BGPE Discussion Paper

No. 242

A Signaling Game for Green Bonds

Fabian Alex

August 2025

ISSN 1863-5733

Editor: Prof. Regina T. Riphahn, Ph.D.
Friedrich-Alexander-Universität Erlangen-Nürnberg

B G P E Discussion Paper Series

A Signaling Game for Green Bonds*

Fabian Alex†

University of Regensburg

Department of Economics

93 040 Regensburg, Germany

Abstract

We build a signaling game model of a firm's decision to acquire a costly green label which enables it to emit a green bond. A greenium may compensate it for the incurred cost. That cost is higher for non-green firms. With an investor that prefers a clean environment and dislikes being fooled into believing in a fabricated green label, there are equilibria featuring green bonds by either both firm types, only the green firm or neither. Allowing side payments undermines stability of all equilibria where a green label is acquired. A neutral rather than a green investor considerably decreases the number of conceivable equilibria, as does uncertainty about the investor type. The equilibria of the baseline model are preserved if we allow two investors, a green and a neutral one, to decide on their respective purchase of the bond sequentially. Lastly, if investors hold all market power, no green labels will be observed at all.

JEL classification: C72, D21, Q5

Key words: Environment, Environmental Economics, Green Economics, Game Theoretic, Game Theory, Two Player, Strategic Game, Signaling Game

*For helpful comments I am indebted to Lutz Arnold. Declarations of interest: none.

†Phone: +49 941 943 2704; e-mail: fabian.alex@ur.de.

1 Introduction

The current green finance literature experiences an ongoing discussion on whether green bonds are priced at a premium, i.e., enable cheaper financing for environmentally sound issuers. Larcker & Watts (2020) provide empirical evidence that the municipal bond market does not feature such a "greenium". Baker et al. (2022) do find a greenium within the same sample. This discrepancy is explained by the varying methodology of both papers: while Baker et al. (2022) compare bonds with similar characteristics in identifying the green bond price differential, Larcker & Watts (2020) require the issuing entity to be identical. The validity of the latter approach may be contested on the grounds of a lack of attributability of bonds to projects: investors will value municipal or corporate greenness, not bond greenness (see also Krahnen et al., 2023).

There are various theoretical reasons to expect a positive greenium. Two out of those are most closely related to our work. For one, investor tastes, dating back to Fama & French (2007), explain voluntary waiver of returns on behalf of socially conscious investors. The other argument rests on the fact that a firm cannot simply declare itself (or any of its projects) to be "green". Rather, it has to verify this state of affairs at a cost. Naturally, no firm would do so if it could not expect any sort of compensation. Empirical evidence indeed suggests that a trustworthy green label is what leads to a positive greenium (Dorfleitner et al., 2022). Other arguments in favor of greenia encompass a hedge against future climate policy restrictions (see Bolton & Kacperczyk, 2021) or simply lower overall riskiness of green projects (see Neagu et al., 2024 on financially sound firms' better ability to issue green bonds and Hoepner et al., 2024 on the reduced tail risk of green firms).

Flammer (2021) argues that "the stock market responds positively to the issuance of green bonds" (p. 507): the cumulative abnormal return of firms issuing a green bond is 0.49 percentage points (pp). This effect is further enhanced to 0.71 pp for certified rather than non-certified green bonds. She interprets these observations

as a signaling effect of green bond issuance, indicating both regards for the environment and a future hedge against climate risks. Consequently, the ownership of such firms by green and long-term investors increases, predominantly so in the presence of a third-party green certification (Flammer, 2021, pp. 512-513). For our modeling approach, it is irrelevant where the additional profits following a certification come from. Whether they are due to a greenium and, hence, reduced costs of capital (an argument disclaimed by Flammer, 2021, see p. 514) or stem from elevated stock demand does not matter as long as the certification cost pays off in some way.

We build a signaling game in line with the seminal contribution by Spence (1973). Comparable applications of signaling games on a firm's decision whether to issue a green bond encompass Yu & Jihan (2022) and Zhu et al. (2023). In Yu & Jihan (2022), environmentally sound firms are left with no choice but to send a "strong" signal of greenness, hence violating the original signaling game structure. As there is then always a chance for a green bond to stem from a green firm by construction, the number and significance of interesting equilibria is thus limited. Zhu et al. (2023) have a strong emphasis on greenwashing. They build a model where nature determines a firm to be a greenwasher or not. The firm then only faces the decision of whether to back up its image via third-party-certification or not. If our model, on the other hand, features greenwashing, it is by rational decision making rather than by coincidence.

The remainder of the paper is organized as follows. Section 2 sets up the model. Section 3 derives a critical level of investor beliefs that incentivizes investment, which will re-occur frequently throughout the paper. Using that, various equilibrium structures are explored in Sections 4 and 5. Section 6 refines these equilibria to collapse them into a smaller number of *perfect Bayesian equilibria* (PBEs) with a notion of *trembling hand perfectness*. Section 7 compares the model to one with a neutral investor and also introduces uncertainty about the investor type. In Section 8, we explore the case of sequential bond offers to two different

investor types. Market power is reversed in Section 9. Finally, Section 10 offers conclusions.

2 Model

The model used is a classical signaling game where nature draws a type for the sender at random who then emits a message to the receiver (see, for example, Section 4.2.A in Gibbons, 2011). Here, the sender is depicted to be a firm while the receiver is a sufficiently large (potentially institutional) investor that the firm hopes to get financed by. It tries to do so by emitting a bond. The investor has both financial and environmental interests, the latter reflecting a sense of social responsibility. The model can readily be thought of as investors operating through a fund manager that is commissioned to adequately reflect the preferences of her investor base.

Nature randomly assigns a type $t \in T = \{t_g, t_b\}$ to the firm, where t_g and t_b refer to "green" and "brown", respectively. The chance of being t_g is denoted p_g. Depending on its type, the firm can generate a gross revenue of R_g or $R_b > R_g$ for an input of D units of capital. Aside from the random assignment of firm types, there is no uncertainty.[1]

The firm observes which t it has been ascribed and then enters the debt market. Since the firm is assumed to command no initial net worth, it has to raise D units of capital to operate its project irrespective of its type. Depending on the kind of bond it places, the repayment of its debt varies. The bond may be "green", promising a percentage return of r_g, or "conventional" (which we set equal to "brown" for the sake of simplicity), in which case it pays the higher return $r_b > r_g$. Both interest rates are exogenous. Irrespective of the chosen mode of financing, both firms' projects have a positive net present value and are, thus, always worth being conducted. A sufficient condition to ensure this is $R_g - (1 + r_b)D > 0$. To qualify

[1] While uncertainty is a crucial feature of financial markets as a whole, it is commonly neglected when dealing with debt markets exclusively (see also Arnold, 2025).

itself as green bond emittent, the firm must invest the amount V into verification. A t_b-firm obtaining a green certificate as fabrication rather than verification is assumed to pay an extra cost of C for counterfeiting. This may be thought of as resources spent on creative accounting or on the acquisition of fraudulent hard- or software as seen in the recent and ongoing VW diesel scandal (see Spiegel, 2024). Another intuitive interpretation of these costs are (expected) costs of being caught greenwashing. As an article by tagesschau (2025) shows for the case of the German company DWS, such punishments may be of considerable size, extending into the eight-digit range with ease. In our model, the bond's label is the message sent by the firm, $m \in M = \{m_g, m_b\}$.

The investor's choice is take-it-or-leave-it. Thus, her available actions are $A = \{a_i, a_n\}$, that is, to invest or not. She observes the bond, including its label, and decides whether to buy it or not. Doing so results in a net payoff of $r_g D$ or $r_b D$, respectively. It is assumed that she has at least D units of capital such that the bond can in fact be purchased. Any additional money could be invested in a different way, but does not affect investment behavior toward the specific firm under consideration and is, hence, ignored. After the bond is purchased, the firm's type is revealed to the lender. Conditional on investment, if $t = t_g$, the investor internalizes positive externality E. It results either from a tangibly improved environment or, more likely, from personal warm glow on behalf of the investor (or her base). However, if the investor turns out to have invested into a brown firm through a green bond, it receives a disutility of F for being deceived. Her outside option when choosing a_n is to buy some market portfolio, which pays an (exogenous) intermediate rate of $r_M \in (r_g, r_b)$ and offers no pro-social returns.

Both agents are assumed to be risk-neutral. So both their financial and non-financial payoffs translate one-to-one into utility and are comparable on an expected-value-basis. The firm's utility given its search for investment succeeds

is

$$U_{firm|a_i} = \begin{cases} R_g - (1+r_g)D - V & \text{if } t = t_g \wedge m = m_g \\ R_b - (1+r_g)D - V - C & \text{if } t = t_b \wedge m = m_g \\ R_g - (1+r_b)D & \text{if } t = t_g \wedge m = m_b \\ R_b - (1+r_b)D & \text{if } t = t_b \wedge m = m_b \end{cases}. \tag{1}$$

If its venture is fruitless, only the certification costs are borne:

$$U_{firm|a_n} = \begin{cases} -V & \text{if } t = t_g \wedge m = m_g \\ -V - C & \text{if } t = t_b \wedge m = m_g \\ 0 & \text{if } m = m_b \end{cases}. \tag{2}$$

The investor obtains

$$U_{investor|a_i} = \begin{cases} r_g D + E & \text{if } t = t_g \wedge m = m_g \\ r_g D - F & \text{if } t = t_b \wedge m = m_g \\ r_b D + E & \text{if } t = t_g \wedge m = m_b \\ r_b D & \text{if } t = t_b \wedge m = m_b \end{cases} \tag{3}$$

from investing. Note that the investor values actual environmental benefits E caused by the firm regardless of the issued bond-type. Hence, she may be regarded as consequentialist (see, for example, Broccardo et al., 2022, p. 3103). Inactivity yields a constant utility of

$$U_{investor|a_n} = r_M D. \tag{4}$$

Knowing about the payoffs to both players contingent on any possible state of the world allows us to illustrate the signaling game as in Figure 1, where we denote firm payoffs above investor payoffs at each terminal node. This game can now be solved for possible subgame perfect equilibria using backward induction. These will be refined to PBEs (and further) in Section 6.

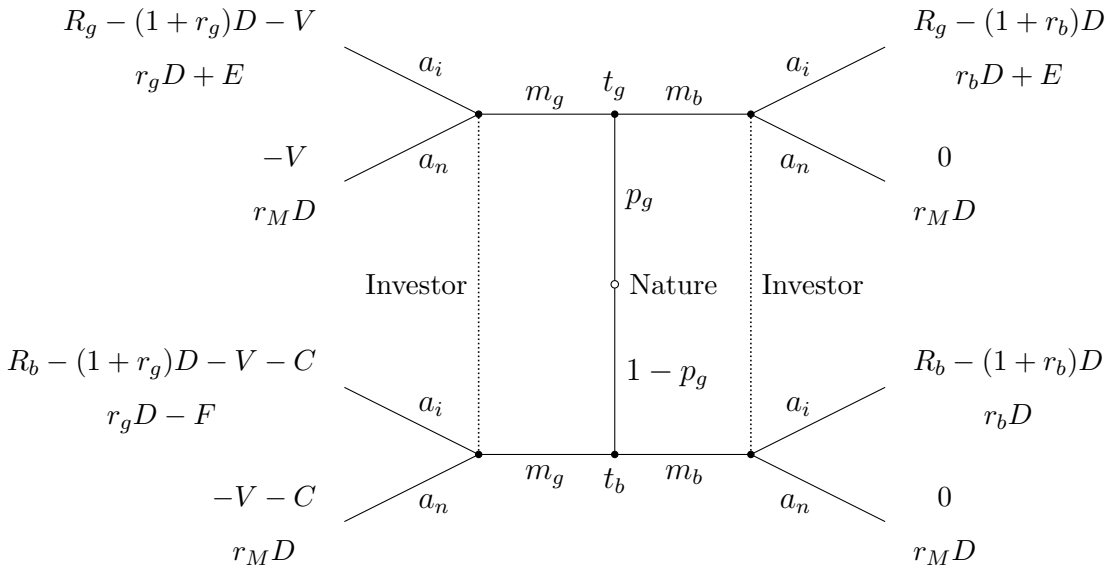

Figure 1: General green bond signaling game (extensive form)

3 Beliefs

The equilibria of the game under consideration have an important determinant other than pure payoffs, which is the investor's assessed probability of the firm being of a certain type given the message the latter sends out. As those subjective probabilities indicate the investor's presumed placement within her information set, they are commonly referred to as *beliefs*. We do not state them in the Figure for reasons of lucidity. We denote with $\mu_g \in [0,1]$ the investor's belief that the firm is of type t_g given it issues the signal m_g. Consequently, the investor thinks that she is in the north-western game node with probability μ_g if the bond has a green label, and that she is to the south-west with $1 - \mu_g$. Her corresponding expected utility from investing after observing a green signal is thus

$$E\left(U_{investor|a_i,m_g}\right) = \mu_g(r_gD + E) + (1 - \mu_g)(r_gD - F),$$

or

$$E\left(U_{investor|a_i,m_g}\right) = r_gD - F + \mu_g(E + F). \tag{5}$$

Similarly, a brown bond makes the investor assign probability μ_b to the firm being of the green type t_g and $1-\mu_b$ to it being t_b instead, placing him onto the north-eastern or south-eastern node, respectively. So expected investor utility for this situation is given by

$$E\left(U_{investor|a_i,m_b}\right) = \mu_b(r_b D + E) + (1-\mu_b) r_b D.$$

It simplifies to

$$E\left(U_{investor|a_i,m_b}\right) = r_b D + \mu_b E. \tag{6}$$

The constant utility of inactivity as indicated by (4) can be directly compared to these expected utility terms in decision making as the former involves no uncertainty and the investor is risk-neutral. The investment will always occur whenever (4) is not larger than (5) or (6), respectively. For the case of message m_g, this implies

$$r_g D - F + \mu_g(E+F) \geq r_M D,$$

or

$$\mu_g \geq \frac{(r_M - r_g)D + F}{E+F}. \tag{7}$$

Inequality (7) shows that the decision to invest in a green bond requires $E > (r_M - r_g)D$, that is, a sufficient level of green benefits to the investor. Otherwise, no conceivable value of μ_g will trigger demand for the investment. Whether or not this is the case hinges on the extent of the investor's pro-sociality.

Analogously, the brown bond is better than inactivity to the investor if

$$r_b D + \mu_b E \geq r_M D.$$

As this merely implies

$$\mu_b \geq \frac{-(r_b - r_M)D}{E},$$

where the RHS is negative, we arrive at the result that anything goes. In other words, this specific firm's regular bond without social benefits will always be

purchased. With our assumption that $r_b > r_M$, the intuition is blatantly clear: it offers an above-market rate of return and has the chance of providing warm-glow at no cost.

4 Separating equilibria

In a separating equilibrium, the sender reveals her type by the message she sends. Consequently, separating equilibria should always feature beliefs of 0 and 1 to be consistent with rational decision making, as a PBE requires. Nevertheless, we begin by assessing all valid beliefs for the various possible subgame perfect equilibrium structures and save refinements for Section 6. The used equilibrium concept for now is thus the notion of subgame perfect Nash equilibrium introduced by Selten (1965): equilibrium strategies must constitute a Nash-equilibrium in every subgame.[2]

4.1 Truthful type revelation

The first conceivable type-revealing equilibrium is straightforward. Assume a t_g-firm decides to incur the cost of certification for issuing a green bond (m_g), while a t_b-firm does not (m_b).

Investments will be made as long as (7) and $\mu_b \geq 0$ hold, as has been argued in Section 3. Naturally, as μ_b is an assessed probability and must thus be located within the real line $[0, 1]$, the latter posits no limitation. Assuming $E > (r_M - r_g)D$ seems natural when considering a socially responsible investor who is prepared to sacrifice capital income in her venture for climate protection, so we take both as given. Hence, she will always decide to invest.

The firm, anticipating this, must determine whether this truthful signaling is

[2] Any signaling game has three subgames. Applied to our model, there are two for anything that happens after the firm sends m_g or m_b, respectively, and the entire game. Note that the first two are irrespective of the firm's type in order to adequately reflect the investor's information set. Graphically, they are the left and right half of Figure 1 (starting at Nature's draw).

indeed a best response to investments for each type. If it is of type t_g, through (1), there is no incentive to deviate if

$$R_g - (1 + r_g)D - V \geq R_g - (1 + r_b)D.$$

This can be shortened to yield the condition

$$r_b - r_g \geq \frac{V}{D}. \tag{8}$$

In words, the benefits from achieving cheaper financing must compensate it for the cost of green certification.

Similarly, the t_b-firm does not wish to change its bond label whenever

$$R_b - (1 + r_b)D \geq R_b - (1 + r_g)D - V - C$$

holds. Manipulation obtains

$$r_b - r_g \leq \frac{V + C}{D}. \tag{9}$$

So the question whether to counterfeit collapses to whether doing so provides a superior way of financing, that is, interest savings from the green label must exceed the costs of dishonestly obtaining it.

It becomes apparent that the greenium $r_b - r_g$ must be within the range of parameters implied by (8) and (9), i.e., $V/D < r_b - r_g < (V+C)/D$. If it is too low, on the one hand, neither firm type will find it worthwhile to acquire the necessary label to reduce its interest payments under the pretext of being pro-social. If the greenium is too high, on the other hand, even non-green firms will want to get it, while being prepared to undergo additional costs of counterfeiting. A truth-telling separating equilibrium can thus only exist with $V/D \leq r_b - r_g \leq (V + C)/D$. We thus arrive at the following:

Proposition 1: Assume $V/D \leq r_b - r_g \leq (V + C)/D$. Then the set of subgame perfect separating equilibria with truthful type revelation is

$$\left[(m_g, m_b), (a_i, a_i), \mu_g, \mu_b\right], \ \mu_g \geq \frac{(r_M - r_g)D + F}{E + F}, \ \mu_b \geq 0.$$

So the firm truthfully reveals its type by the message it sends (the bond it issues) and the investor is prepared to buy either type of bond. As long as we are willing to assume that the investor values greenness heavier than the market premium on top of the green return, $E > (r_M - r_g)D$ (as described above), this set of equilibria is non-empty.

While investment into a regular bond always happens, it need not for the green bond. Assume (7) is violated. Then the investor optimally reacts to m_g with a_n. Clearly, the firm can now improve compared to truth-telling by sending m_b regardless of its type because $R_g - (1 + r_b)D > 0 > -V$. Thus, there can be no separating equilibrium without unconditional investment.

4.2 Deceiving type proclamation

Equilibria where the firm tries to lie about its type are also conceivable. Intuitively, the green label may be shunned even by a green firm due to the fixed costs involved in obtaining it. Conversely, however, even a dirty firm may wish to acquire that label in order to save on interest payments.

The investor will invest in the green bond as long as (7) holds and in the brown one for $\mu_b \geq 0$, as known. Given that the investor would purchase either type of bond from the firm, when will the latter find it optimal not to acquire the green label despite being of type t_g? The condition obtained from (1) reads

$$R_g - (1 + r_b)D \geq R_g - (1 + r_g)D - V,$$

which is just the converse of (8):

$$r_b - r_g \leq \frac{V}{D}. \tag{10}$$

At the same time, t_b should find it optimal to choose m_g, i.e.,

$$R_b - (1 + r_g) - V - C \geq R_b - (1 + r_b)D.$$

Manipulation yields

$$r_b - r_g \geq \frac{V + C}{D}. \tag{11}$$

It becomes apparent that no such situation can arise as (10) contradicts (11). It is intuitive why no separating equilibrium with everyone lying can persist: The acquisition of a green label is more costly for the t_b-firm than it is for the t_g-firm while the interest reduction is the same in both cases. So it cannot be the case that the green label is only procured when it is expensive. The firm has nothing to gain from such an unconditional lie.

Again, the picture might look different if the investor is unwilling to purchase the green bond, perhaps anticipating the deception. Without (7), the firm will, however, wish to depart from the above strategy because $R_b - (1 + r_b)D > 0 > -V - C$, that is, after finding itself to be of type t_b, it would also issue the regular bond to achieve any financing at all and not be left with the costs of acquiring the green label. As the resulting equilibrium would feature identical strategies of both firm types, it is a pooling rather than a separating equilibrium. Hence, we can conclude that there is no separating subgame perfect equilibrium of the deceiving type irrespective of the investor's decision when observing a green bond. This also rules out the existence of corresponding PBEs.

Proposition 2: There does not exist a separating equilibrium with deceiving type proclamation.

5 Pooling equilibria

Aside from separating equilibria, there may be pooling equilibria where the sender elects the same messsage regardless of the type she is assigned. That is, the information on t is not made use of by the firm.

5.1 Green bond

It may be desirable for both firm types to obtain the green certificate. The intuition is that the costs of refinancing are thus reduced from r_b to r_g and that, if this cost reduction is sufficiently large, everyone will want to get it.

Start, as usual by now, by assuming that (7) holds and $\mu_b \geq 0$. The investor will choose a_i regardless of the kind of signal she observes. The conditions for m_g being the best strategy choice for both t_g and t_b in this case have already been derived: they are (8) and (11), respectively. As $(V + C)/D > V/D$, validity of (11) implies validity of (8). So if the greenium is at least $(V + C)/D$, there is green bond pooling.

Proposition 3: Assume $r_b - r_g \geq (V + C)/D$. Then the set of subgame perfect pooling equilibria with green bonds is

$$[(m_g, m_g), (a_i, a_i), \mu_g, \mu_b], \ \mu_g \geq \frac{(r_M - r_g)D + F}{E + F}, \ \mu_b \geq 0.$$

This kind of equilibrium seems odd. Both firm types invest into a label that was meant to be distinguishing but now does nothing. The costs of verification constitute a deadweight loss compared to the equilibrium with two regular bonds. The firm commands this societal loss because its personal gains from cheap financing reimburse it for the incurred burden. Since the empirically found greenium is usually small even when labels are acquired, this seems unrealistic. For example, the overall green bond premium found by Dorfleitner et al. (2022) is statistically significant, but below one basis point.

Assuming again the converse of (7), that is, no investment into a green bond for lack of credibility, challenges the firm's uncontingent decision to label itself as green. Indeed, observing a_n after sending m_g constitutes the worst possible outcome for the firm independently of its type. So it will choose not to get the certificate because this at least saves it the cost from doing so while potentially enabling its project to become financed. Hence, there is no pooling equilibrium that features both a green bond and no investment in that green bond.

5.2 Conventional bond

The previous Subsection has shown that, with a sufficiently high belief μ_g, there is an equilibrium with green bonds. Reciprocally, there is room for an equilibrium

without green labels if that belief is low enough. Start again by assuming that (7) holds and $\mu_b \geq 0$. Then the investor will always choose a_i. Given this, the firm maximizes profits by sending m_b as type t_g as long as

$$R_g - (1 + r_b)D \geq R_g - (1 + r_g)D - V,$$

which simplifies to (10). m_b also maximizes profits for the t_b-firm as long as

$$R_b - (1 + r_b)D \geq R_b - (1 + r_g)D - V - C,$$

or (9). Since this is already implied by (10), the latter on its own suffices to guarantee unconditional optimality of the regular bond. Simply put, if the green certificate is too expensive, it will not be purchased by anyone.

Proposition 4: Assume $r_b - r_g \leq V/D$. Then the set of subgame perfect conventional bond pooling equilibria is

$$[(m_b, m_b), (a_i, a_i), \mu_g, \mu_b], \ \mu_g \geq \frac{(r_M - r_g)D + F}{E + F}, \ \mu_b \geq 0.$$

All equilibria thus far have featured choice of a_i by the investor in every situation. Pooling equilibria with regular bonds where (7) is violated constitue the only (set of) exceptions. If the firm foresees the thus following unwillingness of investors to buy a green bond, it will simply refrain from issuing one, irrespective of the potential greenium since the latter can never be reaped if investor beliefs satisfy the converse of (7). So subgame perfect equilibria without investment in the green bond are also conceivable.

Proposition 5: There exists a set of subgame perfect conventional bond pooling equilibria when the green bond is shunned by investors,

$$[(m_b, m_b), (a_n, a_i), \mu_g, \mu_b], \ \mu_g \leq \frac{(r_M - r_g)D + F}{E + F}, \ \mu_b \geq 0.$$

These equilibria are intuitively appealing: if the investor is so scared of the possibility of becoming a victim of greenwashing that she will generically shirk green bonds, the best the firm can do is to not issue one even if it is t_g. Both players make a rational decision based exclusively on financial motives.

6 Equilibrium refinements

An open question remains on what determines μ_g. Clearly, if V and C are public information, validity of either inequality from (8) through (11) may be checked by the investor to find out about the motives leading a firm to issue a green or regular bond, respectively. At the very least, investors should have rational expectations about everything that happens on the equilibrium path. Denote by S the set of firm types issuing signal m_s, where $s \in \{g, b\}$. That is, with $p_b = 1 - p_g$, we superimpose

$$\mu_s = \frac{p_g}{\sum_{s' \in S} p_{s'}}$$

as long as the green firm is a candidate issuer of m_s. Otherwise, that belief is zero. Note that the numerator is always p_g as both beliefs μ_s refer to the belief of being confronted with a green firm, which has a base rate probability of p_g, given the signal m_s. To sum up, Bayesian beliefs on the equilibrium path take the form

$$\mu_s = \begin{cases} 1, & \text{if } \ S = \{g\} \\ p_g, & \text{if } \ S = \{g, b\} \\ 0, & \text{if } \ S = \{b\} \end{cases}, s \in \{g, b\}\,. \tag{12}$$

This specific process of belief formation constitutes the backbone of perfect Bayesian equilibria (see Gibbons, 2011, p. 188). Thus, applying this refinement to the subgame perfect equilibria found in the preceding Sections delivers PBEs. Beliefs off the equilibrium path will be addressed below.

Truthful type revelation: In a seperating equilibrium, the most plausible beliefs are 0 and 1 because the message serves as a unique identifier. This can be verified to be PBE-beliefs by application of (12).

Proposition 6: Assume $V/D \leq r_b - r_g \leq (V + C)/D$. Then the unique PBE with truthful type revelation is

$$[(m_g, m_b), (a_i, a_i), \mu_g = 1, \mu_b = 0]\,.$$

Green bond pooling: Observing structural parameters, the investor can infer that the model setup implies choice of m_g by both types. Thus, the probability she should assign to the firm being type t_g is the factual probability of this being the case, p_g, as implied by (12). So $\mu_g = p_g$ is rational, implying further that p_g must weakly exceed the lower bound of μ_g in (7) for such an equilibrium to exist. This reduces the set of equilibria:

Proposition 7: Assume $r_b - r_g \geq (V+C)/D$ and $p_g \geq [(r_M - r_g)D + F]/(E+F)$. Then the set of green bond pooling PBEs is

$$[(m_g, m_g), (a_i, a_i), \mu_g, \mu_b], \ \mu_g = p_g, \ \mu_b \geq 0.$$

For pooling equilibria, off-equilibrium beliefs can play a decisive role. Since, in our model, the firm always wants to make the investor choose a_i, which she does on the equilibrium path of any equilibrium, no outcome is based on threats that can be rationalized away (as they can be, for example, in the beer-quiche-game of Cho & Kreps, 1987). Thus, we use the following equilibrium refinement that is meant to express that leaving the equilibrium path is not informative about the sender's type:[3]

Definition 1: A pooling equilibrium with $\mu_g = \mu_b$ is *trembling hand perfect.*

That is, to achieve *trembling hand perfectness*, we have to assume equality of the factual probability of the firm being green, p_g, not just to on-equilibrium beliefs μ_g, but also to off-equilibrium beliefs μ_b. The intuition is that if a message is observed that neither type would like to send, it must be the result of an

[3]We refer to this refinement concept as *trembling hand perfectness* in reference to Selten (1975). He introduced perturbed games where the probability distribution of strategy choices are exogenously altered slightly, such that off-equilibrium strategies may be played with a small but positive probability. Assuming a miniscule probability of the sender emitting an off-equilibrium strategy (independently of her type) requires Bayesian belief formation anywhere, i.e., not just on the equilibrium path, thus resolving the multitude of PBEs with pooled sender strategy choices. To the author's knowledge, the term *trembling hand perfectness* is not rooted in the signaling game literature.

error. Therefore, the question is, at first, not just about the probability of each type, but about the error probability of each type as well. Since any statement other than that they are the same is somewhat arbitrary, those error probabilities should be identical across types. So $\mu_b = p_g$, which places no further restriction on parameter ranges due to $\mu_b \geq 0$, achieves trembling hand perfectness.

Proposition 8: Assume $r_b - r_g \geq (V+C)/D$ and $p_g \geq [(r_M - r_g)D + F]/(E+F)$. Then the *trembling hand perfect* green bond pooling PBE is

$$[(m_g, m_g), (a_i, a_i), \mu_g, \mu_b], \ \mu_g = \mu_b = p_g.$$

Conventional bond pooling: The exact same arguments made for green pooling can be applied to the case of expensive labels leading to brown bond pooling with unconditional investment. We omit the intermediate step of a non-trembling-hand-perfect PBE.

Proposition 9: Assume $r_b - r_g \leq V/D$ and $p_g \geq [(r_M - r_g)D + F]/(E+F)$. Then the *trembling hand perfect* conventional bond pooling PBE is

$$[(m_b, m_b), (a_i, a_i), \mu_g, \mu_b], \ \mu_g = \mu_b = p_g.$$

Note that the restriction to sufficiently high levels pf p_g is merely a necessity as long as we require the investor to respond to m_g with a_i, which is off equilibrium. The condition can be disposed of if there would not be investment into a green bond.

Proposition 10: Assume $p_g \leq [(r_M - r_g)D + F]/(E+F)$. Then the *trembling hand perfect* conventional bond pooling PBE when the green bond is shunned by investors is

$$[(m_b, m_b), (a_n, a_i), \mu_g, \mu_b], \ \mu_g = \mu_b = p_g.$$

We can trivially constitute that the level of p_g plays no role as long as beliefs are formed correctly. Only if $p_g = [(r_M - r_g)D + F]/(E+F)$, the two equilibria coexist.

7 Upsetting equilibria

Rothschild & Stiglitz (1976) look at an insurance market with different accident probabilities of individuals. They show that contracts with non-negative profits from insurance companies can upset pooling equilibria, i.e., contracts that make one risk type better-off than the pooling contract could draw away those customers. Thus, the alleged equilibrium is challenged as a newly entering firm enhances welfare.

We wish to check whether similar considerations can upset equilibria in our model as well. To do so, we have to grant individuals the capability to make voluntary side payments to the firm.[4] Of course, it is not common practice that investors pay firms directly to incentivize certain behavior. One can, instead, imagine the discussed side payments as a voluntary reduction in the demanded r_g or r_b in size of the side payment divided by D.

Take the pooling PBE from Proposition 9 where the firm sends m_b irrespective of its own type. In principle, it appears conceivable that the investor wishes to have certainty about the firm type. Consider her expected payoff from investing given we are in the PBE,

$$r_b D + p_g E. \tag{13}$$

She gets the high interest rate for sure and has the chance of financing a green firm along the way. Were the firm to reveal its type via emission of a green bond, on the other hand, her payoff would be

$$r_g D + E$$

with certainty. The latter can be higher and, thus, preferrable even without any notion of risk aversion. Her willingness to pay for incentivizing the firm to change its behavior is, however, zero. To see this, note that the uncertainty about the

[4] Side payments from firms to investors are, in principle, just as conceivable. However, it can be seen below that such transfers would never be made.

firm's type does not vanish ex ante, that is, viewed from before the start of the game. Knowing that the equilibrium will be type-revealing, the expected investor profit ex ante is

$$p_g(r_g D + E) + (1 - p_g) r_b D, \tag{14}$$

which is always below (13) since $r_g < r_b$. So the investor has no incentive to breach this equilibrium via side payments as she gets E with a chance of p_g irrespective of the interest rate she pays on either bond. Neither does the firm as it acts in a profit-maximizing way given (10), which is in turn a precondition for the equilibrium under consideration.

Consider now the separating equilibrium from Proposition 6. There, the firm decides to acquire a green label if it is genuinely green because only then is that label cost-efficient. So the firm is not willing to pay for a change in the equilibrium structure. The investor, on the other hand, would mostly love to obtain a high interest rate in addition to the good conscience for being invested into a green firm. In fact, it has been shown above by means of comparing (13) with (14) that, ex ante, expected investor profits are higher under brown bond pooling than they are under truthful type revelation given all parameters. The amount by which (14) exceeds (13) gives the investor's willingness to pay for such an equilibrium change. The side payment thus made to a firm sending m_b can be considered as a reduction in r_b. If that reduction can lead (8) to be violated, that is, (10) to be true, the type-revealing equilibrium is upset by allowing for side payments.

Via the usual argument, a firm in a green bond pooling equilibrium like in Proposition 8 has no willingness to pay for a different equilibrium outcome as it is already maximizing profits there. However, the investor is confronted with a useless green label that does not reveal any information to her and that further forces her to accept a reduced interest rate. Again, the equilibrium structure can be changed by transferring wealth to the m_b-emitting firm. The investor has an additional valuation for the truthtelling equilibrium if

$$r_g D + p_g E + (1 - p_g) F < p_g r_g D + p_g E + (1 - p_g) r_b D,$$

which is always true. So if the difference of the above LHS and RHS is at least $r_b - r_g - (V+C)/D$, she can afford to break the firm's lying incentive given by (11) while increasing her expected equilibrium payoff. If she can afford to even pay $r_b - r_g - V/D$ without making herself worse-off, conventional bond pooling can be reached as well. Note that the investor's willingness to pay for going straight from green to conventional bond pooling exceeds the greenium $r_b - r_g$ as she thus avoids being fooled and, thus, suffering a utility loss of F.

The above analysis shows that allowing for side payments leads equilibria to be drawn away from green labels. The reason for this lies within the deadweight loss imposed on society if the green label is acquired. If both actors can agree on disposing of that costly label and allocate the resulting surplus among them, a pareto-improvement happens, which both parties will gladly accept.

8 Investor types

The model framework can be readily adjusted to include a neutral investor who does not value the externality rather than a respnsible one as contemplated so far. For her, set $E = 0$. We still include $F \neq 0$, indicating that, although she does not care about the environment, the investor dislikes being fooled. It seems intuitive that this eliminates incentives for the firm to brand itself as green. Notably, however, including uncertainty about whether the investor is responsible or not may not influence the resulting equilibria at all.

8.1 Neutral investor

For a neutral investor, the belief that supports investment into a green bond is a modified version of (7). It reads

$$\mu_g \geq \frac{(r_M - r_g)D + F}{F}. \tag{15}$$

Since the RHS of (15) is greater than one, it is impossible for this condition to be given. This can also be taken immediately from Figure 1: clearly, with $E = 0$,

a_n becomes a strictly dominant strategy in response to m_g. The investor would never forego additional profits compared to the market rate without compensation. Similarly, a_i remains an unconditional best-response to m_b. The firm's situation remains identical. It is faced with the same projects and verification costs. Therefore, it will still wish to send a message that grants its financing at the lowest possible cost. Trivially, this requires choice of the regular bond by both types because else no funds can be drawn.

Proposition 11: With a neutral investor, the set of subgame perfect equilibria is

$$[(m_b, m_b), (a_n, a_i), \mu_g, \mu_b], \ \mu_g \leq 1, \ \mu_b \geq 0.$$

Trembling hand perfectness refinements and Bayesian belief formation from (12) can be applied in the familiar manner:

Proposition 12: With a neutral investor, the unique *trembling hand perfect* PBE is

$$[(m_b, m_b), (a_n, a_i), \mu_g, \mu_b], \ \mu_g = \mu_b = p_g.$$

8.2 Uncertainty about the investor's type

In principle, it is also conceivable that the firm does not know ex ante whether the investor it faces is environmentally concerned or not. Let it estimate the chance of a green investor to be ι_g. Then the "expected game" from the firm's perspective features this ι_g as a factor on E at investor payoffs and is else equal to Figure 1. Thus, the relevant new version of (7) is

$$\hat{\mu}_g \geq \frac{(r_M - r_g)D + F}{\iota_g E + F}, \tag{16}$$

where the hat denotes expectations formed by the firm.

Two cases have to be distinguished. First, assume that ι_g is low enough to make the firm believe that it will not be able to sell the green bond at an expected profit (that is, the LHS of (16) is at least one). Then the only equilibria parallel

those from Propositions 11 and 12.

Second, assume the opposite: ι_g is sufficiently close to one such that the RHS of (16) is strictly below one. Then complements of all equilibria stated in Propositions 1 and 3 through 5 exist, where the precise structure depends on the level of the greenium. Whether or not the corresponding PBEs from Propositions 6 through 10 also exist hinges on the level of p_g.

In short, we can conclude that the inclusion of uncertainty on behalf of the firm does not enrich the set of possible equilibria. They always parallel the respective sets from facing a responsible or neutral investor with certainty, depending on ι_g, although the sets will in the green-investor-case generally be smaller or more restrictive on the allowed levels of p_g.

9 Coexistence of investor types

In a last step, we vary our model to include both types of investors simultaneously. The idea thus depicted is an application of Arnold's (2023) result that in a (general) equilibrium with social (e.g. green) and neutral investors, portfolios may be switched compared to a benchmark with no social criteria involved, but the financing of firms at unchanged terms is always warranted.

Specifically, the firm in our model can now conduct two rounds of capital acquisition. It tries to acquire funds from the green investor, but has a neutral investor functioning as its double bottom in case capital cannot be raised from the green one. That is, the game from Figure 1 is extended such that the a_n-nodes are not terminal. Instead, they offer the chance for the firm to rethink its branding, i.e., whether to issue bonds at r_g or r_b, indicated by strategies m'_g and m'_b, respectively. Notably, m'_g after m_g is costless as the label has already been acquired, whereas the cost of V and, potentially, C, has to be paid to play m'_g after m_b. Furthermore, it is intuitively appealing to make the costs from m_g sunk: if the green label has been paid for by the firm, that does not prevent it from offering the higher rate r_b via signal m'_b in the second round, but it cannot recover those

expenses that way. The neutral investor cannot observe the firm's type but only its m'-signal. She responds by an investment a'_i or by choosing her outside option a'_n. The game trees following each a_n of the original game are depicted in Figure 2. There, payoff vectors are ordered as showing the firm's payoffs first, the green investor's second and the neutral investor's last.[5]

The neutral investor does not need information about the firm's true type as it is only the message itself that influences her payoffs. She will always buy the conventional bond and avoid the green bond to maximize her return. Anticipating this, the firm will discard of the green label – should it have acquired one in the first round – as a means of damage containment.[6] Doing so ensures choice of a'_i by the neutral investor, thus granting certainty of financing. Denoting subgame perfect equilibria as an array of firm's choice of m', neutral investor's reaction to m'_g and neutral investor's reaction to m'_b in that order (i.e., as a complete plan of play, neglecting the green investor who does not draw), we thus have the following result:

Proposition 13: The subgame perfect equilibrium of the game following a_n of the green investor is

$$[m'_g, (a'_i, a'_i)]$$

irrespective of t and m.

Interestingly, the neutral investor need not even know the firm's message sent to the green investor (although, in equilibrium, she may infer it to be m_g from

[5]If we wish to consider these trees as expansions of the baseline game, the latter has to feature the neutral investors' payoffs, too. One can think of this as the payoff $r_M D$ or (any other reservation value) being added as the third component of each payoff vector following an a_i. Figure 3 below does that. Naturally, as it is never the neutral investor's turn in that case, equilibria are not influenced.

[6]Aside from the disutility caused by a falsely procured green label, the message choice of the firm in the second round following m_g is essentially the choice of its offered interest rate. One may therefore re-interpret the model as the firm sticking to its label, but offering different interest rates on its green bonds.

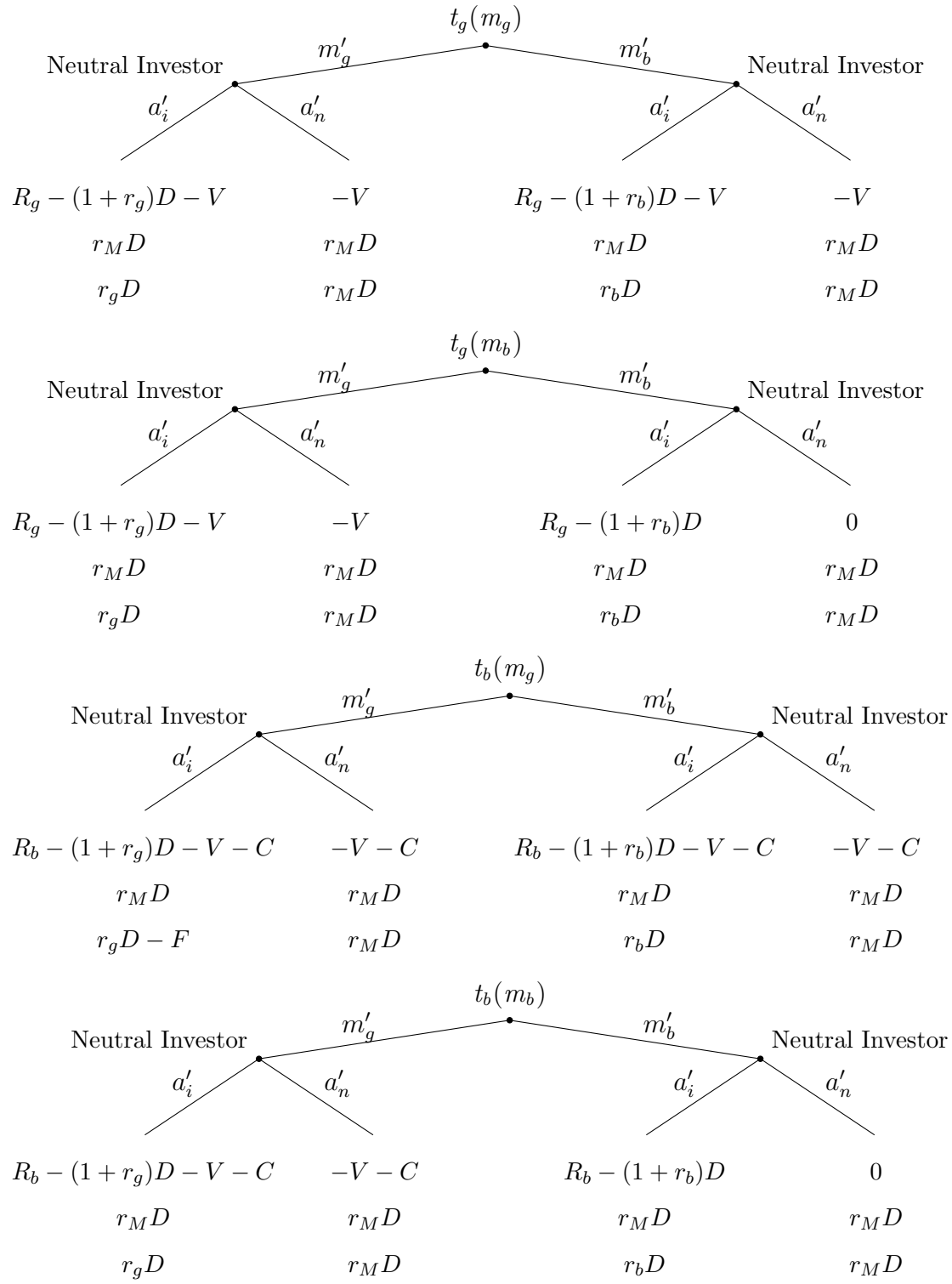

Figure 2: Game trees for the firm's type-message-pairs after green investor's choice of a_n (extensive forms)

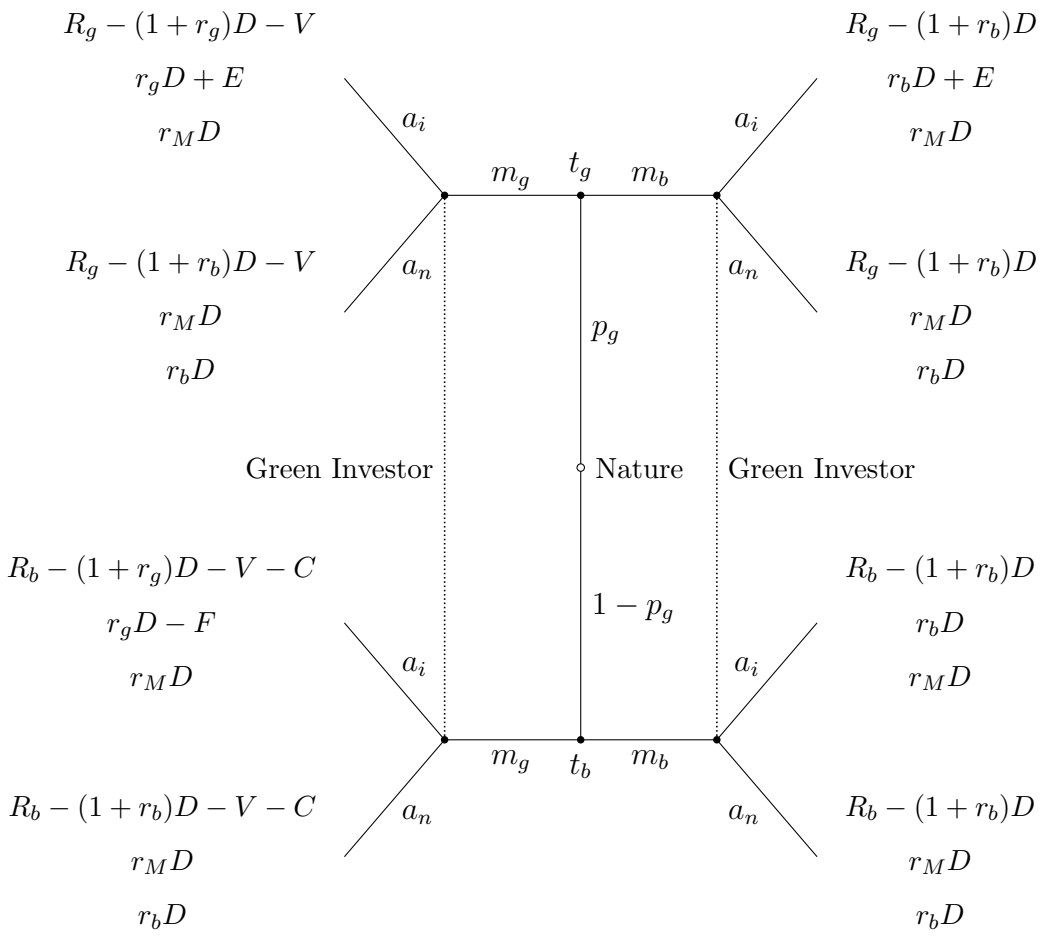

Figure 3: Green bond signaling game with two investors (reduced extensive form)

observing a_n). Furthermore, the second and last game trees from Figure 2 are never reached in an equilibrium as a_n can only be an equilibrium reaction to m_g. We can combine the derived subgame perfect equilibrium with the PBE from the baseline signaling game by replacing the extended branches of the latter with equilibrium payoffs derived from the former. This yields the reduced picture of the extended game depicted in Figure 3.

Note that the green investor's situation is unchanged, while the neutral investor cannot influence the path of the game followed by the other two actors (she may only react optimally if a_n is observed, as discussed above). Hence, we need only occupy ourselves with how the firm's situation has changed compared to Figure

1.

As can readily be seen, the overall situation of the firm improves due to the double bottom provided by the existence of a neutral investor. However, it is still always weakly preferrable to become financed in the first round, that is, by the green investor. Only if the firm is t_b and would have sent m_b anyways are the two modes of financing equally attractive to it. As the choice of bond still hinges on the greenium levels as implied by (8) through (11) as long as a_n is not chosen by the green investor, we can conclude that most equilibria (in either of their versions, i.e., subgame perfect, PBE, and *trembling hand perfect* PBE) carry over to the two-investor model.

Equilibria with a_n (see Proposition 5) deserve further analysis. The question is whether the green investor's (credible) threat of not buying green bonds due to risk of greenwashing still induces the firm to issue conventional bonds. As the green bond comes at a deadweight loss for the firm in either case (costs of V or $V + C$ are borne while the rate r_b, i.e., a conventional bond, must still be offered in the second round), the answer is in the affirmative.

Proposition 14: The equilibria from Propositions 1 and 3 through 5 extend to the two-investor model. Their refinements from Propositions 6 through 10 apply accordingly.

Note that a PBE with both investor types must feature the neutral investor's beliefs about t as well. We omit this discussion here as they may be anything and their levels are utterly irrelevant. As a matter of fact, in equilibrium, the neutral investor never gets to purchase the bond anyway.

We could, in principle, consider the converse situation as well. That is, assume the firm tries to acquire funds from the neutral investor first and, if necessary, from the green one second. Then we arrive at the situation where the choice of bond is essentially also the choice of an investor: the firm knows that the neutral investor would buy (only) the conventional bond. If it issues a green bond, on the other hand, the neutral investor will resort to the market rate instead. Thus,

the firm turns to the green investor, who will find it optimal to buy that green bond under the usual requirement taken from (7).[7] So all equilibria are exactly parallel to those from the baseline game, with the only exception being that the (potentially greenwashed) green bond is always purchased by a green investor and the conventional bond always by a neutral investor.

10 Endogenous interest rates

The model can also be modified to include an endogenous choice of bond interest rates. We shift market power from the firm to the investor by introducing an intermediate step: after the firm has sent its message m_s, the investor indicates an interest rate she is willing to accept as payment on the capital lent, $r_s^*, s \in \{g, b\}$. The firm can either accept that offer (a_a), in which case the money is lent at the proposed interest rate, or decline it (a_d), implying no financing and no recovery of potentially incurred sunk costs V and C from acquisition of the green label via m_g. We restrict r_s^* to be chosen from the subinterval $[0, R_b/D - 1 + \varepsilon]$ of the real line, where ε is positive and small. The upper limit of that interval lies just above the brown project's percentage return. It thus denotes a prohibitively high interest rate that induces a_d by either firm type irrespective of its message. Hence, the investor is not obliged to finance the firm and can still resort to the reservation value $r_M D$. The extensive form of this game is depicted in Figure 4, where the investor's choice of $r_s^*, s \in \{g, b\}$, is indicated by the shaded regions.
Note that each message is followed by a dictator game with the investor proposing a split. We wish to depict an investor that suffers a lot from being fooled into buying a greenwashed bond again. If the bond is known to be greenwashed, it will be avoided by the investor as long as $r_g^* D - F < r_M D$. Assuming that the investor can extract all profits from the t_b-firm sending m_g, which is done by

[7] Naturally, the firm would never abandon its green label in that case.

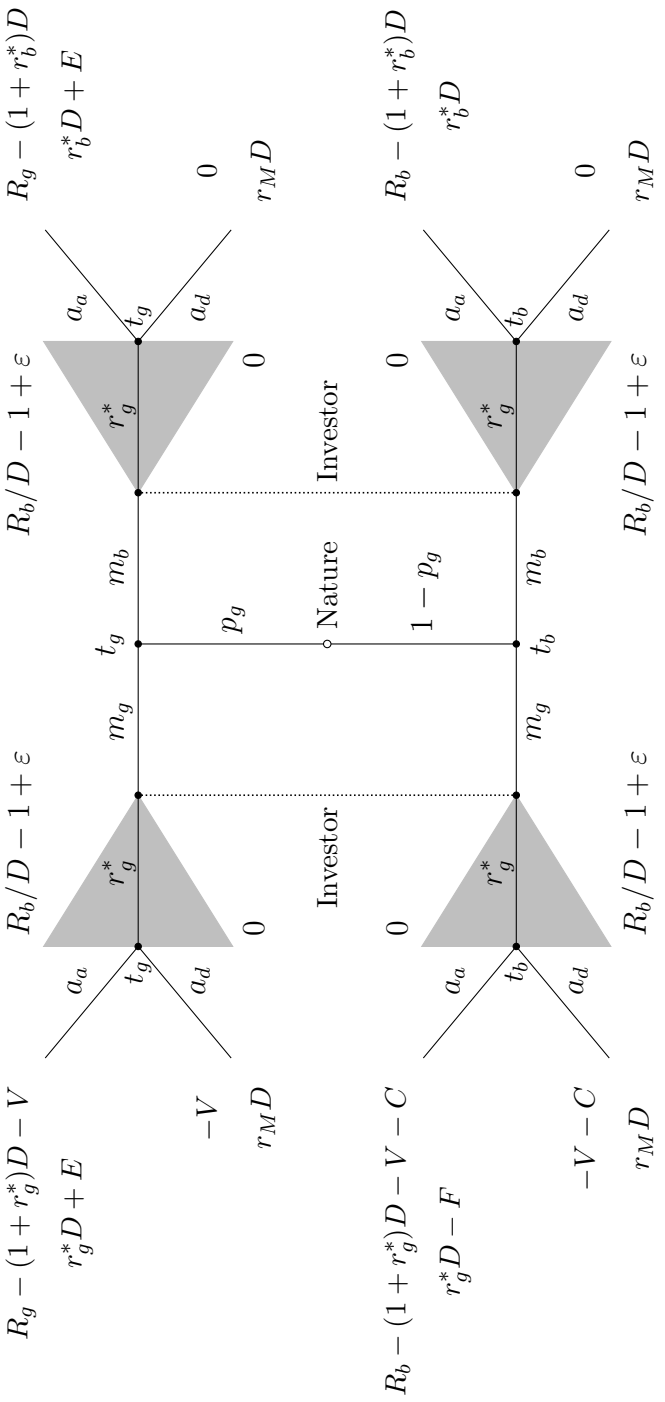

Figure 4: Endogenous interest rate game tree (extensive form)

setting $r_g^* = R_b/D - 1$, thus yields

$$F > R_b - (1 + r_M)D \tag{17}$$

as the condition for not wishing to finance that firm. If the investor is sure to be confronted with a greenwashed firm, she will thus respond optimally by setting the prohibitive rate. This is assumed to hold throughout the following analysis.

To find all subgame perfect equilibria that exist in the game depicted in Figure 4, solve backward. That is, start at the firm observing an offer of r_s^*. To this, the firm always reacts with a_a as long as $r_s^* \leq R_g/D - 1$ if it is of type t_g, and as long as $r_s^* \leq R_b/D - 1$ if it is t_b, $s \in \{g, b\}$. Note that potentially incurred certification costs C and V do not appear in this threshold as the firm cannot recover them at this stage.

As the investor now holds all the market power, she, in fact, has the chance to extract all profits gross of sunk costs from the firm. When being offered a brown bond (m_b), she will always try to do so. However, as the firm's type is not revealed, there are two conceivable ways of how to achieve this. The first is to set $r_b^* = R_g/D - 1$, which leaves zero profits to the green firm and positive profits of $R_b - R_g$ to the brown firm. If, on the other hand, she sets $r_b^* = R_b/D - 1$, she manages to obtain all profits from the brown firm, but also makes the green firm decline the offer, in which case she must resort to the market rate r_M. That is, with a belief of μ_b that the m_b-sending firm is type t_g, the investor's expected payoff from $r_b^* = R_g/D - 1$ is

$$\mu_b \left[\left(\frac{R_g}{D} - 1 \right) D + E \right] + (1 - \mu_b) \left(\frac{R_g}{D} - 1 \right) D,$$

or

$$R_g - D + \mu_b E.$$

The more aggressive rate $r_b^* = R_b/D - 1$, on the other hand, delivers

$$(1 - \mu_b) \left(\frac{R_b}{D} - 1 \right) D.$$

So beliefs

$$\mu_b \geq \frac{R_b - R_g}{R_b - (D - E)} \tag{18}$$

lead to choice of the lower rate $r_b^* = R_g/D - 1$. Denote this threshold as μ_b^*. Due to $R_b > R_g > D > D - E$, it is indeed between zero and one. So there exist beliefs that support either interest choice of the investor.

If a green bond (m_g) is offered, the investor has the same alternatives, but will never choose to set $r_b^* = R_b/D - 1$. To see this, note that, while the high rate extracts all profits from the t_b-firm, the investor is still worse-off in that case compared to not financing the firm because of (17). As a t_g firm would decline that rate, expected investor profits are below $r_M D$ (given $\mu_g < 1$). So the investor will either set $r_g^* = R_g/D - 1$, thus extracting all profits from the green firm, or the prohibitive rate $r_g^* = R_b/D - 1 + \varepsilon$ to ensure rejection by t_b. Hence, if the investor thinks the m_g-sending firm is indeed t_g with probability μ_g, she has to compare the risky payoff

$$\mu_g \left[\left(\frac{R_g}{D} - 1 \right) D + E \right] + (1 - \mu_g) \left[\left(\frac{R_g}{D} - 1 \right) D - F \right],$$

which can be simplified to

$$R_g - D + \mu_g(E + F) - F,$$

against the safe market rate obtained from setting the prohibitive rate. The former is superior to simply opting for the market rate

$$r_M D$$

as long as

$$\mu_g \geq \frac{(1 + r_M)D - R_g + F}{E + F}. \tag{19}$$

Refer to the threshold implied by (19) as μ_g^*. It is positive due to (17) and below one because $(1+r_M)D - R_g < 0 < E$. So there are again beliefs supporting either reaction of the investor.

Having established optimum investor reactions, as a last step it remains to show which firm type sends what message. Given the rates it will face, it can only be rational for t_g to choose m_b. The green firm can never make positive profits and will thus refrain from additionally sinking the verification cost V. We can thus rule out equilibria where t_g sends m_g, which leaves only two candidates: deceiving type proclamation and brown pooling.

Interestingly, this Section's model excludes the possibility of an investor-offered greenium. To see this, consider the following candidate equilibrium: The firm truthfully reveals its type by sending m_g if green and m_b if brown. The investor promises to make the green label worth its while by setting $r_g^* = (R_g - V)/D - 1$. With $r_b^* = R_b/D - 1$, this ensures that the firm has no incentive to switch to the brown label if it observes itself being of type t_g as both options then yield zero. However, when m_g is observed, the investor has no incentive to stick to its promise and will instead demand $r_g^* = R_g/D - 1$ (or, possibly, the prohibitive rate if beliefs are not Bayesian). Anticipating this, the firm will never send m_g in the first place. We can thus conclude that, in a model with market power concentrated on the investor side, a greenium cannot be explained by investors shouldering verification costs. This makes intuitive sense as firms' ability to roll over costs (for example taxes or innovation costs) tends to increase with their own market power.

Deceiving type proclamation: Consider now the separating equilibrium where t_g sends m_b and t_b sends m_g. The baseline game from Section 2 did not even allow for a subgame perfect equilibrium of this strategy (let alone a PBE). Here, such an equilibrium can exist as long as (19) holds as the implied optimum investor reaction $r_g^* = R_g/D - 1$ gives the firm $R_b - R_g - V - C$, which may be positive. However, the t_b-firm has an incentive to deviate to m_b if that makes the investor choose the same interest rate. So we instead have to require $r_b^* = R_b/D - 1$. While this indicates inactivity of the t_g-firm, the latter will accept this as it surpasses incurring the verification cost and being drained of all profits. Lastly, the investor

will only choose $r_b^* = R_b/D - 1$ if the converse of (18) holds.

Proposition 15: In the model with interest rates set by the investor, assume $R_b - R_g - V - C > 0$. Then the set of subgame perfect deceiving type proclamation equilibria is[8]

$$[(m_b, m_g), (R_g/D - 1, R_b/D - 1), (a_a \text{ if } r_s^* \leq R_g/D - 1,\ a_a \text{ if } r_s^* \leq R_b/D - 1),$$
$$\mu_g, \mu_b], \mu_g \geq \frac{(1 + r_M)D - R_g + F}{E + F},\ \mu_b \leq \frac{R_b - R_g}{R_b - (D - E)}.$$

While these equilibria do not seem very appealing, they have an important implication: If investors hold all market power, the only purpose green labels could fulfill is to serve as a means for firms to fool gullible investors. So leaving some of the capital market's market power on the firms' side might be beneficial to investors. If, however, investors are rational in a Bayesian sense, they cannot hold the beliefs supporting this equilibrium: A PBE requires $\mu_g = 0$, which is not included in the set of equilibria permitted by Proposition 15.

Proposition 16: In the model with interest rates set by the investor, there does not exist a separating PBE with deceiving type proclamation.

Conventional bond pooling: Lastly, consider the case of m_b sent by either firm type. We have to distinguish four cases depending on whether and in what combinations (18) and (19) hold. If both do hold, the investor will optimally set $r_s^* = R_g/D - 1, s \in \{g, b\}$. The firm naturally cannot improve by defecting to m_g as this would then simply mean to sink the certification costs without gains from doing so independently of its type. In this equilibrium, the firm gets financed irrespective of its type.

Next, assume (18) does not hold while (19) does. That is, if the investor were

[8]The equilibrium notation in this Section follows the form [(message sent by t_g, message sent by t_b), (rate set in response to m_g, rate set in response to m_b), (condition for choice of a_a if of type t_g, condition for choice of a_a if of type t_b), beliefs]. We omit the influence of the firm's own message on its decision to accept a rate or not as the costs of acquiring a label are sunk at this stage and thus exert no influence.

to observe the (off-equilibrium) signal m_g, she would interpret this as a reason to believe in the firm being genuinely green. She would reward this by offering $r_g^* = R_g/D-1$. However, if $R_b-R_g-V-C>0$, this incentivizes the t_b-firm to lie, which leads us to the separating equilibrium with deceiving type proclamation discussed above. This incentive can be eliminated by setting $R_b-R_g-V-C\leq 0$ instead.

Validity of (18), but not (19), on the other hand, describes a situation where the investor considers the off-equilibrium signal to be deceiving. As she would respond to this with the prohibitive rate, neither firm type would respond to those beliefs with defection. Thus, those beliefs also support a pooling equilibrium of conventional bonds.

Finally, the investor may consider it more likely to face a brown firm regardless of the issued kind of bond, that is, both (18) and (19) could be violated. As the investor would then again set the prohibitive rate in response to m_g, neither firm type will wish to sink the additional verification costs associated with doing so.

As the above logic shows, beliefs can be virtually anything and still support the conventional bond pooling equilibrium as long as $R_b-R_g-V-C\leq 0$. Only if this inequality is violated and beliefs are as in Proposition 15 will the corresponding separating equilibrium replace the subgame perfect pooling equilibrium.

Proposition 17: In the model with interest rates set by the investor, the set of subgame perfect conventional bond pooling equilibria encompasses arbitrary beliefs. Equilibria take on the forms

$$[(m_b,m_b),(R_g/D-1,R_g/D-1),(a_a \text{ if } r_s^*\leq R_g/D-1,\ a_a \text{ if } r_s^*\leq R_b/D-1),$$
$$\mu_g,\mu_b],\mu_g\geq\frac{(1+r_M)D-R_g+F}{E+F},\ \mu_b\geq\frac{R_b-R_g}{R_b-(D-E)},$$

or

$$[(m_b,m_b),(R_b/D-1+\varepsilon,R_g/D-1),(a_a \text{ if } r_s^*\leq R_g/D-1,\ a_a \text{ if } r_s^*\leq R_b/D-1),$$
$$\mu_g,\mu_b],\mu_g\leq\frac{(1+r_M)D-R_g+F}{E+F},\ \mu_b\geq\frac{R_b-R_g}{R_b-(D-E)},$$

or

$$[(m_b, m_b), (R_b/D - 1 + \varepsilon, R_b/D - 1), (a_a \text{ if } r_s^* \leq R_g/D - 1,\ a_a \text{ if } r_s^* \leq R_b/D - 1),$$
$$\mu_g, \mu_b], \mu_g \leq \frac{(1 + r_M)D - R_g + F}{E + F},\ \mu_b \leq \frac{R_b - R_g}{R_b - (D - E)}.$$

Assume in addition that $R_b - R_g - V - C \leq 0$. Then we have further equilibria of the form

$$[(m_b, m_b), (R_g/D - 1, R_b/D - 1), (a_a \text{ if } r_s^* \leq R_g/D - 1,\ a_a \text{ if } r_s^* \leq R_b/D - 1),$$
$$\mu_g, \mu_b], \mu_g \geq \frac{(1 + r_M)D - R_g + F}{E + F},\ \mu_b \leq \frac{R_b - R_g}{R_b - (D - E)}.$$

Starting from these subgame perfect equilibria, our goal is to again apply refinements of PBE and *trembling hand perfectness* as in Section 6. Before doing so, the beliefs necessary to support certain strategy choices as implied by (18) and (19) are worth exploring further. In particular, note that due to (17), we have

$$\mu_g^* > \frac{R_b - R_g}{E + R_b - (1 + r_M)D}.$$

Comparing this to the RHS of (18), i.e., μ_b^*, reveals that the latter is always smaller:

$$\mu_b^* < \mu_g^*. \tag{20}$$

So it takes more certainty of actually facing t_g in order to entice the investor to offer the lower candidate interest rate $R_g/D - 1$ when she observes a green bond (m_g) as compared to a conventional bond (m_b). The rationale for this is that she could be financing a green or brown firm either way, but will only suffer from investing in the brown one if the latter has fooled her into believing it were actually green.

It seems intuitive that, in a world where green labels are costly and can only fulfill the purpose of greenwashing, such green labels are not purchased in most equilibria. In fact, this is true for all PBEs:

Proposition 18: In the model with interest rates set by the investor, the unique

trembling hand perfect conventional bond pooling PBE depends on p_g. Assume $p_g \geq \mu_g^*$. Then the *trembling hand perfect* PBE is

$$[(m_b, m_b), (R_g/D - 1, R_g/D - 1), (a_a \text{ if } r_s^* \leq R_g/D - 1,\ a_a \text{ if } r_s^* \leq R_b/D - 1),$$

$$\mu_g, \mu_b], \mu_g = \mu_b = p_g.$$

Assume $p_g \in [\mu_b^*, \mu_g^*]$. Then the *trembling hand perfect* PBE is

$$[(m_b, m_b), (R_b/D - 1 + \varepsilon, R_g/D - 1), (a_a \text{ if } r_s^* \leq R_g/D - 1,\ a_a \text{ if } r_s^* \leq R_b/D - 1),$$

$$\mu_g, \mu_b], \mu_g = \mu_b = p_g.$$

Assume $p_g \leq \mu_b^*$. Then the *trembling hand perfect* PBE is

$$[(m_b, m_b), (R_b/D - 1 + \varepsilon, R_b/D - 1), (a_a \text{ if } r_s^* \leq R_g/D - 1,\ a_a \text{ if } r_s^* \leq R_b/D - 1),$$

$$\mu_g, \mu_b], \mu_g = \mu_b = p_g.$$

The *trembling hand perfect* PBE is thus unique save for knife-edge cases.[9] Interestingly, we no longer need to distinguish whether $R_b - R_g - V - C \leq 0$ holds or not. This is the case because the subgame perfect equilibria of the form

$$[(m_b, m_b), (R_g/D - 1, R_b/D - 1), (a_a \text{ if } r_s^* \leq R_g/D - 1,\ a_a \text{ if } r_s^* \leq R_b/D - 1),$$

$$\mu_g, \mu_b], \mu_g \geq \frac{(1 + r_M)D - R_g + F}{E + F},\ \mu_b \leq \frac{R_b - R_g}{R_b - (D - E)}$$

required impossible beliefs. With $\mu_g = \mu_b$ in pooling equilibria, there is no way to have μ_g above some threshold (μ_g^*) when at the same time μ_b needs to deceed a comparatively lower threshold (μ_b^*). One could object that such an equilibrium

[9] This statement rests on a mathematical vagueness. Indeed, the investor, in her endeavor to set a prohibitive rate, could in principle choose from the uncountably infinite set $(R_b/D - 1, R_b/D - 1 + \varepsilon]$. That is, we assume away an unnecessarily low prohibitive rate. This is consistent with the investor setting this rate because she knows that it leads to an unconditional refusal of her offer by the firm. Giving her the option to withdraw from the market and restrict rates to $r_s^* \in [0, R_b/D - 1]$ instead would in principle resolve this problem, but at the expense of having to specify additional strategy choices and, thus, a less neat form of Figure 4.

cannot be excluded without *trembling hand perfectness*. However, knowing the game structure, there is no rational reason for the investor to believe in elevated credibility of the green signal. The separating equilibrium is right out as it fails to support Bayesian beliefs, which would have to be $\mu_g = 0$ and $\mu_b = 1$ here (see Proposition 16). Thus, if market power were concentrated at the investors' side, no equilibria with both rational beliefs and green labels can arise.

11 Conclusion

This paper investigates the signaling channel of green bonds and a firm's strategic decision to acquire a green label. The corporation will do so if the label is sufficiently cheap such that the associated costs are offset by reduced capital costs when a greenium can be reaped. Depending on that greenium, we can distinguish three staggered cases: either no firm will purchase the green label, or only the green firm will, or both will. Acquisition of a green certificate by a brown firm can only be valuable if the green firm would purchase that label as well. The intuition is not the identification as a fake by this fraudulent message, but rather a mechanic cost argument: if a green firm cannot afford a green label, neither can a brown one due to additional costs of counterfeiting. A caveat on this argument is, however, that it is made from an within-firm perspective. Different cost structures may explain the coexistence of greenwashers and uncertified genuinely green firms we observe in the real world.

As the certification costs borne by the firm to label its bond as green are a deadweight loss, there are incentives for the investor to pay the firm for not purchasing that label and rather distribute the surplus among them. Conversely, the investor would never pay the firm for acquiring the green label.

Extending the model to include a neutral rather than a green investor eliminates most equilibria and leads to an automatic process of regular bond emission and investment into the latter. Uncertainty on the investor type does not yield qualitatively new equilibria. Rather, the model solutions revert to either of the

certainty cases, possibly with a reduced size of the sets of equilibria.

Including both investor types simultaneously and leaving them the sequential possibility to purchase the bond provides a double bottom to the firm. In equilibrium, however, it does not need to make use of this. If a green investor draws first, the equilibrium is always such that she will indeed purchase the bond. If the neutral investor draws first, the firm implicitly chooses its investor by the message it sends.

Green labels are not purchased if investors hold all market power. This is the case because those investors can squeeze all profits out of the firm, ignoring sunk costs that have to be incurred when declaring a bond to be green. Thus, the mere existence of green bonds in our world hints at the fact that investors do not hold too much market power in fixed-rate capital markets.

The policy implication of our paper is concentrated on the greenium. The results show that, if a green bond is meant to be understood as a credible signal of actual firm greenness, the green bond premium must take on intermediate values. If it is too high, there will be greenwashing. If it is too low, no green bonds will be emitted. The real world observation of positive greenia may thus be benevolently acknowledged as long as foregone investor returns keep on being restricted to small values.

References

Arnold, Lutz G. (2023), "On the neutrality of socially responsible investing: The general equilibrium perspective", *Theoretical Economics* 18 (2023), 65-95.

Arnold, Lutz G. (2025), "The Economics of Green Debt", *SSRN* Working Paper.

Baker, Malcolm, Daniel Bergstresser, George Serafeim, and Jeffrey Wurgler (2022), "The Pricing and Ownership of US Green Bonds", *Annual Review of Financial Economics* 14: 10.1-10.23.

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

Broccardo, Eleonora, Oliver Hart, and Luigi Zingales (2022), "Exit versus voice", *Journal of Political Economy*, volume 130, number 12, December 2022.

Cho, In-Koo, and David M. Kreps (1987), "Signaling Games and Stable Equilibria", *The Quarterly Journal of Economics*, Vol. CII, May 1987, Issue 2.

Dorfleitner, Gregor, Sebastian Utz, and Rongxin Zhang (2022), "The pricing of green bonds: external reviews and the shades of green", *Review of Managerial Science* (2022) 16:797–834.

Fama, Eugene F., and Kenneth R. French (2007), "Disagreement, tastes, and asset prices", *Journal of Financial Economics* 83 (2007) 667–689.

Flammer, Caroline (2021), "Corporate green bonds", *Journal of Financial Economics* 142, 499-516.

Gibbons, Robert (2011), "A primer in game theory", Harlow: Pearson Education Limited.

Hoepner, Andreas G. F., Ioannis Oikonomou, Zacharias Sautner, Laura T. Starks, and Xiao Y. Zhou (2024), "ESG shareholder engagement and downside risk", *Review of Finance*, 2024, 28, 483-510.

Krahnen, Jan, Jörg Rocholl, and Marcel Thum (2023), "A Primer on Green Finance: From Wishful Thinking to Marginal Impact", *Review of Economics* 2023; 74(1): 1–19.

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

Neagu, Florian, Tatarici Luminita, Florian Dragu, and Amalia Stamate (2024), "Are green loans less risky? Micro-evidence from a European Emerging Economy", *Journal of Financial Stability* 70 (2024) 101208.

Rothschild, Michael, and Joseph E. Stiglitz (1976), "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information", *The Quarterly Journal of Economics*, Volume 90, Issue 4, November 1976, Pages 629–649.

Selten, Reinhard (1965), "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit", *Zeitschrift für Gesamte Staatswissenschaft* 121: 301-24.

Selten, Reinhard (1975), "A Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games", *International Journal of Game Theory*, 4 (1): 25–55.

Spence, Michael (1973), "Job Market Signaling", *The Quarterly Journal of Economics*, Vol. 87, No. 3 (Aug., 1973), pp. 355-374.

Spiegel, Der (2024), "Abgasaffäre bei Volkswagen", URL: https://www.spiegel.de/thema/abgasaffaere_bei_volkswagen/ (last sighted: January 14th, 2025).

tagesschau (2025), " Bußgeld wegen Greenwashing – Millionenstrafe für Deutsche-Bank-Tochter DWS", URL: https://www.tagesschau.de/investigativ/ndr/wwf-greenwashing-dws-bussgeld-strafe-100.html (last sighted: April 14th, 2025).

Yu, Dai, and Liu Jihan (2022), "Signaling Game Analysis between Enterprises and Green Investors under Green Bond Financing", *Financial Engineering and Risk Management* (2022), Vol. 5 Num. 5, pp. 130-139.

Zhu, Qianjing, Xianglian Zhao, and Meihua Wu (2023), "Third-party certification: how to effectively prevent greenwash in green bond market? –analysis based on signalling game", *Environment, Development and Sustainability*.