Green Moral Hazard: Estimating the Financial and the Green Impacts of ESG Incentives

Kyle (Jaehoon) Jung*

October 8, 2024

Please find the most recent version here!

Abstract

I structurally estimate the financial and the environmental implications of the actions induced by executive compensation contracts involving ESG (Environmental, Social, and Governance) incentives. I find that such contracts incentivize CEOs to make tangible financial sacrifices to improve ESG performance: approximately 1.3% of firm value for carbon emission intensity reduction of 1.8% per year. I then examine the extent of moral hazard and the value of information associated with the ESG incentives. Through counterfactual analyses, I find that the cost of incentivizing improvement in ESG performance on top of maximizing financial performance, i.e., "green moral hazard", is substantial: \$1.72 million, out of the total moral hazard cost of \$2.05 million, could be attributed to green moral hazard.

*Stanford Graduate School of Business. (kjjung@stanford.edu)
I am grateful to Ivan Marinovic, Maureen McNichols, Kevin Smith, Peter Reiss, and Colleen Honigsberg for their invaluable guidance. I also thank Christopher Armstrong, Brandon Gipper, Suzie Noh, Anne Beyer, Stefan Reichelstein, Zhiguo He, Susanna Gallani, Rodrigo Verdi, Anastasia Zakolyukina, Christopher Ittner, Paul Ma, Jeffrey Wurgler, Tim Baldenius, Hans Christensen, Hao Xue, Lin Qiu, Ilan Guttman, Michael Kimbrough, and participants of the AES Junior Accounting Theory Conference, Stanford Accounting Brown Bag seminar, the Stanford Corporate Finance Reading Group, and the AAA Doctoral Consortium for their helpful comments. All errors are my own.

1 Introduction

A fundamental question in information economics is the impact of information asymmetry on economic decisions. Principal-agent theory addresses how principals incentivize agents' unobservable actions using compensation based on imperfect signals. So far, the literature has focused primarily on frameworks where contracts are designed to maximize firm value. However, executive compensation is increasingly targeting multiple objectives beyond profit maximization.

In particular, the proportion of firms incorporating environmental, social, and governance (ESG) objectives into executive compensation contracts has been steadily increasing, from only 1% in 2011 to over 40% in 2023. Despite its growing prevalence, little research has focused on agency friction in contracts with both financial and ESG objectives. Currently, it is an open question in the literature, whether the existence of ESG incentives in executive compensation indicates that environmental performance is informative about the CEO's effort to improve financial performance or that improving ESG outcome is an objective on its own. This has important implications for investors: both whether and extent to which firms are willing to trade financial outcomes for environmental outcomes are critical factors in assessing how firms' interests align with their own, and thus in making investment decisions. To fill this gap, I develop a structural framework to identify moral hazard in contracts aiming to achieve financial and non-financial objectives. Then, I apply my framework to examine the financial and ESG impacts of executive compensation contracts with ESG incentives, as well as the extent of associated moral hazard.

From my estimation, I find that firms with ESG incentives are willing to forgo their market value to improve their ESG performance. By showing that firms are making financial sacrifices beyond capital market incentives, I contribute to the ongoing research on the extent to which the stakeholders are willing to sacrifice financial benefits for improvements in environmental performance (e.g. "greenium"). With the estimates, I conduct counterfactual analysis to quantify the extent of moral hazard associated with the two objectives: financial performance and ESG performance. From this analysis, I find that firms with ESG incentives pay substantial premiums to incentivize CEOs to improve ESG performance, even more than those for financial performance. The result highlights the severity of the information asymmetry around CEOs' actions regarding ESG outcomes.

It is worth noting that, without a structural model, it is challenging to (1) identify moral hazard in contracts with multidimensional objectives, (2) estimate the tradeoff between financial and ESG outcomes, and (3) quantify the extent of moral hazard associated with each incentive. This is because examining the effect of compensation contracts on firm activities involves multiple challenges. First, the adoption of ESG incentives in compensation contract is in itself an inherently endogenous decision. As the decision to compensate based on ESG performance would depend heavily on the tradeoffs between financial and ESG outcomes, one cannot use the outcomes of firms without ESG incentives as proper counterfactuals. Second, it is highly unlikely that shocks to contracts do not simultaneously affect other important aspects of firms, especially the economic tradeoffs that efforts to improve ESG performance would entail. Then, one cannot identify the nature of the action incentivized by the ESG incentive. I use my structural framework to address these challenges.

In this paper, I examine the extent to which compensation contracts incentivize managers to invest in improving firms' ESG performance and quantify the economic magnitude of the moral hazard problem associated with such incentives. I begin by writing a simple principal-agent model in which the agent can both exert effort and make an investment decision; both actions are unobservable and can only be inferred from performance measures. Then, I estimate this model from the data of realized outcomes and compensations for firms that implement ESG pay, to uncover the underlying parameters including the cost of effort and the value of outside options, as well as the counterfactual outcome distributions. With the estimates, I perform counterfactual analyses to quantify the extent of moral hazard associated with each action.

I start by constructing a two-signal, multitasking model of contracting. In my model, the principal designs a contract with an agent that can perform two types of actions that impact the distribution of the principal's value. The agent's action choice is unobservable and the principal can only infer it from realized outcomes, the joint distribution of which varies by the agent's actions.[1] My model allows me to separately identify the financial and non-financial implications of ESG-related investment decisions from that of the manager's personally costly effort to improve financial performance. The intuition behind this is a lá Holmström (1979) that the shape of the contract reflects the shape of the likelihood ratios of outcomes across actions. Moreover, my model enables me to

[1] As in Holmström (1979), I do not assume any functional form for the joint distribution of outcomes. Yet, I do make parametric assumptions regarding how manager's actions transform the joint distribution.

extract the unobserved distribution of outcomes under counterfactual contract, that does not incentivize improvement in ESG performances, from the observed compensation.

To take my model to the data, I merge datasets from Executive Compensation Analytics (ECA), Execucomp, Trucost, CRSP, and Compustat and construct a firm-year panel of compensation, financial performance, and environmental performance covering over 600 U.S. firms from 2012 to 2022. To measure the impact of ESG metrics, I confine my main analyses to firms having ESG metrics in compensation contracts. I use abnormal stock return as a proxy for financial performance and log reduction in carbon emission intensity as a proxy for environmental performance.

The structural estimation is then applied to the constructed dataset. The estimation process is as follows. First, I nonparametrically estimate the joint distribution of financial and ESG outcomes and the wage function from the sample. Then, I estimate the parameters from the moments computed from the estimated distribution and the wage function. Finally, based on the parameters, I infer the counterfactual distributions under only financial effort and that under only ESG investment from the wage function.

As a result of the estimation, I find that firms are sacrificing substantial financial value to improve their ESG performance: to reduce carbon intensity by more than 1.7%, firms are willing to forgo over 1.3% of stock return. To the extent that the stock market efficiently prices firms' ESG investments, the result is contrary to the claim that firms are paying CEOs on ESG performance only for financial gains. The result also sheds light on the willingness of these firms to sacrifice financial gains for improvements in ESG performance, relative to that of marginal investor in the capital market: firms are willing to forgo at least 0.74% more financial value of the firm, for a percenage reduction in carbon emission intensity.

Then, I perform counterfactual analyses using the estimated parameters, as well as the counterfactual distribution without financial effort and without ESG investment. Specifically, I infer what the optimal contract would have been when one of the actions by the CEO is observable. This allows me to decompose the wage we observe in the real world into three components: (1) first-best wage, which compensates for the cost of effort and outside option, (2) cost of financial moral hazard, which is the cost of incentivizing unobservable financial effort, and (3) cost of green moral hazard, which is the cost of incentivizing unobservable ESG investments on top of financial effort.

My counterfactual analyses have important implications. Incentivizing executives to invest in improving ESG performance, on top of exerting financial effort, is substantially costly: the cost of green moral hazard is estimated at more than 1.7 million U.S. Dollars (USD), which is more than 7% of CEOs' annual compensation. In contrast, the cost of financial moral hazard is estimated at less than 0.4 million USD, only around 1.5%. Therefore, my findings suggest that stock returns better resolve uncertainty regarding CEOs' financial efforts than do carbon emissions about ESG project decisions.

Taken together, my findings suggest that firms are sacrificing tangible financial gains to improve their ESG performance and that a significant portion of executive compensation is devoted to inducing CEOs to execute costly ESG projects.

Contribution to Literature This paper adds to the literature on the structural estimation of contracting models. Prior research in the context of moral hazard includes Margiotta and Miller (2000), Gayle and Miller (2009), Gayle and Miller (2015), Gayle et al. (2022). In particular, a closely related paper is Bertomeu et al. (2023a), which estimates the value of accounting information to contracting efficiency. This paper adds to the literature by providing a framework for disentangling the implications and incentives associated with multiple dimensions of actions that the agent can take. While this paper is applied to the context of ESG pay, the methodology can be readily applied to different settings in which CEOs can take multidimensional actions that affect the joint distribution of financial and non-financial outcomes. Compared to Bertomeu et al. (2023a), my model involves *multi-dimensional action space* for the agent and therefore I estimate multiple counterfactual distributions.

This paper contributes to the literature on ESG and compensation contracts. A big debate in the literature regards the role of ESG metrics in compensation contracts. While works such as Flammer et al. (2019) suggest that firms are using ESG metrics to improve financial outcomes, works such as Cohen et al. (2023) suggest that ESG pay does indeed improve ESG outcomes.[2] I contribute to this literature by finding that ESG signals contribute substantially to incentivizing costly ESG investments that improve ESG performance, which is more consistent with the latter. I also provide an economic magnitude of the financial sacrifice that firms are making to improve ESG

[2]See Acharya et al. (2023) for a literature review on the topic of climate finance and regulatory frameworks.

performance and of the ESG signal's contribution to mitigating information asymmetry regarding ESG investment.

More broadly, this paper contributes to the large literature on moral hazard and contracting efficiency. The seminal work by Holmström (1979) provides the informativeness criterion, for when and how a piece of imperfect information should be used to determine compensation. Gayle and Miller (2009) find that moral hazard has become the most important factor in explaining managerial compensation. My findings not only confirm that the moral hazard explains a substantial portion of executive compensation, but also *quantifies* the cost of moral hazard associated with ESG investments, yielding especially important implications on climate context. Moreover, I also quantify the contracting value of ESG signals for implementing ESG investments.

The interpretation of the results in this paper is subject to caveats. First, my approach assumes that observed contracts are indeed optimal, and my model does not account for any spillover across firms; only the implications of CEOs' actions on their own firms can be identified. Second, my analysis excludes the possibility that CEOs are wealthy enough to pursue their own interests against those of the principal, disregarding the financial incentives provided by the contract. Third, this paper does not provide inference on how well financial and ESG performance metrics capture the underlying fundamental performance in each dimension; I assume that the principals care about the stock price and the carbon emission, as opposed to the fundamental economic value and the true ESG performance of the firms.

Outline of the Paper The remainder of the paper proceeds as follows: Section 2 provides institutional background regarding ESG incentives. Section 3 describes the model and the assumptions for identification. Section 4 describes the sample and data. Section 5 develops the estimation methodology, reports the results, and offers explanations for the findings. Section 6 presents the counterfactual analyses based on the estimation results. Section 7 provides cross-sectional and robustness analyses. Section 8 concludes.

2 ESG Incentives

2.1 What do ESG incentives look like?

I define ESG incentives as the component of compensation that varies with an ESG performance. An ESG incentive often involves an ESG metric, realization of which is assessed based on a set of targets. In terms of structure, it typically consists of a (1) threshold, a minimum level of performance that warrants any amount of compensation, a (2) target, the expected level of performance, and a (3) maximum, beyond which performance is no longer rewarded through compensation.

ESG incentives involve a wide variety of metrics. They include carbon emission intensity, energy efficiency, frequency of chemical leaks, water usage, and recycling. They are assessed on either an absolute or a relative basis, scaled by the firm's past performance (target ratcheting) or concurrent performance of comparable firms in the industry (relative performance evaluation). Contrary to the skepticism that ESG incentives are abstract and subjective, many firms use ESG incentives that are built on concrete structures with objective and measurable metrics. For example, a company using carbon emission as the metric has the following structure. It has a threshold of 2,124 kilotons (kt), a target of 1,865 kt, and a maximum of 1,772 kt. This means that the CEO will receive a bonus for any emission below 2,124 kt, increasing up to emissions below 1,772 kt.

I focus on carbon emission metrics because it is one of the most common metrics for environmental performance, and is observable to econometricians. My framework can account for complex incentive structures and be calibrated at the firm level to incorporate the heterogeneous incentive structures across firms, provided that I can perfectly observe the structures of contracts.

2.2 Compensation Structure with both ESG and Financial Incentives

No firm uses ESG incentives without any financial incentives. How do the ESG incentives affect compensation, combined with traditional financial incentives?

2023 annual bonus scorecard and outcome

For 2023, the committee assessed performance against a bonus scorecard of seven measures across three categories: safety and sustainability, operations and financials. These measures align with our strategy (see page 12) and were set out under the terms of our 2023 policy.

Safety and sustainability		Operations		Financials		Formulaic score
0.41	+	0.24	+	0.99	=	1.64 out of 2.0

Measures		Weighting	Threshold (0)	Target (1)	Maximum (2)	Outcome
Safety and sustainability (30% weight) ESG	Tier 1 process safety events★	15%	12 0	10 0.075	8 0.15	39 (tier 1: 9, tier 2: 30)
	Tier 2 process safety events★		47 0	39 0.075	31 0.15	0.26
	Sustainable emissions reductions★ (million tonnes)	15%	<7.77 0	7.97 0.15	8.27 0.3	7.973 0.15
Operations (20% weight)	bp-operated reliability and availability	10%	94.5% 0	95.5% 0.1	96.4% 0.2	95.7% 0.13
	Convenience & EV gross margin % growth	10%	4% 0	10% 0.1	16% 0.2	10.6% 0.11
Financials (50% weight)	Adjusted free cash flow★	25%	$12.6bn 0	$13.6bn 0.25	$14.6bn 0.5	$15.1bn 0.50
Financial/ Operational	Adjusted EBITDA	25%	$40.7bn 0	$42.2bn 0.25	$43.7bn 0.5	$43.67bn[a] 0.49
Formulaic score						1.64 out of 2.0

Formulaic scorecard outcome **1.64 out of 2.00** → Committee judgement for fatalities **- 0.05** → Final scorecard outcome **1.59 out of 2.00** → **79.5% of maximum**

Figure 1: Compensation Structure of British Petroleum

For illustrative purposes, I provide the compensation scheme of BP p.l.c in 2023[3], which consists of both ESG metrics and financial incentives. Within the performance range, the contract is linear in performance measures. Specifically, the compensation is a weighted average of ESG performance and financial (and operational) performance with weights of 30% and 70%, respectively. Two points are worth noting. First, the ESG incentive constitutes a substantial portion (30%) of variable compensation. Second, it is not trivial to meet ESG targets; CEOs at times fail to achieve them and lose a considerable amount of bonus for such failures. In this example, the CEO lost 7.5% of the maximum compensation because the firm's sustainable emission reduction of 7.973 million tonnes fell short of the maximum level of 8.27 million tonnes.

In my framework, I focus on two dimensions, environmental performance and financial performance for the feasibility of the estimation. In fact, in terms of the model, adding a third dimension

[3] https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/investors/bp-directors-remuneration-report-2023.pdf

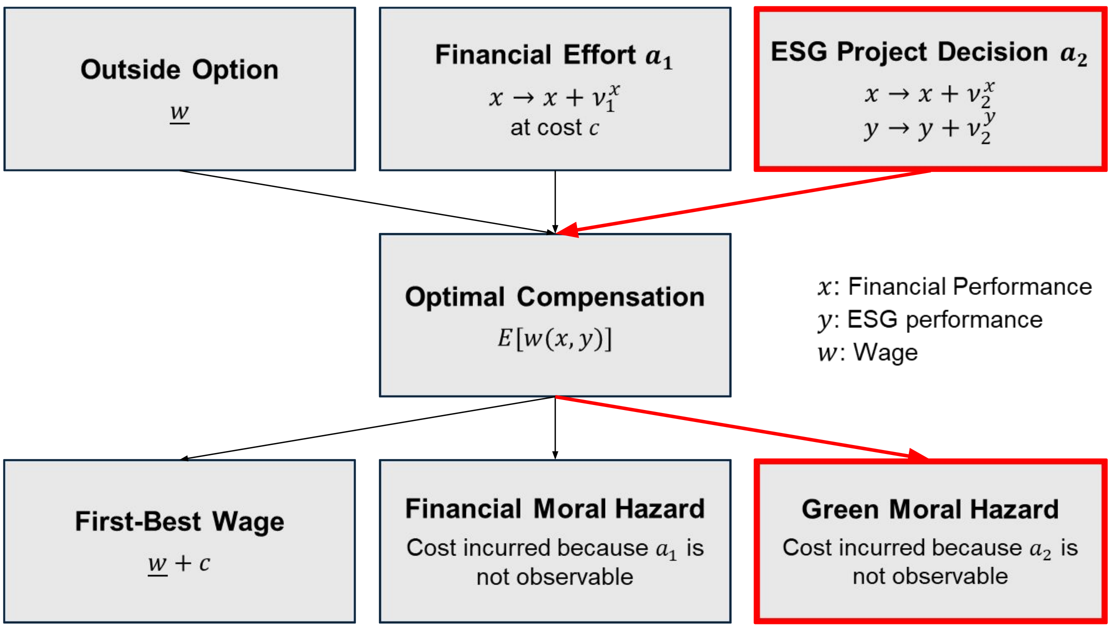
and beyond does not qualitatively change the dynamics or the implication. However, in terms of estimation, the quality is compromised by the curse of dimensionality, as I rely on nonparametric approach.

3 Model

Answering my research question, what are ESG incentivizing and at what cost, involves multiple challenges. First, adoption of ESG metrics in executive compensation contracts is inherently endogenous. Firms with ESG metrics and firms without are therefore not comparable, especially in terms of tradeoff between financial and ESG outcomes. In addition, it is unlikely that shocks to contracts do **not** simultaneously affect underlying economic tradeoffs. To address these challenges, I employ a structural approach.

In this section, I construct a conceptual framework for analyzing compensation contracts that incentivize both financial effort and ESG investment. I then solve the model and characterize the optimal contract.

3.1 Theoretical Framework and Model Setup



My conceptual framework features a simple principal-agent model, in which the agent's action is unobservable and can only be inferred from two observable and contractible signals: financial performance and non-financial performance. This setup is motivated by the fact that many firms, almost 40% by 2020, have started to explicitly include non-financial measures, on top of more traditional financial measures, in their compensation contracts.

The agent is risk-averse and therefore requires a premium on the risk coming from uncertainty in outcome realizations conditional on her effort. Given that the principal seeks to induce the agent's effort under the second-best, this risk premium constitutes the cost of moral hazard to the principal, incurred due to the effort being unobservable. Information about the agent's effort in the two signals, stock performance, and ESG performance, can mitigate the cost of moral hazard by reducing the uncertainty in wage faced by the agent conditional on her effort.

My model features a pure moral hazard problem in which the agent can take multi-dimensional actions. Specifically, the agent can take two types of actions: she can (1) choose to either exert costly effort to improve the financial performance of the firm or shirk and (2) choose to either accept or reject an investment project that affects both financial and non-financial outcomes.

Principal's Problem The principal is risk-neutral and has the objective $V(x, y)$, which is a function of both financial performance x and non-financial performance y. For simplicity, let the principal's objective $V(x, y)$ be a linear combination of financial outcome x and non-financial outcome y:

$$V(x, y) = x + ky \tag{1}$$

where k denotes the marginal loss in financial performance that the principal is willing to sacrifice for a marginal improvement in the non-financial performance. The principal maximizes her expected value less the expected wage to the agent:

$$\max_{w(\cdot)} \mathbb{E}[x - w(x, y)]. \tag{2}$$

Agent's Actions The agent can take two types of actions: $a = (a_1, a_2)$, where a_1 denotes financial effort that improves financial performance and a_2 denotes project choice that jointly affects financial and non-financial outcomes. As I assume a binary action space in each dimension, there are four combinations of actions: $a \in \{(0,0),(1,0),(0,1),(1,1)\}$.

Each combination of effort and investment decision yields a joint distribution $f_a(x,y)$ of the two outcomes. For tractability, I impose restrictions on how the agent's actions affect the outcome distribution. On one hand, I assume that financial effort a_1 only affects financial outcomes. With this assumption, I can disentangle incentives for actions that do not involve any tradeoff between financial and non-financial performances. Specifically, financial effort shifts the mean of financial outcome x by ν_1^x without affecting the unconditional distribution of y:

$$x_{11} = x_{01} + \nu_1^x \tag{3}$$

where x_a denotes a level of financial outcome x under effort a. In terms of joint density, the effect of financial effort a_1 can be expressed as:

$$f_{01}(x,y) = f_{11}(x + \nu_1^x, y) \tag{4}$$

On the other hand, I allow the ESG investment decision a_2 to have both financial and non-financial implications. Specifically, it shifts the means of financial outcome x and non-financial outcome y by ν_2^x and ν_2^y, respectively.

$$x_{11} = x_{10} + \nu_2^x \tag{5}$$

$$y_{11} = y_{10} + \nu_2^y \tag{6}$$

In terms of joint density, the effect of project decision a_2 can be expressed as:

$$f_{10}(x,y) = f_{11}(x + \nu_2^x, y + \nu_2^y) \tag{7}$$

Following the standard approach in the moral hazard literature, I assume that the agent's action involves personal cost, c_a. Specifically, agent's action $a = (a_1, a_2)$ imposes personal cost c_a to the agent, with c_{00} normalized to 0. Given the nature of each decision, I assume that financial effort a_1 is personally costly to the agent, whereas project choice a_2 is not. Let c denote the personal cost of financial effort. Then, effort cost can be summarized as follows:

$$c_{01} \equiv c_{00} = 0 \tag{8}$$

$$c_{11} \equiv c_{10} \equiv c \tag{9}$$

That the ESG investment does not incur a personal cost to the agent, however, does not necessarily mean that project choice a_2 is not costly to the agent: as a_2 affects the joint distribution of x and y, it thereby affects the distribution of wage $w(x, y)$ conditional on the choice of action.

To summarize, the financial effort is personally costly to the manager and only has financial implications, while the project decision imposes no direct cost to the manager and has both financial and non-financial implications.

Agent's Preference Finally, the agent is risk-averse and has a CARA utility:

$$u(w, a) \equiv -e^{-\rho(w-c_a)} \tag{10}$$

with c being cost of effort in "dollars" and ρ is risk-aversion. Let $C \equiv e^{\rho c}$ be the cost in utility. This assumption, used in a number of other structural works (Gayle and Miller (2009), Gayle and Miller (2015), Bertomeu et al. (2023a)) in the executive compensation literature, helps make the estimation feasible, as the wealth of executives is often unobservable. This also allows for dynamic implications, as shown by Holmstrom and Milgrom (1991).

Principal's Preferred Action I assume that it is optimal for the principal to induce both financial effort and project acceptance: $a^* = (1, 1)$. This assumption is based on two relevant features of the data: (1) weight on non-financial outcome is positive and (2) financial performance and non-financial performance are positively correlated.[4] Had the principal been using non-financial performance to induce financial effort, the weight on the non-financial performance should have been negative given its positive correlation with the financial performance.[5]

[4]One potential explanation for the positive correlation is that, for the same level of cash flow performance, investors may have preference for favorable non-financial performance and therefore reward it with stock returns.

[5]I provide a more detailed discussion of this argument in Appendix C.

Discussion of Model Assumptions The assumption of the principal's value $V(x, y)$ being a linear combination of financial performance x and y does not play a significant role in the generalized model, because I am not estimating the principal's objective function. Any value function that is increasing in non-financial performance y at a sufficient rate (i.e. "cares sufficiently about y") for the principal to prefer implementing the ESG project will yield the same optimal contract in the generalized model. I make this assumption for its intuitive appeal and tractability in the simplified model.

Recall that I make two sets of assumptions regarding the agent's actions: first on how they transform the outcome distributions and second on how they fundamentally differ from each other. While the assumption that both actions affect only the means of performances x and y abstracts away from agent's actions having risk (and higher moment) implications, it ensures that the model is identified and thus can be estimated from data.

The assumption that, financial effort only affects financial outcome x and ESG project selection has both financial and non-financial implications, might seem as an oversimplification. However, this setting can be mapped into the following in practice: ESG project selection corresponds to decisions by the manager to improve ESG outcomes that **can** be optimally implemented with a contract.

For instance, ESG project selection could be a firm's decision to install a costly air purifier in its incinerator, which will reduce carbon emissions but also reduce financial profits. Note that this project will likely not be accepted without an ESG incentive that rewards ESG performance. In contrast, financial effort in my model refers to actions that will be taken regardless of ESG incentives. For example, a decision to build a new plant for the firm's main operation, which may have ESG implications, will likely not be deterred by ESG incentives. This is how my model distinguishes ESG projects from financial efforts. This distinction, along with the assumption that ESG project selection is costless, allows for disentangling incentives for ESG outcomes from those for financial outcomes.

Other main assumptions, including actions being binary, are standard in the literature on structural estimation of compensation contracts.

3.2 Contracting Problem

The problem of the principal, who wants to implement both financial effort and project acceptance, is as follows:

$$\max_{w(\cdot)} \mathbb{E}[V(x,y) - w(x,y)|a = (1,1)]. \tag{11}$$

s.t.

$$\mathbb{E}[u(w(x,y),1)|a = (1,1)] \geq \mathbb{E}[u(w(x,y),1)|a = (1,0)] \tag{IC10}$$

$$\mathbb{E}[u(w(x,y),1)|a = (1,1)] \geq \mathbb{E}[u(w(x,y),0)|a = (0,1)] \tag{IC01}$$

$$\mathbb{E}[u(w(x,y),1)|a = (1,1)] \geq \mathbb{E}[u(w(x,y),0)|a = (0,0)] \tag{IC00}$$

$$\mathbb{E}[u(w(x,y),1)|a = (1,1)] \geq u(\underline{w},(0,0)) \tag{P}$$

The first order condition provides the relation among the outcome distributions, one under the optimal effort and others under the alternative levels of effort:

$$\mu_{10} C \frac{f_{10}(x,y)}{f_{11}(x,y)} + \mu_{01} \frac{f_{01}(x,y)}{f_{11}(x,y)} = C(\lambda + \mu_{10} + \mu_{01}) - \frac{1}{\rho} e^{\rho w(x,y)} \tag{FOC}$$

Binding incentive compatibility constraints provide:

$$C \int_x \int_y e^{-\rho w(x,y)} f_{11}(x,y) dydx = C \int_x \int_y e^{-\rho w(x,y)} f_{10}(x,y) dydx \tag{IC10}$$

$$C \int_x \int_y e^{-\rho w(x,y)} f_{11}(x,y) dydx = \int_x \int_y e^{-\rho w(x,y)} f_{01}(x,y) dydx \tag{IC01}$$

Binding participation constraint gives:

$$C \int_x \int_y e^{-\rho w(x,y)} f_{11}(x,y) dydx = e^{-\rho \underline{w}} \tag{P}$$

Moreover, as $f_{10}(x,y)$ and $f_{01}(x,y)$ are probability distribution functions, they should integrate to 1:

$$\int_x \int_y f_{10}(x,y) dydx = 1 \tag{12}$$

$$\int_x \int_y f_{01}(x,y)dydx = 1 \tag{13}$$

3.3 Optimal Contract

From the first order condition, the optimal wage is given as follows:

$$w(x,y) = \frac{1}{\rho}\log\left(\rho C(\lambda + \mu_{10} + \mu_{01}) - \rho C\mu_{10}\frac{f_{10}(x,y)}{f_{11}(x,y)} - \rho\mu_{01}\frac{f_{01}(x,y)}{f_{11}(x,y)}\right) \tag{14}$$

A key observation from the equation above is that the more likely an outcome (x,y) is under actions other than the one prescribed by the contract, the lower the wage. This means that the shape of the wage function is informative about the likelihood ratio across different actions, and therefore the shapes of the counterfactual distributions.

Based on the structure of the compensation in the equation above, the highest possible wage $\overline{w}$ is rewarded to (x,y) that perfectly signals $a = (1,1)$:

$$w(x,y) \leq \overline{w} = \frac{1}{\rho}\log\left(\rho C(\lambda + \mu_{10} + \mu_{01})\right) \tag{15}$$

It can also be seen that, given the base parameters ρ and C, the wage function is determined by shadow costs λ, μ_{10}, and μ_{01}. λ can be readily solved for by combining the first order condition with the binding participation constraint and the incentive compatibility constraints:

$$\lambda = \frac{1}{\rho}e^{\rho\underline{w}} \tag{16}$$

The equation above is consistent with the intuition that the higher value of outside options to the agent makes it costlier to induce the agent to participate in the contract.

On the other hand, it is difficult to obtain analytical expressions for μ_{10} and μ_{01} without making additional assumptions regarding the likelihood ratios across actions. Therefore, for the analysis of the optimal contract to follow, I numerically solve for μ_{10} and μ_{01} that jointly satisfy the binding participation constraint and the incentive compatibility constraints.

In order to verify the optimality of the contract, I examine the second-order condition. Given that $\rho > 0$, $f_{11}(x,y) > 0$ for all (x,y) within support, and $e^{-\rho w(x,y)} > 0$ for any real $w(x,y)$, the second-order condition can be written as:

$$\rho C(\lambda + \mu_{10} + \mu_{01}) - \rho C \mu_{10} \frac{f_{10}(x,y)}{f_{11}(x,y)} - \rho \mu_{01} \frac{f_{01}(x,y)}{f_{11}(x,y)} > 0 \tag{SOC}$$

As any wage $w(x, y)$ that violates the SOC is complex, any wage $w(x, y)$ that is real for every (x, y) should satisfy the SOC.

The figure below plots a sample optimal wage. First, it can be seen that the wage increases both in financial performance x and ESG performance y. This is because higher (x, y) strongly signals both the financial effort and execution of the ESG project. Second, the wage exhibits a non-linear, concave structure, as implied by equation 14.

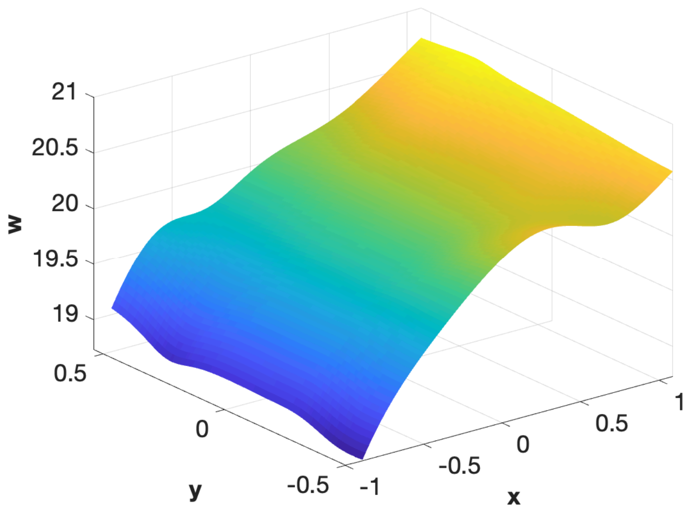
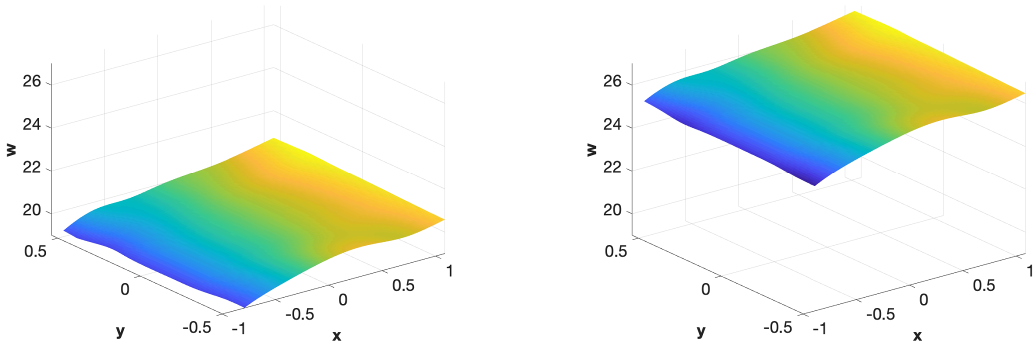


Figure 2: Optimal compensation $w(x, y)$ for a sample set of parameters

3.4 Comparative Statics

In this section, I provide comparative statics of the model, to provide better understanding of how each parameter affects the optimal contract.

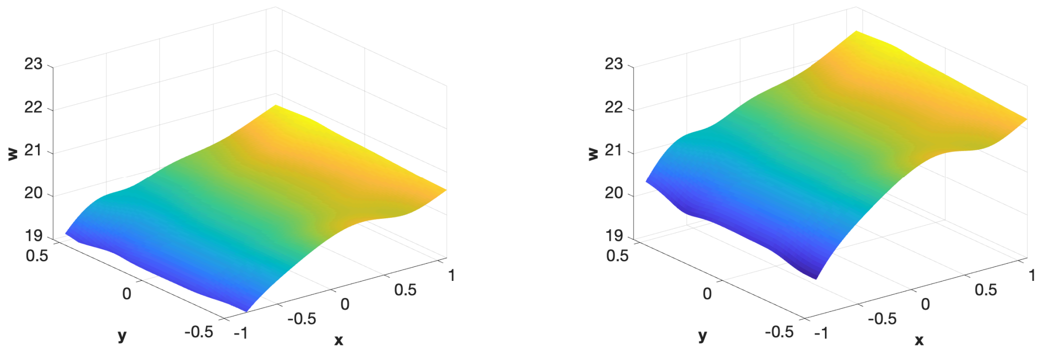
Figures below show how the value of outside option $\underline{w}$ affects the optimal compensation. It can be seen that the value of outside option shifts the level of the wage without affecting the shape. In fact, increase in the value of outside option results in a dollar-for-dollar increase in the level of wage. This is natural, considering that the outside option affects only the incentive to participate in the contract.



(a) Baseline $w(x, y)$ (b) $w(x, y)$ with higher $\underline{w}$

Figure 3: **Optimal compensations $w(x, y)$ under baseline parameters (Panel a) and under higher outside option $\underline{w}$ (Panel b)**

Figures below show how the cost of effort c affects the optimal compensation. It can be seen that an increase in the cost of effect increases both the variance and the level of the wage. For the contract to be incentive compatible with respect to financial effort a_1, the sensitivity of the wage with respect to financial performance x increases in the cost of effort, thus increasing the variance of the wage. The risk-averse agent should then be offered risk premium for this added risk in wage, thereby increasing the level of the wage.



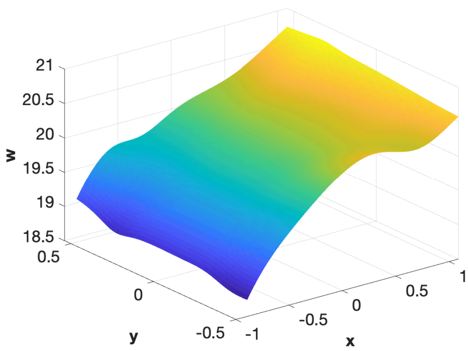
(a) Baseline $w(x, y)$ (b) $w(x, y)$ with higher c

Figure 4: **Optimal compensations $w(x, y)$ under baseline parameters (Panel a) and under higher cost of effort c (Panel b)**

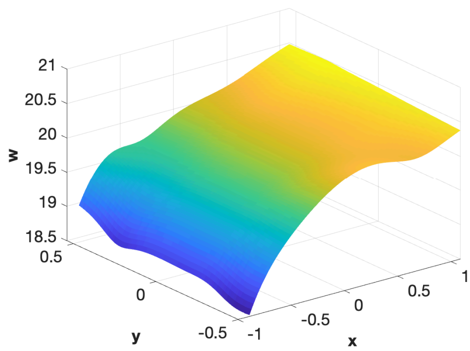
In the remainder of this section, I discuss how effects of agents actions, ν_1^x, ν_2^x, and ν_2^y, affect the optimal contract. An important caveat worth noting is that their effects come primarily through the changes in the likelihood ratios, which depend heavily on the shape of the distribution function

$f_{11}(x, y)$ and the location of the parameters. Therefore, I focus only on the local effects around the given parameters, for the empirical distribution observed in the data.

Figures below show how the effect of financial effort (ν_1^x) affects the optimal compensation. It can be seen that an increase in the effect of financial effort reduces both the variance and the level of the wage. Higher effect of financial effort locally amplifies the difference between $f_{11}(x, y)$ and $f_{01}(x, y)$, and thus the likelihood ratio between the two distributions. In other words, financial performance better signals financial effort, allowing the compensation to be less sensitive with respect to financial performance. As a result, both the variance and the risk premium in wage are lower.



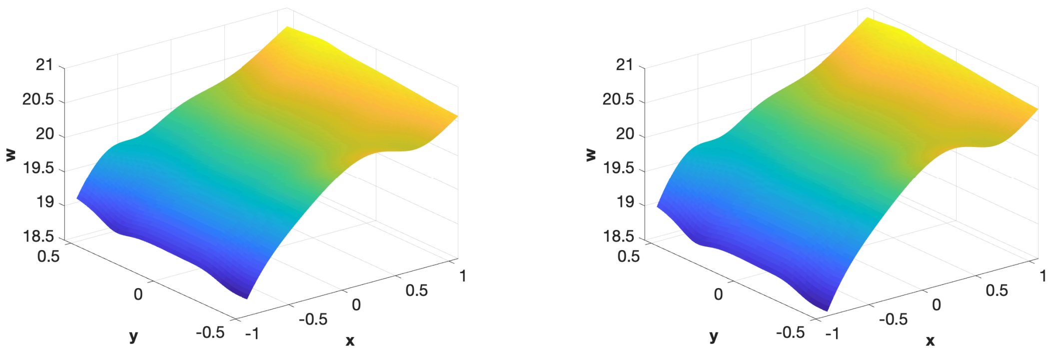
(a) Baseline $w(x, y)$



(b) $w(x, y)$ with higher ν_1^x

Figure 5: **Optimal compensations $w(x, y)$ under baseline parameters (Panel a) and under higher effect of effort ν_1^x (Panel b)**

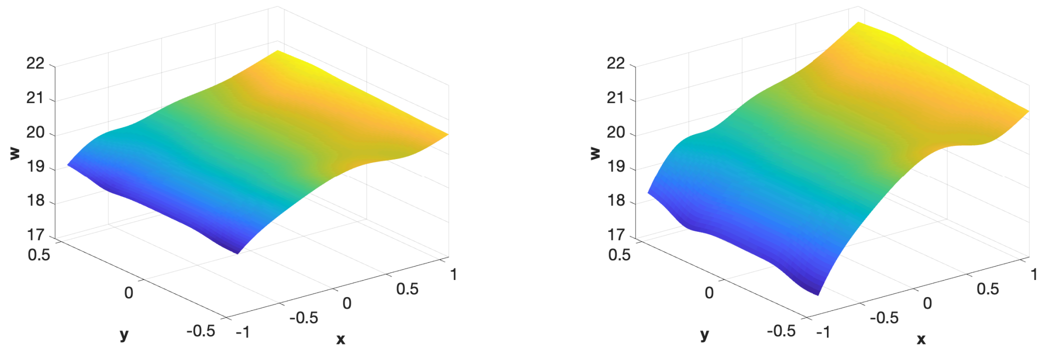
Figures below show how the financial cost of ESG project ($|\nu_2^x|$) affects the optimal compensation. It can be seen that an increase in the financial cost of ESG project increases both the variance and the level of the wage. As the ESG project entails steeper financial sacrifices, the agent will require greater rewards to ESG performances to counteract the disincentive from financial incentives, for the ESG project to be incentive compatible. As a result, the risk premium should also increase to cover the risk added by the incentive on ESG performance.



(a) Baseline $w(x, y)$ (b) $w_{cf}(x, y)$ with higher $|\nu_2^x|$

Figure 6: **Optimal compensations $w(x, y)$ under baseline parameters (Panel a) and under higher financial cost of ESG project $|\nu_2^x|$ (Panel b)**

Figures below show how the ESG effect of ESG project (ν_2^y) affects the optimal compensation. It can be seen that an increase in the effect of financial effort increases both the variance and the level of the wage.



(a) Baseline $w(x, y)$ (b) $w_{cf}(x, y)$ with higher ν_2^y

Figure 7: **Optimal compensations $w(x, y)$ under baseline parameters (Panel a) and under higher ESG effect of ESG project ν_2^y (Panel b)**

3.5 Identification and Assumptions

For the estimation to be feasible, I make one additional assumption. I assume that a high enough outcome in each dimension must be due to high effort in each dimension:

$$\lim_{y \to \infty} \frac{f_{10}(x, y)}{f_{11}(x, y)} = 0 \tag{17}$$

$$\lim_{x\to\infty} \frac{f_{01}(x,y)}{f_{11}(x,y)} = 0 \tag{18}$$

This means that extremely favorable outcome in financial performance x and non-financial performance y perfectly signals financial effort a_1 and ESG project selection a_2, respectively. The assumption allows me to use wages for extremely favorable outcomes to infer the benchmark when moral hazard in each dimension is not present.

From the first order condition (FOC), binding constraints, and the assumption above, I obtain the following five moment conditions, from which I estimate $(C, \underline{w}, \lambda, \mu_{10}, \mu_{01})$ from the following five moment conditions.

$$\begin{bmatrix} \frac{1}{C}e^{-\hat{\rho}\underline{w}} \\ \hat{\rho}(\lambda C + \mu_{01}(C-1)) \\ \frac{1}{\hat{\rho}\lambda C} \\ \hat{\rho}(\lambda + \mu_{10} + \mu_{01})C \\ \hat{\rho}(\lambda + \mu_{10} + \mu_{01})C - \mu_{01} \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \alpha \\ \gamma \\ \delta \end{bmatrix}, \tag{19}$$

where

$$\alpha = \mathbb{E}[e^{-\hat{\rho}w(x,y)}] \tag{20}$$

$$\beta = \mathbb{E}[e^{\hat{\rho}w(x,y)}] \tag{21}$$

$$\gamma = e^{\hat{\rho}\bar{w}} \tag{22}$$

$$\delta = \lim_{y\to\infty} \mathbb{E}[e^{\rho w(x,y)}|y] \tag{23}$$

The first moment α is the agent's expected utility (reversed sign) given wage $w(x,y)$ and outcome distribution $f_{11}(x,y)$. The second moment β captures the expected level of the wage to the agent. The third moment γ effectively represents the theoretical upper bound of the wage. The fourth moment δ captures the expected level of wage under extremely high ESG performance.

By inverting the moment conditions above, I obtain the following analytical expressions for the parameters:

$$\begin{bmatrix} \underline{w} \\ c \\ \lambda \\ \mu_{10} \\ \mu_{01} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\hat{\rho}} \log\left(\frac{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma-\delta)}{\hat{\rho}(\gamma-\delta)}\right) \\ \frac{1}{\hat{\rho}} \log\left(\frac{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma-\delta)}{\hat{\rho}\alpha(\gamma-\delta)}\right) \\ \frac{(\gamma-\delta)}{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma-\delta)} \\ \frac{\alpha(\gamma-\delta)(\hat{\rho}\delta-\beta)}{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma-\delta)} \\ \gamma - \delta \end{bmatrix} \tag{24}$$

Then, I estimate the shift parameters $(\nu_1^x, \nu_2^x, \nu_2^y)$ with the following moment conditions derived from the incentive compatibility condition with respect to financial effort (IC01) and the first order condition (FOC):

$$\begin{bmatrix} \frac{1}{C}\int_x \int_y e^{-\hat{\rho}w(x,y)} f_{11}(x+\nu_1^x, y)dydx \\ \mu_{10}C\nu_2^x + \mu_{01}\nu_1^x \\ \mu_{10}C\nu_2^y \end{bmatrix} = \begin{bmatrix} \alpha \\ \frac{1}{\rho}\eta_x - (C\lambda + (C-1)\mu_{01})m_x \\ \frac{1}{\rho}\eta_y - (C\lambda + (C-1)\mu_{01})m_y \end{bmatrix}, \tag{25}$$

where

$$\eta_x = \int_x \int_y x e^{\hat{\rho}w(x,y)} f_{11}(x,y)dydx \tag{26}$$

$$\eta_y = \int_x \int_y y e^{\hat{\rho}w(x,y)} f_{11}(x,y)dydx \tag{27}$$

$$m_x = \int_x \int_y x f_{11}(x,y)dydx \tag{28}$$

$$m_y = \int_x \int_y y f_{11}(x,y)dydx \tag{29}$$

By substituting the above expression for C into (IC01), I get the following condition for ν_1^x:

$$\int_x \int_y e^{-\hat{\rho}w(x,y)} f_{11}(x+\nu_1^x, y)dydx = \frac{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma-\delta)}{\hat{\rho}(\gamma-\delta)} \tag{30}$$

While ν_1^x cannot be analytically solved for without distributional assumptions, it can still be numerically estimated.

With ν_1^x pinned down, I can solve for ν_2^x and ν_2^y:

$$\nu_2^x = \frac{1}{\mu_{10} C}\left(\frac{1}{\rho}\eta_x - (C\lambda + (C-1)\mu_{01})m_x - \mu_{01}\nu_1^x\right) \tag{31}$$

$$\nu_2^y = \frac{1}{\mu_{10} C}\left(\frac{1}{\rho}\eta_y - (C\lambda + (C-1)\mu_{01})m_y\right) \tag{32}$$

3.6 Intuition for Identification

In this section, I provide intuition for how the features of the observed wage function map into the underlying parameters of the model.

I begin with the parameter that relates to the participation constraint, the value of outside option $\underline{w}$ and the shadow cost of participation λ. Recall from equation 14 that the participation constraint affects the wage function only through the level, not the shape. It is therefore clear that the role of λ is merely matching the level of the observed wage that is not explained by other parameters that simultaneously affect both the level and the variation of the wage. From equation 16, it is clear that $\underline{w}$ and λ are effectively interchangeable, given risk aversion parameter ρ.

Identification for the shadow costs of incentive compatibility conditions, μ_{01} for the financial effort and μ_{10} for the ESG project, comes from the differences in the level of wages under normal vs extremely favorable outcomes that almost perfectly signal agent's actions. When both financial performance x and ESG performance y are extremely favorable ($x \to \infty$ and $y \to \infty$), it is clear that the agent took both financial effort a_1 and ESG project a_2. (Assumptions in equations 17 and 18) Then, maximum wage γ paid to the agent, would reflect neither μ_{01} nor μ_{10}. When ESG performance is extremely favorable ($y \to \infty$), the outcome only signals that the agent took a_2, but not necessarily a_1. Here, expected wage under extremely favorable ESG performance δ would reflect only μ_{01}. Therefore, the difference between γ and δ provides μ_{01}. Across all outcomes, expected level of wage β should reflect both shadow costs, μ_{01} and μ_{10}. Thus, the difference between δ and β provides μ_{10}. The figure below summarizes this intuition.

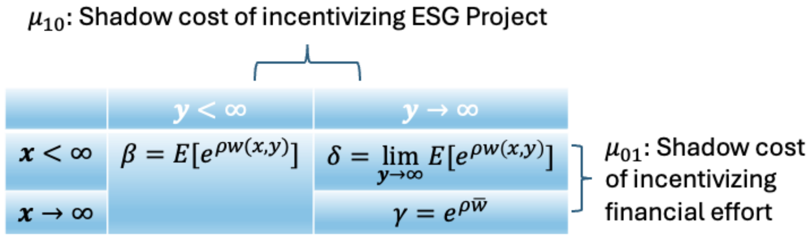


Figure 8: Identification of IC Shadow Costs

The cost of effort c is related to the wage variance. Product between moments α, which is the the (negative) expected utility of the agent and β, which is the exponential trasnformation of the wage that captures the level, provides insight into identifying $C = e^{\rho c}$:

$$\alpha\beta = \mathbb{E}[e^{-\rho w(x,y)}] \cdot \mathbb{E}[e^{\rho w(x,y)}] = 1 + \frac{\mu_{01}}{\lambda}\left(1 - \frac{1}{C}\right) \tag{33}$$

While it is difficult to analyze the product of the expectation for any arbitrary distribution, restricting to normal distributions for outcomes and linear compensation schemes provides the following equality:

$$e^{\rho Var(w(x,y))} = 1 + \frac{\mu_{01}}{\lambda}\left(1 - \frac{1}{C}\right) \tag{34}$$

The lefthand side captures the disutility of wage risk to the agent, while the righthand side increases in the cost of effort. This suggests that volatile wage is consistent with high cost of effort. The intuition is that the agent requires high-powered incentives when the cost of effort is high. For the same effect of financial effort ν_1^x, the agent should get higher rewards to favorable outcomes when the cost of effort is higher, for the contract to be incentive compatible. This in turn increases the incentive power for ESG project, as the agent should be compensated for the loss of financial performance caused by the ESG project. In summary, increase in the cost of effort leads to increase in the incentive power for both financial and ESG outcomes, resulting in a more volatile compensation structure overall.

Effect of financial effort ν_1^x is identified from the incentive compatibility condition on financial effort. When the constraint binds, the expected utility of the agent should be equal between exerting financial effort and shirking, as shown in equation IC01. Rewriting the equation through change of variables gives:

$$C = \frac{\mathbb{E}\left[e^{\rho w(x-\nu_1^x, y)}\right]}{\mathbb{E}\left[e^{\rho w(x,y)}\right]} \tag{35}$$

The intuition here is simply that increase in the expected utility from wage should match the disutility of exerting costly effort. It can be seen here that the sensitivity of wage to financial outcome and ν_1^x are substitutes, in terms of how they affect the expected utility under shirking relative to that under exerting financial effort. Therefore, low sensitivity of wage to financial effort is consistent with high ν_1^x.

Financial and ESG effects of ESG project, ν_2^x and ν_2^y, come from the covariance between the level of the wage and performance in each dimension. Rewriting equations 31 and 32 gives:

$$Cov(e^{\rho w(x,y)}, x) = (\gamma - \delta)\nu_1^x + (\delta - \beta)\nu_2^x \tag{36}$$

$$Cov(e^{\rho w(x,y)}, y) = (\delta - \beta)\nu_2^y \tag{37}$$

Recall that $\gamma - \delta$ and $\delta - \beta$ capture shadow costs of incentivizing financial effort and ESG project, respectively. It can be seen that the covariance between the level of the wage and each performance outcome is a linear combination of effects of effort, weighed by respective shadow costs.

4 Data

4.1 Sample Construction

I merge various datasets to construct a firm-year panel of compensation, financial performance, and environmental performance covering over 600 firms in U.S. from 2012 to 2022, the longest overlapping time period. The five main sources of data are as follows:

Measurement of Compensation To measure the change of CEO's wealth due to compensation, I follow the standardized approach introduced in Bertomeu et al. (2023a). First, I begin with all cash and non-equity compensation from Execucomp, including salary, bonus, and long-term incentives. Second, I add the change in wealth due to stock compensation, both restricted and owned. To this end, I use the stock holdings from Execucomp, as well as the return information from CRSP-

Compustat. Third, I add the change in wealth due to option compensation. I use the option holdings from Execucomp, and inputs of the Black-Scholes formula from CRSP.

Compensation Metrics Data To obtain firm compensation metric data, I use Executive Compensation Analytics (ECA) aggregated by the Institutional Shareholder Services (ISS). This data set is annual, from 2009 to 2022, and it comes from firms' disclosures of executive compensation.

Carbon Emission Data For data on firms' greenhouse gas (GHG) emissions, I use Trucost aggregated by the S&P Global. To construct the measure of firms' ESG performance, I use the scope 1 and 2 emission intensity, following the literature.[6] This data set is annual. For the performance measure, I use the negative log change in emission intensity, to capture the reduction in emissions.

Stock Return Data I obtain stock return information from CRSP. As a measure of financial performance, I construct the abnormal return as the return over the firm-year less the concurrent market return, following Gayle and Miller (2015).

Firm Financial Data I obtain accounting and financial information from the Compustat.

4.2 Descriptive Statistics

Table 1 displays the summary statistics of key variables. I make use of the following three variables in estimation: wage, abnormal return as a proxy of financial performance, and emission reduction as a proxy of ESG performance. I construct the abnormal return variable by subtracting the contemporaneous market return. To construct emission reduction, I compute the *negative* of the log change in scope 1 and 2 emission intensity. I take the negative so that the positive value of the variable can be interpreted as an improvement in terms of environmental performance.

[6]e.g., Bolton and Kacperczyk (2021) and Jung et al. (2021) among others.

	Mean	St.Dev.	25th percentile	75th percentile	Count
Total Pay	22.15	31.82	3.30	28.40	1419
Abnormal Return	0.00	0.29	-0.21	0.19	1419
Emission Reduction	0.03	0.12	-0.03	0.07	1419
Log Size	8.81	1.60	7.66	10.01	1419
ROA	0.04	0.11	0.01	0.08	1419
Log Emission	12.30	2.68	10.45	14.19	1419
Observations	1419				

Table 1: Summary Statistics

To verify the identification assumptions, I examine the correlations among wage, abnormal return, and emission reduction. Table 2 shows that although the pair-wise correlations are low, they are positive. Following from the discussion in Appendix C, these positive correlations suggest that the incentive compatibility for ESG investment (IC10) is indeed binding and that ESG investment entails a negative financial impact ($\nu_2^x < 0$).

	Total Pay	Abnormal Return	Emission Reduction
Total Pay	1.00		
Abnormal Return	0.49	1.00	
Emission Reduction	0.11	0.08	1.00

Table 2: Correlations

5 Estimation

5.1 Non-parametric Estimation of Density and Wage Functions

The first step of the estimation is estimating $f_{11}(x, y)$, the joint density of (x, y) conditional on action $a = (1, 1)$ stipulated in the contract, and $w(x, y)$, the wage function. For the joint density, I use a bivariate kernel density estimator with a standard normal kernel and bandwidths (h_x, h_y):

$$\hat{f}_{11}(x, y) = \frac{1}{N h_x h_y} \sum_{i=1}^{N} \phi\left(\frac{x - X_i}{h_x}\right) \phi\left(\frac{y - Y_i}{h_y}\right) \tag{38}$$

Where bandwidths (h_x, h_y) with smoothing factor f_f are given as:

$$h_x = f_f \cdot \hat{\sigma}_x \cdot N^{\frac{1}{6}} \tag{39}$$

$$h_y = f_f \cdot \hat{\sigma}_y \cdot N^{\frac{1}{6}} \tag{40}$$

For the wage function, I use a bivariate Nadaraya-Watson Estimator with a standard normal kernel and bandwidths (h'_x, h'_y) :

$$\hat{w}(x,y) = \frac{\sum_{i=1}^{N} \phi\left(\frac{x-X_i}{h'_x}\right) \phi\left(\frac{y-Y_i}{h'_y}\right) W_i}{\sum_{i=1}^{N} \phi\left(\frac{x-X_i}{h'_x}\right) \phi\left(\frac{y-Y_i}{h'_y}\right)} \tag{41}$$

Where bandwidths (h'_x, h'_y) with smoothing factor f_w are given as:

$$h'_x = f_w \cdot \hat{\sigma}_x \cdot N^{\frac{1}{6}} \tag{42}$$

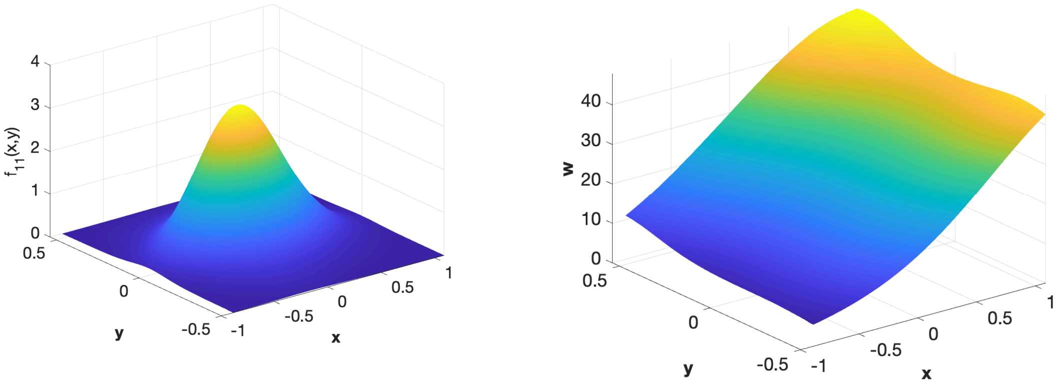
$$h'_y = f_w \cdot \hat{\sigma}_y \cdot N^{\frac{1}{6}} \tag{43}$$

I use a smoothed bandwidth for estimations of the wage function and the distribution function, as the rule-of-thumb bandwidth tends to over-fit the data.

5.2 Parameter Estimation

The second step is to estimate the parameters $(C, \underline{w}, \nu_1^x, \nu_2^x, \nu_2^y, \lambda, \mu_{10}, \mu_{01})$ from the estimated joint density $\hat{f}_{11}(x,y)$ and wage function $\hat{w}(x,y)$. For a given level of risk aversion ρ, I compute the moments $(\alpha, \beta, \gamma, \delta)$ in equation 20. From equation 24, I get estimates for $(C, \underline{w}, \lambda, \mu_{10}, \mu_{01})$, immediately from moments $(\alpha, \beta, \gamma, \delta)$. As there is no analytical expression for ν_1^x, I numerically estimate the parameter from the condition in Equation 30. With ν_1^x pinned down, I estimate the remaining parameters, ν_2^x and ν_2^y, from equations 31 and 32.

5.3 Estimation Results



(a) $\hat{f}_{11}(x, y)$ (b) $\hat{w}(x, y)$

Figure 9: **Nonparametric estimation of density $\hat{f}_{11}(x, y)$ (Panel a) and wage $\hat{w}(x, y)$ (Panel b)** x denotes stock performance, y denotes ESG signal,and a denotes agent's effort

The figures above present the nonparametrically estimated joint density function $\hat{f}_{11}(x, y)$ and wage function $\hat{w}(x, y)$. As discussed in Section 2.2, the wage function is increasing both in the financial signal x and non-financial signal y. This, in conjunction with x and y being positively correlated, suggests that incentive compatibility with respect to ESG investment binds and that ESG investment entails a negative financial impact ($\nu_2^x < 0$).

Parameter	Estimate
$\underline{w}$: Value of outside option (\$mil.)	19.0356
c : Effort cost (\$mil.)	0.80851
ν_1^x : Financial effect of financial effort a_1	0.052
ν_2^x : Financial effect of ESG project a_2	-0.013059
ν_2^y : ESG effect of ESG project a_2	0.017575

Table 3: Parameter estimates

Table 3 above the parameter estimates for the sample with ESG compensation and the entire sample, respectively. For the benchmark risk aversion of $\rho = 0.08$[7], the value of outside option $\underline{w}$ and cost of effort c are estimated at \$19.04 million and \$808,510, respectively.

[7]Based on the median risk aversion of 1 found by Brenner (2015), I adjust my risk aversion parameter at $1/12 \approx 0.08$.

I find that financial effort substantially improves financial performance by 5.2% of stock return. The magnitude is consistent with estimates from prior literature, including Gayle and Miller (2015).

As for the ESG project, I find that it entails a tradeoff between 1.3% loss of stock return and 1.76% improvement in carbon emission intensity reduction per year. Based on this estimates, I infer that firms with ESG metrics in their compensation are making tangible financial sacrifices, around 25% of the value created by CEO's financial effort, to improve ESG performances.

In terms of the principal's preference, the result provides a lower bound on the value that the principal places on improvement in ESG performance. Specifically, the principal values 1% reduction in carbon emission intensity at approximately 0.74% of firm value.

This result suggests that the board values ESG performance more than do the general investors (or the marginal asset pricer) in the capital market. While this does not necessarily indicate a misalignment as the preference of those investing in firms with ESG incentives may differ from that of the general investors, it does still suggest that the "greenium" may be larger for the boards of these firms than for green investors.

5.4 Inference

Based on 1,000 bootstrap simulations, I construct confidence intervals around the parameter estimates. As a result, all parameters' signs are statistically significant at the 90% level.

Parameter	90 Percent CI	95 Percent CI
$\underline{w}$: Value of outside option ($mil.)	(17.8 20.4)	(17.6 20.8)
c : Effort cost ($mil.)	(0.261 1.18)	(0.207 1.33)
ν_1^x : Financial effect of financial effort a_1	(0.02 0.076)	(0.016 0.085)
ν_2^x : Financial effect of ESG project a_2	(-0.049 -0.002)	(-0.060 0.001)
ν_2^y : ESG effect of ESG project a_2	(0.010 0.035)	(0.009 0.040)

Table 4: Estimated 90% and 95% confidence intervals for parameters

Figure 10 below plots the distribution of parameters, as well as the cost of moral hazard, from the bootstrap simulations. I find that simulation estimates are generally distributed around the parameters from my main estimation.

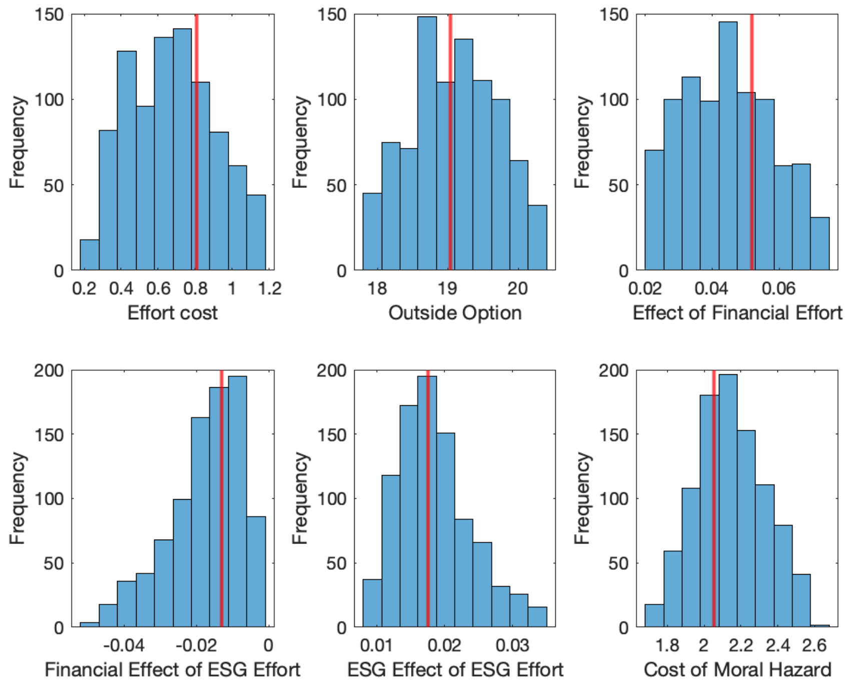


Figure 10: Bootstrap Results

6 Counterfactual Analysis: Decomposing Moral Hazard

I define the cost of moral hazard ΔV as the expected wage the principal should offer the agent in excess of her first-best wage, which is the sum of the effort cost and the value of the outside option:

$$\Delta V = \mathbb{E}[w(x, y)|a = (1, 1)] - (c + \underline{w}) \tag{44}$$

To answer the question of how costly it is to incentivize a manager to exert ESG effort on top of financial effort, I decompose the cost of moral hazard separately for each effort. I define the cost of green moral hazard as the cost incurred to the principal because the principal cannot observe the agent's ESG investment. Let $w_{cf}(x, y)$ denote the counterfactual wage necessary to implement both financial and ESG investment when ESG investment is observable but financial effort is not. The cost of green moral hazard ΔV_G is therefore given as:

$$\Delta V_G = \mathbb{E}[w(x, y) - w_{cf}(x, y)|a = (1, 1)] \tag{45}$$

Then, the cost of financial moral hazard ΔV_F is naturally given as the remaining portion of the cost of moral hazard:

$$\Delta V_F = \Delta V - \Delta V_G = \mathbb{E}[w_{cf}(x,y)|a = (1,1)] - (c + \underline{w}) \tag{46}$$

In order to compute the cost of green moral hazard, I solve for the counterfactual contract that implements both financial and ESG efforts when ESG effort is observable but financial effort is not. This counterfactual contract should solve:

$$\max_{w(\cdot)} \mathbb{E}[V(x,y) - w(x,y)|a = (1,1)]. \tag{47}$$

s.t.

$$\mathbb{E}[u(w(x,y),1)|a = (1,1)] \geq \mathbb{E}[u(w(x,y),0)|a = (0,1)] \tag{IC}$$

$$\mathbb{E}[u(w(x,y),1)|a = (1,1)] \geq u(\underline{w},(0,0)) \tag{P}$$

The first order condition then gives:

$$\begin{aligned} 1 = {} & \lambda_{cf}\rho C e^{-\rho w_{cf}(x,y)} \\ & + \mu_{cf}\rho \left(C e^{-\rho w_{cf}(x,y)} - e^{-\rho w_{cf}(x,y)} \frac{f_{01}(x,y)}{f_{11}(x,y)} \right) \end{aligned} \tag{FOC'}$$

The above can be rearranged to yield the counterfactual wage function $w_{cf}(x,y)$:

$$w_{cf}(x,y) = \frac{1}{\rho} \log \left(\rho \left(C(\lambda_{cf} + \mu_{cf}) - \mu_{cf} \frac{f_{01}(x,y)}{f_{11}(x,y)} \right) \right) \tag{48}$$

The shadow costs λ_{cf} and μ_{cf} remain to be determined. As shown in Equation 16 in the main model, λ can be solved for by combining the FOC' with the binding participation and incentive compatibility constraints:

$$\lambda_{cf} = \frac{1}{\rho} e^{\rho \underline{w}} \tag{49}$$

Solving for $w_{cf}(x,y)$ is then reduced to finding μ_{cf} that satisfies both the binding incentive compatibility constraint and the binding participation constraint:

$$\int_x \int_y e^{-\rho w_{cf}(x,y)} f_{11}(x,y) dy dx = \int_x \int_y e^{-\rho w_{cf}(x,y)} f_{10}(x,y) dy dx \tag{IC'}$$

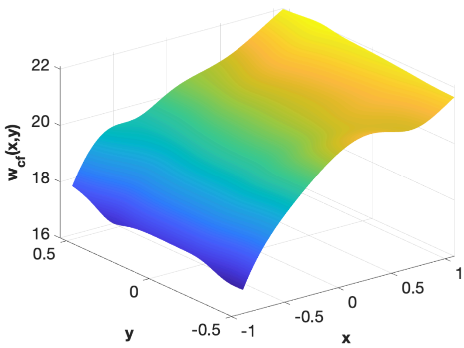
$$C \int_x \int_y e^{-\rho w_{cf}(x,y)} f_{11}(x,y) dy dx = e^{-\rho \underline{w}} \tag{P'}$$

I tabulate the results as follows:

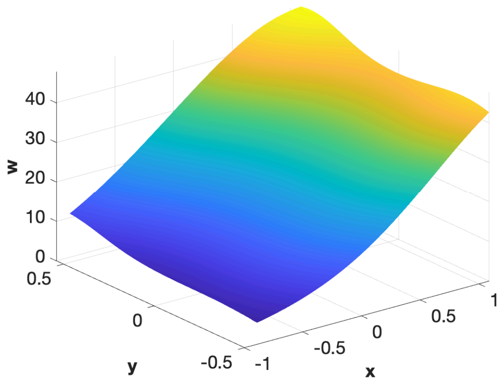
Cost of moral hazard	Estimate
ΔV : Total cost of moral hazard (\$mil.)	2.0537
ΔV_G : Cost of green moral hazard (\$mil.)	1.7248
ΔV_F : Cost of financial moral hazard (\$mil.)	0.3289

Table 5: Estimated cost of moral hazard

Out of the total cost of moral hazard of $\$2.05$ million, I find that the green moral hazard explains around 84%, of $\$1.72$ million. The result suggests that stock return better resolves uncertainty regarding the manager's financial effort than does carbon emission regarding the manager's ESG project decision.



(a) $w_{cf}(x,y)$



(b) $\hat{w}(x,y)$

Figure 11: **Nonparametric estimation of density** $\hat{f}_{11}(x,y)$ **(Panel a) and wage** $\hat{w}(x,y)$ **(Panel b)** x denotes stock performance, y denotes ESG signal,and a denotes agent's effort

7 Cross-sectional and Robustness Analyses

7.1 Cross-sectional Analyses

In this section, I run the estimation on subsamples divided by firm characteristics, to shed light on the mechanism by understanding how the estimates differ across firms of different types.

Parameter	Large Firms	Small Firms
$\underline{w}$: Value of outside option ($mil.)	28.456	11.144
c : Effort cost ($mil.)	1.058	0.223
ν_1^x : Financial effect of financial effort a_1	0.049	0.028
ν_2^x : Financial effect of ESG project a_2	-0.006	-0.015
ν_2^y : ESG effect of ESG project a_2	0.026	0.007

Table 6: Parameter estimates across firm size

Table 6 presents the estimates for subsample with large firms and that with small firms, respectively. Consistent with large firms paying higher and more volatile wages, both the value of the outside option and the effort cost are higher for larger firms. The effect of financial effort is also higher for large firms, aligned with the finding that its cost is higher for larger firms as well. In contrast, the result suggests that financial sacrifices to improve ESG performances are costlier for smaller firms. There are two potential explanations. First, smaller firms may have less technology to reduce carbon emissions, resulting in lower efficiency. This potential explanation is also consistent with ESG projects being more effective for larger firms. Second, the capital market may be less forgiving for smaller firms making financial sacrifices, as they are more likely to be capital-constrained.

7.2 Robustness Test

Stock returns can reflect the long-term value of ESG investment. However, stock returns may also reflect the market's preference for ESG which does not necessarily translate to financial value. In order to address this concern, I use ROE, which is a measure of financial performance unaffected by the belief or preference of the capital market participants, instead of stock return.

Again, I find consistent results: ESG investment forgoes ROE of 2.60% and reduces carbon emissions by 1.18%. These results reinforce my previous finding that firms do not appear to enjoy a rise in stock returns *nor ROE* when they improve ESG performance.

Parameter	Estimate
$\underline{w}$: Value of outside option ($mil.)	21.6212
c : Effort cost ($mil.)	0.61634
ν_1^x : Financial effect of financial effort a_1	0.026
ν_2^x : Financial effect of ESG investment a_2	-0.011847
ν_2^y : ESG effect of ESG investment a_2	0.012855

Table 7: Estimated parameters with ROE as financial metric

8 Conclusion

In this paper, I examine the extent to which CEOs are incentivized through compensation contracts to improve firms' ESG performance, the cost of implementing such investment in ESG by CEOs, and the contribution of ESG signals to contracting efficiency. I first construct a two-signal pure moral hazard model a la Holmström (1979), and allow the agent to separately exert financial effort that only improves financial outcomes and invest in a project that has both financial and non-financial implications. I estimate the model to uncover counterfactual outcome distributions under only financial effort or project acceptance, as well as the cost of incentivizing CEOs to make ESG investments on top of exerting financial effort.

I first find that firms are sacrificing tangible amounts of firm value to improve ESG outcomes. To the extent that the stock market efficiently prices ESG investments, this suggests that firms do care about ESG performance beyond profit maximization. Consistent with the steep sacrifice, I find that a significant portion of executive compensation can be explained by moral hazard associated with ESG projects.

This paper opens a number of promising avenues of research. First avenue of research would be to study the role of accounting information in ESG contracts. Given that the stock price may reflect not only the economic value of ESG investments but also the preference of investors for improvements in ESG performance, accounting signals could be helpful in disentangling the economic value from the preference reflected in prices. Second avenue would be to study the joint problem of trading and contracting in the context of ESG, as trading costs incurred to acquire sufficient shares to influence the contract would be another important cost of incentivizing firms to improve their ESG performances. Third would be examining various frictions, such as CEO's personal preference and misaligned objectives among investors, that prevents the principal from setting up a contract that optimally implements the desired ESG investment. Fourth would be to study how ESG pay affects CEOs' actions that would affect ESG performance of other firms and how ESG incentives interact across firms.

References

Acharya, Viral V., Richard Berner, Robert F. Engle, Hyeyoon Jung, Johannes Stroebel, Xuran Zeng, and Yihao Zhao, “Climate Stress Testing,” *Annual Review of Financial Economics*, 2023, *15* (1), 291–326.

Armstrong, Christopher S., David F. Larcker, and Che-Lin Su, “Endogenous Selection and Moral Hazard in Compensation Contracts,” *Operations Research*, 2010, *58* (4), 1090–1106.

Baker, Malcolm, Daniel Bergstresser, George Serafeim, and Jeffrey Wurgler, “The Pricing and Ownership of US Green Bonds,” *Annual Review of Financial Economics*, 2022, *14* (Volume 14, 2022), 415–437.

Bebchuk, Lucian and Roberto Tallarita, “The Perils and Questionable Promise of ESG-Based Compensation,” *Journal of Corporation Law*, 2022.

Bertomeu, Jeremy, Jaehoon Jung, and Iván Marinovic, “Moral Hazard and the Value of Information: A Structural Approach,” *Working Paper*, 2023.

_ , Ying Liang, and Iván Marinovic, “A Primer on Structural Estimation in Accounting Research,” *Foundations and Trends® in Accounting*, 2023, *18* (1–2), 1–137.

Bolton, Patrick and Marcin Kacperczyk, “Do investors care about carbon risk?,” *Journal of Financial Economics*, 2021, *142* (2), 517–549.

Bonham, Jonathan and Amoray Riggs-Cragun, “Motivating ESG Activities Through Contracts, Taxes and Disclosure Regulation,” *Chicago Booth Research Paper No. 22-05*, 2022.

_ and _ , “Contracting on Information about Value,” *Chicago Booth Research Paper No. 22-03*, 2023.

Brenner, Steffen, “The Risk Preferences of U.S. Executives,” *Management Science*, 2015, *61* (6), 1344–1361.

Carter, Mary Ellen, Andrea Pawliczek, and Rong (Irene) Zhong, “Say on ESG: The Adoption of Say-on-Pay Laws, ESG Contracting, and Firm ESG Performance,” *European Corporate Governance Institute – Finance Working Paper No. 886*, 2023.

Chernozhukov, Victor, Han Hong, and Elie Tamer, “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 2007, *75* (5), 1243–1284.

Cohen, Shira, Igor Kadach, Gaizka Ormazabal, and Stefan Reichelstein, “Executive Compensation Tied to ESG Performance: International Evidence,” *Journal of Accounting Research*, 2023, *61* (3), 805–853.

Core, John E., Wayne R. Guay, and Robert E. Verrecchia, “Price versus Non-Price Performance Measures in Optimal CEO Compensation Contracts,” *The Accounting Review*, 2003, *78* (4), 957–981.

Feltham, Gerald A. and Jim Xie, “Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations,” *The Accounting Review*, 1994, *69* (3), 429–453.

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

__ , **Bryan Hong, and Dylan Minor**, "Corporate governance and the rise of integrating corporate social responsibility criteria in executive compensation: Effectiveness and implications for firm outcomes," *Strategic Management Journal*, 2019, *40* (7), 1097–1122.

Gayle, George-Levi and Robert A. Miller, "Has Moral Hazard Become a More Important Factor in Managerial Compensation?," *American Economic Review*, December 2009, *99* (5), 1740–69.

__ **and** __ , "Identifying and Testing Models of Managerial Compensation," *The Review of Economic Studies*, 2015, *82* (3 (292)), 1074–1118.

__ , **Chen Li, and Robert A. Miller**, "Was Sarbanes–Oxley Costly? Evidence from Optimal Contracting on CEO Compensation," *Journal of Accounting Research*, 2022, *60* (4), 1189–1234.

Holmstrom, Bengt and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, Organization*, 1991, *7*, 24–52.

Holmström, Bengt, "Moral Hazard and Observability," *The Bell Journal of Economics*, 1979, *10* (1), 74–91.

Jung, Hyeyoon, Robert F. Engle, and Richard Berner, "CRISK: Measuring the Climate Risk Exposure of the Financial System," *FRB of New York Staff Report*, 2021, (No. 977).

__ , __ , **Shan Ge, and Xuran Zeng**, "Measuring the Climate Risk Exposure of Insurers," *FRB of New York Staff Report*, 2023, (No. 1066).

Larcker, David F. and Edward M. Watts, "Where's the greenium?," *Journal of Accounting and Economics*, 2020, *69* (2), 101312.

Margiotta, Mary M. and Robert A. Miller, "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review*, 2000, *41* (3), 669–719.

Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang, "Firm-Level Climate Change Exposure," *The Journal of Finance*, 2023.

Tang, Dragon Yongjun and Yupu Zhang, "Do shareholders benefit from green bonds?," *Journal of Corporate Finance*, 2020, *61*, 101427. Environmental, Social, and Governance Issues: Emerging Markets and Beyond.

Zerbib, Olivier David, "A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from Environmental Integration and Sin Stock Exclusion*," *Review of Finance*, 07 2022, *26* (6), 1345–1388.
■Ľuboš Pástor et al.

Ľuboš Pástor, Robert F. Stambaugh, and Lucian A. Taylor, "Sustainable investing in equilibrium," *Journal of Financial Economics*, 2021, *142* (2), 550–571.
■Ľuboš Pástor et al.

__ , __ , **and** __ , "Dissecting green returns," *Journal of Financial Economics*, 2022, *146* (2), 403–424.

Appendix

A Proofs

A.1 Proofs for General Model

Now, I generalize the stylized framework above by (1) relaxing the distributional assumptions on the signals and (2) removing the focus on linear contracts. Specifically, as in Holmström (1979), I allow for arbitrary outcome distributions of $f_a(x, y)$ and arbitrary functional form of wage $w(x, y)$. For the purpose of identification, however, I maintain the assumption on how effort transforms the outcome distribution.

Assume that it is optimal for the principal to induce both financial effort and project acceptance: $a^* = (1, 1)$. Then, the principal's problem becomes:

$$\max_{w(\cdot)} \mathbb{E}[V(x, y) - w(x, y)|a = (1, 1)]. \tag{50}$$

s.t.

$$\mathbb{E}[u(w(x, y), 1)|a = (1, 1)] \geq \mathbb{E}[u(w(x, y), 1)|a = (1, 0)] \tag{IC10}$$

$$\mathbb{E}[u(w(x, y), 1)|a = (1, 1)] \geq \mathbb{E}[u(w(x, y), 0)|a = (0, 1)] \tag{IC01}$$

$$\mathbb{E}[u(w(x, y), 1)|a = (1, 1)] \geq \mathbb{E}[u(w(x, y), 0)|a = (0, 0)] \tag{IC00}$$

$$\mathbb{E}[u(w(x, y), 1)|a = (1, 1)] \geq u(\underline{w}, (0, 0)) \tag{P}$$

Given the optimal effort, the principal's problem can be further simplified as a wage minimization problem:

$$\max_{w(\cdot)} \mathbb{E}[-w(x, y)|a = (1, 1)]. \tag{51}$$

subject to the incentive compatibility constraints and the participation constraint above.

Assume further that under the optimal compensation scheme, $a = (1, 0)$ and $a = (0, 1)$ are the best alternatives. Then, only (IC10) and (IC01) will bind and (IC00) will be a strict inequality:

$$\mathbb{E}[u(w(x, y), 1)|a = (1, 1)] = \mathbb{E}[u(w(x, y), 1)|a = (1, 0)] \tag{IC10}$$

$$\mathbb{E}[u(w(x, y), 1)|a = (1, 1)] = \mathbb{E}[u(w(x, y), 0)|a = (0, 1)] \tag{IC01}$$

$$\mathbb{E}[u(w(x, y), 1)|a = (1, 1)] > \mathbb{E}[u(w(x, y), 0)|a = (0, 0)] \tag{IC00}$$

The first order condition then gives:

$$\begin{aligned}1 &= \lambda\rho Ce^{-\rho w(x,y)} \\ &+ \mu_{10}\rho\left(Ce^{-\rho w(x,y)} - Ce^{-\rho w(x,y)}\frac{f_{10}(x,y)}{f_{11}(x,y)}\right) \\ &+ \mu_{01}\rho\left(Ce^{-\rho w(x,y)} - e^{-\rho w(x,y)}\frac{f_{01}(x,y)}{f_{11}(x,y)}\right)\end{aligned} \tag{52}$$

The first order condition above provides the relation among the outcome distributions, one under the optimal effort and others under the alternative levels of effort:

$$\mu_{10}C\frac{f_{10}(x,y)}{f_{11}(x,y)} + \mu_{01}\frac{f_{01}(x,y)}{f_{11}(x,y)} = C(\lambda + \mu_{10} + \mu_{01}) - \frac{1}{\rho}e^{\rho w(x,y)} \tag{FOC}$$

Binding constraints provide:

$$C\int_x\int_y e^{-\rho w(x,y)}f_{11}(x,y)dydx = C\int_x\int_y e^{-\rho w(x,y)}f_{10}(x,y)dydx \tag{IC10}$$

$$C\int_x\int_y e^{-\rho w(x,y)}f_{11}(x,y)dydx = \int_x\int_y e^{-\rho w(x,y)}f_{01}(x,y)dydx \tag{IC01}$$

$$C\int_x\int_y e^{-\rho w(x,y)}f_{11}(x,y)dydx = e^{-\rho \underline{w}} \tag{P}$$

Moreover, as $f_{10}(x,y)$ and $f_{01}(x,y)$ are probability distribution functions, they should integrate to 1:

$$\int_x\int_y f_{10}(x,y)dydx = 1 \tag{53}$$

$$\int_x\int_y f_{01}(x,y)dydx = 1 \tag{54}$$

Finally, the prescribed effort choice should indeed be optimal for the principal:

$$\mathbb{E}[V(x,y) - w(x,y)|a = (1,1)] \geq \mathbb{E}[V(x,y) - \underline{w}|a = (0,0)] \tag{55}$$

$$\mathbb{E}[V(x,y) - w(x,y)|a = (1,1)] \geq \mathbb{E}[V(x,y) - w_{10}(x,y)|a = (1,0)] \tag{56}$$

$$\mathbb{E}[V(x,y) - w(x,y)|a = (1,1)] \geq \mathbb{E}[V(x,y) - w_{01}(x,y)|a = (0,1)] \tag{57}$$

Where $w_{10}(x,y)$ and $w_{01}(x,y)$ denotes the contracts that optimally induces the alternative effort of $a = (1,0)$ and $a = (0,1)$, respectively. If the incentive compatibility constraint between $a = (1,0)$ and $a = (1,1)$ is binding under alternative contract $w_{10}(x,y)$ and that between $a = (0,1)$ and $a = (1,1)$ is binding under alternative contract $w_{01}(x,y)$, it is only marginally different from the optimal contract. Then, Equations 16 and 17 can be rewritten as:

$$\mathbb{E}[V(x,y) - w(x,y)|a = (1,1)] \geq \mathbb{E}[V(x,y) - w(x,y)|a = (1,0)] \tag{58}$$

$$\mathbb{E}[V(x,y) - w(x,y)|a = (1,1)] \geq \mathbb{E}[V(x,y) - w(x,y)|a = (0,1)] \tag{59}$$

A.1.1 Optimal Contract

From the first order condition, the optimal wage is given as follows:

$$w(x,y) = \frac{1}{\rho}\log\left(\rho C(\lambda + \mu_{10} + \mu_{01}) - \rho C\mu_{10}\frac{f_{10}(x,y)}{f_{11}(x,y)} - \rho\mu_{01}\frac{f_{01}(x,y)}{f_{11}(x,y)}\right) \tag{60}$$

An immediate observation from the equation above is that the more likely an outcome (x, y) is under actions other than the one prescribed by the contract, the lower the wage. Therefore, the highest possible wage $\overline{w}$ is rewarded to (x, y) that perfectly signals $a = (1, 1)$:

$$w(x,y) \leq \overline{w} = \frac{1}{\rho}\log\left(\rho C(\lambda + \mu_{10} + \mu_{01})\right) \tag{61}$$

It can also be seen that, given the base parameters ρ and C, the wage function is determined by shadow costs λ, μ_{10}, and μ_{01}.

λ can be readily solved for by combining the first order condition with the binding participation constraint and the incentive compatibility constraints:

$$\lambda = \frac{1}{\rho}e^{\rho\underline{w}} \tag{62}$$

The equation above is consistent with the intuition that the higher value of outside options to the agent makes it costlier to induce the agent to participate in the contract.

On the other hand, it is difficult to obtain analytical expressions for μ_{10} and μ_{01} without making additional assumptions regarding the likelihood ratios across actions. Therefore, for the analysis of the optimal contract to follow, I numerically solve for μ_{10} and μ_{01} that jointly satisfy the binding participation constraint and the incentive compatibility constraints.

In order to verify the optimality of the contract, I examine the second-order condition. Given that $\rho > 0$, $f_{11}(x, y) > 0$ for all (x, y) within support, and $e^{-\rho w(x,y)} > 0$ for any real $w(x, y)$, the second-order condition can be written as:

$$\rho C(\lambda + \mu_{10} + \mu_{01}) - \rho C\mu_{10}\frac{f_{10}(x,y)}{f_{11}(x,y)} - \rho\mu_{01}\frac{f_{01}(x,y)}{f_{11}(x,y)} > 0 \tag{SOC}$$

As any wage $w(x, y)$ that violates the SOC is complex, any wage $w(x, y)$ that is real for every (x, y) should satisfy the SOC.

A.2 Proofs for Identification

For the estimation to be feasible, I make one additional assumption.

I assume that a high enough outcome in each dimension must be due to high effort in each dimension:

$$\lim_{y\to\infty}\frac{f_{10}(x,y)}{f_{11}(x,y)}=0 \tag{63}$$

$$\lim_{x\to\infty}\frac{f_{01}(x,y)}{f_{11}(x,y)}=0 \tag{64}$$

This means that extremely favorable outcome in financial performance x and non-financial performance y perfectly signals financial effort a_1 and ESG project selection a_2, respectively. The assumption allows me to use wages for extremely favorable outcomes to infer the benchmark when moral hazard in each dimension is not present.

From the binding participation constraint, I get the first moment condition:

$$\int_x\int_y e^{-\rho w(x,y)}f_{11}(x,y)dydx=\frac{1}{C}e^{-\rho\underline{w}} \tag{65}$$

As the moment condition follows directly from the participation constraint, the immediate intuition is that the principal should reward the agent in utility for the effort cost and the outside option available to the agent.

By integrating both sides of the first order condition, I get the second moment condition:

$$\int_x\int_y e^{\rho w(x,y)}f_{11}(x,y)dydx=\rho(\lambda C+\mu_{01}(C-1)) \tag{66}$$

The intuition here is that the level of wage is determined by three factors: risk aversion, cost of participation, and cost of incentivizing costly effort. Note that, as the investment decision is personally costless, its incentive does not affect the overall level of the compensation. Instead, the incentives for investment should come from the relative distribution of the wage.

Combining the first order condition with binding incentive compatibility constraints yields the third moment condition:

$$\int_x\int_y e^{-\rho w(x,y)}f_{11}(x,y)dydx=\frac{1}{\rho\lambda C} \tag{67}$$

Given the first moment condition, this moment condition is in fact equivalent to Equation 16, the intuition of which is that inducing participation grows costly in the value of outside option.

The assumption that an extremely favorable outcome perfectly signals high effort, along with the first order condition, provides the fourth moment condition:

$$\frac{1}{\rho}e^{\rho\overline{w}}=(\lambda+\mu_{10}+\mu_{01})C \tag{68}$$

Given the analysis of the theoretical upper bound on the wage in Equation 61, this moment condition is simply stating the implicit assumption that the highest observed wage approximates the theoretical upper bound.

The assumption provides additional information on the relation between outcome distributions for extreme outcomes in each dimension:

$$(\lambda+\mu_{10}+\mu_{01})C = \lim_{y\to\infty}\left(e^{\rho w(x,y)}+\rho\mu_{01}\frac{f_{01}(x,y)}{f_{11}(x,y)}\right) \tag{69}$$

$$= \lim_{x\to\infty}\left(e^{\rho w(x,y)}+\rho\mu_{10}C\frac{f_{10}(x,y)}{f_{11}(x,y)}\right) \tag{70}$$

By combining the IC01 with the assumption that financial effort has no ESG implication, I get the fifth moment condition:

$$\int_x\int_y e^{-\rho w(x,y)} f_{11}(x,y)dydx = \frac{1}{C}\int_x\int_y e^{-\rho w(x,y)} f_{11}(x+\nu_1^x,y)dydx \tag{71}$$

As this moment condition follows directly from the incentive compatibility condition, the intuition is simply that the improvement in financial performance due to financial effort and thus the increase in wage should compensate for the agent's effort cost.

By combining the FOC with the assumption that financial effort has no ESG implication, I get the following expression for the counterfactual distribution under only financial effort:

$$f_{10}(x,y) = \frac{1}{\mu_{10}C}\left(C(\lambda+\mu_{10}+\mu_{01}) - \frac{1}{\rho}e^{\rho w(x,y)} - \mu_{01}\frac{f_{11}(x+\nu_1^x,y)}{f_{11}(x,y)}\right) f_{11}(x,y) \tag{72}$$

The assumption that financial effort has no ESG assumption, along with the assumption that extremely favorable outcome in each dimension perfectly signals effort in each dimension, provides:

$$(\lambda+\mu_{10}+\mu_{01})C = \lim_{y\to\infty}\left(e^{\rho w(x,y)}+\rho\mu_{01}\frac{f_{11}(x+\nu,y)}{f_{11}(x,y)}\right) \tag{73}$$

Let $\overline{w}(x) = \lim_{y\to\infty} w(x,y)$ and $\overline{f}_{11} = \lim_{y\to\infty} f_{11}(x,y)$ denote wage and probability density under both efforts as functions of financial performance x for asymptotically high level of ESG performance y. Then, the equation above provides the final set of moment conditions:

$$\frac{1}{\rho}e^{\rho\overline{w}(x)} = (\lambda+\mu_{10}+\mu_{01})C - \mu_{01}\frac{\overline{f}_{11}(x+\nu)}{\overline{f}_{11}(x)} \tag{74}$$

As the equation above provides a continuum of moment conditions, I collapse them by integrating w.r.t. x, in order to avoid overidentification:

$$\frac{1}{\rho}\mathbb{E}[e^{\rho w(x,y)}|y=\infty] = C(\lambda+\mu_{10}+\mu_{01}) - \mu_{01} \tag{75}$$

Comparing with the fourth moment condition in Equation 68, intuition here is that the difference between the highest wage and the expected wage under extremely favorable non-financial outcome can be explained by the cost of inducing financial effort.

Finally, I get a set of moment conditions for the effects of ESG investment a_2, ν_2^x and ν_2^y, by multiplying x and y, respectively, and then integrating both sides of FOC:

$$\mu_{10}C(\mathbb{E}[x]-\nu_x^2)+\mu_{01}(\mathbb{E}[x]-\nu_x^1)=C(\lambda+\mu_{10}+\mu_{01})\mathbb{E}[x]-\frac{1}{\rho}\mathbb{E}[xe^{\rho w(x,y)}] \tag{76}$$

$$\mu_{10}C(\mathbb{E}[y]-\nu_y^2)+\mu_{01}(\mathbb{E}[y])=C(\lambda+\mu_{10}+\mu_{01})\mathbb{E}[y]-\frac{1}{\rho}\mathbb{E}[ye^{\rho w(x,y)}] \tag{77}$$

The equations above shows that covariance between level of wage and each performance metrics reveals the extent to which actions shift the mean of each performance metric.

Therefore, I begin by estimating $(C, \underline{w}, \lambda, \mu_{10}, \mu_{01})$ from the following five moment conditions.

$$\begin{bmatrix} \frac{1}{C}e^{-\hat{\rho}\underline{w}} \\ \hat{\rho}(\lambda C+\mu_{01}(C-1)) \\ \frac{1}{\hat{\rho}\lambda C} \\ (\lambda+\mu_{10}+\mu_{01})C \\ (\lambda+\mu_{10}+\mu_{01})C-\mu_{01} \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \alpha \\ \gamma \\ \delta \end{bmatrix}, \tag{78}$$

where

$$\alpha=\int_x\int_y e^{-\hat{\rho}w(x,y)}f_{11}(x,y)dydx \tag{79}$$

$$\beta=\int_x\int_y e^{\hat{\rho}w(x,y)}f_{11}(x,y)dydx \tag{80}$$

$$\gamma=\frac{1}{\hat{\rho}}e^{\hat{\rho}\bar{w}} \tag{81}$$

$$\delta=\frac{1}{\hat{\rho}}\mathbb{E}[e^{\rho w(x,y)}|y=\infty] \tag{82}$$

The first moment α is the agent's expected utility (reversed sign) given wage $w(x,y)$ and outcome distribution $f_{11}(x,y)$. The second moment β captures the expected level of the wage to the agent. The third moment γ effectively represents the theoretical upper bound of the wage. The fourth moment δ captures the expected level of wage under extremely high ESG performance.

From the fourth and the fifth moment condition in Equation 40, I immediately get an expression for μ_{01}:

$$\mu_{01}=\gamma-\delta \tag{83}$$

Substituting the above into the combination of the second and the third moment conditions, I find an expression for $C=e^{\hat{\rho}c}$:

$$C=\frac{\alpha\beta-1+\hat{\rho}\alpha(\gamma-\delta)}{\hat{\rho}\alpha(\gamma-\delta)} \tag{84}$$

Therefore, $c=c_{11}=c_{10}$ can be expressed as:

$$c=\frac{1}{\hat{\rho}}\log\left(\frac{\alpha\beta-1+\hat{\rho}\alpha(\gamma-\delta)}{\hat{\rho}\alpha(\gamma-\delta)}\right) \tag{85}$$

By substituting the above expression for C into the first moment condition, I get the following for $\underline{w}$:

$$\underline{w} = -\frac{1}{\hat{\rho}} \log \left(\frac{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma - \delta)}{\hat{\rho}(\gamma - \delta)} \right) \tag{86}$$

Substituting the above expression for C into the combination of the second and the fifth moment conditions provides an expression for μ_{10}:

$$\mu_{10} = \frac{\alpha(\gamma - \delta)(\hat{\rho}\delta - \beta)}{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma - \delta)} \tag{87}$$

By substituting the above expression for C into the third moment condition, I get the following for λ:

$$\lambda = \frac{(\gamma - \delta)}{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma - \delta)} \tag{88}$$

Then, I estimate the shift parameters $(\nu_1^x, \nu_2^x, \nu_2^y)$ from the remaining three moment conditions:

$$\begin{bmatrix} \frac{1}{C} \int_x \int_y e^{-\hat{\rho} w(x,y)} f_{11}(x + \nu_1^x, y) dy dx \\ \mu_{10} C \nu_2^x + \mu_{01} \nu_1^x \\ \mu_{10} C \nu_2^y \end{bmatrix} = \begin{bmatrix} \alpha \\ \frac{1}{\rho}\alpha_x - (C\lambda + (C-1)\mu_{01})\mu_x \\ \frac{1}{\rho}\alpha_y - (C\lambda + (C-1)\mu_{01})\mu_y \end{bmatrix}, \tag{89}$$

where

$$\eta_x = \int_x \int_y x e^{\hat{\rho} w(x,y)} f_{11}(x, y) dy dx \tag{90}$$

$$\eta_y = \int_x \int_y y e^{\hat{\rho} w(x,y)} f_{11}(x, y) dy dx \tag{91}$$

$$m_x = \int_x \int_y x f_{11}(x, y) dy dx \tag{92}$$

$$m_y = \int_x \int_y y f_{11}(x, y) dy dx \tag{93}$$

By substituting the above expression for C into the sixth condition, I get the following condition for ν_1^x:

$$\int_x \int_y e^{-\hat{\rho} w(x,y)} f_{11}(x + \nu_1^x, y) dy dx = \frac{\alpha\beta - 1 + \hat{\rho}\alpha(\gamma - \delta)}{\hat{\rho}(\gamma - \delta)} \tag{94}$$

With ν_1^x pinned down, I can solve for ν_2^x and ν_2^y:

$$\nu_2^x = \frac{1}{\mu_{10} C} \left(\frac{1}{\rho} \eta_x - (C\lambda + (C-1)\mu_{01}) m_x - \mu_{01} \nu_1^x \right) \tag{95}$$

$$\nu_2^y = \frac{1}{\mu_{10} C} \left(\frac{1}{\rho} \eta_y - (C\lambda + (C-1)\mu_{01}) m_y \right) \tag{96}$$

B ESG Contract Data

In this section, I describe the characteristics of compensation contracts from the ECA in detail.

B.1 ESG Compensation Scheme

B.1.1 Disclosed Metrics

Here are examples of commonly used ESG-related metrics in compensation contracts.

- E Examples: GHG Emission (scope 1 and scope 2, intensity, percentage reduction), Waste Management (percentage reduction, percentage recycled), Water Consumption (intensity, freshwater withdrawal), Environmental Spills and Contamination (# of class 4+ spills or level 3+ environmental incidents), Share of Electricity from Renewable Sources (%)
- S Examples: Employee Health and Safety (OSHA-recordable injuries, lost workdays away, severe injury and fatality rate), Diversity Equity Inclusion (Veteran representation, Women in senior management, ESG Index), Customer satisfaction, COVID 19 Response, Corporate Social Responsibility (CSR Index)

B.1.2 Compensation Structure

Examples: Multiple Targets (Reduction of GHG emission by 6% 8% 10%, GHG intensity reduction by 16% 18% 20%, Projects in bio-fuel 1 2 3), Long-term Target (80% reduction in carbon emissions by 2030), Relative Target (Within 5% of industry leader in terms of Dow Jones Sustainability Index), Qualitative Target ("operate sustainability by delivering world-class end-to-end performance in safety resource efficiency and environmental protection")

B.2 Characteristics of Firms with vs without ESG Compensation Contracts

	FullSample		ESG		Non-ESG		Difference	
	mean	sd	mean	sd	mean	sd	b	t
Log Size	8.621	1.43	8.743	1.55	8.582	1.39	-0.16***	(-3.35)
Abnormal Return	0.013	0.26	0.004	0.27	0.015	0.25	0.01	(1.31)
Log Emission	11.828	2.32	12.233	2.65	11.697	2.19	-0.54***	(-6.66)
Emission Reduction	0.026	0.10	0.030	0.10	0.024	0.09	-0.01	(-1.60)
Total Pay	21.230	28.87	21.452	29.35	21.158	28.71	-0.29	(-0.32)
Observations	5403		1319		4084		5403	

Table A.1: Summary Statistics of ESG vs non-ESG

C Intuitions from a Stylized Framework

Before presenting the full model, I show a simplified version under the framework of linear compensation, exponential utility, and normally distributed performance measures, in the spirit of Holmstrom and Milgrom (1991) and Feltham and Xie (1994), to provide intuition for the generalized model used for the estimation.

Information Structure In this stylized LEN framework, I assume that the errors ϵ_x and ϵ_y in signals x and y, follow a joint normal distribution. The signal structure can therefore be expressed as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = a_1 \begin{bmatrix} \nu_1^x \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} \nu_2^x \\ \nu_2^y \end{bmatrix} + \begin{bmatrix} \epsilon_x \\ \epsilon_y \end{bmatrix} \tag{97}$$

where components (ϵ_x, ϵ_y) are mean-zero errors that are jointly normally distributed with a correlation of r:

$$\begin{bmatrix} \epsilon_x \\ \epsilon_y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \right) \tag{98}$$

Agent's Certainty Equivalent Here, I focus on linear contracts $w(x, y)$ given outcome (x, y):

$$w(x, y) = \alpha + \beta_x x + \beta_y y \tag{99}$$

where β_x and β_y are incentive coefficients for performances x and y, respectively. Note that coefficients $(\alpha, \beta_x, \beta_y)$ sufficiently summarize the contract. Owing to the LEN setup, the agent's certainty equivalent $CE(a)$ for action $a = (a_1, a_2)$ given a linear contract $(\alpha, \beta_x, \beta_y)$ can be simplified as follows:

$$CE(a) = E[w|a] - \frac{1}{2}\rho Var(w|a) - a_1 c \tag{100}$$

$$= a_1 \underbrace{(\beta_x \nu_1^x - c)}_{\textbf{welfare impact of } a_1} + a_2 \underbrace{(\beta_x \nu_2^x + \beta_y \nu_2^y)}_{\textbf{welfare impact of } a_2} + \underbrace{\alpha - \frac{1}{2}\rho(\beta_x^2\sigma_x^2 + \beta_y^2\sigma_y^2 + 2\beta_x\beta_y r\sigma_x\sigma_y)}_{\text{constant w.r.t. action}} \tag{101}$$

From the expression above, incentive compatibility conditions for actions a_1 and a_2 are immediately clear. To induce financial effort ($a_1 = 1$), the incentive β_x for financial outcome x should at least compensate for the cost of effort:

$$\beta_x \geq \frac{c}{\nu_1^x} > 0 \tag{IC1}$$

Contract inducing ESG Investment ("Green Contract") To induce ESG investment (a_2=1), the incentive β_y for non-financial outcome y should at least counteract the disincentive caused by the financial incentive β_x:

$$\beta_y \geq \beta_x \cdot -\frac{\nu_2^x}{\nu_2^y} \tag{IC2}$$

The agent has an outside option offering $\underline{w}$ with certainty. Therefore, to ensure that the agent prefers to participate in the contract, certainty equivalent from wage should at least match the outside option:

$$E[w|a] \geq \underline{w} + a_1 c + \frac{1}{2}\rho Var(w|a) \tag{P}$$

Intuitively, the principal should reward the agent for participation, exerting effort, and taking risks. The constant portion of the wage α is thus determined so that the expected wage is sufficient:

$$\alpha = \underline{w} + \frac{1}{2}\rho\left(\beta_x^2\sigma_x^2 + \beta_y^2\sigma_y^2 + 2\beta_x\beta_y r\sigma_x\sigma_y\right) \tag{102}$$

Based on the constraints above, the optimal contract depends on the action that the principal seeks to implement through the contract. Suppose the principal seeks to implement both financial effort and ESG investment (i.e. $a = (1, 1)$). Then, the principal's problem is reduced to minimizing expected wage subject to the incentive compatibility constraints IC1 and IC2, and the participation constraint P above:

$$\max_{\alpha,\beta_x,\beta_y} -(\alpha + \beta_x(\nu_1^x + \nu_2^x) + \beta_y\nu_2^y) \tag{103}$$

Binding incentive compatibility IC1 for financial effort a_1 gives incentive β_x on financial outcome x:

$$\beta_x = \frac{c}{\nu_1^x} \tag{104}$$

If incentive compatibility IC2 for ESG investment a_2 binds, incentive β_y on non-financial outcome y is given as:

$$\beta_y = -\frac{\nu_2^x}{\nu_2^y}\beta_x \tag{105}$$

However, if IC2 does not bind, β_y should be determined from first-order conditions. The lagrangian of the problem is then given as follows:

$$\begin{aligned}\mathcal{L} = &-(\alpha + \beta_x\nu_2^x + \beta_y\nu_2^y)\\ &+ \mu_1(\beta_x\nu_1^x - c)\\ &+ \lambda\left(\beta_x\nu_2^x + \beta_y\nu_2^y - \frac{1}{2}\rho(\beta_x^2\sigma_x^2 + \beta_y^2\sigma_y^2 + 2\beta_x\beta_y r\sigma_x\sigma_y) + \alpha - \underline{w} - c\right)\end{aligned} \tag{106}$$

Where μ_1 and λ are shadow costs of IC1 and P, respectively.

λ is given from the first-order condition w.r.t. α:

$$\frac{\partial}{\partial\alpha}\mathcal{L} = -1 + \lambda = 0 \tag{107}$$

Substituting λ above into the first-order condition w.r.t. β_y yields:

$$\frac{\partial}{\partial\beta_y}\mathcal{L} = -\rho\sigma_y^2\left(\beta_y + r\frac{\sigma_x}{\sigma_y}\beta_x\right) = 0 \tag{108}$$

Considering both cases, when IC2 binds and when it does not, β_y is given as:

$$\beta_y = \max\left(-r\frac{\sigma_x}{\sigma_y}, -\frac{\nu_2^x}{\nu_2^y}\right) \cdot \beta_x \tag{109}$$

The intuition for the result above is as follows. If the financial incentive β_x is sufficient for inducing both the financial effort a_1 and ESG investment a_2 (i.e. IC2 is not binding), the role of non-financial performance y in the contract is minimizing the risk borne by the agent. Therefore, if non-financial performance y is positively correlated with financial performance x, non-financial incentive β_y should be negative, in order to hedge the agent's exposure to financial performance x. On the contrary, if IC2 is binding, the sign of the non-financial incentive β_y depends on whether the financial impact ν_2^x of ESG investment is positive or negative. On one hand, if ESG investment boosts financial performance ($\nu_2^x > 0$), non-financial incentive β_y should still be negative to hedge the agent's exposure to financial performance x. On the other hand, if ESG investment is financially costly, non-financial incentive β_y should be positive, in order to counteract the disincentive caused by the financial incentive.

Two relevant features of the data are: (1) weight on non-financial outcome is positive ($\beta_y > 0$) and (2) financial performance and non-financial performance are positively correlated ($r > 0$).[8] Reconciling these facts with the model suggests that: (1) Incentive compatibility for ESG investment, IC2, is binding and (2) ESG investment has a negative impact on financial performance. On these grounds, I assume that incentive compatibility for ESG investment binds and exclude the case in which ESG investment boosts financial performance in the analyses to follow.

This framework also allows me to compare how the optimal contract differs by how valuable ESG performance is to the principal (k in Equation (1)). Given the assumptions above that ESG investment is costly, the principal would prefer to induce both financial effort and ESG investment if and only if k is large enough; otherwise, the principal would only induce financial effort and avoid the costly ESG investment.

Contract discouraging ESG Investment ("Brown Contract") To discourage ESG investment (a_2=0), the incentive β_y' for non-financial outcome y should never be strong enough to counteract the disincentive caused by the financial incentive β_x:

$$\beta_y' \leq \beta_x \cdot -\frac{\nu_2^x}{\nu_2^y} \tag{IC2'}$$

Considering both cases, when IC2' binds and when it does not, β_y' is given as:

$$\beta_y' = \min\left(-r\frac{\sigma_x}{\sigma_y}, -\frac{\nu_2^x}{\nu_2^y}\right) \cdot \beta_x \tag{110}$$

[8] One potential explanation for the positive correlation is that, for the same level of cash flow performance, investors may have preference for favorable non-financial performance and therefore reward it with stock returns.

Given the assumptions that financial performance x and non-financial performance y are positively correlated ($r > 0$) and that ESG investment a_2 is costly to the firm ($\nu_2^x < 0$), coefficient β_y is given as:

$$\beta_y' = -r\frac{\sigma_x}{\sigma_y}\beta_x \tag{111}$$

As incentive compatibility w.r.t. financial effort a_1 remains the same, coefficient β_x does not change.

Then, the optimal compensation $w'(x, y)$ that induces $a = (1, 0)$ is given as:

$$w'(x, y) = \alpha' + \beta_x x + \beta_y' y \tag{112}$$

The principal's value net of wage to the agent under the contract that induces ESG investment is as follows:

$$\begin{aligned} &E[V(x, y) - w(x, y)|a = (1, 1)] \\ &= \underbrace{\nu_1^x - c}_{\textbf{Net Value of } a_1} + \underbrace{k\nu_2^y + \nu_2^x}_{\textbf{Net Value of } a_2} - \underline{w} - \underbrace{\frac{1}{2}\rho\left(\frac{c}{\nu_1^x}\right)^2\left(\sigma_x^2 + \left(\frac{\nu_2^x}{\nu_2^y}\right)^2\sigma_y^2 - 2\left(\frac{\nu_2^x}{\nu_2^y}\right)r\sigma_x\sigma_y\right)}_{\textbf{Risk Premium}} \end{aligned} \tag{113}$$

The principal's value net of wage to the agent under the contract that *does not* induce ESG investment is as follows:

$$E[V(x, y) - w'(x, y)|a = (1, 0)] = \underbrace{\nu_1^x - c}_{\textbf{Net Value of } a_1} + \underbrace{\frac{1}{2}\rho\left(\frac{c}{\nu_1^x}\right)^2(1 - r^2)\sigma_x^2}_{\textbf{Risk Premium}} \tag{114}$$

Therefore, the principal chooses to induce ESG investment if and only if:

$$k \geq \frac{1}{\nu_2^y}\left(\underbrace{-\nu_2^x}_{\textbf{Direct Cost of } a_2} + \underbrace{\frac{1}{2}\rho\left(\frac{c}{\nu_1^x}\right)^2\left(r\sigma_x - \frac{\nu_2^x}{\nu_2^y}\sigma_y\right)^2}_{\textbf{Premium for risk added by } a_2}\right) \tag{115}$$

The equation above illustrates that the cost of implementing ESG investment to the principal is twofold: (1) direct financial cost of ESG investment and (2) compensation for the additional risk posed by the ESG incentive.

C.0.1 Comparative Statics

Based on the assumption that ESG project is net costly to the firm ($\nu_x^2 < 0$), I examine how the key parameters, cost of effort (c), effect of financial effort (ν_1^x), financial effect of ESG project (ν_2^x), and ESG effect of ESG project (ν_2^y) impact the cost of moral hazard in the contract that induces ESG project ("Green Contract") versus the contract that discourages ESG project ("Brown Contract").

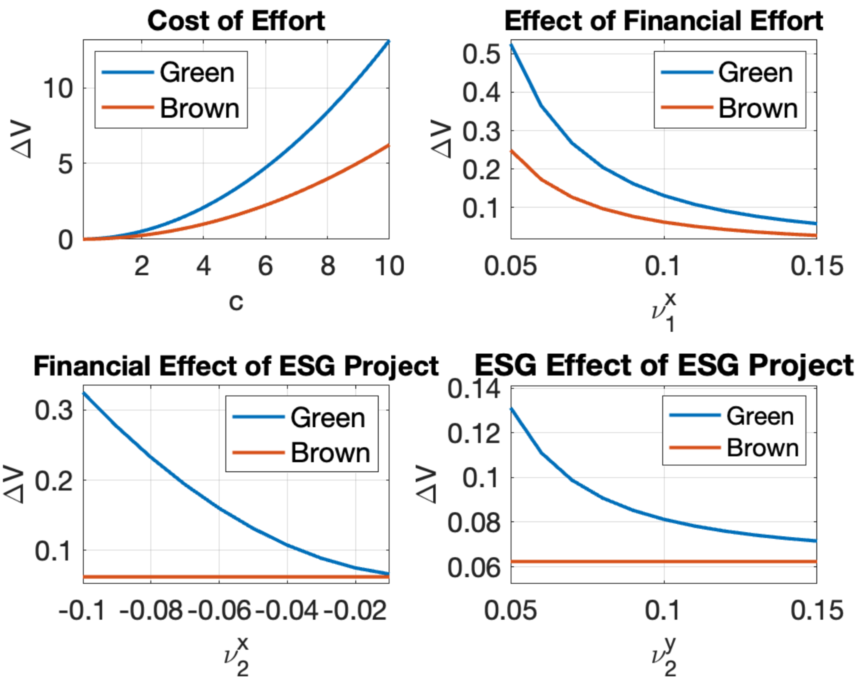


Figure A.1: Cost of Moral Hazard w.r.t. Key Parameters

The top-left panel shows that the cost of moral hazard (ΔV) increases in the cost of effort (c) for both contracts. When the cost of effort increases, the contract becomes more sensitive to financial outcome x (β_x increases in c), leaving the agent more exposed to variation in x. This dynamic is weaker for the "brown contract", in which the non-financial outcome y is used to hedge the agent's exposure to variation in x.

The top-right panel shows that the cost of moral hazard (ΔV) decreases in the effect of financial effort (ν_1^x) for both contracts. This is because ν_1^x plays the exact opposite role of c; higher ν_1^x means cheaper cost of effort for the same level of improvement in x.

The bottom-left panel shows that the cost of moral hazard decreases in the financial effect of ESG project ν_2^x (increases in the financial cost of ESG project) for the "green contract". When the financial cost of ESG project increases, the contract becomes more sensitive to non-financial outcome y (β_y increases in the magnitude of ν_2^x), leaving the agent more exposed to variation in y. In contrast, ν_2^x has no effect on the "brown contract", as it becomes irrelevant when the ESG project is not implemented.

The bottom-right panel shows that the cost of moral hazard decreases in the ESG effect of ESG project ν_2^y for the "green contract". This is because ν_2^y plays the exact opposite role of ν_2^x; higher ν_2^x means smaller financial disincentive for the same level of improvement in y.