=\sigma_\varepsilon^2/N+\eta_i^2$. This implies:

$$\int_{\eta_i}\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i^2\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i=\int_{\eta_i}\left(\sigma_\varepsilon^2/N+\eta_i^2\right)\phi(\eta_i)d\eta_i=\sigma_\varepsilon^2/N+\int_{\eta_i}\eta_i^2\phi(\eta_i)d\eta_i=\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2 \quad (59)$$

The expression in equation (57) is globally convex with respect to β_i. The FOC w.r.t. β_i is:

$$\begin{aligned}
&\left(\frac{\beta_i}{\beta_x+\beta_p}+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\int_{\eta_i}\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i^2\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \\
&=\left(\alpha_B-\frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\int_{\eta_i}\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \\
\Leftrightarrow\quad &\frac{\beta_i}{\beta_x+\beta_p}=\left(\alpha_B-\frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2}+\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right) \qquad (60)
\end{aligned}$$

Substituting in equation (57) gives:

$$\min_{\beta_x,\beta_p}\left(\alpha_B-\frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)^2\frac{\sigma_\varepsilon^4}{(\sigma_\eta^2+\sigma_\varepsilon^2)^2/N}\sum_{i=1,2}\int_{\eta_i}\int_{\varepsilon_i}\left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2}\varepsilon_i-\bar{\eta}\right)^2\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \quad (61)$$

The expression under the integral sign is positive and independent from the contract. The expression in equation (61) is minimized by minimizing $\left(\alpha_B-\frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)^2$. With a nonnegativity constraint on β_x, this is achieved by setting $\beta_x=\max\{\frac{\beta_p\alpha_I}{\alpha_B}-\beta_p,0\}$. As above, for $\alpha_B\leq\alpha_I$, we have $\beta_p>0$, $\beta_x\geq 0$, β_i as defined as in equation (60) is equal to zero, and investment in dimension i of SEP is:

$$y_i(\bar{\varepsilon}_i)\;=\;\frac{\beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\tilde{\bar{\varepsilon}}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)}{\beta_x+\beta_p}\frac{1}{2\theta_i}=\frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\tilde{\bar{\varepsilon}}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right),$$

which is the same as y_i^* as defined in equation (56). Moreover, expected social output when $\alpha_B\leq\alpha_I$ is:

$$\mathbb{E}[\tilde{\eta}_i\tilde{y}_i]=\frac{\beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\mathbb{E}[\tilde{\eta}_i\tilde{\bar{\varepsilon}}_i]+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\mathbb{E}[\tilde{\eta}_i]\right)}{\beta_x+\beta_p}\frac{1}{2\theta_i}=\frac{\alpha_B}{2\theta_i}\left(\frac{\sigma_\eta^4}{\sigma_\eta^2+\sigma_\varepsilon^2/N}+\bar{\eta}^2\right) \quad (62)$$

In the limit:

$$\lim_{N\to\infty}\mathbb{E}[\tilde{\eta}_i\tilde{y}_i]\;=\;\frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2+\sigma_\eta^2\right)$$

For $\alpha_B>\alpha_I$, we have $\beta_x=0$, and to elicit high effort, use equation (52) to set:

$$\beta_p\frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2}=\frac{c_e}{\bar{e}-\underline{e}}\quad\Leftrightarrow\quad\beta_p=\frac{c_e}{\bar{e}-\underline{e}}\left(1+\frac{\sigma_z^2}{\sigma_x^2}\right) \quad (63)$$

In this case, $\beta_i > 0$ and is defined as in equation (60). Substituting for β_x and β_p in equation (60), in this case we have:

$$\beta_i = (\alpha_B - \alpha_I)\frac{c_e}{\bar{e} - \underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\right) \tag{64}$$

In the case $\alpha_B > \alpha_I$, substituting for β_i in equation (51):

$$y_i(\bar{\varepsilon}_i) = \frac{1}{2\theta_i}\frac{(\alpha_B - \alpha_I)\frac{c_e}{\bar{e}-\underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i + \beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)}{\beta_x + \beta_p}$$

Substituting for β_x and β_p, investment in dimension i of SEP is:

$$y_i(\bar{\varepsilon}_i) = \frac{1}{2\theta_i}\Bigg((\alpha_B - \alpha_I)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i + \alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{\varepsilon}_i + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{\eta}\right)\Bigg)$$

In the limit: $\lim_{N\to\infty}(y_i(\bar{\varepsilon}_i) - y_i^*(\bar{\varepsilon}_i)) = 0$, as defined in equation (56). Moreover, expected social output when $\alpha_B > \alpha_I$ is:

$$\begin{aligned}
\mathbb{E}[\tilde{\eta}_i\tilde{y}_i] &= \frac{1}{2\theta_i}\Bigg((\alpha_B - \alpha_I)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\right)\mathbb{E}[\tilde{\eta}_i\tilde{\bar{\varepsilon}}_i] \\
&\quad + \alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\mathbb{E}[\tilde{\eta}_i\tilde{\bar{\varepsilon}}_i] + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{\eta}\mathbb{E}[\tilde{\eta}_i]\right)\Bigg) \\
&= \frac{1}{2\theta_i}\Bigg((\alpha_B - \alpha_I)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\right)\left(\bar{\eta}^2 + \sigma_\eta^2\right) \\
&\quad + \alpha_I\left(\bar{\eta}^2 + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\sigma_\eta^2\right)\Bigg)
\end{aligned}$$

In the limit:

$$\lim_{N\to\infty}\mathbb{E}[\tilde{\eta}_i\tilde{y}_i] = \frac{1}{2\theta_i}\left((\alpha_B - \alpha_I)\left(\bar{\eta}^2 + \sigma_\eta^2\right) + \alpha_I\left(\bar{\eta}^2 + \sigma_\eta^2\right)\right) = \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sigma_\eta^2\right)$$

■

Proof of Proposition 4: see Online Appendix.

E Stock price

This section derives the stock price in different settings. It also analyzes the portfolio optimization problem of socially conscious investors with "warm-glow" preferences which can be used to microfound the stock price function postulated in equation (4).

Assume that investors can invest their wealth either at the riskfree rate, which is zero for simplicity, or in the firm's stock. They are risk neutral, and they have preferences over the firm's cash flows and its social output: they assign a weight of $\alpha_I \geq 0$ to the firm's social output relative to its cash flows.[25] As further discussed in section C below, this can be for several reasons: committing to being socially responsible can be useful for hiring or funding purposes; the board might internalize the future penalties that the firm might incur for the externalities that it generates; or it might simply act in the best interests of shareholders who are themselves socially responsible. They trade at $t = 1$ after observing the financial report z, social investments $\{y_1, y_2\}$, and SEP measures. The firm's stock price is set by market clearing.

E.1 Stock price determination

First, we consider stock price determination when investors observe the firm's social investments, $\{y_1, y_2\}$, and the financial report z. Note that all investors have the same information and therefore do not learn from the stock price. A risk neutral price-taking investor chooses the quantity q of stock to buy to maximize:

$$\mathbb{E}\left[\tilde{x} + \alpha_I \tilde{y} - qp | z\right] = q\left(\mathbb{E}\left[\tilde{x}|z\right] + \alpha_I \mathbb{E}\left[\tilde{y}|z\right]\right) - qp \quad (65)$$

where the third equality relies on the normal distribution assumption. For $v \in \{x, y\}$, define:

$$\mu_{v|\omega} \equiv \mathbb{E}\left[\tilde{v}|t = 1\right] \quad (66)$$

[25]This is the simplest specification of investors' social preferences. If investors have heterogeneous social preferences, the parameter α_I is the social weight of the marginal investor. The marginal investor, who is indifferent between buying the stock or not, is such that the stock market clears (i.e. total investor demand equals supply), see Bucourt and Inostroza (2023). Note that, in this type of model, there must be constraints on portfolio choice for the stock market to clear, for example a short-selling constraint and a borrowing constraint.

where ω denotes the set of performance measures observable at $t = 1$. Maximizing the expression in equation (65) with respect to q shows that the only stock price compatible with market clearing for the firm stock is:

$$p = \mathbb{E}\left[\tilde{x}|z\right] + \alpha_I \mathbb{E}\left[\tilde{y}|z\right] = \mu_{x|z} + \alpha_I \mu_{y|z} \tag{67}$$

where $\mu_{x|z}$ and $\mu_{y|z}$ are respectively as in equations (72) and (73) below. The stock price p is linear in the report z: for two constants a and b, we can write $p = a + b \times z$

Second, we consider stock price determination when investors observe measures m_1 and m_2 of the firm's SEP (a similar reasoning applies when investors observe a set of measures on each dimension of SEP instead; for brevity, we only refer to the baseline case). Let $m \equiv \{m_1, m_2\}$. As above, when all investors have the same information and therefore do not learn from the stock price, the only market clearing stock price is:

$$p = \mu_{x|z,m} + \alpha_I \mu_{y|z,m} \tag{68}$$

where, with one set of ESG scores, $\mu_{x|z,m}$ and $\mu_{y|z,m}$ are described respectively in equations (74) and (75). The cases with additional sets of ESG scores are analyzed in sections (E.4) and (E.5).

The following subsections analyze how investors update their beliefs after observing the signal z and ESG scores, depending on the availability of ESG scores.

E.2 No ESG scores

The manager is compensated with a fixed wage and stock price-based compensation, with a linear contract with a sensitivity β_p of the manager's compensation to the stock price. Investors believe that the manager exerts some effort $\hat{e}$. Consider the perspective of the manager at the time ($t = 0$) of choosing y_1 and y_2 given his knowledge of $\{\eta_1, \eta_2\}$:

$$\begin{aligned} \mathbb{E}[u(x, y, e)] &= \mathbb{E}\left[w + \beta_p \tilde{p} - C(e)\right] \\ &= \beta_p \left(\overline{e} - \theta_1 y_1^2 - \theta_2 y_2^2 + \alpha_I \left(\mathbb{E}[\tilde{\eta}_1|I] y_1 + \mathbb{E}[\tilde{\eta}_2|I] y_2\right)\right) + w - C(e) \end{aligned} \tag{69}$$

where we used $\mathbb{E}[\tilde{z}] = \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2$ since investors observe $\{y_1, y_2\}$, and $\mathbb{E}[\hat{e}|e = \overline{e}] = \overline{e}$. Optimizing from equation (69) with respect to y_i taking the manager's contract as given, the manager will choose y_i such that:

$$\beta_p \left(-2\theta_1 y_i + \alpha_I \mathbb{E}[\tilde{\eta}_i|I]\right) = 0 \quad \Leftrightarrow \quad y_i = \alpha_I \frac{\mathbb{E}[\tilde{\eta}_i|I]}{2\theta_i} \tag{70}$$

Without ESG scores, investors only observe y_i on each dimension i of SEP whereas the manager knows ε_i and might use this knowledge when choosing y_i. Accordingly, let investors update their beliefs such that $\mathbb{E}[\tilde{\eta}_i|I] = \mathbb{E}[\tilde{\eta}_i|y_i] = f(y_i)$ for a given function f. Equation (70) then gives: $y_i = \frac{\alpha_I}{2} \frac{f(y_i)}{\theta_i}$. This shows that, for any f, the optimal y_i chosen by the manager does not depend on ε_i. Thus, y_i is independent from ε_i in equilibrium, i.e. $\mathbb{E}[\tilde{\eta}_i|y_i]$ is a constant which does not depend on the manager's actions. This implies: $\mathbb{E}[\tilde{\eta}_i|y_i] = \mathbb{E}[\tilde{\eta}_i] = \bar{\eta}$, so that, using equation (70): $y_i = \frac{\alpha_I}{2} \frac{\bar{\eta}}{\theta_i}$. Thus, the stock price is defined by:

$$\begin{aligned}
\mu_x &= \hat{e} - \mathbb{E}[\theta_1 \tilde{y}_1^2 + \theta_2 \tilde{y}_2^2] && (71) \\
\mu_{x|z} &= \mathbb{E}[\tilde{x}] + \frac{cov(\tilde{z}, \tilde{x})}{var(\tilde{z})} \left(z - \mathbb{E}[\tilde{z}]\right) \\
&= \hat{e} - \theta_1 y_1^2 - \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) && (72) \\
\mu_{y|z} &= \mathbb{E}\left[\tilde{\eta}_1 \tilde{y}_1 | y_1\right] + \mathbb{E}\left[\tilde{\eta}_2 \tilde{y}_2 | y_2\right] + \mathbb{E}\left[\tilde{\epsilon}_y\right] = \bar{\eta} y_1 + \bar{\eta} y_2 && (73)
\end{aligned}$$

Thus, without ESG scores, a manager with stock price-based incentives does not implement social investments contingent on ε_i. The reason is that the level of social investment on dimension i that is optimal from the manager's perspective only depends on investors' beliefs about $\tilde{\eta}_i$ and on the cost of social and environmental investments, θ_i. Regardless of how investors update their beliefs when they observe y_i, this level of social investment is independent from ε_i.

E.3 One set of ESG scores

We consider the equilibrium in which the board delegates the social investment decision to the firm's manager. Investors believe that the manager exerts some effort $\hat{e}$. With one set of ESG scores, investors update their beliefs about the firm's technology for social output after observing

ESG scores as follows:

$$\begin{aligned}
\mu_{x|z,m} &= \mathbb{E}[\tilde{x}] + \frac{cov(\tilde{z},\tilde{x})}{var(\tilde{z})}\left(z - \mathbb{E}[\tilde{z}]\right) \\
&= \hat{e} - \theta_1 y_1^2 + \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2}\left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) \qquad (74)\\
\mu_{y|z,m} &= \mathbb{E}\left[\tilde{\eta}_1\tilde{y}_1 + \tilde{\eta}_2\tilde{y}_2 + \tilde{\epsilon}_y | z, m\right] \\
&= y_1 \mathbb{E}\left[\tilde{\eta}_1 | m_1, y_1\right] + y_2 \mathbb{E}\left[\tilde{\eta}_2 | m_2, y_2\right] \qquad (75)
\end{aligned}$$

where:

$$\mathbb{E}\left[\tilde{\eta}_i | m_i, y_i\right] = \mathbb{E}\left[\tilde{\eta}_i | \varepsilon_i\right] = \mathbb{E}\left[\tilde{\eta}_i\right] + \frac{cov\left(\tilde{\eta}_i, \tilde{\varepsilon}_i\right)}{var\left(\tilde{\varepsilon}_i\right)}\left(\varepsilon_i - \mathbb{E}[\tilde{\varepsilon}_i]\right) = \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\varepsilon_i - \bar{\eta}\right) \qquad (76)$$

Substituting into equation (75):

$$\begin{aligned}
\mu_{y|z,m} &= y_1\left(\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\varepsilon_1 - \bar{\eta}\right)\right) + y_2\left(\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\varepsilon_2 - \bar{\eta}\right)\right) \\
&= \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} m_1 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} y_1\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} m_2 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} y_2\bar{\eta} \qquad (77)
\end{aligned}$$

E.4 N sets of ESG scores with i.i.d. noise terms

As above, we consider the equilibrium in which the board delegates the social investment decision to the firm's manager. As in the baseline model, the distribution of ESG scores j on dimension i is centered on $\eta_i y_i$ such that $\tilde{m}_i^j \sim \mathcal{N}(\eta_i y_i, \sigma_\varepsilon^2)$. Denoting $\tilde{\zeta}_i^j \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, for any score j on dimension i we can decompose the score as $\tilde{m}_i^j = y_i(\eta_i + \tilde{\zeta}_i^j)$, where the superscript j indicates that the noise term $\tilde{\zeta}_i^j$ is different for each rating j. With N ESG scores on each ESG dimension i whose noise terms are i.i.d., the average score is a sufficient statistic for the mean.[26] Define:

$$\bar{m}_i \equiv \frac{1}{N}\sum_{j=1}^{N} m_i^j \qquad (78)$$

[26]This is a standard application of the factorization theorem. For a source, see https://www.math.arizona.edu/∼tgk/466/sufficient.pdf.

We have:

$$\frac{1}{N}\sum_{j=1}^{N}\tilde{m}_i^j = \frac{1}{N}\sum_{j=1}^{N} y_i\left(\eta_i + \tilde{\zeta}_i^j\right) = \eta_i y_i + \frac{y_i}{N}\sum_{j=1}^{N}\tilde{\zeta}_i^j \tag{79}$$

where the random variable $\frac{1}{N}\sum_{j=1}^{N}\tilde{\zeta}_i^j$ is normally distributed with the following mean and variance:

$$\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N}\tilde{\zeta}_i^j\right] = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left[\tilde{\zeta}_i^j\right] = 0 \tag{80}$$

$$var\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{\zeta}_i^j\right) = \frac{1}{N^2}\sum_{j=1}^{N} var\left(\tilde{\zeta}_i^j\right) = \frac{1}{N}\sigma_\varepsilon^2 \tag{81}$$

where we used the i.i.d. assumption about the noise terms. Therefore, the unconditional variance of $\tilde{\bar{m}}_i/y_i$ is:

$$var\left(\tilde{\bar{m}}_i/y_i\right) = \sigma_\eta + \frac{1}{N}\sigma_\varepsilon^2 \tag{82}$$

Investors update their beliefs about the firm's technology for social output after observing ESG scores as follows:

$$\begin{aligned}
\mu_{x|z,\bar{m}} &= \mathbb{E}[\tilde{x}] + \frac{cov(\tilde{z},\tilde{x})}{var(\tilde{z})}\left(z - \mathbb{E}[\tilde{z}]\right) \\
&= \hat{e} - \theta_1 y_1^2 + \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2}\left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) && (83) \\
\mu_{y|z,\bar{m}} &= \mathbb{E}\left[\tilde{\eta}_1\tilde{y}_1 + \tilde{\eta}_2\tilde{y}_2 + \tilde{\epsilon}_y | z, \bar{m}\right] = y_1\mathbb{E}\left[\tilde{\eta}_1|\bar{m}_1, y_1\right] + y_2\mathbb{E}\left[\tilde{\eta}_2|\bar{m}_2, y_2\right] && (84)
\end{aligned}$$

Moreover, for any $i \in \{1, 2\}$:

$$\mathbb{E}\left[\tilde{\eta}_i|\bar{m}_i\right] = \mathbb{E}\left[\tilde{\eta}_i\right] + \frac{cov\left(\tilde{\eta}_i, \tilde{\bar{m}}_i/y_i\right)}{var\left(\tilde{\bar{m}}_i/y_i\right)}\left(\bar{m}_i/y_i - \mathbb{E}[\tilde{\bar{m}}_i/y_i]\right) = \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta + \sigma_\varepsilon^2/N}\left(\bar{m}_i/y_i - \bar{\eta}\right) \tag{85}$$

Substituting into equation (84):

$$\begin{aligned}
\mu_{y|z,\bar{m}} &= y_1\left(\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\left(\bar{m}_1/y_1 - \bar{\eta}\right)\right) + y_2\left(\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\left(\bar{m}_2/y_2 - \bar{\eta}\right)\right) \\
&= \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{m}_1 + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}y_1\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{m}_2 + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}y_2\bar{\eta} && (86)
\end{aligned}$$

E.5 Two sets of ESG scores with correlated noise terms

As above, we consider the equilibrium in which the board delegates the social investment decision to the firm's manager. Assume that there are two ESG scores on each SEP dimension i with the following distributions conditional on η_i: $\tilde{\varepsilon}_i^1 \sim \mathcal{N}(\eta_i, \sigma_{\varepsilon_i}^{1^2})$, $\tilde{\varepsilon}_i^2 \sim \mathcal{N}(\eta_i, \sigma_{\varepsilon_i}^{2^2})$. As above, let $\tilde{\zeta}_i^j \sim \mathcal{N}(0, \sigma_{\varepsilon_i}^{j^2})$; for any score j on dimension i, we can decompose the score as $\tilde{m}_i^j = y_i(\eta_i + \tilde{\zeta}_i^j)$ or equivalently as $\tilde{\varepsilon}_i^j = \eta_i + \tilde{\zeta}_i^j$, where the superscript j indicates that the noise term $\tilde{\zeta}_i^j$ is different for each score j. Moreover, let the correlation coefficient between $\tilde{\zeta}_i^1$ and $\tilde{\zeta}_i^2$ be denoted by $\rho \in (-1, 1)$. The unconditional distribution of $\tilde{\varepsilon}_i^j$ is given by $\tilde{\varepsilon}_i^j = \tilde{\eta}_i + \tilde{\zeta}_i^j$.

Bayesian updating by investors is similar to equations (83)-(84), albeit with changes in notations that reflect changes in the relevant information set (now investors observe $\{\varepsilon_i^1, \varepsilon_i^2\}$ on each dimension i, as opposed to $\bar{m}_i$). The important change in this subsection is the conditional distribution of $\tilde{\eta}_i$ after the observations of $\{\varepsilon_i^1, \varepsilon_i^2\}$. To alleviate notations in the calculations that follow, we henceforth omit the subscript i and denote the latter set of variables as $\{\varepsilon_1, \varepsilon_2\}$. The variables $\{\tilde{\eta}, \tilde{\varepsilon}_1, \tilde{\varepsilon}_2\}$ follow a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ such that:[27]

$$\boldsymbol{\mu} = \begin{pmatrix} \mathbb{E}[\tilde{\eta}] \\ \mathbb{E}[\tilde{\varepsilon}_1] \\ \mathbb{E}[\tilde{\varepsilon}_2] \end{pmatrix} = \begin{pmatrix} \bar{\eta} \\ \bar{\eta} \\ \bar{\eta} \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \tag{87}$$

where

$$\boldsymbol{\Sigma}_{11} = var\left(\tilde{\eta}\right), \quad \boldsymbol{\Sigma}_{12} = \begin{pmatrix} cov\left(\tilde{\eta}, \tilde{\varepsilon}_1\right) & cov\left(\tilde{\eta}, \tilde{\varepsilon}_2\right) \end{pmatrix}, \quad \boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T, \quad \boldsymbol{\Sigma}_{22} = \begin{pmatrix} var\left(\tilde{\varepsilon}_1\right) & cov\left(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2\right) \\ cov\left(\tilde{\varepsilon}_2, \tilde{\varepsilon}_1\right) & var\left(\tilde{\varepsilon}_2\right) \end{pmatrix}$$

$$\Rightarrow \qquad \boldsymbol{\Sigma}_{11} = \sigma_\eta^2, \quad \boldsymbol{\Sigma}_{12} = \begin{pmatrix} \sigma_\eta^2 & \sigma_\eta^2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T, \quad \boldsymbol{\Sigma}_{22} = \begin{pmatrix} \sigma_\eta^2 + \sigma_{\varepsilon_1}^2 & \sigma_\eta^2 + \rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2} \\ \sigma_\eta^2 + \rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2} & \sigma_\eta^2 + \sigma_{\varepsilon_2}^2 \end{pmatrix}$$

The posterior distribution of $\tilde{\eta}$ after observing $\{\varepsilon_1, \varepsilon_2\}$ is normal with mean:

[27] A source for the following formulas can be found at: https://online.stat.psu.edu/stat505/lesson/6/6.1.

$$
\begin{aligned}
\mathbb{E}[\tilde{\eta}|\varepsilon_1,\varepsilon_2] &= \bar{\eta} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\begin{pmatrix}\varepsilon_1 - \bar{\eta}\\ \varepsilon_2 - \bar{\eta}\end{pmatrix}\\
&= \bar{\eta} + \begin{pmatrix}\sigma_\eta^2 & \sigma_\eta^2\end{pmatrix}\begin{pmatrix}\frac{\sigma_\eta^2+\sigma_{\varepsilon_2}^2}{\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2-\rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2+\sigma_\eta^2(\sigma_{\varepsilon_1}^2+\sigma_{\varepsilon_2}^2-2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})} & \frac{-\sigma_\eta^2-\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}}{\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2-\rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2+\sigma_\eta^2(\sigma_{\varepsilon_1}^2+\sigma_{\varepsilon_2}^2-2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})}\\ \frac{-\sigma_\eta^2-\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}}{\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2-\rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2+\sigma_\eta^2(\sigma_{\varepsilon_1}^2+\sigma_{\varepsilon_2}^2-2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})} & \frac{\sigma_\eta^2+\sigma_{\varepsilon_1}^2}{\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2-\rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2+\sigma_\eta^2(\sigma_{\varepsilon_1}^2+\sigma_{\varepsilon_2}^2-2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})}\end{pmatrix}\begin{pmatrix}\varepsilon_1 - \bar{\eta}\\ \varepsilon_2 - \bar{\eta}\end{pmatrix}\\
&= \bar{\eta} + \frac{\sigma_\eta^2(\sigma_{\varepsilon_2}^2 - \rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})}{\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 - \rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 + \sigma_\eta^2(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 - 2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})}(\varepsilon_1 - \bar{\eta})\\
&\quad + \frac{\sigma_\eta^2(\sigma_{\varepsilon_1}^2 - \rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})}{\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 - \rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 + \sigma_\eta^2(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 - 2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2})}(\varepsilon_2 - \bar{\eta}) \qquad (88)
\end{aligned}
$$

and variance:

$$
\begin{aligned}
\sigma_{\eta|\varepsilon_1,\varepsilon_2}^2 &= \sigma_\eta^2 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\\
&= \sigma_\eta^2 - \frac{\sigma_\eta^4\left(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 - 2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}\right)}{\sigma_\eta^2\sigma_{\varepsilon_1}^2 + \sigma_\eta^2\sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 - \rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 - 2\rho\sigma_\eta^2\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}} \qquad (89)
\end{aligned}
$$

Online Appendix
Executive Compensation with Social and Environmental Performance

1 Additional proofs

Proof of Proposition 2:

The objective function of the board now rewrites as:

$$\mathbb{E}[V(x,y,e)] \;=\; \mathbb{E}\left[\tilde{x} + \sum_{i=1,2} \alpha_B^i \tilde{\eta}_i \tilde{y}_i - \left(w + \beta_x \tilde{x} + \beta_p \tilde{p} + \sum_{i=1,2} \beta_i \tilde{m}_i\right)\right] \tag{90}$$

The first part of the proof involves straightforward notational adjustments to the proof of Proposition 1. Equation (41) rewrites as:

$$\min_{\beta_x,\beta_p,\beta_i} \frac{1}{\theta_i} \sum_{i=1,2} \int_{\eta_i} \int_{\varepsilon_i} \left(\left(\frac{\beta_i}{\beta_x+\beta_p} + \left(\frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} - \alpha_B^i \right) \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \varepsilon_i \right.$$
$$\left. + \left(\frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} - \alpha_B^i \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \tag{91}$$

This expression is globally convex with respect to β_i. As above, the FOC w.r.t. β_i is:

$$\int_{\eta_i} \int_{\varepsilon_i} \frac{2}{\beta_x+\beta_p} \varepsilon_i \left(\left(\frac{\beta_i}{\beta_x+\beta_p} + \left(\frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} - \alpha_B^i \right) \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \varepsilon_i \right.$$
$$\left. + \left(\frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} - \alpha_B^i \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \bar{\eta} \right) \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i = 0$$
$$\Leftrightarrow \quad \frac{\beta_i}{\beta_x+\beta_p} = \left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \left(\alpha_B^i - \frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} \right) \tag{92}$$

Given the nonnegativity constraint and the global convexity of the objective function with respect to β_i, there are two cases. If equation (92) gives a positive β_i given the optimal values of β_x and β_p derived below, then the optimum is $\beta_i^* = \beta_i$ as in equation (92). If equation (92) gives a nonpositive β_i, then the optimum is $\beta_i^* = 0$.

Define:

$$\gamma_i \;\equiv\; \int_{\eta_i} \int_{\varepsilon_i} \left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \tag{93}$$

$$g_i \;\equiv\; \frac{\sigma_\varepsilon^4}{(\sigma_\eta^2+\sigma_\varepsilon^2)^2} \int_{\eta_i} \int_{\varepsilon_i} \left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} \varepsilon_i - \bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i, \tag{94}$$

Both γ_i and g_i are strictly positive. Since the distribution of $\tilde{\eta}_i$ is independent from i, and the distribution of $\tilde{\varepsilon}_i$ conditional on η_i is independent from i, the variables γ_i and g_i are independent from i and independent from the contract. Let $\Gamma_i = \gamma_i$ if the nonnegativity constraint on β_i is binding, and $\Gamma_i = g_i$ if the nonnegativity constraint on β_i is nonbinding.

There are three cases.

Case 1. We conjecture that $\beta_i > 0$ for $i = 1, 2$, so that $\Gamma_i = g_i$ for $i = 1, 2$. Substituting for β_i in equation (91), the optimization problem is:

$$\min_{\beta_x,\beta_p} \sum_{i=1,2} \Gamma_i \left(\frac{\beta_p}{\beta_x + \beta_p} \alpha_I^i - \alpha_B^i \right)^2 \tag{95}$$

where $\Gamma_i = g_i$ in case 1 (we henceforth keep the Γ_i notation because we will refer to equations below for cases other than case 1). This is a sum weighted by Γ_i of the quadratic distance between the board's preference for dimension i of SEP, and a fraction $\frac{\beta_p}{\beta_x+\beta_p}$ of investors' preference for same dimension of SEP. For a given β_p, this is equivalent to choosing:

$$\min_{\beta} \sum_{i=1,2} \Gamma_i \left(\alpha_B^i - \beta \alpha_I^i \right)^2 \quad \text{s.t.} \quad \beta \leq 1 \tag{96}$$

Denote by $\delta \geq 0$ the Lagrange multiplier associated with the constraint. The Lagrangian is:

$$\mathcal{L} = \sum_{i=1,2} \Gamma_i \left(\alpha_B^i - \beta \alpha_I^i \right)^2 + \delta(\beta - 1) \tag{97}$$

The FOC with respect to β is:

$$-2 \sum_{i=1,2} \Gamma_i \alpha_I^i \left(\alpha_B^i - \beta \alpha_I^i \right) + \delta = 0 \quad \Leftrightarrow \quad \beta = \frac{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i - \delta}{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2} \tag{98}$$

where $\delta > 0$ if and only if:

$$\sum_{i=1,2} \Gamma_i \alpha_I^i \left(\alpha_B^i - \alpha_I^i \right) > 0 \tag{99}$$

For $\delta > 0$, because of the complementary slackness condition we have $\beta = 1 \Leftrightarrow \frac{\beta_p}{\beta_x+\beta_p} = 1 \Leftrightarrow \beta_x = 0$. The value of β_p is determined according to incentive compatibility with respect to the manager's effort. Substituting in equation (28):

$$\beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} = \frac{c_e}{\overline{e} - \underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\overline{e} - \underline{e}} \left(1 + \frac{\sigma_z^2}{\sigma_x^2} \right) \tag{100}$$

For $\delta = 0$, the FOC rewrites as:

$$\sum_{i=1,2} \Gamma_i \alpha_I^i \left(\alpha_B^i - \beta \alpha_I^i\right) = 0 \quad \Leftrightarrow \quad \beta = \frac{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i}{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2} \quad \Leftrightarrow \quad \beta_x = \beta_p \left(\frac{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2}{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i} - 1 \right) \tag{101}$$

This equation gives the optimal value of β_x as long as this value is nonnegative, so that the nonnegativity constraint is nonbinding (i.e. $\delta = 0$). The value of β_p is determined according to incentive compatibility with respect to the manager's effort. Substituting in equation (28):

$$\begin{aligned} &\beta_p \left(\frac{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2}{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i} - 1 \right) + \beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} = \frac{c_e}{\bar{e} - \underline{e}} \\ \Leftrightarrow \quad &\beta_p = \frac{c_e}{\bar{e} - \underline{e}} \Big/ \left(\frac{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2}{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i} - 1 + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \right) \end{aligned} \tag{102}$$

Finally, we verify the conjecture that $\beta_i > 0$ for $i = 1, 2$, by plugging β_x and β_p thus derived into equation (92). If the conjecture is verified, the algorithm stops. Otherwise we move to case 2 below.

Case 2. We conjecture that $\beta_1 > 0$ and $\beta_2 = 0$ if $\frac{\alpha_B^1}{\alpha_I^1} > \frac{\alpha_B^2}{\alpha_I^2}$, in which case $\Gamma_1 = g_1$ and $\Gamma_2 = \gamma_2$, and $\beta_1 = 0$ and $\beta_2 > 0$ otherwise, in which case $\Gamma_1 = \gamma_1$ and $\Gamma_2 = g_2$. The proof follows the same steps as in case 1 above except for the different values of Γ_i. If the conjecture is verified, the algorithm stops. Otherwise we move to case 3 below.

Case 3. We conjecture that $\beta_i = 0$ for $i = 1, 2$, so that $\Gamma_i = \gamma_i$ for $i = 1, 2$. The proof follows the same steps as in case 1 above except for the different values of Γ_i.

We now rely on this algorithm to establish the four points of Proposition 2.

(i) If $\alpha_B^i \geq \alpha_I^i$ for $i = 1, 2$, then β_i is as in equation (92) such that $\beta_i \geq 0$ for $i = 1, 2$ even without the nonnegativity constraint, i.e. this constraint does not bind. As a result, we have $\Gamma_i = g_i$ for $i = 1, 2$ and case 1 as described above is relevant. Moreover, with $\alpha_B^i \geq \alpha_I^i$ for $i = 1, 2$, the inequality in equation (99) holds, so that $\beta_x = 0$, and β_p is as in equation (100).

(ii) If $\alpha_B^i < \alpha_I^i$ for $i = 1, 2$, the inequality in equation (99) does not hold, so that $\beta_x > 0$, and β_p is as in equation (102). Given these values of β_x and β_p, β_i is as in equation (92) if the expression is positive, and zero otherwise.

(iii) If $\alpha_I^i = \alpha_I$ for $i = 1, 2$, $\alpha_B^1 < \alpha_I \leq \alpha_B^2$ with $\frac{\alpha_I}{\alpha_B^1}$ sufficiently large and $\frac{\alpha_B^2}{\alpha_I}$ sufficiently close to 1, the inequality in equation (99) does not hold:

$$\sum_{k=1,2} \Gamma_k \alpha_I^k \left(\alpha_B^k - \alpha_I^k\right) = \alpha_I \sum_{k=1,2} \Gamma_k \left(\alpha_B^k - \alpha_I\right) < 0,$$

so that $\beta_x > 0$, and β_p is as in equation (102). With $\alpha_B^1 < \alpha_I \leq \alpha_B^2$ such that $\frac{\alpha_I}{\alpha_B^1}$ is sufficiently large and $\frac{\alpha_B^2}{\alpha_I}$ is sufficiently close to 1, we have $\frac{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2}{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i} > 1$, i.e. $\beta_p > 0$. With $\alpha_B^2 \geq \alpha_I$ and

$\beta_x > 0$, which implies $\beta < 1$, we have $\beta_2 > 0$. Using equation (101):

$$\alpha_B^i - \frac{\beta_p}{\beta_x + \beta_p}\alpha_I^i = \alpha_B^i - \beta\alpha_I = \alpha_B^i - \frac{\sum_{k=1,2}\Gamma_k \alpha_I^2 \alpha_B^k}{\sum_{k=1,2}\Gamma_k \alpha_I^2} = \alpha_B^i - \frac{\sum_{k=1,2}\Gamma_k \alpha_B^k}{\sum_{k=1,2}\Gamma_k} \tag{103}$$

By contradiction, suppose that $\beta_1 > 0$ so that $\Gamma_1 = \Gamma_2 \equiv \Gamma$. Then, substituting in equation (103):

$$\alpha_B^1 - \frac{\beta_p}{\beta_x + \beta_p}\alpha_I^1 = \alpha_B^1 - \frac{\sum_{k=1,2}\Gamma \alpha_B^k}{\sum_{k=1,2}\Gamma} = \alpha_B^1 - \frac{\alpha_B^1 + \alpha_B^2}{2} = \frac{\alpha_B^1 - \alpha_B^2}{2} < 0,$$

so that from equation (92) we would have $\beta_1 < 0$, a contradiction. Thus, $\beta_1 = 0$.

(iv) If α_B^i is sufficiently large and $\alpha_I^i > 0$, then β_i is as in equation (92) such that $\beta_i > 0$ even without the nonnegativity constraint, i.e. this constraint does not bind for β_i, and we have $\Gamma_i = g_i$. Moreover, when α_B^i is sufficiently large and $\alpha_I^i > 0$, the inequality in equation (99) holds:

$$\sum_{k=1,2}\Gamma_k \alpha_I^k \left(\alpha_B^k - \alpha_I^k\right) > g_i \alpha_I^i \left(\alpha_B^i - \alpha_I^i\right) - \Gamma_j {\alpha_I^j}^2 > 0$$

Thus, $\beta_x = 0$, and β_p is as in equation (100). ■

Proof of Proposition 4:

Rewriting the manager's objective function in the case with two sets of ESG scores gives:

$$\arg\max_{e,y_1,y_2} \mathbb{E}\left[w + \beta_x\left(e - \theta_1 y_1^2 - \theta_2 y_2^2\right) + \beta_p \tilde{p} + \sum_{i=1,2}\sum_{k=1,2}\beta_i^k m_i^k\right] - C(e) \tag{104}$$

where the stock price $\tilde{p}$ is as in section E.5.

For $\beta_x + \beta_p > 0$, the manager's objective function is concave in y_i, for $i = 1, 2$. The FOC with respect to y_i is (for $j \neq k$):

$$\beta_x\left(-2\theta_i y_i\right) + -2\beta_p \theta_i y_i + \alpha_I \beta_p \left(\bar{\eta} + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right)\right) + \sum_{k=1,2}\beta_i^k \varepsilon_i^k = 0$$

$$\Leftrightarrow \quad y_i = \frac{1}{2\theta_i}\frac{\sum_{k=1,2}\beta_i^k\varepsilon_i^k + \alpha_I\beta_p\left(\bar{\eta} + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right)\right)}{\beta_x + \beta_p} \tag{105}$$

Given his contract, the manager will optimally exert high effort ($e = \bar{e}$) if and only if equation (52) is satisfied. The fixed component of pay, w, is set to guarantee the manager's participation for a given $e \in \{\underline{e}, \overline{e}\}$:

$$\begin{aligned}
\mathbb{E}[u(x,y,e)|e] &= \bar{W} \\
\Leftrightarrow \quad w + \beta_x \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x | e\right] + \beta_p \mathbb{E}[\tilde{p}|e] + \sum_{i=1,2}\sum_{k=1,2} \beta_i^k \mathbb{E}[\tilde{m}_i^j] &= \bar{W} + C(e) \qquad (106)
\end{aligned}$$

Thus, the board's objective function can be rewritten as:

$$\begin{aligned}
\mathbb{E}[V(x,y,e)] &= \mathbb{E}\left[(1-\beta_x)\left(e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x\right) + \alpha_B\left(\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y\right) - \left(w + \beta_p \tilde{p} + \sum_{i=1,2}\sum_{k=1,2} \beta_i^k \tilde{m}_i^k\right)\right] \\
&= \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x + \alpha_B\left(\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y\right) - \left(\bar{W} + C(e)\right)\right] \qquad (107)
\end{aligned}$$

Following the same steps as in the proof of Proposition 1, maximizing the board's objective function is equivalent to minimizing the expected quadratic distance between $y_i(\varepsilon_i^1, \varepsilon_i^2)$ and $y_i^*(\varepsilon_i^1, \varepsilon_i^2)$, where:

$$\begin{aligned}
\mathbb{E}\left[\tilde{\eta}_i | \varepsilon_i^1, \varepsilon_i^2\right] &= \bar{\eta} + \sum_{k=1,2} \frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right) \\
y_i^*(\varepsilon_i^1, \varepsilon_i^2) &= \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\bar{\eta} + \sum_{k=1,2} \frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right)\right) \qquad (108)
\end{aligned}$$

This problem is:

$$\begin{aligned}
&\min_{\beta_x,\beta_p,\beta_i^k} \sum_{i=1,2}\sum_{k=1,2}\int_{\eta_i}\int_{\varepsilon_i^k}\left(\frac{1}{2}\frac{1}{\theta_i}\frac{\beta_i^k\varepsilon_i^k + \alpha_I\beta_p\left(\frac{\bar{\eta}}{2} + \frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right)\right)}{\beta_x + \beta_p}\right. \\
&\left.-\frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\bar{\eta}}{2} + \frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right)\right)\right)^2 \varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i \\
&\min_{\beta_x,\beta_p,\beta_i^k} \sum_{i=1,2}\frac{1}{2}\frac{1}{\theta_i}\sum_{k=1,2}\int_{\eta_i}\int_{\varepsilon_i^k}\left(\left(\frac{\beta_i^k}{\beta_x + \beta_p} + \left(\frac{\alpha_I\beta_p}{\beta_x + \beta_p} - \alpha_B\right)\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\right)\varepsilon_i^k\right. \\
&\left.+\left(\frac{\alpha_I\beta_p}{\beta_x + \beta_p} - \alpha_B\right)\left(\frac{1}{2} - \frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\right)\bar{\eta}\right)^2 \varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i \qquad (109)
\end{aligned}$$

The expression in equation (109) is globally convex with respect to β_i^k. The FOC w.r.t. β_i^k is:

$$
\begin{aligned}
&\frac{1}{2}\frac{1}{\theta_i}\int_{\eta_i}\int_{\varepsilon_i^k}\frac{2}{\beta_x+\beta_p}\varepsilon_i^k\left(\left(\frac{\beta_i^k}{\beta_x+\beta_p}+\left(\frac{\alpha_I\beta_p}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\right)\varepsilon_i^k\right.\\
&\left.+\left(\frac{\alpha_I\beta_p}{\beta_x+\beta_p}-\alpha_B\right)\left(\frac{1}{2}-\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\right)\bar{\eta}\right)\varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i=0\\
\Leftrightarrow\quad&\left(\frac{\beta_i^k}{\beta_x+\beta_p}+\left(\frac{\alpha_I\beta_p}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\right)\int_{\eta_i}\int_{\varepsilon_i^k}{\varepsilon_i^k}^2\varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i\\
&=\left(\alpha_B-\frac{\alpha_I\beta_p}{\beta_x+\beta_p}\right)\left(\frac{1}{2}-\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\right)\bar{\eta}\int_{\eta_i}\int_{\varepsilon_i^k}\varepsilon_i^k\varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i\\
\Leftrightarrow\quad&\frac{\beta_i^k}{\beta_x+\beta_p}=\left(\alpha_B-\frac{\alpha_I\beta_p}{\beta_x+\beta_p}\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}+\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(1-\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}\right)\right) \quad (110)
\end{aligned}
$$

Letting $\Sigma_k\equiv\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}$ and substituting in equation (109) when β_i^k is as in equation (110):

$$
\begin{aligned}
&\min_{\beta_x,\beta_p}\sum_{i=1,2}\frac{1}{2}\frac{1}{\theta_i}\sum_{k=1,2}\int_{\eta_i}\int_{\varepsilon_i^k}\left(\left(\left(\alpha_B-\frac{\alpha_I\beta_p}{\beta_x+\beta_p}\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}+\Sigma_k\left(1-\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}\right)\right)\right.\right.\\
&\left.\left.+\left(\frac{\alpha_I\beta_p}{\beta_x+\beta_p}-\alpha_B\right)\Sigma_k\right)\varepsilon_i^k+\left(\frac{\alpha_I\beta_p}{\beta_x+\beta_p}-\alpha_B\right)\left(\frac{1}{2}-\Sigma_k\right)\bar{\eta}\right)^2\varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i\\
\Leftrightarrow\quad&\min_{\beta_x,\beta_p}\left(\alpha_B-\frac{\alpha_I\beta_p}{\beta_x+\beta_p}\right)^2\sum_{i=1,2}\frac{1}{2}\frac{1}{\theta_i}\sum_{k=1,2}\int_{\eta_i}\int_{\varepsilon_i^k}\left(\left(\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}+\Sigma_k\left(1-\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}\right)\right)-\Sigma_k\right)\varepsilon_i^k\right.\\
&\left.-\left(\frac{1}{2}-\Sigma_k\right)\bar{\eta}\right)^2\varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i \quad (111)
\end{aligned}
$$

The expression under the integral signs in equation (111) is positive and independent from the contract. The expression in equation (111) is minimized by minimizing $\left(\alpha_B-\frac{\alpha_I\beta_p}{\beta_x+\beta_p}\right)^2$. This is achieved as in the proof of Proposition 1.

Substituting in equation (105) when $\alpha_B\leq\alpha_I$ so that $\beta_i^k=0$:

$$
\begin{aligned}
y_i(\varepsilon_i^1,\varepsilon_i^2)&=\alpha_B\left(\bar{\eta}+\sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k-\bar{\eta}\right)\right)\frac{1}{2\theta_i}\\
&=\alpha_B\left(\bar{\eta}+\sum_{k=1,2}\Sigma_k\left(\varepsilon_i^k-\bar{\eta}\right)\right)\frac{1}{2\theta_i}, \quad (112)
\end{aligned}
$$

which is as in equation (108). In this case, expected social output on dimension i is:

$$\begin{aligned}\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right] &= \frac{\alpha_B}{2\theta_i}\left(\mathbb{E}\left[\tilde{\eta}_i\right]\bar{\eta} + \sum_{k=1,2}\Sigma_k\left(\mathbb{E}\left[\tilde{\eta}_i\tilde{\varepsilon}_i^k\right] - \mathbb{E}\left[\tilde{\eta}_i\right]\bar{\eta}\right)\right)\\ &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\sigma_\eta^2 + \sigma_{\varepsilon_k}^2\right)\right)\end{aligned}$$

For $\rho = -1$:

$$\begin{aligned}\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right] &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 + \sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 + 2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\sigma_\eta^2 + \sigma_{\varepsilon_k}^2\right)\right)\\ &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sum_{k=1,2}\frac{\sigma_{\varepsilon_j}}{\sigma_{\varepsilon_k} + \sigma_{\varepsilon_j}}\left(\sigma_\eta^2 + \sigma_{\varepsilon_k}^2\right)\right)\\ &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sigma_\eta^2 + \frac{\sigma_{\varepsilon_2}\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_1}\sigma_{\varepsilon_2}^2}{\sigma_{\varepsilon_1} + \sigma_{\varepsilon_2}}\right),\end{aligned} \tag{113}$$

which is larger than in equation (48) (the case with one ESG score per dimension) for any parameter values. Moreover, $\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right]$ is continuous in ρ, so that the result about ρ in Proposition 4 when $\alpha_B \leq \alpha_I$ holds by a continuity argument. As $\sigma_2 \to 0$:

$$\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right] \;\to\; \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_2}^2\right)$$

which is larger than in equation (48) for any parameter values. Moreover, $\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right]$ is continuous in σ_2, so that the result about σ_2 in Proposition 4 when $\alpha_B \leq \alpha_I$ holds by a continuity argument.

We have:

$$\begin{aligned}&\frac{\partial}{\partial\rho}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\\ &= \frac{-\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\right) - \sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\left(-2\rho\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - 2\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}\right)}{\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\right)^2}\\ &= \frac{-\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2\sigma_{\varepsilon_k}^2\right) + \sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})2\rho\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2\sigma_{\varepsilon_j}^2\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}}{\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\right)^2}\end{aligned} \tag{114}$$

The sign of the derivative is the same as the numerator's on the RHS of the equation. For $\rho \in (-1, 0)$, the numerator is negative.

For $\rho = 0$:

$$\begin{aligned}\mathbb{E}\left[\tilde{\eta}_i \tilde{y}_i\right] &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sum_{k=1,2} \frac{\sigma_\eta^4 \sigma_{\varepsilon_j}^2 + \sigma_\eta^2 \sigma_{\varepsilon_j}^2 \sigma_{\varepsilon_k}^2}{\sigma_{\varepsilon_k}^2 \sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2)}\right) \\ &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sum_{k=1,2} \frac{\sigma_\eta^4 + \sigma_\eta^2 \sigma_{\varepsilon_k}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2\left(\frac{\sigma_{\varepsilon_k}^2}{\sigma_{\varepsilon_j}^2} + 1\right)}\right)\end{aligned}$$

Substituting in equation (105) when β_i^k is as in equation (110) and $\alpha_B \geq \alpha_I$ so that $\beta_x = 0$:

$$\begin{aligned}y_i(\varepsilon_i^1, \varepsilon_i^2) &= \frac{1}{2\theta_i}\left(\sum_{k=1,2}\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2} + \Sigma_k\left(1 - \frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2}\right)\right)\varepsilon_i^k + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\bar{\eta} + \sum_{k=1,2} \Sigma_k\left(\varepsilon_i^k - \bar{\eta}\right)\right)\right) \\ &= \frac{1}{2\theta_i}\sum_{k=1,2}\left(\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2} + \Sigma_k\left(1 - \frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2}\right)\right)\varepsilon_i^k + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\frac{\bar{\eta}}{2} + \Sigma_k\left(\varepsilon_i^k - \bar{\eta}\right)\right)\right)\end{aligned}$$

In this case, expected social output on dimension i is:

$$\begin{aligned}\mathbb{E}\left[\tilde{\eta}_i \tilde{y}_i\right] &= \frac{1}{2\theta_i}\sum_{k=1,2}\left(\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2} + \Sigma_k\left(1 - \frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2}\right)\right)\mathbb{E}\left[\tilde{\eta}_i \tilde{\varepsilon}_i^k\right]\right. \\ &\quad \left. + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\frac{\bar{\eta}^2}{2} + \Sigma_k\left(\mathbb{E}\left[\tilde{\eta}_i \tilde{\varepsilon}_i^k\right] - \bar{\eta}^2\right)\right)\right) \\ &= \frac{1}{2\theta_i}\sum_{k=1,2}\left(\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2} + \Sigma_k\left(1 - \frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2}\right)\right)\left(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_k}^2\right)\right. \\ &\quad \left. + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\frac{\bar{\eta}^2}{2} + \Sigma_k\left(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_k}^2 - \bar{\eta}^2\right)\right)\right)\end{aligned}$$

For $\rho = -1$:

$$\begin{aligned}\mathbb{E}\left[\tilde{\eta}_i \tilde{y}_i\right] &= \frac{1}{2\theta_i}\sum_{k=1,2}\left(\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_{\varepsilon_k}}{\sigma_{\varepsilon_2} + \sigma_{\varepsilon_1}}\left(1 - \frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2}\right)\right)\left(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_k}^2\right)\right. \\ &\quad \left. + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\frac{\bar{\eta}^2}{2} + \frac{\sigma_{\varepsilon_k}}{\sigma_{\varepsilon_2} + \sigma_{\varepsilon_1}}\left(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_k}^2 - \bar{\eta}^2\right)\right)\right) \\ &= \frac{1}{2\theta_i}\left(\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\bar{\eta}^2 + \sigma_\eta^2 + \sum_{k=1,2}\frac{\sigma_{\varepsilon_k}}{\sigma_{\varepsilon_2} + \sigma_{\varepsilon_1}}\sigma_{\varepsilon_k}^2\right) + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\bar{\eta}^2 + \sigma_\eta^2 + \sum_{k=1,2}\frac{\sigma_{\varepsilon_k}}{\sigma_{\varepsilon_2} + \sigma_{\varepsilon_1}}\sigma_{\varepsilon_k}^2\right)\right)\end{aligned}$$

which is larger than in equation (48) for any parameter values. Moreover, $\mathbb{E}\left[\tilde{\eta}_i \tilde{y}_i\right]$ is continuous in ρ, so that the result about ρ in Proposition 4 when $\alpha_B > \frac{\beta_p \alpha_I}{\beta_x + \beta_p}$ holds by a continuity argument. As $\sigma_2 \to 0$:

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\eta}_i \tilde{y}_i\right] &\rightarrow \frac{1}{2\theta_i}\left(\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2 + \sigma_\eta^2 + \bar{\eta}^2} + \left(1 - \frac{\bar{\eta}^2}{\sigma_{\varepsilon_1}^2 + \sigma_\eta^2 + \bar{\eta}^2}\right)\right)\left(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_1}^2\right)\right. \\
&\quad \left. + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\bar{\eta}^2 + \bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_1}^2 - \bar{\eta}^2\right)\right) \\
&= \frac{1}{2\theta_i}\left(\left(\alpha_B - \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\right)\left(\bar{\eta}^2 + \bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_1}^2 - \bar{\eta}^2\right) + \frac{\beta_p \alpha_I}{\beta_x + \beta_p}\left(\bar{\eta}^2 + \bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_1}^2 - \bar{\eta}^2\right)\right) \\
&= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_1}^2\right)
\end{aligned}
$$

which is larger than in equation (48) for any parameter values. Moreover, $\mathbb{E}\left[\tilde{\eta}_i \tilde{y}_i\right]$ is continuous in σ_2, so that the result about σ_2 in Proposition 4 when $\alpha_B \leq \frac{\beta_p \alpha_I}{\beta_x + \beta_p}$ holds by a continuity argument. ■

2 Case with two imperfectly correlated sets of ESG scores

For simplicity, we study the case without stock price-based compensation ($\beta_p = 0$) and no intrinsic managerial preference for social responsibility ($\alpha_M = 0$).

Rewriting the manager's objective function when there are two sets of ESG scores gives:

$$
\arg\max_{e, y_1, y_2} \mathbb{E}\left[w + \beta_x\left(e - \theta_1 y_1^2 - \theta_2 y_2^2\right) + \beta_1^1 m_1^1 + \beta_1^2 m_1^2 + \beta_2^1 m_2^1 + \beta_2^2 m_2^2\right] - C(e) \tag{115}
$$

The first-order conditions with respect to y_1 and y_2 are:

$$
y_1 = \frac{1}{2}\frac{1}{\theta_1}\frac{\beta_1^1 \varepsilon_1^1 + \beta_1^2 \varepsilon_1^2}{\beta_x} \tag{116}
$$

$$
y_2 = \frac{1}{2}\frac{1}{\theta_2}\frac{\beta_2^1 \varepsilon_2^1 + \beta_2^2 \varepsilon_2^2}{\beta_x} \tag{117}
$$

Incentive compatibility with respect to effort e is achieved as in equation (28) with $\beta_p = 0$, which requires $\beta_x \geq \frac{c_e}{\bar{e} - \underline{e}}$. Setting w so that the manager is at his reservation utility at the contracting phase, the board's objective function at the contracting phase can be rewritten as:

$$
\begin{aligned}
\mathbb{E}[V(x,y,e)] &= \mathbb{E}\Bigg[\bar{e} - \theta_1\left(\frac{1}{2}\frac{1}{\theta_1}\frac{\beta_1^1\tilde{\varepsilon}_1^1+\beta_1^2\tilde{\varepsilon}_1^2}{\beta_x}\right)^2 - \theta_2\left(\frac{1}{2}\frac{1}{\theta_1}\frac{\beta_2^1\tilde{\varepsilon}_2^1+\beta_2^2\tilde{\varepsilon}_2^2}{\beta_x}\right)^2 + \tilde{\epsilon}_x \\
&\quad +(1-\beta_x)\alpha\left(\tilde{\eta}_1\frac{1}{2}\frac{1}{\theta_1}\frac{\beta_1^1\tilde{\varepsilon}_1^1+\beta_1^2\tilde{\varepsilon}_1^2}{\beta_x} + \tilde{\eta}_2\frac{1}{2}\frac{1}{\theta_1}\frac{\beta_2^1\tilde{\varepsilon}_2^1+\beta_2^2\tilde{\varepsilon}_2^2}{\beta_x} + \tilde{\epsilon}_y\right) - \left(\bar{W}+c_e\right)\Bigg] \\
&= \bar{e} - \frac{1}{4\theta_1}\frac{\beta_1^{1^2}\mathbb{E}\left[\tilde{\varepsilon}_1^{1^2}\right] + \beta_1^{2^2}\mathbb{E}\left[\tilde{\varepsilon}_1^{2^2}\right] + 2\beta_1^1\beta_1^2\mathbb{E}\left[\tilde{\varepsilon}_1^1\tilde{\varepsilon}_1^2\right]}{\beta_x^2} \\
&\quad -\frac{1}{4\theta_2}\frac{\beta_2^{1^2}\mathbb{E}\left[\tilde{\varepsilon}_2^{1^2}\right] + \beta_2^{2^2}\mathbb{E}\left[\tilde{\varepsilon}_2^{2^2}\right] + 2\beta_2^1\beta_2^2\mathbb{E}\left[\tilde{\varepsilon}_2^1\tilde{\varepsilon}_2^2\right]}{\beta_x^2} \\
&\quad +(1-\beta_x)\alpha\left(\frac{1}{2\theta_1}\frac{\beta_1^1\mathbb{E}\left[\tilde{\eta}_1\tilde{\varepsilon}_1^1\right]+\beta_1^2\mathbb{E}\left[\tilde{\eta}_1\tilde{\varepsilon}_1^2\right]}{\beta_x} + \frac{1}{2\theta_2}\frac{\beta_2^1\mathbb{E}\left[\tilde{\eta}_2\tilde{\varepsilon}_2^1\right]+\beta_2^2\mathbb{E}\left[\tilde{\eta}_2\tilde{\varepsilon}_2^2\right]}{\beta_x}\right) - \left(\bar{W}+c_e\right) \\
&= \bar{e} - \frac{1}{4\theta_1}\frac{\beta_1^{1^2}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2}) + \beta_1^{2^2}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2}) + 2\beta_1^1\beta_1^2(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)}{\beta_x^2} \\
&\quad -\frac{1}{4\theta_2}\frac{\beta_2^{1^2}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2}) + \beta_2^{2^2}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2}) + 2\beta_2^1\beta_2^2(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)}{\beta_x^2} \\
&\quad +(1-\beta_x)\alpha\left(\frac{1}{2\theta_1}\frac{(\beta_1^1+\beta_1^2)(\sigma_\eta^2+\bar{\eta}^2)}{\beta_x} + \frac{1}{2\theta_2}\frac{(\beta_2^1+\beta_2^2)(\sigma_\eta^2+\bar{\eta}^2)}{\beta_x}\right) - \left(\bar{W}+c_e\right) \qquad (118)
\end{aligned}
$$

where we used:

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\varepsilon}_i^{j^2}\right] &= \bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{j^2} \\
\mathbb{E}\left[\tilde{\varepsilon}_i^1\tilde{\varepsilon}_i^2\right] &= cov\left(\tilde{\varepsilon}_i^1,\tilde{\varepsilon}_i^2\right) + \mathbb{E}\left[\tilde{\varepsilon}_i^1\right]\mathbb{E}\left[\tilde{\varepsilon}_i^2\right] = \rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2 \\
\mathbb{E}\left[\tilde{\eta}_i\tilde{\varepsilon}_i^j\right] &= cov\left(\tilde{\eta}_i,\tilde{\varepsilon}_i^j\right) + \mathbb{E}\left[\tilde{\eta}_i\right]\mathbb{E}\left[\tilde{\varepsilon}_i^j\right] = \sigma_\eta^2+\bar{\eta}^2
\end{aligned}
$$

The objective function is concave with respect to β_i^j. The first-order condition with respect to β_i^j is:

$$
\begin{aligned}
&\frac{1}{2\theta_i}\left(\frac{\beta_i^j}{\beta_x^2}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{j^2}) + \frac{\beta_i^k}{\beta_x^2}(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2)\right) = \alpha(1-\beta_x)\frac{1}{2\theta_i}\frac{\sigma_\eta^2+\bar{\eta}^2}{\beta_x} \\
\Leftrightarrow\quad &\begin{cases} \frac{\beta_i^j}{\beta_x}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{j^2}) + \frac{\beta_i^k}{\beta_x}(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2) = \alpha\left(1-\frac{c_e}{\bar{e}-\underline{e}}\right)(\sigma_\eta^2+\bar{\eta}^2) \\ \frac{\beta_i^k}{\beta_x}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{k^2}) + \frac{\beta_i^j}{\beta_x}(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2) = \alpha\left(1-\frac{c_e}{\bar{e}-\underline{e}}\right)(\sigma_\eta^2+\bar{\eta}^2) \end{cases} \qquad (119)
\end{aligned}
$$

where $k \neq j$, i.e. $k=2$ if $j=1$ and $k=1$ if $j=2$. With $c_e \to 0$, this can be rewritten as:

$$
\begin{cases} \frac{\beta_i^k}{\beta_x} = \frac{1}{\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2}\left(\alpha(\sigma_\eta^2+\bar{\eta}^2) - \frac{\beta_i^j}{\beta_x}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{j^2})\right) \\ \frac{1}{\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2}\left(\alpha(\sigma_\eta^2+\bar{\eta}^2) - \frac{\beta_i^j}{\beta_x}(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{j^2})\right)(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{k^2}) + \frac{\beta_i^j}{\beta_x}(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2) = \alpha(\sigma_\eta^2+\bar{\eta}^2) \end{cases}
$$

$$
\begin{aligned}
\Rightarrow \quad \frac{\beta_i^j}{\beta_x} &= \alpha \frac{(\sigma_\eta^2+\bar{\eta}^2) - \frac{(\sigma_\eta^2+\bar{\eta}^2)(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{k^2})}{\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2}}{(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2) - \frac{(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{j^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{k^2})}{\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2}} \\
&= \alpha \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_i}^{k^2}-\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_i}^{j^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_i}^{k^2}) - (\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2+\bar{\eta}^2)^2}
\end{aligned}
\tag{120}
$$

As $c_e \to 0$, the optimum for the board is thus given by setting $\beta_x \to 0$ and β_1 and β_2 are as in equation (120). Substituting for β_1/β_x and β_2/β_x in equations (116) and (117), respectively:

$$
y_1 = \frac{\alpha}{2}\frac{1}{\theta_1}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\varepsilon_1^1 + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\varepsilon_1^2\right) \tag{121}
$$

$$
y_2 = \frac{\alpha}{2}\frac{1}{\theta_2}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\varepsilon_2^1 + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\varepsilon_2^2\right) \tag{122}
$$

We now derive expected social output in the case with a set of N ESG scores when the board uses an explicit contract for social investments. Expected social output in this case:

$$
\begin{aligned}
\mathbb{E}[\tilde{y}] &= \mathbb{E}[\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y] \\
&= \mathbb{E}\Bigg[\tilde{\eta}_1\frac{\alpha}{2}\frac{1}{\theta_1}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\tilde{\varepsilon}_1^1 + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\tilde{\varepsilon}_1^2\right) \\
&\quad + \tilde{\eta}_2\frac{\alpha}{2}\frac{1}{\theta_2}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\tilde{\varepsilon}_2^1 + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\tilde{\varepsilon}_2^2\right) + \tilde{\epsilon}_y\Bigg] \\
&= \frac{\alpha}{2}\frac{1}{\theta_1}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\mathbb{E}[\tilde{\eta}_1\tilde{\varepsilon}_1^1] + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\mathbb{E}[\tilde{\eta}_1\tilde{\varepsilon}_1^2]\right) \\
&\quad + \frac{\alpha}{2}\frac{1}{\theta_2}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\mathbb{E}[\tilde{\eta}_2\tilde{\varepsilon}_2^1] + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\mathbb{E}[\tilde{\eta}_2\tilde{\varepsilon}_2^2]\right) \\
&= \frac{\alpha}{2}\frac{1}{\theta_1}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2} + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\right)(\sigma_\eta^2+\bar{\eta}^2) \\
&\quad + \frac{\alpha}{2}\frac{1}{\theta_2}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2} + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\right)(\sigma_\eta^2+\bar{\eta}^2) \\
&= \frac{\alpha}{2}\frac{1}{\theta_1}\frac{2\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2}+\sigma_{\varepsilon_1}^{2^2}-2\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\left(\sigma_\eta^2+\bar{\eta}^2\right)^2 \\
&\quad + \frac{\alpha}{2}\frac{1}{\theta_2}\frac{2\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2}+\sigma_{\varepsilon_2}^{2^2}-2\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\left(\sigma_\eta^2+\bar{\eta}^2\right)^2
\end{aligned}
$$

Comparing with the corresponding equation for the case with one set of ESG scores, expected social output when the board delegates the social investment decisions to the manager is higher with a second scores if:

$$\left\{\begin{array}{l} \frac{2\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}+\sigma^{2^2}_{\varepsilon_1}-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}}{(\sigma_\eta^2+\bar{\eta}^2+\sigma^{1^2}_{\varepsilon_1})(\bar{\eta}^2+\sigma_\eta^2+\sigma^{2^2}_{\varepsilon_1})-(\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+\bar{\eta}^2)^2}\left(\sigma_\eta^2+\bar{\eta}^2\right) > \frac{\bar{\eta}^2+\sigma_\eta^2}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}} \\ \frac{2\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_2}+\sigma^{2^2}_{\varepsilon_2}-2\rho\sigma^1_{\varepsilon_2}\sigma^2_{\varepsilon_2}}{(\sigma_\eta^2+\bar{\eta}^2+\sigma^{1^2}_{\varepsilon_2})(\bar{\eta}^2+\sigma_\eta^2+\sigma^{2^2}_{\varepsilon_2})-(\rho\sigma^1_{\varepsilon_2}\sigma^2_{\varepsilon_2}+\bar{\eta}^2)^2}\left(\sigma_\eta^2+\bar{\eta}^2\right) > \frac{\bar{\eta}^2+\sigma_\eta^2}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_2}} \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} \frac{2\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}+\sigma^{2^2}_{\varepsilon_1}-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{2^2}_{\varepsilon_1}-\frac{(\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+\bar{\eta}^2)^2}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}}} > 1 \\ \frac{2\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_2}+\sigma^{2^2}_{\varepsilon_2}-2\rho\sigma^1_{\varepsilon_2}\sigma^2_{\varepsilon_2}}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{2^2}_{\varepsilon_2}-\frac{(\rho\sigma^1_{\varepsilon_2}\sigma^2_{\varepsilon_2}+\bar{\eta}^2)^2}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_2}}} > 1 \end{array}\right.$$

The first condition is equivalent to:

$$\begin{aligned}
& 2\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}+\sigma^{2^2}_{\varepsilon_1}-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1} > \bar{\eta}^2+\sigma_\eta^2+\sigma^{2^2}_{\varepsilon_1}-\frac{(\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+\bar{\eta}^2)^2}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}} \\
\Leftrightarrow\quad & \sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1} > \bar{\eta}^2-\frac{(\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+\bar{\eta}^2)^2}{\bar{\eta}^2+\sigma_\eta^2+\sigma^{1^2}_{\varepsilon_1}} \\
\Leftrightarrow\quad & \bar{\eta}^2\sigma_\eta^2+\sigma_\eta^4+\sigma^{1^2}_{\varepsilon_1}\sigma_\eta^2+\bar{\eta}^2\sigma^{1^2}_{\varepsilon_1}+\sigma_\eta^2\sigma^{1^2}_{\varepsilon_1}+\sigma^{1^4}_{\varepsilon_1}-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\bar{\eta}^2-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\sigma^2_\eta-2\rho\sigma^{1^3}_{\varepsilon_1}\sigma^2_{\varepsilon_1} \\
& > \bar{\eta}^4+\sigma_\eta^2\bar{\eta}^2+\sigma^{1^2}_{\varepsilon_1}\bar{\eta}^2-\rho^2\sigma^{1^2}_{\varepsilon_1}\sigma^{2^2}_{\varepsilon_1}-\bar{\eta}^4-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\bar{\eta}^2 \\
\Leftrightarrow\quad & \sigma_\eta^4+2\sigma^{1^2}_{\varepsilon_1}\sigma_\eta^2+\sigma^{1^4}_{\varepsilon_1}-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\sigma^2_\eta-2\rho\sigma^{1^3}_{\varepsilon_1}\sigma^2_{\varepsilon_1} > -\rho^2\sigma^{1^2}_{\varepsilon_1}\sigma^{2^2}_{\varepsilon_1} \qquad (123)
\end{aligned}$$

The condition in equation (123) is always satisfied with $\rho \leq 0$ since in this case the LHS is positive and the RHS is nonpositive. More generally, the condition in equation (123) is satisfied for any ρ if it is satisfied for $\rho = 1$. For $\rho = 1$, it is satisfied if and only if:

$$\begin{aligned}
& \sigma_\eta^4+2\sigma^{1^2}_{\varepsilon_1}\sigma_\eta^2+\sigma^{1^4}_{\varepsilon_1} > 2\sigma_\eta^2\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+2\sigma^{1^3}_{\varepsilon_1}\sigma^2_{\varepsilon_1}-\sigma^{1^2}_{\varepsilon_1}\sigma^{2^2}_{\varepsilon_1} \\
\Leftrightarrow\quad & \sigma_\eta^2(\sigma_\eta^2+2\sigma^{1^2}_{\varepsilon_1}-2\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}) > \sigma^{1^2}_{\varepsilon_1}(2\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}-\sigma^{2^2}_{\varepsilon_1}-\sigma^{1^2}_{\varepsilon_1}) \\
\Leftrightarrow\quad & \sigma_\eta^2(\sigma_\eta^2+2\sigma^{1^2}_{\varepsilon_1}-2\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}) > -\sigma^{1^2}_{\varepsilon_1}(\sigma^2_{\varepsilon_1}-\sigma^1_{\varepsilon_1})^2 \qquad (124)
\end{aligned}$$

The LHS is quadratic with respect to σ_η^2, with a minimum at

$$\sigma_\eta^2+2\sigma^{1^2}_{\varepsilon_1}-2\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+\sigma_\eta^2 = 2\sigma_\eta^2+2\sigma^{1^2}_{\varepsilon_1}-2\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1} = 0 \quad\Leftrightarrow\quad \sigma_\eta^2 = \sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}-\sigma^{1^2}_{\varepsilon_1}$$

Replacing in equation (124), this condition is generically satisfied at $\rho = 1$ for any σ_η^2 if and only if:

$$\begin{aligned}
& (\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}-\sigma^{1^2}_{\varepsilon_1})(\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}-\sigma^{1^2}_{\varepsilon_1}+2\sigma^{1^2}_{\varepsilon_1}-2\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}) \geq -\sigma^{1^2}_{\varepsilon_1}(\sigma^2_{\varepsilon_1}-\sigma^1_{\varepsilon_1})^2 \\
\Leftrightarrow\quad & (\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}-\sigma^{1^2}_{\varepsilon_1})(\sigma^{1^2}_{\varepsilon_1}-\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}) \geq -\sigma^{1^2}_{\varepsilon_1}(\sigma^2_{\varepsilon_1}-\sigma^1_{\varepsilon_1})^2 \\
\Leftrightarrow\quad & -\sigma^{1^2}_{\varepsilon_1}(\sigma^2_{\varepsilon_1}-\sigma^1_{\varepsilon_1})^2 \geq -\sigma^{1^2}_{\varepsilon_1}(\sigma^2_{\varepsilon_1}-\sigma^1_{\varepsilon_1})^2
\end{aligned}$$

which is true. The same reasoning can be applied to the second condition. ■