A Theory of Socially Responsible Investment *

Martin Oehmke[†] Marcus Opp[‡]

November 4, 2020

Abstract

We characterize necessary conditions for socially responsible investors to impact firm behavior in a setting in which firm production generates social costs and is subject to financing constraints. Impact requires a *broad mandate*, in that socially responsible investors need to internalize social costs irrespective of whether they are investors in a given firm. Impact is optimally achieved by enabling a scale increase for clean production. Socially responsible and financial investors are complementary: jointly they can achieve higher surplus than either investor type alone. When socially responsible capital is scarce, it should be allocated based on a *social profitability index* (SPI). This micro-founded ESG metric captures not only a firm's social status quo but also the counterfactual social costs produced in the absence of socially responsible investors.

Keywords: Socially responsible investing, ESG, SPI, capital allocation, sustainable investment, social ratings.

JEL Classification: G31 (Capital Budgeting; Fixed Investment and Inventory Studies; Capacity), G23 (Non-bank Financial Institutions; Financial Instruments; Institutional Investors).

*For helpful comments and suggestions, we thank Ulf Axelson, Paul Beaumont, Roland Bénabou, Bruno Biais, Markus Brunnermeier, Patrick Bolton, Peter DeMarzo, Diego Garcia, Deeksha Gupta, Anastasia Kartasheva, Lukasz Pomorski, Ailsa Roëll, Joel Shapiro, Lin Shen, Paul Woolley, and seminar participants at ASU Sonoran, BIS, Bonn, Cambridge, Copenhagen Business School, CU Boulder, Deutsche Bundesbank, EFA, École Polytechnique, Erasmus University, EWFC, Frankfurt School of Finance, the FTZ Seminar, Humboldt Universität zu Berlin, IDC Financial Economics Conference, INSEAD Finance Symposium, LSE, MIT OrgEcon lunch, OxFIT, Princeton, Schroders, Stanford, Stockholm School of Economics, UCLA, USC, Warwick Business School, and WFA.

[†]London School of Economics and CEPR, e-mail: m.oehmke@lse.ac.uk.

[‡]Stockholm School of Economics and CEPR, e-mail: marcus.opp@hhs.se.

In recent years, the question of the social responsibility of business, famously raised by Friedman (1970), has re-emerged in the context of the spectacular rise of socially responsible investment (SRI). Assets under management in socially responsible funds have grown manifold,[1] and many traditional investors consider augmenting their asset allocation with environmental, social, and governance (ESG) scores (Pastor, Stambaugh and Taylor, 2020, Pedersen, Fitzgibbons and Pomorski, 2019). From an asset management perspective, this trend raises immediate questions about the financial performance of such investments (Hong and Kacperczyk, 2009, Chava, 2014, Barber, Morse and Yasuda, 2018). However, if socially responsible investing is to generate real impact, it must affect firms' production decisions. This raises additional, fundamental questions: Under which conditions can socially responsible investors impact firm behavior? And how should scarce socially responsible capital be allocated across firms?

Answering these questions requires taking a corporate finance view of socially responsible investment. To this end, we incorporate socially responsible investors and the choice between clean and dirty production into an otherwise standard model of corporate financing with agency frictions, building on Holmström and Tirole (1997). The model's main results are driven by the interaction of negative production externalities (which can lead to overinvestment in socially undesirable dirty production) and financing constraints (leading to underinvestment in socially desirable clean production). Such financing frictions are not only empirically relevant for young firms (an important source of clean innovation), but they also matter for mature firms that seek to replace profitable dirty production with more expensive clean production technologies.

We find that socially responsible investors can indeed push firms to adopt clean production. They optimally do so by raising a firm's financing capacity under clean production beyond the amount that purely profit-motivated investors would provide. The resulting increase in clean production raises total surplus, even compared to the scenario

[1] For example, the Global Sustainable Investment Alliance (2018) reports sustainable investing assets of \$30.7tn at the beginning of 2018, an increase of 34% relative to two years prior.

where all capital is held by socially responsible investors. However, increasing clean production beyond the scale that profit-motivated investors would fund is only possible if socially responsible make financial losses. Therefore, a necessary condition for socially responsible investors to break even, in social terms, on their impact investments is that they follow a broad mandate, in the sense that they internalize social costs generated by firms regardless of whether they are investors in these firms. When faced with an investment decision across many heterogeneous firms, scarce socially responsible capital should be allocated according to a social profitability index (SPI). One key feature of this micro-founded ESG metric is that avoided social costs are relevant for ranking investments. Hence, investments in "sin" industries are not necessarily inconsistent with the mandate of being socially responsible.

We develop these results in a parsimonious model, initially focusing on the investment decision of a single firm. The firm is owned by an entrepreneur with limited wealth, who has access to two constant-returns-to-scale production technologies, dirty and clean. Dirty production has a higher per-unit financial return, but entails significant social costs. Clean production is financially less attractive but socially preferable, because it generates lower (although not necessarily zero) social costs. Production under either technology requires the entrepreneur to exert unobservable effort, so that not all cash flows are pledgeable to outside investors. The firm can raise funding from (up to) two types of outside investors. Financial investors behave competitively and, as their name suggests, care solely about financial returns. Socially responsible investors also care about financial returns, but, in addition, they satisfy two conditions that, as our analysis reveals, are necessary for impact. First, they care *unconditionally* about external social costs generated by the firm (i.e., irrespective of whether they are investors in the firm). Second, they act in a coordinated fashion, so that they internalize the effect of their investments on production decisions. Socially responsible investors in our model are therefore most easily interpreted as a large (e.g., sovereign wealth) fund.

As a benchmark, we initially consider a setting in which only financial investors are present. Because of the entrepreneur's moral hazard problem, the amount of outside financing and, hence, the firm's scale of production are limited. Since the dirty production technology is financially more attractive, financial investors offer better financing terms for dirty production, enabling a larger production scale than under the clean technology. As a result, the entrepreneur may adopt the socially inefficient dirty production technology, even if she partially internalizes the associated externalities due to an intrinsic preference for clean production.

We then analyze whether and how socially responsible investors can address this inefficiency. We show that the optimal way to achieve impact (i.e., induce a change in the firm's production technology) is to relax financing constraints for clean production, thereby enabling additional value creation. One way the firm can implement the resulting financing agreement is by issuing two bonds, a green bond purchased by socially responsible investors and a regular bond purchased by financial investors. However, because financial investors are not willing to provide this scale on their own, the extra financing must involve a financial loss to socially responsible investors. Therefore, the green bond is issued at a premium, consistent with evidence in Baker et al. (2018) and Zerbib (2019).

Our results highlight a complementarity between socially responsible and financial investors. Because of this complementarity, total surplus (which, in our model, is determined by the total scale of clean production) is generally higher when both types of investors are present. The complementarity arises because of financial investors' disregard for externalities, which allows dirty production at a larger scale than the entrepreneur could achieve under self-financing. The resulting threat of dirty production relaxes the participation constraint for socially responsible investors and, thereby, generates additional financing capacity. Since financing constraints imply that clean production is below the socially optimally scale, this additional financing capacity enables a surplus-enhancing increase in the scale of clean production.

Our analysis identifies three necessary conditions for this complementarity to arise. First, socially responsible investors need to follow a broad mandate. This means that they must care unconditionally about external social costs of production, whether or not they are investors in the firm that produces them. If socially responsible investors follow a narrow mandate, so that they only care about social costs generated by their own investment, their presence does not reduce the social costs generated by the firm, since dirty production will then simply be financed by financial investors.[2] Second, the clean technology must be subject to financing constraints, so that additional clean scale is socially valuable. Third, socially responsible capital needs to be in sufficient supply to be able to discipline the threat of dirty production. If this is not the case, dirty production is not merely an off-equilibrium threat, but occurs in equilibrium.

While socially responsible capital has seen substantial growth over the last few years, it is likely that such capital remains scarce relative to financial capital that only chases financial returns. This raises the question of how scarce socially responsible capital is invested most efficiently. Which firms should impact investors target? A multi-firm extension of our model yields a micro-founded investment criterion for scarce socially responsible capital, the *Social Profitability Index* (SPI).

Similar to the profitability index, the SPI measures "bang for buck"— i.e., value created for socially responsible investors per unit of socially responsible capital consumed. However, unlike the conventional profitability index, the SPI not only reflects the (social) return of the project that is being funded, but also the counterfactual social costs that a firm would have generated in the absence of investment by socially responsible investors. For example, investment metrics for socially responsible investors should include estimates of carbon emissions that can be avoided if the firm adopts a cleaner production

[2] Our analysis is motivated by social costs such as carbon emissions, so that socially responsible investors are driven by the mitigation of negative production externalities. We discuss positive production externalities in Section 4.1. This extension reveals an interesting asymmetry: In the case of positive externalities, a narrow mandate (i.e., only accounting for the positive externalities generated by one's own investment) is more effective.

technology. Because avoided externalities matter, it can be efficient for socially responsible investors to invest in firms that, in an absolute sense, generate a high level of social costs even under clean production. Accordingly, investments in sin industries (see Hong and Kacperczyk (2009)) can be consistent with socially responsible investing. In contrast, it is efficient to not invest in firms that are already committed to clean production (e.g., because, intrinsically, the entrepreneur cares sufficiently about the environment), because clean production will occur regardless of investment by socially responsible investors.

Throughout the paper, we abstract from government intervention. One way to interpret our results is therefore as characterizing the extent to which the market can fix problems of social cost before the government imposes regulation or Pigouvian taxes. Another interpretation is that our analysis concerns those social costs that remain after the government has intervened. For example, informational frictions and political economy constraints may make it difficult for governments to apply Pigouvian taxes or ban dirty production (see, e.g., Tirole, 2012).[3] However, even if government intervention is possible, our analysis reveals that text-book regulation, in the form of Pigouvian taxes or a ban on dirty production, may result in lower total surplus than the allocation achieved via co-investment by financial and socially responsible investors. While Pigouvian taxes or bans on dirty production would certainly ensure the adoption of the clean technology (even when financing is provided by financial investors only), such regulation also eliminates the threat of dirty production, which is necessary to unlock additional capital by socially responsible investors. The broader point is that regulation that targets one source of inefficiency (externalities) but does not address the other (financing constraints) has limited effectiveness in our setting. Optimal regulation, which is beyond the scope of this paper, needs to account for both sources of inefficiencies.

[3] Examples of social costs for which government intervention is likely to be particularly difficult include those where the relevant externality is global in nature, as is the case for carbon emissions or systemic externalities caused by large financial institutions.

Related Literature. The theoretical literature on socially responsible investing consists of two main strands. Following the pioneering paper by Heinkel et al. (2001), the first strand studies the effects of *exclusion*, such as investor boycotts, divestment, or portfolio underweighting of dirty firms. Whether the threat of exclusion impacts a firm's production decisions depends on the cost imposed on the firm by not being able to (fully) access capital from socially responsible investors. The vast majority of this literature (see e.g., Pastor et al., 2020, Pedersen et al., 2019, De Angelis et al., 2020, Broccardo et al., 2020 , Zerbib, 2020) generates this effect via a reduction in risk sharing that raises the firm's cost of capital. However, Heinkel et al. (2001) and Broccardo et al. (2020) point out that the effect on risk premia is small when profit-seeking investors can substitute for divested capital.[4] In Landier and Lovo (2020), divestment can be effective because of a matching friction between firms and investors, which implies that a boycott by socially responsible investors (probabilistically) leaves the firm without access to financing.[5]

Our model completely shuts off the exclusion channel by considering a risk-neutral setting with perfectly elastic supply of profit-motivated capital. Our setting therefore echoes Broccardo et al. (2020), who, building on Heinkel et al. (2001) and Hart and Zingales (2017), conclude that "voice" (engagement) is more effective than "exit" (divestment).

Consequently, our paper is more closely related to the second strand of the literature, which studies the effects of *engagement* by activist investors who intrinsically care about social costs, thereby providing a corporate finance perspective on the economics of motivated agents (see e.g., Besley and Ghatak, 2005; Bénabou and Tirole, 2006). Rather than imposing costs on dirty firms via the threat of divestment, socially responsible activists effectively subsidize firms to adopt clean technologies through the direct pricing of social preference. Chowdhry et al. (2018) show that such subsidies optimally take the form of investment by socially-minded activists when firms cannot credibly commit to pursuing

[4] As Davies and Van Wesep (2018) point out, divestment can also have other, unintended consequences, for example, by inducing firms to prioritize short-term profit at the expense of long-term value.

[5] If divestment is optimally chosen to achieve impact, it remains an off-equilibrium threat, so that in equilibrium socially responsible investors earn the same return as financial investors.

social goals (there is no such commitment problem in our setting). Roth (2019) also compares impact investing with grants, highlighting the ability of investors to withdraw capital as an advantage over grants.[6]

Our contribution relative to this strand of literature is twofold. First, we highlight that socially responsible investors can achieve impact by relaxing financing constraints for clean production and that, in doing so, they generate a complementarity with financial investors. Second, our framework endogenizes the allocation of socially responsible capital across firms using a micro-founded investment criterion, the social profitability index.

1 Model Setup

Our modeling framework aims to uncover the role of socially responsible investing in a setting in which *production externalities* interact with *financing constraints*. It builds on the canonical model of corporate financing in the presence of agency frictions laid out in Holmström and Tirole (1997) and Tirole (2006). The main innovation is that the firm has access to two different production technologies, one of them "clean" (i.e., associated with low social costs) and the other "dirty" (i.e., associated with larger social costs). To focus on the role that socially responsible investors can play in alleviating distortions, we abstract from government intervention for most of the paper.[7] In Section 4.2, we discuss the effects of standard regulatory policies, such as Pigouvian taxes or banning the dirty production technology.

[6] A corollary of subsidizing clean production is that socially responsible investors sacrifice financial return. In contrast, in Gollier and Pouget (2014) a large activist investor can generate positive abnormal returns by reforming firms and selling them back to the market.

[7] The two technologies can therefore be interpreted as those available to the firm after government intervention has taken place. Because government intervention is usually subject to informational and political economy constraints, it seems reasonable that the social costs of production cannot be dealt with by the government alone, creating a potential role for socially responsible investors. Alternatively, our analysis can be interpreted as establishing what market forces (in the form of socially responsible investors) can achieve before government intervention takes place.

The entrepreneur, production, and moral hazard. Our setting considers a risk-neutral entrepreneur who is protected by limited liability and endowed with initial liquid assets of A. The entrepreneur has access to two production technologies $\tau \in \{C, D\}$, each with constant returns to scale.[8] The technologies are identical in terms of revenue generation. Denoting firm scale (capital) by K, the firm generates positive cash flow of RK with probability p (conditional on effort by the entrepreneur, as discussed below) and zero otherwise.

Where the technologies differ is with respect to the required ex-ante investment and the social costs they generate. In particular, the dirty technology D generates a non-pecuniary negative externality of $\phi_D > 0$ per unit of scale and requires a per-unit upfront investment of k_D. The clean technology results in a lower per-unit social cost $0 \leq \phi_C < \phi_D$, but requires a higher per-unit upfront investment of $k_C > k_D$.[9] The entrepreneur internalizes a fraction $\gamma^E \in [0, 1)$ of social costs, capturing potential intrinsic motives not to cause social harm.[10]

To generate a meaningful trade-off in the choice of technologies, we assume that the ranking of the two technologies differs depending on whether it is based on financial or social value. Denoting the per-unit financial value by $\pi_\tau := pR - k_\tau$ and the per-unit social value (surplus) by $v_\tau := \pi_\tau - \phi_\tau$, we assume that the dirty technology has higher financial value, $\pi_D > \pi_C$, but clean production generates higher social value, $v_C > 0 > v_D$.[11] The final inequality implies that the social value of the dirty production technologies is negative, meaning that the externalities caused by dirty production outweigh its financial

[8] In Section 4.1, we discuss robustness of our results to $N > 2$ technologies and decreasing returns to scale.

[9] The assumption that $0 \leq \phi_C < \phi_D$ reflects that our analysis focuses on the mitigation of negative production externalities by socially responsible investors. We discuss the case of positive production externalities in Section 4.1.

[10] We allow for social responsibility on account of the entrepreneur for added realism. All of our main results hold when the entrepreneur ignores externalities and simply maximizes profit (i.e., $\gamma^E = 0$).

[11] Once we allow for N technologies (see Section 4.1), we show how our results readily extend to cases where the dirtiest technology may no longer be the profit-maximizing technology. The case where the clean technology also maximizes profits is uninteresting for our analysis of socially responsible investment, since even purely profit-motivated capital would ensure clean production in this case.

value. The assumption that the dirty production technology has negative social value is not necessary for our results, but it simplifies the exposition because it implies that, from a social perspective, the dirty technology should never be adopted.

As in Holmström and Tirole (1997), the entrepreneur is subject to an agency problem. Whereas the choice of production technology is assumed to be observable (and, hence, contractible) effort is assumed to be unobservable (and, therefore, not contractible). Under each technology, the investment pays off with probability p only if the entrepreneur exerts effort ($a = 1$). The payoff probability is reduced to $p - \Delta p$ when the entrepreneur shirks ($a = 0$), where $p > \Delta p > 0$. Shirking yields a per-unit non-pecuniary benefit of B to the entrepreneur, for a total private benefit of BK. A standard result (which we will show below) is that this agency friction reduces the firm's unit pledgeable income by $\xi := p\frac{B}{\Delta p}$, the per-unit agency cost. A high value of ξ can be interpreted as an indicator of poor governance, such as large private benefits or weak performance measurement. We make the following assumption on the per-unit agency cost:

Assumption 1 *For each technology τ, the agency cost per unit of capital $\xi := p\frac{B}{\Delta p}$ satisfies*

$$\pi_\tau < \xi < pR - \frac{p}{\Delta p}\pi_\tau. \tag{1}$$

This assumption states that the moral hazard problem, as characterized by the agency cost per unit of capital ξ, is neither too weak nor too severe. The first inequality implies a finite production scale. The second inequality is a sufficient condition that rules out equilibrium shirking and ensures feasibility of outside financing. To streamline notation, π and v are defined conditional on the relevant case, in which the entrepreneur exerts effort (as usual, shirking is an off-equilibrium action).

Outside investors and securities. The entrepreneur can raise financing from (up to) two types of risk-neutral outside investors $i \in \{F, SR\}$, *financial investors* and *socially responsible investors*. Both investor types care about expected cash flows, but only socially responsible investors internalize social costs of production (i.e., $\gamma^{SR} > 0$, whereas $\gamma^{F} = 0$).[12,13] Regardless of whether the entrepreneur raises financing from both investor types or just one, it is without loss of generality to restrict attention to financing arrangements in which the entrepreneur issues securities that pay a total amount of $X := X^F + X^{SR}$ upon project success and 0 otherwise, where X^F and X^{SR} denote the payments promised to financial and socially responsible investors, respectively. Given that the firm has no resources in the low state, this security can be interpreted as debt or equity. The entrepreneur's utility can then be written as a function of the investment scale K, the total promised repayment X, the effort decision a, upfront consumption by the entrepreneur c, and the technology choice $\tau \in \{C, D\}$,

$$\begin{aligned} U^E(K, X, \tau, c, a) =& p(RK - X) - (A - c) - \gamma^E \phi_\tau K \\ & + \mathbb{1}_{a=0}[BK - \Delta p(RK - X)]. \end{aligned} \tag{U^E}$$

The first two terms of this expression, $p(RK - X) - (A - c)$, represent the project's net financial payoff to the entrepreneur under high effort, where $A - c$ can be interpreted as the upfront co-investment made by the entrepreneur. The third term, $\gamma^E \phi_\tau K$, measures the social cost internalized by the entrepreneur. The final term, $BK - \Delta p(RK - X)$, captures the incremental payoff conditional on shirking ($a = 0$). Exerting effort is incentive compatible if and only if $U^E(K, X, \tau, c, 1) \geq U^E(K, X, \tau, c, 0)$, which limits the total amount X that the entrepreneur can promise to repay to outside investors to

$$X \leq \left(R - \frac{B}{\Delta p}\right) K. \tag{IC}$$

Per unit of scale, the entrepreneur's pledgeable income is therefore given by $pR - \xi$.

[12] The assumption that at least some investors internalize social costs is consistent with evidence in Riedl and Smeets (2017), Bonnefon et al. (2019), and Hartzmark and Sussman (2019).

[13] In our model, financial investors literally do not care about social costs. However, an alternative setting, in which financial investors do care about social costs but do not act on them because of a free-rider problem, would yield equivalent results.

The resource constraint at date 0 implies that capital expenditures, Kk_τ, must equal the total investments made by the entrepreneur and outside investors,

$$Kk_\tau = A - c + I^F + I^{SR}, \tag{2}$$

where I^F and I^{SR} represent the investments of financial and socially responsible investors, respectively.

We impose two conditions on the behavior of socially responsible investors. As we will show later, both of these conditions are *necessary* for socially responsible investors to have impact.

Condition 1 (Broad Mandate) *Socially responsible investors are affected by externalities $\gamma^{SR}\phi_\tau K$ regardless of whether they invest in the firm.*

Condition 2 (Coordination) *Socially responsible investors allocate their capital in a coordinated fashion.*

Effectively, these conditions imply that socially responsible investors care unconditionally about externalities (Condition 1) and are willing to act on them (Condition 2).[14] Because the benefits of socially responsible investment (in the form of reduced externalities) are non-rival and non-excludable, coordination is necessary to ensure that socially responsible investors take into account that their investment affects the social cost generated by the firm (see literature on public goods following Samuelson, 1954). This condition is naturally satisfied if socially responsible capital is directed by one large fund, such as the Norwegian sovereign wealth fund. Another interpretation is that socially responsible investors are dispersed, but have found a way to overcome the free-rider

[14] Socially responsible therefore take a consequentialist view of their actions. Moisson (2020) studies divestment and shareholder activism considering various alternative moral criteria, including consequentialism, rule consequentialism, and shared responsibility.

problem that would usually arise.[15,16]

Under Conditions 1 and 2, the respective utility functions of outside investors, given an incentive-compatible financing arrangement, can be written as

$$U^F = pX^F - I^F, \tag{U^F}$$

$$U^{SR} = pX^{SR} - I^{SR} - \gamma^{SR}\phi_\tau K, \tag{U^{SR}}$$

where γ^{SR} captures the degree to which socially responsible investors internalize the social costs generated by the firm.

Let us make two brief observations regarding the objective function (U^{SR}) and what it means to be "socially responsible" in our model. First, in contrast to a planner, socially responsible investors do not maximize total surplus. Socially responsible investors care about their financial return and externalities generated by the firm, but they do not value rents that accrue to the entrepreneur.[17] Second, note that the sum $\gamma^E + \gamma^{SR} \in (0, 1]$ represents the fraction of total externalities that is taken into account by investors and the entrepreneur. When $\gamma^E + \gamma^{SR} = 1$, agents in the model jointly internalize all externalities. In the case $\gamma^E + \gamma^{SR} < 1$, some externalities are not internalized. The latter can be interpreted as a situation in which some externalities (e.g., those imposed on future generations) are simply unaccounted for. Alternatively, the partial internalization of externalities may capture, in reduced form, the effect of imperfect coordination among socially responsible investors. (In the extreme case $\gamma^{SR} = 0$, the free-rider problem among

[15] One such example is the establishment of the Poseidon Principles, an initiative by eleven major to promote green shipping, see Nauman (2019). Morgan and Tumlinson (2019) provide a model in which shareholders of a company value public good production but are subject to free-rider problems. Dimson et al. (2019) empirically document coordinated engagements by large investors.

[16] Another implication of the non-rival nature of the social benefit is that, unlike financial investors, socially responsible investor do not have an incentive to compete with each other. Because the reduction in social cost is non-rival and non-excludable, all socially responsible investors profit from it. Moreover, as we will show below, the (excludable) financial part of their overall return is negative, so that no socially responsible investors has an incentive to undercut on this dimension.

[17] The recent wave of inflows into ESG funds as well as the empirical evidence cited in footnote 12 all suggest that some investors explicitly care about externalities, in addition to caring about their financial payoffs. There is no evidence that these investors also value rents to insiders.

socially responsible investors is then so severe that they effectively act like financial investors.)

2 The Effect of Socially Responsible Investment

In this section, we investigate whether and how socially responsible investors can impact a single firm's investment choice, assuming that socially responsible capital is abundant relative to the funding needs of the firm. Our subsequent multi-firm setting in Section 3 analyzes how scarce socially responsible capital should be optimally allocated across firms. In Section 2.1, we first solve a benchmark case of firm financing without socially responsible investors. This benchmark shows that, when investors care exclusively about financial returns, the dirty technology may be chosen even when the entrepreneur has some concern for the higher social cost generated by dirty production (i.e., when $\gamma^E > 0$). In Section 2.2, we add socially responsible investors to the model and characterize conditions under which their presence has impact, in the sense that it changes the firm's production decision.

2.1 Benchmark: Financing from Financial Investors Only

The setting in which the entrepreneur can borrow exclusively from competitive financial investors corresponds to the special case $I^{SR} = X^{SR} = 0$. Then, the entrepreneur's objective is to choose a financing arrangement (consisting of scale $K \geq 0$, repayment $X^F \in [0, R]$, upfront consumption by the entrepreneur $c \geq 0$, and technology choice $\tau \in \{C, D\}$) that maximizes the entrepreneur's utility U^E subject to the entrepreneur's IC constraint and financial investors' IR constraint, $U^F \geq 0$.

As a preliminary step, it is useful analyze the financing arrangement that maximizes scale for a given technology τ.[18] Following standard arguments (see Tirole, 2006), this

[18] As discussed below, given competitive investors, maximizing scale is indeed optimal if the technology generates positive surplus from the entrepreneur's perspective.

agreement requires the entrepreneur to co-invest all her wealth (i.e., $c = 0$) and that the entrepreneur's IC constraint as well as the financial investors' IR constraint bind. The binding IC constraint ensures that the firm optimally leverages its initial resources A, whereas the binding IR constraint is a consequence of competition among financial investors.

When all outside financing is raised from financial investors, the maximum firm scale under production technology τ is then given by

$$K_\tau^F = \frac{A}{\xi - \pi_\tau}. \tag{3}$$

This expression shows that the entrepreneur can scale her initial assets A by a factor that depends on the *agency cost per unit of investment*, $\xi := p\frac{B}{\Delta p}$, and the financial value under technology τ, π_τ. As $\xi > \pi_D$ (see Assumption 1), the maximum investment scale is finite under either technology. The key observation from Equation (3) is that the maximum scale that the entrepreneur can finance from financial investors is larger under dirty than under clean production, $K_D^F > K_C^F$, since dirty production generates larger financial value, $\pi_D > \pi_C$, and financial investors only care about financial returns.

The following lemma highlights that the entrepreneur's technology choice τ_F is then driven by a trade-off between achieving scale and her concern for externalities. Of course, if the entrepreneur completely disregards externalities $\left(\gamma^E = 0\right)$, no trade-off arises and the entrepreneur always chooses dirty production to maximize scale.

Lemma 1 (Benchmark: Financial Investors Only) *When only financial investors are present, the entrepreneur chooses*

$$\tau_F = \arg\max_\tau (\xi - \gamma^E \phi_\tau) K_\tau^F. \tag{4}$$

The firm operates at the maximum scale that allows financial investors to break even,

$K^F_{\tau_F}$. *The entrepreneur's utility is given by*

$$\underline{U}^E = (\xi - \gamma^E \phi_{\tau_F}) K^F_{\tau_F} - A. \tag{5}$$

According to Lemma 1, when financing is raised from financial investors only, the entrepreneur chooses the technology τ_F that maximizes her payoff, which is given by the product of the per-unit payoff to the entrepreneur (agency rent net of internalized social cost) and the maximum scale under technology τ (given by Equation (4)). Maximum scale is optimal because, under the equilibrium technology τ_F, the project generates positive surplus for the entrepreneur and financial investors. It follows that the entrepreneur adopts the dirty technology whenever

$$(\xi - \gamma^E \phi_D) K^F_D > (\xi - \gamma^E \phi_C) K^F_C. \tag{6}$$

Given that the maximum scale is larger under the dirty technology, $K^F_D > K^F_C$, this condition is satisfied whenever the entrepreneur's concern for externalities γ^E lies below a strictly positive cutoff $\bar{\gamma}^E$.

Corollary 1 (Benchmark: Conditions for Dirty Production) *When only financial investors are present, the entrepreneur adopts the dirty production technology when* $\gamma^E < \bar{\gamma}^E := \frac{\xi(\pi_D - \pi_C)}{\phi_D(\xi - \pi_C) - \phi_C(\xi - \pi_D)}$.

Note that the entrepreneur can be induced to choose the dirty technology when financing from financial investors is available even if she were to choose the clean technology under autarky (i.e., self-financing).[19] Given that financing from financial investors is always available, this benchmark case shows that there is a potential role for socially responsible investors to steer the entrepreneur towards the clean production technology

[19] The entrepreneur prefers the clean technology under self-financing if and only if $\frac{A}{k_C}\left(\pi_C - \gamma^E \phi_C\right) > \frac{A}{k_D}\left(\pi_D - \gamma^E \phi_D\right)$. Hence, the entrepreneur is "corrupted" by financial markets when $\gamma^E \in \left(\widetilde{\gamma}^E, \bar{\gamma}^E\right)$ where $\widetilde{\gamma}^E := \frac{k_C \pi_D - k_D \pi_C}{k_C \phi_D - k_D \phi_C}$.

as long as $\gamma^E < \bar{\gamma}^E$.

2.2 Equilibrium with Socially Responsible Investors

We now analyze whether and how the financing arrangement and the resulting technology choice are altered when socially responsible investors are present. Because the entrepreneur could still raise financing exclusively from financial investors, the utility she receives under the financing arrangement with financial investors only, $\underline{U}^E$, now takes the role of an outside option to the entrepreneur.

2.2.1 Optimal Financial Contract with Socially Responsible Investors

Due to the broad mandate (Condition 1), socially responsible investors are affected by the social costs of production regardless of whether they have a financial stake in the firm. In particular, if socially responsible investors remain passive, their (reservation) utility is given by

$$\underline{U}^{SR} = -\gamma^{SR}\phi_{\tau_F}K^F_{\tau_F} < 0, \tag{7}$$

which reflects the social costs generated when the entrepreneur chooses the optimal production technology τ_F and scale $K^F_{\tau_F}$ (see Lemma 1) when raising financing from financial investors only.[20] To generate Pareto improvements relative to their respective outside options $\underline{U}^{SR}$ and $\underline{U}^E$, socially responsible investors can engage with the entrepreneur and agree on a contract, which specifies the technology τ, scale K, as well as the required financial investments and cash flow rights for all investors and the entrepreneur. We adopt the bargaining procedure of Hart and Moore (1998): With probability η, the entrepreneur gets to make a take-it-or-leave-it offer, giving her the maximum payoff, denoted by $\overline{U}^E$, while socially responsible investors remain at their reservation utility $\underline{U}^{SR}$. With probability $1-\eta$, socially responsible investors get to make a take-it-or-leave-it offer,

[20] If the entrepreneur could not raise financing from financial investors, the outside option for socially responsible investors would be determined by the entrepreneur's technology choice under self-financing.

leading to the analogous respective payoffs of $\overline{U}^{SR}$ and $\underline{U}^{E}$.[21] We augment this bargaining game by allowing socially responsible to make an offer before the above bargaining game starts.[22] Then, for a given surplus division parameter η, the optimal bilateral agreement can be found by maximizing the payoff to socially responsible investors, subject to the entrepreneur's acceptance of the proposed contract:

Problem 1 (Optimal Bilateral Agreements)

$$\max_{I^F, I^{SR}, X^{SR}, X^F, K, c, \tau} pX^{SR} - I^{SR} - \gamma^{SR}\phi_\tau K \tag{8}$$

subject to the entrepreneur's IR constraint:

$$U^E\left(K, X^{SR} + X^F, \tau, c, 1\right) \geq (1-\eta)\,\underline{U}^E + \eta\bar{U}^E, \tag{IR$^{\text{E}}$}$$

as well as the entrepreneur's IC constraint, the resource constraint (2), the financial investors' IR constraint $U^F \geq 0$*, and the non-negativity constraints* $K \geq 0, c \geq 0$*.*

The constraint IR^{E} ensures that, under the proposed contract, the entrepreneur receives at least as much as her expected payoff from playing the bargaining game that follows if the initial offer were to be rejected, $(1-\eta)\,\underline{U}^E + \eta\bar{U}^E$.[23] Note that this formulation permits the possibility of compensating the entrepreneur with sufficiently high upfront consumption ($c > 0$) in return for smaller scale K, possibly even shutting down production completely (as in the typical Coasian solution, see Coase, 1960).[24]

[21] The respective values of $\overline{U}^E$ and $\overline{U}^{SR}$ are given in equations (A.17) and (A.18) in the appendix.

[22] The role of the augmentation is to ensure deterministic contracts and allocations. In equilibrium, the initial offer made by socially responsible investors reflects the surplus division parameter η and will be accepted by the entrepreneur. Due to the linearity of the Pareto frontier, this augmentation does not yield any efficiency benefits, in contrast to Hart and Moore (1998).

[23] A simple special case is $\eta = 0$. In this case, socially responsible investors have all the bargaining power and the entrepreneur is held to her outside option of raising financing from financial investors only.

[24] However, such a complete shut down does not occur in equilibrium. This follows from the assumption that at least one of the technologies generates positive joint surplus, i.e., $v_C > 0$. We discuss other cases in the robustness section 4.1.

Proposition 1 (Technology and Scale with Socially Responsible Investors) *Let* $\hat{v}_\tau := \pi_\tau - \left(\gamma^E + \gamma^{SR}\right)\phi_\tau \geq v_\tau := \pi_\tau - \phi_\tau$ *denote bilateral surplus per unit of scale.*[25] *Then, in any optimal financing agreement, the technology choice is independent of the surplus division parameter* η *and given by*

$$\hat{\tau} = \arg\max_\tau \hat{v}_\tau \hat{K}_\tau(\eta) = \arg\max_\tau \frac{\hat{v}_\tau}{\xi - \gamma^E \phi_\tau}. \tag{9}$$

The associated equilibrium scale $\hat{K}_{\hat{\tau}}(\eta)$ *is increasing in* η *where*

$$\hat{K}_\tau(\eta) = \eta K_\tau^{\max} + (1-\eta) K_\tau^{\min}, \tag{10}$$

$K_\tau^{\min} = \frac{\xi - \gamma^E \phi_{\tau_F}}{\xi - \gamma^E \phi_\tau} K_{\tau_F}^F$ *and* $K_\tau^{\max} = \frac{\xi - \gamma^E \phi_{\tau_F} - \hat{v}_{\tau_F}}{\xi - \gamma^E \phi_\tau - \hat{v}_\tau} K_{\tau_F}^F$*. The entrepreneur consumes no resources upfront,* $\hat{c} = 0$ *.*

Proposition 1 contains the main theoretical result of the paper. Intuitively, the optimal technology choice $\hat{\tau}$ maximizes total bilateral surplus, which is driven both by the per-unit surplus $\hat{v}_\tau$ and the relevant scale $\hat{K}_\tau(\eta)$. The equilibrium scale, $\hat{K}_{\hat{\tau}}(\eta)$, is increasing in the entrepreneur's bargaining power η. It takes on its minimum value, $K_{\hat{\tau}}^{\min}$ when all the bargaining power rests with socially responsible investors ($\eta = 0$) and its maximum value $K_{\hat{\tau}}^{\max}$ when the entrepreneur has all the bargaining power ($\eta = 1$).

In contrast, the optimal technology choice $\hat{\tau}$ is independent of the division of surplus. Mathematically, this invariance result follows from the observation that $\arg\max_\tau \hat{v}_\tau K_\tau^{\min} = \arg\max_\tau \hat{v}_\tau K_\tau^{\max}$. Economically, it follows because Pareto improvements can only be achieved via changes in the production technology, but not via changes in scale: While increases in scale do indeed generate higher bilateral surplus $\hat{v}_\tau K$ for any fixed technology with $\hat{v}_\tau > 0$, the moral hazard friction implies that the entrepreneur's payoff must increase by more than the associated increase in bilateral surplus, which prohibits Pareto

[25] Since financial investors break even, this also corresponds to the joint surplus, per unit of scale, accruing to all investors and the entrepreneur.

improvements.

When a Pareto improvement is feasible by switching to the clean technology, an increase in the entrepreneur's bargaining power translates into a strictly larger equilibrium scale. Intuitively, increased bargaining power implies that more resources need to be shifted to the entrepreneur. Rather than consuming these resources upfront, the underinvestment problem makes it optimal to use these additional resources to increase scale rather than to consume them upfront.

Implementation. While the optimal financing arrangement uniquely pins down the production side (i.e., technology choice and scale), there exists a continuum of co-investment arrangements between financial and socially responsible investors that solve Problem 1 for any given surplus division parameter η. This indeterminacy arises because any increase in cash flows accruing to financial investors, $\hat{X}^F$, translates at competitive terms into higher upfront investment by financial investors, $\hat{I}^F$.

Corollary 2 (Optimal Co-investment Arrangements) *The total payout (to both investors) satisfies* $\hat{X} = \left(R - \frac{B}{\Delta p}\right) \hat{K}_{\hat{\tau}}(\eta)$ *in exchange for their joint upfront investment of* $\hat{I} = \hat{K}_{\hat{\tau}}(\eta)\, k_{\hat{\tau}} - A$*. The set of optimal co-investment arrangements can be obtained by tracing out the cash-flow share accruing to socially responsible investors* $\lambda \in [0,1]$ *and setting* $\hat{X}^{SR} = \lambda \hat{X}$*,* $\hat{X}^F = (1-\lambda)\hat{X}$*,* $\hat{I}^F = p\hat{X}^F$ *and* $\hat{I}^{SR} = \hat{I} - \hat{I}^F$*.*

There are two particularly intuitive ways in which the optimal financing arrangement characterized in Proposition 1 and Corollary 2 can be implemented.[26]

Corollary 3 (Implementation) *The following securities implement the optimal financing agreement:*

1. Green bond and regular bond: *The entrepreneur issues two bonds with respective*

[26] Under both implementations, the security targeted at socially responsible investors is issued at a premium in the primary market (see Corollary 5 below), ensuring that only socially responsible investors have an incentive to purchase this security.

face values $\hat{X}^F$ and $\hat{X}^{SR}$ at respective prices $\hat{I}^F$ and $\hat{I}^{SR}$. The green bond contains a technology-choice covenant specifying technology $\hat{\tau}$.

2. Dual-class share structure: *The entrepreneur issues voting and non-voting shares, where shares with voting rights yield an issuance amount of $\hat{I}^{SR}$ in return for control rights and a fraction λ of dividends. The remaining proceeds $\hat{I}^F$ are obtained in return for non-voting shares with a claim on a fraction $1-\lambda$ of dividends.*

It is worthwhile pointing out that it is not necessary for the optimal financing agreement to restrict the entrepreneur from seeking financing for additional dirty production. The financing agreement described in Proposition 1, Corollaries 2 and 3 exhausts all pledgeable assets, so that financial investors would not provide any additional financing for the dirty technology.

2.2.2 Impact

To highlight the economic mechanism behind Proposition 1, this section provides a more detailed investigation of the case in which socially responsible investors have impact, which we define as an induced change in the firm's production decision, through a switch in technology from $\tau_F = D$ to $\hat{\tau} = C$ and/or a change in production scale.[27] Based on Proposition 1, the following corollary summarizes the conditions for impact.

Corollary 4 (Impact) *Socially responsible investors who follow a broad mandate have impact if and only if $\gamma^E < \bar{\gamma}^E$ and $\gamma^{SR} \geq \bar{\gamma}^{SR}$, where the threshold $\bar{\gamma}^{SR}$ is decreasing in γ^E.*

Intuitively, impact requires that, absent socially responsible investors, the firm chooses the dirty technology ($\gamma^E < \bar{\gamma}^E$) and that socially responsible investors care sufficiently

[27] If investment by socially responsible investors does not result in a change in production technology compared to the benchmark case (i.e., $\hat{\tau} = \tau_F$), there is no impact and we obtain the same scale, $\hat{K}_\tau(\eta) = K^F_{\tau_F}$, and utility for all agents in the economy as in the benchmark case. This (less interesting) situation occurs either if the entrepreneur adopts the clean production technology even in the absence of investment by socially responsible investors, or if the entrepreneur adopts the dirty technology irrespective of whether socially responsible investors provide funding.

about the externality to change the entrepreneur's mind ($\gamma^{SR} \geq \bar{\gamma}^{SR}$). The latter cutoff is lower, the more the entrepreneur herself cares about the externality. Of course, if the entrepreneur and socially responsible investors jointly internalize all externalities, $\gamma^E + \gamma^{SR} = 1$, production will always be clean, since, in this case, their bilateral surplus coincides with total surplus (i.e., $\hat{v}_C = v_C > 0 > v_D$).

Complementarity between Financial and Socially Responsible Capital. When these conditions for impact are satisfied, then the equilibrium of our model features a complementarity between financial and socially responsible investors.

Proposition 2 (Complementarity) *Suppose that $\gamma^E < \bar{\gamma}^E$ and $\gamma^{SR} \geq \bar{\gamma}^{SR}$, then financial capital and socially responsible capital act as complements: The equilibrium clean scale with both investor types, $\hat{K}_C(\eta)$, is larger than the clean scale that can be financed in an economy with only one of the two investor types,*

$$\hat{K}_C(\eta) > \max\left\{K_C^F, K_C^{SR}(\eta)\right\}. \tag{11}$$

Total surplus $v_C \hat{K}_C(\eta)$ is increasing in the entrepreneur's bargaining power η.

The key feature of Proposition 2 is that the equilibrium clean scale in the presence of both investor types strictly exceeds the scale that is attainable with only one investor type. Let us first highlight why the equilibrium clean scale with both investors exceeds the maximum clean scale that can be funded by financial investors, $\hat{K}_C(\eta) > K_C^F$. When $\gamma^E < \bar{\gamma}^E$, a clean scale of K_C^F is not large enough to induce clean production when only financial investors are present—the entrepreneur prefers dirty production at scale K_D^F. Therefore, to induce the entrepreneur to switch to the clean production technology, socially responsible investors need to inject additional resources into the firm. Due to the moral hazard friction and ensuing underinvestment problem, this capital injection is optimally used to raise the scale of clean production above and beyond what financial investors are willing to offer, so that $\hat{K}_C(\eta) > K_C^F$.

Perhaps more surprisingly, $\hat{K}_C(\eta)$ also exceeds the scale that could be financed if only socially responsible investors were present. The reason is that financial investors' disregard for externalities allows dirty production at a larger scale than the entrepreneur could achieve under self-financing (i.e., when no financial investors are around). The resulting pollution threat relaxes the participation constraint for socially responsible investors, through its effect on their reservation utility, $\underline{U}^{SR} = -\gamma^{SR}\phi_D K_D^F$. This unlocks additional financing capacity, so that $\hat{K}_C(\eta) > K_C^{SR}(\eta)$.

Since a larger scale of clean production is socially valuable, Proposition 2 implies that social surplus, $v_C\hat{K}_C(\eta)$, is strictly higher when both financial and socially responsible investors deploy capital, relative to the case in which all capital is allocated by either financial or socially responsible investors. Since the equilibrium scale is increasing in the entrepreneur's bargaining power, total surplus is highest when all the bargaining power rests with the entrepreneur.

Abstracting from specific modeling details, two basic ingredients are necessary for the interaction of the two investor types to generate additional social value. First, there must be underinvestment in the clean technology. In our setup, this arises because both investor types do not value agency rents that accrue to the entrepreneur because of the moral hazard friction.[28] Second, socially responsible investors need to care about the externality regardless of their investments (the "broad mandate"). This ingredient implies that the threat of dirty production enabled by financial investors acts as a *quasi asset* to the firm, generating additional financing capacity from socially responsible investors. Because of underinvestment (the first ingredient), the additional financing result in an increase in clean scale, which is socially valuable.

[28] Note that the objective of socially responsible investors (see U^{SR}) differs from the maximization of social surplus $v_C K$ even if $\gamma^{SR} = 1$.

The cost of impact. Even though socially responsible investors only invest if doing so increases their utility relative to the case in which they remain passive,

$$\Delta U^{SR} := \bar{U}^{SR} - \underline{U}^{SR} = (1-\eta)\left(\hat{v}_C K_C^{\min} - \hat{v}_D K_D^F\right) \geq 0, \tag{12}$$

they do not break even in financial terms on their impact investment.

Corollary 5 (Socially Responsible Investors Make a Financial Loss) *Impact (a switch from $\tau_F = D$ to $\hat{\tau} = C$) requires that socially responsible investors make a financial loss. That is, in any optimal financing arrangement as characterized in Proposition 1,*

$$p\hat{X}^{SR} < \hat{I}^{SR}. \tag{13}$$

Intuitively, to induce a change from dirty to clean production, socially responsible investors need to enable a scale for the clean technology greater than the clean scale offered by competitive financial investors in isolation. Because financial investors just break even at that scale, socially responsible investors must make a financial loss on any additional scale they finance.[29] Empirically, Corollary 5 therefore predicts that impact funds must have a negative alpha or, equivalently, that funds generating weakly positive alpha cannot generate (real) impact.

Our model also predicts that the financial loss for socially responsible investors, $p\hat{X}^{SR} - \hat{I}^{SR}$, occurs at the time when the firm seeks to finance investment in the primary market, consistent with evidence on the at-issue pricing of green bonds in Baker et al. (2018) and Zerbib (2019). However, if socially responsible investors were to sell their cash flow stake $\hat{X}^{SR}$ to financial investors after the firm has financed the clean technology, our model does not predict a price premium for the "green" security in the secondary market (i.e., in the secondary market, the security would be fairly priced at $p\hat{X}^{SR}$).[30]

[29] Socially responsible investors are nevertheless willing to provide financing because their financial loss, $p\hat{X}^{SR} - \hat{I}^{SR}$, is outweighed by the utility gain resulting from reduced social costs, $\gamma^{SR}\left(\phi_D K_D^F - \phi_C \hat{K}\right)$.

[30] Intuitively, in our static model, control (or a technology covenant) only matters once, at the time

Necessary conditions for impact. The analysis above reveals why Conditions 1 and 2 are both necessary for socially responsible investors to have impact. To see the necessity of the broad mandate, suppose first that Condition 1 is violated and that socially responsible investors follow a narrow mandate, in that they only care about social costs that are a direct consequence their own investments. Because, under the narrow mandate, socially responsible investors ignore the social costs of firms that are financed by financial investors, the threat of dirty production does not relax their participation constraint. Hence, the key force that generates the additional financing capacity for clean production (see Proposition 2) is absent and, therefore, no impact can be achieved.[31]

Next suppose that socially responsible investors are infinitesimal and uncoordinated, so that Condition 2 is violated. Then, due to the resulting free-rider problem, each individual investor takes social costs generated by the firm as given and, therefore, behaves as if $\gamma^{SR} = 0$. No impact can be achieved because socially responsible investors behave like financial investors.

Finally, a third necessary condition for impact is that socially responsible capital is available in sufficient amounts to ensure adoption of the clean production technology. When this is not the case, the presence of financial capital can induce firms to adopt the dirty production technology, leading to a social loss. We discuss this case in Section 3, where we consider an economy with multiple firms and limited socially responsible capital. This analysis will shed further light on how the composition of investor capital (and not just the aggregate amount) matters for total surplus.

---

of initial investment. In a more general, dynamic setting, control could matter multiple times (e.g., whenever investment technologies are chosen).

[31] In Section 4.1, we revisit the necessity of the broad mandate in the context of social goods (i.e., technologies with positive production externalities).

3 The Social Profitability Index

Based on the framework presented above, we now derive a micro-founded investment criterion to guide scarce socially responsible capital. To do so, we extend the single-firm analysis presented in Section 2 to a multi-firm setting with limited socially responsible capital. For ease of exposition, we focus on the case in which socially responsible investors make a take-it-or-leave-it offer (i.e., the entrepreneur's bargaining power is $\eta = 0$) and use the shorthand notation $\hat{K} = \hat{K}_C(0)$.

Let κ be the aggregate amount of socially responsible capital (we continue to assume that financial capital is abundant) and consider an economy with a continuum of infinitesimal firms grouped into distinct firm types.[32] Firms that belong to the same firm type j are identical in terms of all relevant parameters of the model, whereas firms belonging to distinct types differ according to at least one dimension (with Assumption 1 satisfied for all types). Let $\mu(j)$ denote the distribution function of firm types, then the aggregate social cost in the absence of socially responsible investors is given by

$$\int_{\gamma_j^E < \bar{\gamma}_j^E} \phi_{D,j} K_{D,j}^F d\mu(j) + \int_{\gamma_j^E \geq \bar{\gamma}_j^E} \phi_{C,j} K_{C,j}^F d\mu(j). \tag{14}$$

The first term of this expression captures the social cost generated by firms that, in the absence of socially responsible investors, choose the dirty technology ($\gamma_j^E < \bar{\gamma}_j^E$), whereas the second term captures firm types run by entrepreneurs that have enough concern for external social costs that they choose the clean technology even in absence of socially responsible investors ($\gamma_j^E \geq \bar{\gamma}_j^E$).

Given this aggregate social cost, how should socially responsible investors allocate their limited capital? One direct implication of Proposition 1 is that any investment in firm types with $\gamma_j^E \geq \bar{\gamma}_j^E$ cannot be optimal as these firms adopt the clean technology even when raising financing from financial investors only. For the remaining firm types,

[32] The assumption that firms are infinitesimally small is made only to rule out well-known difficulties that can arise when ranking investment opportunities of discrete size.

the payoff to socially responsible investors from reforming a firm of type j is given by:

$$\Delta U_j^{SR} = (\pi_{C,j} - \xi_j)\,\hat{K}_j + A_j + \gamma^{SR}\left[\phi_{D,j}K_{D,j}^F - \phi_{C,j}\hat{K}_j\right]. \tag{15}$$

The first two terms of this expression capture the total financial payoff to socially responsible investors, net of the agency cost that is necessary to incentivize the entrepreneur. The third term captures the (internalized) change in social cost that results from inducing a firm of type j to adopt the clean production technology.

Given limited capital, socially responsible investors are generally not able to reform all firms. They should therefore prioritize investments in firm types that maximize the *impact per dollar invested*. This is achieved by ranking firms according to a variation on the classic profitability index, the *social profitability index* (SPI). The SPI divides the change in payoffs to socially responsible investors, ΔU_j^{SR}, by the amount socially responsible investors need to invest to impact the firm's behavior, I^{SR}.[33]

$$\mathrm{SPI}_j = \mathbb{1}_{\gamma_j^E < \bar{\gamma}_j^E}\frac{\Delta U_j^{SR}}{I_j^{SR}}. \tag{16}$$

Proposition 3 (The Social Profitability Index (SPI)) *Socially responsible investors should rank firms according to the social profitability index, SPI_j. There exists a threshold $SPI^*(\kappa) \geq 0$ such that socially responsible investors with scarce capital κ should invest in all firms for which $SPI_j \geq SPI^*(\kappa)$.*

According to Proposition 3, it is optimal to invest in firms ranked by the SPI until no funds are left, which happens at the cutoff $SPI^*(\kappa)$. Social capital is scarce if and only if the amount κ is not sufficient to fund all firm types with $SPI_j > 0$. The SPI links the attractiveness of an investment by socially responsible investors to underlying parameters of the model, thereby shedding light on the types of investments that socially

[33] The change in the payoff to socially responsible investors, ΔU_j^{SR}, is the same across all financing agreements characterized in Proposition 1. Absent other constraints, it is therefore optimal for socially responsible investors to choose the minimum co-investment that implements clean production.

responsible investors should prioritize.

Proposition 4 (SPI Comparative Statics) *As long as* $\gamma_j^E < \bar{\gamma}_j^E$*, the SPI is increasing in the avoided social cost,* $\Delta\phi := \phi_D - \phi_C$*, and the entrepreneur's concern for social cost,* γ^E*, and decreasing in the financial cost associated with switching to the clean technology,* $\Delta\pi := k_C - k_D$*.*

Proposition 4 states that socially responsible investors should prioritize in firms for which avoided social cost is high, as reflected in the difference in social costs under the clean and the dirty technology, $\Delta\phi$. Because the SPI reflects the relative social cost, it can be optimal for socially responsible investors to invest in firms that generate significant social costs, provided that these firms would have caused even larger social costs in the absence of engagement by socially responsible investors. The avoided social cost $\Delta\phi_j$ has to be traded off against the associated financial costs, as measured by the resulting reduction in profits $\Delta\pi_j$.

Another implication is that, as long as $\gamma_j^E < \bar{\gamma}_j^E$, firms with more socially minded entrepreneurs should be prioritized, because they require a smaller investment from socially responsible investors to be convinced to reform. However, as soon as the entrepreneur internalizes enough of the externalities, so that she chooses the clean technology even if financed by financial investors (i.e., $\gamma_j^E \geq \bar{\gamma}_j^E$), the SPI drops discontinuously to zero. Socially responsible investors should not invest in these firms.

To obtain a closed-form expression for the SPI, it is useful to consider the special case of $\gamma^E = 0$ and $\gamma^{SR} = 1$. Moreover, while strictly speaking it is optimal to minimize socially responsible investors' investment by selling all cash flow rights to financial investors, suppose that socially responsible investors need to receive a fraction λ_j of a firm's cash flow rights. This minimum cash-flow stake pins down I_j^{SR}. The assumption of a required cash-flow stake for socially responsible investors can be justified on two grounds. First, it is natural that socially responsible investors cannot rely purely on utility derived from the non-pecuniary benefits of reducing social costs, but require a certain amount of financial payoffs alongside non-pecuniary payoffs. Second, the minimum cash flow share λ_j can be interpreted as a reduced form representation of the control rights that are necessary to implement ensure that firm j implements the clean technology.[34] Given these assumptions, the SPI takes the following simple expression,

$$\mathrm{SPI}_j = \frac{\Delta\phi_j - \Delta\pi_j}{\Delta\pi_j + \lambda_j\left(p_j R_j - \xi_j\right)}. \tag{17}$$

This expression reveals the trade-off between the two key ingredients of the SPI, avoided pollution $\Delta\phi_j$ and foregone profits $\Delta\pi_j$ (see Proposition 4).[35] Moreover, as can be seen from Equation (17), our model rationalizes why environmental, social, and governance issues are usually bundled into one ESG score. A connection between these distinct aspects of corporate behavior arises because the severity of the manager's agency problem (a proxy for governance) determines the financing constraints the firm faces with respect to both financial and socially responsible investors. The SPI reflects these financing constraints because they interact with the (environmental and social) externalities generated by the firm. Since financing constraints apply to both the clean and the dirty technology, the total effect on the SPI is driven by two channels. Ceteris paribus, higher agency rents make it more expensive for socially responsible investors to finance clean production, leading to a lower SPI. However, higher agency rents also limit the scale that is offered by financial investors for dirty production, thereby reducing the necessary financial subsidy to induce clean production, leading to a higher SPI. In our baseline setup, this second effect dominates, which explains why the SPI given in Equation (17) is increasing in ξ.[36]

[34] This could be the case because the entrepreneur cannot commit to the adoption of the clean technology. In this case, a cash-flow stake for socially responsible investors and blunt the entrepreneur's profit motive (see Chowdhry et al., 2018) or may allow socially responsible investors to enforce appropriate technology adoption, for example, via voting rights.

[35] While these ingredients are conceptually intuitive, implementation of the SPI requires relatively detailed knowledge of the production process within a given industry, in order to be able to estimate avoided emissions.

[36] In an extension presented in Section 4.1, we allow for agency costs that are technology-specific. In this case, the SPI decreases in the agency cost under the clean technology and increases in the agency

To conclude this section, we briefly revisit the complementarity result given in Proposition 2 in a setting with limited socially responsible capital. The change in total surplus relative to the case without socially responsible investors, $\Delta\Omega$, results purely from the set of reformed firms: firms for which $\gamma_j^E < \bar{\gamma}_j^E$ and $SPI_j \geq SPI^*(\kappa)$. We can therefore write the change in total surplus as

$$\Delta\Omega = \int_{j:\gamma_j^E<\bar{\gamma}_j^E \ \& \ SPI_j \geq SPI^*(\kappa)} \left(v_{C,j}\hat{K}_j - v_{D,j}K_{D,j}^F \right) d\mu(j). \tag{18}$$

Clearly, if socially responsible capital is abundant, the results of Proposition 2 still apply: Total surplus is strictly higher in an economy with both types of investors than in an economy where all capital is held exclusively by either financial or socially responsible investors. In contrast, when socially responsible capital is scarce there is a trade-off. On the one hand, the set of reformed firms contributes towards higher total surplus, as before. On the other hand, the set of unreformed dirty firms exhibit overinvestment in the dirty technology due to the presence of competitive financial capital without regard for externalities. This trade-off implies that the right *balance* between socially responsible and financial capital is important for a complementarity between the two types of capital to arise.

4 Discussion

4.1 Generalizing the production technology

In our baseline model, we considered the choice between two constant-returns-to-scale production technologies with identical cash flows and agency rents. Moreover, we focused on the case, in which externalities of production are negative for all production technologies. As we show in this section, these assumptions can be relaxed relatively

cost under the dirty technology.

straightforwardly. In particular, Proposition 1 generalizes to multiple technologies and social goods. Moreover, even when the production technology exhibits decreasing returns to scale, it remains optimal to reward the entrepreneur with additional scale as long as financing frictions for the clean technology are significant.

Many (heterogeneous) technologies and social goods. Let us first retain the assumption of constant returns to scale, but generalize all other dimensions of the available production technologies. In particular, suppose that the entrepreneur has access to N production technologies characterized by technology-specific cash flow, cost, and moral hazard parameters R_τ, k_τ, p_τ, Δp_τ, and B_τ. The differences in parameters could reflect features such as increased willingness to pay for goods produced by firms with clean production technologies, implying $R_C > R_D$ (for models with this feature, see Aghion et al., 2019, Albuquerque et al., 2019), or they could capture differences in competition or market structure that affect the relative payoffs of behaving ethically (Dewatripont and Tirole, 2020). In addition, in contrast to the baseline model, we now allow for the technology-specific social cost parameter ϕ_τ to be negative, in which case the technology generates a positive externality (a social good).

In analogy to the baseline model, we can then define, for each technology $\tau \in \{1, ..., N\}$, the financial value π_τ, the agency rent ξ_τ, and the maximum scale available from financial investors K_τ^F, maintaining the assumption that $\xi_\tau > \pi_\tau$ for all τ, so that the maximum scale of production is finite. A straightforward extension of Lemma 1 then implies that, in the absence of investment by socially responsible investors, the entrepreneur chooses the technology τ_F such that:

$$\tau_F = \arg\max_\tau \frac{\pi_\tau - \gamma^E \phi_\tau}{\xi_\tau - \pi_\tau}. \tag{19}$$

Equation (19) shows that even with N general technologies, the entrepreneur's choice of technology is essentially the same as in Lemma 1, with the exception that the agency cost ξ_τ is now project specific. Moreover, Equation (19) clarifies the entrepreneur's relevant outside option in the presence of N technologies: In particular, adopting any production technology dirtier than τ_F is not a credible threat.

The induced technology choice in the presence of socially responsible investors $\hat{\tau}$ and the associated capital stock $\hat{K}$ are given by

$$\hat{\tau} = \arg\max_{\tau} \frac{\hat{\upsilon}_\tau}{\xi_\tau - \gamma^E \phi_\tau}, \tag{20}$$

$$\hat{K}(\eta) = \begin{cases} \left[(1-\eta) \frac{\xi_{\tau_F} - \gamma^E \phi_{\tau_F}}{\xi_{\hat{\tau}} - \gamma^E \phi_{\hat{\tau}}} + \eta \frac{\xi - \gamma^E \phi_{\tau_F} - \hat{\upsilon}_{\tau_F}}{\xi - \gamma^E \phi_{\hat{\tau}} - \hat{\upsilon}_{\hat{\tau}}} \right] K^F_{\tau_F} & \text{if } \hat{\upsilon}_{\hat{\tau}} > 0 \\ 0 & \text{if } \hat{\upsilon}_{\hat{\tau}} \leq 0 \end{cases}. \tag{21}$$

These expressions mirror Proposition 1, except that technology choice and scale in the presence of socially responsible investors now also depend on the technology-specific severity of the agency problem ξ_τ. Ceteris paribus, a smaller agency problem makes it more likely that a technology is adopted, both in the presence of financial investors only and when there is co-investment by socially responsible investors (Equations (19) and (20), respectively).

Whereas the formal expressions are unaffected by whether the externality is negative or positive, there is one important difference between these two cases. When externalities are negative, a broad mandate (Condition 1) is necessary to ensure that socially responsible investors have impact. A broad mandate reduces the outside option for socially responsible investors (see Equation (7)), thereby unlocking the required additional financing capacity. In contrast, when the externalities under technology D are positive, $\phi_D < 0$, the outside option for socially responsible investors is higher under a broad mandate than under a narrow mandate (the outside option is positive under a broad mandate, whereas it is zero under a narrow mandate). Therefore, in the presence of positive externalities, impact is possible and, in fact, more likely to occur under a narrow mandate, revealing an interesting asymmetry between preventing social costs and encouraging social goods.

The more general technology specification yields some additional insights about cases that we previously excluded. First, it is possible that for some industries the cleanest technology also maximizes financial value (e.g., because of demand by socially responsible consumers). In this case, there is no trade-off between doing good and doing well and, hence, socially responsible investors play no role. Second, the dirty technology may be the socially optimal technology when cleaner technologies are too expensive. Also in this case, there is no role for socially responsible investors. Finally, it is possible that, for some industries, any feasible technology τ yields negative bilateral surplus (i.e., $\hat{v}_{\hat{\tau}} < 0$). In this case, the socially optimal scale is zero and the entrepreneur is optimally rewarded with a transfer $\hat{c} > 0$ to shut down production.

Decreasing returns to scale. We now consider the case in which the two production technologies $\tau \in \{C, D\}$ exhibit decreasing returns to scale. In particular, suppose that the *marginal financial value* $\pi_\tau(K)$ is strictly decreasing in K. Then, the first-best scale K_C^{FB} under the (socially efficient) clean technology is characterized by the first-order condition

$$\pi_C\left(K_C^{FB}\right) = \phi_C. \tag{22}$$

Now consider the scenario in which technology D is chosen in the absence of socially responsible investors, with an associated scale of K_D^F. Moreover, for ease of exposition, focus on the case $\gamma^E + \gamma^{SR} = 1$, so that socially responsible investors have incentives to implement first-best scale. The optimal financing agreement that socially responsible investors offer to induce the entrepreneur to switch to the clean technology then comprises three cases.

1. If the financing constraints generated by the agency problem are severe, so that the maximum clean scale under the benchmark financing agreement with financial investors lies below a cutoff $\bar{K}$, i.e., $K_C^F \leq \bar{K} < K_C^{FB}$, the optimal agreement offered

by socially responsible investors rewards the entrepreneur exclusively through an increase in scale (rather than upfront consumption), as in Proposition 1. Even with socially responsible investors, the resulting clean scale, $\hat{K}$, is smaller than first-best scale (i.e., $\hat{K} \leq K_C^{FB}$), with equality when $K_C^F = \bar{K}$.

2. If the financing constraints generated by the agency problem are intermediate, i.e., $\bar{K} < K_C^F \leq K_C^{FB}$, the optimal agreement specifies the first-best scale, $\hat{K} = K_C^{FB}$. In this case, it is efficient to increase clean scale up to the first-best level but no further, since scale above and beyond K_C^{FB} would reduce joint surplus. Inducing the entrepreneur to switch technologies solely through an increase in scale would require a production scale exceeds the first-best level K_C^{FB}. It is therefore optimal to partially compensate the entrepreneur through an upfront consumption transfer.

3. If financing constraints are mild, so that $K_C^F > K_C^{FB}$, then financial investors alone would provide funding above and beyond the first-best scale of the clean production technology (note that this case can only occur if $\phi_C > 0$). In this case, the optimal financing agreement with socially responsible investors ensures that the clean production technology is run at the first-best scale, $\hat{K} = K < K_C^F$. To induce the entrepreneur to switch to the clean technology at a lower scale than financial investors would finance, the agreement includes an upfront consumption transfer to the entrepreneur.

The above reasoning shows that, as long as financing constraints are significant (as in cases 1 and 2), the main insights of the baseline analysis continue to hold even under decreasing returns to scale: Socially responsible investors optimally achieve impact by relaxing financing constraints and increasing scale for the clean technology. Only when financing constraints are mild (or absent), socially responsible investors optimally achieve impact by reducing firm investment. Note that this latter case resembles a Coasian solution in a setting without financing constraints. For example, a downstream fishery might pay an upstream factory to reduce production (see Coase, 1960).[37]

4.2 Regulation

Our analysis so far focused on what socially responsible investors can achieve in the absence of regulation (or when regulation is not optimally set). While the design of optimal regulation is beyond the scope of this paper, our analysis reveals that standard policy interventions (e.g., an outright ban of dirty production or a Pigouvian tax) generally cannot restore efficiency in the presence of financing constraints.[38] In fact, in some situations such interventions can lead to worse outcomes than no intervention at all.

Banning the dirty technology. Suppose regulators could simply ban the dirty production technology. With clean production as the only feasible technology, this intervention certainly ensures that the entrepreneur adopts the socially efficient production technology. However, in the context of our baseline model, the scale of production under the clean technology would be strictly lower, equal to K_C^F, compared to the case when dirty production is allowed but socially responsible investors ensure that the entrepreneur adopts the clean production technology with scale $\hat{K}_C(\eta) > K_C^F$. Intuitively, the key ingredient for the additional financing capacity from socially responsible investors is the (credible) threat of dirty production (see Proposition 2). Banning dirty production eliminates this threat, so that socially responsible investors are unwilling to extend financing above and beyond the scale offered by financial investors, K_C^F. Of course, we do not argue that a production ban is socially harmful in all scenarios. In particular, if socially responsible investors do not have enough capital or if their investment mandate does not satisfy Conditions 1 and 2, they are unable to steer the entrepreneur towards the socially

[37] Note that in cases 2 and 3, the agreement would need to explicitly limit the amount of firm investment (and not just specify the technology). Otherwise, the entrepreneur would find it privately optimal to convert the upfront consumption into additional firm investment.

[38] See Tirole (2010) and Hoffmann et al. (2017) for models of optimal Pigouvian taxation in imperfect capital markets with financial constraints.

efficient choice. In this case, production bans can increase total surplus as they may prevent the adoption of the dirty technology by firms that are not disciplined by socially responsible investors.

Pigouvian taxes. Now suppose that the regulator imposes a tax on the social cost generated by the firm's production (e.g., a tax assessed on the firm's carbon emissions), resulting in a total tax of $\phi_\tau K$ for a firm producing with technology τ at scale K. Such a tax makes dirty production financially nonviable. While this prevents dirty production, it reduces or eliminates the threat of dirty production, resulting in similar effects to a production ban. However, total surplus can be even lower than under a production ban, because the firm is taxed also on the clean technology, by an amount $\phi_C K$, lowering the maximum feasible scale of clean production below K_C^F.

5 Conclusion

A key question in today's investment environment is to understand conditions under which socially responsible investors can achieve impact. For example, can investors with social concerns influence firms to tilt their production technologies towards lower carbon emissions? To shed light on this question, this paper develops a parsimonious theoretical framework, based on the interaction of production externalities and corporate financing constraints.

Our analysis uncovers the necessity of a *broad mandate* for socially responsible investors. Given an abundant supply of profit-motivated capital, it is not enough for socially responsible investors to simply internalize the social costs generated by the firms they are invested in. Rather, their concern for social costs must be unconditional—independent of their own investment. This condition generates both normative and positive implications. From a positive perspective, our model implies that if current ESG funds lack such a broad mandate, they do not have impact. From a normative perspective, it states that, if society wants responsible investors to have impact, then their mandate needs to be broad. Moreover, because a broad mandate entails the sacrifice of financial returns, socially responsible funds need to be evaluated according to broader measures, explicitly accounting for real impact rather than focusing solely on financial metrics.

To achieve impact in the most efficient way, it is optimal for socially responsible investors to relax firm financing constraints for clean production, thereby enabling a scale increase of clean technology relative to what financial capital is willing to offer. This generates a complementarity between financial and socially responsible capital: Total surplus is generally highest in an economy in which there is a balance between financial and socially responsible capital.

From a practical investment perspective, our model implies a micro-founded investment criterion for scarce socially responsible capital, the *social profitability index* (SPI), which summarizes the interaction of environmental, social and governance (ESG) aspects. Importantly, in line with the broad mandate, the SPI accounts for social costs that would have occurred in the absence of engagement by socially responsible investors. Accordingly, it can be optimal to invest in firms that generate relatively low social returns (e.g., a firm with significant carbon emissions), provided that the potential increase in social costs, if only financially-driven investors were to invest, is sufficiently large. This contrasts with many common ESG metrics that focus on firms' social status quo.

To highlight these ideas in a transparent fashion, our model abstracts from a number of realistic features which could be analyzed in future work. First, our model considers a static framework, where investment is best interpreted as new (greenfield) investment. In a dynamic setting, a number of additional interesting questions would arise: How to account for dirty legacy assets? How to ensure the timely adoption of novel (and cleaner) production technologies as they arrive over time? Because the adoption of future green technologies may be hard to contract ex ante, a dynamic theory might yield interesting implications on the issue of control.

Second, our model considers the natural benchmark case where socially responsible investors are homogeneous. These results can be extended in a straightforward way if socially responsible investors have the same directional preferences (e.g., to lower carbon emissions), albeit with different intensity. More challenging is the case in which socially responsible investors' objectives conflict or are multi-dimensional (e.g., there is a agreement on the goal of lowering carbon emissions, but disagreement on the social costs imposed by nuclear energy).

Third, while our analysis is motivated by the large rise in the demand for ESG investments, the regulatory landscape is changing simultaneously. Regulatory measures that have been discussed include the taxation of carbon emissions as well as subsidies for investments in clean technology (e.g., subsidized loans for the purchases of electric cars or lower capital requirements for bank loans to clean firms). It would therefore be interesting to understand the conditions under which regulation and impact by socially responsible investors are substitutes or complements.

Finally, we excluded the possibility that firms may interact (e.g., as part of a supply chain or as competitors). Yet it is plausible that the financing of a green technology by one firm may impact other firms (e.g., through cross-firm externalities related to production technologies or by alleviating or worsening financing constraints). It would be interesting to study such spillovers in future work.

A Proofs

Proof of Lemma 1. The Proof of Lemma 1 follows immediately from the proof of Proposition 1 given below. First, set $\gamma^{SR} = 0$ (so that socially responsible investors have the same preferences as financial investors). Second, to obtain the competitive financing arrangement (i.e., the agreement that maximizes the utility of the entrepreneur subject to the investors' participation constraint) one needs to choose the utility level of the entrepreneur u in (A.10) such that $\hat{v}_\tau K^*_\tau(u) - u = 0$.[39]

Proof of Proposition 1. The Proof of Proposition 1 will make use of Lemmas A.1 to A.5.

Lemma A.1 *In any solution to Problem 1, the IR constraint of financial investors,* $pX^F - I^F \geq 0$ *must bind,*

$$pX^F - I^F = 0. \tag{A.1}$$

Proof: The proof is by contradiction. Suppose there was an optimal contract for which $pX^F - I^F > 0$. Then, one could increase X^{SR} while lowering X^F by the same amount (until (A.1) holds). This perturbation strictly increases the objective function of socially responsible investors in (8), satisfies by construction the IR constraint of financial investors, whereas all other constraints are unaffected since $X = X^{SR} + X^F$ is unchanged. Hence, we found a feasible contract that increases the utility of socially responsible investors, which contradicts that the original contract was optimal. ■

Lemma A.2 *There exists an optimal financing arrangement with* $I^F = X^F = 0$.

Proof: Take an optimal contract $\left(I^F, I^{SR}, X^{SR}, X^F, K, c, \tau\right)$ with $I^F \neq 0$. Now consider the following "tilde" perturbation of the contract (leaving K, c and τ unchanged). Set $\tilde{X}^F$ and $\tilde{I}^F$ to 0 and set $\tilde{I}^{SR} = I^{SR} + I^F$ and $\tilde{X}^{SR} = X^{SR} + X^F$. The objective of socially responsible investors in (8) is unaffected since

$$p\tilde{X}^{SR} - \tilde{I}^{SR} - \gamma^{SR}\phi_\tau K = pX^{SR} - I^{SR} + \underbrace{pX^F - I^F}_{0} - \gamma^{SR}\phi_\tau K \tag{A.2}$$

$$= pX^{SR} - I^{SR} - \gamma^{SR}\phi_\tau K, \tag{A.3}$$

where the second line follows from Lemma A.1. All other constraints are unaffected since $\tilde{X}^F + \tilde{X}^{SR} = X^F + X^{SR}$ and $\tilde{I}^F + \tilde{I}^{SR} = I^F + I^{SR}$ ■

[39] Note that $\hat{v}_\tau = \pi_\tau - \gamma^E\phi_\tau$ in the special case when $\gamma^{SR} = 0$.

Lemma A.2 implies that we can phrase Problem 1 in terms of total investment I and total repayment to investors X in order to determine the optimal consumption c, technology τ, and scale K. However, to make the proof most instructive, it is useful to replace X and I as control variables and instead use the expected repayment to investors Ξ and expected utility provided to the entrepreneur u, which satisfy

$$\Xi := pX, \tag{A.4}$$

$$u := \left(pR - k_\tau - \gamma^E \phi_\tau\right) K + I - pX. \tag{A.5}$$

Then, using the definition $\hat{v}_\tau := \pi_\tau - \left(\gamma^E + \gamma^{SR}\right) \phi_\tau \geq v_\tau$, we can write Problem 1 as:

Problem 1*

$$\max_{\tau} \max_{u \geq \eta \bar{U}^E + (1-\eta) \underline{U}^E} \max_{K, \Xi} \hat{v}_\tau K - u \tag{A.6}$$

subject to

$$K \geq 0 \tag{A.7}$$

$$\Xi \leq (pR - \xi) K \tag{IC}$$

$$\Xi \geq -(A + u) + \left(pR - \gamma^E \phi_\tau\right) K \tag{LL}$$

Here, the last constraint (LL) can be interpreted as a limited liability constraint, since it refers to the constraint that upfront consumption is weakly greater than zero (using the aggregate resource constraint in (2)). As the problem formulation suggests, it is useful to sequentially solve the optimization in 3 steps to exploit the fact that Ξ only enters the linear program via the constraints (IC) and (LL), but not the objective (A.6).

As is obvious from Problem 1*, only a technology that delivers positive surplus to investors and the entrepreneur (i.e., $\hat{v}_\tau > 0$) is a relevant candidate for the equilibrium technology.[40] Now consider the inner problem, i.e., for a fixed technology τ with $\hat{v}_\tau > 0$

[40] Note that $\hat{v}_C$ is unambiguously positive whereas $\hat{v}_D$ could be negative or positive depending on

and a fixed utility $u \geq \eta \bar{U}^E + (1-\eta)\underline{U}^E$ we solve for the optimal vector (K, Ξ) as a function of τ and u.

Lemma A.3 *For any τ with $\hat{v}_\tau > 0$ and $u \geq \eta \bar{U}^E + (1-\eta)\underline{U}^E$, the solution to the inner problem, i.e., $\max_{K,\Xi} \hat{v}_\tau K - u$ subject to (A.7), (IC) and (LL), implies maximal scale, i.e.,*

$$K_\tau^*(u) = \frac{A+u}{\xi - \gamma^E \phi_\tau} > 0.\text{[41]} \tag{A.8}$$

The expected payment to investors is:

$$\Xi_\tau(u) = (pR - \xi) K_\tau^*(u). \tag{A.9}$$

Proof: The feasible set for (K, Ξ) as implied by the three constraints (A.7), (IC) and

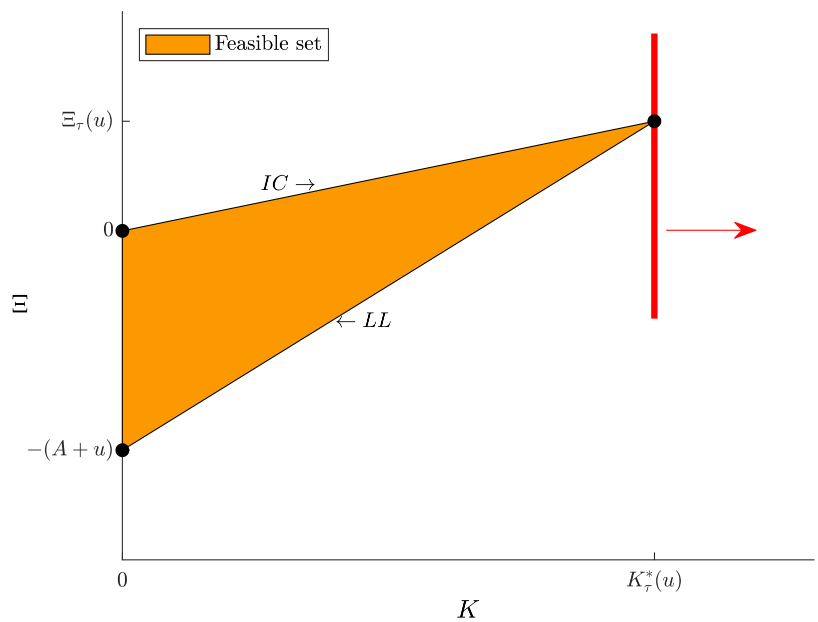

Figure 1. Feasible set of the inner problem: The set of feasible solutions is depicted in orange and forms a polygon. The objective function is represented by the red line and the arrow: The red line is a level set of the objective function of socially responsible investors, and the arrow indicates the direction in which we are optimizing.

whether the sum $\gamma^E + \gamma^{SR}$ is sufficiently close to 1.

[41] Note that this maximum $K_\tau^*(u)$ is distinct from $\hat{K}_\tau(\eta)$ defined in the main text. The former is a function of the utility to the entrepreneur, whereas the latter is a function of the bargaining power of the entrepreneur.

(LL) forms a polygon (see orange region in Figure 1). The upper bound on Ξ in (IC) is an affine function of K through the origin (i.e., linear in K) whereas the lower bound in Equation (LL) is an affine function of K (with negative intercept $-(A+u)$). The slope of the lower bound in Equation (LL) is strictly greater than the slope of the upper bound in Equation (IC) since

$$\begin{aligned}\left(pR-\gamma^E\phi_\tau\right)-(pR-\xi) &= \xi-\gamma^E\phi_\tau \\ &> \pi_\tau-\gamma^E\phi_\tau \\ &> \pi_\tau-\left(\gamma^E+\gamma^{SR}\right)\phi_\tau=\hat{v}_\tau>0,\end{aligned}$$

where the second line follows from the finite scale that is implied by Assumption 1 (i.e., $\xi>\pi_\tau$). Therefore, the intersection of the upper bound (IC) and the lower bound in (LL) defines the maximal feasible scale of K. Choosing the maximal scale $K_\tau^*(u)$ is optimal, since for any given τ with $\hat{v}_\tau>0$ and any fixed $u\geq\eta\bar{U}^E+(1-\eta)\underline{U}^E$, the objective function $\hat{v}_\tau K-u$ is strictly increasing in K and independent of Ξ. The expression for $K_\tau^*(u)$ in Equation (A.8) is obtained from $(pR-\xi)K=-(A+u)+\left(pR-\gamma^E\phi_\tau\right)K$. ■

Given the solution to the inner problem, $(K_\tau^*(u),\Xi_\tau(u))$, we now turn to the optimal choice of u which maximizes $\hat{v}_\tau K_\tau^*(u)-u$ subject to $u\geq\eta\bar{U}^E+(1-\eta)\underline{U}^E$.

Lemma A.4 *In any solution to Problem 1*, the entrepreneur obtains her reservation utility from the bargaining game* $u=\eta\bar{U}^E+(1-\eta)\underline{U}^E$.

Proof: It suffices to show that the objective is strictly decreasing in u. Using $K_\tau^*(u)=\frac{A+u}{\xi-\gamma^E\phi_\tau}$ and $\hat{v}_\tau=\pi_\tau-\left(\gamma^E+\gamma^{SR}\right)\phi_\tau$, we obtain that:

$$\hat{v}_\tau K_\tau^*(u)-u=\frac{\hat{v}_\tau}{\xi-\gamma^E\phi_\tau}A-\frac{\xi+\gamma^{SR}\phi_\tau-\pi_\tau}{\xi-\gamma^E\phi_\tau}u \tag{A.10}$$

Since $\xi>\pi_\tau$ and $\xi>\gamma^E\phi_\tau$ (both by Assumption 1), both the numerator and the denominator of $\frac{\xi+\gamma^{SR}\phi_\tau-\pi_\tau}{\xi-\gamma^E\phi_\tau}$ are positive, so that Equation (A.10) is strictly decreasing in

u. ■

Given that the utility to the entrepreneur is given by $u = \eta \bar{U}^E + (1-\eta)\underline{U}^E$ we can now define the (relevant) scale as a function of the bargaining power η, i.e.,

$$\hat{K}_\tau(\eta) := K^*_\tau\left(\eta \bar{U}^E + (1-\eta)\underline{U}^E\right) \tag{A.11}$$

$$= \eta K^{\max}_\tau + (1-\eta) K^{\min}_\tau \tag{A.12}$$

where the second line follows from linearity of K^*_τ in u and the following definitions:

$$K^{\min}_\tau := K^*_\tau\left(\underline{U}^E\right) = \frac{\xi - \gamma^E \phi_{\tau_F}}{\xi - \gamma^E \phi_\tau} K^F_{\tau_F} \tag{A.13}$$

$$K^{\max}_\tau := K^*_\tau\left(\bar{U}^E\right) = \frac{\xi - \gamma^E \phi_{\tau_F} - \hat{v}_{\tau_F}}{\xi - \gamma^E \phi_\tau - \hat{v}_\tau} K^F_{\tau_F} \tag{A.14}$$

Here, $K^{\min}_\tau$ can be interpreted as the minimum required scale to induce the entrepreneur switch to technology τ. It is obtained by evaluating $K^*_\tau(u)$ (see A.8) at the entrepreneur's outside option $\underline{U}^E = (\pi_{\tau_F} - \gamma^E \phi_{\tau_F}) K^F_{\tau_F}$. The maximum scale that socially responsible investors are willing to offer for technology τ, $K^{\max}_\tau$, is determined by their break-even condition as given by their outside option $\underline{U}^{SR} = -\gamma^{SR} \phi_{\tau_F} K^F_{\tau_F}$.

Using the definition of (A.12), the payoff to socially responsible investors *for a given* τ (at the optimal scale) is given by:

$$U^{SR} = \hat{v}_\tau \hat{K}_\tau(\eta) - \left[\eta \bar{U}^E + (1-\eta)\underline{U}^E\right]. \tag{A.15}$$

We now turn to the final step, i.e., the optimal technology choice.

Lemma A.5 *The optimal technology choice is independent of the bargaining power and given by*

$$\hat{\tau} = \arg\max_\tau \hat{v}_\tau \hat{K}_\tau(\eta). \tag{A.16}$$

Proof: In the relevant case where $\hat{v}_D > 0$, we need to compare payoffs in (A.15).

The clean technology is chosen if and only if $\hat{v}_C \hat{K}_C(\eta) > \hat{v}_D \hat{K}_D(\eta)$, which simplifies to (A.16). If $\hat{v}_D \leq 0$, then A.16 trivially holds as only $\hat{v}_C > 0$. Finally, we need to prove that the technology choice is independent of the bargaining power η (for the relevant case where $\hat{v}_D > 0$). This follows from the fact whenever $\hat{v}_C K_C^{\max} > \hat{v}_D K_D^{\max}$, then also $\hat{v}_C K_C^{\min} > \hat{v}_D K_D^{\min}$ and vice versa. Hence, $\arg\max_\tau \hat{v}_\tau K_\tau^{\max} = \arg\max_\tau \hat{v}_\tau K_\tau^{\min} = \arg\max_\tau \hat{v}_\tau \hat{K}_\tau(\eta)$. ■

Lemmas A.3 to A.5, thus, jointly characterize the solution to Problem 1*, which, in turn, allows us to retrieve the solution to the original Problem 1. That is, the equilibrium technology $\hat{\tau}$ is run at scale $\hat{K}_{\hat{\tau}}(\eta)$. Moreover, since ($LL$) binds, we obtain that $\hat{c} = 0$. Finally, we also obtain intuitive expression for the respective maximum feasible utilities:

$$\bar{U}^E = \underline{U}^E + \hat{v}_{\hat{\tau}} K_{\hat{\tau}}^{\max} - \hat{v}_{\tau_F} K_{\tau_F}^F \tag{A.17}$$

$$\bar{U}^{SR} = \underline{U}^{SR} + \hat{v}_{\hat{\tau}} K_{\hat{\tau}}^{\min} - \hat{v}_{\tau_F} K_{\tau_F}^F \tag{A.18}$$

Proof of Corollary 2. The aggregate resource constraint in (2) then implies that total investment by both investors must satisfy $\hat{I} = \hat{K}_{\hat{\tau}}(\eta)\, k_{\hat{\tau}} - A$, whereas (IC) implies that $\hat{X} = \left(R - \frac{B}{\Delta p}\right) \hat{K}_{\hat{\tau}}(\eta)$. Since any agreement must satisfy $X^F + X^{SR} = \hat{X}$ and $I^F + I^{SR} = \hat{I}$, we can trace out all possible agreements using the fact that financial investors break even (Lemma A.1), meaning that $pX^F - I^F = 0$ and $X^F \in [0, R]$.

Proof of Proposition 2. See discussion in main text.

Proof of Proposition 3. See discussion in main text.

Proof of Proposition 4. The social profitability index is defined as:

$$\text{SPI} = \frac{\Delta U^{SR}}{I^{SR}} \tag{A.19}$$

Using Proposition 1, we obtain that the minimum investment that is sufficient to induce a change in production technology is given by

$$I_{\min}^{SR} = (\xi - \pi_C)\,\hat{K} - A. \tag{A.20}$$

The corresponding (maximal) SPI is, hence, given by

$$\mathrm{SPI}_{\max} = \gamma^{SR} \frac{\Delta\phi}{\Delta\pi - \frac{\gamma^E}{\xi}\left(\Delta\phi\,(\xi - \pi_C) + \Delta\pi\phi_C\right)} - 1 \tag{A.21}$$

which is increasing in $\Delta\phi$, ξ, and γ^E and decreasing in $\Delta\pi$.

References

Aghion, Philippe, Roland Bénabou, Ralf Martin, and Alexandra Roulet, "Environmental Preferences and Technological Choices: Is Market Competition Clean or Dirty?," Technical Report 2019.

Albuquerque, Rui, Yrjö Koskinen, and Chendi Zhang, "Corporate social responsibility and firm risk: Theory and empirical evidence," *Management Science*, 2019, *65* (10), 4451–4469.

Angelis, Tiziano De, Peter Tankov, and Olivier David Zerbib, "Environmental Impact Investing," 2020. Working Paper, Tilburg University.

Baker, Malcolm, Daniel Bergstresser, George Serafeim, and Jeffrey Wurgler, "Financing the response to climate change: The pricing and ownership of US green bonds," 2018. Working paper, Harvard Business School.

Barber, Brad M, Adair Morse, and Ayako Yasuda, "Impact investing," 2018. Working paper.

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Besley, Timothy and Maitreesh Ghatak, "Competition and incentives with motivated agents," *American economic review*, 2005, *95* (3), 616–636.

Bonnefon, Jean-Francois, Augustin Landier, Parinitha Sastry, and David Thesmar, "Do Investors Care About Corporate Externalities? Experimental Evidence," *Working Paper, TSE, MIT and HEC*, 2019.

Broccardo, Eleonora, Oliver Hart, and Luigi Zingales, "Exit vs. Voice," 2020. Working Paper, Harvard University and University of Chicago.

Chava, Sudheer, "Environmental externalities and cost of capital," *Management Science*, 2014, *60* (9), 2223–2247.

Chowdhry, Bhagwan, Shaun William Davies, and Brian Waters, "Investing for impact," *Review of Financial Studies*, 2018, *32* (3), 864–904.

Coase, R. H., "The Problem of Social Cost," *The Journal of Law and Economics*, 1960, *3*, 1–44.

Davies, Shaun William and Edward Dickersin Van Wesep, "The unintended consequences of divestment," *Journal of Financial Economics*, 2018, *128* (3), 558–575.

Dewatripont, Mathias and Jean Tirole, "The Morality of Markets and the Nature of Competition," 2020. Working Paper, University of Brussels and Toulouse School of Economics.

Dimson, Elroy, Oğuzhan Karakaş, and Xi Li, "Coordinated engagements," *Working Paper, Cambridge University*, 2019.

Friedman, Milton, "A Friedman doctrine: The social responsibility of business is to increase its profits," 1970. The New York Times Magazine, September 13, 1970.

Gollier, Christian and Sébastien Pouget, "The "Washing Machine": Investment Strategies and Corporate Behavior with Socially Responsible Investors," 2014. Working Paper, Toulouse School of Economics.

Hart, Oliver and John Moore, "Default and Renegotiation: A Dynamic Model of Debt," *The Quarterly Journal of Economics*, 1998, *113* (1), 1–41.

_ **and Luigi Zingales**, "Companies Should Maximize Shareholder Welfare Not Market Value," *Journal of Law, Finance, and Accounting*, 2017, *2* (2), 247–275.

Hartzmark, Samuel M and Abigail B Sussman, "Do investors value sustainability? A natural experiment examining ranking and fund flows," *Journal of Finance*, 2019, *74* (6), 2789–2837.

Heinkel, Robert, Alan Kraus, and Josef Zechner, "The effect of green investment on corporate behavior," *Journal of financial and quantitative analysis*, 2001, *36* (4), 431–449.

Hoffmann, Florian, Roman Inderst, and Ulf Moslener, "Taxing Externalities Under Financing Constraints," *The Economic Journal*, 2017, *127* (606), 2478–2503.

Holmström, Bengt and Jean Tirole, "Financial intermediation, loanable funds, and the real sector," *Quarterly Journal of Economics*, 1997, *112* (3), 663–691.

Hong, Harrison and Marcin Kacperczyk, "The price of sin: The effects of social norms on markets," *Journal of Financial Economics*, 2009, *93* (1), 15–36.

Landier, Augustin and Stefano Lovo, "ESG Investing: How to Optimize Impact," 2020. Working Paper, HEC.

Moisson, Paul-Henri, "Ethics and Impact Investment," 2020. Working Paper, TSE.

Morgan, John and Justin Tumlinson, "Corporate provision of public goods," *Management Science*, 2019, *65* (10), 4489–4504.

Nauman, Billy, "Global bank agree on framework to promote green shipping," *Financial Times*, 2019, *June 18.*

Pastor, Lubos, Robert F. Stambaugh, and Lucian Taylor, "Sustainable Investing in Equilibrium," *Journal of Financial Economics*, 2020.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski, "Responsible Investing: The ESG-Efficient Frontier," 2019. Working Paper, Copenhagen Business School and AQR.

Riedl, Arno and Paul Smeets, "Why do investors hold socially responsible mutual funds?," *Journal of Finance*, 2017, *72* (6), 2505–2550.

Roth, Benjamin N., "Impact Investing: A Theory of Financing Social Enterprises," 2019. Working Paper, Harvard Business School.

Samuelson, Paul A., "The Pure Theory of Public Expenditure," *The Review of Economics and Statistics*, 1954, *36* (4), 387–389.

Tirole, Jean, *The theory of corporate finance*, Princeton University Press, 2006.

_ , "From Pigou to extended liability: on the optimal taxation of externalities under imperfect financial markets," *Review of Economic Studies*, 2010, *77* (2), 697–729.

_ , "Some Political Economy of Global Warming," *Economics of Energy & Environmental Policy*, 2012, *1* (1), 121–132.

Zerbib, Olivier David, "The effect of pro-environmental preferences on bond prices: Evidence from green bonds," *Journal of Banking and Finance*, 2019, *98*, 39 – 60.

_ , "A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from green investing and sin stock exclusion," 2020. Working Paper, Tilburg University.