# Executive Compensation with Social and Environmental Performance*

Pierre Chaigneau† Nicolas Sahuguet‡

Queen's University HEC Montréal

April 4, 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 uses measures of social performance when the level of corporate social responsibility preferred by the board exceeds the one that maximizes the stock price. Thus, explicit social incentives and socially responsible investors are substitutes. Relying on multiple measures based on different methodologies will generally mitigate inefficiencies due to gaming, i.e. harmonization of social performance measurement can backfire. This has normative implications for the regulation and harmonization of ESG measurement.

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

---

*For useful comments and suggestions we thank Jeremy Bertomeu, Milo Bianchi, Jörg Budde (discussant), Alex Edmans, Thomas Geelen (discussant), Christian Hofmann, Nicolas Inostroza, Jan Mahrt-Smith, Lucas Mahieux, Günter Strobl, John Van Reenen, and participants at the Accounting and Economics Society webinar, Accounting Research Workshop, Asia Meeting of the Econometric Society, Erasmus Corporate Governance Conference, the Québec Political Economy Conference, 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]

At the same time, prominent scholars have argued that metrics of social and environmental performance are imperfect and easily gamed, and therefore should not be used for incentive purposes (Edmans (2021), Bebchuk and Tallarita (2022)). 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."

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? To answer this question, we consider a stylized principal-agent model of multitasking in which a 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 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

[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.

knows the bias at the time of making investment decisions, SEP-based incentives will lead him to game the incentive system.

These assumptions allow to take into account a frequent criticism leveled at SEP-based incentives (that they will encourage "gaming"). They also represent a departure from standard models of multitasking, manipulation, or signal-jamming, in which the principal is aware of any biases that the manager may have at the contracting phase. In particular, existing models of multitasking address the consequences of imperfect ex-post measurement for incentive provision (Holmström and Milgrom (1991), Feltham and Xie (1994)), whereas we address the consequences of biased measurement which is privately known ex-ante by the manager.

The manager's effort and investment decisions will affect the firm's stock price, which is an important performance metric in incentive systems (Holmström and Tirole (1993)).[2] In recent years, there has been a significant increase in the propensity of investors to be socially responsible. There is now empirical evidence that shareholders are willing to sacrifice financial returns to invest in more sustainable or socially responsible firms (Barber, Morse, and Yasuda (2021), Bauer, Ruof, and Smeets (2021)), and that their demand for shares depends on the externalities generated by firms such as carbon emissions (Bolton and Kacperczyk (2021)). In response, ESG raters have started providing measures of SEP known as ESG scores and ratings that allow investors to assess firms' SEP. Accordingly, we let stock market investors be "socially responsible", i.e., their objective function puts some weight on the externalities generated by the firm.

To start, suppose that the board's objective is to maximize either firm output or the stock price, consistently with a narrow interpretation of its fiduciary duty. In this case, we show that the compensation contract does not include SEP measures. Moreover, when investors are not socially responsible, the manager's compensation is only contingent on the stock price. Intuitively, the stock price aggregates information about the firm's performance efficiently and consistently with the board's preferences. On the contrary, if the stock price is set by socially responsible investors, then the manager's compensation should also be positively related to a measure of the firm's profitability. Intuitively, the level of investment in externalities mitigation that maximizes the stock price exceeds the level of investment in externalities mitigation that the board would prefer, and profits-based compensation discourages such costly investment. In this case, the board

[2] 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.

can still induce the first-best level of investment – as defined as the hypothetical outcome in the absence of an agency problem.

This model cannot rationalize why some firms include SEP metrics in executive compensation contracts. To explain this practice, we show that the board must want to encourage the manager to improve not just financial performance but also SEP, i.e. it must be "socially responsible". As further discussed in section 1.4 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.

With a socially responsible board, the model can rationalize the use of SEP measures in executive compensation. However, this is only the case if the board is *more* socially responsible than stock market investors. In this case, compensation is explicitly based on the stock price and on SEP measures, even though these measures may have a low quality and will lead the manager to "game" the incentive system. These measures are used to supplement incentives for investment in externalities mitigation already embedded in stock price-based compensation. Moreover, the sensitivity of compensation to SEP measures is decreasing in the variance of their bias (i.e., increasing in their quality).[3] For 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 especially useful for incentive purposes because it efficiently aggregates all available information. By contrast, explicit incentives based on SEP measures put all the weight on the signal provided by these measures. As a result, in the aforementioned case, the board cannot offer a contract that induces the first-best level of investment in externalities mitigation.[4] Depending on the (random) bias in a SEP measure, the actual level of investment can be higher or

[3]It should be emphasized that the low quality of these measures is not a problem per se. In section B of the Appendix, we study a setting similar in every respect except that 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 does not depend on their quality, and it is higher than in the baseline model that features "gaming".

[4]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. Intuitively, the stock price aggregates information in a way that cannot be replicated by a contract.

lower than the first-best level. On average, the level of investment is then lower than the first-best level. Intuitively, less efficient incentive provision on this dimension tends to reduce investment.

We analyze the outcome when the board and stock market investors don't put the same weight on all SEP measures. For example, the board might care more about the firm's carbon emissions but less about working conditions than stock market investors or have different beliefs in that regard. 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 it could be complemented with targeted compensation based on specific SEP measures.

The model generates empirical implications about the use of SEP measures in incentive programs. A measure of SEP is only used when the board is more socially responsible than stock market investors on this dimension. Thus, 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.[5]

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 improve externalities mitigation, it can alternatively worsen it if the quality of additional SEP measures is sufficiently low.

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

This is in contrast to models of contracting or multitasking in which the value of a new performance measure is always nonnegative (e.g. Holmström (1979), Feltham and Xie (1994)). Intuitively, additional SEP measures will affect the stock price and therefore 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 results in greater externalities mitigation on this dimension, and that the distorting 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), and it suggests that regulatory efforts should instead focus on improving the quality of ESG scores or ratings. 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.[6]

**Related literature**

John, Saunders, and Senbet (2000) have explored how executive compensation can be used to mitigate another inefficiency due to misaligned incentives – risk shifting by banks. They argue that capital and asset choice regulations only indirectly affect the incentives of bank managers, but that changes in managerial compensation directly and effectively affect their decisions. Contracting based on ratings is related to Rajan and Parlour (2020), who analyze a principal-agent model with a contract that can be contingent on credit ratings, and to Hörner and Lambert (2021), who study how to optimally design a rating for incentive purposes.

The multitasking literature has already analyzed incentive provision when there are more tasks than performance measures, as well as the importance of maintaining balanced incentives across

[6]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. In the latter article, 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."

tasks (Holmström and Milgrom (1991), Dewatripont, Jewitt, and Tirole (2000)).[7] 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 (i.e. the manager observes the "noise" before taking his action: he can predict the future realization of performance measures, as opposed to their expected future realizations in other models), which introduces new possibilities from the manager to game a performance-based contract when allocating firm resources.

The possibility of gaming is related to the vast literature on manipulation, but it involves different mechanisms: "manipulation" typically means altering the value of a signal such as a "report" which is related to a contractible performance measure to inflate managerial pay (e.g. Beyer, Guttman, and Marinovic (2014)). By contrast, "gaming" refers to the manager using his private information to mostly spend resources on dimensions where the impact on measurable performance is the greatest, and conversely to avoid spending resources on dimensions where the impact on measurable performance is minimal. The optimal contract is designed in part to mitigate this distortion in resource allocation.

Our paper takes an optimal contracting perspective to the provision of incentives for corporate investment in social goods.[8] Baron (2008) analyzes how the manager's compensation contract changes depending on the CSR preferences of several key stakeholders of the firm, especially the balance between "profit incentives" and "social incentives". In his model, the manager has private information about her ability, but 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. 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

[7] This framework has been applied to accounting-based and market-based measures of performance (Lambert and Larcker (1987)), balanced scorecards (Budde (2007), Kvaløy and Olsen (2022)), short-term and long-term performance (Holmström and Tirole (1993)), information aggregation into the stock price (Paul (1992)), and performance measure congruity (Feltham and Xie (1994), Datar, Kulp, and Lambert (2001), Bonham and Riggs-Cragun (2021)). Although this is not the primary focus of our analysis, we find that optimal incentives for social investments are increasing in profits-based incentives, which is consistent with the importance of balancing incentives across tasks.

[8] Corporate social responsibility can also be fostered by investor activism: Landier and Lovo (2020), Edmans, Levit, and Schneemeier (2022), and Gupta, Kopytov, and Starmans (2022) study how socially conscious investors will force profit-maximizing companies to partially internalize externalities. Chang, Rhee, and Yoon (2023) study financing of ESG projects by socially conscious lenders when the firm can renege on its ESG promises.

shareholder value.

Compensation based on measures of social or environmental performance, including ESG-based compensation, is still in its infancy. Flammer, Hong, and Minor (2019), Ikram, Li, and Minor (2019), Homroy, Mavruk, and Nguyen (2023), and Pawliczek, Carter, and Zhong (2023) provide empirical evidence on CSR-contingent executive compensation contracts, and argue that it solves an agency problem. On the contrary, Bebchuk and Tallarita (2022) have criticized companies that choose a narrow set of ESG metrics which correspond to easily achievable objectives but do not capture what is most consequential to their stakeholders. They also criticize the use of vague metrics that help corporate leaders retain discretion over the final amount of compensation. Even if these metrics can be poorly used, the notions that ESG-based compensation is a manifestation of poor governance or that ESG targets are easily achieved are inconsistent with the empirical findings of Homroy, Mavruk, and Nguyen (2023). There is also some evidence that the adoption of "say-on-pay" voting laws and greater shareholder engagement lead to an increase in ESG contracting (Cohen et al. (2023), Pawliczek, Carter, and Zhong (2022)). Recent empirical studies commonly use ESG scores as measures of a corporation's ESG performance or ESG quality, e.g. Giese et al. (2019), Lopez de Silanes, McCahery, and Pudschedl (2022), Homroy, Mavruk, and Nguyen (2023), Pawliczek, Carter, and Zhong (2023).

# 1 The model

Consider a firm run by a manager and controlled by a socially responsible board on behalf of shareholders.

## 1.1 Technology

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 investment decisions, $y_1$ and $y_2$, that improve social and environmental outcomes but decrease the firm's profits. We refer to them as "social investments". Cash flows or "profits" $\tilde{x}$ and overall "social output" $\tilde{y}$ are respectively defined as:

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

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 $\tilde{\eta}_2 \sim \mathcal{N}(\bar{\eta}, \sigma_\eta^2)$. All random variables, $\tilde{\eta}_1$, $\tilde{\eta}_2$, $\tilde{\epsilon}_x$, $\tilde{\epsilon}_y$, are independent. The firm's technology for social output is fully described by its costs $\{\theta_1, \theta_2\}$, which are assumed to be positive, and its multidimensional "social productivity" $\{\eta_1, \eta_2\}$.[9]

The board's objective function given its information $J$ is:

$$\mathbb{E}[V(x, y, e)|J] = \mathbb{E}\left[\tilde{x} + \alpha_B \tilde{y} - \tilde{W} \middle| J\right], \tag{2}$$

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 section 1.4, 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.

## 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 $i \in \{1, 2\}$) 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 $\eta_i$ is the realization of $\tilde{\eta}_i$. That is, with $\sigma_\varepsilon^2 > 0$, $\varepsilon_i$ is a noisy measure of the firm's social productivity on dimension $i$.

[9] 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.

At $t = 0$, after contracting but before the social investment decision, the manager observes the nonverifiable variable $\varepsilon_i$. This assumption allows to parsimoniously capture the notion that, because of his on-the-job expertise and his understanding of the methodology used to determine each measure of SEP, the manager understands how investment decisions will affect SEP measures.[10] The assumption that the manager makes decisions after observing $\varepsilon_i$ is similar to the assumption in Edmans and Gabaix (2011) that the agent chooses his action after observing the noise.

We will henceforth refer to SEP measures as "ESG scores" for brevity (see section A of the Appendix for institutional details), but our model can be applied more broadly to other standardized measures of SEP.

As in Holmström and Tirole (1993), cash flows are not yet realized when the stock price is set, but a publicly observable financial report $z$, which is imperfectly informative about the firm's profitability, is realized at $t = 1$ and is such that:

$$\tilde{z} = e - \theta_1 y_1^2 - \theta_2 y_2^2 + \epsilon_x + \tilde{\epsilon}_z \qquad \text{where } \tilde{\epsilon}_z \sim \mathcal{N}(0, \sigma_z^2). \tag{4}$$

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.[11] 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.

## 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 in principal-agent models, the manager is self-interested. He receives a fixed wage $w$, and his compensation is linear in the following performance measures: firm cash flows (with sensitivity $\beta_x$), stock price (with sensitivity $\beta_p$), and ESG score $i$ (with sensitivity

[10] It is worth noting that both conditions are needed. First of all, the manager must understand how SEP measures are constructed. However, this understanding is only useful because of his practical experience running the firm, which allows him to understand how various business decisions will affect the firm's SEP measures.

[11] This is the simplest specification of investors' social preferences. If investors have heterogeneous social preferences, the parameter $\alpha_I$ 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.

$\beta_i$) for $i = 1, 2$.[12] Following Holmström and Tirole (1993), linear contracts are commonly assumed in agency models with several performance measures. In practice, some executive compensation contracts include ESG ratings and scores as performance metrics (Cohen et al. (2023), Table 3).

We assume that the sensitivity of pay to cash flows and ESG scores must be nonnegative, i.e. $\beta_x \geq 0$ and $\beta_i \geq 0$ for $i = 1, 2$.[13] The former ($\beta_x \geq 0$) can be motivated similarly to Innes (1990): the manager would destroy output if the sensitivity of his compensation to cash flows were negative. 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).

If the manager accepts the contract, he learns $\{\varepsilon_1, \varepsilon_2\}$ from his practical experience on the job and his understanding of the ESG scoring methodology, chooses social investments $\{y_1, y_2\}$, and privately chooses effort $e$. 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. Unless otherwise specified, the discount rate is zero.

The timeline of the model is depicted in Figure 1.

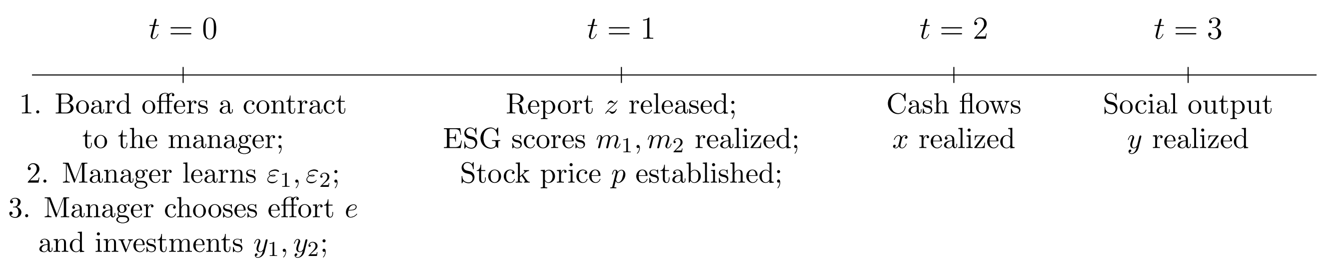


Figure 1: Timeline of the model.

[12]This implies that investments $y_i$ for $i = 1, 2$ are not included in the contract. In practice, explicit incentives are typically not based on investment decisions, possibly because of perverse incentives. For example, contracting based on social investment $i$ might incentivize the manager to funnel money to a third party under the guise of increasing $y_i$ and then get the money back, possibly at a small fee, so that the firm's cash flows are not significantly affected. Although it does not consider social investments, the literature on underinvestment and overinvestment usually assume that corporate investment is noncontractible (e.g. Stulz (1990), Inderst and Müller (2003)).

[13]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 $\beta_x \geq 0$.

## 1.4 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.[14] 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 $\alpha_I$ on social output relative to cash flows. 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 cash flows 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.[15] This is related to the notion of "enlarged fiduciary duty" proposed by Tirole

[14] 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.

[15] 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

(2001), in which stakeholders could sue a firm whose actions did not "follow the 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 $\alpha_B$ and $\alpha_I$ 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)).[16] 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.[17]

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 (Hartzmark and Sussman (2019), Barber, Morse, and Yasuda (2021), Bauer, Ruof, and Smeets (2021), 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).[18]

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 $\alpha_B$

[16] 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.

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

[18] 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)).

and $\alpha_I$ are respectively the board's and investors' preference for social output relative to cash flows. A discrepancy between these two preference parameters can be justified because the firm's shareholders are not necessarily the same economic agents as investors who actively trade on financial markets.

# 2 Social performance in a corporate governance system

## 2.1 Baseline model

We solve the model by backward induction. Section D of the Appendix derives the stock price in several cases of interest. For publicly traded firms, the stock price is established on stock markets. For private firms, it represents the implied stock valuation established in a sale or a funding round. Lemma 1, proven in section D.2 of the Appendix, determines the $t = 1$ stock price in the baseline model.

**Lemma 1** *Given a financial report $z$, ESG scores $m_1$ and $m_2$, and expected managerial effort $\hat{e}$, the stock price $p$ is:*

$$p = \mu_{x|z,m} + \alpha_I \mu_{y|z,m} \tag{5}$$

$$\text{where} \quad \mu_{x|z,m} = \hat{e} - \theta_1 y_1^2 + \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2}\left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) \tag{6}$$

$$\mu_{y|z,m} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} m_1 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} y_1 \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} m_2 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} y_2 \bar{\eta} \tag{7}$$

The stock price is additive in two terms. The first is expected cash flows conditional on investors' information. 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 cash flows $\tilde{x}$. 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. Since investors with $\alpha_I > 0$ have social preferences, the stock price depends on the perceived SEP of the firm as reflected in ESG scores. 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 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, investors understand that ESG scores are noisy measures and that social investments are distorted accordingly, so that neither these investments nor these scores reveal the firm's actual social productivity. As a result, denoting by $I$ investors' information set at the time of setting the stock price, their expectation of the firm's social productivity ($\mathbb{E}[\tilde{\eta}_i|I]$) does not only rely on the firm's ESG scores and its social investments, but also on their 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 cash flows, and its ESG scores.

As a first preliminary step, in section D.1 of the Appendix, we consider the case without ESG scores or ratings. In this case, we show that the stock price is informative about the levels of social investments ($\{y_1, y_2\}$), which are observed by investors. However, in the absence of measures of the firm's SEP, the stock price is then 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 provides a useful contrast with the rest of the analysis, and 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 an agency problem, e.g. when the firm is owned by a shareholder who is also the firm's manager. Let $y_i(\varepsilon_i)$ be the social investment in dimension $i$ optimally chosen by the manager given his contract and his signal $\varepsilon_i$, let $\varphi$ denote the conditional density function of $\tilde{\varepsilon}_i$, and $\phi$ denote the density function of $\tilde{\eta}_i$.

**Lemma 2** *The first-best social investment given information $J$ is:*

$$y_i^* = \frac{\alpha_B}{2} \frac{\mathbb{E}[\tilde{\eta}_i|J]}{\theta_i} \quad \textit{where} \quad \mathbb{E}[\tilde{\eta}_i|\varepsilon_i] = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta} \tag{8}$$

*At the second-best, the board sets the manager's contract to induce effort and minimize the agency* *cost in equation (9):*

$$\sum_{i=1,2} \theta_i \int_{\eta_i} \int_{\varepsilon_i} \left(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)\right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \tag{9}$$

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 between the firm and the manager or rent extraction.[19] Instead, the agency cost reflects the extent of the inefficiency in resource allocation. As can be seen in equation (9), it measures the deviation between social investments and their first-best levels, which are defined in equation (8).

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 $\theta_i$. 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 (8). The agency cost is proportional to the monetary cost of social investments, as measured by $\theta_i$. Intuitively, if social investments were costless, there would be no tradeoff between social investments and cash flows. 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, investors who set the stock price and shareholders who design the contract do not necessarily have the same preferences with respect to social and environmental factors. Fourth, the sensitivity of the manager's compensation to cash flows 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 cash flows be non-negative.

[19]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)).

**Lemma 3** *Without a nonnegativity constraint on $\beta_x$, 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:*

$$\beta_x = \left(\frac{\alpha_I}{\alpha_B} - 1\right)\beta_p \quad \text{and} \quad \beta_p = \frac{c_e}{\overline{e} - \underline{e}}\left(\frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} - 1\right)^{-1} \tag{10}$$

Intuitively, if the board cares less about the firm's social output than investors ($\alpha_B < \alpha_I$), 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 cash flows-based incentives ($\beta_x > 0$), which discourages costly social investments. On the contrary, if the board cares more about social output than investors ($\alpha_B > \alpha_I$), 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 be achieved by punishing the manager for achieving high cash flows, 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 (8).

Proposition 1 describes the optimal contract when the sensitivity of pay to cash flows must be nonnegative.

**Proposition 1**

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

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

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

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

*and the optimal 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 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 cash flows encourages effort and discourages social investments. A positive sensitivity of pay to ESG scores tends to encourage social investment.[20] A positive sensitivity of pay to the stock price encourages effort and tends to encourage the level of social investment preferred by stock market investors.

It is worth emphasizing that our assumption that the manager is ex-ante aware of the bias in performance measurement, as opposed to imperfect ex-post performance measurement, makes some of the driving forces behind our results different from a standard multitasking model (Holmström and Milgrom (1991)), as further explained below.

To start, consider the optimal contract when the board (which represents the firm's shareholders) and stock market investors have the same preference for social output, i.e. $\alpha_B = \alpha_I$. In this case, the manager's compensation is only contingent on the stock price: $\beta_i = 0$, $\beta_x = 0$, and $\beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)$. In particular, this describes the contract in the special case when neither the board nor stock market investors care about social output ($\alpha_B = \alpha_I = 0$). Intuitively, if stock market investors have the same preference for social output as the board, then the level of social investment preferred by the board is also the one that maximizes the stock price, i.e. it is the one preferred by stock market investors. When managerial compensation is only sensitive to the stock price, the manager optimally chooses the levels of 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 board is less socially responsible than stock market investors

[20]Because of the imperfection of ESG scores, they do not *always* encourage social investment. In the case when the sign of $\varepsilon_i$ is opposite the sign of $\eta_i$, imperfect measurement that can be anticipated ex-ante gives rise to counterproductive incentives. Since ESG scores also affect the stock price, a similar caveat applies to the effect of stock price compensation on social investments.

($\alpha_B < \alpha_I$). In this case, the manager's compensation is only contingent on cash flows and the stock price. A positive sensitivity of compensation to cash flows discourages social investment relative to the level that would maximize the stock price. Since both cash flows-based and stock price-based compensation encourage managerial effort, a positive sensitivity of compensation to cash flows reduces the sensitivity of compensation to the stock price required to elicit managerial effort. In this case, the board can 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 board is more socially responsible than stock market investors ($\alpha_B > \alpha_I$). In this case, the manager's compensation is only 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 be an increasing function of two variables: the average productivity $\bar{\eta}$ of social investments, and the signal $\varepsilon_i$ 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 ratings on investors' 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 cash flows 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 cash flows maximization relative to the board's preference. This cannot be remedied by a negative sensitivity of managerial compensation to cash flows: the nonnegativity constraint on cash flows binds.

This can be partly remedied by also making managerial compensation contingent on the firm's ESG scores: ESG scores-based compensation can be used 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 $\varepsilon_i$. 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 $\varepsilon_i$, 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 $\sigma_\varepsilon$. 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.[21]

ESG scores are noisy not because they are unpredictable from the manager's perspective (instead, they rely on publicly known formulas) but because their formulas make them imperfect indicators of social output. This imperfection combined with their predictability leads to distorted incentives whenever the scores are used directly or indirectly for incentive purposes. For example, when $\varepsilon_i$ 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 $\varepsilon_i$ 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 the ESG score $m_i$.[22]

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 knowledge of the biases of each SEP measure, which results in either excessive or inadequate investment on each dimension. In section B of the Appendix, we study the same setting with multitasking but without gaming (i.e. the manager does not observe $\varepsilon_i$ at the time of making investment decisions), and we show that the first-best outcome can then always be obtained.

[21] This also explains why the nonnegativity constraints on the sensitivity of compensation to ESG scores never bind, as established in the proof of Proposition 1. In an earlier version of the paper in which pay could only be contingent on the stock price and ESG scores, this nonnegativity constraint would bind if and only if $\alpha_B < \alpha_I$, i.e. the board is less concerned about the firm's social output than stock market investors. Intuitively, if stock market investors care more about social output than the board, then a manager who is only compensated based on the stock price will have excessive incentives to invest in social investments from the board's perspective.

[22] Even though measures of an ESG score could be distorted in other ways, our assumptions capture the distortionary effect of ESG scores for social investment.

When the board is more socially responsible than stock market investors ($\alpha_B > \alpha_I$), the firm can either underinvest or overinvest in social investments, depending on the realization of $\tilde{\varepsilon}_i$. Indeed, in this case, the optimal 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 $\alpha_B \gtrless \alpha_I$ is asymmetric. When $\alpha_B \leq \alpha_I$, it is possible to reach the first-best outcome using available compensation instruments – with cash flows-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. Thus, for incentive alignment purposes, it is better for investors to care "too much" rather than "too little" about social output. Indeed, it is easier for the board to align the manager's interests in the former case than in the latter case. 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.

Proposition 1 gives a condition under which, as opposed to relying exclusively on financial performance measures for pay-for-performance, the firm uses ESG scores-based managerial compensation. This happens if and only if the board cares more about social output than stock market investors. Even in this case, Proposition 1 emphasizes that the manager's compensation needs not be highly sensitive to ESG scores. This is for two reasons. First, what matters is the balance of incentives across different dimensions: SEP-based incentives are only required to offset incentives to increase cash flows. Second, when stock market investors care about social output ($\alpha_I > 0$), SEP-based incentives are already embedded in the stock price. When the board cares even more about social output, ESG scores-based incentives are only used as a complement to stock price-based incentives. Moreover, the level of ESG scores-based incentives is proportional to the difference in the preference for social output between the board and stock market investors, as measured by $\alpha_B - \alpha_I$. In sum, 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 is "economically insignificant", is ineffective (Walker (2022)).

In subsections 2.2 and 2.3, we extend the baseline model in two directions. First, we let both the board and investors have heterogeneous preferences with respect to different 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 let the board and investors have a different preference with respect to each social dimension $i$ relative to cash flows, as measured by $\alpha_B^i \geq 0$ for the board, and by $\alpha_I^i \geq 0$ for investors, for $i = 1, 2$ (by contrast, in the baseline model, $\alpha_B^1 = \alpha_B^2$). For simplicity, in this subsection we assume $\sigma_y = 0$.

**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 $i = 1, 2$:*

$$\frac{\beta_i}{\beta_p} = \left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2 + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\right)(\alpha_B^i - \alpha_I^i); \tag{11}$$

*(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 $\alpha_B^i$ 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 (11).*

In part (i), when the board is more socially responsible than investors on all dimensions, there is no cash flows-based compensation.[23] On the contrary, in part (ii), when the board is strictly less

[23] More generally, we show in the proof of Proposition 2 that the firm will not use cash flows-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 $\beta$ 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 $\Gamma_i$ 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

socially responsible than investors on all dimensions, then it uses cash flows-based compensation. In both cases, the intuition is as in the previous section.

In part (iii), the board is much less socially responsible than investors with respect to dimension 1, and slightly more socially responsible than investors on dimension 2. Then, compensation is sensitive to cash flows 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. By contrast, this outcome is not possible in Proposition 1, when an economic agent's preferences for social output are homogeneous across dimensions of social output. This is illustrated in Example 1 below.

**Example 1** *Suppose that* $\alpha_B^1 = 0$, $\alpha_I^1 = 1$, $\alpha_B^2 = 1.1$, $\alpha_I^2 = 1$, $\sigma_\eta = 1$, $\sigma_\varepsilon = 1$, $\sigma_x = 1$, $\sigma_z = 1$, $\bar{\eta} = 2$, $c_e/(\bar{e} - \underline{e}) = 0.1$. *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 is very socially responsible on some dimension $i$, then it uses ESG ratings-based compensation on dimension $i$. Moreover, there is no cash flows-based compensation, even if the board is much less socially responsible than investors on the other dimension ($j$). This is surprising because the board could use cash flows-based compensation to discourage social investments on dimension $j$ ($\beta_x > 0$), and partly offset the effect of cash flows-based compensation on dimension $i$ of SEP by incentivizing investment on this dimension with the corresponding ESG score ($\beta_i > 0$). 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, the contract does not simply become more complex with added performance measures when the divergence in preferences increases. This is illustrated in Example 2 below.

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 cash flows-based compensation to discourage investment in the first dimension – even though it can separately encourage investment in the second dimension by increasing $\beta_2$.

---

optimal ratio is 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 cash flows-based incentives to the manager to counterbalance stock price-based incentives.

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

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: cash flows, the stock price, and some ESG scores.

## 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)). This suggests that additional ESG scores are not redundant.

We now assume that there are $N$ ESG raters. Each 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 $\sigma_\varepsilon^2$ (see section D.3 in the Appendix for additional details). This implies that ESG scores on dimension $i$ are uncorrelated conditional on $\eta_i$, but they are unconditionally positively correlated because of their dependence on $\eta_i$. Let $\bar{\varepsilon}_i = \frac{1}{N}\sum_{j=1}^{N} \varepsilon_i^j$ be the average signal on the firm's social productivity extracted from ESG scores. This average signal is a sufficient statistic for the mean of the distribution (see section D.3).

**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:*

$$\lim_{N\to\infty} y_i(\bar{\varepsilon}_i) = \frac{\alpha_B}{2}\frac{\bar{\varepsilon}_i}{\theta_i} \qquad \text{and} \qquad \lim_{N\to\infty} \mathbb{E}[\tilde{\eta}_i\tilde{y}_i] = \frac{\alpha_B}{2}\frac{\bar{\eta}^2 + \sigma_\eta^2}{\theta_i}. \tag{12}$$

Proposition 3 implies that increasing the number of ESG scores allows to overcome the agency problem. 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. Indeed, each score can be viewed as being biased ex-post, with a bias equal to the realization of each score's noise term, $\varepsilon_i^j - \eta_i$. Moreover, the average score 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 ex-post bias of the average score. This increases the quality of the average score, and diminishes the manager's propensity to game scores. In sum, the manager does not game the ESG scoring system as much when there are more ESG scores. According to Proposition 3, as the number of ESG scores gets very large, the incentives to game the system become negligible, and vanish in the limit.

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 $\eta_i$ and variance $\sigma_{\varepsilon_i}^{j^2}$ (or equivalently precision $1/\sigma_{\varepsilon_i}^{j^2}$); conditional on $\eta_i$, these two scores are correlated with correlation coefficient $\rho \in (-1, 1)$ (see section D.4 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 section D.4 of the Appendix that neither $\varepsilon_i^1$ nor $\varepsilon_i^2$ is a sufficient statistic for $\tilde{\eta}_i$, regardless of the scores' distribution. 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.*

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 (94) in the Appendix. 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. Proposition 4 also shows that the addition of ESG scores can decrease the expected social output of the firm. Intuitively, additional scores affect the stock price and therefore stock price-based incentives, so that the board does not have the option to simply "ignore" additional scores.[24]

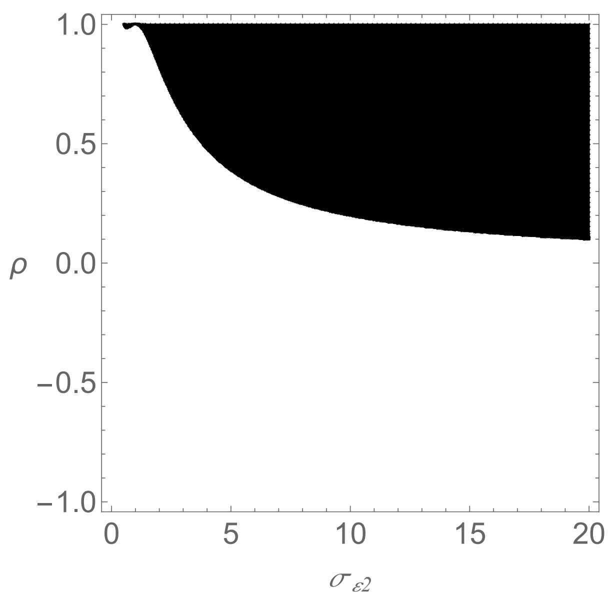


Figure 2: The black area is the subset of parameter values $\{\sigma_{\varepsilon_2}, \rho\}$ such that expected social output $\mathbb{E}[\tilde{\eta}_i \tilde{y}_i]$ on dimension $i$ is higher with one rather than two ESG scores on this dimension. The initial score is characterized by $\sigma_{\varepsilon_1}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

Propositions 3 and 4 have important lessons for SEP measurement "convergence" or "harmo-

[24]In section 1 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 cash flows and ESG scores).

nization", which is frequently advocated and debated.[25] On the one hand, Proposition 3 shows 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. 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.[26]

[25]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 stop 'greenwashing', *Reuters* February 12 2020, Greenwashing And ESG: What You Need To Know, *Forbes* August 25 2022, 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)).

[26]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.

# 3 Conclusion

This paper has studied how to use measures of corporate SEP in a compensation contract, and the effect of these measures on social and environmental investments. Perhaps surprisingly, the need to align interests with respect to this additional dimension does not necessarily require the use of SEP measures in a contract. As long as the board and stock market investors have the same preferences for social output, compensation is only contingent on the stock price. By contrast, when the board and stock market investors have different preferences for social output, the manager's compensation contract is contingent on at least two performance measures: the stock price, and another performance measure which is used to counterbalance the incentives embedded in the stock price. A necessary but insufficient condition for the manager's compensation contract to be simultaneously contingent on the stock price, the firm's profitability, and some SEP measures, is that preferences for social output be heterogeneous across dimensions of SEP.

A divergence between the social preferences of the board and investors does not necessarily make it harder to provide SEP-based incentives. Indeed, the first-best level of social investments can still be induced when the board is less socially responsible than stock market investors, but not in the opposite case. In that latter case, 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.

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.

# References


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.

Beyer, A., Guttman, I., Marinovic, I., 2014. Optimal contracts with performance manipulation. *Journal of Accounting Research*, 52, 817-847.

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

Bonham, J., Riggs-Cragun, A., 2021. Contracting on what firm owners value. Working paper, University of Chicago.

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.

Budde, J., 2007. Performance measure congruity and the balanced scorecard. *Journal of Accounting Research*, 45, 515-539.

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.

Chang, D., Rhee, K., Yoon, A., 2023. ESG integration under asymmetric information. Working paper, CUHK.

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.

Datar, S., Kulp, S.C., Lambert, R.A., 2001. Balancing performance measures. *Journal of Accounting Research*, 39, 75-92.

Dewatripont, M., Jewitt, I., Tirole, J., 2000. Multitask agency problems: Focus and task clustering. *European Economic Review*, 44, 869-877.

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.

Edmans, A., Levit, D., Schneemeier, J., 2022. Socially responsible divestment. Working paper, LBS.

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*, forthcoming.

Flammer, C., Hong, B., Minor, D., 2019. Corporate governance and the rise of integrating corporate social responsibility criteria in executive compensation: Effectiveness and implications for firm outcomes. *Strategic Management Journal*, 40, 1097-1122.

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.

Giese, G., Lee, L.E., Melas, D., Nagy, Z., Nishikawa, L., 2019. Foundations of ESG investing: How ESG affects equity valuation, risk, and performance. *Journal of Portfolio Management*, 45, 69-83.

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

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

Gupta, D., Kopytov, A. and Starmans, J., 2022. The pace of change: Socially responsible investing in private markets. Working paper, Carnegie Mellon University.

Haber, S., Kepler, J.D., Larcker, D.F., Seru, A., Tayan, B., 2022. ESG Investing: What Shareholders Do Fund Managers Represent? Working paper, Stanford University.

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.

Ikram, A., Li, Z.F., Minor, D., 2019. CSR-contingent executive compensation contracts. *Journal of Banking & Finance*, 105655.

Inderst, R., Müller, H.M., 2003. Internal versus external financing: An optimal contracting approach. *Journal of Finance*, 58, 1033-1062.

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.

John, K., Saunders, A., Senbet, L.W., 2000. A theory of bank regulation and management compensation. *Review of Financial Studies*, 13, 95-125.

Kvaløy, O., Olsen, T.E., 2022. Balanced scorecards: A relational contract approach. *Journal of Accounting Research*, forthcoming.

Lambert, R.A., Larcker, D.F., 1987. An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research*, 25, 85-125.

Landier, A., Lovo, S., 2020. ESG investing: How to optimize impact? Working paper, HEC Paris.

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

Lopez de Silanes, F., McCahery, J.A., Pudschedl, P.C, 2022. Institutional Investors and ESG Preferences. Working paper, ECGI.

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.

Paul, J.M., 1992. On the efficiency of stock-based compensation. *Review of Financial Studies*, 5, 471-502.

Pawliczek, A., Carter, M.E., Zhong, R.I., 2023. Say on ESG: The adoption of say-on-pay laws, ESG contracting, and firm ESG performance. Working paper, University of Colorado Boulder.

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

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

Stulz, R., 1990. Managerial discretion and optimal financing policies. *Journal of Financial Economics*, 26, 3-27.

Tirole, J., 2001. Corporate Governance. *Econometrica*, 69, 1-35.

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

Walker, D.I., 2022. The economic (in)significance of executive pay ESG incentives. Working paper, Boston University.

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 diversity, 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."[27] 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)).

## B Case with no gaming

We assume that the manager does not observe $\varepsilon_i$ 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:

$$\mathbb{E}\left[w + \beta_x \left(e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x\right) + \beta_p \tilde{p} + \beta_1 y_1 \varepsilon_1 + \beta_2 y_2 \varepsilon_2 | e\right] - C(e) \tag{13}$$

[27]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:

$$\beta_x \left(-2\theta_i y_i\right) + \beta_p \left(-2\theta_i y_i + \alpha_I \bar{\eta}\right) + \beta_i \bar{\eta} = 0 \quad \Leftrightarrow \quad y_i = \frac{\beta_i + \beta_p \alpha_I}{\beta_x + \beta_p} \frac{\bar{\eta}}{2\theta_i} \tag{14}$$

The first-best optimal value of $y_i$ is:

$$y_i^*(\varepsilon_i) = \alpha_B \frac{\bar{\eta}}{2\theta_i} \tag{15}$$

This can be achieved by setting:

$$\frac{\beta_i + \beta_p \alpha_I}{\beta_x + \beta_p} = \alpha_B. \tag{16}$$

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

$$\beta_x + \beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \geq \frac{c_e}{\bar{e} - \underline{e}} \tag{17}$$

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

$$\beta_x = \frac{c_e}{\bar{e} - \underline{e}} \tag{18}$$

$$\beta_p = 0 \tag{19}$$

$$\beta_i = \alpha_B \beta_x = \alpha_B \frac{c_e}{\bar{e} - \underline{e}} \qquad \text{for } i = 1, 2 \tag{20}$$

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:

$$\beta_x = 0 \tag{21}$$

$$\beta_p = \left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right) \frac{c_e}{\bar{e} - \underline{e}} \tag{22}$$

$$\beta_i = (\alpha_B - \alpha_I) \left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right) \frac{c_e}{\bar{e} - \underline{e}} \qquad \text{for } i = 1, 2 \tag{23}$$

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 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.

$$\max_{e,\beta_x,\beta_p,\beta_1,\beta_2} \mathbb{E}[V(x,y,e)] \quad \text{s.t.} \quad \{e^*, y_1^*, y_2^*\} = \arg\max_{e,y_1,y_2} \mathbb{E}[u(x,y,e)], \quad \text{and } \beta_x \geq 0, \beta_p \geq 0 \tag{24}$$

where:

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

$$\mathbb{E}[u(x,y,e)] = \mathbb{E}\left[w + \beta_x \tilde{x} + \beta_p \tilde{p} + \sum_{i=1,2} \beta_i \tilde{m}_i \middle| \varepsilon_1, \varepsilon_2\right] - C(e) \tag{26}$$

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

$$\mathbb{E}\left[w + \beta_x\left(e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x\right) + \beta_p \tilde{p} + \beta_1 y_1 \varepsilon_1 + \beta_2 y_2 \varepsilon_2 | e\right] - C(e) \tag{27}$$

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:

$$\beta_x(-2\theta_i y_i) + \beta_p\left(-2\theta_i y_i + \alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta}\right)\right) + \beta_i \varepsilon_i = 0 \tag{28}$$

$$\Leftrightarrow \quad y_i = \frac{\beta_i \varepsilon_i + \beta_p \alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)}{\beta_x + \beta_p}\frac{1}{2\theta_i} = \frac{\left(\beta_i + \beta_p \alpha_I \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\varepsilon_i + \beta_p \alpha_I \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}}{\beta_x + \beta_p}\frac{1}{2\theta_i} \tag{29}$$

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

$$\begin{aligned}
& \mathbb{E}\left[u(x,y,e)|e = \overline{e}\right] - c_e \geq \mathbb{E}\left[u(x,y,e)|e = \underline{e}\right] \\
\Leftrightarrow\ & w + \beta_x \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x | e = \overline{e}\right] + \beta_p \mathbb{E}\left[\mu_{x|z} + \alpha_I \mu_{y|z} \middle| e = \overline{e}\right] + \beta_1 y_1 \varepsilon_1 + \beta_2 y_2 \varepsilon_2 - c_e \\
& \geq w + \beta_x \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x | e = \underline{e}\right] + \beta_p \mathbb{E}\left[\mu_{x|z} + \alpha_I \mu_{y|z} \middle| e = \underline{e}\right] + \beta_1 y_1 \varepsilon_1 + \beta_2 y_2 \varepsilon_2 \\
\Leftrightarrow\ & \beta_x + \beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \geq \frac{c_e}{\overline{e} - \underline{e}}
\end{aligned} \tag{30}$$

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

$$\begin{aligned} \mathbb{E}[u(x,y,e)|e] &= \bar{W} \\ \Leftrightarrow \quad w + \beta_x \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x | e\right] + \beta_p \mathbb{E}[\tilde{p}|e] + \sum_{i=1,2} \beta_i \mathbb{E}[\tilde{m}_i] &= \bar{W} + C(e) \end{aligned} \tag{31}$$

Thus, equation (25) can be rewritten as:

$$\begin{aligned} \mathbb{E}[V(x,y,e)] &= \mathbb{E}\left[(1-\beta_x)\left(e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x\right) + \alpha_B\left(\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y\right) - \left(w + \beta_p \tilde{p} + \sum_{i=1,2} \beta_i \tilde{m}_i\right)\right] \\ &= \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x + \alpha_B\left(\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y\right) - \left(\bar{W} + C(e)\right) | e\right] \end{aligned} \tag{32}$$

We derive the "first-best" outcome without an agency problem (i.e. the incentive constraint in equation (30) can be ignored):

$$\begin{aligned} & \max_{y_1, y_2} \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x + \alpha_B\left(\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y\right) - \left(\bar{W} + C(e)\right) | \varepsilon_1, \varepsilon_2\right] \\ \Leftrightarrow \quad & \max_{y_1, y_2} \left\{e - \theta_1 y_1^2 - \theta_2 y_2^2 + \alpha_B\left(\mathbb{E}\left[\tilde{\eta}_1 | \varepsilon_1\right] y_1 + \mathbb{E}\left[\tilde{\eta}_2 | \varepsilon_2\right] y_2\right) - \left(\bar{W} + C(e)\right)\right\} \end{aligned} \tag{33}$$

where

$$\mathbb{E}\left[\tilde{\eta}_i | \varepsilon_i\right] = \mathbb{E}\left[\tilde{\eta}_i\right] + \frac{cov(\tilde{\eta}_i, \tilde{\varepsilon}_i)}{var(\tilde{\varepsilon}_i)}(\varepsilon_i - \mathbb{E}[\tilde{\varepsilon}_i]) = \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}(\varepsilon_i - \bar{\eta}) = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta} \tag{34}$$

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

$$y_i^*(\varepsilon_i) = \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta}\right) \tag{35}$$

We now consider the second-best outcome with an agency problem. Denote:

$$f(\varepsilon_i, y_i) \equiv -\theta_i y_i^2 + \alpha_B \mathbb{E}\left[\tilde{\eta}_i | \varepsilon_i\right] y_i \tag{36}$$

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

$$f_y(\varepsilon_i, y_i) = -2\theta_i y_i + \alpha_B \mathbb{E}\left[\tilde{\eta}_i | \varepsilon_i\right] \tag{37}$$

$$f_{yy}(\varepsilon_i, y_i) = -2\theta_i \tag{38}$$

The objective function of the board can be written as:

$$\mathbb{E}[V(x,y,e)] = e + \mathbb{E}\left[\sum_{i=1,2} f(\varepsilon_i, y_i)\right] - \left(\bar{W} + C(e)\right) \tag{39}$$

where

$$\mathbb{E}\left[\sum_{i=1,2} f(\varepsilon_i, y_i)\right] = \sum_{i=1,2} \int_{\eta_i} \int_{\varepsilon_i} f(\varepsilon_i, y_i)\varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \tag{40}$$

Thus, for a given effort $e$ and a given $\varepsilon_i$, the value of $y_i$ that maximizes the board's objective function is the value of $y_i$ that maximizes the expression in equation (40). By definition of $y_i^*(\varepsilon_i)$, for a given $\varepsilon_i$, the function $f(\varepsilon_i, y_i)$ is maximized by setting $y_i(\varepsilon_i) = y_i^*(\varepsilon_i)$ for $i = 1, 2$. Accordingly, define the loss function for a given $\varepsilon_i$ as $f(\varepsilon_i, y_i^*(\varepsilon_i)) - f(\varepsilon_i, y_i)$. The value of $y_i$ that maximizes the board's objective function is the value of $y_i$ that maximizes:

$$\max_{y_i} \mathbb{E}\left[\sum_{i=1,2} \left(f(\varepsilon_i, y_i) - f(\varepsilon_i, y_i^*(\varepsilon_i))\right)\right] = \sum_{i=1,2} \mathbb{E}\left[\left(f(\varepsilon_i, y_i) - f(\varepsilon_i, y_i^*(\varepsilon_i))\right)\right] \tag{41}$$

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

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

Thus:

$$\begin{aligned} f(\varepsilon_i, y_i) - f(\varepsilon_i, y_i^*(\varepsilon_i)) &= \left(-2\theta_i y_i^*(\varepsilon_i) + \alpha_B\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta}\right)\right)(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)) - \theta_i(y_i(\varepsilon_i) - y_i^*(\varepsilon_i))^2 \\ &= 2\theta_i\left(\left(-y_i^*(\varepsilon_i) + \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\bar{\eta}\right)\right)(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)) - \frac{1}{2}(y_i(\varepsilon_i) - y_i^*(\varepsilon_i))^2\right) \\ &= 2\theta_i\left(-\frac{1}{2}(y_i(\varepsilon_i) - y_i^*(\varepsilon_i))^2\right) \\ &= -\theta_i\left(y_i(\varepsilon_i) - y_i^*(\varepsilon_i)\right)^2 \end{aligned} \tag{42}$$

■

**Proof of Lemma 3 and Proposition 1:**

Plugging the optimal investment $y_i$ from equation (29) in the optimization problem using Lemma 2, gives:

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

$$\Leftrightarrow \quad \min_{\beta_x,\beta_p,\beta_i} \sum_{i=1,2} \theta_i \int_{\eta_i} \int_{\varepsilon_i} \left( \left( \frac{\beta_i}{\beta_x+\beta_p} + \left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\varepsilon_i + \left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B\right)\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \quad (43)$$

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

Without nonnegativity constraints on contracting, simply set $\beta_i = 0$ and $\frac{\beta_p}{\beta_x+\beta_p}\alpha_I = \alpha_B \Leftrightarrow \beta_x = \left(\frac{\alpha_I}{\alpha_B} - 1\right)\beta_p$, so that the expression under the integral sign in equation (43) is zero for any $\varepsilon_i$. Since this expression (a quadratic function) is nonnegative for any $\varepsilon_i$, achieving a value of zero for this expression at any $\varepsilon_i$ 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 (42)). To elicit high effort, use equation (30) to set:

$$\left(\frac{\alpha_I}{\alpha_B} - 1\right)\beta_p + \beta_p\frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} = \frac{c_e}{\bar{e}-\underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(\frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} - 1\right)^{-1} \quad (44)$$

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

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

$$\int_{\eta_i}\int_{\varepsilon_i} \varepsilon_i\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i = \int_{\eta_i}\left(\int_{\varepsilon_i}\varepsilon_i\varphi(\varepsilon_i|\eta_i)d\varepsilon_i\right)\phi(\eta_i)d\eta_i = \int_{\eta_i}\eta_i\phi(\eta_i)d\eta_i = \bar{\eta} \quad (45)$$

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:

$$\int_{\eta_i}\int_{\varepsilon_i} \varepsilon_i^2\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i = \int_{\eta_i}\left(\sigma_\varepsilon^2+\eta_i^2\right)\phi(\eta_i)d\eta_i = \sigma_\varepsilon^2 + \int_{\eta_i}\eta_i^2\phi(\eta_i)d\eta_i = \sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2 \quad (46)$$

The expression in equation (43) is globally convex with respect to $\beta_i$. The FOC w.r.t. $\beta_i$ is:

$$
\begin{aligned}
& \theta_i \int_{\eta_i}\int_{\varepsilon_i} \frac{2}{\beta_x+\beta_p}\varepsilon_i \left( \left( \frac{\beta_i}{\beta_x+\beta_p} + \left( \frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B \right) \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \varepsilon_i + \left( \frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta} \right) \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i = 0 \\
\Leftrightarrow \quad & \left( \frac{\beta_i}{\beta_x+\beta_p} + \left( \frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B \right) \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) (\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2) = \left( \alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p} \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}^2 = 0 \\
\Leftrightarrow \quad & \beta_i = (\alpha_B(\beta_x+\beta_p) - \beta_p\alpha_I) \left( \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \qquad (47)
\end{aligned}
$$

We verify later that the nonnegativity constraint on $\beta_i$ does not bind. Substituting for $\frac{\beta_i}{\beta_x+\beta_p}$ in equation (43):

$$
\begin{aligned}
& \sum_{i=1,2} \theta_i \int_{\eta_i}\int_{\varepsilon_i} \left( \left( \left( \alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p} \right) \left( \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) + \left( \frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B \right) \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \varepsilon_i \right. \\
& \left. + \left( \frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \\
= \quad & \left( \alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p} \right)^2 \frac{\sigma_\varepsilon^4}{(\sigma_\eta^2+\sigma_\varepsilon^2)^2} \sum_{i=1,2} \theta_i \int_{\eta_i}\int_{\varepsilon_i} \left( \frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2}\varepsilon_i - \bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \qquad (48)
\end{aligned}
$$

The expression under the integral sign is positive and independent from the contract. The expression in equation (48) is minimized by minimizing $\left( \alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p} \right)^2$. With a nonnegativity constraint on $\beta_x$, this is achieved by setting $\beta_x = \left( \frac{\alpha_I}{\alpha_B} - 1 \right) \beta_p$ if $\alpha_B \leq \alpha_I$, and $\beta_x = 0$ otherwise.

In the case $\alpha_B \leq \alpha_I$, we have $\beta_x = \left( \frac{\alpha_I}{\alpha_B} - 1 \right) \beta_p$ so that $\beta_i = 0$ from equation (47). Substituting for $\beta_i$ in equation (43):

$$
\sum_{i=1,2} \theta_i \int_{\eta_i}\int_{\varepsilon_i} \left( \left( \frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B \right) \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\varepsilon_i + \left( \frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \qquad (49)
$$

Also substituting for $\beta_x$ shows that the expression under the integral sign of equation (49) is zero for any $\varepsilon_i$. Since this expression (a quadratic function) is nonnegative for any $\varepsilon_i$, achieving a value of zero for this expression at any $\varepsilon_i$ 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 (42)).

Still in the case $\alpha_B \leq \alpha_I$, to elicit high effort, use equation (30) to set:

$$
\left( \frac{\alpha_I}{\alpha_B} - 1 \right) \beta_p + \beta_p \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} = \frac{c_e}{\bar{e} - \underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\bar{e} - \underline{e}} \left( \frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} - 1 \right)^{-1}, \qquad (50)
$$

where $\beta_p > 0$ since $\alpha_B \leq \alpha_I$. In this case, $\alpha_B(\beta_x+\beta_p) - \alpha_I\beta_p = \alpha_B \left( \frac{\alpha_I}{\alpha_B} - 1 + 1 - \frac{\alpha_I}{\alpha_B} \right) \beta_p = 0$ so that $\beta_i$

as defined as in equation (47) is equal to zero. Expected social output when $\alpha_B \leq \alpha_I$ is:

$$\begin{aligned}
\mathbb{E}[\tilde{\eta}_i \tilde{y}_i] &= \frac{\beta_p \alpha_I \left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\mathbb{E}[\tilde{\eta}_i\tilde{\varepsilon}_i] + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\mathbb{E}[\tilde{\eta}_i]\right)}{\beta_x+\beta_p}\frac{1}{2\theta_i} \\
&= \frac{\alpha_B}{2\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\bar{\eta}^2+\sigma_\eta^2\right) + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}^2\right) \\
&= \frac{\alpha_B}{2\theta_i}\left(\frac{\sigma_\eta^4}{\sigma_\eta^2+\sigma_\varepsilon^2} + \bar{\eta}^2\right) \qquad (51)
\end{aligned}$$

In the case $\alpha_B > \alpha_I$, we have $\beta_x = 0$ and $\beta_i$ is as in equation (47). From equations (42) and (48) we have:

$$\begin{aligned}
f(\varepsilon_i, y_i) - f(\varepsilon_i, y_i^*(\varepsilon_i)) &= \theta_i\left(\left(\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2}\varepsilon_i - \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)\right)^2 \\
&= \theta_i\left(\alpha_B-\alpha_I\right)^2\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2}\varepsilon_i - \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\right)^2 \qquad (52)
\end{aligned}$$

and

$$\begin{aligned}
y_i(\varepsilon_i) - y_i^*(\varepsilon_i) &= \underbrace{(\alpha_B-\alpha_I)}_{>0}\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2}\varepsilon_i - \bar{\eta}\right) \\
\Rightarrow \quad \mathbb{E}[y_i(\tilde{\varepsilon}_i)] - \mathbb{E}[y_i^*(\tilde{\varepsilon}_i)] &= \underbrace{(\alpha_B-\alpha_I)}_{>0}\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} - 1\right)\bar{\eta} \qquad (53)
\end{aligned}$$

Still in the case $\alpha_B > \alpha_I$, to elicit high effort, use equation (30) to set:

$$\beta_p\frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} = \frac{c_e}{\bar{e}-\underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(1+\frac{\sigma_z^2}{\sigma_x^2}\right) \qquad (54)$$

In this case, $\alpha_B(\beta_x+\beta_p) - \alpha_I\beta_p = \beta_p(\alpha_B-\alpha_I) > 0$ so that $\beta_i > 0$ and is defined as in equation (47). Substituting for $\beta_x$ and $\beta_p$ in equation (47), in this case we have:

$$\beta_i = (\alpha_B-\alpha_I)\frac{c_e}{\bar{e}-\underline{e}}\left(1+\frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right) \qquad (55)$$

Expected social output when $\alpha_B > \alpha_I$ is:

$$\begin{aligned}
\mathbb{E}[\tilde{\eta}_i\tilde{y}_i] &= \frac{(\alpha_B-\alpha_I)\frac{c_e}{\bar{e}-\underline{e}}\left(1+\frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\mathbb{E}[\tilde{\eta}_i\tilde{\varepsilon}_i] + \beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\mathbb{E}[\tilde{\eta}_i\tilde{\varepsilon}_i] + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\bar{\eta}\mathbb{E}[\tilde{\eta}_i]\right)}{\beta_x+\beta_p}\frac{1}{2\theta_i} \\
&= \frac{1}{2\theta_i}\left((\alpha_B-\alpha_I)\left(\frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\right)\left(\bar{\eta}^2+\sigma_\eta^2\right) + \alpha_I\left(\frac{\sigma_\eta^4}{\sigma_\eta^2+\sigma_\varepsilon^2} + \bar{\eta}^2\right)\right) \qquad (56)
\end{aligned}$$

■

**Proof of Proposition 2:**

The objective function of the board in equation (25) now rewrites as:

$$\mathbb{E}[V(x,y,e)] \quad = \quad \mathbb{E}\left[\tilde{x} + \sum_{i=1,2} \alpha_B^i \tilde{\eta}_i \tilde{y}_i - \left(w + \beta_x \tilde{x} + \beta_p \tilde{p} + \sum_{i=1,2} \beta_i \tilde{m}_i\right)\right] \tag{57}$$

The first part of the proof involves straightforward notational adjustments to the proof of Proposition 1. Equation (43) rewrites as:

$$\min_{\beta_x,\beta_p,\beta_i} \sum_{i=1,2} \int_{\eta_i} \int_{\varepsilon_i} \left( \left( \frac{\beta_i}{\beta_x+\beta_p} + \left( \frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} - \alpha_B^i \right) \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \varepsilon_i + \left( \frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} - \alpha_B^i \right) \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \tag{58}$$

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

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

$$\Leftrightarrow \quad \frac{\beta_i}{\beta_x+\beta_p} = \left( \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \right) \left( \alpha_B^i - \frac{\beta_p \alpha_I^i}{\beta_x+\beta_p} \right) \tag{59}$$

Given the nonnegativity constraint and the global convexity of the objective function with respect to $\beta_i$, there are two cases. If equation (59) gives a positive $\beta_i$ given the optimal values of $\beta_x$ and $\beta_p$ derived below, then the optimum is $\beta_i^* = \beta_i$ as in equation (59). If equation (59) gives a nonpositive $\beta_i$, then the optimum is $\beta_i^* = 0$.

Define:

$$\gamma_i \quad \equiv \quad \int_{\eta_i} \int_{\varepsilon_i} \left( \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \varepsilon_i + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2} \bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i \tag{60}$$

$$g_i \quad \equiv \quad \frac{\sigma_\varepsilon^4}{(\sigma_\eta^2+\sigma_\varepsilon^2)^2} \int_{\eta_i} \int_{\varepsilon_i} \left( \frac{\bar{\eta}^2}{\sigma_\varepsilon^2+\sigma_\eta^2+\bar{\eta}^2} \varepsilon_i - \bar{\eta} \right)^2 \varphi(\varepsilon_i|\eta_i)\phi(\eta_i) d\varepsilon_i d\eta_i, \tag{61}$$

Both $\gamma_i$ 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 $\eta_i$ is independent from $i$, the variables $\gamma_i$ and $g_i$ are independent from $i$ and independent from the contract. Let $\Gamma_i = \gamma_i$ if the nonnegativity constraint on $\beta_i$ is binding, and $\Gamma_i = g_i$ if the nonnegativity constraint on $\beta_i$ 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 $\beta_i$ in equation (58), the optimization problem is:

$$\min_{\beta_x,\beta_p} \sum_{i=1,2} \Gamma_i \left( \frac{\beta_p}{\beta_x+\beta_p} \alpha_I^i - \alpha_B^i \right)^2 \tag{62}$$

where $\Gamma_i = g_i$ in case 1 (we henceforth keep the $\Gamma_i$ notation because we will refer to equations below for cases other than case 1). This is a sum weighted by $\Gamma_i$ 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 $\beta_p$, this is equivalent to choosing:

$$\min_{\beta} \sum_{i=1,2} \Gamma_i \left( \alpha_B^i - \beta \alpha_I^i \right)^2 \quad \text{s.t.} \quad \beta \leq 1 \tag{63}$$

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

$$\mathcal{L} = \sum_{i=1,2} \Gamma_i \left( \alpha_B^i - \beta \alpha_I^i \right)^2 + \delta(\beta - 1) \tag{64}$$

The FOC with respect to $\beta$ is:

$$-2 \sum_{i=1,2} \Gamma_i \alpha_I^i \left( \alpha_B^i - \beta \alpha_I^i \right) + \delta = 0 \quad \Leftrightarrow \quad \beta = \frac{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i - \delta}{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2} \tag{65}$$

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

$$\sum_{i=1,2} \Gamma_i \alpha_I^i \left( \alpha_B^i - \alpha_I^i \right) > 0 \tag{66}$$

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 $\beta_p$ is determined according to incentive compatibility with respect to the manager's effort. Substituting in equation (30):

$$\beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} = \frac{c_e}{\overline{e} - \underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\overline{e} - \underline{e}} \left( 1 + \frac{\sigma_z^2}{\sigma_x^2} \right) \tag{67}$$

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

$$\sum_{i=1,2} \Gamma_i \alpha_I^i \left( \alpha_B^i - \beta \alpha_I^i \right) = 0 \quad \Leftrightarrow \quad \beta = \frac{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i}{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2} \quad \Leftrightarrow \quad \beta_x = \beta_p \left( \frac{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2}{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i} - 1 \right) \tag{68}$$

This equation gives the optimal value of $\beta_x$ as long as this value is nonnegative, so that the nonnegativity constraint is nonbinding (i.e. $\delta = 0$). The value of $\beta_p$ is determined according to incentive compatibility with respect to the manager's effort. Substituting in equation (30):

$$\beta_p \left( \frac{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2}{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i} - 1 \right) + \beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} = \frac{c_e}{\bar{e} - \underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\bar{e} - \underline{e}} \Big/ \left( \frac{\sum_{i=1,2} \Gamma_i {\alpha_I^i}^2}{\sum_{i=1,2} \Gamma_i \alpha_I^i \alpha_B^i} - 1 + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \right) \tag{69}$$

Finally, we verify the conjecture that $\beta_i > 0$ for $i = 1, 2$, by plugging $\beta_x$ and $\beta_p$ thus derived into equation (59). 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 $\Gamma_i$. 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 $\Gamma_i$.

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 $\beta_i$ is as in equation (59) 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 (66) holds, so that $\beta_x = 0$, and $\beta_p$ is as in equation (67).

(ii) If $\alpha_B^i < \alpha_I^i$ for $i = 1, 2$, the inequality in equation (66) does not hold, so that $\beta_x > 0$, and $\beta_p$ is as in equation (69). Given these values of $\beta_x$ and $\beta_p$, $\beta_i$ is as in equation (59) 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 (66) 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 $\beta_p$ is as in equation (69). 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 (68):

$$\alpha_B^i - \frac{\beta_p}{\beta_x + \beta_p} \alpha_I^i = \alpha_B^i - \beta \alpha_I = \alpha_B^i - \frac{\sum_{k=1,2} \Gamma_k \alpha_I^2 \alpha_B^k}{\sum_{k=1,2} \Gamma_k \alpha_I^2} = \alpha_B^i - \frac{\sum_{k=1,2} \Gamma_k \alpha_B^k}{\sum_{k=1,2} \Gamma_k} \tag{70}$$

By contradiction, suppose that $\beta_1 > 0$ so that $\Gamma_1 = \Gamma_2 \equiv \Gamma$. Then, substituting in equation (70):

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

so that from equation (59) we would have $\beta_1 < 0$, a contradiction. Thus, $\beta_1 = 0$.

(iv) If $\alpha_B^i$ is sufficiently large and $\alpha_I^i > 0$, then $\beta_i$ is as in equation (59) such that $\beta_i > 0$ even without the nonnegativity constraint, i.e. this constraint does not bind for $\beta_i$, and we have $\Gamma_i = g_i$. Moreover, when $\alpha_B^i$ is sufficiently large and $\alpha_I^i > 0$, the inequality in equation (66) 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 $\beta_p$ is as in equation (67). ■

**Proof of Proposition 3:**

The steps are similar to the ones in Proposition 1.

A preliminary step is to establish that the average ESG score is a sufficient statistic: see section D.3. 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 (27) accordingly gives:

$$\arg\max_{e,y_1,y_2}\mathbb{E}\left[w + \beta_x\left(e - \theta_1y_1^2 - \theta_2y_2^2\right) + \beta_p\tilde{p} + \beta_1\bar{m}_1 + \beta_2\bar{m}_2\right] - C(e) \tag{71}$$

where the stock price $\tilde{p}$ is as in section D.3.

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:

$$\beta_x\left(-2\theta_iy_i\right) + \beta_p\left(-2\theta_iy_i + \alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{\varepsilon}_i + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\bar{\eta}\right)\right) + \beta_i\bar{\varepsilon}_i = 0 \tag{72}$$

$$\Leftrightarrow \quad y_i = \frac{\beta_i\bar{\varepsilon}_i + \beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)}{\beta_x + \beta_p}\frac{1}{2\theta_i} = \frac{\left(\beta_i + \beta_p\alpha_I\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i + \beta_p\alpha_I\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}}{\beta_x + \beta_p}\frac{1}{2\theta_i} \tag{73}$$

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

$$\begin{aligned}
& \mathbb{E}\left[u(x,y,e)|e=\overline{e}\right]-c_e \geq \mathbb{E}\left[u(x,y,e)|e=\underline{e}\right] \\
\Leftrightarrow \quad & w+\beta_x\mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x|e=\overline{e}\right]+\beta_p\mathbb{E}\left[\mu_{x|z,\bar{m}}+\alpha_I\mu_{y|z,\bar{m}}\middle|e=\overline{e}\right]+\beta_1 y_1\bar{\varepsilon}_1+\beta_2 y_2\bar{\varepsilon}_2-c_e \\
& \geq w+\beta_x\mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x|e=\underline{e}\right]+\beta_p\mathbb{E}\left[\mu_{x|z,\bar{m}}+\alpha_I\mu_{y|z,\bar{m}}\middle|e=\underline{e}\right]+\beta_1 y_1\bar{\varepsilon}_1+\beta_2 y_2\bar{\varepsilon}_2 \\
\Leftrightarrow \quad & \beta_x+\beta_p\frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2}\geq\frac{c_e}{\overline{e}-\underline{e}} \qquad (74)
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}[u(x,y,e)|e] &= \bar{W} \\
\Leftrightarrow \quad w+\beta_x\mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x|e\right]+\beta_p\mathbb{E}[\tilde{p}|e]+\sum_{i=1,2}\beta_i\mathbb{E}[\tilde{\bar{m}}_i] &= \bar{W}+C(e) \qquad (75)
\end{aligned}$$

Thus, equation (25) can be rewritten as:

$$\begin{aligned}
\mathbb{E}[V(x,y,e)] &= \mathbb{E}\left[(1-\beta_x)\left(e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x\right)+\alpha_B\left(\tilde{\eta}_1 y_1+\tilde{\eta}_2 y_2+\tilde{\epsilon}_y\right)-\left(w+\beta_p\tilde{p}+\sum_{i=1,2}\beta_i\tilde{\bar{m}}_i\right)\right] \\
&= \mathbb{E}\left[e-\theta_1 y_1^2-\theta_2 y_2^2+\tilde{\epsilon}_x+\alpha_B\left(\tilde{\eta}_1 y_1+\tilde{\eta}_2 y_2+\tilde{\epsilon}_y\right)-\left(\bar{W}+C(e)\right)|e\right] \qquad (76)
\end{aligned}$$

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 $y_i(\bar{\varepsilon}_i)$ and $y_i^*(\bar{\varepsilon}_i)$, where:

$$\begin{aligned}
\mathbb{E}\left[\tilde{\eta}_i|\bar{\varepsilon}_i\right] &= \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta} \qquad (77) \\
y_i^*(\bar{\varepsilon}_i) &= \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right) \qquad (78)
\end{aligned}$$

This problem is:

$$\begin{aligned}
& \min_{\beta_x,\beta_p,\beta_i}\sum_{i=1,2}\int_{\eta_i}\int_{\bar{\varepsilon}_i}\left(\frac{1}{2}\frac{1}{\theta_i}\frac{\left(\beta_i+\beta_p\alpha_I\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i+\beta_p\alpha_I\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}}{\beta_x+\beta_p}-\frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)\right)^2\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \\
\Leftrightarrow \quad & \min_{\beta_x,\beta_p,\beta_i}\sum_{i=1,2}\int_{\eta_i}\int_{\bar{\varepsilon}_i}\left(\left(\frac{\beta_i}{\beta_x+\beta_p}+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i+\left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p}-\alpha_B\right)\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)^2\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \qquad (79)
\end{aligned}$$

We have:

$$\int_{\eta_i}\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i=\int_{\eta_i}\left(\int_{\bar{\varepsilon}_i}\bar{\varepsilon}_i\varphi(\bar{\varepsilon}_i|\eta_i)d\bar{\varepsilon}_i\right)\phi(\eta_i)d\eta_i=\int_{\eta_i}\eta_i\phi(\eta_i)d\eta_i=\bar{\eta} \qquad (80)$$

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:

$$\int_{\eta_i}\int_{\bar{\varepsilon}_i} \bar{\varepsilon}_i^2 \varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i) d\bar{\varepsilon}_i d\eta_i = \int_{\eta_i} \left(\sigma_\varepsilon^2/N + \eta_i^2\right)\phi(\eta_i) d\eta_i = \sigma_\varepsilon^2/N + \int_{\eta_i} \eta_i^2 \phi(\eta_i) d\eta_i = \sigma_\varepsilon^2/N + \sigma_\eta^2 + \bar{\eta}^2 \quad (81)$$

The expression in equation (79) is globally convex with respect to $\beta_i$. The FOC w.r.t. $\beta_i$ is:

$$\begin{aligned}
&\left(\frac{\beta_i}{\beta_x+\beta_p} + \left(\frac{\beta_p\alpha_I}{\beta_x+\beta_p} - \alpha_B\right)\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\int_{\eta_i}\int_{\bar{\varepsilon}_i} \bar{\varepsilon}_i^2\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \\
&= \left(\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\int_{\eta_i}\int_{\bar{\varepsilon}_i} \bar{\varepsilon}_i\varphi(\bar{\varepsilon}_i|\eta_i)\phi(\eta_i)d\bar{\varepsilon}_i d\eta_i \\
\Leftrightarrow\quad & \frac{\beta_i}{\beta_x+\beta_p} = \left(\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right) \qquad (82)
\end{aligned}$$

Substituting in equation (79) gives:

$$\min_{\beta_x,\beta_p}\left(\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)^2\frac{\sigma_\varepsilon^4}{(\sigma_\eta^2+\sigma_\varepsilon^2)^2/N}\sum_{i=1,2}\int_{\eta_i}\int_{\varepsilon_i}\left(\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2}\varepsilon_i - \bar{\eta}\right)^2\varphi(\varepsilon_i|\eta_i)\phi(\eta_i)d\varepsilon_i d\eta_i \qquad (83)$$

The expression under the integral sign is positive and independent from the contract. The expression in equation (48) is minimized by minimizing $\left(\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)^2$. With a nonnegativity constraint on $\beta_x$, this is achieved by setting $\beta_x = \left(\frac{\alpha_I}{\alpha_B} - 1\right)\beta_p$ if $\alpha_B \le \alpha_I$, and $\beta_x = 0$ otherwise.

In the former case ($\alpha_B \le \alpha_I$), to elicit high effort, use equation (74) to set:

$$\left(\frac{\alpha_I}{\alpha_B} - 1\right)\beta_p + \beta_p\frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} = \frac{c_e}{\bar{e}-\underline{e}} \quad\Leftrightarrow\quad \beta_p = \frac{c_e}{\bar{e}-\underline{e}}\left(\frac{\alpha_I}{\alpha_B} + \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2} - 1\right)^{-1} \qquad (84)$$

In this case, $\alpha_B(\beta_x+\beta_p) - \alpha_I\beta_p = \alpha_B\left(\frac{\alpha_I}{\alpha_B} - 1 + 1 - \frac{\alpha_I}{\alpha_B}\right)\beta_p = 0$ so that $\beta_i$ as defined as in equation (82) is equal to zero. 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 (78). Moreover, expected social output when $\alpha_B \le \alpha_I$ is:

$$\mathbb{E}[\tilde{\eta}_i\tilde{y}_i] = \frac{\beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\mathbb{E}[\tilde{\eta}_i\tilde{\bar{\varepsilon}}_i] + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\mathbb{E}[\tilde{\eta}_i]\right)}{\beta_x+\beta_p}\frac{1}{2\theta_i} = \frac{\alpha_B}{2\theta_i}\left(\frac{\sigma_\eta^4}{\sigma_\eta^2+\sigma_\varepsilon^2/N} + \bar{\eta}^2\right) \qquad (85)$$

In the limit:

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

In the latter case ($\alpha_B > \alpha_I$), to elicit high effort, use equation (74) to set:

$$\beta_p \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} = \frac{c_e}{\bar{e} - \underline{e}} \quad \Leftrightarrow \quad \beta_p = \frac{c_e}{\bar{e} - \underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right) \tag{86}$$

In this case, $\alpha_B(\beta_x + \beta_p) - \alpha_I \beta_p = \beta_p(\alpha_B - \alpha_I) > 0$ so that $\beta_i > 0$ and is defined as in equation (47). Substituting for $\beta_x$ and $\beta_p$ in equation (82), in this case we have:

$$\beta_i = (\alpha_B - \alpha_I)\frac{c_e}{\bar{e} - \underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\right) \tag{87}$$

In the case $\alpha_B > \alpha_I$, substituting for $\beta_i$ in equation (73):

$$y_i(\bar{\varepsilon}_i) = \frac{(\alpha_B - \alpha_I)\frac{c_e}{\bar{e}-\underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\bar{\varepsilon}_i + \beta_p \alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\varepsilon}_i + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\right)}{\beta_x + \beta_p}\frac{1}{2\theta_i}, \tag{88}$$

where $\beta_x$ and $\beta_p$ are as in the proof of Proposition 1. Substituting for $\beta_x$ and $\beta_p$, investment in dimension $i$ of SEP is:

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

In the limit: $\lim_{N\to\infty}(y_i(\bar{\varepsilon}_i) - y_i^*(\bar{\varepsilon}_i)) = 0$, as defined in equation (78). Moreover, expected social output when $\alpha_B > \alpha_I$ is:

$$\begin{aligned}
\mathbb{E}[\tilde{\eta}_i \tilde{y}_i] &= \frac{(\alpha_B - \alpha_I)\frac{c_e}{\bar{e}-\underline{e}}\left(1 + \frac{\sigma_z^2}{\sigma_x^2}\right)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N+\sigma_\eta^2+\bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\right)\mathbb{E}[\tilde{\eta}_i\tilde{\bar{\varepsilon}}_i] + \beta_p\alpha_I\left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\mathbb{E}[\tilde{\eta}_i\tilde{\bar{\varepsilon}}_i] + \frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{\eta}\mathbb{E}[\tilde{\eta}_i]\right)}{\beta_x + \beta_p}\frac{1}{2\theta_i} \\
&= \left((\alpha_B - \alpha_I)\left(\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\frac{\bar{\eta}^2}{\sigma_\varepsilon^2/N + \sigma_\eta^2 + \bar{\eta}^2} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\right)\left(\bar{\eta}^2 + \sigma_\eta^2\right) + \alpha_I\left(\bar{\eta}^2 + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2/N}\sigma_\eta^2\right)\right)\frac{1}{2\theta_i} \qquad (89)
\end{aligned}$$

In the limit:

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

■

**Proof of Proposition 4:**

The steps are similar to the ones in Proposition 1.

Rewriting the manager's objective function in equation (27) in the case with two sets of ESG scores gives:

$$\arg\max_{e,y_1,y_2} \mathbb{E}\left[w + \beta_x\left(e - \theta_1 y_1^2 - \theta_2 y_2^2\right) + \beta_p \tilde{p} + \sum_{i=1,2}\sum_{k=1,2} \beta_i^k m_i^k\right] - C(e) \tag{90}$$

where the stock price $\tilde{p}$ is as in section D.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 (for $j \neq k$):

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

$$\Leftrightarrow \quad y_i = \frac{\sum_{k=1,2}\beta_i^k\varepsilon_i^k + \beta_p\alpha_I\left(\bar{\eta} + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right)\right)}{\beta_x + \beta_p}\frac{1}{2\theta_i} \tag{91}$$

Given his contract, the manager will optimally exert high effort ($e = \overline{e}$) if and only if equation (74) is satisfied. The fixed component of pay, $w$, is set to guarantee the manager's participation for a given $e \in \{\underline{e}, \overline{e}\}$:

$$\begin{aligned} \mathbb{E}[u(x,y,e)|e] &= \bar{W} \\ \Leftrightarrow \quad w + \beta_x\mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x|e\right] + \beta_p\mathbb{E}[\tilde{p}|e] + \sum_{i=1,2}\sum_{j=1,2}\beta_i^j\mathbb{E}[\tilde{m}_i^j] &= \bar{W} + C(e) \end{aligned} \tag{92}$$

Thus, equation (25) can be rewritten as:

$$\begin{aligned} \mathbb{E}[V(x,y,e)] &= \mathbb{E}\left[(1-\beta_x)\left(e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x\right) + \alpha_B\left(\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y\right) - \left(w + \beta_p\tilde{p} + \sum_{i=1,2}\sum_{j=1,2}\beta_i^j\tilde{m}_i^j\right)\right] \\ &= \mathbb{E}\left[e - \theta_1 y_1^2 - \theta_2 y_2^2 + \tilde{\epsilon}_x + \alpha_B\left(\tilde{\eta}_1 y_1 + \tilde{\eta}_2 y_2 + \tilde{\epsilon}_y\right) - \left(\bar{W} + C(e)\right)|e\right] \end{aligned} \tag{93}$$

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:

$$\begin{aligned}\mathbb{E}\left[\tilde{\eta}_i|\varepsilon_i^1,\varepsilon_i^2\right] &= \bar{\eta} + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k-\bar{\eta}\right)\\ y_i^*(\varepsilon_i^1,\varepsilon_i^2) &= \frac{\alpha_B}{2}\frac{1}{\theta_i}\left(\bar{\eta}+\sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2-\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2-\rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2+\sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2-2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k-\bar{\eta}\right)\right) \quad (94)\end{aligned}$$

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+\beta_p\alpha_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)}{\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{\beta_p}{\beta_x+\beta_p}\alpha_I-\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{\beta_p}{\beta_x+\beta_p}\alpha_I-\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 \quad (95)\end{aligned}$$

The expression in equation (95) is globally convex with respect to $\beta_i^k$. The FOC w.r.t. $\beta_i^k$ 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{\beta_p}{\beta_x+\beta_p}\alpha_I-\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{\beta_p}{\beta_x+\beta_p}\alpha_I-\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{\beta_p}{\beta_x+\beta_p}\alpha_I-\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{\beta_p}{\beta_x+\beta_p}\alpha_I\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{\beta_p}{\beta_x+\beta_p}\alpha_I\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 (96)\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 (95) when $\beta_i^k$ is as in equation (96):

$$\min_{\beta_x,\beta_p} \sum_{i=1,2} \frac{1}{2}\frac{1}{\theta_i} \sum_{k=1,2} \int_{\eta_i}\int_{\varepsilon_i^k} \left(\left(\left(\alpha_B - \frac{\beta_p}{\beta_x+\beta_p}\alpha_I\right)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2} + \Sigma_k\left(1-\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}\right)\right)\right.\right.$$
$$\left.\left. + \left(\frac{\beta_p}{\beta_x+\beta_p}\alpha_I - \alpha_B\right)\Sigma_k\right)\varepsilon_i^k + \left(\frac{\beta_p}{\beta_x+\beta_p}\alpha_I - \alpha_B\right)\left(\frac{1}{2}-\Sigma_k\right)\bar{\eta}\right)^2 \varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i$$
$$\Leftrightarrow \quad \min_{\beta_x,\beta_p}\left(\alpha_B - \frac{\beta_p}{\beta_x+\beta_p}\alpha_I\right)^2 \sum_{i=1,2}\frac{1}{2}\frac{1}{\theta_i}\sum_{k=1,2}\int_{\eta_i}\int_{\varepsilon_i^k}\left(\left(\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2} + \Sigma_k\left(1-\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}\right)\right) - \Sigma_k\right)\varepsilon_i^k\right.$$
$$\left. - \left(\frac{1}{2}-\Sigma_k\right)\bar{\eta}\right)^2 \varphi(\varepsilon_i^k|\eta_i)\phi(\eta_i)d\varepsilon_i^k d\eta_i \tag{97}$$

The expression under the integral signs in equation (97) is positive and independent from the contract. The expression in equation (97) is minimized by minimizing $\left(\alpha_B - \frac{\beta_p\alpha_I}{\beta_x+\beta_p}\right)^2$. With a nonnegativity constraint on $\beta_x$, this is achieved by setting $\beta_x = \left(\frac{\alpha_I}{\alpha_B}-1\right)\beta_p$ if $\alpha_B \leq \alpha_I$, and $\beta_x = 0$ otherwise.

Substituting in equation (91) when $\alpha_B \leq \alpha_I$ so that $\beta_i^k = 0$ and $\beta_x = \left(\frac{\alpha_I}{\alpha_B}-1\right)\beta_p$:

$$\begin{aligned} y_i(\varepsilon_i^1,\varepsilon_i^2) &= \alpha_B\left(\bar{\eta} + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\varepsilon_i^k - \bar{\eta}\right)\right)\frac{1}{2\theta_i} \\ &= \alpha_B\left(\bar{\eta} + \sum_{k=1,2}\Sigma_k\left(\varepsilon_i^k - \bar{\eta}\right)\right)\frac{1}{2\theta_i}, \end{aligned} \tag{98}$$

which is as in equation (94). 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 $\rho = -1$:

$$\begin{aligned} \mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right] &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sum_{k=1,2}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 + \sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2+\sigma_{\varepsilon_j}^2 + 2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}\left(\sigma_\eta^2+\sigma_{\varepsilon_k}^2\right)\right) \\ &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sum_{k=1,2}\frac{\sigma_{\varepsilon_j}}{\sigma_{\varepsilon_k}+\sigma_{\varepsilon_j}}\left(\sigma_\eta^2+\sigma_{\varepsilon_k}^2\right)\right) \\ &= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2 + \sigma_\eta^2 + \frac{\sigma_{\varepsilon_2}\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_1}\sigma_{\varepsilon_2}^2}{\sigma_{\varepsilon_1}+\sigma_{\varepsilon_2}}\right), \end{aligned} \tag{99}$$

which is larger than in equation (51) for any parameter values. Moreover, $\mathbb{E}[\tilde{\eta}_i \tilde{y}_i]$ is continuous in $\rho$, so that the result about $\rho$ in Proposition 4 when $\alpha_B \le \alpha_I$ holds by a continuity argument. As $\sigma_2 \to 0$:

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

which is larger than in equation (51) for any parameter values. Moreover, $\mathbb{E}[\tilde{\eta}_i \tilde{y}_i]$ is continuous in $\sigma_2$, so that the result about $\sigma_2$ in Proposition 4 when $\alpha_B \le \alpha_I$ holds by a continuity argument.

We have:

$$\begin{aligned}
&\frac{\partial}{\partial\rho}\frac{\sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})}{\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})} \\
= \; &\frac{-\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\right) - \sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\left(-2\rho\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - 2\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}\right)}{\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\right)^2} \\
= \; &\frac{-\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2\sigma_{\varepsilon_k}^2\right) + \sigma_\eta^2(\sigma_{\varepsilon_j}^2 - \rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})2\rho\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2\sigma_{\varepsilon_j}^2\sigma_\eta^2\sigma_{\varepsilon_k}\sigma_{\varepsilon_j}}{\left(\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 - \rho^2\sigma_{\varepsilon_k}^2\sigma_{\varepsilon_j}^2 + \sigma_\eta^2(\sigma_{\varepsilon_k}^2 + \sigma_{\varepsilon_j}^2 - 2\rho\sigma_{\varepsilon_k}\sigma_{\varepsilon_j})\right)^2}
\end{aligned} \tag{100}$$

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 $\rho = 0$:

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

Substituting in equation (91) when $\beta_i^k$ is as in equation (96) and $\alpha_B \ge \alpha_I$ so that $\beta_x = 0$:

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

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

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

$$= \frac{1}{2\theta_i}\sum_{k=1,2}\left((\alpha_B-\alpha_I)\left(\frac{1}{2}\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}+\Sigma_k\left(1-\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}\right)\right)\left(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_k}^2\right)\right.$$
$$\left.+\alpha_I\left(\frac{\bar{\eta}^2}{2}+\Sigma_k\left(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_k}^2-\bar{\eta}^2\right)\right)\right)$$

For $\rho=-1$:

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

which is larger than in equation (56) for any parameter values. Moreover, $\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right]$ is continuous in $\rho$, so that the result about $\rho$ in Proposition 4 when $\alpha_B>\alpha_I$ holds by a continuity argument. As $\sigma_2\to 0$:

$$\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right] \to \frac{1}{2\theta_i}\left((\alpha_B-\alpha_I)\left(\frac{\bar{\eta}^2}{\sigma_{\varepsilon_k}^2+\sigma_\eta^2+\bar{\eta}^2}+\left(1-\frac{\bar{\eta}^2}{\sigma_{\varepsilon_1}^2+\sigma_\eta^2+\bar{\eta}^2}\right)\right)\left(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^2\right)+\alpha_I\left(\bar{\eta}^2+\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^2-\bar{\eta}^2\right)\right)$$
$$= \frac{1}{2\theta_i}\left((\alpha_B-\alpha_I)\left(\bar{\eta}^2+\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^2-\bar{\eta}^2\right)+\alpha_I\left(\bar{\eta}^2+\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^2-\bar{\eta}^2\right)\right)$$
$$= \frac{\alpha_B}{2\theta_i}\left(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^2\right)$$

which is larger than in equation (56) for any parameter values. Moreover, $\mathbb{E}\left[\tilde{\eta}_i\tilde{y}_i\right]$ is continuous in $\sigma_2$, so that the result about $\sigma_2$ in Proposition 4 when $\alpha_B\le\alpha_I$ holds by a continuity argument. ■

# D Stock price

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:

$$\mathbb{E}\left[(\tilde{x}+\alpha_I\tilde{y})-qp|z\right]=q\left(\mathbb{E}\left[\tilde{x}|z\right]+\alpha_I\mathbb{E}\left[\tilde{y}|z\right]\right)-qp \tag{101}$$

where the third equality relies on the normal distribution assumption. Maximizing the expression in equation (101) with respect to $q$ shows that the only stock price compatible with market clearing for the firm stock is:

$$p=\mathbb{E}\left[\tilde{x}|z\right]+\alpha_I\mathbb{E}\left[\tilde{y}|z\right]=\mu_{x|z}+\alpha_I\mu_{y|z} \tag{102}$$

where $\mu_{x|z}$ and $\mu_{y|z}$ are respectively as in equations (107) and (108) 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:

$$p=\mu_{x|z,m}+\alpha_I\mu_{y|z,m} \tag{103}$$

where, with one set of ESG scores, $\mu_{x|z,m}$ and $\mu_{y|z,m}$ are described respectively in equations (109) and (110). The stock price $p$ is linear in the report $z$ and in the measures $m_1$ and $m_2$: for three constants $a$ and $b$, we can write $p=a+b_0z+b_1m_1+b_2m_2$, with $b_1=b_2$. The cases with additional sets of ESG scores are analyzed in sections (D.3) and (D.4).

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

## D.1 No ESG scores

The manager is compensated with a fixed wage and stock price-based compensation, with a linear contract with a sensitivity $\beta_p$ of the manager's compensation to the stock price. Consider the perspective of the manager at the time ($t=0$) of choosing $y_1$ and $y_2$ given its knowledge of $\{\eta_1,\eta_2\}$ and the information

$I$ of investors at $t = 1$ (remember that all economic agents observe $\{y_1, y_2\}$).

$$\begin{aligned}
\mathbb{E}[u(x,y,e)|J] &= \mathbb{E}\left[w + \beta_p \tilde{p} - C(e)\right] \\
&= \beta_p \left(\bar{e} - \theta_1 y_1^2 - \theta_2 y_2^2 + \alpha_I \left(\mathbb{E}[\tilde{\eta}_1|I] y_1 + \mathbb{E}[\tilde{\eta}_2|I] y_2\right)\right) + w - C(e) \qquad (104)
\end{aligned}$$

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 $\mathbb{E}[\hat{e}|e = \bar{e}] = \bar{e}$. Optimizing from equation (104) with respect to $y_i$ taking the manager's contract as given, the manager will choose $y_i$ such that:

$$\beta_p \left(-2\theta_1 y_i + \alpha_I \mathbb{E}[\tilde{\eta}_i|I]\right) = 0 \quad \Leftrightarrow \quad y_i = \alpha_I \frac{\mathbb{E}[\tilde{\eta}_i|I]}{2\theta_i} \qquad (105)$$

Without ESG scores, investors only observe $y_i$ on each dimension $i$ of SEP whereas the manager knows $\varepsilon_i$ 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 (105) 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 $\varepsilon_i$. Thus, $y_i$ is independent from $\varepsilon_i$ 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 (105): $y_i = \frac{\alpha_I}{2} \frac{\bar{\eta}}{\theta_i}$. Thus, the stock price is defined by:

$$\begin{aligned}
\mu_x &= \hat{e} - \mathbb{E}[\theta_1 \tilde{y}_1^2 + \theta_2 \tilde{y}_2^2] \qquad (106) \\
\mu_{x|z} &= \mathbb{E}[\tilde{x}] + \frac{cov(\tilde{z}, \tilde{x})}{var(\tilde{z})} \left(z - \mathbb{E}[\tilde{z}]\right) \\
&= \hat{e} - \theta_1 y_1^2 - \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) \qquad (107) \\
\mu_{y|z} &= \mathbb{E}\left[\tilde{\eta}_1 \tilde{y}_1 | y_1\right] + \mathbb{E}\left[\tilde{\eta}_2 \tilde{y}_2 | y_2\right] + \mathbb{E}\left[\tilde{\epsilon}_y\right] \; = \bar{\eta} y_1 + \bar{\eta} y_2 \qquad (108)
\end{aligned}$$

Thus, without ESG scores, a manager with stock price-based incentives does not implement social investments contingent on $\varepsilon_i$. 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, $\theta_i$. Regardless of how investors update their beliefs when they observe $y_i$, this level of social investment is independent from $\varepsilon_i$.

## D.2 One set of ESG scores

We consider the equilibrium in which the board delegates the social investment decision to the firm's manager. With one set of ESG scores, investors update their beliefs about the firm's technology for social output after observing ESG scores as follows:

$$\begin{aligned}
\mu_{x|z,m} &= \mathbb{E}[\tilde{x}] + \frac{cov(\tilde{z},\tilde{x})}{var(\tilde{z})}\left(z - \mathbb{E}[\tilde{z}]\right) \\
&= \hat{e} - \theta_1 y_1^2 + \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2}\left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) \qquad (109) \\
\mu_{y|z,m} &= \mathbb{E}\left[\tilde{\eta}_1\tilde{y}_1 + \tilde{\eta}_2\tilde{y}_2 + \tilde{\epsilon}_y | z, m\right] \\
&= y_1\mathbb{E}\left[\tilde{\eta}_1|m_1,y_1\right] + y_2\mathbb{E}\left[\tilde{\eta}_2|m_2,y_2\right] \qquad (110)
\end{aligned}$$

where:

$$\mathbb{E}\left[\tilde{\eta}_i|m_i,y_i\right] = \mathbb{E}\left[\tilde{\eta}_i|\varepsilon_i\right] = \mathbb{E}\left[\tilde{\eta}_i\right] + \frac{cov\left(\tilde{\eta}_i,\tilde{\varepsilon}_i\right)}{var\left(\tilde{\varepsilon}_i\right)}\left(\varepsilon_i - \mathbb{E}[\tilde{\varepsilon}_i]\right) = \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\varepsilon_i - \bar{\eta}\right) \qquad (111)$$

Substituting into equation (110):

$$\begin{aligned}
\mu_{y|z,m} &= y_1\left(\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\varepsilon_1 - \bar{\eta}\right)\right) + y_2\left(\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\left(\varepsilon_2 - \bar{\eta}\right)\right) \\
&= \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}m_1 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}y_1\bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}m_2 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}y_2\bar{\eta} \qquad (112)
\end{aligned}$$

## D.3 $N$ 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.[28] Define:

$$\bar{m}_i \equiv \frac{1}{N}\sum_{j=1}^{N} m_i^j \tag{113}$$

We have:

$$\frac{1}{N}\sum_{j=1}^{N}\tilde{m}_i^j = \frac{1}{N}\sum_{j=1}^{N} y_i\left(\eta_i + \tilde{\zeta}_i^j\right) = \eta_i y_i + \frac{y_i}{N}\sum_{j=1}^{N}\tilde{\zeta}_i^j \tag{114}$$

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

$$\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N}\tilde{\zeta}_i^j\right] = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left[\tilde{\zeta}_i^j\right] = 0 \tag{115}$$

$$var\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{\zeta}_i^j\right) = \frac{1}{N^2}\sum_{j=1}^{N} var\left(\tilde{\zeta}_i^j\right) = \frac{1}{N}\sigma_\varepsilon^2 \tag{116}$$

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

$$var\left(\tilde{\bar{m}}_i/y_i\right) = \sigma_\eta + \frac{1}{N}\sigma_\varepsilon^2 \tag{117}$$

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

$$\begin{aligned} \mu_{x|z,\bar{m}} &= \mathbb{E}[\tilde{x}] + \frac{cov(\tilde{z},\tilde{x})}{var(\tilde{z})}\left(z - \mathbb{E}[\tilde{z}]\right) \\ &= \hat{e} - \theta_1 y_1^2 + \theta_2 y_2^2 + \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2}\left(z - \hat{e} + \theta_1 y_1^2 + \theta_2 y_2^2\right) \end{aligned} \tag{118}$$

$$\mu_{y|z,\bar{m}} = \mathbb{E}\left[\tilde{\eta}_1\tilde{y}_1 + \tilde{\eta}_2\tilde{y}_2 + \tilde{\epsilon}_y | z,\bar{m}\right] = y_1\mathbb{E}\left[\tilde{\eta}_1|\bar{m}_1, y_1\right] + y_2\mathbb{E}\left[\tilde{\eta}_2|\bar{m}_2, y_2\right] \tag{119}$$

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

$$\mathbb{E}\left[\tilde{\eta}_i|\bar{m}_i\right] = \mathbb{E}\left[\tilde{\eta}_i\right] + \frac{cov\left(\tilde{\eta}_i, \tilde{\bar{m}}_i/y_i\right)}{var\left(\tilde{\bar{m}}_i/y_i\right)}\left(\bar{m}_i/y_i - \mathbb{E}[\tilde{\bar{m}}_i/y_i]\right) = \bar{\eta} + \frac{\sigma_\eta^2}{\sigma_\eta + \sigma_\varepsilon^2/N}\left(\bar{m}_i/y_i - \bar{\eta}\right) \tag{120}$$

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

Substituting into equation (119):

$$\begin{aligned}\mu_{y|z,\bar{m}} &= y_1\left(\bar{\eta}+\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\left(\bar{m}_1/y_1-\bar{\eta}\right)\right)+y_2\left(\bar{\eta}+\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\left(\bar{m}_2/y_2-\bar{\eta}\right)\right)\\ &= \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{m}_1+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}y_1\bar{\eta}+\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2/N}\bar{m}_2+\frac{\sigma_\varepsilon^2/N}{\sigma_\eta^2+\sigma_\varepsilon^2/N}y_2\bar{\eta} \qquad (121)\end{aligned}$$

## D.4 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 $\eta_i$: $\tilde{\varepsilon}_i^1 \sim \mathcal{N}(\eta_i, \sigma_{\varepsilon_i}^{1^2})$, $\tilde{\varepsilon}_i^2 \sim \mathcal{N}(\eta_i, \sigma_{\varepsilon_i}^{2^2})$. 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 (118)-(119), 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:[29]

$$\boldsymbol{\mu} = \begin{pmatrix}\mathbb{E}[\tilde{\eta}]\\ \mathbb{E}[\tilde{\varepsilon}_1]\\ \mathbb{E}[\tilde{\varepsilon}_2]\end{pmatrix} = \begin{pmatrix}\bar{\eta}\\ \bar{\eta}\\ \bar{\eta}\end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix} \qquad (122)$$

where

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

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

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

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

$$\begin{aligned}
\mathbb{E}[\tilde{\eta}|\varepsilon_1, \varepsilon_2] &= \bar{\eta} + \boldsymbol{\Sigma}_{12}\boldsymbol{\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 (123)
\end{aligned}$$

and variance:

$$\begin{aligned}
\sigma_{\eta|\varepsilon_1,\varepsilon_2}^2 &= \sigma_\eta^2 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\
&= \sigma_\eta^2 - \frac{\sigma_\eta^4\left(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 - 2\rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}\right)}{\sigma_\eta^2\sigma_{\varepsilon_1}^2 + \sigma_\eta^2\sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 - \rho^2\sigma_{\varepsilon_1}^2\sigma_{\varepsilon_2}^2 - 2\rho\sigma_\eta^2\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}} \qquad (124)
\end{aligned}$$

Online Appendix
Executive Compensation with Social and Environmental Performance

# 1 Case with two imperfectly correlated sets of ESG scores

We study the case without stock price-based compensation ($\beta_p = 0$).

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

$$\arg\max_{e,y_1,y_2} \mathbb{E}\left[w + \beta_x\left(e - \theta_1 y_1^2 - \theta_2 y_2^2\right) + \beta_1^1 m_1^1 + \beta_1^2 m_1^2 + \beta_2^1 m_2^1 + \beta_2^2 m_2^2\right] - C(e) \tag{125}$$

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

$$y_1 = \frac{1}{2}\frac{1}{\theta_1}\frac{\beta_1^1 \varepsilon_1^1 + \beta_1^2 \varepsilon_1^2}{\beta_x} \tag{126}$$

$$y_2 = \frac{1}{2}\frac{1}{\theta_2}\frac{\beta_2^1 \varepsilon_2^1 + \beta_2^2 \varepsilon_2^2}{\beta_x} \tag{127}$$

Incentive compatibility with respect to effort $e$ is achieved as in equation (30) 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)
\end{aligned} \tag{128}$$

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 $\beta_i^j$. The first-order condition with respect to $\beta_i^j$ is:

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

$$\Leftrightarrow \quad \begin{cases} \frac{\beta_i^j}{\beta_x}(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_i}^{j^2}) + \frac{\beta_i^k}{\beta_x}(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2 + \bar{\eta}^2) = \alpha\left(1 - \frac{c_e}{\bar{e}-\underline{e}}\right)(\sigma_\eta^2 + \bar{\eta}^2) \\ \frac{\beta_i^k}{\beta_x}(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_i}^{k^2}) + \frac{\beta_i^j}{\beta_x}(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2 + \bar{\eta}^2) = \alpha\left(1 - \frac{c_e}{\bar{e}-\underline{e}}\right)(\sigma_\eta^2 + \bar{\eta}^2) \end{cases} \tag{129}$$

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:

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

$$\begin{aligned}
\Rightarrow \quad \frac{\beta_i^j}{\beta_x} &= \alpha\frac{(\sigma_\eta^2 + \bar{\eta}^2) - \frac{(\sigma_\eta^2 + \bar{\eta}^2)(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_i}^{k^2})}{\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2 + \bar{\eta}^2}}{(\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2 + \bar{\eta}^2) - \frac{(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_i}^{j^2})(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_i}^{k^2})}{\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2 + \bar{\eta}^2}} \\
&= \alpha\frac{(\sigma_\eta^2 + \bar{\eta}^2)(\sigma_\eta^2 + \sigma_{\varepsilon_i}^{k^2} - \rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2)}{(\sigma_\eta^2 + \bar{\eta}^2 + \sigma_{\varepsilon_i}^{j^2})(\bar{\eta}^2 + \sigma_\eta^2 + \sigma_{\varepsilon_i}^{k^2}) - (\rho\sigma_{\varepsilon_i}^1\sigma_{\varepsilon_i}^2 + \bar{\eta}^2)^2}
\end{aligned} \tag{130}$$

As $c_e \to 0$, the optimum for the board is thus given by setting $\beta_x \to 0$ and $\beta_1$ and $\beta_2$ are as in equation (130). Substituting for $\beta_1/\beta_x$ and $\beta_2/\beta_x$ in equations (126) and (127), respectively:

$$y_1 = \frac{\alpha}{2}\frac{1}{\theta_1}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{2^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\varepsilon_1^1 + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2}-\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_1}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_1}^{1^2})-(\rho\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2+\bar{\eta}^2)^2}\varepsilon_1^2\right) \tag{131}$$

$$y_2 = \frac{\alpha}{2}\frac{1}{\theta_2}\left(\frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{1^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{2^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\varepsilon_2^1 + \frac{(\sigma_\eta^2+\bar{\eta}^2)(\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2}-\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2)}{(\sigma_\eta^2+\bar{\eta}^2+\sigma_{\varepsilon_2}^{2^2})(\bar{\eta}^2+\sigma_\eta^2+\sigma_{\varepsilon_2}^{1^2})-(\rho\sigma_{\varepsilon_2}^1\sigma_{\varepsilon_2}^2+\bar{\eta}^2)^2}\varepsilon_2^2\right) \tag{132}$$

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

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

The first condition is equivalent to:

$$
\begin{aligned}
& 2\sigma_\eta^2+{\sigma^{1}_{\varepsilon_1}}^2+{\sigma^{2}_{\varepsilon_1}}^2-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1} > \bar{\eta}^2+\sigma_\eta^2+{\sigma^{2}_{\varepsilon_1}}^2-\frac{(\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+\bar{\eta}^2)^2}{\bar{\eta}^2+\sigma_\eta^2+{\sigma^{1}_{\varepsilon_1}}^2} \\
\Leftrightarrow\quad & \sigma_\eta^2+{\sigma^{1}_{\varepsilon_1}}^2-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1} > \bar{\eta}^2-\frac{(\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}+\bar{\eta}^2)^2}{\bar{\eta}^2+\sigma_\eta^2+{\sigma^{1}_{\varepsilon_1}}^2} \\
\Leftrightarrow\quad & \bar{\eta}^2\sigma_\eta^2+\sigma_\eta^4+{\sigma^{1}_{\varepsilon_1}}^2\sigma_\eta^2+\bar{\eta}^2{\sigma^{1}_{\varepsilon_1}}^2+\sigma_\eta^2{\sigma^{1}_{\varepsilon_1}}^2+{\sigma^{1}_{\varepsilon_1}}^4-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\bar{\eta}^2-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\sigma_\eta^2-2\rho{\sigma^{1}_{\varepsilon_1}}^3\sigma^2_{\varepsilon_1} \\
& > \bar{\eta}^4+\sigma_\eta^2\bar{\eta}^2+{\sigma^{1}_{\varepsilon_1}}^2\bar{\eta}^2-\rho^2{\sigma^{1}_{\varepsilon_1}}^2{\sigma^{2}_{\varepsilon_1}}^2-\bar{\eta}^4-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\bar{\eta}^2 \\
\Leftrightarrow\quad & \sigma_\eta^4+2{\sigma^{1}_{\varepsilon_1}}^2\sigma_\eta^2+{\sigma^{1}_{\varepsilon_1}}^4-2\rho\sigma^1_{\varepsilon_1}\sigma^2_{\varepsilon_1}\sigma_\eta^2-2\rho{\sigma^{1}_{\varepsilon_1}}^3\sigma^2_{\varepsilon_1} > -\rho^2{\sigma^{1}_{\varepsilon_1}}^2{\sigma^{2}_{\varepsilon_1}}^2 \qquad (133)
\end{aligned}
$$

The condition in equation (133) 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 (133) is satisfied for any $\rho$ if it is satisfied for $\rho = 1$. For $\rho = 1$, it is satisfied if and only if:

$$\begin{aligned}
& \sigma_\eta^4 + 2\sigma_{\varepsilon_1}^{1^2}\sigma_\eta^2 + \sigma_{\varepsilon_1}^{1^4} > 2\sigma_\eta^2\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2 + 2\sigma_{\varepsilon_1}^{1^3}\sigma_{\varepsilon_1}^2 - \sigma_{\varepsilon_1}^{1^2}\sigma_{\varepsilon_1}^{2^2} \\
\Leftrightarrow \quad & \sigma_\eta^2(\sigma_\eta^2 + 2\sigma_{\varepsilon_1}^{1^2} - 2\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2) > \sigma_{\varepsilon_1}^{1^2}(2\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2 - \sigma_{\varepsilon_1}^{2^2} - \sigma_{\varepsilon_1}^{1^2}) \\
\Leftrightarrow \quad & \sigma_\eta^2(\sigma_\eta^2 + 2\sigma_{\varepsilon_1}^{1^2} - 2\sigma_{\varepsilon_1}^1\sigma_{\varepsilon_1}^2) > -\sigma_{\varepsilon_1}^{1^2}(\sigma_{\varepsilon_1}^2 - \sigma_{\varepsilon_1}^1)^2
\end{aligned} \tag{134}$$

The LHS is quadratic with respect to $\sigma_\eta^2$, with a minimum at

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

Replacing in equation (134), this condition is generically satisfied at $\rho = 1$ for any $\sigma_\eta^2$ if and only if:

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

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