Rating Under Asymmetric Information[1]

Christian Hilpert[2] Stefan Hirth[3] Alexander Szimayer[4]

Abstract

We study a dynamic signaling game where a firm, by its decision to stay solvent, signals its quality to a rating agency with the rating feeding back into the firm's cost of capital. Observing the firm's true cash flow blurred by a persistent measurement error, the error-minimizing rating agency learns dynamically through the firm's solvency decision. Firms observed with higher measurement error default earlier, inducing directional learning by successively eliminating measurement errors which are too high to be feasible. In a partially separating perfect Bayesian equilibrium in Markov strategies, the firm employs a measurement-error dependent cut-off strategy. We discuss the extensive economic consequences of such a learning mechanism.

JEL classification: G14, G24, G32, G33

Keywords: Rating, Asymmetric Information, Learning, Feedback Effect, Markov Perfect Bayesian Equilibrium

[1]We thank Antje Berndt, Stine Louise Daetz (SGF discussant), Herbert Dawid, Robert Durand, Alexander Eisl (DGF discussant), Sivan Frenkel (WFA discussant), Nils Friewald, Sebastian Gryglewicz, Michael Halling, Andrey Malenko, Angelo Ranaldo, Karl Schlag, Günter Strobl, Yuri Tserlukevich (FDU discussant), Peter Zweifel, and seminar participants at Aarhus University, Australian National University, Cass Business School, Curtin University, Hamburg University, and the University of Southern Denmark for helpful comments and discussions. Furthermore, we thank participants of the Finance Down Under 2018 Conference, German Finance Association (DGF) Conference 2017, Western Finance Association (WFA) Conference 2017, and the Swiss Society for Financial Market Research (SGF) Conference 2017 for helpful comments. This work was supported by Australian Research Council [DP160104737] and Deutsche Forschungsgemeinschaft [DFG282079427]. A previous version of this paper was titled "History Matters: Credit Rating under Asymmetric Information".

[2]Lingnan (University) College, Sun Yat-sen University, Guangzhou 510275, PR China. Email: `martin@mail.sysu.edu.cn`.

[3]University of Southern Denmark (SDU), Aarhus University and Danish Finance Institute. Address: University of Southern Denmark (SDU), Department of Business and Economics, Campusvej 55, 5230 Odense M, Denmark. Email: `shirth@sam.sdu.dk`.

[4]Hamburg University, Department of Economics and Social Science, Von-Melle-Park 5, 20146 Hamburg, Germany. Email: `alexander.szimayer@wiso.uni-hamburg.de`.

1 Introduction

Which information value does the strategic timing of option exercise contain? How and to which degree can signaling through option exercise mitigate the measurement error on an imperfectly observed state variable over time? Many economic problems require an option exercise decision in a situation in which information unfolds over time. Usually the information can be decomposed into the true realization of a state variable that is available only to informed players, and a measurement error that explains why outsiders only have access to an imperfect observation of the same variable.

In this paper, we focus on the case of the informed player being a firm's manager-owner choosing whether to stay solvent by injecting new equity in low cash-flow states. We face the question of how corporate bond markets and rating agencies as the outsiders learn the firm's quality and creditworthiness in presence of a measurement error regarding the firm's cash flow. Moreover, how does this learning feed back into the firm's strategic choice of staying solvent or defaulting on its debt? Answering these questions leads to a complex interaction, as the cost at which the firm can finance itself in the future will affect the firm's default decision, which again is the determinant of the firm's cost of debt capital.

We introduce a rating agency as a general representative of the bond market, as opposed to the firm's insiders who know the firm's true quality. Thus, we take a closer look at the strategic interaction between a rating agency and a firm in a setting in which the rating agency assesses imperfectly measured information relevant to the rating. Specifically, we consider a dynamic signaling game in continuous time in a random environment. The firm aims to maximize its stream of cash flows net of rating dependent interest payments. To this end, the firm's management chooses the optimal liquidation time given by a default threshold. The rating agency estimates the firm's default level with the aim of minimizing the forecasting error of this default threshold, with the rating feeding back into the firm's interest payments as in Manso (2013). We abstract from agency conflicts between the firm's management and the current owners in spirit of Ross (1977), and focus on the interaction with outsiders as in Morellec and Schürhoff (2011).

We model the relevant set of information in reduced form as the firm's cash flow. This information is available to the public as well as the rating agency, but only as an imperfect observation, because its assessment of the firm's information carries a measurement error. The measurement error can be induced by soft information that cannot be assessed accurately by the rating agency. The liquidation decision of the firm is the signaling device of the firm's management: A firm can signal solvency by not defaulting for low observed cash flow levels, in particular by means of the firm's equity owners injecting additional funds. We show that the firm's optimal strategy is a cut-off rule which depends on the measurement error. The firm's management will specify a cut-off level for the observed cash flow at which the present value of future expected cash flows is too low to justify the firm staying solvent, given the high interest payments due to the low rating and low prospects of recovery.

The central mechanism of our game is the way the rating agency learns and reduces the measurement error in the observed cash flow process that is interpreted as type in the signaling game. This measurement error, or type, is persistent as in Fershtman and Pakes (2012) which allows us to isolate the effect of learning out of observing the firm's strategy. To this end, we depart from previous models of learning by observing the cash flow dynamics with parameter uncertainty in financial markets that a player could resolve over time, see Pastor and Veronesi (2009) for an overview. Instead, we consider a stochastic setup similar to Grenadier, Malenko, and Malenko (2016), who study real options for a principal agent problem. In such a setting, the imperfectly observed cash flow dynamics does not deliver information in its own right, so that the rating agency only updates its information as it observes action by the firm's management, namely to default or stay solvent. We enrich this setup by rating-dependent performance-sensitive debt as in Manso, Strulovici, and Tchistyi (2010) and Manso (2013). Performance-sensitive debt allows us to analyze effects similar to those occurring with the roll-over of short-term or finite-maturity debt. Hence, it should be seen as a modeling alternative to the classical papers introducing finite maturity into infinite-horizon structural models, such as Leland and Toft (1996), Leland (1998), He and Xiong (2012a,b), and He and Milbradt (2016).

In our model, the rating agency learns from observing non-default as the observed cash flow's historic low deteriorates. We show that the rating agency's optimal strategy consists of issuing a higher rating for the same current cash flow, if the historical minimum has been sufficiently low. Because the rating agency estimates default thresholds based on its belief over possible measurement errors, only historic lows of imperfectly observed cash flow allow to infer the measurement error. This result is intuitive: Bayesian updating of its beliefs over the measurement error leads to its assessment that it cannot have grossly overestimated the cash flow, because such a firm would have defaulted in the previous low cash flow state. This leads the rating agency to rule out the most extreme measures of positive measurement errors. If the observed cash flow then hits further historic lows, the potentially possible overestimation of the cash flow keeps dropping, which we coin directional learning.

We obtain a perfect Bayesian equilibrium in Markov strategies as introduced by Maskin and Tirole (1988) in spirit of Grenadier, Malenko, and Malenko (2016), Acharya and Ortner (2017), and Halac and Kremer (2018). The directional learning causes our equilibrium to be only partially separating, because the rating agency cannot fully pin down the measurement error, or types, prior to the firm's default, but may rule out some measurement errors, or types, by observing non-default. Moreover, we derive constraints for such a rating policy such that it maintains the incentives for the rated firm not to mimic another firm that has a higher true cash flow. While we employ a similar equilibrium concept as Grenadier, Malenko, and Malenko (2016), we extend the setup substantially by a feedback effect between the cost structure and the decision to default. In this paper, we restrict our analysis to a signaling game of two players with a continuous flow of information. Dynamic learning in a setting with a large player facing a mass of small players has been studied in Faingold and Sannikov (2011), while Sannikov and Skrzypacz (2010) extend the classical stream of a continuous Brownian motion to be interrupted by discontinuous information shocks. In this spirit, we contribute also to the literature on the strategic exercise of real options in finance, which has previously been studied in the context of IPOs by Bustamante (2012), for corporate investments by Hirth and Uhrig-Homburg (2010), Grenadier and Malenko (2011), Morellec and Schürhoff (2011), and Grenadier, Malenko, and Strebulaev (2014), for dynamic agency problems for real options in Gryglewicz and Hartman-Glaser (2014) or, in a more general setting by Kruse and Strack (2015).

Rather than focusing on learning, many recent contributions in the credit rating literature assume perfect observation of issuer quality, see, e.g., Bolton, Freixas, and Shapiro (2012) and Hirth (2014). While Manso (2013) models the rated entity's cash flow process in continuous time as we do, he does not consider information asymmetry between the rated entity and either the rating agency or the financial market. Consequently, he does not analyze learning.[1] In contrast, the directional learning in our model can explain what we coin “ex-post rating inflation”. In our structural model setup, the rating agency produces an unbiased estimate of the firm’s default level and refines the cash flow estimation by ruling out the most overestimated cash flows in its learning over time. At default, the firm’s cash flow is therefore overestimated, leading to a delay of default and lower asset values being available to creditors upon default. Note that the ex-post rating inflation occurs even though the rating agency’s objective is to produce unbiased ratings.

Our paper is related to Duffie and Lando (2001), who study a structural model of debt valuation featuring imperfect information. What distinguishes our contribution from Duffie and Lando (2001) is that in their model, the debt contract (in particular, the interest payment) is fixed ex ante, and the creditors’ (secondary bond market’s) learning of the asset process feeds back neither into the cost of debt nor the firm’s endogenous default decision. In contrast, our model is more suitable in describing a repeated interaction in which the firm’s observable credit quality today affects its future financing conditions, and thus also its considerations whether to inject new equity or to default already today. Papers that model learning with Bayesian updating include Board and Meyer-ter Vehn (2013), Fulghieri, Strobl, and Xia (2014), Frenkel (2015), and Thomas (2018). However, these neither focus on the imperfectly observed state variables of a rated entity, nor would their frameworks be suitable for explaining the dynamics of credit spreads. As Bolton, Freixas, and Shapiro (2012) put it, “there are very few such models in the industrial organization literature for obvious reasons of tractability”.

Our framework allows not only qualitative statements on the behavior of firm and rating agency. Moreover, we can explain the dynamic evolution of ratings and credit spreads over time, similar to, e.g., Jarrow, Lando, and Turnbull (1997) and Duffie and Lando (2001). Thus, we also relate to the asset pricing perspective on credit risk. We calibrate our model to actual credit spreads for the respective rating classes, and we can make predictions on how much learning matters in terms of dollar value for the rated entity. In a specific example we show that learning can decrease the credit spread by more than 300 basis points. This is the motivation why we rely on a rather

[1] Note further that “When the cash-flow process of the firm follows a geometric Brownian motion, equilibrium of the game is unique” (Manso (2013, p.543)), just as in our model. The case of multiple equilibria, which drives a large part of his paper’s results, requires a mean-reverting cash-flow process.

involved mathematical framework in continuous time. Still, we are able to stay within the class of structural credit risk models with endogenous default, in the tradition of Leland (1994) and Goldstein, Ju, and Leland (2001).

As an extension, we deviate from the assumption of an unbiased rating agency and examine the effect of the rating agency taking a biased view on the firm in two possible directions: Firstly, a rating agency attempting to protect investors can take a conservative estimate of the default threshold by avoiding an overestimation of the measurement error at the expense of the firm. In this case the agency assumes worse measurement errors, leading to higher interest payments and by this to earlier defaults. Consequently, the directional learning is accelerated. In the extreme case, this policy implements the perfect-information default threshold and eliminates ex-post rating inflation completely. Secondly, a rating agency aiming to maximize revenues by providing progressive ratings will provided lower estimated default thresholds, which slows down directional learning. In the extreme case, such an agency maximally contributes to ex-post rating inflation, because it refuses to learn.

The remainder of the paper is organized as follows: In Section 2, we introduce the rating game between the rating agency and the firm. Section 3 introduces the best response strategies of both players. Subsequently, we derive and compute the rating game equilibrium in Section 4. In Section 5, we provide an extensive analysis. Section 6 concludes.

2 A Rating Model under Asymmetric Information

We set off by formally describing the interaction between a levered firm and a rating agency. The firm's payoff to its equity holders depends on its continuous stream of cash flows net of its interest payments on outstanding debt. The rating agency's success is measured by its reputation, namely its ability to deviate as little as possible from a precise and unbiased rating. As a motivation for this objective function, note that on a competitive debt market, creditors have to price credit risk as accurate as possible to be successful.[2]

[2] See also Section 5.6, in which we extend the model interpretation to a setting without a rating agency. Then, creditors assess the credit risk themselves, and the effect on the cost of debt occurs when the borrower rolls over maturing debt.

The rated firm generates a stream of non-negative cash flow at the rate $X = (X_t)_{t\geq 0}$ (hereafter: cash flow) and pays interest on its outstanding debt at the rate $C = (C_t)_{t\geq 0}$. The interest payment rate C depends on the rating agency's rating of the firm $R = (R_t)_{t\geq 0}$, which is detailed below. The firm's debt is modelled as performance-sensitive debt (PSD), see Manso, Strulovici, and Tchistyi (2010). In a more general interpretation, this can be seen as a way to model the roll-over of maturing debt, which will make the firm's cost of debt capital similarly dependent of on the outsiders' current perception of the firm's credit quality. The rating agency's rating R is based on the imperfectly observed cash flow $D = (D_t)_{t\geq 0}$ of the firm's cash flow X, which is given by $D = \tilde{\theta}\, X$. The persistent measurement error $\tilde{\theta}$ is drawn and learned by the firm at initial date $t = 0$. The rating agency does not know $\tilde{\theta}$. However, the law of $\widetilde{\theta}$ is common knowledge, see, e.g., Grenadier, Malenko, and Malenko (2016) for a related setup. Based on the information from observing D, the rating agency estimates the firm's critical cash flow level where default occurs, i.e., $\hat{D}^\star = (\hat{D}^\star_t)_{t\geq 0}$, and issues the rating as distance to default $R_t = D_t/\hat{D}^\star_t$, which we discuss in more detail below.

The firm is risk-neutral and aims to maximize the net present value of the cash flows net the interest payments. For a realization θ, the firm thus identifies the optimal time $\tau(\theta)$ when to default with observed default threshold $D^\star(\theta) = D_{\tau(\theta)}$, $\theta \in \Theta$. The rating agency's objective is to minimize the reputation costs, which we explain by reduced future business in case the rating is not accurate. It learns from the firm's endogenous survival of low cash flow states. Apart from the current cash flow level, an important firm characteristic is therefore the lowest observed cash flow $E = (E_t)_{t\geq 0}$, with $E_t = \inf_{0\leq s\leq t} D_s$, $t \geq 0$. From the perspective of the rating agency, E is the information generating process, and the rating agency's estimate of the firm's default threshold is a function of this variable, i.e. $\hat{D}^\star = g(E)$.

Specifically, the firm's cash flow rate X satisfies

$$\mathrm{d}X_t = \mu\, X_t\, \mathrm{d}t + \sigma\, X_t\, \mathrm{d}W_t\,, \text{ for } t > 0\,, X_0 \in \mathbb{R}^+\,, \tag{1}$$

where μ and σ represent the cash flow's growth rate and volatility, respectively, with $\mu < r$, r being the risk-free interest rate, and $W = (W_t)_{t\geq 0}$ is a Wiener process. The growth rate μ and volatility σ are common knowledge, but X is known only by the firm. The imperfect observation D of the cash flow is

$$D_t = \tilde{\theta}\, X_t \,, t \geq 0 \,, \tag{2}$$

which depends on the measurement error $\tilde{\theta}$, but has identical dynamics as the cash flow X, i.e., $\mathrm{d}D_t = \mu\, D_t\, \mathrm{d}t + \sigma\, D_t\, \mathrm{d}W_t$, for $t > 0$. As a consequence, no information on $\tilde{\theta}$ and hence X can be obtained from the imperfect observation D. The persistent measurement error $\tilde{\theta}$ is drawn by nature and learned by the firm at initial date $t = 0$. The rating agency does not know $\tilde{\theta}$. Rather, it overestimates the true cash flow for a type $\theta > 1$, while a type $\theta < 1$ leads the rating agency to underestimate the true cash flow. To put it the other way round, given a specific observation D_t of the cash flow, a higher θ means that the actual cash flow X_t is lower. It is common knowledge that $\tilde{\theta}$ is a random draw from the distribution $\mathbb{P}_{\tilde{\theta}}$ on $\Theta = [\underline{\theta}, \overline{\theta}]$, with $0 < \underline{\theta} < \overline{\theta} < \infty$, and is independent of the Wiener process W. Therefore, the rating agency does not have to learn the parameters of the cash flow dynamics, but the learning in our model solely concerns the persistent measurement error $\tilde{\theta}$ as in Fershtman and Pakes (2012). Learning in this setting is restricted to learning from strategic behavior, while pure statistical learning in the from of observing the imperfectly observed cash flow process alone can by definition not reveal any information. In contrast, think of the well-known case of a mean reverting process with unknown mean. In such a situation, a mere observation of the process realization over time can be useful to find better estimates of the unknown parameter. We assume that the distribution $\mathbb{P}_{\tilde{\theta}}$ admits a density, which is bounded from above and away from zero, which is our prior ϕ. We write $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ and $\mathbb{G} = (\mathcal{G}_t)_{t \geq 0}$ to designate the filtration generated by X and D, respectively. Cash flow X and type $\tilde{\theta}$ are independent by assumption.[3] For an interpretation of the model, one could think of the "cash flow" more broadly as some characteristics determining firm value.

The firm issues performance sensitive debt as in Manso, Strulovici, and Tchistyi (2010), which is specified by the interest payment rate C depending on the rating (performance) $R \geq 1$. In particular, C is a non-increasing function $C : [1, \infty] \to \mathbb{R}_0^+$ which we assume bounded away from zero and bounded from above. Hence satisfies $0 < \underline{C} = C(\infty) \leq \overline{C} = C(1) = C < \infty$. The debt contract is perpetual in nature and

[3] Formally, the firm's information set at t is given by $\sigma(\tilde{\theta}) \vee \mathcal{F}_t$, for $t \geq 0$. Since $\tilde{\theta}$ is known to the firm at $t = 0$ as well as $\tilde{\theta}$ and X are independent, we can condition on $\tilde{\theta} = \theta$ and work with $\mathbb{F}$.

does not specify a repayment of principal. Thus, apart from the rating-dependence of C, it can be seen as a consol bond as in Black and Cox (1976) and Leland (1994).

The rating R is specified as follows. Given the available information, the rating agency assigns the firm's rating based on the estimated default level $\hat{D}^\star$. Formally, the default level $\hat{D}^\star$ is strictly positive and $\mathbb{G}$-adapted, that is, the rating agency only employs the information contained in the observed cash flow D, but cannot access the firm's private information of the real cash flow X, or, equivalently $\tilde{\theta}$ (with $\tilde{\theta} = D/X$). Thus, the rating agency can enhance its default assessment by continuously learning the measurement error from observing the firm's solvency signal in low cash flow states. We define the rating R issued by the rating agency as the distance to default, which we identify as

$$R_t = \frac{D_t}{\hat{D}_t^\star}\,, t \geq 0\,. \tag{3}$$

The higher the firm's observed cash flow D is relative to the predicted default level $\hat{D}^\star$, the higher will the firm's rating be. The rating R takes values in $[1, \infty]$, where $R_t = 1$ implies a firm that is expected to default immediately, and $R_t = \infty$ corresponds to a default-free firm from the rating agency's perspective. As the rating agency notices changes of the imperfectly observed cash flow over time, it adjusts the rating accordingly, because the estimated default threshold $\hat{D}^\star$ changes.

The firm observes the cash flow process X and knows the realization of the measurement error θ. Based on this information, the firm chooses the time to liquidate, denoted by $\tau(\theta)$. Because the firm is aware of the measurement error at the start and $\tilde{\theta}$ and the Wiener process W are independent, we can write $\tau = (\tau(\theta))_{\theta\in\Theta}$, where $\tau(\theta)$ is an $\mathbb{F}$-stopping time, for $\theta \in \Theta$, see also Footnote 3.

Observing the firm's decision either to default or to signal solvency by not defaulting, the rating agency gradually learns the soft information of the firm's cash flow over time. Hence, the rating agency forms its belief $\pi = (\pi_t)_{t\geq 0}$ regarding the measurement error $\tilde{\theta}$ which we interpret as type in the given signaling game. We restrict our analysis to beliefs which are absolutely continuous with respect to the prior $\mathbb{P}_{\tilde{\theta}}$, and thus also π has a density, say ϕ^π, with

$$\phi_t^\pi(\theta) = L_t^\pi(\theta)\,\phi(\theta)\,, \text{ for } \theta \in \Theta, t \geq 0\,, \tag{4}$$

where $L_t^\pi = (L_t^\pi(\theta))_{\theta\in\Theta}$ describes the evolution of the probabilities for each measurement error by using the information available in the market.[4] While we formulate the beliefs here in general, we will later use perfect Bayesian Markov equilibrium as the equilibrium concept. Thus, the rating agency will update its belief about the measurement error according to Bayes rule whenever possible. Because Bayesian updating is infeasible if the rating agency observes actions which are not used by a firm with any possible measurement error, we need to specify off-path beliefs.[5] Following Grenadier, Malenko, and Malenko (2016), we make the standard assumption that in such a case the beliefs remain unchanged:

Assumption 1. *If at any t, the rating agency's belief π_t and the firm's action are such that no possible measurement error could use this action in equilibrium, then the belief is unchanged.*

By this learning mechanism, the strategy of the firm in all states $\tau(\theta)$ affects the rating agency's belief ϕ^π, hence the rating agency's strategy in form of the estimated default level $\hat{D}^\star$ feeds back into the specific strategy $\tau(\theta)$. Thus, the firm's strategy is a measurement error-dependent stopping time, which we will later assume to be a Markov strategy in a suitably extended state space. Figure 1 displays an example of the rating agency's Bayesian updating of its belief: Setting off from a prior (solid line), by observing the firm's default or survival, the rating agency learns and reassesses the probabilities for each measurement error, which in turn sharpens the rating agency's belief. As the cash flow evolves, the beliefs become more and more precise (dotted and dashed lines).

The strategies for the firm $(\tau(\theta))_{\theta\in\Theta}$ and the rating agency $\hat{D}^\star$, respectively, are now specified in a general form. The Markov property of the state processes, (X, Y) with running minimum $Y = (Y_t)_{t\geq 0}$ of X, i.e. $Y_t = \inf_{0\leq s\leq t} X_s$, $t \geq 0$, or, (D, E) as observed by the rating agency, suggests that it is sufficient to consider Markov strategies. The set of admissible Markov strategies $\mathcal{A}_f$ for the firm is given by default levels $f(\theta)$ of firm cash flow as observed by the rating agency, for $\theta \in \Theta$. The first time

[4] L_t^π is a family of non-negative $\mathcal{G}_t$-measurable random variables with $\int_\Theta L_t^\pi(\theta)\,\phi(\theta)\,\mathrm{d}\theta = 1$, $t \geq 0$.

[5] For example, we will show in later sections that for each measurement error, there will be default threshold for the firm. If the rating agency observes that the firm does not default, even though firms with all measurement errors should have defaulted, this strategy should not occur in equilibrium, so that we need to specify beliefs in this case.

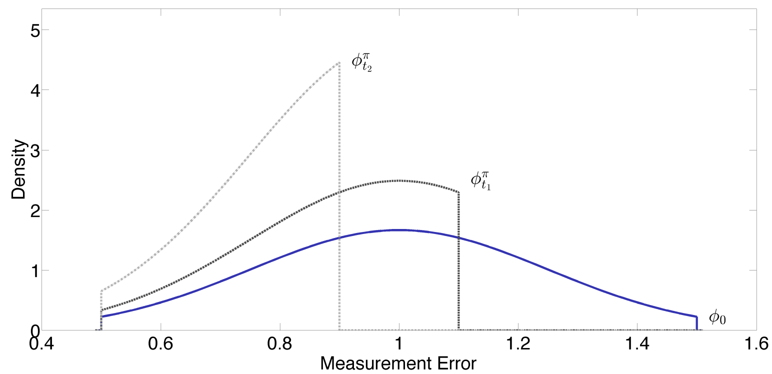


Figure 1: Bayesian updating of beliefs. *This figures displays an example of how the rating agency updates the prior (solid line) to the updated belief at times t_1 and t_2 (dotted and dashed line, respectively). It plots the density for each measurement error of the soft information $\tilde{\theta}$, which has a support of $\Theta = [\underline{\theta}, \overline{\theta}] = [0.5, 1.5]$. As time evolves, the rating agency rules out overestimated soft information.*

the cash flow as observed by the rating agency D falls below $f(\theta)$, the firm defaults, i.e. $\tau(\theta) = \inf\{t \geq 0 : D_t \leq f(\theta)\}$, for $\theta \in \Theta$.[6] The set of admissible strategies is[7]

$$\mathcal{A}_f = \left\{ f : \Theta \to \mathbb{R}_0^+, f \text{ is measurable} \right\}. \tag{5}$$

The admissible strategies for the rating agency $\mathcal{A}_g$ are functions of the information generating process, namely the minimum observed cash flow E, i.e. $\hat{D}^\star = g(E)$, and

$$\mathcal{A}_g = \left\{ g : \mathbb{R}_0^+ \to \mathbb{R}_0^+, g \text{ is measurable} \right\}. \tag{6}$$

In addition, we require that g is reasonable from a financial economics perspective. In (D_t, E_t), the predicted default at $g(E_t)$ should be attainable, and thus $g(E_t) \leq E_t$, or, $g \leq Id$, where Id is the identity on $\mathbb{R}_0^+$, $Id(x) = x$, for $x \in \mathbb{R}_0^+$. Firm survival in bad times, that is for a decreasing running minimum E_t, potentially signals quality. Then the estimated default threshold $\hat{D}^\star = g(E_t)$ should be adjusted downwards or remain constant, i.e. g is non-decreasing. However, we demand that a rating $R = D/g(E)$

[6] We are taking the perspective of the rating agency to avoid problems when discussing the matters from the perspective of both parties, firm and rating agency. The critical default level in the firm cash flow is then $f(\theta)/\theta$, i.e. $\tau(\theta) = \inf\{t \geq 0 : X_t \leq f(\theta)/\theta\}$, since $D = \theta\, X$.

[7] Strictly speaking, the set of Markov strategies is much larger and consists of stopping times that are given by first entry times in a measurable set $B(\theta) \subseteq \mathbb{R}^2$, i.e. $\tau(\theta) = \inf\{t \geq 0 : (X_t, Y_t) \in B(\theta)\}$, $\theta \in \Theta$. However, the subsequent Proposition 2 shows that this restriction is innocent.

should not improve when the cash flow hits a new all-time low. Formally, a new all-time low occurs at time t for $D_t = E_t$. Then the rating is $R_t = D_t/g(E_t) = E_t/g(E_t)$, and has to be non-increasing in E_t to avoid a better rating in case of a new all-time low. The latter is equivalent to the constraint that g/Id being non-increasing. Thus we state reasonable strategies for the rating agency $\mathcal{A}_g^{\mathbf{C}}$ that avoid incentives for the firm to intentionally downward-adjust its observable cash flow status with the aim of reducing its long-term interest payments

$$\mathcal{A}_g^{\mathbf{C}} = \{g \in \mathbf{C}(\mathbb{R}_0^+, \mathbb{R}_0^+) : g \text{ non-decreasing, } g/Id \text{ non-increasing, } g \leq Id\}, \tag{7}$$

where $\mathbf{C}(\mathbb{R}_0^+, \mathbb{R}_0^+)$ denotes the set of continuous functions mapping $\mathbb{R}_0^+$ onto $\mathbb{R}_0^+$.[8] Restricting the rating agency to strategies contained in $\mathcal{A}_g^{\mathbf{C}}$ excludes for example that actively reducing cash flows would improve ratings. It also avoids accounting manipulation in the form of excessive use of deferred income with the purpose of rating optimization.

We now specify the expected payoffs as functions of both the firm's and the rating agency's strategies, $(\tau, \hat{D}^\star)$, with $\tau = (\tau(\theta))_{\theta \in \Theta}$ and $\tau(\theta) = \inf\{t \geq 0 : D_t \leq f(\theta)\}$, $\theta \in \Theta$, and $\hat{D}^\star = g(E)$, where $f \in \mathcal{A}_f$ and $g \in \mathcal{A}_g$. The firm has a measurement-dependent expected payoff equal to the discounted stream of real cash flow minus rating-dependent interest C, as the perceived soft information feeds back into the interest payments. Specifically, noting that $X = D/\tilde{\theta}$ and $R = D/\hat{D}^\star$, we have

$$U_F^{(\theta)}(\tau, \hat{D}^\star) = \mathbb{E}\left[\int_0^{\tau(\theta)} e^{-r\,t}\left(D_t/\theta - C(D_t/\hat{D}_t^\star)\right) \mathrm{d}t\right], \text{ for } \theta \in \Theta\,. \tag{8}$$

If the interest payments exceed the true cash flow at a given time, the firm does not have to default, but the owners can inject further capital. The firm chooses the default time τ with the objective to maximize $U_F^{(\theta)}(\tau, \hat{D}^\star)$, namely the present value of cash flows after interest payments to creditors (see Equation (16) below for the formal specification of the firm's optimization problem).

This rating agency aims to maximize its reputation as an objective certifier of the firm's default risk in spirit of the fair certifier in Lerner and Tirole (2006). The rating agency chooses the estimated default threshold $\hat{D}^\star$ and thus its rating scale with

[8] If g is non-decreasing and g/Id is non-increasing, then g is continuous.

the objective to minimize its reputation costs, which we explain by reduced future business in case the rating is not accurate. Conditional on its belief π and the firm's measurement error-dependent liquidation strategy $\tau(\theta)$, the rating agency expects negative costs from a rating with an estimated default threshold $\hat{D}^\star$ to be, given the rating agency's information,

$$U^\pi_{RA}(\tau, \hat{D}^\star) = -\mathbb{E}\left[\int_0^\tau e^{-\rho\, t} k^\pi_t \,\mathrm{d}t\right] , \tag{9}$$

with the cost rate

$$k^\pi_t = \int_\Theta (\hat{D}^\star_t - f(\theta))^2\, \phi^\pi_t(\theta)\,\mathrm{d}\theta\,,\ \text{for } t \geq 0\,, \tag{10}$$

and the rating agency's discount parameter ρ, which measures the time preference over future reputation losses. For the rating agency, we define the expected reputation costs as the squared deviation of the estimated default threshold, conditional on the belief, and the true default threshold, dependent on the perceived soft information.

Later, we will generalize this cost rate by allowing for an asymmetric effect on the cost rate. This generalized structure captures different attitudes of the rating agency towards its stakeholders: If it is more concerned with protecting investors, the rating agency would be biased towards avoiding overestimations and estimate the default threshold conservatively. In the extreme case, it would always assume the worst possible measurement error from the investors's perspective. On the other hand, a rating agency aiming for maximizing revenues and its market share, would rate more progressively to please their clients. A detailed discussion of this aspect postponed to Section 5.5.

We focus on equilibria in pure strategies. The equilibrium concept is perfect Bayesian equilibrium in Markov strategies, which requires that the rating agency's strategies are sequentially optimal, beliefs are updated according to Bayes' rule whenever possible, and the equilibrium strategies are Markov. In particular, the Markov property requires that the firm's and the rating agency's strategies are only functions of the payoff-relevant information at any time t, i.e., measurement error $\tilde{\theta}$ and the current value of the state process (D_t, E_t) (equivalently (X_t, Y_t)) for the firm, and the beliefs ϕ^π_t about $\tilde{\theta}$ and the current value of the state process (D_t, E_t)

for the rating agency. The formal definition of the perfect Bayesian equilibrium in Markov strategies is presented in Appendix B and is subsequently referred to simply as equilibrium. Note that our state space processes for firm and rating agency contain both the observed cash flow process and its running minimum. Therefore, our players have a memory beyond the current value of the cash flow process, although the Markov property holds for the extended state space.

We consider an exogenously specified interest payment rate C which depends on the firm's rating R, and we assume C to be sufficiently sensitive to the rating.[9] However, for the subsequent equilibrium analysis the sensitivity of C cannot be overly excessive. In order to control the sensitivity of C, we make the following assumption.

Assumption 2. *Assume that the interest payment rate C satisfies for some $0 < L_C < 1$ that*

$$C(z) \leq C(z') \leq (z/z')^{L_C} C(z)\,, \text{ for } 1 \leq z' \leq z\,. \tag{11}$$

To confirm the validity of Assumption 2, we give an outlook on the parametrization based on market data that we will use in our analysis in Section 5. Figure 2 illustrates the interest rate and rating structure obtained from the perfect-information case. Appendix A.1 shows that this interest rate structure indeed satisfies Assumption 2.

3 Best Responses and Learning

Having specified the strategies of both the firm and the rating agency in Section 2, we can now turn to the best responses of both players, and describe how the rating agency learns the firm's true cash flow. In particular, the rating agency's best estimate of the firm's optimal liquidation decision is a continuously updated default barrier,

[9] In contrast, if the interest payment rate is not sensitive to the firm's rating, then the firm's payoff is not rating-dependent, leaving the firm indifferent to the rating. Then, the rating issued by the rating agency has no feedback effect on the firm strategy and becomes irrelevant, and thus we have no game between the two parties. In this degenerated case, the firm's default level is a constant f_1 in the firm's real cash flow X and hence linear in the rating agency's imperfectly observed cash flow $D = \tilde{\theta}\, X$, i.e. $f(\theta) = \theta\, f_1$.

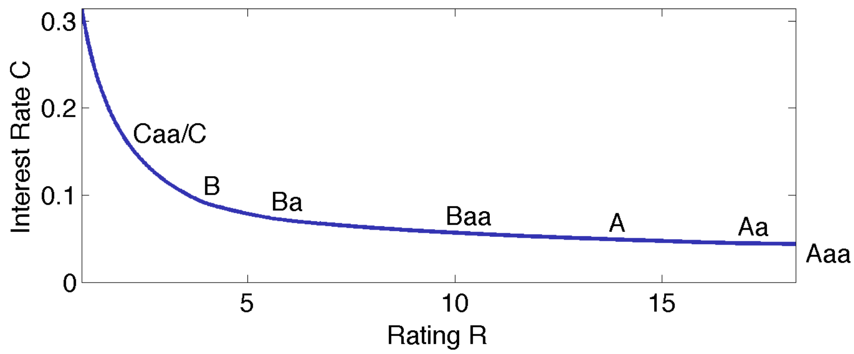


Figure 2: Interest rates for different ratings. *This graph displays the average interest payment rate C as a function of the rating R.*

dependent on the observed cash flow trajectory, and especially on its running minimum. For the observed cash flow trajectory, the rating agency infers up to which degree of overestimation the firm would have defaulted, ruling out the most overestimated types in its consistent belief. In its best response, the firm specifies its optimal liquidation decision as a response to a rating strategy, based on the type, or measurement error. The firm's strategy consists of the classical equity holder trade-off between the cost of injecting additional funds to cover presently negative net cash flows and the benefit of receiving positive net cash flows in the future, when the cash flow generating process recovers, see, e.g., Black and Cox (1976), Leland (1994), and Goldstein, Ju, and Leland (2001). Our new contribution is that the injection of additional funds makes the rating agency (and other external stakeholders) update their beliefs about the firm's quality and thus leads to lower financing costs in the future. These lower future financing costs again feed back into the equity holder trade-off today, because the injection of additional funds remains attractive for lower present cash flows. Thus, liquidation can be triggered at a later time.

3.1 Best Response and Learning of the Rating Agency

In this section, we characterize the best response of the rating agency $\hat{D}^\star = g(E)$ and its consistent belief π given the firm's liquidation strategy τ. The firm's strategy is type-dependent and is given by $f \in \mathcal{A}_f$ with $\tau(\theta) = \inf\{t \geq 0 : D_t \leq f(\theta)\}$, for all $\theta \in \Theta$. The rating agency's consistent belief is driven by E. At time t, types θ with $f(\theta) \geq E_t$ can be discarded, since default obviously has not occurred yet. The consistent belief is therefore $\pi_t = \mathbb{P}_{\tilde{\theta}}\big|_{f(\tilde{\theta})<E_t}$ with density given in (13) below. For a given belief π, the best response g is the solution to the maximization problem given in (12). Recalling the rating agency's expected payoff given in Equation (9), it optimizes its expected payoff over admissible estimated default thresholds of the firm

$$\sup_{\hat{D}^\star = g(E),\, g\in\mathcal{A}_g} U^{\pi}_{RA}(\tau, \hat{D}^\star) = -\inf_{g\in\mathcal{A}_g} \mathbb{E}\left[\int_0^\tau e^{-\rho t}\int_\Theta (g(E_t) - f(\theta))^2\, \phi^\pi_t(\theta)\,\mathrm{d}\theta\,\mathrm{d}t\right]. \quad (12)$$

For each t, the optimal rating agency strategy $g(E_t)$ minimizes the mean squared error for estimating the type-dependent default threshold $f(\theta)$ using the current belief. The latter is given by the respective conditional expectation, see (14).

This allows us to formulate both a consistent belief and the rating agency's best response in the following proposition.

Proposition 1 (Rating Agency's Best Response)**.** *Let a firm strategy τ be given by a function $f \in \mathcal{A}_f$, with $\tau(\theta) = \inf\{t \geq 0 : D_t \leq f(\theta)\}$, $\theta \in \Theta$. Then the rating agency's consistent belief is given by*

$$\phi^\pi_t(\theta) = \frac{\mathbf{1}_{f(\theta)<E_t}}{\int_\Theta \mathbf{1}_{f(\theta')<E_t}\,\phi(\theta')\,\mathrm{d}\theta'}\,\phi(\theta)\,, \text{ for } \theta\in\Theta \text{ and } 0 \leq t < \tau\,. \quad (13)$$

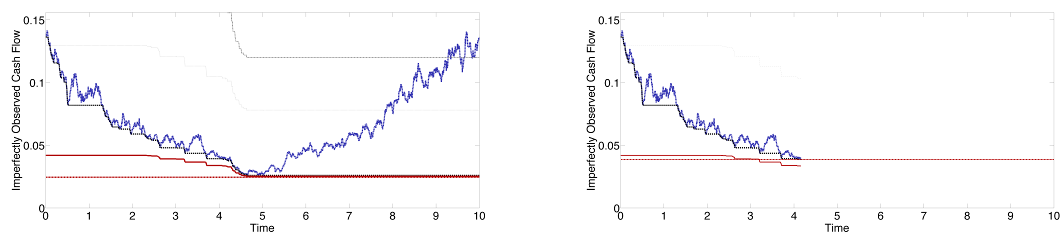
The rating agency's corresponding best response to f is given by $\hat{D}^\star = g(E; f)$, with

$$g(e; f) = \begin{cases} \mathbb{E}\left[f(\tilde{\theta})\,\middle|\, f(\tilde{\theta}) < e\right], & \text{for } e > \inf_{\theta\in\Theta} f(\theta)\,, \\ e\,, & \text{else.} \end{cases} \quad (14)$$

and is bounded by Id, i.e. $g(e; f) \leq e$, for $e \geq 0$, as well as non-decreasing. Moreover, if f is strictly increasing, then $g(\cdot; f)$ is continuous and strictly increasing on $f(\Theta)$.

Proposition 1 specifies that for the rating agency, the optimal response to a type-dependent firm liquidation strategy is issue a rating based on an estimated liquidation barrier in accordance with its consistent belief about the type, namely by observing which types should have defaulted for the observed cash flow process. Economically speaking, a rating agency's belief is consistent if the likelihood for the type is in accordance with firm's default behavior. Hence, if the rating agency sets the rating scale inducing a specific default barrier and the imperfectly observed cash flow reaches this level, then a firm of a sufficiently overestimated type defaults while a firm of an underestimated type may not. This observed default behavior provides useful information to the rating agency, and allows it to update its strategy, i.e. the rating scale, consistent with its belief about the types.

Comparing two states with identical current observed cash flow but with different historical cash flow paths, the state with lower historical minimum cash flow exhibits a better rating. However, this does not imply that a decreasing cash flow improves the current rating, in contrary the rating is worsening for a decreasing cash flow, see the specification of the set of admissible rating strategies $\mathcal{A}_g^{\mathbf{C}}$ in (7) and the discussion there. However, the survival of low observed cash flow levels in the past improves future ratings spurred by the rating agency's learning. Note that learning occurs when the observed cash flow falls to a level where a given type, or, measurement error, may default, which are typically very low levels with a short distance to default.



(a) *No Default.* ***(b)*** *Default.*

Figure 3: Best responses of the rating agency. *This graph displays the best response of the rating agency as a default threshold, dependent on the imperfectly observed cash flow. The firm is initially rated B and has an underestimated cash flow with $\theta = 0.5$ (left graph) and an overestimated cash flow with $\theta = 1.2$ (right graph). The blue line represents the imperfectly observed cash flow, and the black line its running minimum. The gray dashed line is the imperfectly observed cash flow limit for the B-rating, while the gray dotted line represents the cash flow limit for the BB-rating. The solid red line is the estimated default threshold by the rating agency, and the dashed red line is the true default threshold of the firm, which is unobservable to the rating agency.*

Figure 3 illustrates the best response of the rating agency, as well as the continuous learning. Panel 3a shows a sample path of underestimated cash flow ($\theta = 0.50$). The parametrization is taken from our analysis in Section 5. The rating agency observes an initial cash flow of 0.1366, for which the rating agency assigns just a B-rating and the firm pays interests $C = 0.1470$.[10] The true cash flow is $0.1366/0.5 = 0.2732$. Under

[10]We have $R = 0.1366/0.0419 = 3.26$, see Section 5. For $C(R) = 0.1084$, see Figure 2 and Table 1.

perfect information, the firm would have a BB-rating and pay interests of $C = 0.0732$, see Table 1. In the graph, the blue line represents the imperfectly observed cash flow, and the black line its running minimum. The gray dashed line is the imperfectly observed cash flow limit for the B-rating, while the gray dotted line represents the cash flow limit for the BB-rating. The solid red line presents the default threshold estimated by the rating agency, and the dashed red line is the true default threshold of the firm, which is unobservable to the rating agency.

As the imperfectly observed cash flow decreases, the rating worsens to a C/CCC-rating already in the first year. However, as the running minimum falls, the rating agency continuously adjusts the best response default threshold, as it rules out the most overestimated types, who would have defaulted for the observed path. As the imperfectly observed cash flow increases again, a rating update from C/CCC to B takes place when the observed cash flow exceeds a level of around 0.08, and from B to BB when the observed cash flow exceeds a level of around 0.12. Consequently, when the observed cash flow again reaches the level of 0.1366 as at the start of the observation period, the rating corresponds now to the perfect information rating of BB, rather than the B-rating applied earlier for the same observed cash flow level, and effectively reducing the interest by 352 basis points. This substantial change is the direct result of the rating agency's updated beliefs, after observing a temporarily low cash flow and ruling out the most severe cash flow overestimation (see also Figure 1).

In contrast, Panel 3b illustrates the best response of the rating agency for a sample path of an initially overestimated cash flow ($\theta = 1.2$). The true cash flow is now $0.1366/1.2 = 0.1138$ and the perfect information rating would be C/CCC. In that case, the firm would have to pay interests of $C = 0.1276$.[11] Note though that the rating agency observes again the cash flow of 0.1366, just as in the underestimated cash flow case. Therefore it again assigns a B-rating, although the true cash flow is much lower in this scenario. Once the observed cash flow deteriorates sufficiently in this case, the firm defaults as the cash flow hits the true default threshold and the rating game ends. For this case, the firm will stay alive longer than in the perfect information case, as it has to pay lower interest (initially $C = 0.1084$ instead of $C = 0.1276$). Still, we ensure by Eq. (7) that the firm cannot mimic a situation with less overestimated cash flow.

[11] We have $R = 0.1138/0.0419 = 2.72$, see Section 5. For $C(R) = 0.1276$, see Figure 2 and Table 1.

Panel 3b also indicates what we coin “ex-post rating inflation”.[12] As the minimum observed firm cash flow deteriorates, there will be a time from which on the rating agency will unconsciously inflate the firm's ratings until the end of the rating game. In Panel 3b, this corresponds to the predicted default threshold falling below the true default threshold from around $t = 3.25$ onwards. The reason is the rating agency's directional learning, which results in a predicted default threshold corresponding to the average of the potentially surviving types, see Eq. (14). For a fixed type θ with default threshold $f(\theta)$, the predicted default threshold $g(E)$ is smaller or equal than the true threshold if $g(E_t) \leq f(\theta)$. We capture this event by a stopping time $T_f(\theta)$ defined as $T_f(\theta) = \inf\{t \geq 0 : g(E_t) \leq f(\theta)\}$. From this time $T_f(\theta)$ on, the predicted default threshold is lower or equal than the true one, that is, $\hat{D}_t^\star \leq f(\theta)$, for $T_{f,\theta} \leq t < \tau(\theta)$, and $T_f(\theta) < \tau(\theta)$ almost surely. Accordingly, the rating of the firm for type θ is inflated from $T_f(\theta)$ on, that is in $[T_f(\theta), \tau(\theta)]$. Note that this inflation is persistent if the firm recovers. Not only will the firm have to pay lower coupons if it stays in a regime of low current cash flow. Also when the cash flow recovers, will the firm not only enjoy the well-deserved relief in debt payments given the higher cash flow, but also an additional relief due to the persistent inflation. We label this “ex-post rating inflation” as the persistent inflation kicks in at the stopping time $T_f(\theta)$ not known by the rating agency, since the true θ is not known until default. Consequently, the ex-post rating inflation cannot be identified by the rating agency until default. At default, the true measurement error is revealed. In contrast to reduced-form intensity models, in which default occurs by chance, in this paper default is determined by non-observable firm characteristics.

Proposition 1 characterizes the rating agency's best response to an arbitrary firm strategy $f \in \mathcal{A}_f$. The best response $g(\cdot; f) \in \mathcal{A}_g$ satisfies almost all conditions in $\mathcal{A}_g^{\mathbf{C}}$, that is, $g(\cdot; f)$ is non-decreasing and bounded by Id. However, the important constraint that $g(\cdot; f)/Id$ is non-decreasing, which prohibits inappropriate incentives resulting from the rating agency's strategy, cannot be shown to hold in general. More specifically, the rating agency may not provide incentives that prevent the firm from decreasing its cash flow to immediately profit from reduced interest payments. Instead of imposing the constraint by requiring $g(\cdot; f)/Id$ to be non-decreasing in (12), we

[12] This is further illustrated by Figures A.2 and A.3 in the Appendix.

rather push $g(\cdot; f)$ into $\mathcal{A}_g^{\mathbf{C}}$ by the following transformation[13]

$$\mathcal{R}(g)(e) = \begin{cases} e \inf\{g(z)/z : 0 < z \leq e\}\,, & \text{for } e > 0, \\ 0\,, & \text{for } e = 0. \end{cases} \tag{15}$$

The transform $\mathcal{R}$ maps $\{g \in \mathcal{A}_g : g \text{ non-decreasing}, g \leq Id\}$ to $\mathcal{A}_g^{\mathbf{C}}$ and is the identity on $\mathcal{A}_g^{\mathbf{C}}$, i.e. $\mathcal{R}(g) = g$, for $g \in \mathcal{A}_g^{\mathbf{C}}$, see Lemma 1 in the Appendix.

However, note that although the constraint that $g(\cdot; f)/Id$ is non-decreasing cannot be shown to hold in general, our numerical implementation and all cases that we could think of as practically relevant lie on $\mathcal{A}_g^{\mathbf{C}}$. That is, the transform is formally needed but often will just turn out to be the identity.

3.2 Best Response of the Firm

In the following we characterize the best response of the firm τ for an admissible rating strategy g as specified in Equation (7). Since the firm knows its own cash flow and the type θ, we can specify the firm's type-dependent best response $\tau(\theta; g)$.

For a given rating agency strategy $g \in \mathcal{A}_g^{\mathbf{C}}$ and type θ, the best response $\tau(\theta; g)$ is the solution to the optimal stopping problem given in (16) below. Recall the firm's expected payoff given in Eq. (8). Denote by $v(\cdot, \cdot; \theta, g)$ the value function, which is given by

$$v(d, e; \theta, g) = \sup_{\tau \in \mathcal{T}_{(d,e)}} \mathbb{E}_{(d,e)} \left[\int_0^\tau e^{-r\,t} \left(D_t/\theta - C(D_t/g(E_t)) \right) \mathrm{d}t \right], \text{ for } (d, e) \in \mathcal{C}, \tag{16}$$

where $\mathcal{C}\{(d, e) \in \mathbb{R}^2 : 0 \leq e \leq d\}$ is the convex cone on which the imperfectly observed cash flow D and its running minimum E take values in and $\mathcal{T}_{(d,e)}$ is the set of all stopping times with respect to the information generated by (D, E) with starting

[13] Alternatively, we can characterize the rating agency's best response in the constrained set $\mathcal{A}_g^{\mathbf{C}}$ as the solution of the related constrained optimization problem. It can be shown that this problem is of the form $\hat{v}(e, g; f) = \max_u \int_{e_0}^e h(\tilde{e}, \tilde{g}; f)\, \mathrm{d}\tilde{e}$, where $g'(e) = u(e, g) \in \hat{\mathcal{A}}(e, g; f)$ and $e_0 = \inf_{\theta \in \Theta} f(\theta)$. Starting at $t = 0$ with initial values $D_0 = E_0$ for the imperfectly observed cash flow, the rating agency solves for the optimal value function $\hat{v}(E_0, g_0; f)$ and then chooses the initial value g_0 as to maximize the latter expression. Thus the rating agency's best response depends on the initial value, i.e. $g(\cdot; E_0, f)$. In particular, $g(\cdot; E_0, f) \neq g(\cdot; E_0', f)$ on $[0, \min(E_0, E_0')]$ in general, for $E_0 \neq E_0'$. An equilibrium analysis of this situation is beyond the scope of this paper and left for future research.

value (d, e).

Optimal stopping problems are connected to free-boundary value problems, see Peskir and Shiryaev (2006). For the case that the running minimum of a diffusion process is included in the state variables, Heinricher and Stockbridge (1991) and Barron (1993) provide a characterization of the solution of optimal control problems in terms of the solution of an associated free boundary value problem. While we do have a partial differential equation system for the firm's optimal stopping problem in Equation (16), this is a non-standard system of partial differential equations, as we have to incorporate not only the cash flow fixed, but also the running minimum of the imperfectly observed cash flow the rating agency uses. The value function v, we characterize through a viscosity solution to the following free boundary problem.

Define the differential operator $\mathcal{L}^{(\theta,g)}$ by

$$\mathcal{L}^{(\theta,g)}h = \mu\, d\, \frac{\partial h}{\partial d} + \frac{1}{2}\sigma^2\, d^2\, \frac{\partial^2 h}{\partial d^2} + k^{(\theta,g)} - r\, h\,, \tag{17}$$

where $k^{(\theta,g)}(d, e) = d/\theta - C(d/g(e))$, for $(d, e) \in \mathcal{C}$. Then

$$\mathcal{L}^{(\theta,g)}v \leq 0\,, \tag{18}$$

$$v \geq 0\,, \tag{19}$$

$$v \cdot \mathcal{L}^{(\theta,g)}v = 0\,, \tag{20}$$

with boundary conditions

$$0 = \frac{\partial v}{\partial d} \text{ on } \partial C^{(\theta,g)}\,, \text{ and } 0 = \frac{\partial v}{\partial e} \text{ on } \mathcal{D}\,, \tag{21}$$

where $C^{(\theta,g)} = \{(d, e) \in \mathcal{C} : v(d, e; \theta, g) > 0\}$ is the continuation region and $\mathcal{D} = \{(d, d) \in \mathcal{C} : d > 0\}$ is the diagonal. The first boundary condition is due to smooth fit at the edge of the continuation region $\partial C^{(\theta,g)}$ and the second condition is normal reflection on the diagonal $\mathcal{D}$.

The best response of the firm to a given rating agency rating strategy g is the collection of optimal stopping times $(\tau(\theta; g))_{\theta\in\Theta}$. Under the condition $g \in \mathcal{A}_g^{\mathbf{C}}$ and g is non-decreasing as well as bounded by Id, we show that the optimal stopping rule

$\tau(\theta; g)$ is a cut-off rule. Specifically, the firm liquidates at the first hitting time of a threshold $f(\theta; g)$, for all $\theta \in \Theta$, as is shown in the following proposition. The cut-off rule balances the firm's trade-off between continuing in unfavorable conditions now in order to reap lower interest payments once the rating agency updates the imperfectly observed cash flow. If the imperfectly observed cash flow is above the threshold, the prospect of continuing is attractive. Once it hits the cut-off, the firm liquidates.

Proposition 2 (Firm's Best Response)**.** *For $g \in \mathcal{A}_g^{\mathbf{C}}$, $\theta \in \Theta$, and $d > 0$, the optimal stopping time of* (16) *is given by*

$$\tau_{(d,d)}(\theta; g) = \inf\{t \geq 0 : D(t) \leq f(\theta; g)\}\,, \tag{22}$$

where $f(\theta; g)$ is some positive real constant, i.e. $\tau_{(d,d)}(\theta; g)$ is the first hitting time of the imperfectly observed cash flow D with respect to the barrier $f(\theta; g)$.

This result implies that the best response of the firm $\tau(\theta; g)$ for a rating agency's strategy g is characterized by a specific default barrier $f(\theta; g)$, for each $\theta \in \Theta$.[14] For the path of cash flow, the firm is aware of its over- or underestimation by the rating agency based on the observed cash flow. If the imperfectly observed cash flow decreases beyond a type-dependent level, the firm defaults.

For our specific default barrier, we can provide more structure on how the barrier changes in the type, see Lemma 2 in the Appendix. In particular: (i) The best response barrier $f(\cdot; g)$ is non-decreasing. It is based on the observed cash flow, as in Equation (16), such that the measurement error θ only enters as the scaling factor of the observed cash flow D. Hence, a higher measurement error θ implies a lower true cash flow X (note that $X = D/\theta$), and subsequently a lower default threshold.

(ii) The default barrier in terms of the true cash flow $f(\cdot; g)/Id$ is non-increasing and uniformly bounded in $g \in \mathcal{A}_g^{\mathbf{C}}$.[15] A firm subject to an overestimated type has a lower true default barrier, as it largely underpays its interest. (iii) The best response barrier $f(\cdot; g)$ is uniformly Lipschitz continuous in $g \in \mathcal{A}_g^{\mathbf{C}}$. (iv) Assumption 2 yields that $f(\cdot; g)$ is strictly increasing with an uniformly lower bound in $g \in \mathcal{A}_g^{\mathbf{C}}$. The two

[14]Solving the associated free boundary value problem determines the point-wise solution.

[15]Default happens if $E_t \leq f(\theta; g)$, or, equivalently $Y_t = E_t/\theta \leq f(\theta; g)/\theta$.

latter properties are crucial for the subsequent equilibrium analysis as it clears the way for applying the Schauder fixed point theorem.

4 Rating Equilibrium

The results of the previous section are fundamental for the subsequent equilibrium analysis. Both the rating agency and the firm use Markov strategies, which are both characterized by real-valued functions, see Proposition 1 and Proposition 2. We obtain a perfect Bayesian equilibrium in Markov strategies as introduced by Maskin and Tirole (1988) in spirit of Grenadier, Malenko, and Malenko (2016) by the Schauder fixed point theorem. However, the application of this fixed point theorem requires that the interest payment is well behaved. Specifically, the existence of the rating game requires a growth constraint on the rating-dependent interest payments, see Assumption 2. Given this assumption, we arrive in the present section at our equilibrium on the rating market. First, Proposition 3 below provides the existence of an equilibrium candidate. Then, Proposition 4 characterizes the solution and verifies the existence and uniqueness, provided specific technical conditions hold.

In particular, the latter non-trivially extends the result of Manso (2013) giving the existence and uniqueness of an equilibrium for the cash flow that follows a geometric Brownian motion to the case of information asymmetry. The underlying reason for the uniqueness is the non-stationarity assumption for the cash flow process. Multiple equilibria would be likely for a mean-reverting specification, see Manso (2013). Furthermore, notice that $f(\theta)$ is strictly increasing according to Lemma 2, that is, a best response has a minimal slope of $l_f > 0$ for all given strategies g of the rating agency. This eliminates classical semi-pooling equilibria, in which some observed types default at the same time, i.e. play the same strategy. While we do not have a semi-pooling equilibrium, the rating agency's learning implies that the information is only revealed fully at the time the firm actually defaults, prior to which the rating agency can only rule out some types, but cannot fully infer the firm's observed type. The rating agency learns the exact type at default, but the game ends simultaneously, leaving the rating agency without the possibility to react.

To establish our equilibrium, we apply the Schauder fixed point theorem to the mapping $T : (f, g) \mapsto (f(\cdot; g), \mathcal{R}(g(\cdot; f)))$, where $f(\cdot; g)$ is the firm's best response given in Proposition 2, $g(\cdot; f)$ is the rating agency's best response given in Proposition 1, and $\mathcal{R}$ is the respective transformation defined in (15). To ensure that the best responses are well-defined, we need to restrict the set of admissible strategies to

$$\mathcal{A}_f^\star = \{f \in \mathbf{C}(\Theta, \mathbb{R}_0^+) : f \text{ strictly increasing}\}\,, \text{ and } \mathcal{A}_g^\star = \mathcal{A}_g^{\mathbf{C}}\,. \tag{23}$$

Proposition 3. *Suppose that Assumption 2 holds, then the mapping* $T : \mathcal{A}_f^\star \times \mathcal{A}_g^\star \to \mathcal{A}_f^\star \times \mathcal{A}_g^\star, (f, g) \mapsto (f(\cdot; g), \mathcal{R}(g(\cdot; f)))$ *has at least one fixed point. Let* $(f^\star, g^\star)$ *be such a fixed point, if* $\mathcal{R} \circ g(\cdot; f^\star) = g(\cdot; f^\star)$ *then* $(f^\star, g^\star)$ *is an equilibrium.*

Proposition 3 provides us with a candidate for an equilibrium in this very general setup with function-valued strategies, which then has to be verified to be an equilibrium. Practically, for the subsequent analysis of special cases of the rating game's equilibrium, the transform $\mathcal{R}$ regulating the rating agency's strategy g always takes the form of the identity function, so that no transformation is necessary. In turn, without employing the transformation, we immediately end up with an equilibrium for the rating game without further ado for all relevant cases.

This result is a pure existence result and no particular guidance is given on how to actually compute such an equilibrium candidate. Now, we provide with Proposition 4 below a result that characterizes an equilibrium candidate as the solution to a two-dimensional ordinary differential equation (ODE) under some technical assumption, see Assumption 3 below. Further, an inequality condition is stated, and given this condition holds, the candidate is indeed an equilibrium, which is then also unique. This result is the basis for computing an equilibrium strategy in a fast and efficient way, which we rely on heavily in the following analysis.

For $g \in \mathcal{A}_g^\star$ and $\theta \in \Theta$ consider the value function $v(\cdot, \cdot; \theta, g)$. Its boundary $\partial C^{(\theta,g)}$ can be described by a function $b(\cdot, \theta; g)$ that assigns to each minimum observed cash flow e the critical d-value where the firm defaults, i.e.

$$b(e, \theta; g) = \inf\{d \geq e : v(d, e; \theta, g) > 0\}\,, \text{ for } e \in [0, f(\theta; g)]\,. \tag{24}$$

See Lemma 3 in Appendix C for the properties of b. In order to facilitate the subsequent analysis, we make the following assumptions.

Assumption 3. *For $g \in \mathcal{A}_g^\star$, assume that the collection of solutions $v(\cdot,\cdot;\cdot,g)$ to the boundary value problem (17-21) indexed by θ, $\theta \in \Theta$, is continuously differentiable in θ as well as allows for interchanging the order of differentiation with respect to d and θ on the interior of $\bigcup_{\theta\in\Theta} C^{(\theta,g)} \times \{\theta\}$, and that the collection of boundary functions $b(\cdot,\cdot;g)$ is continuously differentiable with respect to e and θ.*

Under the given assumption, the following proposition characterizes an equilibrium candidate as the solution to an implicit two-dimensional ODE and gives a condition for the candidate being indeed an equilibrium, which then turns out to be unique. This ODE is the centerpiece to the numerical computation of the equilibrium. Its proof is given in Appendix C.

Proposition 4. *Given the setting of Proposition 3, denote by $(f^\star, g^\star)$ a fixed point of T. Suppose that Assumption 3 holds for $g^\star$, that $f^\star$, $g^\star$ and ϕ are continuously differentiable, as well as that the technical condition in* (50) *of Corollary 2 holds. Denote by $(f,\hat{g})$ the solution of the implicit two-dimensional ODE*

$$\begin{pmatrix} f'(\theta) \\ \hat{g}'(\theta) \end{pmatrix} = \begin{pmatrix} \dfrac{(1+\eta)\,\sigma^2}{2\,(r-\mu)} \, \dfrac{f(\theta)^2/\theta^2}{C(f(\theta)/\hat{g}(\theta)) - f(\theta)/\theta} \, \dfrac{1}{1 - h(f(\theta),\hat{g}(\theta),f'(\theta),\hat{g}'(\theta),\theta)} \\ \dfrac{\phi(\theta)}{\Phi(\theta)}\,(f(\theta) - \hat{g}(\theta)) \end{pmatrix},$$

on $(\underline{\theta},\overline{\theta})$ with initial condition

$$\begin{pmatrix} f(\underline{\theta}) \\ \hat{g}(\underline{\theta}) \end{pmatrix} = \underline{\theta} \begin{pmatrix} f_1^\star \\ f_1^\star \end{pmatrix},$$

where $\eta = \frac{\mu - \frac{1}{2}\sigma^2}{\sigma^2} + \sqrt{\left(\frac{\mu-\frac{1}{2}\sigma^2}{\sigma^2}\right)^2 + \frac{2\,r}{\sigma^2}} > 0$, $h(f(\theta),\hat{g}(\theta),f'(\theta),\hat{g}'(\theta),\theta) = \frac{\partial b}{\partial e}(f(\theta),\theta;\hat{g}\circ f^{-1})$, and $(f_1^\star, g_1^\star)$ denotes the equilibrium of the perfect information case, i.e. $\Theta_1 = \{1\}$ and hence $D = X$, with $g_1^\star = f_1^\star$, which exists and is unique under the given *assumptions, as well as* $\Phi(\theta) = \int_{\underline{\theta}}^{\theta} \phi(t)\,\mathrm{d}t$. *If*

$$\hat{g}' \leq f' \, \frac{\hat{g}}{f}\,, \ \ \textit{on}\ (\underline{\theta}, \overline{\theta})\,,$$

then $(f^\star, g^\star)$ *is the unique equilibrium satisfying the aforementioned assumptions and is given by* $(f^\star, g^\star) = (f, \hat{g} \circ f^{-1})$ *on* $\Theta \times f(\Theta)$.

5 Analysis of Rating Equilibrium

In this section, we turn to the analysis of the equilibrium derived in the previous sections. In particular, we study numerically the variation in values which the imperfectly observed cash flow introduces. To this end, we compare a range of firms with different distances to default. The heterogeneity is caused by different true cash flows. Thus, for the same imperfectly observed initial cash flow, they have a different deviation of the actual firm value from the expected firm value that the rating agency perceives.

The firm has debt outstanding with face value scaled to unity such that C then denotes the interest rate in percentage terms payable by the firm to the debt holders depending on the rating R. The firm generates cash flow per unit debt at the rate X. We assume that an equilibrium exists in our setup, see Proposition 4 for sufficient conditions.

5.1 Parametrization based on Market Data

We consider a firm's cash flow is characterized by an expected growth rate of $\mu = 0$ and a volatility of $\sigma = 0.30$, and combine it with a risk-free rate r set to 0.0211, which is the average 3-month T-Bill rate (DTB3) over the time period 01 January 1997 through 31 December 2016 provided by Federal Reserve Economic Data. For this specification, the interest payment rate determining function C is calculated under the assumption of perfect information, i.e. $D = X$, or $\tilde{\theta} = 1$ and $\Theta_1 = \{1\}$, from available market data. For seven rating classes AAA, AA, A, BBB, BB, B and C/CCC, the

Table 1: *Rating-Dependent Interest Rate*

Rating Class	10-Year Default Probability	Implied Rating R	Interest Rate C
AAA	0.0086	18.2180	0.0439
AA	0.0109	16.8281	0.0447
A	0.0195	13.7125	0.0494
BBB	0.0464	9.7811	0.0573
BB	0.1527	5.5756	0.0732
B	0.2746	3.9336	0.0916
C/CCC	0.5584	2.2507	0.1513

Notes: *The 10-year default probabilities are obtained from Table 25 in Standard and Poor's (2016) and are referring to rated US companies in the period 1981 to 2015. The implied rating is determined using* (27). *The interest rates are the average of effective corporate yields provided by Federal Reserve Economic Data.*

interest rate C_i is determined by the average effective yield for each rating class, i.e., averaging over the time period 01 January 1997 through 31 December 2016.[16] [17] For each rating class i, the distance-to-default type rating R_i is extracted from the 10-year probability of default PD_i, which is collected from Table 25 of Standard and Poor's (2016) based on U.S. data in the period 1981 through 2015.[18]

5.2 Type Specification and Equilibrium Computation

Before we study the impact of learning and uncertainty about the firm's observed type, we consider the perfect information case as benchmark, i.e. $\Theta_1 = \{1\}$. In this case, the equilibrium $(f_1^\star, g_1^\star)$ is given by $f_1^\star = g_1^\star = 0.0490$. Comparing this number to Table 1, we see that in the base case a company being rated B or better is having a cash flow X of $f_1^\star R_B = 0.0490 \times 3.9336 = 0.1928$ or greater. This cash flow exceeds the interest payment rate C, which amounts to 0.0916 for a B rated firm or is smaller for a firm with a better rating. However, for the C/CCC rated firm, the corresponding

[16]The respective Federal Reserve Economic Data identifiers for yield data are BAMLC0A1CAAAEY, BAMLC0A2CAAEY, BAMLC0A3CAEY, BAMLC0A4CBBBEY, BAMLH0A1HYBBEY, BAMLH0A2HYBEY, and BAMLH0A3HYCEY. Note that the data for calibration stems predominantly from non-PSD instruments with a higher default risk than comparable PSD instruments. However, the higher credit risk of non-PSD instruments is perhaps accompanied by a higher interst rate rewarding for the additional risk. Thus, the effect of using non-PSD instruments for calibration does not seem to be critical.

[17]This case is an equilibrium, see Proposition 4.

[18]For the exact specification, refer to AppendixA.1

cash flow amounts to $f_1^\star R_{C/CCC} = 0.0490 \times 2.2507 = 0.1103$, which is not sufficient to cover the interest of 0.1513. In this case, the equity holders keep the company alive by injecting additional funds. If the cash flow drops further and reaches the default threshold $f_1^\star = 0.0490$, the firm defaults on its debt and creditors receive the assets.

The rating agency faces a firm whose cash flow it observes imperfectly following the type $\tilde{\theta}$ which is distributed according to a truncated normal distribution with parameters μ_θ, σ_θ, and truncation to $\Theta = [\underline{\theta}, \overline{\theta}]$. Accordingly, the density of $\tilde{\theta}$ is given by

$$\phi(\theta) = \begin{cases} 0, & \text{for } \theta < \underline{\theta}, \\ \frac{1}{\sqrt{2\,\pi}\,\sigma_\theta\,c_\theta} \exp\left(-\frac{1}{2}\frac{(\theta-\mu_\theta)^2}{\sigma_\theta^2}\right), & \text{for } \underline{\theta} \leq \theta \leq \overline{\theta}, \\ 0, & \text{for } \theta > \overline{\theta}. \end{cases} \tag{25}$$

where $c_\theta = N((\overline{\theta} - \mu_\theta)/\sigma_\theta) - N((\underline{\theta} - \mu_\theta)/\sigma_\theta)$ and N is the standard normal cumulative distribution function. The base case parameters are $\mu_\theta = 1$ (unbiased), $\sigma_\theta = 0.25$ and $\Theta = [\underline{\theta}, \overline{\theta}] = [0.50, 1.50]$.

Solving the two-dimensional ODE of Proposition 4 with initial condition

$$(f(\underline{g}), \hat{g}(\underline{g})) = (\underline{g}\, f_1^\star, \underline{g}\, f_1^\star) = (0.5 \times 0.0490, 0.5 \times 0.0490) = (0.0245, 0.0245)$$

gives the equilibrium strategy $(f^\star, g^\star) = (f, \hat{g} \circ f^{-1})$.

5.3 Firm Strategies and Learning in Equilibrium

This section illustrates how both the firm and rating agency specify their strategies in equilibrium and how the rating agency learns by inferring the type from the firm's decision to survive or default given the observed cash flow trajectory. Taking the firm strategy $f^\star$ and the rating agency's strategy $g^\star$ from the previous section, Figure 4 displays the equilibrium strategies of the firm and rating agency. Panel 4a displays the equilibrium strategy of the rating agency $g^\star$ (solid) and its continuation (dashed) depending on the smallest imperfectly observed cash flow e. Panel 4b displays the equilibrium strategy of the firm $f^\star$ in terms of the default threshold depending on

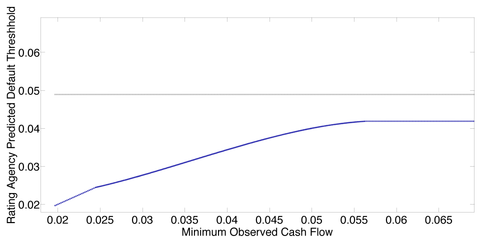
Table 2: *Equity Values*

	Perfect Information Case		Imperfect Information Equity Value		
Rating	Cash Flow	Equity Value	Expected Value	Minimum	Maximum
AAA	0.8927	38.8615	39.2513	38.4905	39.6371
AA	0.8246	35.6461	36.0189	35.2411	36.4153
A	0.6719	28.4281	28.7643	27.9459	29.1885
BBB	0.4793	19.2942	19.5860	18.7128	20.0535
BB	0.2732	9.4892	9.7350	8.8125	10.2504
B	0.1927	5.6681	5.8969	5.0026	6.4246
C/CCC	0.1103	1.8827	2.0814	1.4380	2.5644
D	0.0490	0.0000	0.0429	0.0000	0.1544

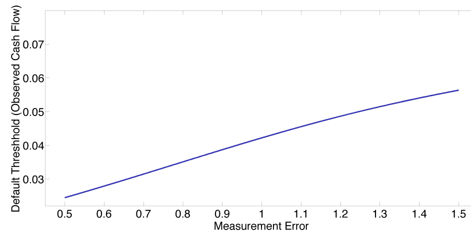
Notes: *Perfect information cash flow (column 2) for base case parameters computed from Table 1 and* $X_i = R_i f_1^\star$ *where* $f_1^\star = 0.0490$ *is the default threshold, for all rating classes* i *including default (D) with* $R = 1$. *The firm value (column 3) is the perfect information value of equity* w, *which is computed numerically based on equations (17–21), for* $\theta = 1$. *The same equation system is solved for base parameters and various* $\theta \in [\underline{\theta}, \overline{\theta}] = [0.5, 1.5]$ *and the expected value (column 4) is calculated, as well as minimum (column 5) and maximum (column 6) are given.*

the type θ. Panel 4c displays the equilibrium strategy of the firm $f^\star$ in terms of the default threshold in the firm's true cash flow depending on the type (solid line) and the equilibrium default threshold in the perfect information case (dashed).

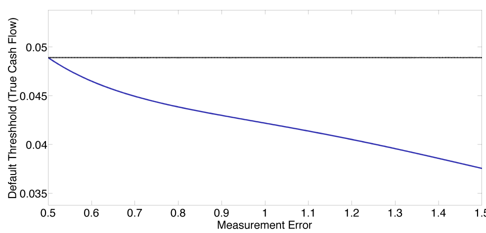
Panel 4a displays how the rating agency learns in equilibrium in terms of the predicted default threshold $g^\star$ depending on the minimum observed cash flow e. It draws the agency's estimated firm default threshold as a function of the minimum observed cash flow from the imperfectly observed cash flow trajectory. For a minimum observed cash flow greater than 0.0563, the firm does not default regardless of the realization of the type θ. The predicted default threshold for the observed cash flow is then given as 0.0419, which is somewhat smaller than the default threshold of 0.0490 in case of perfect information. The dashed line on the right part of Panel 4a indicates the area for which the rating agency cannot learn yet because the firm has not been in enough trouble for even the most overestimated type to default. Once the minimum observed cash flow falls below 0.0563 and the firm has not defaulted yet, the rating agency learns about the true type θ and the belief is updated as indicated in (13). Based on the updated belief, the predicted default threshold is adjusted according to $g^\star(e) = \mathbb{E}[f^\star(\tilde{\theta}) | f^\star(\tilde{\theta}) < e]$, see (14), which the solid line in the center



(a) *Equilibrium Strategy of Rating Agency*



(b) *Equilibrium Strategy of Firm (Observed Cash Flow)*



(c) *Equilibrium Strategy of Firm (True Cash Flow)*

Figure 4: *Panel a displays the equilibrium strategy of the rating agency $g^\star$ (solid) and its continuation (dashed) depending on the smallest imperfectly observed cash flow e. Also, the default threshold under perfect information is included. Panel b displays the equilibrium strategy of the firm $f^\star$ in terms of the default threshold depending on the type θ. Panel c displays the equilibrium strategy of the firm $f^\star$ in terms of the default threshold in the firm's true cash flow depending on the type (solid line) and the equilibrium default threshold in the perfect information case (dashed). The graphs are based on the base case parameters.*

of Panel 4a indicates. With decreasing minimum observed cash flow e, the rating agency's predicted default threshold $g^\star$ decreases to 0.0245. At that cash flow level, the firm has defaulted regardless of the realization of the type θ, and thus $g^\star(e) \searrow 0.0245$, as $e \searrow 0.0245$. For the case that the minimum cash flow is below 0.0245, the rating agency thus predicts immediate default and thus $g^\star(e) = e$, for $0 < e \leq 0.0245$. The dashed line on the left part of Panel 4a displays the minimal observed cash flow region for which the rating agency predicts immediate default.

Consider the firm's strategy in Panel 4b. For the most severe overestimation $\theta = \overline{\theta} = 1.5$, the firm cash flow is overstated by 50%, hence its rating is biased upwards and it pays a lower interest rate than it should pay. When the observed cash flow D hits $f^\star(1.5) = 0.0563$, the firm defaults. The true cash flow X at default amounts to $f^\star(1.5)/1.5 = 0.0376$, which is smaller than the perfect information default threshold of $f_1^\star = 0.0490$. For a correct estimate, i.e. $\theta = 1$, the observed cash flow D coincides with the true cash flow X, i.e. $D = X$. The firm defaults when the cash flow drops to $f^\star(1) = 0.0422$. The latter is also smaller than the perfect information default threshold of $f_1^\star = 0.0490$. This seemingly surprising result makes sense since the rating agency initially has an unbiased estimate of the firm's cash flow, i.e. $\mathbb{E}[\tilde{\theta}] = 1$, according to the rating agency's belief π as long as the observed cash flow D is above the default threshold for the most severe overestimation, i.e., $f^\star(\overline{\theta}) = f^\star(1.5) = 0.0563$. The rating agency updates its belief π according to (13), and then $\mathbb{E}[\tilde{\theta}|f^\star(\tilde{\theta}) < e] < 1$, for $e < f^\star(\overline{\theta}) = 0.0563$. Therefore, the rating agency then has an upward bias on the cash flow estimate and a surviving firm pays a lower interest rate. In turn, this leads to a lower default threshold than in the perfect information case. For the most severe underestimation $\theta = \underline{\theta} = 0.5$, the firm's cash flow is understated by 50%, hence its rating is biased downwards and it pays a higher interest rate than it should. The rating agency learns about the true cash flow once the observed cash flow starts falling below $f^\star(1.5) = 0.0563$ and updates the belief to some π, but $\mathbb{E}[\tilde{\theta}|f^\star(\tilde{\theta}) < e] \geq 0.5$, for $e \geq f^\star(0.5) = 0.0245$. The true cash flow at default amounts to $f^\star(\underline{\theta})/\underline{\theta} = f^\star(0.5)/0.5 = 0.0490$ and is identical to the perfect information default threshold of $f_1^\star = 0.0490$. This implies that the firm pays a higher interest rate than it should, but as the cash flow tends to the default threshold, the rating becomes accurate and hence the default threshold coincides with the perfect information case.

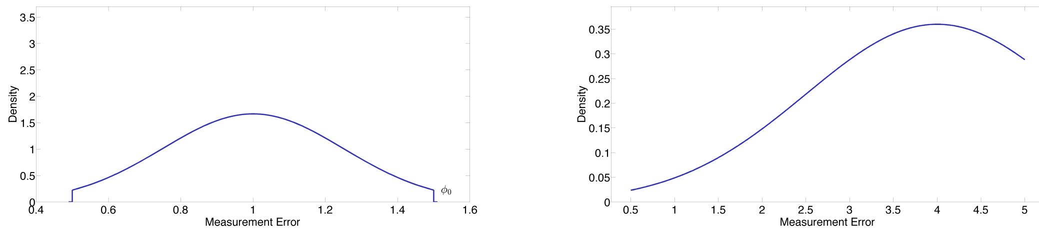
Panel 4c shows the functional form of the firm's default threshold in the terms of the firm's true cash flow depending on θ. The default threshold in this perspective is always smaller or equal than in the perfect information case. For $\theta = \underline{\theta} = 0.5$, the default threshold based on the firm's true cash flow coincides with the perfect information case, i.e. $f^\star(\underline{\theta})/\underline{\theta} = f_1^\star = 0.0490$. Opposed to the firm's cut-off strategy expressed in the imperfectly observed cash flow by the rating agency, the firm's default threshold in terms of its true cash flow is decreasing in θ with a minimum value $f^\star(\overline{\theta})/\overline{\theta} = f^\star(1.5)/1.5 = 0.0376$. Therefore, the inclusion of imperfect observation of the cash flow leads to a lower default threshold of the firm and hence to default occurring later, compared to the perfect information case. Firms defer the liquidation decision to profit from lower interest payments for all types, as, once the rating agency starts to learn by observing the firm's liquidation decision, the updating of the rating agency's belief results in an upwards biased rating. The firm then faces the trade off of deferring liquidation to profit from lower interest payments against the potential injections of further capital. The higher the rating agency overestimates the firm, the more profitable deferring becomes, leading to the downward slope in Panel 4c.

However, having a lower default threshold alone is not necessary profitable for the firm: Although it defers liquidation, the deferral is acquired at the cost of higher interest payments for the time prior to liquidation if the rating agency infers a low cash flow. Again, this effect is largest for the most underrated type: As the rating agency heavily underestimates the firm's cash flow, its assessment leads to higher interest payments lasting also the longest, as the rating agency needs the most time to infer the true cash flow. This effect is dampened because of the asymmetry of the learning. While the rating agency learns over time, it can only rule out the most overrated types by observing the absence of default. As it adjusts the distribution of types, the mean type is decreasing over time, leading to an overestimation of any type shortly before the firm's default.

5.4 Value of Uncertainty

This section studies the value of uncertainty that the rating bias creates after the rating agency starts learning by observing the firm's liquidation decisions. In particular, firms with a low true initial cash flow profit from the uncertainty, which substantially increases their firm value relative to the perfect information case if they are overestimated. Firms with a healthy cash flow level still display this effect but do not experience an increase of their firm value if overrated as much, because they set off with a strong rating to begin with anyway. Furthermore, this section sheds light on the value of uncertainty for different observed types of firms: A stable and mature firm will display less uncertainty about the soft information than a start up firm, which poses a much greater challenge to the rating agency. Figure 5 shows the two different cases.

First, we turn to the base case of a mature firm, which we model by a measurement error distribution with $\mu_\theta = 1$ and $\sigma_\theta = 0.25$ on $\Theta = [\underline{\theta}, \overline{\theta}] = [0.50, 1.50]$. Panel 5a displays the corresponding density, which is symmetric. Economically speaking, the rating agency may over- or underestimate the soft information, but large surprises are unlikely, that is, the distribution is not skewed. Figure 6 displays the value function in terms of the firm's true cash flow depending on the realization of the



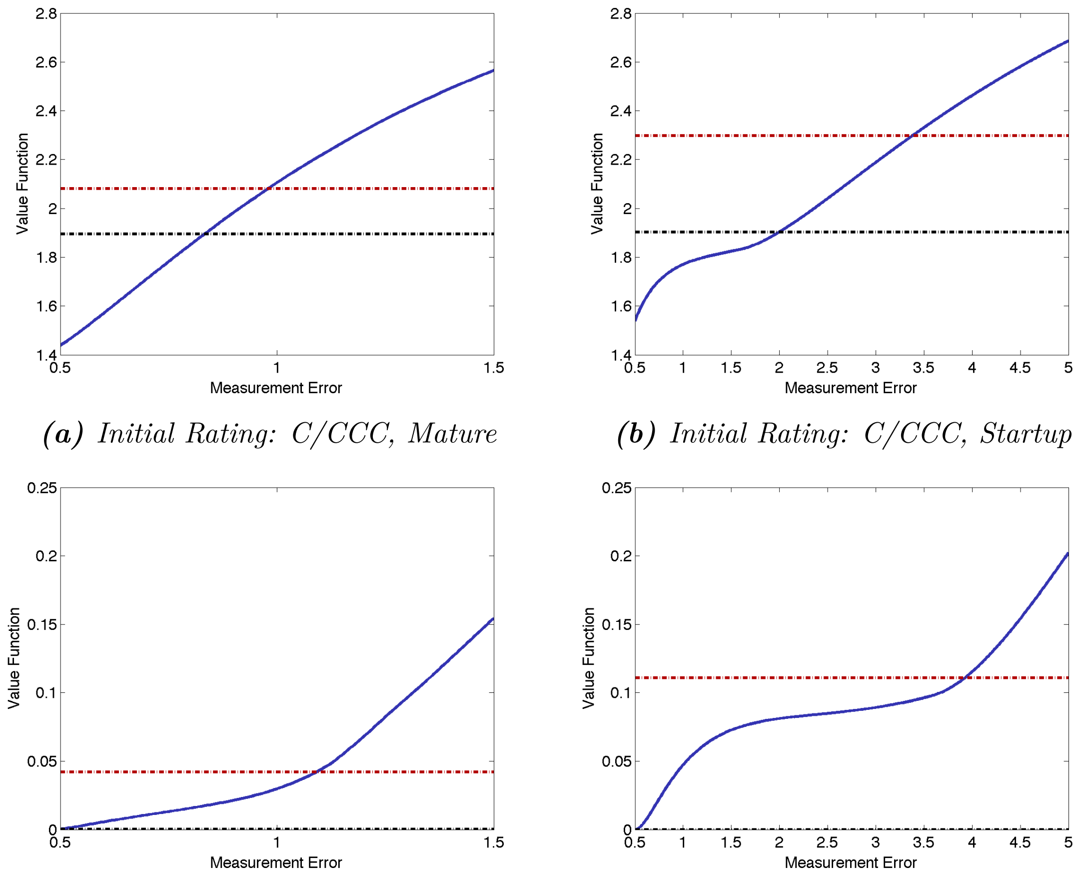
(a) *Mature firm.* ***(b)*** *Startup firm.*

Figure 5: Densities for different firms. *This graph displays the density of measurement error distribution for the soft information for mature firm (Panel a) and a startup firm (Panel b). Specifically, the mature firm has a measurement error distribution ranging from* $\Theta = [\underline{\theta}, \overline{\theta}] = [0.50, 1.50]$*, with* $\mu_\theta = 1.0$ *and* $\sigma_\theta = 0.25$*, while the startup firm has a wider range of values, i.e.* $\Theta = [\underline{\theta}, \overline{\theta}] = [0.50, 5.0]$*, with* $\mu_\theta = 4.5$ *and* $\sigma_\theta = 0.5$*.*

type θ, for an initial cash flow corresponding to a specific rating in the perfect information case, see Table 1. For a specific rating R_i (column 3, Table 1), the initial cash flow X_i is computed by $X_i = R_i\, f_1^\star$, where $f_1^\star$ is the default threshold in the perfect information case, e.g., $f_1^\star = 0.0490$ for the base case parameters. Note that also the rating class default with letter D is included, which corresponds to a rating of $R_D = 1$. Table 2 summarizes the key facts of Figure 6. In particular, it contains the rating (column 1), the initial cash flow X_i (column 2), and the equity value v (column 3), which is computed numerically based on equations (17–21), for $\theta = 1$, all for the perfect information case, $\Theta_1 = \{1\}$ and base case parameters. For varying θ, the expected value $\mathbb{E}[v(X_i, X_i; \tilde{\theta}, g^\star)]$ (column 4) is computed, as well as the minimum $\min_{\theta\in\Theta} v(X_i, X_i; \theta, g^\star)$ (column 5) and the maximum $\max_{\theta\in\Theta} v(X_i, X_i; \theta, g^\star)$ (column 6) are given.

In general, we observe that the expected equity value increases when comparing the firm value under perfect information (column 3) with the expected firm value under imperfect information (column 4). For all rating classes other than the default class D, the firm may benefit or be disadvantaged compared to the perfect information case, since the minimum value (column 5) is smaller and the maximum value (column 6) is greater than the perfect information firm value (column 3), respectively.

Now consider a start up firm, with a measurement error distribution on $\Theta = [\underline{\theta}, \overline{\theta}] = [0.50, 5.0]$, which Panel 5b shows. This distribution is strongly skewed with a mean of 4.5, so that the rating agency with a large probability overestimates the value of the firm, but potentially it heavily underestimates the firms value. Recall from



(a) Initial Rating: C/CCC, Mature *(b) Initial Rating: C/CCC, Startup*

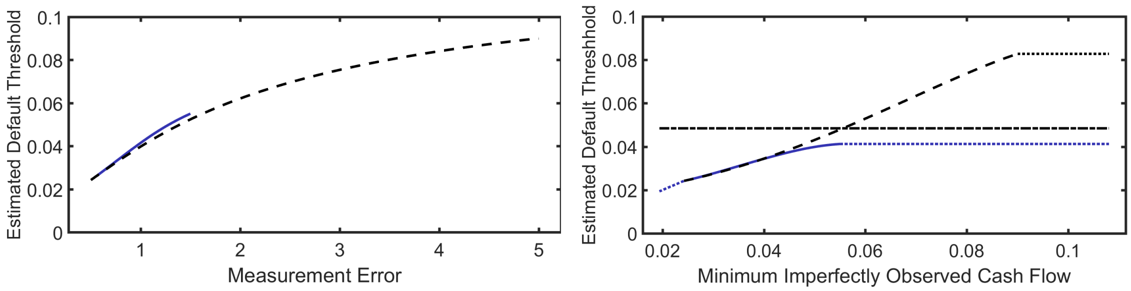
(c) Firm at default, Mature. *(d) Firm at default, Startup.*

Figure 6: *The graphs display value functions w for various initial values depending on the type θ for the base case parameters, see also Table 2. The initial cash flows imply a rating of AAA (a), C/CCC (b), and at the point of default (c). The value function w is computed numerically based on equations (17–21).*

Section 5.2 that the type $\tilde{\theta}$ is distributed according to a truncated normal distribution with parameters $\mu_\theta = 4.5$ and $\sigma_\theta = 0.5$.

Consequently, Figure 7 shows how the firm's and the rating agency's strategies from the base case (recall Figure 4) are affected by the change of the measurement error distribution. In both cases, the solid line displays the strategy for the mature firm, while the dashed line represents the strategy for the start up firm.

If we consider the mature firm, the type distribution is narrower, and we note that the rating agency has to observe a lower minimum cash flow, until it adjusts the default threshold prediction downwards. On the other hand, for the start up firm the



(a) *Firm strategy.* ***(b)*** *Rating Agency's strategy.*

Figure 7: *The graphs show how the firm's (Panel a) and rating agency's strategy (Panel b) from Figure 4 are affected by changes in the measurement error distribution implied by different economic scenarios. Specifically, it shows strategies for a mature firm with a parameter range in* $\Theta = [\underline{\theta}, \overline{\theta}] = [0.50, 1.50]$ *with* $\mu_\theta = 1.0$ *and* $\sigma_\theta = 0.25$ *(blue, solid line), and a startup firm with a wider parameter range of* $\Theta = [\underline{\theta}, \overline{\theta}] = [0.50, 5.0]$ $\mu_\theta = 4.5$ *and* $\sigma_\theta = 0.5$ *(black, dashed line). For both strategies in Panel b, the dotted lines represent areas in which the rating agency cannot learn. The dashed-dotted line represents the default threshold under perfect information.*

rating agency updates the predicted default threshold much sooner. As in the majority of cases it faces a grossly overrated firm, these firms fail already at early stages and for relatively small drops in observed cash flows. There are, however, a small number of scenarios in which the rating agency does not observe failure for a beginning crisis. In this cases there is a substantial uncertainty about the measurement error of the soft information, because some gems are included in the distribution, leading to a large set of observed cash flows for which the rating agency learns.

We also observe different strategies for start up firms compared to mature companies. Because the rating agency has much more difficulties in assessing the measurement error caused by the soft information, the firm faces the tradeoff between liquidating and facing adverse conditions if staying in the game on a higher scale. Hence, if rated adversely, firms may liquidate much sooner.

5.5 Alternative Incentives: Generalized Reputation Costs

So far, the cost rate k^π in Eq. (10) has been defined such that reputation losses are symmetric for an over- or underestimation of the true default threshold. In reality, the rating agency could be more concerned about making a mistake in either direction. On the one hand, consider the case that its objective function is not only driven by the goal of building up a reputation for precise ratings, but also its revenue originates from the rated entities as in the issuer-pays model. Then, the rating agency will be particularly concerned about avoiding an overestimation of the true default threshold ($\hat{D}_t^\star > f(\theta)$). Such an overestimation leads to a higher cost of debt, thus hurts the issuers and, if they have a choice (rating shopping) among several rating agencies applying different degrees of overestimation, potentially lead to lower fee income for the more overestimating agency. On the other hand, regulatory authorities will typically be particularly concerned about avoiding an underestimation of the true default threshold ($\hat{D}_t^\star < f(\theta)$). An underestimation leads to more defaults happening without the appropriate warning signs and can thus be dangerous especially in economic downturns. Thus, regulators might want to incentivize rating agencies rather to over- than underestimate the true default threshold. The investors' preferences regarding over- or underestimation are ambiguous. If debt is fairly priced and the investors can obtain adequate compensation, then they are indifferent regarding the rating agency's assessment. In our framework, we assume that investors and rating agency have the same information level. Thus, the debt investors can just ask for the appropriate cost of debt, regardless of which scheme the rating agency applies. If debt investors have regulatory disadvantages of holding low-rated debt, they might even prefer an underestimation of the true default threshold and as such collude with the issuers, see Opp, Opp, and Harris (2013). However, in our way of modeling, the cost of debt is determined directly from the rating. Then, debt investors will prefer an overestimation of the true default threshold, as it yields higher cash flows to debt.

To allow for such alternative incentives for the rating agency, we generalize the cost rate k^π to $k^\pi(\alpha)$ defined by

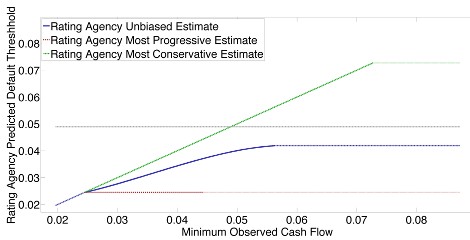
$$k_t^\pi(\alpha) = \int_\Theta Q(\hat{D}_t^\star, f(\theta); \alpha)\, \phi_t^\pi(\theta)\, \mathrm{d}\theta\,, \text{ for } t \geq 0\,, \tag{26}$$

with $Q : \mathbb{R} \times \mathbb{R} \times [0,1]$ given by $Q(d,f;\alpha) = 2\,(1-\alpha)\,\mathbf{1}_{d \leq f}(d-f)^2 + 2\,\alpha\,\mathbf{1}_{d>f}(d-f)^2\,.$ For $\alpha = 0.5$, Eq. (26) reduces to Eq. (10), i.e., the case of symmetric reputation losses. For $\alpha = 0$, reputation costs are only driven by underestimation of the default threshold ($\hat{D}_t^\star \leq f(\theta)$). In that case, the rating agency can avoid any costs by setting the rating $D_t^\star$ such that $D_t^\star > f(\theta)$ for all θ with non-zero belief, i.e. for all θ with $f(\theta) < E_t$. The rating agency's best response then changes from (14) in Proposition 2 to $g_0(e; f) = \min\left(e, \sup_{\theta\in\Theta} f(\theta)\right)\,,$ for $e \geq 0\,.$ The interpretation of the case $\alpha = 0$ is that the rating agency plays safe by overestimating the default threshold and therefore the default risk by assuming the worst case. This most conservative rating leads to higher interest payments for the firm and hence to earlier defaults accelerating the rating agency's learning. A further consequence of the most conservative rating approach is that the firm's credit risk at default is assessed accurately. Accordingly, firms default at the perfect information default threshold regardless of the observed type. While firms are still in business, they make higher interest payments than under perfect information, where the interest differential is increasing for decreasing type. Accordingly, the firm has no benefit from the information asymmetry. In the worst case specification, debt holders benefit at the expense of equity holders.

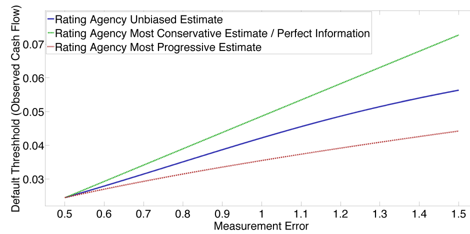
For $\alpha = 1$, reputation costs are only driven by overestimation of the default threshold ($\hat{D}_t^\star > f(\theta)$). The rating agency can avoid any costs by setting the rating $D_t^\star$ such that $D_t^\star \leq f(\theta)$ for all θ with non-zero belief, i.e. for all θ with $f(\theta) < E_t$. The rating agency's best response then changes from (14) in Proposition 2 to $g_1(e; f) = \min\left(e, \inf_{\theta\in\Theta} f(\theta)\right)$, for $e \geq 0$. The interpretation of $\alpha = 1$ is that the rating agency underestimates the default threshold and therefore the default risk by assuming the best case. As a consequence, the firm has the full benefit from the information asymmetry. The rating is always better than under perfect information. Accordingly, firms pay less interest than under perfect information, where the interest differential is increasing for increasing type, and firms delay default compared to perfect information regardless of the observed type. In the most progressive rating specification, equity holders benefit at the expense of debt holders.

The equilibria for both cases, most progressive and most conservative, respectively, are obtained in a similar fashion as in Proposition 3 and Proposition 4. In fact, the derivation is much simpler, as the best response of the rating agency has a simple and explicit structure and is therefore omitted here.

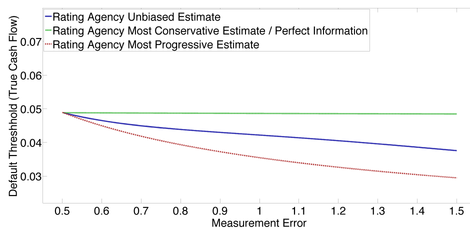
Figure 8 displays the impact of alternative incentives of the rating agency on the strategies in equilibrium. Panel 8a shows the equilibrium strategy of the rating agency: The solid line displays the rating agency's strategy as discussed previously. The red and green lines represent the most progressive and conservative cases, respectively. The rating agency that maximally overestimates the type learns much earlier compared to the base case. In contrast, the most progressive rating agency always assumes the best case, that is, it maximally underestimates the type. As the firm subject to the



(a) *Equilibrium Strategy of Rating Agency*



(b) *Equilibrium Strategy of Firm (Observed Cash Flow)*



(c) *Equilibrium Strategy of Firm (True Cash Flow)*

Figure 8: *Panel a displays the equilibrium strategy of the rating agency $g^\star$ for the base case (blue) and the most progressive and conservative rating agency (red and green) depending on the smallest imperfectly observed cash flow e. Also, the default threshold under perfect information is included. Panel b displays the equilibrium strategy of the firm $f^\star$ in terms of the default threshold depending on the type θ for the base case (blue) and the most progressive and conservative rating agency (red and green). Panel c displays the equilibrium strategy of the firm $f^\star$ in terms of the default threshold in the firm's true cash flow depending on the type for the base case (blue) and the most progressive and conservative rating agency (red and green), and the equilibrium default threshold in the perfect information case (dashed).*

most serious underestimation of type defaults last, the rating agency never updates its estimate and hence does not learn.

Panels 8b and 8c show the firm's strategy for a maximally progressive and conservative rating agency in both the observed cash flow (Panel 8b), and the true cash flow (Panel 8c), respectively. In Panel 8b, we see the firm default thresholds as a function of type, for the base case (blue) and maximally over- (red) and underestimated cash flow (green). A firm which faces a progressive rating agency defaults for a lower observed cash flow, because it profits more from the reduced interest payments. A firm facing a conservative rating agency defaults for a higher observed cash flow, as the cash flow underestimation and the subsequent lower rating and adverse interest payments make it more attractive to stop injecting additional liquidity. In Panel 8c, we show the firm's strategy in terms of the true cash flow. Comparing the firm's equilibrium strategy in terms of observed and true cash flow, we observe that for a given type, the ordering of the default thresholds is the same, that is, the firm defaults earliest facing a conservative rating agency, followed by the unbiased and the progressive case. This is intuitive, as the more conservative the rating agency, the higher is the cost of debt for a firm of any given type (except for the lowest possible type). However, the slopes of the default thresholds as a function of type are opposite for the observed and true cash flow representation. This is due to the fact that the measurement error θ enters as the scaling factor of the observed cash flow D.

When analyzing the unbiased rating strategy in Section 3.1, we have pointed out that it results in lower predicted default thresholds than the true default threshold of a fixed type from some type-dependent time before actual default, which we coin "ex-post rating inflation". For the generalized reputation cost function proposed in the current section, Panel 8c illustrates that ex-post rating inflation is only eliminated for the most conservative rating agency ($\alpha = 0$). Such an agency's policy makes the firm default at the perfect-information default threshold regardless of the observed type. Otherwise, all types and all reputation cost functions except for the one with $\alpha = 0$ induce ex-post rating inflation to varying degrees.

5.6 Generalizations of Model Implications

Our model is formulated as a game between rating agency and rated entity. However, its implications can be understood much more broadly. Note that the rating agency in our model faces the same differential of information towards the firm as the capital market participants in general, specifically the investors on the bond market. The function of the rating agency is rather a didactical one, namely to simplify the exposition of the interaction between outsiders' perception of the firm's quality and the firm's cost of debt financing. We use the concept of performance-sensitive debt (PSD) as in Manso, Strulovici, and Tchistyi (2010) to formalize this interaction. In a more general interpretation, think of a firm that has to roll over maturing debt in regular intervals. Then, for the new debt issue the firm will face higher yields, if its credit quality has deteriorated. Therefore, also a firm issuing plain vanilla corporate bonds repeatedly faces similar effects on the cost of debt financing as one using performance-sensitive debt. The overall capital market, assuming competitive pricing, will just require the yield for the new issue that is appropriate given the firm's endogenous default strategy. Thus, performance-sensitive debt should be seen as a modeling alternative to the classical approaches introducing finite maturity into infinite-horizon structural models, such as Leland and Toft (1996) and Leland (1998), and it allows us to analyze effects similar to those occurring with the roll-over of short-term or finite-maturity debt.

According to our model, one of the hidden pieces of public information that the market participants should take into account, or that possibly proxies for an unobservable factor that market participants take into account already today, is the historical minimum of the firm's observable status, possibly combined with the information that equity issuance has been used, to overcome the situation with low observables. On the other hand, the firm's equity holders take into account these rational judgments of the capital market participants, when estimating the firm's future cost of debt financing, and basing the decision of whether to default or to inject new equity capital. This interpretation holds either when the equity holders are indeed better informed than debt investors, e.g., for a firm with privately held or at least not widely dispersed equity ownership. Otherwise, one should think of better informed managers acting in the interest of equity holders.

6 Conclusion

This paper analyzes a continuous-time rating game between a rating agency and a rated firm. The main friction in the model is a measurement error of soft information, which implies imperfectly observed cash flows. Both the rating agency and the firm's decision are linked through a feedback effect: The rating influences the firm's capital costs, to which the rating responds again. The firm's non-default in periods of apparent crisis signals that over-estimation of the cash flow is only possible to a certain extent. Hence, the rating agency's optimal strategy is to issue a higher rating for the same current cash flow, if the historical minimum has been sufficiently low. The firm responds to the rating strategy by maximizing its firm value by defaulting at a type-dependent default threshold.

The central mechanism of the game is what we coin “directional learning”. The rating agency rules out types over time by noting that the firm does not default for observed cash flow levels too low for more overestimated types to survive. This one-sided narrowing of the measurement error implies "ex-post rating inflation". As the rating agency rules out more and more types, at some point the true type is above the rating agency's estimate, leaving the rating agency to unconsciously overestimate the measurement error. Furthermore, the uncertainty over measurement errors delays the firm's default, and enhances the firm's equity value at the expense of debt-holders.

The paper provides a rich framework for studying feedback effects in dynamic structural models in potential extensions. Our structural model framework allows for a broader concept of asymmetric information by adding ambiguity on the type, so that the rating agency is not only limited by its imperfect observation of the firm's cash flow, but also uncertain about the exact distribution of the measurement error of the soft information, which better captures the nature of this vaguer concept. In another direction, our model framework carries over to the valuation of real options in structural models in corporate finance, which allows for embedding feedback effects into these modeling approaches.

References

Acharya, Avidit and Juan Ortner (2017). "Progressive Learning". In: *Econometrica* 85.6, pp. 1965–1990.

Barron, E. N. (1993). "The Bellman equation for control of the running max of a diffusion and applications to look-back options". In: *Applicable Analysis* 48.1-4, pp. 205–222.

Bielecki, Tomasz R and Marek Rutkowski (2004). *Credit risk: modeling, valuation and hedging*. Springer.

Black, Fischer and John C Cox (1976). "Valuing corporate securities: Some effects of bond indenture provisions". In: *The Journal of Finance* 31.2, pp. 351–367.

Board, Simon and Moritz Meyer-ter Vehn (2013). “Reputation for quality”. In: *Econometrica* 81.6, pp. 2381–2462.

Bolton, Patrick, Xavier Freixas, and Joel Shapiro (2012). “The Credit Ratings Game”. In: *Journal of Finance* 67.1, pp. 85–112.

Bustamante, M. Cecila (2012). “The Dynamcis of Going Public”. In: *Review of Finance* 16, pp. 577–618.

Duffie, Darrell and David Lando (2001). “Term structures of credit spreads with incomplete accounting information”. In: *Econometrica* 69.3, pp. 633–664.

Faingold, Eduardo and Yuliy Sannikov (2011). “Reputation in Continuous-Time Games”. In: *Econometrica* 79.3, pp. 773–876.

Fershtman, Chaim and Ariel Pakes (2012). “Dynamic Games with Asymmetric Information: A Framework for Empirical Work”. In: *Quarterly Journal of Economics*, pp. 1611–1661.

Frenkel, Sivan (2015). “Repeated interaction and rating inflation: A model of double reputation”. In: *American Economic Journal: Microeconomics* 7.1, pp. 250–280.

Fulghieri, Paolo, Günter Strobl, and Han Xia (2014). “The economics of solicited and unsolicited credit ratings”. In: *Review of Financial Studies* 27.2, pp. 484–518.

Goldstein, Robert, Nengjiu Ju, and Hayne Leland (2001). “An EBIT-based model of dynamic capital structure”. In: *The Journal of Business* 74.4, pp. 483–512.

Grenadier, Steven R and Andrey Malenko (2011). “Real options signaling games with applications to corporate finance”. In: *The Review of Financial Studies* 24.12, pp. 3993–4036.

Grenadier, Steven R, Andrey Malenko, and Nadya Malenko (2016). “Timing decisions in organizations: Communication and authority in a dynamic environment”. In: *The American Economic Review* 106.9, pp. 2552–2581.

Grenadier, Steven R, Andrey Malenko, and Ilya A Strebulaev (2014). "Investment Busts, Reputation, and the Temptation to Blend in with the Crowd". In: *Journal of Financial Economics* 111.1, pp. 137–157.

Gryglewicz, Sebastian and Barney Hartman-Glaser (2014). "Dynamic Agency and Real Options". Working Paper.

Halac, Marina and Ilan Kremer (2018). "Experimenting with Career Concerns".

He, Zhiguo and Konstantin Milbradt (2016). "Dynamic debt maturity". In: *The Review of Financial Studies* 29.10, pp. 2677–2736.

He, Zhiguo and Wei Xiong (2012a). "Dynamic debt runs". In: *The Review of Financial Studies* 25.6, pp. 1799–1843.

— (2012b). "Rollover risk and credit risk". In: *The Journal of Finance* 67.2, pp. 391–430.

Heinricher, Arthur C. and Richard H. Stockbridge (1991). "Optimal control of the running max". In: *SIAM Journal on Control and Optimization* 29.4, pp. 936–953.

Hirth, Stefan (2014). "Credit rating dynamics and competition". In: *Journal of Banking & Finance* 49, pp. 100–112.

Hirth, Stefan and Marliese Uhrig-Homburg (2010). "Investment timing, liquidity, and agency costs of debt". In: *Journal of Corporate Finance* 16.2, pp. 243–258.

Jarrow, Robert A, David Lando, and Stuart M Turnbull (1997). "A Markov model for the term structure of credit risk spreads". In: *Review of Financial Studies* 10.2, pp. 481–523.

Kruse, Thomas and Philipp Strack (2015). "Optimal stopping with private information". In: *Journal of Economic Theory* 159, pp. 702–727.

Leland, Hayne E (1994). “Corporate debt value, bond covenants, and optimal capital structure”. In: *The journal of finance* 49.4, pp. 1213–1252.

— (1998). “Agency costs, risk management, and capital structure”. In: *The Journal of Finance* 53.4, pp. 1213–1243.

Leland, Hayne E and Klaus Bjerre Toft (1996). “Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads”. In: *The Journal of Finance* 51.3, pp. 987–1019.

Lerner, Josh and Jean Tirole (2006). “A Model of Forum Shopping”. In: *American Economic Review* 96.4, pp. 1091–1115.

Manso, Gustavo (2013). “Feedback effects of credit ratings”. In: *Journal of Financial Economics* 109.2, pp. 535–548.

Manso, Gustavo, Bruno Strulovici, and Alexei Tchistyi (2010). “Performance-Sensitive Debt”. In: *Review of Financial Studies* 23.5, pp. 1819–1854.

Maskin, Eric and Jean Tirole (1988). “A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs”. In: *Econometrica* 56.3, pp. 549–569.

Morellec, Erwan and Norman Schürhoff (2011). “Corporate investment and financing under asymmetric information”. In: *Journal of Financial Economics* 99, pp. 262–288.

Opp, Christian C., Marcus M. Opp, and Milton Harris (2013). “Rating agencies in the face of regulation”. In: *Journal of Financial Economics* 108.1, pp. 46–61. ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2012.10.011.

Pastor, Lubos and Pietro Veronesi (2009). “Learning in Financial Markets”. In: *Home Annual Review of Financial Economics* 1.1, pp. 361–381.

Peskir, Goran and Albert Shiryaev (2006). *Optimal stopping and free-boundary problems*. Springer.

Ross, Stephen A. (1977). “The Determination of Financial Structure: The Incentive-Signalling Approach”. In: *The Bell Journal of Economics* 8.1, pp. 23–40.

Sannikov, Yuliy and Andrezej Skrzypacz (2010). “The Role of Information in Repeated Games with Frequent Actions”. In: *Econometrica* 78.3, pp. 847–882.

Standard and Poor’s (2016). *Default, Transition, and Recovery: 2015 Annual Global Corporate Default Study And Rating Transitions*.

Thomas, Caroline (2018). “Career Concerns and Policy Persistence - Dynamic Signalling with Changing Types”.

A Additional Results, Tables, and Figures

A.1 Interest Rates

We use Lemma 3.1.2 of Bielecki and Rutkowski (2004),

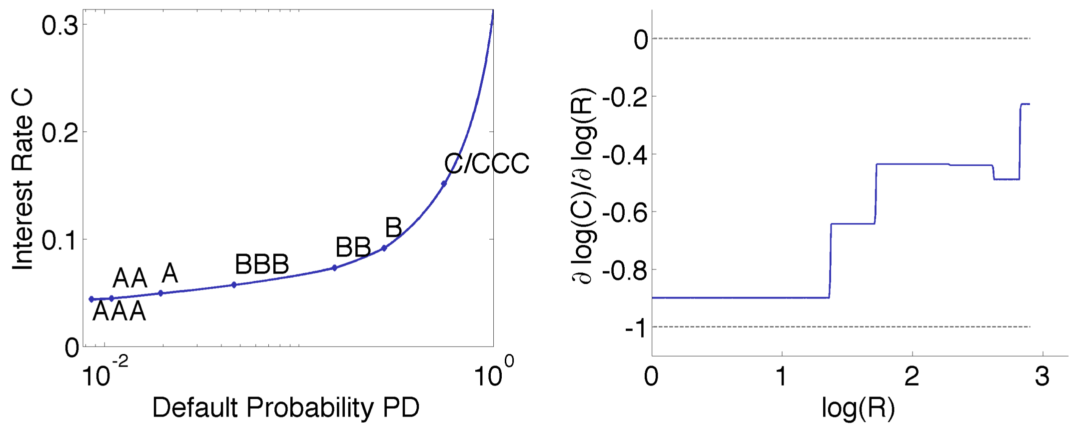
$$
\begin{aligned}
PD(T) &= \mathbb{P}\left(\inf_{t\in[0,T]} X_t \le f_1^\star\right) = \mathbb{P}\left(\inf_{t\in[0,T]} X_t \le X_0/R\right) \\
&= R^{-2\,\frac{\mu-\sigma^2/2}{\sigma^2}} N\left(\frac{-\ln(R) + (\mu-\sigma^2/2)\,T}{\sigma\sqrt{T}}\right) + N\left(\frac{-\ln(R) - (\mu-\sigma^2/2)\,T}{\sigma\sqrt{T}}\right),
\end{aligned}
\tag{27}
$$

where N is the standard normal cumulative distribution function, $f_1^\star$ is the firm's default threshold in the case in which the rating agency can observe the cash flow perfectly, and $R = X_0/f_1^\star$ is the distance-to-default type rating associated with the given default probability $PD(T)$, for time horizon $T > 0$. When we solve this equation for each $PD(10)_i$, with i denoting the rating class, we obtain the corresponding rating R_i in our scale. The results are given in Table 1. The values of C on $[1, \infty)$ are obtained by linear interpolation and extrapolation on the log-scale.

Figure A.1 illustrates Assumption 2 in light of the above interest rate structure. Panel A.1a displays the average interest payment rate C as a function of default probability PD. Figure 2 in the main text displays the average interest payment rate C as a function of rating R. If the rating turns bad, that is, if the rating implies a high default probability, then the firm's interest payments increase sharply. On the other hand, if the firm is far away from default, a further increase in the rating has only a small effect on the firm's interest payments. Panel A.1b draws the slope on the log-log scale of the interest payment rate function C depending on R based on the values given in Table 1. In particular, the slope in the log-log scale lies in $[-L_C, 0)$ with $L_C = 0.8989$ and thus satisfies Assumption 2.

A.2 Figures on Strategies

Figures A.2 and A.3 illustrate the rating agency's learning. In Figure A.2, we see the true (black line) and imperfectly observed cash flow (blue line) from the rating



(a) *Default Probabilities* ***(b)*** *Test of Assumption 2*

Figure A.1: Interest rates for different ratings. *Panel A.1a displays the average interest payment rate C as a function of default probability PD. Linear interpolation and extrapolation, respectively, is applied on the log-log scale. Panel A.1b draws the slope on the log-log scale of the interest payment rate function C depending on R based on the values given in Table 1.*

agency's perspective. The gray line displays the estimated median cash flow based on the rating agency's information. The solid red line is the firm's true default threshold, while the dashed red line is the default level estimated by the rating agency. For a true type of $\theta = 0.5$, the rating agency starts by underestimating the true cash flow. In the beginning, the only available information is the observed cash flow (coinciding with the estimated median cash flow). As time progresses, the observed cash flow decreases, leading the rating agency to update its belief by ruling out the most overestimated types. As Figure A.3 shows, the rating agency updates its estimated median type from the starting value of 1 to a lower value, which increases the estimated cash flow relative to the observed one. During the second crisis, the rating agency learns again and rules out more overestimated types. Because the estimated type is now below the true type, as Figure A.3 illustrates, the estimated cash flow in Figure A.2 is now overestimated by the rating agency. Figures A.2 and A.3 also indicate what we coin "ex-post rating inflation".

Figure A.4 illustrates the firm's best response. The imperfectly observed cash flow which the rating agency observes and its running minimum are given in blue and black, respectively. There are two types in this example: An overestimated type (dashed gray line) and an underestimated type (solid gray line). As in Lemma 2, the

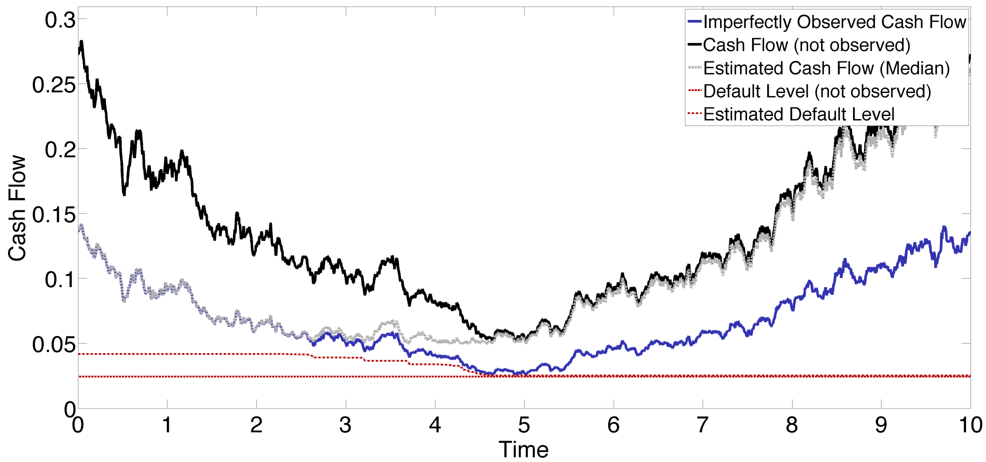


Figure A.2: Learning of the Rating Agency – Cash Flow Process. *This graph displays the true (black) and imperfectly observed cash flow (blue) from the rating agency's perspective. The gray line displays the estimated median cash flow based on the rating agency's information. The solid red line is the firm's true default threshold, while the dashed red line is the default level estimated by the rating agency. The type information is presented in Figure A.3. The true type is* $\theta = 0.5$.

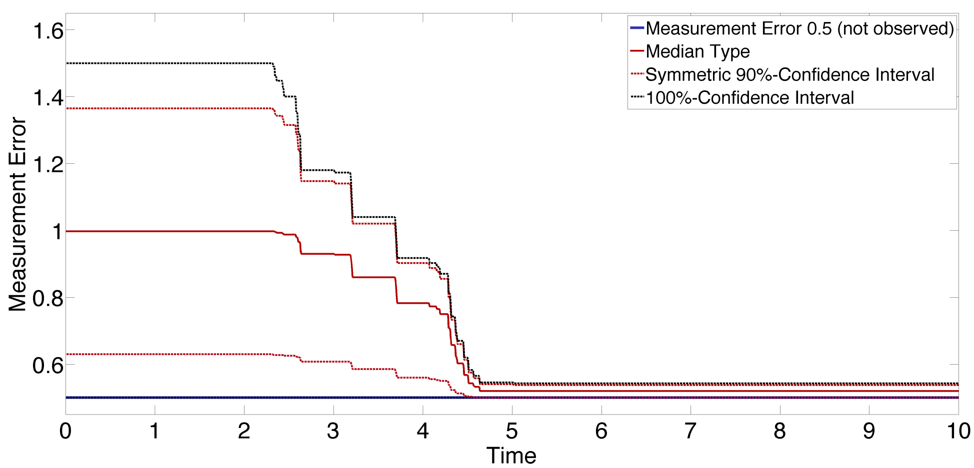


Figure A.3: Learning of the Rating Agency – Type Estimate. *This figure shows the true type of* $\theta = 0.5$ *(black) and the rating agency's estimated median type updated over time (blue). The dashed lines indicate the 90% and 100% confidence intervals based on the rating agency's updated belief. The cash flows for this figure are presented by Figure A.2.*

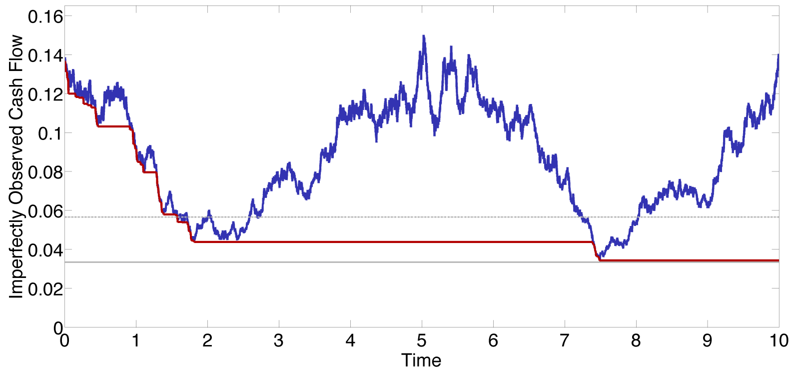


Figure A.4: Best response of firms. *This figure displays the best response to the rating agency's default threshold for a firm of an overestimated type (dashed red line) and a firm of an underestimated type (solid red line) for a specific imperfectly observed cash flow (blue) and its running minimum (black).*

firm has type-dependent thresholds at which it defaults, where the firm in case of underestimation defaults only at a lower imperfectly observed cash flow. When the imperfectly observed cash flow hits the default threshold for the overestimated case, the firm defaults and reveals its true cash flow, while for the underestimated the firm does not.

A.3 Transform of the Rating Agency's Strategy

A.4 Remark ODE

Remark 1. *The function $h(f(\theta), \hat{g}(\theta), f'(\theta), \hat{g}'(\theta), \theta) = \frac{\partial b}{\partial e}(f(\theta), \theta; \hat{g} \circ f)$ needs to be computed to calculate the equilibrium $(f^\star, g^\star)$. The proof of Proposition 6 shows how to obtain this quantity numerically. Define the function $\hat{v} : \mathbb{R}^+ \times \Theta \times \mathbb{R}^+ \to \mathbb{R}$, $(d; \theta, g) \mapsto \hat{v}(d; \theta, g)$ by*

$$\hat{v}(d; \theta, g) = \sup_{\tau \in \mathcal{T}_d} \mathbb{E}_d \left[\int_0^\tau e^{-r\,t} \, (d_t/\theta - C(d_t/g)) \, \mathrm{d}t \right] ,$$

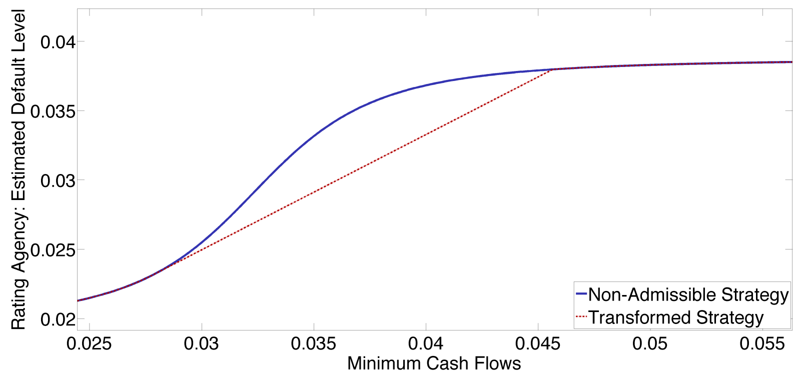


Figure A.5: Transform of the rating agency's strategy. *This figure displays an example of a transform of a non-admissible strategy. It plots the rating strategy, i.e., the estimated default threshold, as a function of the minimum imperfectly observed cash flow. The graph displays the original rating strategy (solid line) and the transformed estimated default threshold (dashed line).*

where $\mathcal{T}_d$ is the set of all stopping times with respect to the information generated by D with starting value d and $\mathbb{E}_d$ is the corresponding expectation. In contrast to the function v defined in (16)*, the direct dependence on the minimum observed cash flow E is eliminated. Instead, a conventional optimal stopping problem in the observed cash flow D is given, parameterized by $\theta \in \Theta$ and $g \in \mathbb{R}^+$. However, for $e \leq f(\theta; g)$ we have $v(d, e; \theta, g) = \hat{v}(d; \theta, g(e))$, for $(d, e) \in \mathcal{C}$ and $\theta \in \Theta$. From the arguments in the proof of Proposition 6, $g(f(\theta)) = \hat{g}(\theta)$ and $(g^\star)'(f(\theta)) = \hat{g}'(\theta)/f'(\theta)$ it follows that*

$$\begin{aligned}\frac{\partial b}{\partial e}(f(\theta), \theta; g^\star) &= \lim_{\Delta e \searrow 0} \frac{1}{\Delta e}\left(f(\theta) - \sup\{d > 0 : \hat{v}(d; \theta, \hat{g}(\theta) - \frac{\hat{g}'(\theta)}{f'(\theta)}\Delta e) = 0\}\right) \\ &\approx \frac{1}{\Delta e}\left(f(\theta) - \sup\{d > 0 : \hat{v}(d; \theta, \hat{g}(\theta) - \frac{\hat{g}'(\theta)}{f'(\theta)}\Delta e) = 0\}\right).\end{aligned}$$

for sufficiently small Δe, where the critical level $\sup\{d > 0 : \hat{v}(d; \theta, \hat{g}(\theta) - \frac{\hat{g}'(\theta)}{f'(\theta)}\Delta e) = 0\}$ is obtained by solving the free-boundary value problem associated optimal stopping problem with value function $\hat{v}(\cdot; \theta, \hat{g}(\theta) - \frac{\hat{g}'(\theta)}{f'(\theta)}\Delta e)$ numerically.

B Proof of Main Results

Definition 1 (Perfect Bayesian Equilibrium in Markov Strategies)**.** *Strategies* $(\tau^\star, \hat{D}^{\star\star})$ *and beliefs* $\pi^\star$ *constitute a perfect Bayesian equilibrium in Markov strategies (PBEM) if:*

1. *For every* $0 \le t < \tau^\star$*,* $\theta \in \Theta$*, and strategy* $\tau(\theta)$

$$\mathbb{E}\left[\int_t^{\tau^\star(\theta)} e^{-r\,s}\left(D_s/\theta - C(D_s/\hat{D}_s^{\star\star})\right)\,\mathrm{d}s\middle|\mathcal{F}_t\right] \\ \ge \mathbb{E}\left[\int_t^{\tau(\theta)} e^{-r\,s}\left(D_s/\theta - C(D_s/\hat{D}_s^{\star\star})\right)\,\mathrm{d}s\middle|\mathcal{F}_t\right].$$

2. *For every* $0 \le t < \tau^\star$ *and strategy* $\hat{D}^\star$

$$-\mathbb{E}\left[\int_t^{\tau^\star} e^{-\rho\,s}\int_\Theta (\hat{D}_s^{\star\star} - \mathbb{E}\left[D_{\tau^\star(\theta)}\middle|\mathcal{G}_s\right])^2\,\phi_s^{\pi^\star}(\theta)\,\mathrm{d}\theta\,\mathrm{d}s\middle|\mathcal{G}_t\right], \\ \ge -\mathbb{E}\left[\int_t^{\tau^\star} e^{-\rho\,s}\int_\Theta (\hat{D}_s^{\star} - \mathbb{E}\left[D_{\tau^\star(\theta)}\middle|\mathcal{G}_s\right])^2\,\phi_s^{\pi^\star}(\theta)\,\mathrm{d}\theta\,\mathrm{d}s\middle|\mathcal{G}_t\right].$$

3. *Bayes rule is used to update beliefs* $\pi^\star$ *with density* $(\phi_t^{\pi^\star})_{t\ge 0} = (L_t^{\pi^\star}\,\phi)_{t\ge 0}$ *whenever possible: For every* $t \ge 0$*, if there exists* $\theta_0 \in \Theta$ *such that* $t < \tau^\star(\theta_0)$*, then*

$$\phi_t^{\pi^\star}(\theta) = \frac{\phi_{t-}^{\pi^\star}(\theta)\,\mathbf{1}_{t<\tau^\star(\theta)}}{\int_\Theta \phi_{t-}^{\pi^\star}(\theta')\,\mathbf{1}_{t<\tau^\star(\theta')}\mathrm{d}\theta'},\ \textit{for } \theta\in\Theta\,,$$

where $\phi_0^{\pi^\star} = \phi$*, i.e.* $L_0^{\pi^\star}(\theta) = 1$*, for* $\theta \in \Theta$*.*

4. *The strategies are Markov, i.e.:*

$$\tau^\star(\theta) = \inf\{t \ge 0 : (D_t, E_t) \in \mathcal{E}(\theta)\}\,,\ \textit{for a Borel set } \mathcal{E}(\theta) \subseteq \mathbb{R}^+, \theta \in \Theta\,,$$
$$\hat{D}_t^{\star\star} = g(D_t, E_t)\,,\ \textit{for some function } g : \mathcal{C} \to \mathbb{R}_0^+ \textit{ for } 0 \le t < \tau^\star\,.$$

Proof of Proposition 1. The function $f \in \mathcal{A}_f$ specifying the firm's strategy is given, which is then $\tau = (\tau(\theta))_{\theta \in \Theta}$ with $\tau(\theta) = \inf\{t \geq 0 : D_t \leq f(\theta)\}$, for $\theta \in \Theta$. The structure of the rating agency's belief π that is consistent with the firm strategy f as given in (13) follows from Bayes' rule. Using the consistent belief π, the rating agency maximizes the respective utility, i.e.

$$\sup_{g \in \mathcal{A}_g} U^{\pi}_{RA}(\tau, g(E)) = - \inf_{g \in \mathcal{A}_g} \mathbb{E}\left[\int_0^{\tau} e^{-\rho t} \int_{\Theta} (g(E_t) - f(\theta))^2 \, \phi^{\pi}_t(\theta) \, \mathrm{d}\theta \, \mathrm{d}t\right] .$$

The above expression is minimized in case $g(E_t)$ minimizes for each $0 \leq t < \tau$

$$\begin{aligned} \int_{\Theta} (g(E_t) - f(\theta))^2 \, \phi^{\pi}_t(\theta) \, \mathrm{d}\theta &= \frac{\displaystyle\int_{\theta} (g(E_t) - f(\theta))^2 \, \mathbf{1}_{f(\theta) < E_t} \, \phi(\theta) \, \mathrm{d}\theta}{\displaystyle\int_{\Theta} \mathbf{1}_{f(\theta) < E_t} \, \phi(\theta) \, \mathrm{d}\theta} \\ &= \mathbb{E}\left[(g(e) - f(\tilde{\theta}))^2 \Big| \, f(\tilde{\theta}) < e \right]\Big|_{e = E_t} . \end{aligned}$$

In fact, we look for $g(E_t)$ which minimizes the squared distance to the random variable $f(\tilde{\theta})|_{f(\tilde{\theta}) < E_t}$, which is a function f of the random variable describing the type $\tilde{\theta}$ based on the consistent belief π_t. Therefore, the optimal $g(E_t; f)$ has to be the expected value of $f(\tilde{\theta})|_{f(\tilde{\theta}) < E_t}$, which is in essence (14). This result can also be obtained by solving the first order condition and checking the second order condition for a minimum. For the irrelevant case $t \geq \tau$, or, equivalently, $E_t < \inf_{\theta \in \Theta} f(\theta)$, we set the critical default level at $g(e; f) = e$, for $e \leq \inf_{\theta \in \Theta} f(\theta)$. Thus, default is predicted to happen immediately as it should have been occurred before. Clearly, $g(\cdot; f) \in \mathcal{A}_g$.

From (14) we see that $g(\cdot; f)$ is bounded by Id, i.e. $g(e; f) \leq e$, for $e \geq 0$, and that $g(\cdot; f)$ is non-decreasing. Assuming that f is strictly increasing, f^{-1} is well defined on

$f(\Theta)$. Take $e > e' > \inf_{\theta\in\Theta} f(\theta)$ and then

$$\begin{aligned}
g(e;f) - g(e';f) =& \frac{\int_{\theta<f^{-1}(e)} f(\theta)\,\phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e)} \phi(\theta)\,\mathrm{d}\theta} - \frac{\int_{\theta<f^{-1}(e')} f(\theta)\,\phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e')} \phi(\theta)\,\mathrm{d}\theta} \\
=& \frac{\int_{\theta<f^{-1}(e')} f(\theta)\,\phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e')} \phi(\theta)\,\mathrm{d}\theta} \frac{\int_{\theta<f^{-1}(e')} \phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e)} \phi(\theta)\,\mathrm{d}\theta} \\
&+ \frac{\int_{f^{-1}(e')}^{f^{-1}(e)} f(\theta)\,\phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e)} \phi(\theta)\,\mathrm{d}\theta} - \frac{\int_{\theta<f^{-1}(e')} f(\theta)\,\phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e')} \phi(\theta)\,\mathrm{d}\theta} \\
=& g(e';f)\,\frac{-\int_{f^{-1}(e')}^{f^{-1}(e)} \phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e)} \phi(\theta)\,\mathrm{d}\theta} + \frac{\int_{f^{-1}(e')}^{f^{-1}(e)} f(\theta)\,\phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e)} \phi(\theta)\,\mathrm{d}\theta} \\
=& \frac{\int_{f^{-1}(e')}^{f^{-1}(e)} (f(\theta) - g(e';f))\,\phi(\theta)\,\mathrm{d}\theta}{\int_{\theta<f^{-1}(e)} \phi(\theta)\,\mathrm{d}\theta}\,.
\end{aligned}$$

The above quantity is non-negative and further strictly positive on the interior of $f(\Theta)$ and converges to zero for $e \searrow e'$. Accordingly, $g(\cdot;f)$ is continuous and strictly increasing as claimed. $\square$

Lemma 1. *Suppose that $g \in \mathcal{A}_g$ is non-decreasing and is bounded by Id, i.e. $g(e) \le e$, for $e \ge 0$, then $\mathcal{R}(g) \in \mathcal{A}_g^{\mathbf{C}}$. Moreover, $\mathcal{R}(g) = g$, for $g \in \mathcal{A}_g^{\mathbf{C}}$.*

Proof of Lemma 1. We see immediately $\mathcal{R}(g) \le g \le Id$. To observe that $\mathcal{R}(g)$ is non-decreasing take $e' \ge e > 0$ and

$$\begin{aligned}
\mathcal{R}(g)(e') &= e' \inf\{t(z)/z : 0 < z \le e'\} \\
&= e' \left(\inf\{t(z)/z : 0 < z \le e\} \wedge \inf\{g(z)/z : e < z \le e'\}\right) \\
&\ge \left(\frac{e'}{e}\,\mathcal{R}(g)(e)\right) \wedge \left(e' \inf\{g(e)/z : e < z \le e'\}\right) \\
&= \left(\frac{e'}{e}\,\mathcal{R}(g)(e)\right) \wedge g(e) \\
&= \mathcal{R}(g)(e) + \left(\frac{e'-e}{e}\,\mathcal{R}(g)(e)\right) \wedge \left(g(e) - \mathcal{R}(g)(e)\right) \\
&\ge \mathcal{R}(g)(e)\,.
\end{aligned}$$

By definition, $\mathcal{R}(g)/Id$ is non-increasing. These two properties rule out negative jumps as well as positive jumps, and hence $\mathcal{R}(g)$ is continuous on $\mathbb{R}^+$. To check the continuity at 0, we observe that $0 \le \mathcal{R}(g)(e) \le e$, and thus $\lim_{e \searrow 0} \mathcal{R}(g)(e) = 0 = \mathcal{R}(g)(0)$, as defined in (15). To address the other claimed properties of $\mathcal{R}(g)$ observe that by g being non-decreasing it holds that

$$\begin{aligned}\mathcal{R}(g)(e) &= e \inf\{g(z)/z : 0 < z \le e\} \\ &\le e \inf\{g(e)/z : 0 < z \le e\} \;=\; e\, g(e)/e = g(e)\,.\end{aligned}$$

□

Proof of Proposition 2. For $g \in \mathcal{A}_g^{\mathbf{C}}$ and $\theta \in \Theta$, the early exercise region associated with the optimal stopping time of (16) is denoted by $\mathcal{E}(\theta; g) = \{(d, e) \in \mathcal{C} : v(d, e; \theta, g) = 0\}$, see also (58). The optimal strategy can be written as first entry time of the state process (D, E) with starting value (d, d), $d \ge 0$, in the early exercise region $\mathcal{E}(\theta; g)$,

$$\tau_{(d,d)}(\theta; g) = \inf\{t \ge 0 : (D(t), E(t)) \in \mathcal{E}(\theta; g)\}\,.$$

Now, $g \in \mathcal{A}_g^{\mathbf{C}}$, and we can apply the results of Lemma 7. Recall the definition in (59)

$$\mathcal{D}(\theta; g) = \{d \in \mathbb{R}_0^+ : (d, d) \in \mathcal{E}(\theta; g)\}\,, \text{ and } D(\theta; g) = \sup \mathcal{D}(\theta; g).$$

According to Lemma 7 we have that $\mathcal{D}(\theta; g) = [0, D(\theta; g)]$. Set $f(\theta; g) = D(\theta; g)$. The stopping time $\tau_{(d,d)}(\theta; g)$ is depending on the starting value (d, d). Consider first the case $d \le D(\theta; g) = f(\theta; g)$. Then $d \in \mathcal{D}(\theta; g)$ and by the definition of $\mathcal{D}(\theta; g)$ we conclude that $(d, d) \in \mathcal{E}(\theta; g)$. Accordingly, we start in the early exercise region and stop immediately at 0, i.e. $\tau_{(d,d)}(\theta; g) = 0$. Since we also have $D_0 = d \le D(\theta; g) = f(\theta; g)$, (22) holds true. Now, consider $d > D(\theta; g) = f(\theta; g)$. The starting value (d, d) of (D, E) is not in $\mathcal{E}(\theta; g)$. The process (D, E) has continuous paths and E is the running minimum of D. Thus, if (D, E) decreases in the second component then it has to travel through the diagonal $\mathcal{D}$. Since $\{(d, d) : d \in \mathcal{D}(\theta; g)\} \subseteq \mathcal{E}(\theta; g)$, we have that

$$\begin{aligned}\tau_{(d,d)}(\theta; g) &= \inf\{t \geq 0 : (D(t), E(t)) \in \mathcal{E}(\theta; g)\} \\ &\leq \inf\{t \geq 0 : (D_t, E_t) \in \{(d, d) : d \in \mathcal{D}(\theta; g)\}\} \\ &= \inf\{t \geq 0 : D_t \leq f(\theta; g)\} .\end{aligned}$$

Suppose now that $\tau_{(d,d)}(\theta; g) < \inf\{t \geq 0 : D_t \leq f(\theta; g)\}$ with some non-negative probability. Thus (D, E) has to hit $\mathcal{E}(\theta; g)$ with some non-negative probability before it eventually hits $(f(\theta; g), f(\theta; g))$. Then there exists an $(d', e') \in \mathcal{E}(\theta; g)$ with $e' > f(\theta; g)$. However, this would contradict (61). Therefore, $\tau_{(d,d)}(\theta; g) = \inf\{t \geq 0 : D_t \leq f(\theta; g)\}$ almost surely and (22) holds also in this case. □

For our specific default barrier, we can provide more structure on how the barrier changes in the type in Lemma 2.

Lemma 2. *Given the setting of Proposition 2 and let $\theta, \theta' \in \Theta$ with $\theta' \leq \theta$, then*

$$f(\theta'; g) \leq f(\theta; g) \,, \text{ and } \underline{f} \leq \frac{f(\theta; g)}{\theta} \leq \frac{f(\theta'; g)}{\theta'} \leq \bar{f} \,, \tag{28}$$

uniformly in g, where $0 < \underline{f} \leq \bar{f} < \infty$. In particular, $f(\cdot; g)$ is Lipschitz continuous

$$|f(\theta; g) - f(\theta'; g)| \leq L_f \, |\theta - \theta'| \,, \text{ for } \theta, \theta' \in \Theta \,, \tag{29}$$

where $L_f = \bar{f} > 0$ is the uniform Lipschitz constant for all g. Moreover, suppose that Assumption 2 holds, then for $\theta, \theta' \in \Theta$ with $\theta' \leq \theta$ it holds that

$$f(\theta; g) - f(\theta'; g) \geq l_f \, (\theta - \theta') \,, \tag{30}$$

where $l_f = (1 - L_C) \, \underline{f} > 0$ is the uniform constant for all g.

Proof of Lemma 2. For $\theta, \theta' \in \Theta$, with $\theta' \leq \theta$, and $g \in \mathcal{A}_g^{\mathbf{C}}$ is non-decreasing and bounded by Id, we have $v(\cdot,\cdot;\theta',g) \geq v(\cdot,\cdot;\theta,g)$ by part 2. of Lemma 6 and $v(\cdot,\cdot;\theta',g), v(\cdot,\cdot;\theta,g) \geq 0$ by part 1. of Lemma 6. From this, the corresponding early exercise regions $\mathcal{E}(\theta;g) = \{(d,e) \in \mathcal{C} : v(d,e;\theta,g) = 0\}$ and $\mathcal{E}(\theta';g) = \{(d,e) \in \mathcal{C} : v(d,e;\theta',g) = 0\}$, see also (58) for definition, satisfy $\mathcal{E}(\theta';g) \subseteq \mathcal{E}(\theta;g)$. This holds in particular on the diagonal $\mathcal{D}$, i.e. $\mathcal{E}(\theta';g) \cap \mathcal{D} \subseteq \mathcal{E}(\theta;g) \cap \mathcal{D}$, and $[0, D(\theta';g)] = \mathcal{D}(\theta';g) \subseteq \mathcal{D}(\theta;g) = [0, D(\theta;g)]$, thus $D(\theta';g) \leq D(\theta;g)$. Identifying $f(\theta';g) = D(\theta';g)$ and $f(\theta;g) = D(\theta;g)$ as is done in the proof of Proposition 2 gives $f(\theta';g) \leq f(\theta;g)$, establishing the first part of (28). For the second part, we rewrite (22) of Proposition 2 in the firm scale (x,y). With $(d,e) = \theta\,(x,y)$, we obtain $\tau_{(d,d)}(\theta;g) = \inf\{t \geq 0 : X_t \leq f(\theta;g)/\theta\}$, and $f(\theta;g)/\theta$ is the default barrier in the firm scale, for $\theta \in \Theta$. Regardless of the rating strategy g, the interest payment rate function C is bounded from below by $\underline{C}$ and from above by $\bar{C}$. Denote by $\underline{f}$ the default threshold in the firm scale for the constant interest $\underline{C}$ (i.e. by setting $\underline{g} = 0$) and by $\bar{f}$ the default threshold in the firm scale for the constant interest $\bar{C}$ (i.e. by setting $\bar{g} = \infty$, which is interpreted as limiting case, and noting that C is as in Lemma 6 extended to $[0,\infty]$ by setting $C|_{[0,1)} = C(1)$). From $\underline{g} \leq g \leq \bar{g}$ we obtain with part 3. of Lemma 6 that $w(\cdot,\cdot;\theta,\underline{g}) \geq w(\cdot,\cdot;\theta,g) \geq w(\cdot,\cdot;\theta,\bar{g})$, and hence $\tilde{\mathcal{E}}(\theta;\underline{g}) \subseteq \tilde{\mathcal{E}}(\theta;g) \subseteq \tilde{\mathcal{E}}(\theta;\bar{g})$, where $\tilde{\mathcal{E}}(\theta;g') = \{(x,y) \in \mathcal{C} : w(x,y;\theta,g') = 0\}$, for $g' \in \{\underline{g}, g, \bar{g}\}$. The boundary cases $\underline{g}$ and $\bar{g}$ also admit a critical default level, which is given by $\underline{f} = \frac{\eta\,(r-\mu)}{(1+\eta)\,r}\underline{C}$ and $\bar{f} = \frac{\eta\,(r-\mu)}{(1+\eta)\,r}\bar{C}$, respectively, where $\eta = \frac{b-\frac{1}{2}\sigma^2}{\sigma^2} + \sqrt{\left(\frac{b-\frac{1}{2}\sigma^2}{\sigma^2}\right)^2 + \frac{2\,r}{\sigma^2}} > 0$, see, e.g., Equation (C.5) in Manso (2013). And thus $\underline{f} \leq f(\theta;g)/\theta \leq \bar{f}$, for all $\theta \in \Theta$ and $g \in \mathcal{A}_g^{\mathbf{C}}$ that are non-decreasing and bounded by Id. Part 2. of Lemma 6 gives $w(\cdot,\cdot;\theta',g) \leq w(\cdot,\cdot;\theta,g)$, yielding $f(\theta;g)/\theta \leq f(\theta';g)/\theta'$, what completes the second part of (28). Finally, recall that $\theta' \leq \theta$ and use the just established results given in (28) to see

$$
\begin{aligned}
0 \le f(\theta; g) - f(\theta'; g) &= \theta \, \frac{f(\theta; g)}{\theta} - f(\theta'; g) \\
&\le \theta \, \frac{f(\theta'; g)}{\theta'} - f(\theta'; g) \;=\; (\theta - \theta') \frac{f(\theta'; g)}{\theta'} \le (\theta - \theta') \, \bar{f} \,,
\end{aligned}
$$

what proves (29). Now, assume that Assumption 2 holds. For $\theta, \theta' \in \Theta$, with $\theta' \le \theta$, and $g \in \mathcal{A}_g^{\mathbf{C}}$ is non-decreasing and bounded by Id, and $(d, e) \in \mathcal{C}$. From part 1. of Lemma 5 we see $g_{\theta'} \le \frac{\theta}{\theta'} g_\theta$ and since C is non-increasing $C(x/g_{\theta'}(y)) \le C((\theta'/\theta)x/g_\theta(y))$, for $(x, y) \in \mathcal{C}$. Assumption 2 gives

$$
C(z) \le C(z') \le (z/z')^{L_C} \; C(z) \,, \text{ for } 1 \le z' \le z \,.
$$

With $z' = (\theta'/\theta)x/g_\theta(y)$ and $z = x/g_\theta(y)$, it holds that $z' \le z$, and we obtain that $C((\theta'/\theta)x/g_\theta(y)) \le (\theta/\theta')^{L_C} C(x/g_\theta(y))$, and thus

$$
C(x/g_{\theta'}(y)) \le (\theta/\theta')^{L_C} C(x/g_\theta(y)) \,, \text{ for } (x, y) \in \mathcal{C} \,. \tag{31}
$$

Consider the optimal stopping problem in (16), but now scaled by $\hat{\theta} = (\theta/\theta')^{L_C} \ge 1$,

$$
\tilde{v}(d, e; \hat{\theta}, g_\theta) = \sup_{\tau \in \mathcal{T}_{(d,e)}} \mathbb{E}_{(d,e)} \left[\int_0^\tau e^{-r\,t} \left(D_t - \hat{\theta} \, C(D_t/g_\theta(E_t)) \right) \mathrm{d}t \right] ,
$$

or, noting that (E, D) has the same distribution as (X, Y) provided the starting values are identical, we can express this alternatively

$$
\tilde{v}(x, y; \hat{\theta}, g_\theta) = \sup_{\tau \in \mathcal{T}_{(x,y)}} \mathbb{E}_{(x,y)} \left[\int_0^\tau e^{-r\,t} \left(X_t - \hat{\theta} \, C(X_t/g_\theta(Y_t)) \right) \mathrm{d}t \right] ,
$$

and as in (52)

$$
w(x, y; \theta', g) = \sup_{\tau \in \mathcal{T}_{(x,y)}} \mathbb{E}_{(x,y)} \left[\int_0^\tau e^{-r\,t} \left(X_t - C(X_t/g_{\theta'}(Y_t)) \right) \mathrm{d}t \right] ,
$$

By Proposition 2, the latter two optimal stopping problems admit a critical default level for describing the optimal default time. The value function $\tilde{v}(\cdot,\cdot;\hat{\theta},g_\theta)$ is in the rating agency-scale (D,E) and the critical level is given by $f(\hat{\theta};g_\theta)$. The value function $w(\cdot,\cdot;\theta',g)$ is in the firm-scale (X,Y) and critical default level is given by $f(\theta';g)/\theta'$. Recalling that the interest payments are ordered uniformly, see (31), we can order the critical default levels, as a higher interest payment leads to a higher critical default value, and hence

$$f(\hat{\theta};g_\theta) \geq f(\theta';g)/\theta' \,.$$

Now, focus on $f(\hat{\theta};g_\theta)$ and recall that $\hat{\theta} = (\theta/\theta')^{L_C} \geq 1$, and since $f(\cdot;g)/Id$ is decreasing by (28)

$$f(\hat{\theta};g_\theta) = \hat{\theta}\frac{f(\hat{\theta};g_\theta)}{\hat{\theta}} \leq \hat{\theta}\,\frac{f(1;g_\theta)}{1} = \hat{\theta}\,f(1;g_\theta) = \hat{\theta}\,\frac{f(\theta;g)}{\theta}\,,$$

where the last step follows from plugging $(1,g_\theta)$ in (16) and comparing this with (52), to see that $f(1,g_\theta)$ is also the optimal default level in firm-scale for (θ,g), what is $f(\theta;g)/\theta$. Using this, we find a lower bound to

$$\begin{aligned} f(\theta;g) - f(\theta';g) &\geq f(\theta;g) - \theta'\,f(\hat{\theta};g_\theta) \geq f(\theta;g) - \theta'\frac{\hat{\theta}}{\theta}f(\theta;g) \\ &= f(\theta;g)\,\left(1-(\theta'/\theta)^{1-L_C}\right) \geq \underline{f}\,\theta\left(1-(\theta'/\theta)^{1-L_C}\right). \end{aligned}$$

Taking the limit gives

$$\begin{aligned} D_+f(\theta';g) &= \liminf_{\theta\searrow\theta'}\frac{f(\theta;g)-f(\theta';g)}{\theta-\theta'} \\ &\geq \liminf_{\theta\searrow\theta'}\frac{\underline{f}\,\theta\left(1-(\theta'/\theta)^{1-L_C}\right)}{\theta-\theta'} \;=\; (1-L_C)\,\underline{f}\,, \end{aligned}$$

implying the claimed inequality. □

Proof of Proposition 3. A fixed point is obtained by the Schauder fixed point theorem: Let $\mathcal{K}$ be a nonempty convex compact subset of a Banach space $\mathcal{V}$, if $\tilde{T} : \mathcal{K} \to \mathcal{K}$ is continuous, then $\tilde{T}$ has a fixed point. Set $\mathcal{V} = C(\Theta, \mathbb{R}) \times C(\Xi, \mathbb{R})$ endowed with the sup-norm, i.e. $\|(f, g)\|_\infty = \max(\|f\|_\infty, \|g\|_\infty)$, for $(f, g) \in \mathcal{V}$, where $\|f\|_\infty = \sup_{\theta \in \Theta} |f(\theta)|$, $\|g\|_\infty = \sup_{\xi \in \Xi} |g(\xi)|$, and the set $\Xi = [\underline{\xi}, \overline{\xi}]$ is defined by $\underline{\xi} = \underline{\theta}\,\underline{f}$ and $\overline{\xi} = \overline{\theta}^2\,\overline{f}^2/(\underline{\theta}\,\underline{f})$. Since Θ and Ξ are closed intervals, $\mathcal{V}$ is a Banach space as required. Next, we define the set $\mathcal{K} = \mathcal{K}_f \times \mathcal{K}_g$. The set $\mathcal{K}_f \subseteq C(\Theta, \mathbb{R}_0^+)$ should contain the firm strategies in $\mathcal{A}_f = C(\Theta, \mathbb{R}_0^+)$ that are relevant for fixed points of T. Based on Lemma 2 define

$$\begin{aligned}\mathcal{K}_f = \{f \in C(\Theta, \mathbb{R}_0^+) :& l_f\,(\theta - \theta') \leq f(\theta') - f(\theta) \leq L_f\,(\theta - \theta'), \\ & \underline{\theta}\,\underline{f} \leq f(\theta) \leq \overline{\theta}\,\overline{f}, \text{ for } \theta, \theta' \in \Theta \text{ with } \theta' \leq \theta\},\end{aligned} \tag{32}$$

where $0 < l_f = (1 - L_C)\,\underline{f} \leq \bar{f} = L_f$, with $0 \leq L_C < 1$. The set $\mathcal{K}_g \subseteq C(\Xi, \mathbb{R}_0^+)$ should contain the rating agency strategies that are relevant for fixed points of T. Proposition 1 and Lemma 4 suggest

$$\mathcal{K}_g = \{g \in C(\Xi, \mathbb{R}_0^+) : \underline{\xi} \leq g \leq Id, g \text{ non-decreasing}, g/Id \text{ non-increasing}\}. \tag{33}$$

It is sufficient to constrain the domain of the rating agency strategy g from originally $\mathbb{R}_0^+$ to Ξ for the following reason. By Proposition 1, we see that $g(\cdot; f)$, for $f \in \mathcal{K}_f \subseteq \mathcal{A}_f$, is continuous, non-decreasing, bounded by Id. Moreover, $g(\cdot; f)$ is determined by f on $f(\Theta)$. For $e \geq \overline{e}_f = \sup_{\theta \in \Theta} f(\theta)$, it holds $g(e; f) = g(\overline{e}_f; f)$, and $g(\cdot; f) = Id$ on $[0, \underline{e}]$, with $\underline{e}_f = \inf_{\theta \in \Theta} f(\theta)$. Further, $f(\Theta) \subseteq [\underline{\theta}\,\underline{f}, \overline{\theta}\,\overline{f}]$, for $f \in \mathcal{K}_f$. This property is preserved when considering $\mathcal{R} \circ g(\cdot; f)$ and the set $[\underline{\theta}\,\underline{f}, \overline{\theta}^2\,\overline{f}^2/(\underline{\theta}\,\underline{f})] = \Xi$ according to Lemma 1, i.e. $\mathcal{R} \circ g(e; f) = g(\overline{e}_f; f)$, for $e \geq \overline{\theta}^2\,\overline{f}^2/(\underline{\theta}\,\underline{f}) = \overline{\xi}$, and $\mathcal{R} \circ g(e; f) = e$, for $0 \leq e \leq \underline{\theta}\,\underline{f} = \underline{\xi}$, for all $f \in \mathcal{K}_f \subseteq \mathcal{A}_f$. For this reason we cut the non-relevant part of T off, i.e. we consider $\tilde{T} : \mathcal{K} \mapsto \mathcal{K}, (f, g) \mapsto (f(\cdot; g|_{\mathbb{R}_0^+}), \mathcal{R}(g(\cdot; f))|_\Xi)$, where $g|_{\mathbb{R}_0^+}$ is understood as the obvious continuation of $g \in \mathcal{K}_g$ from Ξ to $\mathbb{R}_0^+$ with

$$g|_{\mathbb{R}_0^+}(e) = \begin{cases} e & \text{for } 0 \leq e < \underline{\xi}, \\ g(e), & \text{for } e \in \Xi, \\ g(\overline{\xi}), & \text{for } e > \overline{\xi}, \end{cases}$$

and thus $\tilde{T}$ is well-defined. To show that $\tilde{T}$ has a fixed point, we have to prove that $\mathcal{K} \subseteq \mathcal{V}$ is nonempty, convex and compact, as well as $\tilde{T}$ is continuous.

The set $\mathcal{K} = \mathcal{K}_f \times \mathcal{K}_g$ is nonempty, convex and compact, if the factors $\mathcal{K}_f$ and $\mathcal{K}_g$ have these properties. By definition $\mathcal{K}_f$ and $\mathcal{K}_g$ are both nonempty. The convexity and compactness follows from Lemma 8 and Lemma 9, respectively. It remains to show that $g \mapsto f(\cdot; g|_{\mathbb{R}_0^+})$ and $f \mapsto \mathcal{R}(g(\cdot; f))|_{\Xi}$ are both continuous.

Consider the best response of the firm $f(\cdot;\cdot) : \mathcal{K}_g \to \mathcal{K}_f, g \mapsto f(\cdot; g|_{\mathbb{R}_0^+})$, see Proposition 2. Take $g, g' \in \mathcal{K}_g$ with $\|g - g'\|_\infty \leq \varepsilon$, for some $\varepsilon > 0$. Denote the continuations by $\hat{g} = g|_{\mathbb{R}_0^+}$ and $\hat{g}' = g'|_{\mathbb{R}_0^+}$, then $\|\hat{g} - \hat{g}'\|_\infty \leq \varepsilon$. Fix $\theta \in \Theta$, then

$$|\hat{g}_\theta(e) - \hat{g}'_\theta(e)| = \frac{1}{\theta}|\hat{g}(\theta\, e) - \hat{g}'(\theta\, e)| \leq \frac{\varepsilon}{\theta}\,, \text{ for all } e \geq 0\,,$$

and hence $\|\hat{g}_\theta - \hat{g}'_\theta\|_\infty \leq \varepsilon/\theta$, for $\theta \in \Theta$. For now, we fix $\theta \in \Theta$ and estimate $|f(\theta; \hat{g}) - f(\theta; \hat{g}')|$. For doing so, we focus on $\hat{g}_\theta$ and $\hat{g}'_\theta$. For $0 \leq e < \underline{\xi}/\theta$ we have that $\hat{g}_\theta(e) = \hat{g}(\theta\, e)/\theta = \theta\, e/\theta = \hat{g}'(\theta\, e)/\theta = \hat{g}'_\theta(e)$. For $e \geq \underline{\xi}/\theta$, note that $\hat{g}_\theta(e) = \hat{g}(\theta\, e)/\theta \geq \underline{\xi}/\theta$ and

$$\hat{g}'_\theta(e) \leq \hat{g}_\theta(e) + \varepsilon/\theta = \hat{g}_\theta(e)(1 + \varepsilon/(\theta\hat{g}_\theta(e))) \leq \hat{g}_\theta(e)(1 + \varepsilon/(\theta\hat{g}_\theta(e))) \leq \hat{g}_\theta(e)(1 + \varepsilon/\underline{\xi})\,.$$

Define $\lambda(\varepsilon) = 1 + \varepsilon/\underline{\xi} \geq 1$, and by symmetry we have

$$\lambda(\varepsilon)^{-1}\, \hat{g}_\theta \leq \hat{g}'_\theta \leq \lambda(\varepsilon)\, \hat{g}_\theta\,.$$

The interest payment rate function C is non-increasing, thus

$$C(\lambda(\varepsilon)\,x/\hat{g}_\theta(y)) \leq C(x/\hat{g}'_\theta(y)) \leq C(\lambda(\varepsilon)^{-1}\,x/\hat{g}_\theta)\,, \text{ for } (x,y) \in \mathcal{C}\,,$$

and Assumption 2 applied in the same fashion as in the proof of Lemma 2 gives

$$\lambda(\varepsilon)^{-L_C}\,C(x/\hat{g}_\theta(y)) \leq C(x/\hat{g}'_\theta(y)) \leq \lambda(\varepsilon)^{L_C} C(x/\hat{g}_\theta(y))\,.$$

Using the same arguments as in the proof of Lemma 2,we have

$$f(\lambda(\varepsilon)^{-L_C};\hat{g}_\theta) \leq f(\theta;\hat{g}')/\theta \leq f(\lambda(\varepsilon)^{L_C};\hat{g}_\theta)\,.$$

Recalling that $f(\cdot;h_g)/Id$ is non-increasing and $\lambda_g(\varepsilon) \geq 1$, we proceed as in the proof of Lemma 2, to obtain

$$\lambda(\varepsilon)^{-L_C}\,f(\theta;\hat{g}) \leq f(\theta;\hat{g}') \leq \lambda(\varepsilon)^{L_C}\,f(\theta;\hat{g})\,.$$

Now, we can estimate

$$|f(\theta;\hat{g}) - f(\theta;\hat{g}')| \leq \Big(\lambda_g(\varepsilon)^{L_C} - 1\Big)\; f(\theta;\hat{g}) \leq \Big((1+\varepsilon/\underline{\xi})^{L_C} - 1\Big)\,\overline{\theta}\,\overline{f} \leq \frac{L_C\,\overline{\theta}\,\overline{f}}{\underline{\xi}}\,\varepsilon\,,$$

where we used (28) of Lemma 2 in the second step and $0 < L_C < 1$, which is given by Assumption 2, in the third step. Hence we obtain a uniform upper bound in $\theta \in \Theta$, i.e.

$$\|f(\cdot;g|_{\mathbb{R}_0^+}) - f(\cdot;g'|_{\mathbb{R}_0^+})\|_\infty = \|f(\cdot;\hat{g}) - f(\cdot;\hat{g}')\|_\infty \leq \frac{L_C\,\overline{\theta}\,\overline{f}}{\underline{\xi}}\,\varepsilon\,.$$

This implies that $g \mapsto f(\cdot;g|_{\mathbb{R}_0^+})$ is continuous on $\mathcal{K}_g$.

Finally, consider the transformed best response of the rating agency, which is given by $\mathcal{R}(g(\cdot;\cdot))|_\Xi : \mathcal{K}_f \to \mathcal{K}_g, f \mapsto \mathcal{R}(g(\cdot;f))|_\Xi$, see Proposition 1. We focus on the best response $f \mapsto g(\cdot; f)$ for now, the transformation $\mathcal{R}$ is dealt with later using Lemma 4. Take $f, f' \in \mathcal{K}_f$ with $\|f - f'\|_\infty \le \varepsilon$, for some $\varepsilon > 0$. Consider $e \in [0, f(\underline{\theta}) \wedge f'(\underline{\theta}))$, then $g(e; f) = g(e; f') = e$ and

$$|g(e; f) - g(e; f')| = 0\,, \text{ for } e \in [0, f(\underline{g}) \wedge f'(\underline{g}))\,.$$

Now, $e \in [f(\underline{\theta}) \wedge f'(\underline{\theta}), f(\underline{\theta}) \vee f'(\underline{\theta}))$. Without loss of generality assume $f(\underline{\theta}) < f'(\underline{\theta})$, and thus $f(\underline{\theta}) \le e < f'(\underline{\theta})$. From this we see by (14) $g(e; f') = e$ and $f(\underline{\theta}) \le g(e; f) \le e$, and since $e < f'(\underline{\theta})$ as well as we have $|f(\underline{\theta}) - f'(\underline{\theta})| \le \|f - f'\|_\infty = \varepsilon$, we have

$$|g(e; f) - g(e; f')| \le \varepsilon\,, \text{ for } e \in [f(\underline{\theta}) \wedge f'(\underline{\theta}), f(\underline{\theta}) \vee f'(\underline{\theta}))\,.$$

Consider $e \in [f(\underline{\theta}) \vee f'(\underline{\theta}), e(\varepsilon, f, f'))$, where $e(\varepsilon, f, f') = f(\underline{\theta}) \vee f'(\underline{\theta}) + L_f\, \varepsilon^{1/2}$. Then by the uniform Lipschitz continuity of $\mathcal{K}_f$ with Lipschitz constant L_f, see Lemma 8, it follows that $f(\underline{\theta}) \wedge f'(\underline{\theta}) \le g(e; f), g(e; f') \le f(\underline{\theta}) \vee f'(\underline{\theta}) + L_f\, \varepsilon^{1/2}$, and thus

$$|g(e; f) - g(e; f')| \le \varepsilon + L_f\, \varepsilon^{1/2}\,, \text{ for } e[f(\underline{\theta}) \vee f'(\underline{\theta}), e(\varepsilon, f, f'))\,.$$

Consider $e \ge e(\varepsilon, f, f')$. Without loss of generality assume $f^{-1}(e) \le f'^{-1}(e)$, note that by $f, f' \in \mathcal{K}_f$ both functions are continuous and strictly increasing with minimum slope l_f and thus their respective inverse functions exist and are well-defined. Using the definition of $g(\cdot, f)$ in (14) we can write

$$
\begin{aligned}
g(e; f') =& \frac{\int_{\underline{\theta}}^{f'^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \\
=& \frac{\int_{\underline{\theta}}^{f^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} + \frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \\
=& \frac{\int_{\underline{\theta}}^{f^{-1}(e)} f(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} + \frac{\int_{\underline{\theta}}^{f^{-1}(e)} \left(f'(e') - f(e')\right)\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} + \frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \\
=& \frac{\int_{\underline{\theta}}^{f^{-1}(e)} f(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f^{-1}(e)} \phi(e')\,\mathrm{d}e'}\,\frac{\int_{\underline{\theta}}^{f^{-1}(e)} \phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \\
&+ \frac{\int_{\underline{\theta}}^{f^{-1}(e)} \left(f'(e') - f(e')\right)\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} + \frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \\
=& g(e; f) - g(e; f)\,\frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \\
&+ \frac{\int_{\underline{\theta}}^{f^{-1}(e)} \left(f'(e') - f(e')\right)\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} + \frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'}\,.
\end{aligned}
$$

And hence

$$
\begin{aligned}
|g(e; f) - g(e; f')| \leq& g(e; f)\,\frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} + \frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \\
&+ \frac{\int_{\underline{\theta}}^{f^{-1}(e)} \left|f'(e') - f(e')\right|\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'}\,. \qquad (34)
\end{aligned}
$$

All three expressions on the right hand side need to become small for $\varepsilon \searrow 0$. Observe that $0 \leq f'^{-1}(e) - f^{-1}(e) \leq l_f^{-1}\varepsilon$, where the first inequality follows by assumption and the second by the fact that f, f' are strictly increasing with a minimum slope of $l_f > 0$ and $\|f - f'\|_\infty \leq \varepsilon$. Also, f, f' are bounded by $\overline{\theta}\,\overline{f}$ by (28), and so are $g(\cdot; f)$ and $g(\cdot; f')$. Since *phi* is bounded away from zero and bounded from above by assumption, we see that

$$\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e' \ge \int_{\underline{\theta}}^{f'^{-1}(e(\varepsilon,f,f'))} \underline{\phi}\,\mathrm{d}e' \ge \int_{\underline{\theta}}^{f'^{-1}(f'(\underline{\theta})+L_f\,\varepsilon)} \underline{\phi}\,\mathrm{d}e'$$
$$= \underline{\phi}(f'^{-1}(f'(\underline{\theta}) + L_f\,\varepsilon^{1/2}) - \underline{\theta}) \ge \underline{\phi}\,L_f\,\varepsilon^{1/2}/L_f = \underline{\phi}\,\varepsilon^{1/2}.$$

Using this, we can bound the first expression on the right hand side of (34)

$$g(e;f)\,\frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \le \overline{\theta}\,\overline{f}\,\frac{\bar{\phi}\,l_f^{-1}\,\varepsilon}{\underline{\phi}\,\varepsilon^{1/2}} = \frac{\overline{\theta}\,\overline{f}\,\bar{\phi}}{\underline{\phi}\,l_f}\,\varepsilon^{1/2}\,.$$

For the second expression in (34) the same arguments apply, now aimed at f rather than $g(\cdot;f)$, and

$$\frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} f'(e')\,\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \le \overline{\theta}\,\overline{f}\,\frac{\int_{f^{-1}(e)}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \le \frac{\overline{\theta}\,\overline{f}\,\overline{\phi}}{\underline{\phi}\,l_f}\,\varepsilon^{1/2}\,.$$

The third expression in (34) can be estimated as follows

$$\frac{\int_{\underline{\theta}}^{f^{-1}(e)} |f'(e') - f(e')|\;\phi(e')\,\mathrm{d}e'}{\int_{\underline{\theta}}^{f'^{-1}(e)} \phi(e')\,\mathrm{d}e'} \le \frac{\|f - f'\|_\infty}{\underline{\phi}\,\varepsilon^{1/2}} \le \frac{\varepsilon^{1/2}}{\underline{\phi}}\,.$$

Adding up the three expressions we obtain as bound

$$|g(e;f) - g(e;f')| \le \frac{2\,\overline{\theta}\,\overline{f}\,\overline{\phi} + l_f}{\underline{\phi}\,l_f}\,\varepsilon^{1/2}\,,\ \text{ for } e \in [e(\varepsilon,f,f'),\infty)\,.$$

From there the uniform bound can be expressed as follows

$$\|g(\cdot;f) - g(\cdot;f')\|_\infty \le \varepsilon + \frac{2\,\overline{\theta}\,\overline{f}\,\overline{\phi} + l_f + \underline{\phi}\,l_f\,L_f}{\underline{\phi}\,l_f}\,\varepsilon^{1/2}\,. \tag{35}$$

This implies that $f \mapsto g(\cdot;f)$ is continuous. From Lemma 4 and noting that the restriction to Ξ does no harm, the continuity of $\mathcal{R}(g(\cdot;f))|_\Xi$ follows. $\square$

C ODE Characterization of Best Responses and potential Equilibria

Proposition 4 gives an ODE characterization of a equilibrium in case a specified condition holds. In the following, this result is derived. First, the best responses of both, rating agency and firm, are characterized by solutions to ODEs systems. Based on these, the equilibrium characterization is established.

C.1 Best Response of Rating Agency ODE

Proposition 5. *Given a strategy $f \in \mathcal{A}_f^\star$ defined in* (23). *The best response $g(\cdot; f)$ given in* (14) *of Proposition 1 and its transformation $\mathcal{R}(g(\cdot; f))$ defined in* (15) *can be characterized as follows*

$$g(\cdot; f) = \hat{g}_f \circ f^{-1} \;\; \textit{and} \;\; \mathcal{R}(g(\cdot; f)) = \tilde{g}_f \circ f^{-1}\,, \;\; \textit{on}\; f(\Theta)\,, \tag{36}$$

where $\hat{g}_f$ and $\tilde{g}_f$ have initial values

$$\hat{g}_f(\underline{\theta}) = \tilde{\theta}_f(\underline{\theta}) = f(\underline{\theta})\,, \tag{37}$$

and derivative $\hat{g}_f'$ and $\tilde{g}_f'$ that satisfy

$$\hat{g}_f' = \frac{\phi}{\Phi}\,(f - \hat{g}_f) \;\; \textit{and} \;\; \tilde{g}_f' = f'\,\frac{\tilde{g}_f}{f}\,\mathbf{1}_{\tilde{g}_f < \hat{g}_f} + \min\left(\hat{g}_f', f'\,\frac{\tilde{g}_f}{f}\right)\mathbf{1}_{\tilde{g}_f = \hat{g}_f}\,, \tag{38}$$

Lebesgue almost everywhere on $(\underline{\theta}, \overline{\theta})$, where $\Phi(g) = \int_{\underline{\theta}}^{\theta} \phi(t)\,\mathrm{d}t$, $\theta \in \Theta$.

Corollary 1. *Suppose that $f \in \mathcal{A}_f^\star$ and ϕ are both continuously differentiable, then the function pair $(\hat{g}_f, \tilde{g}_f)$ given in Proposition 5 is the solution to the ODE*

$$(\hat{g}_f', \tilde{g}_f') = \left(\frac{\phi}{\Phi}\,(f - \hat{g}_f)\,, f'\,\frac{\tilde{g}_f}{f}\,\mathbf{1}_{\tilde{g}_f < \hat{g}_f} + \min\left(\hat{g}_f', f'\,\frac{\tilde{g}_f}{f}\right)\mathbf{1}_{\tilde{g}_f = \hat{h}_f}\right)\,, \tag{39}$$

on $(\underline{\theta}, \overline{\theta})$ *with initial conditions*

$$\left(\hat{g}_f(\underline{\theta}), \tilde{g}_f(\underline{\theta})\right) = \left(f(\underline{\theta}), f(\underline{\theta})\right) \quad \textit{and} \quad \left(\hat{g}'_f(\underline{\theta}), \tilde{g}'_f(\underline{\theta})\right) = \left(\frac{1}{2}\, f'(\underline{\theta}), \frac{1}{2}\, f'(\underline{\theta})\right) . \tag{40}$$

Proof of Proposition 5. First, define $\hat{g}_f : \Theta \to \mathbb{R}^+$ by $\hat{g}_f(\theta) = g(f(\theta); f)$, for $\theta \in \Theta$, where $g(\cdot; f)$ is given in (14) of Proposition 1. Then $g(\cdot; f) = \hat{g}_f \circ f^{-1}$ as in (36) and furthermore

$$\hat{g}_f(\theta) = \frac{1}{\Phi(\theta)} \int_{\underline{\theta}}^{\theta} f(t)\, \phi(t)\, \mathrm{d}t\,, \ \text{for } \theta \in (\underline{\theta}, \overline{\theta}]\,,$$

and $\hat{g}_f(\underline{\theta}) = g(f(\underline{\theta}); f) = f(\underline{\theta})$, where the latter proves the first part of (37). Noting that $g(\cdot; f)$ is the quotient of two absolutely continuous and strictly positive functions on $(\underline{\theta}, \bar{\theta}]$, we see that $g(\cdot; f)$ is Lebesgue almost everywhere differentiable with derivative $\hat{g}'_f$ that satisfies according to the quotient rule

$$\hat{g}'_f(\theta) = \frac{1}{\Phi(\theta)^2} \left(f(\theta)\, \phi(\theta)\, \Phi(\theta) - \int_{\underline{\theta}}^{\theta} f(t)\, \phi(t)\, \mathrm{d}t\, \phi(\theta) \right) = \frac{\phi(\theta)}{\Phi(\theta)} \left(f(\theta) - \hat{g}_f(\theta)\right)\,,$$

for Lebesgue almost every $\theta \in (\underline{\theta}, \overline{\theta})$, proving the first part of (38). Define $\tilde{g}_f$ by

$$\tilde{g}_f(\theta) = \mathcal{R}(g(\cdot; f))(f(\theta))\,, \ \text{for } \theta \in \Theta\,,$$

then $\mathcal{R}(g(\cdot; f)) = \tilde{g}_f \circ f^{-1}$ on $f(\Theta)$, as claimed in (36). Note $g(z; f)/z = 1$, for $0 < z \le f(\underline{\theta})$, $g(z; f)/z \le 1$, for $z > 0$, and $f \in \mathcal{A}^\star_f$ is continuous and strictly increasing, to see that

$$\begin{aligned}\tilde{g}_f(\theta) = \mathcal{R}(g(\cdot; f))(f(\theta)) &= e \inf_{0<z\le e} \frac{g(z; f)}{z}\bigg|_{e=f(\theta)} = f(\theta) \inf_{f(\underline{\theta})\le z\le f(\theta)} \frac{g(z; f)}{z} \\ &= f(\theta) \inf_{\underline{\theta}\le\theta'\le\theta} \frac{g(f(\theta'); f)}{f(\theta')} = f(\theta) \inf_{\underline{\theta}\le\theta'\le\theta} \frac{\hat{g}_f(\theta')}{f(\theta')}\,, \ \text{for } \theta \in \Theta\,.\end{aligned}$$

The initial value is given by $\tilde{g}_f(\underline{\theta}) = f(\underline{\theta})\, \hat{g}_f(\underline{\theta})/f(\underline{\theta}) = \hat{g}_f(\underline{\theta}) = f(\underline{\theta})$ establishing the second part of (37). Further, from Lemma 1 it follows that $\tilde{g}_f$ is continuous and nondecreasing, bounded by f, as well as, that $\tilde{g}_f/f$ is nonincreasing. For $\theta \geq \theta' \in \Theta$, write

$$\begin{aligned} 0 \leq & \tilde{g}_f(\theta) - \tilde{g}_f(\theta') = f(\theta) \inf_{\underline{\theta} \leq z \leq \theta} \frac{\hat{g}_f(z)}{f(z)} - f(\theta') \inf_{\underline{\theta} \leq z \leq \theta'} \frac{\hat{g}_f(z)}{f(z)} \\ & \leq (f(\theta) - f(\theta')) \inf_{\underline{\theta} \leq z \leq \theta} \frac{\hat{g}_f(z)}{f(z)} \leq (f(\theta) - f(\theta')) \leq L_f \left| \theta - \theta' \right|, \end{aligned}$$

by Lemma 2, for $L_f = \underline{f} > 0$. Thus, $\tilde{g}_f$ is Lipschitz continuous and by this has a derivative $\tilde{g}_f'$ Lebesgue almost everywhere on Θ. By the same rational f has a derivative f' Lebesgue almost everywhere on Θ. Since $\tilde{g}_f/f$ is nonincreasing and $f > 0$, we have Lebesgue almost everywhere on Θ

$$0 \geq (\tilde{g}_f/f)' = \frac{\tilde{g}_f' \, f - \tilde{g}_f \, f'}{f^2} \quad \Longleftrightarrow \quad \tilde{g}_f' \leq f' \, \frac{\tilde{g}_f}{f} \, .$$

Denote by $E_f \subseteq \Theta$ the set where $\hat{g}_f$ equals $\tilde{g}_f$, i.e. $E_f = \{\theta \in \Theta : \hat{g}_f(\theta) = \tilde{g}_f(\theta)\}$. Since $\hat{g}_f$ and $\tilde{g}_f$ are both continuous and Θ is bounded, E_f is compact. On $\mathring{E}_f = E_f \setminus \partial E_f$ we have Lebesgue almost everywhere that $\tilde{g}_f' = \hat{g}_f'$. Now, the boundary ∂E_f has Lebesgue measure 0 and $\tilde{g}_f' \leq f' \, \frac{\tilde{g}_f}{f}$ Lebesgue almost everywhere on Θ, hence we have Lebesgue almost everywhere on E_f

$$\tilde{g}_f' = \min\left(\hat{g}_f', f' \, \frac{\tilde{g}_f}{f} \right) .$$

Next, consider $(\underline{\theta}, \bar{\theta}) \setminus E_f$, which is open. Take $\theta \in (\underline{\theta}, \bar{\theta}) \setminus E_f$, then we find an $\varepsilon > 0$ such that $B_\varepsilon(\theta) = \{\theta' \in \mathbb{R} : |\theta - \theta'| < \varepsilon\} \in (\underline{\theta}, \bar{\theta}) \setminus E_f$, i.e. $\tilde{g}_f < \hat{g}_f$ on $B_\varepsilon(\theta)$. Then

$$\tilde{g}_f(\theta) = f(\theta) \inf_{\underline{\theta} \leq \theta' \leq \theta} \frac{\hat{g}_f(\theta')}{f(\theta')} = f(\theta) \, \frac{\hat{g}_f(\theta^\star)}{f(\theta^\star)}$$

for some $\theta^\star < \theta$, since $\tilde{g}_f(\theta)/f(\theta) < \hat{g}_f(\theta)/f(\theta)$ and $\theta \mapsto \inf_{\underline{\theta} \leq \theta' \leq \theta} \hat{g}_f(\theta)/f(\theta)$ is continuous, where the latter follows from the continuity of $\hat{g}_f$ and f. Moreover, the above equality extend by the latter mentioned continuity to $[\theta, \theta+\varepsilon^\star)$, for some $\varepsilon^\star > 0$. And thus for $\tilde{\varepsilon} = \varepsilon \wedge \varepsilon^\star$, we have

$$\tilde{g}_f(\theta') = f(\theta') \, \frac{\hat{g}_f(\theta^\star)}{f(\theta^\star)} \,, \text{ for } \theta' \in B_{\tilde{\varepsilon}}(\theta) \,.$$

Now, f and $\tilde{h}_f$ are absolutely continuous and their derivatives satisfy Lebesgue almost surely on $B_{\tilde{\varepsilon}}(\theta)$

$$\tilde{g}_f' = f' \, \frac{\hat{g}_f(\theta^\star)}{f(\theta^\star)} = f' \, \frac{\tilde{g}_f}{f} \,.$$

Putting the pieces together gives the second part of (38). □

Proof of Corollary 1. Since ϕ and Φ are continuous, the results in (38) hold for all $\theta \in (\underline{\theta}, \bar{\theta})$, what then also specifies an ODE. The initial values for the derivative of $\hat{g}_f$ is obtained using L'Hospital rule

$$\begin{aligned}
\hat{g}_f'(\underline{\theta}) &= \lim_{\theta \searrow \underline{\theta}} \hat{g}_f'(\theta) = \lim_{\theta \searrow \underline{\theta}} \frac{\phi(\theta)}{\Phi(\theta)} \left(f(\theta) - \hat{g}_f(\theta)\right) = \phi(\underline{\theta}) \lim_{\theta \searrow \underline{\theta}} \frac{f(\theta) - \hat{g}_f(\theta)}{\Phi(\theta)} \\
&= \phi(\underline{\theta}) \lim_{\theta \searrow \underline{\theta}} \frac{f'(\theta) - \hat{g}_f'(\theta)}{\phi(\theta)} = \phi(\underline{\theta}) \frac{f'(\underline{\theta}) - \hat{g}_f'(\underline{\theta})}{\phi(\underline{\theta})} = f'(\underline{\theta}) - \hat{g}_f'(\underline{\theta}) \,,
\end{aligned}$$

and hence $\hat{g}_f'(\underline{\theta}) = f'(\underline{\theta})/2$ as claimed. Then

$$\begin{aligned}
\tilde{g}_f'(\theta) &= \lim_{\theta \searrow \underline{\theta}} f'(\theta) \, \frac{\tilde{g}_f(\theta)}{f(\theta)} \, \mathbf{1}_{\tilde{g}_f(\theta) < \hat{g}_f(\theta)} \min\left(\hat{g}_f'(\theta), f'(\theta) \, \frac{\tilde{g}_f(\theta)}{f(\theta)}\right) \mathbf{1}_{\tilde{g}_f(\theta) = \hat{g}_f(\theta)} \\
&= f'(\underline{\theta}) \, \frac{\tilde{g}_f(\underline{\theta})}{f(\underline{\theta})} \lim_{\theta \searrow \underline{\theta}} \mathbf{1}_{\tilde{g}_f(\theta) < \hat{g}_f(\theta)} + \min\left(\hat{g}_f'(\underline{\theta}), f'(\underline{\theta}) \, \frac{\tilde{g}_f(\underline{\theta})}{f(\underline{\theta})}\right) \lim_{\theta \searrow \underline{\theta}} \mathbf{1}_{\tilde{g}_f(\theta) = \hat{g}_f(\theta)} \\
&= f'(\underline{\theta}) \lim_{\theta \searrow \underline{\theta}} \mathbf{1}_{\tilde{g}_f(\theta) < \hat{g}_f(\theta)} + \min\left(f'(\underline{\theta})/2, f'(\underline{\theta})\right) \lim_{\theta \searrow \underline{\theta}} \mathbf{1}_{\tilde{g}_f(\theta) = \hat{g}_f(\theta)} \,.
\end{aligned}$$

Observe that $\hat{g}_f(\underline{\theta}) = f(\underline{\theta})$ and that the slope $\hat{g}_f'(\underline{\theta}) = f'(\underline{\theta})/2$ is strictly smaller than $f'(\underline{\theta}) > 0$. Therefore, $\hat{g}_f/f$ is strictly decreasing in a neighborhood of $\underline{\theta}$ and thus $\hat{g}_f = \tilde{g}_f$ on $[\underline{\theta}, \underline{\theta} + \varepsilon]$ for some $\varepsilon > 0$. Accordingly, $\lim_{\theta \searrow \underline{\theta}} \mathbf{1}_{\tilde{g}_f(\theta) < \hat{g}_f(\theta)} = 0$ and

$\lim_{\theta\searrow\underline{\theta}} \mathbf{1}_{\tilde{g}_f(\theta)=\hat{g}_f(\theta)} = 1$. From there we conclude

$$\tilde{g}_f'(\theta) = \min\left(f'(\underline{\theta})/2, f'(\underline{\theta})\right) = f'(\underline{\theta})/2\,,$$

finishing the proof. $\square$

C.2 Best Response of Firm ODE

For a given $g \in \mathcal{A}_g^\star$ and $\theta \in \Theta$, consider the free boundary value problem given in (17-21) to characterize the value function $v(\cdot, \cdot; \theta, g)$. The boundary $\partial C^{(\theta,g)}$ of the continuation region $C^{(\theta,g)}$ can be described by a boundary function $b(\cdot, \theta; g)$ with the following properties.

Lemma 3. *For given $g \in \mathcal{A}_g^\star$ and $\theta \in \Theta$, the boundary $\partial C^{(\theta,g)}$ is characterized by a function $b : [0, f(\theta; g)] \times \Theta \times \mathcal{A}_g^\star) : \mathbb{R}_0^+, (e, \theta; g) \mapsto b(e, \theta; g)$, i.e.*

$$\partial C^{(\theta,g)} = \{(b(e, \theta; g), e) : 0 \le e \le f(\theta; g)\}\,,$$

that is non-decreasing and continuous with terminal value $b(f(\theta; g), \theta; g) = f(\theta; g)$. The restriction of $b(\cdot, \theta; g)$ to $[\underline{\xi}, f(\theta; g)]$, with $\underline{\xi} = \underline{\theta}\,\underline{f}$, is Lipschitz continuous with constant $L_b = \bar{f}/\underline{\xi}$.

Proof of Lemma 3. Lemma 6 part 4 implies that $b(\cdot, \theta; g)$ is non-decreasing. Part 5 implies that the slope in e is bounded by $f(\theta; g)/e$, which also implies continuity. The terminal value $b(f(\theta; g), \theta; g) = f(\theta; g)$ follows from Lemma 7. The continuity of b is uniformly for all $\theta \in \Theta$ with maximum slope $\bar{f}/\underline{\xi}$, where $\bar{f}$ is given in Lemma 2 and $\underline{\xi} = \underline{\theta}\,\underline{f}$, see discussion around (33). $\square$

Proposition 6. *For a given continuously differentiable strategy $g \in \mathcal{A}_g^\star$ suppose that Assumption 3 holds. Then the firm's best response $f(\cdot; g)$ given in Proposition 2* *satisfies*

$$f'(\theta;g) = \frac{(1+\eta)\,\sigma^2}{2\,(r-\mu)} \frac{f(\theta;g)^2/\theta^2}{C(f(\theta;g)/g(f(\theta;g))) - f(\theta;g)/\theta} \frac{1}{1-\frac{\partial b}{\partial e}(f(\theta;g),\theta;g)}, \tag{41}$$

for $\theta \in (\underline{\theta}, \bar{\theta})$*, where* $\eta = \frac{\mu - \frac{1}{2}\sigma^2}{\sigma^2} + \sqrt{\left(\frac{\mu-\frac{1}{2}\sigma^2}{\sigma^2}\right)^2 + \frac{2\,r}{\sigma^2}} > 0$*, and the partial derivative of the boundary describing function* b *with respect to* e *in* $(f(\theta;g),\theta)$ *is a function of* $f(\theta), g(f(\theta))$*, and* $g'(f(\theta))$*, i.e.*

$$\frac{\partial b}{\partial e}(f(\theta;g),\theta;g) = h(f(\theta), g(f(\theta)), g'(f(\theta)), \theta)\,, \tag{42}$$

for $\theta \in (\underline{\theta}, \bar{\theta})$*, and some function* $h(\cdot,\cdot,\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \times \Theta \to \mathbb{R}_0^+$.

Proof of Proposition 6. In the following, the value function $v(\cdot,\cdot;\cdot,g)$, $g \in \mathcal{A}_g^\star$, is frequently differentiated on the boundary $\partial C^{(\theta,g)}$. Without stating this explicitly, we assume that the differential is calculated in the interior of $C^{(\theta,g)}$ and by continuity of $v(\cdot,\cdot;\dot{;}g)$, the limit to the boundary is taken. In order to show the claim, we proceed in four steps. First, we differentiate the smooth fit condition in (21) with respect to θ to obtain

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\partial v}{\partial d}(b(e,\theta;g),e,\theta;g) \\ &= \frac{\partial^2 v}{\partial d^2}(b(e,\theta;g),e,\theta;g)\,\frac{\partial b}{\partial \theta}(e,\theta;g) + \frac{\partial^2 v}{\partial \theta \partial d}(b(e,\theta;g),e,\theta;g)\,, \end{aligned} \tag{43}$$

for $e \le f(\theta;g)$. In the second step, we identify $\frac{\partial^2 v}{\partial d^2}$ at the boundary using (18-20). In the third step, we characterize $\frac{\partial v}{\partial \theta}$ in order to calculate $\frac{\partial^2 v}{\partial \theta \partial d}$. In the fourth and final step, we use results of the third step and the normal reflection condition in (21) to determine $\frac{\partial b}{\partial \theta}$.

For the second step observe that $v = \frac{\partial v}{\partial d} = 0$ on $\partial C^{(\theta,g)}$. From this and (18-20) it follows that

$$\frac{1}{2}\,\sigma^2\, b(e,\theta;g)^2\,\frac{\partial^2 v}{\partial d^2}(b(e,\theta;g),e,\theta;g)+\frac{b(e,\theta;g)}{\theta}-C(b(e,\theta;g)/g(e))=0\,,$$

or, equivalently,

$$\frac{\partial^2 v}{\partial d^2}(b(e,\theta;g),e,\theta;g)=2\,\frac{C(b(e,\theta;g)/g(e))-\frac{b(e,\theta;g)}{\theta}}{\sigma^2\, b(e,\theta;g)^2}\,, \tag{44}$$

for $e\le f(\theta;g)$.

Now, the third step is taken. For $\theta\in(\underline{\theta},\bar{\theta})$ denote by u the differential of v with respect to θ, which by Assumption 3 exists is continuous in θ. The continuity of b according to Assumption 3, the fact that the continuation region $C^{(\theta,g)}$ is open in $\mathcal{C}$, and from equations (18-20) we obtain

$$\mu\, d\,\frac{\partial u}{\partial d}(d,e;\theta,g)+\frac{1}{2}\sigma^2\, d^2\,\frac{\partial^2 u}{\partial d^2}(d,e;\theta,g)-\frac{d}{\theta^2}-r\,u(d,e;\theta,g)=0\,,$$

for $(d,e)\in C^{(\theta,g)}$. Similar reasoning implies for (d,e) in the interior of $\mathcal{C}\setminus C^{(\theta,g)}$ that

$$u(d,e;\theta,g)=0\,,$$

which extends to all $(d,e)\in\mathcal{C}\setminus C^{(\theta,g)}$, i.e. also to $\partial C^{(\theta,g)}$, since $v=0$ on the boundary $\partial C^{(\theta,g)}$, and computing the derivative in θ gives

$$\frac{\partial v}{\partial d}(b(e,\theta;g),e;\theta,g)\,\frac{\partial b}{\partial \theta}(e,\theta;g)+\frac{\partial v}{\partial \theta}(b(e,\theta;g),e;\theta,g)=0\,,$$

and recalling that $\frac{\partial v}{\partial d}=0$ on $\partial C^{(\theta,g)}$ by the smooth fit condition in (21) and $\frac{\partial v}{\partial \theta}=u$ by definition. This second order ODE in d is not depending explicitly on e. Therefore, we can interpret e as well as θ as fixed parameters, and u is in its general form given by

$$u(d,e;\theta,g) = -\frac{d}{\theta^2(r-\mu)} + d^{-\eta}\,L(e,\theta,g) + d^{-\tilde{\eta}}\,\tilde{L}(e,\theta,g)\,,$$

for $(d,e) \in C^{(\theta,g)}$, or, equivalently, for $d \geq b(e,\theta;g)$ in case $e \leq f(\theta;g)$ and $d \geq e$ in case $e > f(\theta;g)$, where $L(e,\theta,g)$ and $\tilde{L}(e,\theta,g)$ are constants taking values in $\mathbb{R}$, and

$$\eta = \frac{\mu - \frac{1}{2}\sigma^2}{\sigma^2} + \sqrt{\left(\frac{\mu - \frac{1}{2}\sigma^2}{\sigma^2}\right)^2 + \frac{2\,r}{\sigma^2}} \text{ and } \tilde{\eta} = \frac{\mu - \frac{1}{2}\sigma^2}{\sigma^2} - \sqrt{\left(\frac{\mu - \frac{1}{2}\sigma^2}{\sigma^2}\right)^2 + \frac{2\,r}{\sigma^2}}\,.$$

From $r > 0$, we see directly that $\eta > 0$ and $\tilde{\eta} < 0$. It can also be shown that $\tilde{\eta} < -1$, since $\mu < r$, and hence $x^{-\tilde{\eta}}\,\tilde{L}(e,\theta,g)$ becomes the dominating expression of u for large d. Observe that the value function v for large d is approximately $\frac{d}{\theta(r-\mu)}$, and thus its partial derivative in θ is approximately $-\frac{d}{\theta^2(r-\mu)}$. Accordingly, the weight of the otherwise dominating expression $d^{-\tilde{\eta}}$ has to be zero, i.e. $\tilde{L}(e,\theta,g) = 0$, and hence

$$u(d,e;\theta,g) = -\frac{d}{\theta^2(r-\mu)} + d^{-\eta}\,L(e,\theta,g)\,, \text{ for } (d,e) \in C^{(\theta,g)}\,.$$

Fixing θ and focusing on $\partial C^{(\theta,g)}$, we obtain $L(e,\theta,g)$ in terms of the boundary defining function $b(\cdot,\theta;g)$. On the boundary, i.e. $d = b(e,\theta;g)$ with $e \leq f(\theta;g)$, we have that $u(b(e,\theta;g),e,\theta;g) = 0$. Thus

$$L(e,\theta,g) = \frac{b(e,\theta;g)^{1+\eta}}{\theta^2(r-\mu)}\,, \text{ for } e \leq f(\theta;g)\,,$$

yielding for $e \leq f(\theta;g)$ that

$$u(d,e;\theta,g) = -\frac{d}{\theta^2(r-\mu)}\left(1 - \left[\frac{d}{b(e,\theta;g)}\right]^{-(1+\eta)}\right)\mathbf{1}_{d \geq b(e,\theta;g)}\,. \tag{45}$$

The partial derivative with respect to d is then

$$\frac{\partial u}{\partial d}(d,e;\theta,g) = -\frac{1}{\theta^2(r-\mu)}\left(1+\eta\left[\frac{d}{b(e,\theta;g)}\right]^{-(1+\eta)}\right)\mathbf{1}_{d\geq b(e,\theta;g)}\,,$$

for $e \leq f(\theta;g)$. In particular, we obtain for $d = b(e,\theta;g)$ that

$$\frac{\partial^2 v}{\partial\theta\partial d}(b(e,\theta;g),e;\theta,g) = \frac{\partial u}{\partial d}(b(e,\theta;g),e;\theta,g) = -\frac{1+\eta}{\theta^2(r-\mu)}\,, \tag{46}$$

where the derivative in $(b(e,\theta;g),e) \in \partial C^{(\theta,g)}$ is understood in the limit from the interior of $C^{(\theta,g)}$ and the interchange of the order of differentiation follows from Assumption 3.

For the fourth step, note that the boundary function satisfies $f(\theta;g) = b(f(\theta;g),\theta;g)$, and differentiating this expression with respect to θ gives

$$f'(\theta;g) = \frac{\partial b}{\partial e}(f(\theta;g),\theta;g)\,f'(\theta;g) + \frac{\partial b}{\partial\theta}(f(\theta;g),\theta;g)\,,$$

where $f'(\cdot;g)$ denotes $\frac{\partial f}{\partial\theta}(\cdot;g)$. Now, solve for $\frac{\partial b}{\partial\theta}$ to obtain

$$\frac{\partial b}{\partial\theta}(f(\theta;g),\theta;g) = f'(\theta;g)\left(1-\frac{\partial b}{\partial e}(f(\theta;g),\theta;g)\right)\,. \tag{47}$$

Finally, we set $e = f(\theta;g)$, hence $b(f(\theta;g),\theta;g) = f(\theta;g)$, and plug (44), (46) and (47) in (43), to see that

$$0 = 2\,\frac{C(f(\theta;g)/g(f(\theta;g))) - \frac{f(\theta;g)}{\theta}}{\sigma^2\,f(\theta;g)^2}\,f'(\theta;g)\left(1-\frac{\partial b}{\partial e}(f(\theta;g),\theta;g)\right) - \frac{1+\eta}{\theta^2(r-\mu)}\,,$$

and solving for $f'(\theta;g)$ gives (41).

The partial derivative $\frac{\partial b}{\partial e}(f(\theta;g),\theta;g)$ in (41) is now analyzed. For doing so, define the function $\hat{v} : \mathbb{R}^+ \times \Theta \times \mathbb{R}^+ \to \mathbb{R}$, $(d; \theta, g) \mapsto \hat{v}(d; \theta, g)$ by

$$\hat{v}(d; \theta, g) = \sup_{\tau \in \mathcal{T}_d} \mathbb{E}_d \left[\int_0^\tau e^{-r\,t} \, (d_t/\theta - C(d_t/g))\, \mathrm{d}t \right] ,$$

where $\mathcal{T}_d$ is the set of all stopping times with respect to the information generated by D with starting value d and $\mathbb{E}_d$ is the corresponding expectation. In contrast to the function v defined in (16), the direct dependence on the minimum observed cash flow E is eliminated. Instead, an optimal stopping problem in the observed cash flow D is given parameterized by $g \in \mathbb{R}^+$ and $\theta \in \Theta$. However, for $e \leq f(\theta; g)$, the connection between the coordinates d and e of v is dissolved and we have

$$v(d, e; \theta, g) = \hat{v}(d; \theta, g(e))\,, \text{ for } (d, e) \in \mathcal{C}, e \leq f(\theta; g), \theta \in \Theta\,.$$

Accordingly, we can use $\hat{v}$ instead of v in order to characterize the partial derivative $\frac{\partial b}{\partial e}(f(\theta; g), \theta; g)$. Using Assumption 3 and the differentiability of g, we can write

$$\begin{aligned}
\frac{\partial b}{\partial e}(f(\theta; g), \theta; g) &= \lim_{\Delta e \searrow 0} \frac{1}{\Delta e} \left(b(f(\theta; g), \theta; g) - b(f(\theta; g) - \Delta e, \theta; g)\right) \\
&= \lim_{\Delta e \searrow 0} \frac{1}{\Delta e} \left(\sup\{d \geq e = f(\theta; g) : v(d, e; \theta, g) = 0\}\right. \\
&\qquad\qquad\qquad \left. - \sup\{d \geq e = f(\theta; g) - \Delta e : v(d, e; \theta, g) = 0\}\right) \\
&= \lim_{\Delta e \searrow 0} \frac{1}{\Delta e} \left(\sup\{d > 0 : \hat{v}(d; \theta, g(f(\theta; g))) = 0\}\right. \\
&\qquad\qquad\qquad \left. - \sup\{d > 0 : \hat{v}(d; \theta, g(f(\theta; g) - \Delta e)) = 0\}\right) .
\end{aligned}$$

Noting that $f(\theta; g) = b(f(\theta; g), \theta; g) = \sup\{d > 0 : \hat{v}(d; \theta, g(f(\theta; g))) = 0\}$, gives

$$\begin{aligned}
&\frac{\partial b}{\partial e}(f(\theta; g), \theta; g) \\
&\quad = \lim_{\Delta e \searrow 0} \frac{1}{\Delta e} \left(f(\theta; g) - \sup\{d > 0 : \hat{v}(d; \theta, g(f(\theta; g)) - g'(f(\theta; g))\Delta e) = 0\}\right). \\
&\quad =: h(f(\theta; g), g(f(\theta; g)), g'(f(\theta; g)), \theta)\,,
\end{aligned}$$

defining the function h and thus proving (42). $\square$

C.3 ODE Characterization of Equilibrium

Proposition 7. *Given the setting of Proposition 3, denote by $(f^\star, g^\star)$ a fixed point of T. Suppose that that $f^\star$, $g^\star$, as well as ϕ are continuously differentiable and that Assumption 3 holds for $g^\star$. Then $(f, \tilde{g}, \hat{g}) = (f^\star, g^\star \circ f^\star, g(\cdot; f^\star) \circ f^\star)$ satisfies*

$$\begin{pmatrix} f'(\theta) \\ \tilde{g}'(\theta) \\ \hat{g}'(\theta) \end{pmatrix} = \begin{pmatrix} \frac{(1+\eta)\,\sigma^2}{2\,(r-\mu)}\,\frac{f(\theta)^2/\theta^2}{C(f(\theta)/\tilde{g}(\theta))-f(\theta)/\theta}\,\frac{1}{1-\frac{\partial b}{\partial e}(f(\theta),\theta;g^\star)} \\ f'(\theta)\,\frac{\tilde{g}(\theta)}{f(\theta)}\,\mathbf{1}_{\tilde{g}(\theta)<\hat{g}(\theta)} + \min\left(\hat{g}'(\theta), f'(\theta)\,\frac{\tilde{g}(\theta)}{f(\theta)}\right)\mathbf{1}_{\tilde{g}(\theta)=\hat{g}(\theta)} \\ \frac{\phi(\theta)}{\Phi(\theta)}\,(f(\theta)-\hat{g}(\theta)) \end{pmatrix}, \tag{48}$$

on $(\underline{\theta}, \overline{\theta})$ with initial condition

$$\begin{pmatrix} f(\underline{\theta}) \\ \tilde{g}(\underline{\theta}) \\ \hat{g}(\underline{\theta}) \end{pmatrix} = \underline{\theta} \begin{pmatrix} f_1^\star \\ f_1^\star \\ f_1^\star \end{pmatrix}, \tag{49}$$

where the partial derivative of the boundary describing function $b(\cdot,\cdot;g^\star)$ with respect to e in $(f(\theta),\theta)$ is a function of $f(\theta), \tilde{g}(\theta), f'(\theta), \tilde{g}'(\theta)$ and θ i.e.

$$\frac{\partial b}{\partial e}(f(\theta),\theta;g^\star) = \tilde{h}(f(\theta), \tilde{g}(\theta), f'(\theta), \tilde{g}'(\theta), \theta)\,,$$

for some function $\tilde{h}$, $(f_1^\star, g_1^\star) \in \mathbb{R}^2$ denotes the equilibrium of the perfect information case, i.e. $\Theta_1 = \{1\}$ and hence $D = X$, with $f_1^\star = g_1^\star$, which exists and is unique under the given assumptions, and $\Phi(\theta) = \int_{\underline{\theta}}^{\theta} \phi(t)\,\mathrm{d}t$, $\theta \in \Theta$.

Proof of Proposition 7. A fixed point $(f^\star, g^\star)$ of T, which exists by Proposition 3, satisfies by its very definition that $g^\star = \mathcal{R} \cdot g(\cdot; f^\star)$ and $f^\star = f(\cdot; g^\star)$, i.e. both strategies are their mutual (transformed) best responses. Corollary 1 of Proposition 5 yields the description of $\tilde{g}$ and $\hat{g}$, as well as Proposition 6 that of f, respectively. When looking at $\frac{\partial b}{\partial e}$ as given in Proposition 6, we see that the function h describing $\frac{\partial b}{\partial e}$ in $(f(\theta),\theta)$ depends now on $\tilde{g}(\theta) = g^\star(f(\theta))$ and $\tilde{g}'(\theta) = (g^\star)'(f(\theta))\, f'(\theta)$, where the latter is equivalent to $\frac{\tilde{g}'(\theta)}{f'(\theta)} = (g^\star)'(f(\theta))$, for $\theta \in \Theta$, and thus $\tilde{h}$ defined by

$$\tilde{h}(f(\theta),\tilde{g}(\theta),f'(\theta),\tilde{g}'(\theta),\theta) = h(f(\theta),\tilde{g}(\theta),\frac{\tilde{g}'(\theta)}{f'(\theta)},\theta)\,,$$

is a function of $f(\theta),\tilde{g}(\theta),f'(\theta),\tilde{g}'(\theta)$ and θ as claimed.

It remains to verify that the initial condition. Therefore, we focus on $\tilde{g}_f$ and $\hat{g}_f$ around the starting value $\underline{\theta}$. By assumption, $f^\star$ as well as ϕ are continuously differentiable, and hence by Corollary 1 of Proposition 5 we have $\hat{g}_f'(\underline{\theta}) = \tilde{g}_f'(\underline{\theta}) = \frac{1}{2}\, f'(\underline{\theta})$. Accordingly, $\hat{g}_f'(\underline{\theta}) < f'(\underline{\theta})\frac{\hat{g}(\underline{\theta})}{f(\underline{\theta})} = f'(\underline{\theta})$, where the strict inequality follows from Lemma 2 implying $f' \geq l_f > 0$. By the assumed continuity, there exists an $\varepsilon > 0$ such that $\hat{g}' = \tilde{g}'$ on $[\underline{\theta},\underline{\theta}+\varepsilon]$, where $\varepsilon \leq \overline{\theta} - \underline{\theta}$. It follows that $\hat{g} = \tilde{g}$ on $[\underline{\theta},\underline{\theta}+\varepsilon]$, since also $\hat{g}(\underline{\theta}) = \tilde{g}(\underline{\theta})$ by definition. This implies that $\hat{g}\circ f^{-1} = \tilde{g}\circ f^{-1}$ on $f([\underline{\theta},\underline{\theta}+\varepsilon])$. Using Proposition 5, the best response $g(\cdot,f^\star)$ and its transformation $\mathcal{R}(g(\cdot,f^\star)) = g^\star$ coincide on $f([\underline{\theta},\underline{\theta}+\varepsilon])$. Denote by $\mathbb{P}^{\varepsilon'}_{\tilde{\theta}}$ a modified prior, which is given by $\mathbb{P}^{\varepsilon'}_{\tilde{\theta}}(\cdot) = \mathbb{P}_{\tilde{\theta}}(\cdot\,|\tilde{\theta} \leq \underline{\theta}+\varepsilon')$, for $0 < \varepsilon' \leq \varepsilon$. Then $(f^\star,g^\star)$ restricted to $f^\star([\underline{\theta},\underline{\theta}+\varepsilon])$ and $[\underline{\theta},\underline{\theta}+\varepsilon]$, respectively, is an equilibrium for the prior $\mathbb{P}^{\varepsilon'}_{\tilde{\theta}}$, for all $0 < \varepsilon' \leq \varepsilon$. Now, the best response operators are continuous in the sup-norm as shown in Proposition 3. As $\varepsilon' \searrow 0$, we are tending to the perfect information case, here with known type $\underline{\theta}$, and hence the limit $(f^\star(\underline{\theta}),g^\star(\underline{\theta}))$ is an equilibrium of the perfect information case, here scaled by $\underline{\theta}$. The existence and uniqueness of the equilibrium in the perfect information case, i.e.: $\Theta_1 = \{1\}$ and hence $D = X$, follows from Theorem 1 and the working in Appendix C of Manso (2013). Note that Manso (2013) specifies the coupon function as step function, where our framework allows for a continuous coupon function C, which satisfies Assumption 2. However, using, for example, an approximating sequence of step functions $(C_n)_{n\geq 1}$ to our coupon function C the results carry over. Further, denote by $f_1^\star$ the firm's default threshold in equilibrium. In order to transfer the result from the firm scale to the rating agency scale, we multiply the equilibrium $f_1^\star$ by $\underline{\theta}$ and the initial condition follows as claimed. $\square$

The differential equation satisfied by $(f, \tilde{g}, \hat{g}) = (f^\star, g^\star \circ f^\star, g(\cdot; f^\star) \circ f^\star)$ in (48) and (49) of Proposition 7 can be used to obtain the fixed point $(f^\star, g^\star)$ constructively. The suggested ODE structure is somewhat more complicated, since characterization of the partial derivative f' on the left hand side of (48) also involves the term $\tilde{h}(f(\theta), \tilde{g}(\theta), f'(\theta), \tilde{g}'(\theta), \theta)$ on the right hand side of (48), which depends on f'. Rewriting the first line of (48) as follows $H(f(\theta), \hat{g}(\theta), f'(\theta), \hat{g}'(\theta), \theta) = 0$, is an implicit characterisation of f'. To ensure that this implicit characterization of f' is well-defined, it is required that the partial derivative H with respect to f' is nonzero, such that we can invert this relation locally to back out f'.

Corollary 2. *Given the setting of Proposition 7, suppose that the function H : $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \times \Theta \to \mathbb{R}^+$ given by*

$$H(z_1, z_1, z_3, z_4, z_5) = z_3 - \frac{(1+\eta)\,\sigma^2}{2\,(r-\mu)} \frac{z_1^2/z_5^2}{C(z_1/z_2) - z_1/z_5} \frac{1}{1 - \tilde{h}(z_1, z_2, z_3, z_4, z_5)}\,,$$

is continuously differentiable in all arguments and it holds that

$$\frac{\partial H}{\partial z_3} \neq 0\,, \tag{50}$$

then (48) and (49) is an implicit ODE with solution $(f, \tilde{g}, \hat{g})$. Further, any fixed point $(f^\star, g^\star)$ of T satisfying the conditions above is characterized by $(f^\star, g^\star) = (f, \tilde{g} \circ f^{-1})$ and is thereby unique.

Proof of Corollary 2. The implicit ODE admits a unique solution since it is non-explosive by Lemma 2 and Lemma 9, what show that f and g are both uniformly Lipschitz continuous. $\square$

Proof of Proposition 4. Denote $(f^\star, g^\star)$ a fixed point of T given in Proposition 3. If $g(\cdot; f^\star) = \mathcal{R} \circ g(\cdot; f^\star)$ holds, then $(f^\star, g^\star,)$ is an equilibrium. By Proposition 7 the latter is equivalent to $\tilde{g} = \hat{g}$ which is implied by

$$\hat{g}' \leq f' \, \frac{\hat{g}}{f} \,, \text{ on } (\underline{\theta}, \overline{\theta}) \,.$$

The ODE characterization is identical for all fixed point of T satisfying the assumptions giving the claimed uniqueness, see Corollary 2. $\square$

D Auxiliary Results

From (14) we see that best response of the rating agency $g(\cdot; f)$ with respect to a firm strategy f depends on the image of f, i.e. $f(\Theta) \subset \mathbb{R}_0^+$. In case $f(\Theta)$ is contained in a compact interval $[\underline{e}, \bar{e}]$, which is bounded away from zero, i.e. $0 < \underline{e} \leq \bar{e} < \infty$, then $g(\cdot; f) = Id$ on $[0, \underline{e}]$ and $g(\cdot; f) = g(\bar{e})$ on $[\bar{e}, \infty)$. For this setting, the subsequent lemma shows that convergence in the sup norm $\|\cdot\|_\infty$ is preserved under the functional $\mathcal{R}$. Further, the function $\mathcal{R}(g\cdot; f)$ then acts on a compact interval $[\underline{e}_{\mathcal{R}}, \bar{e}_{\mathcal{R}}]$, with standard continuation outside, i.e. $\mathcal{R}(g) = Id$ on $[0, \underline{e}_{\mathcal{R}}]$ and $\mathcal{R}(g) = g(\bar{e}) =$ constant on $[\bar{e}_{\mathcal{R}}, \infty)$.

Lemma 4. *Suppose that $g, g' \in \mathcal{A}_g$ are continuous, non-decreasing with $0 \leq g, g' \leq Id$, and there exists $0 < \underline{e} \leq \bar{e} < \infty$ such that $g = g' = Id$ on $[0, \underline{e}]$ as well as, $g(e) = g(\bar{e})$ and $g'(e) = g'(\bar{e})$, for $e \in [\bar{e}, \infty)$. Then $\mathcal{R}(g), \mathcal{R}(g') = Id$ on $[0, \underline{e}_{\mathcal{R}}]$ as well as, $\mathcal{R}(g)(e) = g(\bar{e})$ and $\mathcal{R}(g')(e) = g'(\bar{e})$, for $e \in [\bar{e}_{\mathcal{R}}, \infty)$, where $\underline{e}_{\mathcal{R}} = \underline{e}$ and $\bar{e}_{\mathcal{R}} = \frac{\bar{e}^2}{\underline{e}}$, and*

$$\|\mathcal{R}(g) - \mathcal{R}(g')\|_\infty \leq 2 \, \frac{\bar{e}_{\mathcal{R}}}{\underline{e}_{\mathcal{R}}} \, \|g - g'\|_\infty \,. \tag{51}$$

Proof of Lemma 4. Take $e \in [0, \underline{e}_{\mathcal{R}}]$ with $\underline{e}_{\mathcal{R}} = \underline{e}$, then $\mathcal{R}(g)(e) = g(e) = e$ and $\mathcal{R}(g')(e) = g'(e) = e$, hence $|\mathcal{R}(g)(e) - \mathcal{R}(g')(e)| = 0$. Now, take $e \in (\underline{e}, \bar{e})$. Noting that $g(z)/z = g'(z)/z = 1$ on $(0, \underline{e}]$ and in general $g/Id, g'/Id \leq 1$, we see that

$$\begin{aligned}\mathcal{R}(g)(e) &= e \inf\{g(z)/z : \underline{e} \leq z \leq e\}, \text{ and} \\ \mathcal{R}(g')(e) &= e \inf\{g'(z)/z : \underline{e} \leq z \leq e\}.\end{aligned}$$

Without loss of generality assume $\mathcal{R}(g')(e) \leq \mathcal{R}(g)(e)$ and write

$$\mathcal{R}(g)(e) = e\frac{g(z_0)}{z_0}, \text{ and } \mathcal{R}(g')(e) = e\frac{g'(z_0')}{z_0'},$$

where $z_0, z_0' \geq \underline{e}$ are the respective minimizing arguments of the expressions above. These quantities exist due the continuity of g and g', but are perhaps not unique. Then by assumption and the optimality of z_0 we have

$$e\frac{g'(z_0')}{z_0'} = \mathcal{R}(g')(e) \leq \mathcal{R}(g)(e) = e\frac{g(z_0)}{z_0} \leq e\frac{g(z_0')}{z_0'},$$

and

$$0 \leq \mathcal{R}(g)(e) - \mathcal{R}(g')(e) \leq e\frac{g(z_0')}{z_0'} - e\frac{g'(z_0')}{z_0'} \leq \frac{e}{z_0'} \|g - g'\|_\infty \leq \frac{\bar{e}}{\underline{e}} \|g - g'\|_\infty .$$

Finally, we have to check the case $e \in (\bar{e}, \infty)$. We see that

$$\begin{aligned}\mathcal{R}(g)(e) =& e\, (\inf\{g(z)/z : \underline{e} \leq z \leq \bar{e}\} \wedge \inf\{g(z)/z : \bar{e} \leq z \leq e\}) \\ =& \left(\frac{e}{\bar{e}}\, \mathcal{R}(g)(\bar{e})\right) \wedge (e\, \inf\{g(\bar{e})/z : \bar{e} \leq z \leq e\}) \\ =& \left(\frac{\mathcal{R}(g)(\bar{e})}{\bar{e}}\, e\right) \wedge g(\bar{e}), \text{ for } e \geq \bar{e},\end{aligned}$$

and analogously

$$\mathcal{R}(g')(e) = \left(\frac{\mathcal{R}(g')(\bar{e})}{\bar{e}}\, e\right) \wedge g'(\bar{e}), \text{ for } e \geq \bar{e}.$$

Without loss of generality assume that $\mathcal{R}(g')(\bar{e}) \leq \mathcal{R}(g)(\bar{e})$. Define

$$\begin{aligned} e_0 &= \inf\{e \geq \bar{e} : \mathcal{R}(g)(e) = g(\bar{e})\} = \frac{g(\bar{e})\, \bar{e}}{\mathcal{R}(g)(\bar{e})}\,, \text{ and} \\ e_0' &= \inf\{e \geq \bar{e} : \mathcal{R}(g')(e) = g'(\bar{e})\} = \frac{g'(\bar{e})\, \bar{e}}{\mathcal{R}(g')(\bar{e})}\,. \end{aligned}$$

Consider the case $e \geq e_0'$, then $\mathcal{R}(g')(e) = \mathcal{R}(g')(e_0') = g'(\bar{e})$, and

$$|\mathcal{R}(g)(e) - \mathcal{R}(g')(e)| = \begin{cases} |g(\bar{e}) - g'(\bar{e})|\,, & \text{for } e \geq (e_0 \vee e_0')\,, \\ \left|\frac{\mathcal{R}(g)(\bar{e})}{\bar{e}}\, e - \frac{\mathcal{R}(g')(\bar{e})}{\bar{e}}\, e_0'\right|, & \text{for } e_0' \leq e < (e_0 \vee e_0')\,. \end{cases}$$

Focusing on $e_0' \leq e < (e_0 \vee e_0')$, recalling $\mathcal{R}(g')(\bar{e}) \leq \mathcal{R}(g)(\bar{e})$ and using that the assumption g, g' are non-decreasing implies $\mathcal{R}(g)(\bar{e}) \geq \underline{e}$ and $\mathcal{R}(g')(\bar{e}) \geq \underline{e}$, respectively, gives

$$\begin{aligned} |\mathcal{R}(g)(e) - \mathcal{R}(g')(e)| &= \left|\frac{\mathcal{R}(g)(\bar{e})}{\bar{e}}\, e - \frac{\mathcal{R}(g')(\bar{e})}{\bar{e}}\, e_0'\right| \\ &= \frac{1}{\bar{e}} \left(\mathcal{R}(g)(\bar{e})\,(e - e_0') + (\mathcal{R}(g)(\bar{e}) - \mathcal{R}(g')(\bar{e}))\; e_0'\right) \\ &\leq \frac{1}{\bar{e}} \left(\bar{e}\,((e_0 \vee e_0') - e_0') + (\mathcal{R}(g)(\bar{e}) - \mathcal{R}(g')(\bar{e}))\, \frac{g'(\bar{e})\, \bar{e}}{\mathcal{R}(g')(\bar{e})}\right) \\ &\leq ((e_0 - e_0') \vee 0) + (\mathcal{R}(g)(\bar{e}) - \mathcal{R}(g')(\bar{e}))\, \frac{\bar{e}}{\underline{e}} \\ &\leq \left(\left(\frac{g(\bar{e})\, \bar{e}}{\mathcal{R}(g)(\bar{e})} - \frac{g'(\bar{e})\, \bar{e}}{\mathcal{R}(g')(\bar{e})}\right) \vee 0\right) + \frac{\bar{e}^2}{\underline{e}^2}\, \|g - g'\|_\infty \\ &\leq \frac{\bar{e}}{\mathcal{R}(g')(\bar{e})} \left(\left(\frac{\mathcal{R}(g')(\bar{e})}{\mathcal{R}(g)(\bar{e})}\, g(\bar{e}) - g'(\bar{e})\right) \vee 0\right) + \frac{\bar{e}^2}{\underline{e}^2}\, \|g - g'\|_\infty \\ &\leq \frac{\bar{e}}{\underline{e}}\, ((g(\bar{e}) - g'(\bar{e})) \vee 0) + \frac{\bar{e}^2}{\underline{e}^2}\, \|g - g'\|_\infty \\ &\leq 2\, \frac{\bar{e}^2}{\underline{e}^2}\, \|g - g'\|_\infty \end{aligned}$$

And thus for all $e \geq e_0'$ we have

$$|\mathcal{R}(g)(e) - \mathcal{R}(g')(e)| \leq 2\, \frac{\bar{e}^2}{\underline{e}^2}\, \|g - g'\|_\infty \,.$$

For $\bar{e} \leq e \leq e_0'$ compute

$$\begin{aligned} |\mathcal{R}(h)(y) - \mathcal{R}(h')(y)| &\leq (\mathcal{R}(g)(\bar{e}) - \mathcal{R}(g')(\bar{e}))\, \frac{e}{\bar{e}} \;\leq\; \frac{\bar{e}}{\underline{e}}\, \|g - g'\|_\infty\, \frac{e_0'}{\bar{e}} \\ &\leq \frac{\bar{e}}{\underline{e}}\, \|g - g'\|_\infty\, \frac{1}{\bar{e}}\, \frac{g'(\bar{e})\, \bar{e}}{\mathcal{R}(g')(\bar{e})} \;\leq\; \frac{\bar{e}}{\underline{e}}\, \|g - g'\|_\infty\, \frac{1}{\bar{e}}\, \frac{\bar{e}\, \bar{e}}{\underline{e}} \\ &= \frac{\bar{e}^2}{\underline{e}^2}\, \|g - g'\|_\infty \,. \end{aligned}$$

Noting that $e_0, e_0' \leq \bar{e}_{\mathcal{R}} = \frac{\bar{e}^2}{\underline{e}}$ and verifying that $\mathcal{R}(g)$ and $\mathcal{R}(g')$ are constant and equal to $g(\bar{e})$ and $g'(\bar{e})$, respectively, finishes the proof. □

In order to prove Proposition 2, the value function $v(\cdot, \cdot; \theta, g)$ of the optimal stopping problem of the firm in (8), with $\theta \in \Theta$ and $g \in \mathcal{A}_g^{\mathbf{C}}$ such that g is non-decreasing and bounded by Id, has to be characterized. Note that (8) is on the rating agency-scale, using the imperfectly observed cash flow D and its running minimum E. It is helpful to also consider the firm's optimal stopping problem also on the firm-scale, i.e. $X = E/\theta$ and $Y = D/\theta$, for $\theta > 0$. For $(x, y) \in \mathcal{C}$, define

$$w(x, y; \theta, g) = \sup_{\tau \in \mathcal{T}_{(x,y)}} \mathbb{E}_{(x,y)} \left[\int_0^\tau e^{-r\, t} \left(X_t - C(\theta\, X_t / g(\theta\, Y_t))\right)\, \mathrm{d}t \right] , \tag{52}$$

where the firm cash flow X follows (1), its running minimum $Y = (Y_t)_{t \geq 0}$ is given by $Y_t = \min(Y_0, \inf_{0 \leq s \leq t} X_s)$, for $t \geq 0$, as well as $g \in \mathcal{A}_g^{\mathbf{C}}$ is non-decreasing, bounded by Id, and $\theta \in \Theta$.

First properties of the value function $w(\cdot, \cdot; \theta, g)$ defined in (52) are collected in the following Lemma. Therefore, it is useful to define the function g_θ appearing inside the interest payment rate function C in (52) by $g_\theta : \mathbb{R}_0^+ \to \mathbb{R}_0^+, y \mapsto g_\theta(y) = \frac{1}{\theta}\, g(\theta\, y)$.

Lemma 5. *Denote $v(\cdot, \cdot; \theta, g)$ and $w(\cdot, \cdot; \theta, g)$ the value functions specified in* (16) *and* (52)*, respectively, for* $\theta > 0$ *and* $g \in \mathcal{A}_g^{\mathbf{C}}$*, then* $g_\theta \in \mathcal{A}_g^{\mathbf{C}}$ *and*

$$w(x, y; \theta, g) = v(\theta\, x, \theta\, y; \theta, g)\,, \qquad \text{for } (x, y) \in \mathcal{C}\,, \tag{53}$$

$$w(x, y; \theta, g) = v(x, y; 1, g_\theta)\,, \qquad \text{for } (x, y) \in \mathcal{C}\,. \tag{54}$$

Moreover, for $g \in \mathcal{A}_g$ *and* $0 < \theta' \leq \theta$ *it holds:*

1. *If* g *is non-decreasing, then* g_θ *is non-decreasing, and* $\theta'\, g_{\theta'} \leq \theta\, g_\theta$*.*

2. *If* g/Id *is non-increasing, then* g_θ/Id *is non-increasing, and* $g_{\theta'} \geq g_\theta$*.*

3. *If* $g \leq Id$*, then* $g_\theta \leq Id$*.*

Proof of Lemma 5. The equality (53) follows directly from the definition in (16), as (D, E) is obtained from (X, Y) by multiplying by $\theta > 0$. Equality (54) follows likewise, using the definition of g_θ above and (52). To show part 1. observe that for $0 < y' \leq y$, it follows $\theta\, y' \leq \theta\, y$, and by g being non-decreasing, we have

$$g_\theta(y) = \frac{g(\theta\, y)}{\theta} \geq \frac{g(\theta\, y')}{\theta} = g_\theta(y')\,,$$

i.e. g_θ is non-decreasing. Now,

$$\theta\, g_\theta(y) = g(\theta\, y) \geq g(\theta'\, y) = \theta'\, g_{\theta'}(y)\,, \text{ for } y \geq 0\,,$$

since g is non-decreasing and $\theta'\, y \leq \theta\, y$. For part 2., consider

$$\frac{g_\theta(y')}{y'} = \frac{g(\theta\, y')}{\theta\, y'} \geq \frac{g(\theta\, y)}{\theta\, y} = \frac{g_\theta(y)}{y}\,, \text{ for } y > 0\,,$$

since g/Id is non-increasing and $\theta\, y \geq \theta\, y'$. Thus g_θ/Id is non-increasing. Further,

$$g_{\theta'}(y) = y\, \frac{g(\theta'\, y)}{\theta'\, y} \geq y\, \frac{g(\theta\, y)}{\theta\, y} = g_\theta(y)\,, \text{ for } y > 0\,,$$

since g/Id is non-increasing and $\theta' y \le \theta y$. For part 3., write

$$g_\theta(y) = \frac{g(\theta\, y)}{\theta} \le \frac{\theta\, y}{\theta} = y\,,\ \text{for } y \ge 0\,,$$

since $g \le Id$, hence $g_\theta \le Id$. $\square$

The value functions v and w given in (16) and (52), respectively, are described by expectations conditioning on the starting values (e, d) and (x, y), respectively. For the subsequent analysis it is helpful to write the dependence on the starting value directly into the payoff function, which is possible since the driving processes X and E, respectively, are geometric Brownian motions, see (1) and (2), respectively. Denote by $(\tilde{X}, \tilde{Y})$ the process X defined in (16) with starting value 1 and $\tilde{Y} = (\tilde{Y}_t)_{t\ge 0}$ its running minimum, i.e. $\tilde{Y}_t = \inf_{0\le s\le t} \tilde{X}_s$, for $t \ge 0$. Then

$$v(d, e; \theta, g) = \sup_{\tau\in\tilde{\mathcal{T}}} \mathbb{E}\left[\int_0^\tau e^{-r\,t}\left(d\,\tilde{X}_t/\theta - C(d\,\tilde{X}_t/g(\min(e, d\,\tilde{Y}_t)))\right)\mathrm{d}t\right], \tag{55}$$

$$w(x, y; \theta, g) = \sup_{\tau\in\tilde{\mathcal{T}}} \mathbb{E}\left[\int_0^\tau e^{-r\,t}\left(x\,\tilde{X}_t - C(x\,\tilde{X}_t/g_\theta(\min(y, x\,\tilde{Y}_t)))\right)\mathrm{d}t\right], \tag{56}$$

where $\tilde{\mathcal{T}}$ denotes the set of stopping times w.r.t. to the filtration generated by $(\tilde{X}, \tilde{Y})$. To allow for straight-forward calculations subsequently, we extend the interest payment rate function C from $[1, \infty]$ trivially to $[0, \infty]$ by setting $C|_{[0,1)} = C(1)$. This representation allows to establish the following properties of $v(\cdot, \cdot; \theta, g)$ and $w(\cdot, \cdot; \theta, g)$.

Lemma 6. *Denote $v(\cdot, \cdot; \theta, g)$ and $w(\cdot, \cdot; \theta, g)$ the value functions specified in* (16) *and* (52)*, respectively, for $\theta > 0$ and $g, g' \in \mathcal{A}_g^{\mathbf{C}}$, then the following holds true:*

1. *$v(\cdot, \cdot; \theta, g)$ and $w(\cdot, \cdot; \theta, g)$ are non-negative.*
2. *If $0 < \theta' \le \theta$, then $v(\cdot, \cdot; \theta', g) \ge v(\cdot, \cdot; \theta, g)$ and $w(\cdot, \cdot; \theta', g) \le w(\cdot, \cdot; \theta, g)$.*
3. *If $g\prime \le g$, then $v(\cdot, \cdot; \theta, g') \ge v(\cdot, \cdot; \theta, g)$ and $w(\cdot, \cdot; \theta, g') \ge w(\cdot, \cdot; \theta, g)$.*

4. $v(\cdot,\cdot;\theta,g)$ *and* $w(\cdot,\cdot;\theta,g)$ *are non-increasing in* e *and* y*, respectively, i.e.:*

$$v(d,e';\theta,g) \geq v(d,e;\theta,g)\,, \text{ for } 0 \leq e' \leq e \leq d < \infty\,,$$
$$w(x,y';\theta,g) \geq w(x,y;\theta,g)\,, \text{ for } 0 \leq y' \leq y \leq x < \infty\,.$$

5. $v(\cdot,\cdot;\theta,g)$ *and* $w(\cdot,\cdot;\theta,g)$ *are non-decreasing on rays starting in the origin, i.e.:*

$$v(d,e;\theta,g) \leq v(\lambda\,d,\lambda\,e;\theta,g)\,, \text{ for } (d,e)\in\mathcal{C} \text{ and } \lambda \geq 1\,,$$
$$w(x,y;\theta,g) \leq w(\lambda\,x,\lambda\,y;\theta,g)\,, \text{ for } (x,y)\in\mathcal{C} \text{ and } \lambda \geq 1\,.$$

Proof of Lemma 6. Part 1. follows from $\tau = 0$. For the remainder of the proof recall the convention the interest payment rate function C is extended from $[1,\infty]$ trivially to $[0,\infty]$ by setting $C|_{[0,1)} = C(1)$. For part 2., we focus on v as given in (55) and compare the accumulated discounted net income stream until $\tau \in \tilde{\mathcal{T}}$. For $0 < \theta' \leq \theta$, we have almost surely

$$d\,\tilde{X}_t/\theta' - C(d\,\tilde{X}_t/g(\min(e,d\,\tilde{Y}_t))) \geq \tilde{X}_t/\theta - C(d\,\tilde{X}_t/g(\min(e,d\,\tilde{Y}_t)))\,,$$

and hence almost surely

$$\int_0^\tau e^{-r\,t}\left(d\,\tilde{X}_t/\theta' - C(d\,\tilde{X}_t/g(\min(e,d\,\tilde{Y}_t)))\right)\mathrm{d}t \\ \geq \int_0^\tau e^{-r\,t}\left(\tilde{X}_t/\theta - C(d\,\tilde{X}_t/g(\min(e,d\,\tilde{Y}_t)))\right)\mathrm{d}t\,.$$

The inequality is preserved by taking the expectation and the supremum over all $\tau \in \tilde{\mathcal{T}}$, and the first assertion of part 2. follows. For w, we take (52) . For g/Id non-increasing, as assumed, part 2. of Lemma 5 gives that $0 < \theta' \leq \theta$ implies $g_{\theta'} \geq g_\theta$, and since C is non-increasing we have

$$x\,\tilde{X}_t - C(x\,\tilde{X}_t/g_{\theta'}(\min(y,x\,\tilde{Y}_t))) \leq x\,\tilde{X}_t - C(x\,\tilde{X}_t/g_\theta(\min(y,x\,\tilde{Y}_t)))\,,$$

and by similar arguments as before, i.e. integrating the discounted payoff stream over $[0,\tau]$ as well as noting the inequality is preserved by taking the expectation and the supremum over all $\tau \in \tilde{\mathcal{T}}$, it follows that $w(x,y;\theta',g) \leq w(x,y;\theta,g)$, for all $(x,y) \in \mathcal{C}$, as claimed. To show part 3., observe that for $g' \leq g$ we have

$$C(a/g'(b)) \leq C(a/g(b))\,, \text{ for } (a,b) \in \mathbb{R}_0^+ \times \mathbb{R}^+\,,$$

since C is non-increasing, and the first assertion follows using similar arguments as for part 2. For w, we calculate

$$g'_\theta(y) = \frac{g'(\theta\, y)}{\theta} \leq \frac{g(\theta\, y)}{\theta} = g_\theta(y)\,, \text{ for } y > 0\,.$$

Hence, $g'_\theta \leq g_\theta$ and the second assertion follows by identical arguments as the first assertion. To verify assertion 4, observe that g is non-decreasing by assumption, and thus g_θ is also non-decreasing by part 1. of Lemma 5. Therefore, $g(b') \leq g(b)$ and $g_\theta(b') \leq g_\theta(b)$, for $0 \leq b' \leq b$. Since C is non-increasing, it holds

$$C(a/g(b') \leq C(a/g(b)) \text{ and } C(a/g_\theta(b') \leq C(a/g_\theta(b))\,, \text{ for } a \geq 0 \text{ and } 0 < b' \leq b.$$

Applying the same arguments as in part 3. to the representations in (55) and (56) gives the claimed result. Now, part 5. is verified. Note that g/Id is non-increasing by assumption, which implies by part 2. of 5 that $g \geq g_\lambda$, where we set $1 = \theta' \leq \theta = \lambda$, and

$$\frac{a}{g(b)} \leq \frac{a}{g_\lambda(b)} = \frac{\lambda\, a}{g(\lambda\, b)}\,, \text{ for } (a,b) \in \mathbb{R}_0^+ \times \mathbb{R}^+\,. \tag{57}$$

Taking a look at the net income rate in (55) for v with starting value $(\lambda\, d, \lambda\, e)$, where

$(d, e) \in \mathcal{C}$ and $\lambda \geq 1$, we obtain

$$\begin{aligned}
\lambda\, d\, \tilde{X}_t/\theta - C(\lambda\, e\, \tilde{X}_t/g(\min(\lambda\, e, \lambda\, d\, \tilde{Y}_t)))& \\
&= d\, \tilde{X}_t/\theta - C(d\, \tilde{X}_t/g(\min(e, d\, \tilde{Y}_t))) + (\lambda - 1)\, d\, \tilde{X}_t/\theta \\
&\quad + C(d\, \tilde{X}_t/g(\min(e, d\, \tilde{Y}_t))) - C(\lambda\, d\, \tilde{X}_t/g(\lambda\, \min(e, d\, \tilde{Y}_t))) \\
&\geq d\, \tilde{X}_t/\theta - C(d\, \tilde{X}_t/g(\min(e, d\, \tilde{Y}_t)))\,,
\end{aligned}$$

since $(\lambda - 1)\, d\, \tilde{X}_t/\theta$ is greater equal to zero due to $\lambda \geq 1$ and $C(d\, \tilde{X}_t/g(\min(e, d\, \tilde{Y}_t))) - C(\lambda\, d\, \tilde{X}_t/g(\lambda\, \min(e, d\, \tilde{Y}_t))) \geq 0$ thanks to (57), set $a = d\, \tilde{X}_t$ and $b = \min(e, d\, \tilde{Y}_t)$. And by similar arguments as before, i.e. integrating the discounted payoff stream over $[0, \tau]$ as well as noting the inequality is preserved by taking the expectation and the supremum over all $\tau \in \tilde{\mathcal{T}}$, it follows that $v(d, e; \theta, g) \leq v(\lambda\, d, \lambda\, e; \theta, g)$, for all $(d, e) \in \mathcal{C}$, as claimed. When considering w, observe that from Lemma 5 part 2. it follows that g_θ/Id is non-increasing. Now, the similar reasoning as for v applies and the proof is finished. □

For a given rating agency strategy $g \in \mathcal{A}_g^{\mathbf{C}}$ and $\theta > 0$, the early exercise region

$$\mathcal{E}(\theta; g) = \{(d, e) \in \mathcal{C} : v(d, e; \theta, g) = 0\}\,, \tag{58}$$

allows us to characterize the best response of the firm $\tau(\theta; g)$ as first hitting time. An important subset of $\mathcal{E}(\theta; g)$ is that on the diagonal, which is identified with $\mathcal{D}(\theta; g)$ and the corresponding supremum $f(\theta; g)$, i.e.

$$\mathcal{D}(\theta; g) = \{d \in \mathbb{R}_0^+ : (d, d) \in \mathcal{E}(\theta; g)\}\,, \text{ and } D(\theta; g) = \sup \mathcal{D}(\theta; g). \tag{59}$$

Lemma 7. *Let $\mathcal{E}(\theta; g)$, $\mathcal{D}(\theta; g)$, and $D(\theta; g)$ be given by* (58) *and* (59)*, respectively, for $\theta > 0$ and $g \in \mathcal{A}_g^{\mathbf{C}}$. Then*

$$\mathcal{D}(\theta; g) = [0, D(\theta; g)]\,, \tag{60}$$

and for $(d,e) \in \mathcal{E}(\theta; g)$ *it holds that*

$$e \leq d \leq D(\theta; g)\,. \tag{61}$$

Proof of Lemma 7. To see the first assertion, note that $v(0,0;\theta,g) = 0$, and hence $\mathcal{D}(\theta; g)$ is non-empty. For $d \in \mathcal{D}(\theta; g)$, we have $d' \in \mathcal{D}(\theta; g)$, for $d' \in (0, d]$ by part 5. of Lemma 6 by setting $\lambda = d/d' \geq 1$. Further, we have for $d > \theta\,\bar{C}$ that $v(d,d;\theta,g) > 0$, since then the income stream from not defaulting in (d,d) is strictly positive. Accordingly, $\mathcal{D}(\theta; g)$ is a convex and bounded subset of $\mathbb{R}_0^+$. Since $v(\cdot,\cdot;\theta,g)$ is continuous, $\mathcal{D}(\theta; g)$ is also closed and $D(\theta; g) \in \mathcal{D}(\theta; g)$, and (60) follows. For the second assertion, take $(d,e) \in \mathcal{E}(\theta; g)$, then $(d,d) \in \mathcal{E}(\theta; g)$ by part 4. of Lemma 6, what is equivalent to $d \in \mathcal{D}$. Thus, $d \leq D(g; h)$. Since $(d,e) \in \mathcal{E}(\theta; g) \subset \mathcal{C}$ we have $e \leq d$ finishing the proof. □

Lemma 8. *The set* $\mathcal{K}_f$ *is convex and compact in* $(C(\Theta,\mathbb{R}), \|\cdot\|_\infty)$*. Moreover,* $\mathcal{K}_f$ *is uniformly bounded by* $\overline{\theta}\,\overline{f}$ *and uniformly Lipschitz continuous with Lipschitz* $L_f = \bar{f}$*.*

Proof of Lemma 8. To see that $\mathcal{K}_f$ is convex, take $f, f' \in \mathcal{K}_f$, $\lambda \in [0,1]$ and define $f^\lambda = \lambda\, f + (1-\lambda)\, f'$. Now, f^λ is continuous, since f, f' are, hence $f^\lambda \in C(\Theta,,\mathbb{R})$. For $\theta, \theta' \in \Theta$ with $\theta' \leq \theta$ we have

$$\begin{aligned}
f^\lambda(\theta) - f^\lambda(\theta') &= \lambda\, f(\theta) + (1-\lambda)\, f'(\theta) - \lambda\, f(\theta') - (1-\lambda)\, f'(\theta') \\
&= \lambda(f(\theta) - f(\theta')) + (1-\lambda)(f'(\theta) - f'(\theta')) \\
&\leq \lambda\, L_f\,(\theta - \theta') + \lambda\, L_f\,(\theta - \theta') = L_f\,(\theta - \theta')\,.
\end{aligned}$$

Using similar reasoning, one verifies that all conditions of the definition of $\mathcal{K}_f$ in (32) hold for f^λ, and thus $f^\lambda \in \mathcal{K}_f$. Accordingly, $\mathcal{K}_f$ is convex. To see that $\mathcal{K}_f$ is compact it is by Arzela-Ascoli sufficient to show that $\mathcal{K}_f$ is closed, bounded and equicontinuous. To show that $\mathcal{K}_f$ is closed consider a sequence $(f_n)_{n\geq 1}$ in $\mathcal{K}_f$ that converges to some

$f \in C(\Theta, , \mathbb{R})$, i.e. $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$. For $\theta, \theta' \in \Theta$ with $\theta' \leq \theta$ we have

$$\begin{aligned} f(\theta) - f(\theta') &\leq f_n(\theta) + \|f_n - f\|_\infty - f_n(\theta') + \|f_n - f\|_\infty \\ &\leq L_f\,(\theta - \theta') + 2\,\|f_n - f\|_\infty\,. \end{aligned}$$

This holds for all $n \geq 1$. As $n \to \infty$, we obtain $f(\theta) - f(\theta') \leq L_f\,(\theta - \theta')$. Using similar reasoning, one verifies that all conditions of the definition of $\mathcal{K}_f$ in (32) hold for f, and thus $f \in \mathcal{K}_f$. Accordingly, $\mathcal{K}_f$ is closed. That $\mathcal{K}_f$ is bounded follows immediately from the definition with uniform upper bound $\overline{\theta}\,\overline{f}$. The equicontinuity of $\mathcal{K}_f$ is implied if all $f \in \mathcal{K}_f$ are Lipschitz continuous with a common Lipschitz constant L_f, which holds by the very definition of $\mathcal{K}_f$. Note that the common Lipschitz constant is given by $L_f = \overline{f}$. $\square$

Lemma 9. *The set $\mathcal{K}_g$ is convex and compact in $(C(\Xi, \mathbb{R}_0^+), \|\cdot\|_\infty)$. Moreover, $\mathcal{K}_g$ is uniformly bounded by $\overline{\theta}^2\overline{f}^2/(\underline{\theta}\underline{f})$ and uniformly Lipschitz continuous with Lipschitz constant $L_g = 1$.*

Proof of Lemma 9. The proof follows along the same lines as that of Lemma 8 once it is shown that $\mathcal{K}_g$ is uniformly bounded and uniformly Lipschitz continuous. The uniform bound of $\overline{\theta}^2\overline{f}^2/(\underline{\theta}\underline{f})$ follows from $g \leq Id$ for $g \in \mathcal{K}_h \subseteq (\Xi)$ and $\overline{\xi} = \overline{\theta}^2\overline{f}^2/(\underline{\theta}\underline{f})$. For the uniform Lipschitz continuity, observe for $g \in \mathcal{K}_g$ and $e, e' \in \Xi$ with $e' \leq e$ that

$$\begin{aligned} 0 &\leq g(e) - g(e') = e\,g(e)/e - g(e') \\ &\leq e\,g(e')/e' - g(e') = (e - e')\,g(e')/e' \leq (e' - e)\,, \end{aligned}$$

where the first step follows since $g \in \mathcal{K}_g$ is non-decreasing, and the second step from g/Id is non-increasing, and the final step from $g \leq Id$. Accordingly, g is Lipschitz continuous with Lipschitz constant $L_g = 1$, which is common for all $g \in \mathcal{K}_g$. Now, the remaining claims follow by the same arguments as in the proof of Lemma 8. $\square$