Socially Optimal Eligibility Criteria for ESG Funds*

Roman Inderst† Marcus Opp‡

May 2023

Abstract

Our paper analyzes whether and how a planner should design a taxonomy for sustainable investment products in the presence of traditional tools for environmental regulation. Barring policy failures for environmental regulation, such a taxonomy can only achieve positive real effects if financial constraints prevent firms from supplying the socially optimal quantity of output. Then, the planner effectively exploits warm-glow sustainability preferences by retail investors to subsidize firms' sustainability investments, thereby relaxing financial constraints. We characterize how the optimal cut-off determining eligibility for a ESG fund interacts with environmental regulation, social costs of externalities, shifts in social preferences, and social norms.

Keywords: Sustainability; ESG; green financing; EU taxonomy; ESG funds; European Green Deal.

*For helpful comments and suggestions, we thank Ben Hébert, Pablo Ruiz-Verdu, and Per Strömberg as well as participants at Stanford, ASU, Stockholm School of Economics and Universidad Carlos III de Madrid. This project has received funding from the Marianne and Marcus Wallenberg (MMW) foundation.

†Goethe Universität Frankfurt; inderst@finance.uni-frankfurt.de

‡Stockholm School of Economics, CEPR, ECGI, and FTG; e-mail: marcus.opp@hhs.se.

1 Introduction

Over the past two decades, the demand for sustainable investing by retail investors has grown exponentially. At the same time, concerns about vague standards (Berg et al., 2022) and greenwashing (de Freitas Netto et al., 2020) have called into question the real impact of these investments. In response to these concerns, the EU created a uniform and legally binding classification system that defines sustainable economic activities (EU, 2020). The resulting taxonomy is supposed to help channel funding by retail investors towards more sustainable activities, and, hence, incentivize companies to become more climate-friendly in the context of the European Green Deal.

Can such a classification system help achieve the stated objectives and if so, which firms should be classified as "sustainable"? Is there a role for "sustainable finance" even in the absence of policy failures for environmental regulation? Is it even possible that such regulation backfires?

Our theoretical analysis provides the following high-level insights. The introduction of a sustainable investment category is beneficial only if either of the following two conditions are satisfied. First, it is beneficial if environmental regulation is too lax, as e.g., argued by Tirole (2012). In this case, financial regulation is an imperfect substitute for the policy failure to efficiently use more direct policy tools such as production standards or taxes on externalities. Second, it is beneficial even under optimal environmental regulation if the socially efficient aggregate output is prevented by outside financing constraints. In this case, the taxonomy raises welfare by "activating" sustainable investors to subsidize clean production. Then, as preferences for sustainable investments become stronger over time, standards for sustainability should be optimally increased in conjunction. They should also be raised when the thereby addressed externality has higher social costs. In contrast, if neither of these two conditions is satisfied, the introduction of sustainable standards is strictly welfare reducing.

We derive these results in a parsimonious modeling framework with the following features. We assume that a subset of investors depart from the conventional, profit-motivated investor in that these investors experience a (heterogeneous) warm glow boost upon investing in sustainable firms. This warm glow reflects non-consequentialist preferences rather than a concern about the investment's ultimate impact, consistent with experimental evidence by Bonnefon et al. (2019).[1] Investors invest in firms that choose the scale and sustainability of their production (costly abatement of externalities) subject to minimum production standards (or taxes on externalities) and outside financing constraints à la Holmström and Tirole (1997). The relevance of such financial constraints for abatement investments has been empirically documented by Bartram et al. (2022) and Xu and Kim (2022). If a firm's sustainability choice exceeds a more stringent threshold, it is considered taxonomy-conform. The design of this threshold is the key choice variable of the planner. To focus on real effects in our welfare analysis, we treat the resulting warm-glow as a purely decisional utility, so that there are no trivial benefits from the introduction of a sustainable investment classification. Our model is closed on the product market side in that the aggregate scale of production by all firms (and, hence output) determines the price in the product market, and, hence consumer welfare.

We now characterize key features of the equilibrium to illustrate the planner's trade-off. As we focus on the realistic case where the supply of sustainable capital is not large enough to meet the funding requirements by all firms in the economy, equilibrium investment and, hence aggregate output, is solely pinned down by environmental policies (minimum production standards or environmental taxes). In contrast, the taxonomy for sustainable investing only affects the *share* of firms that choose to meet the taxonomy threshold. The regulator now faces the following trade-off. By increasing the sustainability threshold, each abiding firm causes fewer externalities, but also requires a higher financing subsidy to

[1]The regulation is not intended to target large impact investors, as studied by Oehmke and Opp (2019), who can make informed decisions without the need to rely on a regulatory classification.

offset the higher production cost to meet the standard. Given a downward sloping supply of sustainable capital, this leads to a smaller share of sustainable firms in equilibrium.

Based on this trade-off we can now explain the intuition for the first result. If the minimum standard or environmental taxes are too low, aggregate output in the economy is too large from a welfare perspective due to the presence of negative externalities. The introduction of a taxonomy for sustainable investing is now beneficial as it incentivizes a subset of firms to reduce their externalities (even though it does not affect aggregate output). As preferences for sustainable investment become stronger this favors on the margin to increase the sustainability standard. Given the time trend of preferences for sustainability in recent years, our model, thus predicts that optimal sustainability standards should gradually increase over time. Moreover, if households' preferences depend on endogenous social norms, i.e., increase with the fraction of other investors that invest sustainably, the planner exploits the feedback-effect by designing the optimal classification less stringent.

However, is there a role for a sustainable investments category if more direct policy instruments, such as a minimum standard or a tax on externality, are optimally chosen? We show that the introduction of a sustainable investment category is only beneficial if financial constraints prohibit aggregate production at the efficient scale. In this case, stricter environmental regulation in isolation would exacerbate underproduction. By activating a subsidy from "sustainably-oriented" investors, the planner can now mitigate this trade-off between imposing a higher minimum standard and, at the same time, inefficiently shrinking the economy. The joint optimization allows the planner to increase the weighted average sustainability of all firms while increasing aggregate output.

When financial frictions do not prevent the economy from running at the efficient scale, activating sustainable finance is not just unnecessary, it strictly backfires. The reason why it backfires is that it induces some firms to produce at an inefficiently high environmental standard, which is too costly from a welfare perspective. When environmental policy sets a minimum standard (e.g., a maximum level of emissions per output), as is often the case in practice (e.g., catalytic converter), the additional instrument of a sustainability classification is only beneficial when the economy is *sufficiently* financially constrained.

We now provide some additional background for our modeling of households' warm-glow preferences following the work of Andreoni (1990). We assume that households have non-consequentialist preferences, deriving a warm-glow utility from every Euro invested in a "taxonomy-conform" sustainability (ESG) fund. We motivate these preferences based on a fast-growing empirical literature on preferences for sustainable investing, e.g., Riedl and Smeets (2017), Bonnefon et al. (2019) or Humphrey et al. (2021). This survey-based and experimental evidence suggests that investors are overwhelmingly "value aligned" rather than consequentialist (driven by the impact of their investments).[2]

The recent theoretical literature on sustainable investing, see e.g., Oehmke and Opp (2019), Landier and Lovo (2020) as well as Green and Roth (2021), has highlighted how such "value aligned" preferences or "narrow" investment mandates fail to generate real impact. Consistent with the findings of this literature, the non-regulated market outcome in our setting will trivially lead to the lowest possible standard, thereby exhausting the available sustainable capital, but to no real effect. What differentiates our paper from this literature is that our paper features a impact-seeking planner that is able to exploit investors' non-consequentialist preferences for sustainability to direct subsidized funding to firms that make sustainability investments. The assumption that households derive their warm glow through the respective classification can be motivated by retail investors' lack of information or excessive complexity associated with a more detailed assessment. Since the warm-glow effect resulting from a purchase of "moral satisfaction", see Kahneman and Knetsch (1992), remains relatively unaffected by the actually achieved outcome, we model

[2]The specification of such preferences also receives support by a large literature in environmental and resource economics that elicits preferences. For instance, elicited willingness-to-pay frequently fails a so-called "scope test" or "adding-up test": Subjects' willingness-to-pay is relatively insensitive to the actual impact of the respective scenario change, e.g., the number of animals saved, see e.g., Boyle et al. (1994) or Desvousges et al. (2012) for prominent studies on these effects.

it as a purely decisional utility that does not (directly) enter the social planner's objective, consistent with Broccardo et al. (2022).

Our paper builds on a rapidly growing literature on the theory of socially responsible investing. This literature consists mainly of two strands: *exclusion* and *impact investing*, cf. Oehmke and Opp (2019) for a detailed overview. Since we aim to model investor demand for sustainable finance products by small retail investors, our model does not feature large activist investors, which are studied in Chowdhry et al. (2018), Oehmke and Opp (2019), Biais and Landier (2022) and Gupta et al. (2022).[3] Instead, our paper shares the mechanism of the exclusion literature following the pioneering paper by Heinkel et al. (2001). In equilibrium, sustainably-oriented investors exclude unsustainable firms in their portfolio, which generates larger equilibrium funding costs for non-sustainable vs. sustainable firms and provides an incentive for firms to reform.[4]

None of these contributions has addressed the two key questions of our analysis, i.e., whether such subsidized financing is socially beneficial in view of alternative, more direct environmental policy instruments and how to optimally design a sustainable investment classification.[5] The latter issue taps into a large theoretical literature on optimal certification (see Bizzotto and Harstad (2023) for a recent contribution and a detailed survey). This literature, however, typically considers the perspective of a profit-maximizing certifier and the design of labels for products or services when purchasers have limited information about quality.

[3]We share with Oehmke and Opp (2019) the modeling of financial constraints based on Holmström and Tirole (1997). It extends Oehmke and Opp (2019) in two ways. First, we incorporate retail investors with empirically relevant warm-glow preferences in a tractable way. Second, firm payoffs are determined in a product market equilibrium, see also Inderst and Heider (2022), rather than exogenously specified.

[4]More recently, the effects of exclusion, underweighting or divestment are studied in Landier and Lovo (2020), Davies and Van Wesep (2018), Pastor et al. (2021), Pedersen et al. (2021) and Edmans et al. (2022). In particular, we share with Landier and Lovo (2020) that preferences are risk-neutral, which ensures tractability, i.e., there is no effect on risk-premia which result from imperfect risk-sharing.

[5]Biais and Landier (2022), Döttling and Rola-Janicka (2022), and Oehmke and Opp (2022) have analyzed the interaction of environmental regulation with the financial sector when environmental regulation is subject to a commitment problem. There is no commitment problem in our paper.

This paper is organized as follows. In Section 2 we introduce the modeling framework. Section 3 analyzes the market equilibrium, while Section 4 provides conditions under which the introduction of a sustainable investment category is optimal and, if so, which threshold should be optimally chosen. Section 5 embeds this analysis into a framework of endogenous environmental policy, considering a minimum standard or a Pigouvian tax. Section 6 introduces social preferences and Section 7 concludes. All proofs are collected in an appendix.

2 Model framework

We consider a single-sector economy and analyze the technology choice and investment decisions of firms that compete in a product market. Firms are endowed with internal funds that are essential to mitigate a moral-hazard problem vis-à-vis outside investors (households). A subset of these households cares not only about cash flows resulting from these investments in firms but also about their investments' sustainability classification. A regulator designs a taxonomy that determines which firms are certified as "sustainable."

We next describe the model ingredients and objective functions of firms, households and the regulator. In Section 3, we characterize the equilibrium, where households optimally allocate their capital and firms optimally decide which technology to use and how much outside financing to raise. Our main analysis characterizes the optimal sustainability classification by the regulator (Section 4) as well as optimal environmental regulation (Section 5).

Firms. The economy consists of a unit mass of firms indexed by i, each endowed with initial funds of size A (which, thus, also corresponds to the aggregate size of internal funds). Firms compete in a homogeneous product market: To isolate the role of sustainable investors, as opposed to sustainable consumers, we assume that sustainable firms cannot obtain a price premium on the product market.[6] That is, given individual firm output q_i, each firm reaps a market price of $P(q)$ where $q = \int_0^1 q_i di$ denotes aggregate output and P satisfies the usual conditions, i.e., $P' < 0$ and $\lim_{q\to\infty} P(q) = 0$.

The production technology choice features a trade-off between profitability and sustainability $\theta_i \leq \theta^{\max}$, which arises, for example, from the costly installation of air filters and θ_i governing their quality. Concretely, we assume that production generates negative externality of $\rho(\theta^{\max} - \theta_i)$ per unit of output while investment costs per unit of output are given by $c(\theta_i)$, where c is a strictly increasing and convex function with $c'(0) = 0$ and $\lim_{\theta\to\theta^{\max}} c'(\theta) = \infty$.

The resulting total cost of production, $c(\theta_i) q_i$, needs to be financed by a combination of internal funds $A_i \leq A$ and external funds from household investors. External financing is subject to financing frictions adopted from the workhorse model of Holmström and Tirole (1997). Specifically, the sale of output is only successful with probability one if the owner-manager exerts unobservable effort. If she shirks, she obtains a per-unit private benefit $B > 0$, but with probability $\Delta p > 0$ no sale occurs. As is standard in the literature, we assume that the agency rent $\frac{B}{\Delta p}$ is low enough so that shirking is off equilibrium (see exact condition in Appendix-Lemma A.1), but large enough to matter $\frac{B}{\Delta p} > \rho(\theta^{\max} - \hat{\theta})$ where $\hat{\theta}$ solves $c'(\theta) = \rho$.

Environmental regulation. One objective of our analysis is to investigate the relevance of sustainable finance and ESG classification in the presence of alternative policy instruments. We first consider an emission cap of $\theta^{\max} - \theta_m$ or, equivalently, a minimum sustainability standard θ_m. For instance, this may require mandatory air filters of a certain quality. Likewise, it can require investments in protection against health hazards for workers or undue harm on animals. In an extension we also consider a possible tax of τ

[6] See Hakenes and Schliephake (2021), Broccardo et al. (2022), and Piccolo et al. (2022) for models that consider both socially responsible consumption and socially responsible investment.

per unit of social cost and show how our results extend when this instrument is used either alone or in combination with a minimum sustainability standard.[7]

Sustainable finance and ESG classification. Funds for investment are provided by atomistic, risk-neutral households. All of these households have access to a storage technology that offers a fixed return r_0 (which is normalized to zero). A subset of these households corresponds to traditional financial investors that only care about cash flows. We presume that the capital owned by these financial investors is in ample supply. In addition, as motivated by the recent literature on ESG funds, we suppose that a subset of households derives an additional "warm-glow" payoff from owning sustainable firms. Since sustainable capital is still small relative to purely financial capital, we presume that the capital held by households with sustainability preferences, denoted by K, is insufficient to finance all firms in the economy.[8]

Firms are classified as sustainable, or taxonomy-conform, if they exceed the more stringent sustainability standard $\theta_s > \theta_m$. In reduced form, the dependence of preferences on the taxonomy can be motivated by the idea that certification is necessary for investors to trust the sustainability label and, hence, experience the warm glow.[9] The choice of the threshold for taxonomy compliance, θ_s, will represent our main policy variable of interest.

We stipulate that if household j makes an investment in such a firm, this generates a warm-glow effect w_j per unit of investment, which, thus, acts akin to an additional per-unit return. The demand for sustainable investments (and, hence, the extra-willingness to provide financing to sustainable firms) is assumed to be heterogeneous across households.

[7] We note that while (Pigouvian) taxes are a prominent tool for economists, they are not widely used in practice. In fact, outside environmental regulation they seem to be hardly used at all, while there are also only few cases where taxes on pollutants, such as CO_2 or SO_x, are used to govern emissions. One such case is the European Union's Emissions Trading System (EU ETS), though the primary policy instrument are emission rights and thus quantity. While fuel taxes are more common in various jurisdictions, their prominence precedes environmental concerns. These observations justify that we initially focus on a minimum sustainability standard as a primary policy instrument.

[8] See the proof of Proposition 2 for the exact condition.

[9] Put differently, due to concerns about greenwashing (see e.g., the DWS scandal in 2022) households would not trust self-proclaimed ESG funds.

The degree of heterogeneity is captured by the density function $f(w)$ over $[0, \overline{w}]$ with associated CDF $F(w)$.

We note that under the considered warm-glow preferences, private provision of a sustainable investment classification fails, *even* when investment intermediaries could commit to a threshold θ_s. Intuitively, so as to maximize (intermediated) volume, an intermediary would set θ_s as close as possible to θ_m, thereby minimizing firms' incremental costs $c(\theta_s) - c(\theta_m)$ per unit of output.

While the just described model setup formally does not require an intermediary, one should think about our setup in light of the following competitive intermediation arrangement. If a firm obtains the "sustainability" certification by exceeding a more stringent sustainability standard $\theta_s > \theta_m$, it is eligible for investment by an ESG fund. Under the European Union's Sustainable Finance Disclosure Regulation (SFDR), investment products have to disclose whether the investment is aligned with the EU Taxonomy or not.[10] Upon investing in an ESG fund, households derive a warm glow utility (relative to a "normal" investment fund that offers the return $r_0 = 0$).

Social planner's objective function. We envisage an objective function of the social planner that encompasses only the consumption and production side, but not investors' perception of a warm glow. We thus conceive of w_j as merely a decisional utility (as in Broccardo et al., 2022). As a result, the introduction of an ESG classification does not have an obvious, direct welfare effect, and we can focus on the real effects. Given production choices θ_i and q_i, welfare, thus comprises first, gross consumer welfare; second, investment costs; and third, the externality, where we stipulate constant marginal social costs $\rho > 0$:

$$\Omega = \int_0^q P(q)dq - \int q_i c(\theta_i)\, di - \int \rho q_i [(\theta^{\max} - \theta_i)] di, \tag{1}$$

[10] According to the July 2022 Article 9 guidance by the European Commission, a ESG fund is prohibited from investing in any firm that is not considered sustainable by the EU taxonomy.

which already uses the result that shirking is off equilibrium (see Proof of Appendix-Lemma A.1).

Throughout the analysis we restrict consideration to cases where, first, it is profitable for firms to operate given some exogenously set minimum standard θ_m (it will always be profitable under the socially optimal θ_m). Second, we presume that the economic activity is socially valuable, $\Omega > 0$, at least under optimal regulation. It is, thus, necessary that the consumer surplus on the initial unit exceeds the marginal social cost, the sum of the marginal private investment costs $c(\theta)$ and externalities $\theta^{\max} - \theta$,

$$P(0) > \min_{\theta}[c(\theta) + \rho(\theta^{\max} - \theta)]. \tag{A1}$$

Timing. We consider the following logical sequence of events. At $t = 0$ the social planner chooses the minimum production standard θ_m (or, in our extension, a tax on externalities) as well as the threshold for sustainable investing θ_s. At $t = 1$, given this regulatory environment, firms then simultaneously choose their sustainability level θ_i, output q_i as well as their external financing $c(\theta_i)\, q_i - A_i$. Households optimally allocate funds to firms and the storage technology. At $t = 2$, managers exert effort and sell output q_i at price $P(q)$.

3 Market equilibrium

We now characterize the market equilibrium and investment decisions of firms for a given regulatory environment (θ_m, θ_s). In a first step, this requires us to determine the effect of financial constraints for the production capacity of each firm.

3.1 Financial constraints

In our setup, financial constraints arise from the moral hazard problem. Let D_i denote the promised repayment to household investors, then incentive compatibility of effort requires that the owner's payoff under effort, $q_i P(q) - D_i$, exceeds the expected payoff under shirking, $(1-\Delta p)[q_i P(q) - D_i] + Bq_i$, which implies the following upper bound on the pledgable payoff to outside investors,

$$D_i \leq q_i \left[P(q) - \frac{B}{\Delta p} \right]. \tag{IC}$$

where $\frac{B}{\Delta p}$ measures the (minimal) agency rent per unit of output accruing to the manager.

The investors' participation constraint (IR) requires a sufficiently high repayment D_i that ensures that investors earn at least the required return $r(\theta_i)$ on their investment of size $c(\theta_i) q_i - A_i$, where $A_i \leq A$ denotes the insider's coinvestment:

$$D_i \geq (c(\theta_i) q_i - A_i)(1 + r(\theta_i)). \tag{IR}$$

The required return $r(\theta_i)$ will be endogenized below as a function of a firm's sustainability classification which, in turn, is determined by the firm's sustainability choice θ_i. Combining the manager's IC constraint and the investors' IR constraint we obtain an upper bound on production by firm i

$$q_i \leq k_i A_i, \tag{2}$$

where the production capacity multiplier k_i satisfies

$$k_i = \frac{1 + r(\theta_i)}{\frac{B}{\Delta p} - [P(q) - c(\theta_i)(1 + r(\theta_i)]}. \tag{3}$$

In equilibrium, the quantity q will always adjust, so that $P(q) < B/\Delta p + c(\theta_i)(1 + r(\theta_i))$ and the agency constraints imply a finite multiplier k_i for each individual firm (see proof of Proposition 1).[11] However, aggregate economic output is only impacted by financial constraints if firms cannot *jointly* finance the production of zero-profit output $\bar{q}$ solving

$$P(\bar{q}) = c(\theta_m).^{12} \tag{4}$$

If aggregate output is constrained by financial frictions, $q < \bar{q}$, firms earn scarcity rents and, hence, find it optimal to strictly lever up as much as possible by coinvesting all their wealth $A_i = A$. Using binding (IR) then yields the following payoff of the entrepreneur

$$U_i = q_i P(q) - (c(\theta_i) q_i - A)(1 + r(\theta_i)). \tag{5}$$

Interestingly, if economic output is not limited by financial constraints, $q = \bar{q}$, (5) still applies. In this case, firms are indifferent between coinvesting all of their own funds $A_i = A$ and using the storage technology, so that $k_i = \frac{1}{B/\Delta p}$ and $U_i = A$.

The objective in (5) reveals that firms operate a constant-returns-to-scale (CRS) technology and that individual firms treat the product price as a given. In aggregate, the industry exhibits decreasing returns to scale as the inverse demand is downward sloping. Both the individual production capacity as implied by (3) and an individual firm's profitability, see (5), are a decreasing function of aggregate output q via $P(q)$.

3.2 Equilibrium conditions

We first characterize optimal firm behavior and the resulting equilibrium both in the product market and the financial market given exogenous regulation parameters θ_m and

[11] In particular, if the willingness to pay on the initial unit is sufficiently high, i.e., $P(0) > \frac{B}{\Delta p} + c(\theta_m)$, the economy can produce quantity $q = P^{-1}\left(\frac{B}{\Delta p} + c(\theta_m)\right)$ even if firms have no inside funds, $A = 0$.

[12] Intuitively, Proposition 1 shows that if firm's individual assets are sufficiently high then aggregate output is determined by a zero-profit condition, which pins down aggregate output even in the absence of financial constraints.

$\theta_s > \theta_m$. We characterize the optimal planner problem in the subsequent section. In equilibrium, firms and households make optimal decisions.

Supply of (sustainable) financing. Since households only care about whether a firm is certified as sustainable or not, determining eligibility for an ESG fund, we only need to distinguish between two rates of return. Formally, the return function $r(\theta)$ is a step function. For $\theta < \theta_s$, the required rate of return for non-sustainable investments is equal to that obtained from the alternative investment opportunity, in particular, $r(\theta_m) = r_0 = 0$. The required rate of return for sustainable investment, $\theta \geq \theta_s$, is given by $r(\theta_s)$, which will be endogenized below. "*Investor optimality*" requires that a household with warm-glow preferences $w_j \geq 0$ chooses to invest in a sustainable firm if and only if

$$w_j \geq \Delta r := r(\theta_m) - r(\theta_s), \tag{6}$$

where $\Delta r > 0$ measures the return subsidy for sustainable investments by firms. Given this simple cut-off strategy by individual households, the fraction of households that invests sustainably, i.e., all households with $w_j \geq \Delta r$, is $1 - F(\Delta r)$. The total supply of sustainable investment funds, $K[1 - F(\Delta r)] < K$, is, thus, strictly decreasing in the return subsidy Δr for $\Delta r \in (0, \overline{w})$. Intuitively, a smaller return sacrifice Δr attracts a larger set of households.

Demand for (sustainable) financing. We now turn to our second equilibrium condition, "*firm optimality*" regarding the choice of the sustainability level θ_i. An individual firm takes as given both the market conditions of outside financing, $r(\theta_m) = r_0$ and $r(\theta_s)$, as well as the product price $P(q)$. Given the sustainability subsidy Δr, the firm chooses θ_i and q_i to maximize U_i in (5) subject to the outside financing (incentive) constraint (2). Due to the binary nature of the return required by outside investors and the costs of higher sustainability, a firm will optimally either choose the minimum standard θ_m or θ_s. As a result of these optimal financing and production decisions, the aggregate amount of loans demanded by *sustainable* firms is $L_s(\Delta r)$.

Our third equilibrium condition relates to the market for sustainable funding ("*sustainable market clearing*"). The equilibrium sustainability subsidy Δr is determined by the intersection of demand and supply

$$L_s(\Delta r) = K[1 - F(\Delta r)]. \tag{7}$$

Characterization. Since warm-global capital K is insufficient to finance investment of all firms in the economy, a fraction of firms will choose to operate at the minimum production standard θ_m. As a result, both sustainable and non-sustainable firms are active in equilibrium. The equilibrium conditions outlined in the previous section then imply that firms must be indifferent between 1) producing sustainably and obtaining a financing subsidy of Δr as a compensation for higher production cost $c(\theta_s)$ and 2) producing less sustainably and incurring lower production cost $c(\theta_m)$ but unsubsidized financing conditions.

Either way, as firms operate a CRS technology, it is weakly optimal to go for maximal scale. Binding (IC) implies that the payoff of any (sustainable and less sustainable) firm in (5) simplifies to the product of the per-unit agency rent, $\frac{B}{\Delta p}$, and maximal scale, Ak_i:

$$U_i = \frac{B}{\Delta p} A k_i. \tag{8}$$

Given the utility expression (8), the equilibrium indifference condition now requires that the output capacity multiplier k of sustainable and non-sustainable firms be equalized. Therefore, the capacity multiplier for *less sustainable* firms, using $r(\theta_m) = 0$ and $c(\theta_m)$ in

(3), can be used to determine the respective multiplier for *all* firms:

$$k(q) = \frac{1}{\frac{B}{\Delta p} - [P(q) - c(\theta_m)]}. \tag{9}$$

Note that this capacity multiplier is only a function of primitives (in particular, the agency rent $\frac{B}{\Delta p}$ and the minimum standard θ_m) apart from the dependence on q. Intuitively, the dependence of the capacity multiplier on q arises because larger aggregate output pushes down product prices (and, hence profitability).

We now turn to the equilibrium value of aggregate output $q^* \leq \bar{q}$, which implies the equilibrium capacity multiplier $k^* := k(q^*)$ and product price $P(q^*)$.

Proposition 1 (Aggregate output q^*) *If total internal funds A are sufficiently small, $A < \bar{q}B/\Delta p$, the output of the economy is constrained by financial frictions. Equilibrium output q^* is then the unique solution to*

$$q = Ak(q). \tag{10}$$

If $A \geq \bar{q}B/\Delta p$, the economy is not constrained by the moral hazard problem and $q^* = \bar{q}$ as defined in (4).

Intuitively, Proposition 1 highlights that the moral hazard problem constrains aggregate output only if firm assets A are sufficiently low. The threshold $\bar{q}B/\Delta p$ can be interpreted as the total agency rent associated with zero-profit output $\bar{q}$.

We now turn to the equilibrium effect of sustainability preferences. While the distribution of investor preferences F and the sustainability threshold θ_s do not affect aggregate output q^*, they have *compositional* implications for production, i.e., the fraction of firms that produce sustainably versus the ones that produce at the minimum standard. To derive these compositional effects, we initially determine the equilibrium financing subsidy

Δr^* that just outweighs the production cost differential $\Delta c = c(\theta_s) - c(\theta_m) > 0$,

$$\Delta r^* = \frac{\Delta c}{P(q^*) - \frac{B}{\Delta p} + \Delta c}. \tag{11}$$

Intuitively, in equilibrium a larger cost differential Δc must translate into a larger equilibrium financing subsidy Δr^* (this holds strictly as long as the required subsidy meets some demand from sustainable investors, $\Delta r^* \leq \bar{w}$). Given Δr^*, we can now characterize the composition of production.

Proposition 2 (The sustainability subsidy and the composition of production) *Total sustainable output is given by:*

$$q_s^* = K[1 - F(\Delta r^*)]\frac{k^*}{c(\theta_s)k^* - 1}. \tag{12}$$

The equilibrium share of sustainable output, $\omega^* := \frac{q_s^*}{q^*}$*, satisfies*

$$\omega^* = \frac{q_s^*}{\min\{\bar{q}, Ak^*\}} < 1. \tag{13}$$

Proposition 2 highlights that sustainable output, q_s^*, is the product of the equilibrium supply of sustainable capital, $K[1 - F(\Delta r^*)]$, and a term that reflects leverage as well as the cost of sustainable production.

This characterization yields unambiguous comparative statics. We first analyze the effect of a trend in ESG demand by retail investors.

Corollary 1 *An increase in the amount of capital held by investors with sustainability concerns K or a First-Order Stochastic Dominance shift in $F(w)$ increase the share of sustainable investment ω^*, while the financing subsidy Δr^* remains unchanged.*

Intuitively, when ceteris paribus there is a greater supply of sustainable capital, this results in a greater share of sustainable investment. Still the financing subsidy Δr^* remains unchanged, as in equilibrium this is pinned down by firms' endogenous decision to become more sustainable and the resulting indifference condition (11).[13]

Further, we can analyze the effects of policy changes in the form of a stricter threshold for sustainability θ_s.

Corollary 2 *An increase in the sustainability standard θ_s decreases the share of sustainable investment ω^* and increases the equilibrium financing subsidy Δr^*.*

While a higher sustainability standard does not have an effect on aggregate output q^* and hence k^* (see Proposition 1), it increases the cost differential Δc for producing sustainably relative to the minimum standard. This higher cost differential, in turn, requires the capital cost subsidy for sustainable firms to go up, so as to keep sustainable production equally attractive, see (11). The required increase in the subsidy needs to be paid by households and, hence, reduces the attractiveness of the ESG fund for all households. As a result, previously marginal households no longer invest sustainably. This comparative statics highlights a key trade-off that a regulator is facing in our upcoming normative analysis. While increasing the sustainability cutoff reduces the negative externalities of sustainable firms, it reduces the fraction of firms that choose to produce sustainably.[14]

We conclude with the following comparative result on the effect of minimum production standards (see Section 5 for the derivation of the optimal θ_m).

Corollary 3 *If the minimum standard θ_m increases, holding θ_s constant, total output q^* decreases while sustainable investment q_s^* increases (implying also an increase in the respective share ω^*).*

[13]In this sense, Corollary 1 compares the outcome after the respective equilibrium adjustments. If over a shorter time horizon firms' sustainability levels remained unchanged, Δr^* would increase. As we have shown in a working paper version, availability of sustainable funding would still have real effects by increasing the share of sustainable investments by relaxing firm-level financial constraints.

[14]Note that the lower equilibrium differential Δr^* following a reduction of the threshold θ_s is thus not an immediate effect of investors' lower appreciation for sustainability when the threshold is lower. Recall that in our model households have non-consequentialist preferences, so that the warm-glow effect that they experience is independent of the threshold.

Intuitively, an increase in θ_m increases marginal production cost, which lowers aggregate output q^*.[15] Since raising the minimum standard lowers the relative cost of producing sustainably Δc, the required capital cost differential ensuring firm indifference, Δr^*, also decreases in tandem. As a result, sustainable output strictly increases.

3.3 First-best benchmark

Before analyzing how welfare Ω is maximized when the social planner is constrained to use only a given set of instruments, we introduce the first-best benchmark. Recall that total welfare, as given in (1), comprises consumer welfare, investment costs, and the social externality.

As firms are a priori homogeneous, it is sufficient to consider two control variables: total output q and a uniform sustainability standard θ. Then, (1) simplifies to

$$\Omega = \int_0^q P(\widetilde{q})d\widetilde{q} - q\left[c(\theta) + \rho(\theta^{\max} - \theta)\right]. \tag{14}$$

We denote the unique pair of optimizers by (θ_{FB}, q_{FB}). The optimal sustainability level is given by $\theta_{FB} = \hat{\theta}$, solving

$$c'\left(\hat{\theta}\right) = \rho. \tag{15}$$

The optimal output and market size is given by

$$P(q_{FB}) = c(\hat{\theta}) + \rho(\theta^{\max} - \hat{\theta}), \tag{16}$$

where we used $\theta_{FB} = \hat{\theta}$. Hence, at the first best the marginal cost (per unit of output) of increasing sustainability equals the associated marginal social benefit (per unit of out-

[15] The result follows immediately from (4) when the economy is unconstrained. When the aggregate output is constrained by financing frictions, $q^* < \bar{q}$, the result follows from the fact that the multiplier (9) decreases for any given level of q which also implies that the fixed point q^*, see (10) decreases. For details see Proof of Corollary 3.

put), condition (15), while marginal consumer welfare equals the marginal social cost of production, condition (16).

Lemma 1 (First best) *First best welfare Ω_{FB} is achieved when all firms choose the same sustainability standard $\theta_{FB} = \hat{\theta}$, given by (15), and aggregate output q_{FB} satisfies (16).*

4 Sustainable category for investments

4.1 When is a sustainable taxonomy optimal?

We now analyze first when it is optimal to introduce a sustainability classification for investments θ_s, taking as given environmental regulation in the form of the minimum production standard θ_m. This "partial" optimization may also be motivated by the fact that minimum production standards are subject to infrequent changes, and often argued to be not stringent enough, e.g., due to political constraints.[16] Such constraints may also prevent the adjustment of θ_m despite an improved understanding of the negative impact of the respective externalities.

Based on our preceding equilibrium characterization in Propositions 1 and 2, it is instructive to rewrite the social planner's objective function (1) as follows

$$\begin{aligned}\Omega &= \int_0^{q^*} P(q)dq - \rho q^* \left[c\left(\theta_m\right) + \rho(\theta^{\max} - \theta_m)\right] \\ &\quad + q_s^*\left(\theta_s\right)\left(\rho(\theta_s - \theta_m) - \left[c\left(\theta_s\right) - c\left(\theta_m\right)\right]\right),\end{aligned} \tag{17}$$

where we make the dependence on θ_s explicit and exploit the fact that firms optimally either choose θ_m or θ_s. The first term in (17) captures the baseline welfare if all investment

[16] For instance, for firms operating outside the jurisdiction or procuring from other countries, supply chain regulation may prescribe compliance only with a limited set of fundamental human rights, such as prohibiting the use of slave labour. In this case, the standard θ_s could prescribe, in addition, certain rights for employees or trade union activity. In the case of environmental pollution, for a given pollutant θ_m may be chosen only to avoid immediate health risks. Again, θ_s then represents more stringent requirements.

were non-sustainable. The second term captures the incremental effect of sustainable output with quantity $q_s^* (\theta_s)$.

As total output q^* does not depend on θ_s, changes in θ_s only affect the social planner's objective through the second term. As a result, the planner's program reduces to that of maximizing the product of $q_s^* v(\theta_s)$ where $v(\theta_s)$ captures the incremental benefit per unit of output, trading off the reduction of the externality with the increase in cost,

$$v(\theta_s) := \rho(\theta_s - \theta_m) - [c(\theta_s) - c(\theta_m)].$$

It is useful to highlight that $v(\theta_s)$ is maximized for $\theta_s = \hat{\theta}$, see (15), which follows directly from the first-order condition $v'(\theta_s) = 0$.

Proposition 3 (Optimality of a sustainable investment classification) *Keeping the minimum standard θ_m fixed, the introduction of a sustainable investment category is strictly suboptimal if $\theta_m \geq \hat{\theta}$ and strictly optimal if $\theta_m < \hat{\theta}$.*

Intuitively, if the minimum standard θ_m is already very stringent, $\theta_m > \hat{\theta}$, the incremental benefit of introducing an even more stringent classification for sustainable investments is negative, $v(\theta_s) < 0$, for all $\theta_s > \theta_m$. Hence, the regulator should not introduce a category for sustainable investments. While even in this case the availability of sustainable investment opportunities (and the ensuing warm-glow) would attract investors and thereby lead to subsidized capital costs for sustainable firms, it would induce a fraction of firms to overinvest in sustainability from welfare perspective.[17] The first part of Proposition 3 thus qualifies the notion of a general social desirability of an ESG-classification of investment funds. Even when such investment opportunities meet with positive demand, this could represent "too much of a good thing."[18]

[17]Recall that the social planner's objective only accounts for real effects, but not investors' warm-glow perception.

[18]Demand for sustainable investments is positive as along θ_s is not so high that even the "most sustainable" investor with warm glow $\bar{w}$ is unwilling to pay the required subsidy in (11).

4.2 Optimal classification threshold

We now focus on the case where environmental regulation is sufficiently lax, i.e., the minimum standard satisfies $\theta_m < \hat{\theta}$. As discussed by Tirole (2012), this case is arguably more relevant in practice. As the net benefit v is positive (for θ_s not too high) it is now efficient to introduce a sustainability classification for investments. The classification allows the policymaker to harness the households' sustainability preferences so as to (indirectly) subsidize higher sustainability of some firms.

Before characterizing the key welfare trade-offs behind choosing the optimal taxonomy, we note at the outset that this policy tool – used in isolation – should not be expected to restore first-best. The second-best nature is a direct result of Propositions 1 and 2 as well as Lemma 1. First, by Proposition 1, aggregate output in the economy q^*, and, hence, consumer welfare, is independent of the regulatory choice of sustainability classification for investments θ_s. Hence, unless $q^* = q_{FB}$, the choice of the classification for sustainable investments is unable to rectify this. We will revisit the issue of market output when we also endogenize θ_m in the subsequent Section. Second, by Proposition 2 only a fraction of firms $\omega^* < 1$ chooses to produce sustainably at level θ_s in equilibrium (while first-best requires that all firms produce at the sustainability level $\theta_{FB} = \hat{\theta}$).

Proposition 4 (Optimal classification threshold) *Suppose that $\theta_m < \hat{\theta}$, so that it is strictly optimal to introduce a sustainable investment category. Then the optimal threshold, θ_s^*, satisfies*

$$\frac{\partial \ln v(\theta_s)}{\partial \theta_s} = \left| \frac{\partial \ln q_s^*(\theta_s)}{\partial \theta_s} \right| > 0, \tag{18}$$

so that $\theta_m < \theta_s^* < \hat{\theta}$.

Similar to the optimal pricing decision of a monopolist, the optimal calibration of θ_s^* can be expressed in terms of (semi)-elasticities: Ignoring the effect on the supply of sustainable capital $q_s^*(\theta_s)$, it would be optimal to set $\frac{\partial \ln v(\theta_s)}{\partial \theta_s} = 0$ or equivalently $\theta_s^* = \hat{\theta}$. However, because the regulator additionally needs to account for the downward sloping supply of sustainable capital (see Corollary 2) the optimal choice features $v'(\theta_s^*) > 0$ so that $\theta_s^* < \hat{\theta}$). We next conduct several comparative analyses.

Shift in investor preferences. The characterization in (18) can be used to determine the effect of a trend towards more sustainable preferences by investors. Intuitively, stronger preferences for sustainability dampen the negative effect on supply and thus alleviate the documented trade-off of the social planner when choosing θ_s^*.[19]

Corollary 4 *Suppose $\theta_m < \hat{\theta}$. As investor preferences become more sustainable in the sense of a monotone hazard rate shift in $F(w)$, the optimal threshold θ_s^* increases.*

Interpreting Corollary 4 in the time series, the social planner should optimally apply a lower threshold when sustainable preferences for investment are still less prevalent rather than "challenging" the market with a high initial classification threshold. As preferences shift over time towards more sustainability the optimal standard θ_s^* should be increased in conjunction. Below we will further investigate such dynamics by introducing a feedback loop when investors' preferences are shaped by an endogenous social norm.

Comparative analysis in the social cost of the externality. Recall that ρ denotes the social cost of the externality, $\theta^{\max} - \theta_i$ per unit of output. It is intuitive that, holding the minimum standard θ_m constant, higher social costs of the externality render it optimal to increase the threshold for the sustainable investment classification, as the subsequent Corollary confirms.[20]

Corollary 5 *Suppose $\theta_m < \hat{\theta}$. As the social costs of the externality, ρ, increases, the optimal threshold θ_s^* increases.*

[19]Formally, the hazard rate condition is slightly stronger than the FOSD condition we imposed in Corollary 2.

[20]The minimum standard may instead not be sufficiently responsive, e.g., given political constraints, while the introduction of a new policy instrument, such as a sustainable investment taxonomy, may represent an opportunity to take into account new information or changes in societal preferences.

5 Optimal environmental policy

We now consider the central question whether there is a role for a taxonomy for investments even if environmental policy (in the form of the minimum standard) is optimally chosen. Subsequently, we show that our results extend qualitatively when the social planner can (also) use a Pigouvian tax on the externality.

5.1 Optimality of a sustainable investment category

We are thus first interested in endogenizing the condition when a sustainable investment category is optimal, now under the optimal choice of the minimum standard. It turns out that to answer this question, it is sufficient to consider the auxiliary program where the social planner uses only a minimum standard.

Auxiliary program: Optimal minimum standard as sole policy tool. We thus set initially $q_s^* = 0$ in the objective (17). The optimal choice of θ_m^*, characterized by the single first-order condition $\frac{d\Omega}{d\theta_m} = 0$, aims to balance deviations from the two separate first-order conditions for the technology and quantity in the first-best benchmark, see (15) and (16):

$$\rho - c'(\theta_m) = \left| \frac{\frac{dq^*}{d\theta_m}}{q^*(\theta_m)} \right| \left[P(q^*) - c(\theta_m) - \rho(\theta^{\max} - \theta_m) \right]. \tag{19}$$

Here, the left hand-side of (19) captures the socially optimal technology choice, cf. condition (15). The right-hand side captures condition (16), which is the marginal social surplus of an additional unit of output fixing technology θ_m. It is scaled by the semi-elasticity of output to environmental standards, $\left| \frac{dq^*}{d\theta_m} / q^*(\theta_m) \right|$, which results from the feedback effect of the minimum standard on aggregate output. We obtain:

Lemma 2 (Optimal minimum standard as sole policy tool) *Suppose the social planner can only impose a minimum standard θ_m (but no sustainable investment category).*

There exists a threshold for internal funds A_{FB} such that the optimal standard satisfies $\theta_m^* > \hat{\theta}$ for $A > A_{FB}$ and $\theta_m^* < \hat{\theta}$ for $A < A_{FB}$. At $A = A_{FB}$, first-best welfare is achieved, i.e., $\theta_m^* = \hat{\theta}$ and $q^*\left(\hat{\theta}\right) = q_{FB}$.

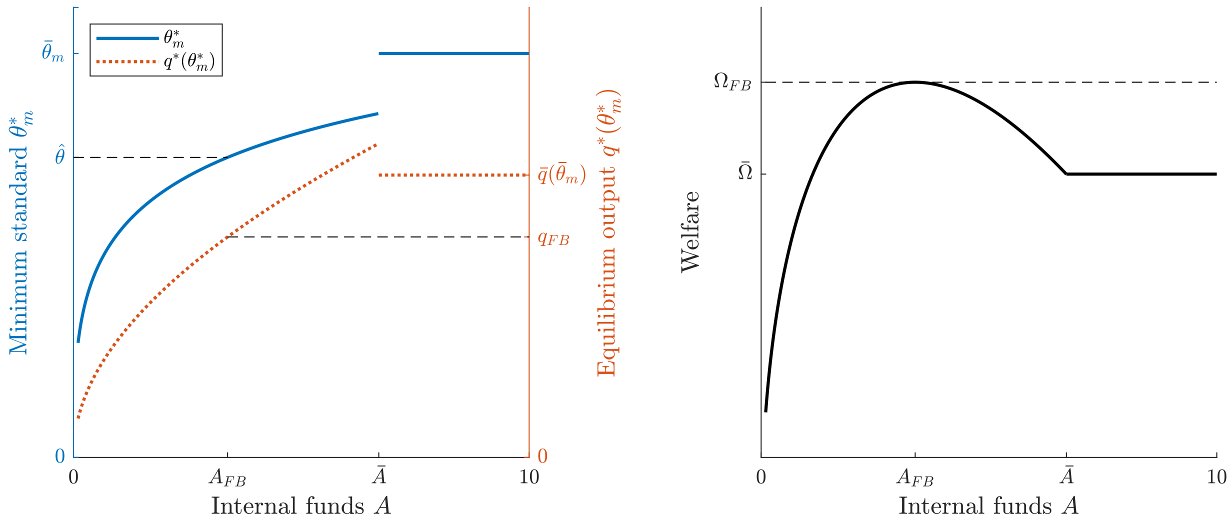

Figure 1. Optimal minimum standard as a sole policy tool. The graph in the left panel plots the optimal minimum standard θ_m^* and the associated equilibrium output $q^*\left(\theta_m^*\right)$ as a function of internal funds A. Note that the respective units on the y-axis have different scales for θ_m^* (left scale) and $q^*\left(\theta_m^*\right)$ (right scale). The right panel plots the resulting welfare under the optimal minimum production standard θ_m^*. For $A = A_{FB}$, first-best welfare is achieved. Financial constraints don't bind for $A \geq \bar{A}$.

Intuitively, if internal funds A are low, $A < A_{FB}$, the economy is below the socially efficient scale, $q^* < q_{FB}$, (see dotted line in left panel of Figure 1) so that the marginal social surplus of an additional unit of output is positive, $P\left(q^*\right) - c\left(\theta_m\right) - \rho\left(\theta^{\max} - \theta_m\right) > 0$. To avoid "overshrinking" the economy, the environmental standard θ_m^*, solving (19), is now optimally chosen to be less stringent $\theta_m^* < \hat{\theta}$. The concern about underproduction due to financial constraints becomes less and less relevant as firm assets A increase. As a result, welfare is initially increasing in A (see right panel of Figure 1). In particular, at $A = A_{FB}$, the planner can choose $\theta_m^* = \hat{\theta}$, achieving technology efficiency, and "the right amount of" financial constraints ensures that firms do not overproduce, $q^* = q_{FB}$. First-best welfare is attained.

As internal funds increase beyond A_{FB}, firms overproduce (from a planner perspective) despite still binding financial constraints for $A < \bar{A}$, $q^* > q_{FB}$. Intuitively, financial constraints are not severe enough to limit overproduction. This overproduction is now optimally countered by "overinvestment" in the abatement technology, $\theta_m^* > \hat{\theta}$. This insight extends intuitively when firms have sufficient internal funds, $A > \bar{A}$, so that they become financially unconstrained (even under the optimal policy $\bar{\theta}_m$) and jointly supply the associated zero-profit output $\bar{q}\left(\bar{\theta}_m\right)$.[21]

In the presence of two frictions – financing constraints and production externalities – one would expect one tool, in this case θ_m, to be insufficient to restore first-best. In particular, while the planner can force all firms to choose $\hat{\theta}$ by setting $\theta_m = \hat{\theta}$, this policy choice does not automatically ensure the socially optimal quantity q_{FB} since output is endogenously supplied by profit-maximizing firms. In fact, unless financial constraints are sufficiently binding, output is oversupplied at $\theta_m = \hat{\theta}$. However, as we have shown, the first best is achieved when the financial constraints are "just right," which stands in stark difference to canonical corporate finance models, where financial constraints reduce total surplus as they prevent firms from exploiting profitable investment opportunities. In the presence of social externalities, financial constraints are thus no longer unambiguously harmful, not even under the optimal (environmental) policy.[22]

Joint optimization. We now let the social planner choose both policy instruments and denote the solution by $(\theta_s^{**}, \theta_m^{**})$. As Proposition 3 implies that the introduction of a sustainable investment category is only beneficial if the minimum standard is sufficiently

[21] The switch from binding financial constraints to non-binding constraints causes a jump in the policy function exactly at the level of internal funds where the planner is indifferent between causing financial constraints or not. This jump arises because, for $A = \bar{A}$, the objective function has two global maxima. See details in Proof of Lemma 2.

[22] As we show below, with financial constraints neither the setting of a minimum standard nor the choice of an optimal Pigouvian tax represents an unambiguously superior policy instrument, as either achieves first best for different values of A.

lax $\theta_m < \hat{\theta}$ and Lemma 2 showed that θ_m^* is optimally lax if $A < A_{FB}$, we intuitively obtain:

Proposition 5 (Taxonomy under optimal environmental policy) *If the social planner can both impose a minimum standard θ_m and a sustainable investment category θ_s, introducing the latter only improves welfare if $A < A_{FB}$ and it is otherwise strictly suboptimal.*

In sum, the introduction of a taxonomy for sustainable investments can only add value on top of *optimal* environmental regulation if "lack of financing" is a source of a social inefficiency. What is more, there must be a sufficiently severe lack of financing, as at $A = A_{FB}$ output is still limited by financial constraints.

We now analyze when it is more likely that the introduction of a sustainable investment category is socially optimal under the optimal choice of environmental policy θ_m:

Corollary 6 *In an economy where environmental policy (θ_m) is optimally chosen, the additional introduction of a sustainable investment category is more likely to be beneficial if, ceteris paribus, firms' internal funds are more limited (lower A) or the agency problem vis-à-vis external investors is more severe (higher $B/\Delta p$).*

Hence, under optimal environmental policy the introduction of a sustainable investment category is more likely to be socially beneficial when (lack of) financial development or the (mal-)functioning of the legal system sufficiently limit internal funding and raise the costs of external financing. For a developed financial system, as prevailing in the European Union, this would thus seem less likely.

We conclude this section with a brief discussion of the properties of the optimal policy under joint optimization of θ_m and θ_s. Since it is suboptimal to introduce a sustainable investment classification if $A \geq A_{FB}$, we immediately obtain that $\theta_m^{**} = \theta_m^*$ (see Figure 2 for $A \geq A_{FB}$). In contrast, when $A < A_{FB}$, the optimal use of the second tool, the sustainable

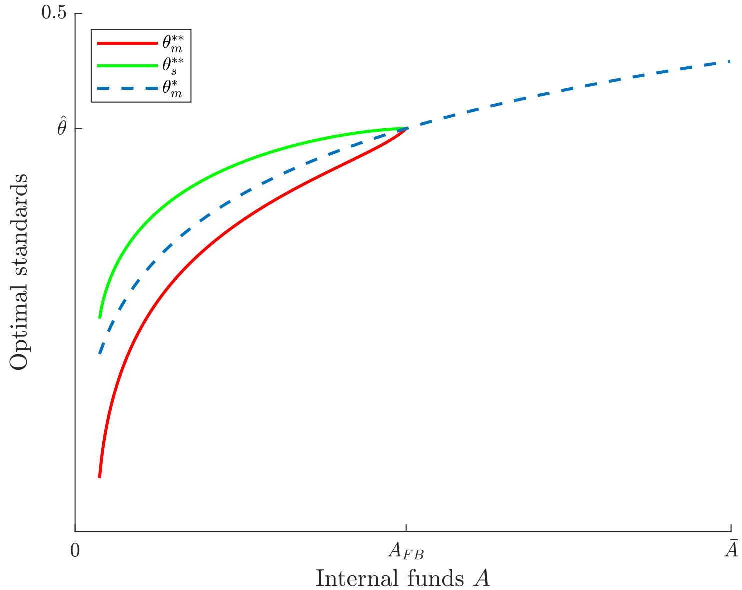

Figure 2. Jointly optimal policy $(\theta_m^{**}, \theta_s^{**})$. The graph plots the optimal choice $(\theta_m^{**}, \theta_s^{**})$ if the planner can both flexibly a choose a minimum standard and a taxonomy for sustainable investments $\theta_s > \theta_m$. For $A < A_{FB}$, $\theta_m^{**} < \theta_m^{*} < \theta_s^{**}$. For $A \geq A_{FB}$, we obtain that $\theta_m^{**} = \theta_m^{*}$ and no taxonomy for investments is introduced.

investments category, feeds back into the optimal choice of the minimum standard, (see Figure 2 for $A < A_{FB}$): The planner can lower the environmental minimum standard $\theta_m^{**} < \theta_m^{*}$ (compare red line for θ_m^{**} to blue dotted line for θ_m^{*}), which alleviates underproduction in the economy ($\frac{dq^*}{d\theta_m} < 0$), and then harnesses the availability of sustainable finance to subsidize a subset of firms to choose a more stringent standard $\theta_s^{**} \in \left(\theta_m^{**}, \hat{\theta}\right)$, which is characterized in Proposition 3 (compare green line for θ_s^{**} relative to red line).

5.2 Pigouvian taxes

While in practice the reduction of harmful emissions or the achievement of other sustainability objectives is often governed by a minimum standard, rather than by (Pigouvian) taxes, the latter represents the main textbook policy instrument. Since both the intuition and formal results are similar to the previous section, we present the insights in a heuristic fashion in the main text and relegate the formal analysis to Appendix B.

Suppose thus that the social planner can impose a tax τ on externalities, first as the sole environmental policy tool. Since the marginal welfare cost of externalities equals ρ in our setting, a Pigouvian tax would stipulate $\tau = \rho$. Let us now consider the case when financial constraints do not constrain production. Given a tax of $\tau = \rho$ a firm would optimally choose θ_i to minimize $c(\theta_i) + \rho(\theta^{\max} - \theta)$, which results in $\theta_i^* = \theta^* = \hat{\theta}$, cf. condition (15). Since financing constraints do not limit aggregate output, equilibrium output is then pinned down by the zero profit condition,

$$P(\bar{q}) = c(\theta^*) + \tau(\theta^{\max} - \theta^*), \tag{20}$$

so that, with $\tau = \rho$ and $\theta^* = \theta_{FB} = \hat{\theta}$, equilibrium output is thus equal to the first-best output q_{FB}, cf. condition (16). In sum, when financial constraints do not limit aggregate output, a Pigouvian tax leads to the first-best choice of both sustainability and aggregate investment so that first-best welfare is achieved. Hence, it is strictly inefficient to introduce an additional sustainable investment category, as this would lead to overinvestment by a fraction of firms.

In contrast, when aggregate output is impacted by financial constraints, the Pigouvian tax on its own is no longer sufficient to achieve first best: At $\tau = \rho$, output in the economy is excessively constrained as the marginal social surplus is still strictly positive. It is then optimal for the social planner to set τ strictly below the Pigouvian level, at which point, however, the thereby implemented sustainability standard falls short of $\hat{\theta}$. This makes it optimal to introduce a sustainable investment category and thereby increase sustainability for a fraction of firms. This confirms our qualitative insights from the preceding analysis with a minimum sustainability standard as the main policy instrument, albeit now the introduction of a sustainable investment classification becomes optimal if and only if financial constraints limit output (while previously the condition was even stricter; cf. Proposition 5 with Lemma 2).

Proposition 6 (Pigouvian taxation) *Suppose the social planner can both impose a tax τ on emissions and a sustainable investment category θ_s. Then, there exists a threshold for internal funds $\widehat{A} > A_{FB}$ such that the following case distinction applies:*

1. *If $A \geq \widehat{A}$, it is strictly suboptimal to introduce a classification for sustainable investment, while the optimal tax is set at the Pigouvian level, $\tau^{**} = \rho$. First-best welfare is attained.*

2. *If $A < \widehat{A}$, the social planner optimally introduces a sustainability category θ_s^{**} for investments and sets a tax strictly below the Pigouvian level, $\tau^{**} < \rho$.*

Note that the previously considered a minimum standard imposed a financial burden on firms only through the respective costs $c(\theta_m)$. Instead, if a tax τ is levied to incentivize adoption of a sustainability level θ, firms incur both the respective costs $c(\theta)$ and the tax burden $\tau(\theta^{\max} - \theta)$.[23] From this observation it follows intuitively that the threshold for assets $\widehat{A}$, at which the first best is achieved with a Pigouvian tax, is strictly above the level A_{FB}, at which a minimum standard achieves the first best. Comparing taxation with a minimum standard, there is thus a range of values for A where the minimum standard strictly outperforms from a welfare perspective. Observe next that under taxation, the first best is, however, achieved for all $A \geq \widehat{A}$, while with a minimum standard the social planner still faces a trade-off for all $A > A_{FB}$ (see right panel of Figure 1), resulting both in an excessively high standard and in excessive production, $\theta_m^* > \hat{\theta}$ and $q^* > q_{FB}$ (see left panel of Figure 1). This in turn implies that for sufficiently high values of A taxation outperforms.

For the sake of completeness, the Appendix also covers the case where the social planner can avail herself of both instruments of environmental policy. Then, first best welfare is feasible for all $A \geq A_{FB}$, where the use of both instruments is strictly optimal over the

[23] In a world without financial constraints (and where also distributive considerations are irrelevant), the additional tax burden would be inconsequential.

range $A_{FB} < A < \widehat{A}$. Intuitively, a binding minimum standard at $\theta_m = \hat{\theta}$ governs the choice of sustainability, while through additional taxation the social planner is able to adjust downwards the size of total production to q_{FB}. We finally note that in this case, when both instruments are available, our conclusion from Proposition 5 on the optimality of a sustainable investment category remain robust: The introduction of a sustainable investment category is beneficial if and only if $A < A_{FB}$.

6 Social network effects

So far, we treated preferences for sustainability, as governed by $F(w)$, as an exogenous primitive of the economy. We now allow for a feedback mechanism via social norms, as we suppose that individual preferences for sustainable investment depend on the observed or anticipated behavior of other investors. We suppose that the willingness-to-pay w for a more sustainable investment increases when more households make such investments.

We first provide some motivation for these preferences. As acknowledged by a large literature on environmental economics, preferences for sustainability essentially derive from non-use benefits[24] and as such are particularly susceptible to social norms. These can be shaped by the anticipated or observed behavior of others. Such an understanding of how social norms change is widely shared in social sciences,[25] and it is also recognized in economics. There, the relevance of the behaviour of others has been confirmed in many experiments, such as in games of contributions to a public good.[26] In addition, various field experiments relate environmentally conscious behavior to that of others.[27] Endogenous preferences and changes in norms have also been incorporated in policy suggestions to fight climate change, see e.g., Stiglitz (2019). Our key premise is thus that households'

[24] On this concept see, for instance, Pearce et al. (2006).

[25] Also the relevance for public policy has been recognized early, see e.g. Cialdini et al. (1990).

[26] See Sugden (1984) for an early example.

[27] For instance, individual recycling behavior has been found to correlate strongly with beliefs about such behavior in the community; see Cialdini (2003) or the various studies quoted in Schultz (2002).

investment preferences are not exogenously fixed but endogenous in that they depend on the observed or anticipated behavior of others.

We capture such social norms in reduced-form by stipulating a dependency of the distribution of preferences on the (anticipated) equilibrium cutoff w^*, i.e., $F(w \mid w^*)$. As w^* decreases, so that sustainable investments are more widespread, the distribution moves upwards in the sense of strict First-Order Stochastic Dominance (FOSD). Relative to our preceding derivations, we only need to modify expression (12), where we now obtain

$$q_s^* = K[1 - F(w^* \mid w^*)] \frac{k^*}{c(\theta_s) k^* - 1} \text{ with } w^* = \Delta r^*. \tag{21}$$

For our normative analysis, we now need to take the total derivative of $F(w^* \mid w^*)$, that is including the feedback effect $\left. \frac{dF(w^* \mid \widetilde{w})}{d\widetilde{w}} \right|_{\widetilde{w}=w^*} < 0$. Hence, when the social planner reduces θ_s and thereby also w^*, this has now also an indirect positive effect through the (social norm) feedback. This makes it optimal to choose a strictly lower threshold.

Proposition 7 *Suppose households' preferences are subject to the described social-norm feedback (in that $F(w \mid w^*)$ shifts with w^* in the sense of strict FOSD). Then, the social planner should optimally choose a strictly lower threshold θ_s^* for the classification of sustainable investment (provided it is optimal to introduce such a classification).*

7 Concluding remarks

Agencies around the world are in the process of developing taxonomies for sustainable (or ESG) investment products. This paper poses two questions. First, is it beneficial to introduce a classification for sustainable investments when a social planner can optimally choose more direct policy instruments, such as a minimum standard or Pigouvian taxes? Second, if so, how sustainable does a firm have to be to be eligible for marketing their securities as sustainable investments?

The answer to the first question is rendered non-trivial as the planner in our model only cares about real effects, but not directly about the warm glow that households experience when investing in a taxonomy-conform investment product. The key mechanism in our model is that the planner can channel non-consequentialist warm glow preferences by retail investors towards subsidizing firms' sustainability investments.

The provision of such subsidies is efficient in two cases. First and intuitively, if policy failures render environmental regulation to be too lax, they can partially mitigate under-investment in sustainability. More interestingly, there is a role for harnessing "sustainable capital" even if traditional environmental tools can be optimally chosen, but this requires financial frictions to prevent the economy (or the considered industry) from running at the socially efficient scale when the socially desirable minimum standard or tax on externalities is set. In this case, the sustainable investment classification adds a valuable second instrument that mitigates the trade-off between achieving higher sustainability for production and generating an inefficient contraction of economic activity. When choosing the optimal threshold for sustainability, the planner accounts for the downward-sloping supply of sustainable capital. One consequence of this is that, as preferences for sustainability become more widespread across investors, it is optimal to increase the stringency of the standard over time.

If environmental policy is neither too lax nor financial constraints cause an inefficient reduction in economic activity, the (additional) introduction of a classification for sustainable investments is strictly suboptimal from a welfare perspective. While the introduction of such a classification would potentially meet the demand by households for sustainable investments, there are no immediate benefits from satisfying such demand (since welfare only accounts for real effects).

Our analysis has normative implications for optimal policy across countries. Economies with higher financial development, for which financial constraints are less relevant, should have stricter environmental regulation. Then, the introduction of a sustainable investments category backfires unless environmental policy is too lax. In contrast, for economies with poorer financial development, binding financial constraints imply a welfare-enhancing role for sustainable finance even under optimal environmental policy as subsidies are required to finance the transition.

More generally, our paper has analyzed the role of financial regulation for supporting a sustainable transition. We focused on one particular tool, a taxonomy for sustainable investments. Other regulatory initiatives are ongoing, e.g., green monetary policy as studied in Papoutsi et al. (2021), or green capital requirements as studied in Oehmke and Opp (2022). The analysis of which (combination of) tools is most impactful is an interesting question for future empirical and theoretical research.[28] The results of this paper suggest that financial regulation is only part of the optimal policy mix if "finance" is the root of the problem. Put differently, absent frictions in the financial sector, environmental regulation should be at the top of the regulatory pecking order. We also note that if the social planner could provide the respective subsidies directly, this would further erode the role for such classification. Raising the necessary financing for a direct subsidy through taxes could, however, lead to additional distortions.

In addition, our model has abstracted from additional (transactional) costs associated with the introduction of the sustainable taxonomy and its monitoring. In practice, sustainability investments by firms are not readily observable and, hence, require active monitoring, which may lead to a revival of actively managed funds with high fees. In addition, the packaging and advertising of such funds may also involve a great deal of greenwashing.[29] Even though such considerations are outside of our model, our analysis cautions against the frequently encountered optimism regarding more sustainable (ESG) investments.

[28]This echoes the recent discussion of Stiglitz (2019), who argues for an "alternative approach" to environmental regulation, which uses various instruments so as to deal with real-world complexities, including distributional issues, imperfect markets, or risk and uncertainty.

[29]Even of fraudulent nature; cf. the recent actions against DWS, the in-house provider of investment funds of Deutsche Bank.

References

Andreoni, James, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal*, 06 1990, *100* (401), 464–477.

Bartram, Söhnke M., Kewei Hou, and Sehoon Kim, "Real effects of climate policy: Financial constraints and spillovers," *Journal of Financial Economics*, 2022, *143* (2), 668–696.

Berg, Florian, Julian F Kölbel, and Roberto Rigobon, "Aggregate Confusion: The Divergence of ESG Ratings*," *Review of Finance*, 05 2022, *26* (6), 1315–1344.

Biais, Bruno and Augustin Landier, "Emission caps and investment in green technologies," 2022. Working Paper, HEC Paris.

Bizzotto, Jacopo and Bård Harstad, "The certifier for the long run," *International Journal of Industrial Organization*, 2023, *87*, 102920.

Bonnefon, Jean-Francois, Augustin Landier, Parinitha Sastry, and David Thesmar, "The Moral Preferences of Investors: Experimental Evidence," 2019. Working Paper, TSE, MIT and HEC.

Boyle, Kevin J, William H Desvousges, F Reed Johnson, Richard W Dunford, and Sara P Hudson, "An investigation of part-whole biases in contingent-valuation studies," *Journal of environmental economics and management*, 1994, *27* (1), 64–83.

Broccardo, Eleonora, Oliver Hart, and Luigi Zingales, "Exit versus Voice," *Journal of Political Economy*, 2022, *130* (12), 3101–3145.

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

Cialdini, Robert B, "Crafting normative messages to protect the environment," *Current directions in psychological science*, 2003, *12* (4), 105–109.

_ , Raymond R Reno, and Carl A Kallgren, "A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places.," *Journal of personality and social psychology*, 1990, *58* (6), 1015.

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

de Freitas Netto, Sebastiao Vieira, Marcos Felipe Falcao Sobral, Ana Regina Bezerra Ribeiro, and Gleibson Robert da Luz Soares, "Concepts and forms of greenwashing: a systematic review," *Environmental Sciences Europe*, 2020, *32* (19).

Desvousges, William, Kristy Mathews, and Kenneth Train, "Adequate responsiveness to scope in contingent valuation," *Ecological Economics*, 2012, *84*, 121–128.

Döttling, Robin and Magdalena Rola-Janicka, "Too Levered for Pigou? A Model of Environmental and Financial Regulation," 2022. Working Paper, Erasmus University Rotterdam and Tilburg University.

Edmans, Alex, Doron Levit, and Jan Schneemeier, "Socially Responsible Divestment," 2022. European Corporate Governance Institute–Finance Working Paper.

Green, Daniel and Benjamin Roth, "The Allocation of Socially Responsible Capital," 2021. Working Paper, Harvard Business School.

Gupta, Deeksha, Alexandr Kopytov, and Jan Starmans, "The Pace of Change: Socially Responsible Investing in Private Markets," *Available at SSRN 3896511*, 2022.

Hakenes, Hendrik and Eva Schliephake, "Responsible Investment and Responsible Consumption," 2021. Working Paper, University of Bonn.

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.

Holmström, Bengt and Jean Tirole, "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 1997, *112* (3), 663–691.

Humphrey, Jacquelyn, Shimon Kogan, Jacob Sagi, and Laura Starks, "The asymmetry in responsible investing preferences," Technical Report, National Bureau of Economic Research 2021.

Inderst, Roman and Florian Heider, "A Corporate Finance Perspective on Environmental Policy," 2022. Working Paper, Goethe University Frankfurt.

Kahneman, Daniel and Jack L Knetsch, "Valuing public goods: The purchase of moral satisfaction," *Journal of Environmental Economics and Management*, 1992, *22* (1), 57–70.

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

Oehmke, Martin and Marcus Opp, "A Theory of Socially Responsible Investment," 2019. Working Paper, LSE and SSE.

_ **and** _ , "Green Capital Requirements," 2022. Working Paper, LSE and SSE.

Papoutsi, Mellina, Monika Piazzesi, and Martin Schneider, "How Unconventional is Green Monetary Policy?," 2021. Working Paper, European Central Bank and Stanford University.

Pastor, Lubos, Robert F. Stambaugh, and Lucian Taylor, "Sustainable Investing in Equilibrium," *Journal of Financial Economics*, 2021, *142* (2), 550–571.

Pearce, David, Giles Atkinson, and Susana Mourato, *Cost-Benefit Analysis and the Environment: recent developments*, OECD, 2006.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski, "Responsible Investing: The ESG-Efficient Frontier," *Journal of Financial Economics*, 2021, *142* (2), 572–597.

Piccolo, Alessio, Jan Schneemeier, and Michele Bisceglia, "Externalities of Responsible Investments," *Available at SSRN 4183855*, 2022.

Riedl, Arno and Paul Smeets, "Why Do Investors Hold Socially Responsible Mutual Funds?," *Journal of Finance*, 2017, *72* (6), 2505–2550.

Schultz, P Wesley, "Knowledge, information, and household recycling: Examining the knowledge-deficit model of behavior change," *New tools for environmental protection: Education, information, and voluntary measures*, 2002.

Stiglitz, Joseph E., "Addressing climate change through price and non-price interventions," *European Economic Review*, 2019, *119*, 594–612.

Sugden, Robert, "Reciprocity: The Supply of Public Goods through Voluntary Contributions," *Economic Journal*, 1984, *94* (376), 772–87.

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

Xu, Qiping and Taehyun Kim, "Financial Constraints and Corporate Environmental Policies [The limits of limited liability: Evidence from industrial pollution]," *Review of Financial Studies*, 2022, *35* (2), 576–635.

A Proofs

Proof of Proposition 1. In equilibrium, aggregate output satisfies $q^* \in [q_{\min}, \bar{q}]$. Here, $q_{\min}$ is the output that can be produced even if firms have no inside funds, $A = 0$, i.e.,

$$q_{\min} = \begin{cases} 0 & \text{if } P(0) \leq \frac{B}{\Delta p} + c(\theta_m) \\ P^{-1}\left(\frac{B}{\Delta p} + c(\theta_m)\right) & \text{if } P(0) > \frac{B}{\Delta p} + c(\theta_m) \end{cases} \tag{A.1}$$

This follows from the fact that as long as $P(q) > \frac{B}{\Delta p} + c(\theta_m)$ each individual firm's borrowing constraint would not bind, see (IC) and (IR) using $r(\theta_m) = 0$.[30] The upper bound $\bar{q}$ follows from the fact that for any $q > \bar{q}$ firms would earn a lower return than their outside option $r_0 = 0$.

We now turn to the question whether output $\bar{q}$ is feasible in the presence of financial constraints. Suppose that aggregate output is at $\bar{q}$, then the associated output capacity multiplier is $k^* = \frac{\Delta p}{B}$. Each individual firm takes this multiplier as given. We now distinguish between two cases.

Case 1) If $A < \bar{q}B/\Delta p$, then even if all firms were to lever up to the maximum (using the candidate multiplier $k^* = \frac{\Delta p}{B}$), aggregate output of $\bar{q}$ would not be feasible. Hence, q^* is simply the unique solution to

$$q = Ak(q)$$

Uniqueness follows from the fact that $k(q)$ is strictly decreasing and continuous in q and $Ak(q_{\min}) > 0$.

Case 2) If $A \geq \bar{q}B/\Delta p$, it is feasible to produce aggregate output of $\bar{q}$. In this case, firms are indifferent between investing and the storage technology. ■

Lemma A.1 (No Shirking) *A sufficient condition to rule out shirking in equilibrium is:*

$$\frac{B}{\Delta p} < P(q_{\min}) - \frac{P(q_{\min}) - \min[\{c(\theta_s)(1-\bar{w}), c(\theta_m)\}]}{\Delta p}, \tag{A.2}$$

where $q_{\min}$ is given by (A.1).

Proof of Lemma A.1. Suppose the entrepreneur shirks in equilibrium, then one only needs to consider the investors' IR constraint:

$$(1 - \Delta p) D_i \geq (c(\theta_i) q_i - A_i)(1 + r(\theta_i)). \tag{IR*}$$

[30] We only need to check this for the unsustainable firm since the sustainable firm's multiplier is identical in equilibrium.

Binding (IR*) now implies that the face value of debt is set to:

$$D_i = \frac{(c(\theta_i) q_i - A)(1 + r(\theta_i))}{1 - \Delta p}.^{31} \tag{A.3}$$

Then, the manager's payoff, including the payoff from the storage technology $(A - A_i)$, is:

$$\begin{aligned} U_i &= (1 - \Delta p)\left[q_i P(q) - \frac{(c(\theta_i) q_i - A_i)(1 + r(\theta_i))}{(1 - \Delta p)}\right] + Bq_i + (A - A_i) \\ &= q_i [(1 - \Delta p) P(q) - c(\theta_i)(1 + r(\theta_i)) + B] + A_i r(\theta_i) + A. \end{aligned}$$

If (A.2) holds, then $(1 - \Delta p) P(q) - c(\theta_i) q_i (1 + r(\theta_i)) + B < 0$ irrespective of the product price P (since P is highest for $q_{\min}$) and regardless of whether the firm produces unsustainably at cost $c(\theta_m)$ with $r(\theta_m) = 0$ or sustainably at cost $c(\theta_s)$ with the highest financing subsidy $\bar{w}$. As a result, the entrepreneur's utility would be below the payoff received from investing in the storage technology, $U_i < A$, whenever $q_i > q_{\min}$. Therefore, (A.2) rules out shirking in equilibrium. ■

Proof of Proposition 2. We first derive the equilibrium financing subsidy. Equating the respective multipliers in (9) for unsustainable and sustainable production yields the condition:

$$\frac{1}{\frac{B}{\Delta p} - [P(q) - c(\theta_m)]} = \frac{1 + r(\theta_s)}{\frac{B}{\Delta p} - [P(q) - c(\theta_s)(1 + r(\theta_s)]}$$

which implies (11) using $r(\theta_s) = -\Delta r^*$.

To provide q_s^* units of sustainable output, each firm demands $c(\theta_s) Ak^* - A$ of capital from outside investors while coinvesting A. Given Δr^*, the total fraction of supplied sustainable capital is $K(1 - F(\Delta r^*))$. The equalization of demand and supply in equilibrium, see condition (7), thus implies that

$$c(\theta_s) Ak^* - A = K(1 - F(\Delta r^*))$$

Solving for A yields the aggregate amount of internal funds provided by all sustainable firms in equilibrium,

$$A_s = \frac{K(1 - F(\Delta r^*))}{c(\theta_s) k^* - 1}.$$

Therefore, sustainable output, $q_s^* = A_s k^*$, is given by

$$q_s^* = K(1 - F(\Delta r^*)) \frac{k^*}{c(\theta_s) k^* - 1}. \tag{A.4}$$

Since aggregate output is given by $q^* = \min\{\bar{q}, Ak^*\}$ (by Proposition 1) and total sustainable output is given by (12), the ratio is given by (13). The condition on the short supply

[31] Because the manager is protected by limited liability, i.e., $D_i \leq P(q)$, there may be an upper bound on the feasible quantity q_i, but this is irrelevant for our argument.

of sustainable capital requires that $\omega^* < 1$, i.e.,

$$K\left(1 - F\left(\Delta r^*\right)\right) \frac{k^*}{c\left(\theta_s\right) k^* - 1} < \min\left\{\bar{q}, Ak^*\right\},$$

which is always satisfied if the amount of capital owned by investors with sustainability concerns, K, is sufficiently small. ■

Proof of Corollary 1. The invariance of Δr^* follows from (11). Equilibrium output q^* and, hence, the multiplier k^* are also unaffected by $F(w)$ as a result of Proposition 1. A first-order stochastic dominance shift of the distribution $F(w)$ must now decrease $F(\Delta r^*)$ and, hence increases q_s^*. The same holds for an increase in K. ■

Proof of Corollary 2. This follows directly from the fact that Δr^* is strictly increasing in Δc, see (11). Formally, $\frac{\partial \Delta r^*}{\partial \Delta c} > 0$ because $P(q^*) - \frac{B}{\Delta p} > 0$ which follows from (IC). ■

Proof of Corollary 3. We again distinguish based on whether aggregate output is constrained by financial frictions:

Case 1) If output is unconstrained, then $P(q^*) = c(\theta_m)$, see (4). An increase in $c(\theta_m)$ decreases aggregate output q^* because $P'(q) < 0$. We now turn to sustainable output q_s^*. Differentiating (A.4) yields

$$\frac{\partial q_s^*}{\partial \theta_m} = K\left(1 - \frac{\partial \Delta r^*}{\partial \theta_m} f\left(\Delta r^*\right)\right) \frac{k^*}{c\left(\theta_s\right) k^* - 1} > 0,$$

where $\frac{\partial \Delta r^*}{\partial \theta_m} < 0$ follows from (11). In this case the capacity multiplier $k^* = \frac{1}{B/\Delta p}$ is just a constant and, is, hence unaffected by changes in θ_m.

Case 2) We now consider the case, where the economy is constrained so that q^* is determined from $Ak(q) = q$ where $k(q) = \left(\frac{B}{\Delta p} - \left[P(q) - c(\theta_m)\right]\right)^{-1}$. The implicit function theorem implies that

$$\frac{\partial q^*}{\partial \theta_m} = \frac{A \frac{\partial k(q)}{\partial \theta_m}}{1 - Ak'(q)} < 0, \tag{A.5}$$

where the sign follows from $k'(q) < 0$ (so that the denominator is positive) and the negative sign of the numerator, $\frac{\partial k(q)}{\partial \theta_m} < 0$, follows from

$$\frac{\partial k(q)}{\partial \theta_m} = -\frac{c'(\theta_m)}{\left(\frac{B}{\Delta p} - \left[P(q) - c(\theta_m)\right]\right)^2} < 0.$$

Let us now consider the effect on $q_s^* = \frac{K(1-F(\Delta r^*))}{c(\theta_s) - \frac{1}{k^*}}$. We obtain:

$$\frac{\partial q_s^*}{\partial \theta_m} = \frac{\left[c\left(\theta_s\right) - \frac{1}{k^*}\right] K\left(1 - \frac{\partial \Delta r^*}{\partial \theta_m} f\left(\Delta r^*\right)\right) - \frac{K(1-F(\Delta r^*))}{k^{*2}} \frac{\partial k^*}{\theta_m}}{\left(c\left(\theta_s\right) - \frac{1}{k^*}\right)^2} > 0,$$

which follows from $c(\theta_s) > \frac{1}{k^*}$ and $\frac{\partial k^*}{\theta_m} < 0$ follows from $\frac{\partial q^*}{\partial \theta_m} < 0$, see (A.5), and $q^* = Ak^*$.

Since, in both cases, q_s^* is strictly increasing and q^* is strictly decreasing in θ_m, the share $\omega^* = \frac{q_s^*}{q^*}$ strictly increases. ■

Proof of Propositions 3 and 4. Recall that the social planner's objective function reduces to that of maximizing

$$q_s^*(\theta_s)\, v(\theta_s)\,, \tag{A.6}$$

where

$$v(\theta_s) = [\rho(\theta_s - \theta_m) - (c(\theta_s) - c(\theta_m))]\,.$$

Proposition 3 follows immediately from the observation that $v'(\theta_s)$ is strictly positive $\theta_s < \hat{\theta}$ and strictly negative for all $\theta_s > \hat{\theta}$.

If $\theta_m < \hat{\theta}$, the first-order condition for θ_s is given by

$$\frac{\partial q_s^*}{\partial \theta_s} v(\theta_s) + q_s^*(\theta_s)\, v'(\theta_s) = 0. \tag{A.7}$$

which can be transformed to obtain (18) in Proposition 4, where $\frac{\partial q_s^*}{\partial \theta_s} < 0$ follows from Corollary 2. As a result $v'(\theta_s) > 0$ at the optimum so that $\theta_s^* < \hat{\theta}$. Moreover, $\theta_s^* > \theta_m$ so that $v(\theta_s) > 0$. ■

Proof of Corollary 4. The left hand side of the first-order condition (18) is independent of the distribution $F(w)$, so that we only need to examine the right hand side. Recall that $q_s^* = K(1 - F(\Delta r^*))\frac{k^*}{c(\theta_s)k^* - 1}$ so that:

$$\begin{aligned} \left|\frac{\partial q_s^*}{\partial \theta_s}/q_s^*\right| &= -\frac{-Kc'(\theta_s)\, f(\Delta r^*)\frac{\partial \Delta r^*}{\partial \Delta c}\frac{k^*}{c(\theta_s)k^*-1} - K(1 - F(\Delta r^*))\frac{k^{*2}c'(\theta_s)}{(c(\theta_s)k^*-1)^2}}{K(1 - F(\Delta r^*))\frac{k^*}{c(\theta_s)k^*-1}} \\ &= c'(\theta_s)\left(\frac{f(\Delta r^*)}{1 - F(\Delta r^*)}\frac{\partial \Delta r^*}{\partial \Delta c} + \frac{k^*}{c(\theta_s)\, k^* - 1}\right) \end{aligned} \tag{A.8}$$

By assumption of monotonicity of the hazard rate, $\frac{f(\Delta r^*)}{1-F(\Delta r^*)}$ decreases implying that $\left|\frac{\partial q_s^*}{\partial \theta_s}/q_s^*\right|$ strictly decreases in the interior of the support. Since a monotone hazard rate shift, thus, strictly decreases the right hand side of (18), the left hand side must be smaller at the optimum, too. Since, the left hand side is a strictly decreasing function of θ_s, $\frac{\partial^2 \ln v^*(\theta_s)}{\partial \theta_s^2} = -\frac{v'(\theta_s)^2 + v(\theta_s)c''(\theta_s)}{v(\theta_s)^2} < 0$, we, thus, obtain that θ_s^* must strictly increase. ■

Proof of Corollary 5. Applying the implicit function theorem on (A.7) and exploiting the second-order condition of θ_s^* implies that the sign of $\frac{d\theta_s^*}{d\rho}$ is determined by the sign of

$$\begin{aligned}\frac{\partial}{\partial \rho}\left[\frac{\partial q_s^*}{\partial \theta_s} v(\theta_s) + q_s^*(\theta_s) v'(\theta_s)\right] &= \frac{\partial q_s^*}{\partial \theta_s}(\theta_s - \theta_m) + q_s^*(\theta_s) \\ &= -\frac{\partial q_s^*}{\partial \theta_s}\frac{c'(\theta_s^*)(\theta_s^* - \theta_m) - [c(\theta_s^*) - c(\theta_m)]}{\rho - c'(\theta_s^*)} > 0\end{aligned}$$

where the second line follows from substituting $q_s^*(\theta_s) = -\frac{\partial q_s^*}{\partial \theta_s}\frac{\rho[\theta_s^* - \theta_m] - [c(\theta_s^*) - c(\theta_m)]}{\rho - c'(\theta_s^*)}$, which follows from rearranging first-order condition (A.7) and using the definition of v. The positive sign in line 2 follows from $c'(\theta_s^*) < \rho$ (by $\theta_s^* < \hat{\theta}$), $\frac{\partial q_s^*}{\partial \theta_s} < 0$, and strict convexity of $c(\theta)$ (together with $\theta_s^* > \theta_m$ and thus also $c'(\theta_s^*) > c'(\theta)$ for all $\theta_m \leq \theta < \theta_s^*$). ■

Proof of Lemma 2. The result follows from Lemmas A.2 to A.4. ■

Lemma A.2 *If $\frac{B}{\Delta p} \geq \rho(\theta^{\max} - \hat{\theta})$, there exists a unique level for internal funds A_{FB} such that the planner can achieve first-best by setting the minimum standard to $\theta_m = \hat{\theta}$. For $A = A_{FB}$, financial constraints limit aggregate output, i.e., $A_{FB} < \bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p}$.*

Proof of Lemma A.2. Recall that first-best requires that all firms choose the sustainability level $\hat{\theta}$ and aggregate output satisfies $q^* = q_{FB}$ (see Proposition 1). The sustainability choice of $\hat{\theta}$ can only be ensured by setting the minimum standard to $\theta_m = \hat{\theta}$. We now prove that there exists a level of internal funds $A_{FB} < \bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p}$ such that this policy choice, $\theta_m = \hat{\theta}$, also ensures the first-best output $q^* = q_{FB}$.

We first show that financial constraints must constrain aggregate output, i.e., $A_{FB} < \bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p}$. Suppose to the contrary that $A > \bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p}$ and the regulator were to set $\theta_m = \hat{\theta}$, then financial constraints would not affect aggregate output so that $q^* = \bar{q}\left(\hat{\theta}\right)$. In this case, there is overproduction, $q^* > q_{FB}$.
We now prove that for some $A \in \left[0, \bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p}\right]$ first-best can be achieved as long as $\frac{B}{\Delta p} \geq \rho(\theta^{\max} - \hat{\theta})$. By Proposition 1, output is characterized by the fixed point

$$q^* = \frac{A}{\frac{B}{\Delta p} - \left[P(q^*) - c\left(\hat{\theta}\right)\right]}. \tag{A.9}$$

(A.9) implies that aggregate output q^* is a continuous, strictly increasing function of A over the domain $A \in \left[0, \bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p}\right]$ with range $\left[q_{\min}\left(\hat{\theta}\right), \bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p}\right]$ where $q_{\min}$ refers to the output that can be produced with zero internal funds. Given continuity and strict monotonicity as well as $\bar{q}\left(\hat{\theta}\right)\frac{B}{\Delta p} > q_{FB}$, it thus, suffices that $q_{\min} < q_{FB} = P^{-1}\left(c(\hat{\theta}) + \rho(\theta^{\max} - \hat{\theta})\right)$. If $q_{\min} = 0$, see definition in (A.1), the result is immediate since first-best output q_{FB} is by assumption positive. If the minimum output satisfies $q_{\min} = P^{-1}\left(c\left(\hat{\theta}\right) + \frac{B}{\Delta p}\right)$, then $q_{\min} < q_{FB}$ if and only if $\frac{B}{\Delta p} > \rho(\theta^{\max} - \hat{\theta})$, which proves A.2. ■

Lemma A.3 *In the absence of financial constraints, the optimal policy minimum standard satisfies $\theta_m^* = \bar{\theta}_m > \hat{\theta}$, where $\bar{\theta}_m$ solves:*

$$\rho - c'(\theta_m) = -\left|\frac{\frac{d\bar{q}(\theta_m)}{d\theta_m}}{\bar{q}(\theta_m)}\right| \rho\left(\theta^{\max} - \theta_m\right).$$

Proof of Lemma A.3. Without financial constraints, equilibrium output satisfies $q^* = \bar{q}(\theta_m)$ so that welfare is given by:

$$\Omega = \int_0^{\bar{q}(\theta_m)} P(q)dq - \bar{q}(\theta_m)\left[\rho(\theta^{\max} - \theta_m) + c(\theta_m)\right],$$

Exploiting that $P(\bar{q}(\theta_m)) = c(\theta_m)$, the necessary first-order condition implies:

$$\rho - c'(\theta_m) = -\left|\frac{\frac{d\bar{q}(\theta_m)}{d\theta_m}}{\bar{q}(\theta_m)}\right| \rho\left(\theta^{\max} - \theta_m\right)$$

Since the right hand side is negative, the left hand side is too. Therefore, $c'(\theta_m) > \rho$ or, equivalently $\bar{\theta}_m > \hat{\theta}$. ■

Lemma A.4 *There exists a threshold $\bar{A}$ such that financial constraints bind (at the optimal choice) for all $A < \bar{A}$. In this case, the optimum minimum standard satisfies.*

$$\rho - c'(\theta_m) = \left|\frac{\frac{dq^*}{d\theta_m}}{q^*(\theta_m)}\right| \left[P(q^*) - c(\theta_m) - \rho\left(\theta^{\max} - \theta_m\right)\right].$$

Proof of Lemma A.4. Take a level of internal funds $\tilde{A}$ for which aggregate output is not impacted by financial constraints (under optimal regulation) so that the optimal minimum standard is $\theta_m^* = \bar{\theta}_m$ (by Lemma A.3). Then, this must also be the optimal minimum standard for any $A > \tilde{A}$, which allows the planner to achieve the same payoff as with assets $\tilde{A}$. As a result, the set of internal funds for which financial constraints do not affect aggregate output (under the optimal choice of $\theta_m^* = \bar{\theta}_m$) is connected. We denote the lowest value of this set as $\bar{A}$ (see characterization below).

As a result, for $A < \bar{A}$, financial constraints constrain aggregate output under the optimal policy. The necessary first-order condition for welfare now satisfies

$$\frac{dq^*}{d\theta_m}\left[P(q^*) - c(\theta_m) - \rho\left(\theta^{\max} - \theta_m\right)\right] + q^*(\theta_m)\left[\rho - c'(\theta_m)\right] = 0. \tag{A.10}$$

This is equivalent to condition (19), which we restate:

$$\rho - c'(\theta_m) = \left| \frac{\frac{dq^*}{d\theta_m}}{q^*(\theta_m)} \right| \left[P(q^*) - c(\theta_m) - \rho(\theta^{\max} - \theta_m) \right]. \tag{A.11}$$

It now follows from (A.11) that $c'(\theta_m^*) \geq \rho$, i.e., $\theta_m^* \geq \hat{\theta}$, if and only if

$$G(\theta_m^*, A) := \rho(\theta^{\max} - \theta_m^*) - [P(q^*(\theta_m^*, A)) - c(\theta_m^*)] \geq 0. \tag{A.12}$$

As proven in Lemma A.2 there exists a unique level A_{FB} for which we obtain first-best, i.e., $\rho - c'(\theta_m^*) = 0$ and $G(\theta_m^*, A_{FB}) = 0$ with $\theta_m^* = \hat{\theta}$.

We now evaluate the derivative $\left.\frac{d\theta_m^*}{dA}\right|_{A=A_{FB}}$ *locally* for $A = A_{FB}$. Using the implicit function theorem and the second-order condition, the sign is determined by

$$\left.\frac{\partial G(\theta_m^*, A)}{\partial A}\right|_{A=A_{FB}} = -P'\left(q^*\left(\hat{\theta}, A\right)\right) \frac{dq^*\left(\hat{\theta}, A\right)}{dA} > 0 \tag{A.13}$$

The sign follows from $P' < 0$ and the fact that for $A \leq \bar{A}$ larger internal funds strictly increase output, $\frac{dq^*\left(\hat{\theta}, A\right)}{dA} > 0$.

Since $\theta_m^*(A)$ is continuous for $A \leq \bar{A}$ and A_{FB} is unique (by Lemma A.2), the positive local derivative $\left.\frac{d\theta_m^*}{dA}\right|_{A=A_{FB}} > 0$ also implies that $\theta_m^* > \hat{\theta}$ for $A > A_{FB}$ and $\theta_m^* < \hat{\theta}$ for $A < A_{FB}$.

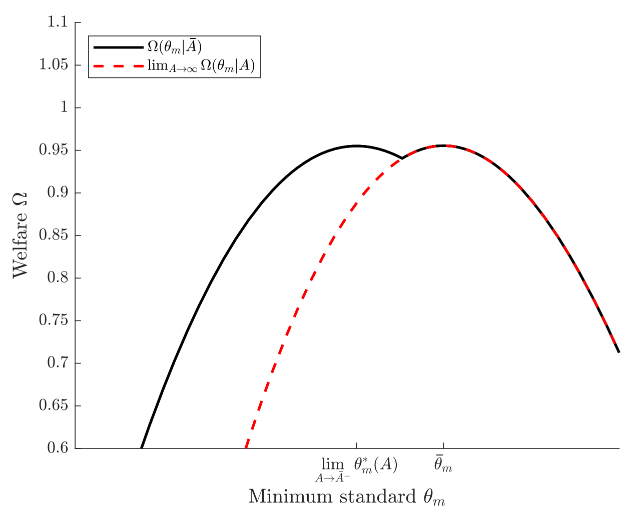

Figure 3. Discontinuity of policy function. The black graph plots welfare $\Omega(\theta_m| A)$ as a function of the minimum standard θ_m conditional on internal funds $A = \bar{A}$. The function has two global maxima. The red dashed line plots the welfare function in the absence of financial constraints, i.e., internal funds approaching infinity.

The threshold $\bar{A}$ is characterized by the condition that the planner is indifferent between setting a low value of the minimum standard $\theta_m^* = \lim_{A\to\bar{A}^-}\theta_m^*$ (so that financial constraints constrain output $q^* < \bar{q}(\theta_m)$) or setting a high value of the minimum standard $\bar{\theta}_m$ so that aggregate output is not constrained by financial constraints and given by $q^* = \bar{q}(\bar{\theta}_m)$. (Recall that $\bar{q}$ is a strictly decreasing function of θ_m). Formally, the objective function $\Omega(\theta_m|A)$ has two global maxima for $A = \bar{A}$ (see Figure 3). ■

Proof of Proposition 5. As the social planner can now avail herself of both instruments, her program is to choose θ_m and potentially a sustainable investment category with threshold $\theta_s > \theta_m$ so as to maximize Ω in (17). The optimal outcome is denoted by $(\theta_m^{**}, \theta_s^{**})$. We note first that the derivative with respect to θ_s is identical to that in the partial problem of Proposition 3, where we took θ_m as given. It is thus strictly optimal to introduce a sustainable investment category $\theta_s^{**} > \theta_m^{**}$ if $\theta_m^{**} \geq \hat{\theta}$ and strictly suboptimal otherwise. It thus remains to show that, also in the presence of both potential instruments, $\theta_m^{**} \geq \hat{\theta}$ ($\theta_m^{**} < \hat{\theta}$) if $A \geq A_{FB}$ ($A < A_{FB}$). We show this by contradiction.

Suppose that $A < A_{FB}$ and suppose, instead that $\theta_m^{**} \geq \hat{\theta}$. Then, it would not be optimal to introduce a sustainable investment category with $\theta_s^{**} > \theta_m^{**}$, in which case we know however from Lemma 2 that $\theta_m^{**} = \theta_m^* < \hat{\theta}$, a contradiction.

For $A = A_{FB}$ the argument is immediate as $\theta_m^{**} = \hat{\theta}$ achieves the first best, while Ω is strictly lower for any choice $\theta_m^{**} < \hat{\theta}$, irrespective of the choice of θ_s.

Turning finally to $A > A_{FB}$, suppose that $\theta_m^{**} < \hat{\theta}$, in which case it would be optimal to choose a threshold $\theta_m^{**} < \theta_s^{**} < \hat{\theta}$. To show that this is not optimal, it is sufficient to argue that welfare is strictly higher by setting instead $\widetilde{\theta}_m = \theta_s^{**}$, without a sustainable investment category. Denote the thereby realized value by $\widetilde{\Omega}$,

$$\widetilde{\Omega} = \int_0^{\widetilde{q}^*} P(q)dq - \rho\widetilde{q}^* \left[c\left(\widetilde{\theta}_m\right) + \rho(\theta^{\max} - \widetilde{\theta}_m)\right].$$

Calculating now the difference $\widetilde{\Omega} - \Omega$, observe the two differences: first, while quantity q_s^* was previously and now generated at standard θ_s^{**}, quantity $\widetilde{q}^* - q_s^*$ is now produced at θ_s^{**} and no longer at θ_m^{**}; second, total output is reduced from q^* to $\widetilde{q}^*$. With this decomposition we have $\widetilde{\Omega} - \Omega = \Delta_1 + \Delta_2$ with

$$\Delta_1 = (\widetilde{q}^* - q_s^*)\left[\rho(\theta_s^{**} - \theta_m^{**}) - \left[c(\theta_s^{**}) - c(\theta_m^{**})\right]\right],$$

which is strictly positive from $\theta_m^{**} < \theta_s^{**} < \hat{\theta}$, and

$$\Delta_2 = -\int_{\widetilde{q}^*}^{q^*} \left[P(q) - \left[c(\theta_m^{**}) + \rho(\theta^{\max} - \theta_m^{**})\right]\right] dq,$$

which is strictly positive when, at the lower boundary,

$$P(\widetilde{q}^*) - \left[c(\theta_m^{**}) + \rho(\theta^{\max} - \theta_m^{**})\right] < 0.$$

But this follows immediately by construction, as with $A > A_{FB}$ the marginal social return is negative at least for all $\theta_m \leq \hat{\theta}$, and thus also for $\widetilde{\theta}_m = \theta_s^{**} < \hat{\theta}$ (and the corresponding quantity $\widetilde{q}^*$). ■

Proof of Corollary 6. We prove that A_{FB} is strictly increasing in the severity of the agency problem, $a = \frac{B}{\Delta p}$. For this we consider the definition of A_{FB} in (A.12), with $G(\theta_m^*, A_{FB}) = 0$ and $\theta_m^* = \hat{\theta}$

$$\frac{dA_{FB}}{da} = \frac{-P'(q^*)\frac{dq^*}{da}}{-P'(q^*)\frac{dq^*}{dA}} = -\frac{dq^*/da}{dq^*/dA} > 0,$$

which uses $\frac{dq^*}{dA} > 0$ and $dq^*/da < 0$ under constrained output (using equation (10) in Proposition 1, $q^* = Ak(q^*)$). ■

B Complementary material on Pigouvian taxes

We first extend the derivation of a market equilibrium to the case with a tax on externalities. For this it is expedient to denote the marginal costs of production, now including taxes on externalities, by $\widetilde{c}(\theta_i) = c(\theta_i) + \tau(\theta^{\max} - \theta_i)$. The firm thus needs to raise now the amount $\widetilde{c}(\theta_i)\, q_i - A$, and with this modification the incentive and participation constraints, (IC) and (IR) respectively, remain unchanged. The entrepreneur's payoff becomes likewise

$$U_i = q_i P(q) - [(\widetilde{c}(\theta_i) q_i - A)(1 + r(\theta_i))].$$

The entrepreneur chooses θ_i to maximize U_i. When the choice of θ_i does not affect financing conditions, it minimizes costs $\widetilde{c}(\theta_i)$. So as not to overburden notation, we slightly abuse notation and denote the unique maximizer by θ_m, i.e., θ_m is now pinned down by the tax τ, according to $\widetilde{c}'(\theta_m) = 0$ (or likewise $c'(\theta_m) = \tau$). Obviously, the introduction of a sustainable investment category that exceeds the prevailing standard again requires that $\theta_s > \theta_m$, and we implicitly stipulate that when setting θ_s, the policymaker knows the (minimum) standard θ_m that her choice of the tax on externality implements. Given $r(\theta_i) = r_0 - \Delta_r$ for all $\theta_i \geq \theta_s$ and $\theta_s > \theta_m$, a firm that wants to tap into cheaper sustainable financing, will again optimally choose $\theta_i = \theta_s$. The equilibrium interest differential, Δr^*, adjusts so as to make firms indifferent between these two choices, where now the cost differential in the definition of Δr^* includes tax payments, so that $\Delta c = \widetilde{c}(\theta_s) - \widetilde{c}(\theta_m)$. Finally, equilibrium output again depends on whether the economy is financially constrained in the aggregate. Then, the respective size $q^* = \bar{q}$ solves $P(\bar{q}) = \widetilde{c}(\theta_m)$. If constrained, we have $k^* := k(q^*)$, where the multiplier adjusts by use of $\widetilde{c}$. Hence, Proposition 1 for the characterization of q^* again applies, replacing c by $\widetilde{c}$. This is also the unique adjustment necessary for the characterization of q_s^* in Proposition 2.

Lemma A.5 *The characterization of the equilibrium with a minimum standard in Propositions 1 and 2 extends to the case with a tax on externalities, after augmenting the costs of production by the respective tax payments, i.e., replacing $c(\theta)$ by the expression $\widetilde{c}(\theta) = c(\theta) + \tau(\theta^{\max} - \theta)$, and using $\widetilde{c}'(\theta_m) = 0$.*

As in the case with a minimum standard in Proposition 3, we first solve for the optimal threshold for the sustainable investment category for a given choice of the tax on externalities. For this note that the social planner's objective function is still given by (17), where θ_m is now a function of the potential instrument τ.[32]

Lemma A.6 (Optimal sustainable investment classification under a tax) *Take now a tax $\tau \leq \rho$ on externalities as given. If τ lies below the Pigouvian level, $\tau < \rho$, it is strictly optimal to introduce a sustainable investment category. If $\tau \geq \rho$ this is strictly suboptimal.*

Proof of Lemma A.6. Holding now the tax fixed, the social planner's objective is to maximize (A.6), where θ_m (as part of $v(\theta_s)$) is obtained from $\widetilde{c}'(\theta_m) = 0$). This implies that optimally $q_s^* = 0$ when $\tau = \rho$ (or even larger). Otherwise, the first-order condition is given by (A.7), with the only difference that now

$$\frac{\partial q_s^*}{\partial \theta_s} = -K\left(\frac{k^*}{k^*-1}\right)^2 g\left(\Delta r^*\right)\widetilde{c}'\left(\theta_s\right) < 0 \tag{B.14}$$

when $\theta_s < \theta_m$. With $\theta_m < \hat{\theta}$ it is then again strictly optimal to set $\theta_m < \theta_s^* < \hat{\theta}$ and with this $q_s^* > 0$. **Q.E.D.** ■

We turn next to the combination of the two instruments. Here, we refer to the intuition that we have already laid out in the main text. As described there, now the key case distinction refers to whether the economy is financially constrained in the aggregate when the Pigouvian level $\tau = \rho$ is chosen, taking into account the additional tax payments $\tau(\theta^{\max} - \hat{\theta})$.

Proof of Proposition 6. Suppose the social planner sets the Pigouvian tax level $\tau = \rho$ with respective marginal costs $\widetilde{c}(\theta_i) = c(\theta_i) + \rho(\theta^{\max} - \theta_i)$. From the equilibrium characterization we know that there exists a level of A, denoted by $\widehat{A}$, so that, with this choice, the economy is not financially constrained if and only if $A \geq \widehat{A}$. In this case the first best is realized, as $\widetilde{c}'(\theta_m) = 0$ and thus $c'(\theta_m) = \theta_m$ and as $q^* = \bar{q} = q_{FB}$ given $P(\bar{q}) = \widetilde{c}(\theta_m)$ (which equals the marginal social costs when $\tau = \rho$). Introducing a sustainable investment classification can thus not improve welfare, and it is from Proposition A.6 also suboptimal.

Take next the case with $A < \widehat{A}$. We argue to a contradiction, supposing thus that it was instead optimal to choose $\tau^{**} \geq \rho$, implying $\theta_m \geq \hat{\theta}$, in which case a sustainable investment category would be suboptimal from Proposition A.6. We show now that without a sustainable investment category, it is however strictly optimal to set $\tau^{**} < \rho$. To see this note that, when $q_s^* = 0$,

$$\frac{d\Omega}{d\tau} = \frac{dq^*}{d\tau}\left[P(q^*) - c\left(\theta_m\right) - \rho\left(\theta^{\max} - \theta_m\right)\right] + \frac{\partial \theta_m}{\partial \tau} q^*[\rho - c'(\theta_m)].$$

At $\tau = \rho$, $\rho - c'(\theta_m) = 0$, while with binding financial constraints, $P(q^*) > \widetilde{c}(\theta_m) = c(\theta_m) + \rho(\theta^{\max} - \theta_m)$, so that together with $\frac{dq^*}{d\tau} < 0$ the first part is strictly negative. That indeed $\tau^{**} < \rho$ follows then from the presumed strict quasiconcavity. ■

[32] We thus treat the tax as a welfare neutral transfer.

We finally discuss the case where the social planner can avail herself of all three instruments, in particular a minimum standard next to a tax on externalities. We choose an informal discussion, notably so as not to impose further notation. A key observation is that now it is feasible to achieve the first-best outcome for all $A \geq A_{FB}$. Recall that with only a minimum standard, this was feasible only when $A = A_{FB}$ ($\theta_m = \hat{\theta}$), while with only a tax, it was feasible for $A \geq \widehat{A}$ (setting $\tau = \rho$), where $A_{FB} < \widehat{A}$. To now cover as well the range $A \in (A_{FB}, \widehat{A})$, the social planner leaves the minimum standard at $\theta_m = \hat{\theta}$ and introduces a tax τ that, moving from A_{FB} to $\widehat{A}$, is gradually increased from zero to ρ. As is immediate, with $\theta_m = \hat{\theta}$ and $\tau < \rho$, firms' actual standard is pinned down by the minimum standard, and not by the tax. The additional tax burden, however, affects profitability and thus the (otherwise excessive) output q^*. From $A \geq \widehat{A}$ onwards, the minimum standard at $\theta_m = \hat{\theta}$ is redundant. And when $A < A_{FB}$, output is not excessive, and the additional tax is not welfare improving, in contrast to the introduction of a sustainable investment category.

Lemma A.7 (Minimum standard plus a tax) *Suppose the social planner can choose both a minimum standard and a tax, as well as a sustainable investment category. Then introducing the latter only improves welfare when $A < A_{FB}$ and it is otherwise strictly suboptimal.*