Rating-Induced Default Risk and Downgrade Hesitation✩

Hermann W. Elendner

*Vienna Graduate School of Finance (VGSF),
Heiligenstädter Straße 46-48, A–1190 Vienna, Austria

hermann.elendner@vgsf.ac.at

Abstract

Credit ratings should reflect credit risk. Mounting evidence implies they also impact credit risk. As strategic agencies will take this into account, I build a model to show how this feedback effect incentivizes raters to postpone or omit downgrades which they know would be warranted. If agencies succumb to the conflict of interest, they restrict self-inflicted fee losses by optimally hesitating to announce a strictly positive fraction of merited announcements. In equilibrium, ratings are informative, but only partially, because the agency withholds information—irrespective of any level of reputation costs.

To devise a test design, I derive empirical predictions: It is shown that the probability of agencies concealing downgrades is increasing with obligors’ proximity to default and their reliance on external financing, while decreasing in distance to the default boundary and subsequent to crises due to higher reputation costs. For opaque firms more information is held back. Finally, I detail the identification strategy for an empirical test of the predictions based on CDS spreads and market-implied ratings.

Key words: credit ratings, rating agencies, CRA, credit cliff, feedback effect

JEL: G14, G24, G38, D8, L14

✩I am thankful to Hamed Ghoddusi, Kurt Hornik, Christian Laux, Stefan Pichler, and Tanja Veža for insightful comments and valuable suggestions, as well as for those received at the VGSF PhD research seminar, the Humboldt-Universität zu Berlin, and the Austrian Working Group 2011.

March 19, 2012
Rating-Induced Default Risk
and Downgrade Hesitation

“I never have bought any bonds over the last 28 years based upon the credit-agency ratings – they are always delayed on the upside or downside.”

Leslie Beck,¹ Wall Street Journal, November 2, 2008

“In recent years bond rating agencies have been under increasing scrutiny because of their obvious failures to accurately predict and warn investors of impending firm-related financial difficulties”

Pinches and Singleton (1978 [sic]), Journal of Finance

1. Introduction

Credit-rating agencies have faced a surge in accusations of dishonest reporting—although such criticism is nothing new (see the quote above). Nonetheless, little evidence exists so far to substantiate such claims convincingly.² This lack is not due to their ratings’ confirmed accuracy: rather, two apparent shortcomings have proven highly effective in justifying doubtful regularities: First, the limited information content of ratings, and second, agencies’ imperfect information.

The fact that ratings provide only limited information makes it difficult to contradict them. As they effectively establish merely an ordinal ranking of an unspecified measure of creditworthiness per point in time, complaints about “too high” or “too low” risk in any class are ill-contrived: Neither can the correct measure of risk be identified, nor is it permissible to criticize any relation among different classes apart from the pure ordering—and that holds remarkably well empirically, in and out of severe crises. (Hanson and Schuermann, 2006)³

For the same reason comparisons across agencies are inconclusive (unless they take potentially differing rating scales into account; see Cantor and Packer, 1994; Eisl et al., 2012). Likewise, the time dimension does not allow for decisive tests, since ratings neither are nor are supposed to be time-invariant (Blume et al., 1998). Actually, agencies explicitly state that standards of creditworthiness per class fluctuate over time, a practice referred to as “rating through

---

¹Cited in Miniter (2008).
²Possibly with the exception of Xia (2011), who documents that Egan-Jones Ratings, hardly at an informational advantage, publishes better ratings than Standard & Poor’s; however, he does not account for potentially different rating scales. Other excellent research like He et al. (2010) or Mählmann (2011), while finding strong indication of an incentive problem, cannot conclusively determine whether it arises at the agencies’ or the ratees’ end, i.e., cannot distinguish biased reporting from “rating shopping,” the practice of obligors to request ratings from different agencies but let only the relatively better ones be published (Mählmann, 2008).
³Although exceptions exist, e.g., Mählmann (2006). Also, monotonicity should only hold for proxies of credit quality as conceived by agencies. Momentum, for instance, appears to exhibit a U-shaped relation to credit ratings (Avramov et al., 2007).
the cycle” (Altman and Rijken, 2004; Löffler, 2004, 2005). Consequently, the majority of objections against credit ratings are rooted in misconceptions.

The second cornerstone in rejecting allegations of deceit has been the obvious fact that agencies themselves are subject to noisy information and prone to estimation errors. Contrary to many influential models of rating agencies, which assume that a dishonest agency will be found out and disregarded eventually (e.g., Bolton et al., 2011), in reality any mis-assessment, no matter how striking, can stem not only from unlikely realizations, but beyond that from honest mistake.4 Along these lines, then-CEO of Standard & Poor’s Deven Sharma responded to perceivedly faulty ratings in the wake of the financial crisis: “We’re very disappointed and embarrassed.” (Lippert, 2011) He acknowledged his firm had been fooled by randomness, but insisted on its uncompromised integrity due to reputational concerns.

This crucial claim is extremely challenging to judge, because rated obligors provide private information to rating agencies,5 which contractually bind themselves to keep such information private (Jorion et al., 2005). In principle, this allows them, in the spirit of Benabou and Laroque (1992), to repeatedly manipulate reports without being found out.

I devise a test of whether they succumb to incentive problems based on regularities in their downgrade hesitation. This time lag from when a downgrade is warranted to when it is announced is well-documented (Weinstein, 1977; Pinches and Singleton, 1978; Hand et al., 1992; Ederington and Goh, 1998; Altman and Rijken, 2005; Güttler and Wahrenburg, 2007; Alsakka and ap Gwilym, 2010a,b). Along the two defense lines illustrated above, agencies attribute this lag to the intent to provide stable ratings and the information imperfections they are subject to, too. As far as feasible under these limitations, the argument goes, reputation costs induce honest reporting. I thus devise a procedure to separate the effects of noisy signals from those of a conflict of interest.

This conflict I focus on arises from a phenomenon which has appeared recently in the literature, theoretically modeled by Manso (2011), while Behr and Güttler (2008) unsuspectingly provide empirical evidence, namely the feedback effect: ratings not only reflect credit risk, but also influence it. From this I derive an incentive problem, as strategic agencies will hesitate to announce warranted downgrades which further impair fee-paying obligors’ default risk.

I thus build a game-theoretic model to illustrate the economic interplay and clarify how arbitrarily large reputation costs can be insufficient to induce honesty. With the model, I derive a set of testable empirical predictions, relating the magnitude of the feedback effect, changes in reputation costs, the present value of future fee income, and firm characteristics such as opaqueness and the level of idiosyncratic risk to the extent of downgrade hesitation.

The outline of the paper follows: Section 2 reviews the related literature

---

4 Opp et al. (2010) present a model that explicitly separates the CRA’s acquisition of a noisy signal from potentially dishonest disclosure; however, they then prove a lemma that under their assumptions honest disclosure is always optimal.

5 An exception are so-called unsolicited ratings, see Fulghieri et al. (2011) for a theoretical model and Poon (2003); Gan (2004); Poon and Firth (2005); Behr and Güttler (2008); Fairchild et al. (2009); Poon et al. (2009) as well as Bannier et al. (2010) for empirical results.
and discusses the mounting relevance of the feedback effect. Section 3 presents the model, the results of which are shown in Section 4. Subsequently, Section 5 derives empirical predictions and devises the test design. Finally, Section 6 concludes.

2. Related Literature

Practically all notable bond issues or issuers around the globe, both corporate and sovereign, are rated by at least one credit-rating agency (CRA). These credit ratings are as fuzzy as pervasive: They represent purely ordinal assessments of at best vaguely determined measures of creditworthiness on unknown scales. Moreover, they are publicly available and rarely change. Nevertheless, they appear to contain information which eludes market prices.

Curiously, common justifications of the importance of rating agencies do not even rely on incremental information content. Simply the fact that obligors, financial institutions and regulators let ratings influence their decisions confers weight to these judgments. All three do so extensively.

First, there is evidence that non-financial firms adjust their capital structure to accommodate credit ratings. This extends beyond lip service by chief financial officers, as collected by Kisgen (2007) or as famously surveyed by Graham and Harvey (2001), to whom CFOs report credit ratings as their second-highest objective in determining their company’s leverage. More formally, Kisgen (2006) shows empirically that firms near a credit-rating upgrade or downgrade issue less debt relative to equity than firms with no rating change in sight. Furthermore, Kisgen (2009) demonstrates that firms reduce leverage following credit-rating downgrades. Similarly, Sufi (2009) finds that the introduction of bank-loan ratings has increased some firms’ access to capital and so fostered real investment, and Tang (2009) reports real effects of unanticipated upgrades on capital investments, cash accumulation, and asset growth.

Second, institutional investors, in particular mutual funds, insurers, and pension funds, frequently enact charters restricting the amounts of debt to hold in certain rating classes, often limiting themselves to so-called investment grade. Empirical evidence suggests that for bonds dropping out of this category selling pressure from insurance companies alone can be strong enough to temporarily distort bond market-prices (Ellul et al., 2010). Moreover, after crossing the investment-grade boundary, firms become sensitive to fund flows; they apparently cut real investment when retail investors withdraw capital from high-yield mutual funds, the main buyers of speculative-grade bonds (Chernenko and Sunderam, 2010). In addition, banks often calibrate their credit-risk models to agencies’ ratings.

Third, regulatory capital requirements in the standardized approach of Basel II (BIS, 2004) directly determine the necessary capital reserve and thus the cost of capital to regulated banks via agency ratings (Rösch, 2002).

Another indication of the importance of ratings is their prevalence despite the fact that most are costly. Apparently, about 98% of issuers find it worthwhile to pay well above half a basis point in fees for an additional bond issue being rated (Klinger and Sarig, 2000), not accounting for the considerable resources
spent on and risks involved in providing private information to the agency. In fact, the vast majority chooses to incur these costs twice or thrice, by acquiring ratings from several rating agencies (Mählmann, 2009; Bongaerts et al., 2010).

Finally, while agencies have regularly been accused of failing to correctly assess credit risk over decades (see the quote), being blamed for aggravating the recent financial crisis by mis-reporting credit risk (especially on structured products) and missing to adjust ratings in a timely manner is yet another striking illustration of the relevance ascribed to them. In fact, these ascriptions are so sweeping that Vink and Fabozzi (2009) and Mählmann (2010) actually test whether spreads of asset-backed securities depend on nothing but ratings.6

For such accusations to be meaningful, however, it is critical that agencies claim to provide incremental information over and above what is already known by the markets. As mentioned, they do. Accordingly, to test the informational content of ratings, the finance literature has studied market-price reactions upon the announcement of rating changes, beginning with Katz (1974).7 While Katz himself found neither pre-announcement effects nor immediate reactions (probably due to econometric limitations of his time), as of Weinstein (1977, on bonds) and Pinches and Singleton (1978, on stocks) the literature has firmly established that rating changes are largely anticipated by the market via abnormal returns already as long as one year, sometimes even more, before the rating event.

Beginning with Griffin and Sanvicente (1982), and Holthausen and Leftwich (1986), studies also find ever more evidence for reactions in security prices upon downgrades; for upgrades results are generally weak and mixed. The effect has been documented for bond returns (Hite and Warga, 1997; May, 2010), for stocks8 (Matolcsy and Lianto, 1995; Dichev and Piotroski, 2001), for commercial paper (Nayar and Rozeff, 1994; Crabbe and Post, 1994), and also for credit-default swaps (Norden and Weber, 2004; Norden, 2008; Dilly and Mählmann, 2010). Moreover, it is stronger in reaction to global rating agencies as compared to local ones (Li et al., 2006), and already manifest for the preceding common warning procedure, i.e., upon announcements of outlook changes or additions to the watch list (Holthausen and Leftwich, 1986; Hand et al., 1992; Bannier and Hirsch, 2010).

In a particularly convincing study, Kliger and Sarig (2000) exploit the natural experiment of Moody’s surprise move to add modifiers to all ratings on April 26, 1982, i.e., to increase the precision of ratings overnight by improving the granularity of their rating scale. They show that debt value increases and equity value falls when Moody’s announces better-than-expected ratings, and vice versa; even implied volatilities of options measurably reflect the informational change.

---

6They do conclude, however, “that the notion of pure reliance on assigned ratings that has been popularized in the market may be overstated” (Vink and Fabozzi, 2009, p. 2).
7Prior studies like Pogue and Soldofsky (1969) or Pinches and Mingo (1973) focused on explaining ratings based on market data and thus, while quite successful at their task, could not draw conclusions about potential additional information content.
8In line with intuition about risk shifting, the effect does not arise, however, if the downgrade is not due to a deteriorating financial position but to increases in firm leverage (Goh and Ederington, 1993). Moreover, Odders-White and Ready (2006) document a relation also to common measures of equity-market liquidity.
Ederington and Goh (1998) relate rating events to revisions in earnings forecasts by stock analysts and find mutual Granger causality. Moreover, while both actual and forecast earnings fall following downgrades, no change in actual earnings is discernible after upgrades—while analysts do revise their forecasts upward.

In a similar vein, Jorion et al. (2005) analyze the effects of Regulation Fair Disclosure, implemented in 2000, which prohibits US public companies from making selective, non-public disclosures to favored investment professionals, but provided an exception for credit-rating agencies. They find the stock-price effect of downgrades and upgrades to be much greater in the post-regulation period, consistent with increased information content of ratings due to credit analysts having access to confidential information not publicly available.

To summarize, the literature has by now reached consensus that while most information about an issue’s creditworthiness is well anticipated by the markets, there remains a significant reaction in terms of abnormal returns following downgrades: the announcement effect.

At this point it is important to stress that this in itself does not imply market inefficiency: The reaction can well be the rational and accurate incorporation of an informative (albeit noisy and imprecise) signal provided by the rating agency’s announcement.

In fact, rating agencies commonly extensively request and obtain confidential information from a rated company’s management, including financial statements in greater detail, capital-spending plans, minutes of board meetings, and planned reactions to various scenarios (Jorion et al., 2005). The agency’s credit analysts meet and discuss regularly with the ratee’s management, and provide a preliminary assessment along with the opportunity to supply further information if the firm believes a better rating to be justified. Only after agreement is reached, the rating—based on all information obtained by the agency—is publicized, together with an explanation that exclusively relates to public information, to keep strict confidentiality on the rest.

From this point of view it is more puzzling why market pressures fail to induce agencies to provide more informative measures9 (for instance, at least ordinal), rather than why an announcement effect can obtain.

However, several recent empirical studies have produced findings which, if announcements purely uncover additional information, are puzzling or outright inconsistent. First, Johnson (2004) studies agencies’ actions around the investment-grade boundary. He documents that while ratings are no more volatile in the lowest investment-grade class, downgrades starting there tend to be more severe than those starting from higher or lower classes. Also, he shows that Egan-Jones Ratings, a smaller rating agency and at the time without the status of national recognition, tends to precede S&P rating changes by 21 to 70 days and even to predict their magnitude, although they do not obtain inside information (Beaver et al., 2006, footnote 5). Xia (2011) confirms this outperformance in terms of default rates, and finds a relation of rating infla-

---

9However, see Doherty et al. (2011) for a model of agency incentives leading to signals of little information
tion to proxies for incentive problems. More specifically, Rösch and Scheule (2011) demonstrate that Moody’s ratings of structured products underestimate impairment risk when fee revenue is high.

Second, there is evidence from unsolicited ratings, i.e., those ratings that agencies produce without being commissioned by the debt issuer (and at times against his expressed will, see Fulghieri et al. (2011) for a theoretical model). As a stylized fact, unsolicited ratings are lower than solicited ones (Poon, 2003; Poon and Firth, 2005; Fairchild et al., 2009; Poon et al., 2009), but as Gan (2004) and Bannier et al. (2010) show, this can be explained by self-selection and agencies’ conservatism when facing a lack of data.

The crucial results are thus those of Behr and Güttler (2008), who study stock-market reactions to changes in unsolicited ratings. While by definition unsolicited ratings are based on public information only, they do find a significant announcement effect. The same phenomenon was observable during the recent government-debt crisis, when it was doubtful that rating agencies could possess substantial additional information, yet their downgrade announcements moved markets strongly.

In light of these findings, therefore, this paper challenges the notion that re-ratings serve exclusively as information revelation. Apparently, the announcements per se are able to influence market prices. Unless we give up market efficiency, this implies the announcements themselves impact the respective obligors’ creditworthiness. Yet this is precisely what the rating reflects. A downgrade, insofar as it is a pure reflection of a given deteriorated credit condition, might thus feed back to additionally diminish creditworthiness even further. Denote this rating-induced default risk as the feedback effect. Recently I have become aware that Manso (2011) has provided a theoretical model of the effect. He uses it to study welfare implications of multiple equilibria. While in his model agencies do take the consequences of their rating announcements into account strategically, by construction all ratings are correct at all times. This effectively assumes agencies resist the arising incentive problem faultlessly. The aims and contributions of the present paper are to show how an equilibrium can exist where agencies do cave in to this incentive up to a degree (which is endogenous and chosen optimally), and to then exploit this insight to devise a test of their honesty in reporting.

There exist many reasons why such a feedback effect would arise. If contracts refer to credit ratings instead of continuous market-determined measures of credit risk, as many do, a downgrade can lead to a discrete jump in firm value due to numerous adverse developments for the obligor: from higher refinancing costs, increased procurement costs, more aggressive competitors due to the signal of higher exit likelihood, to reduced sales due to diminished goodwill, lower market value of service offers and guarantees, loss of human capital, and more. Finally, market-value changes will necessitate capital-structure adjustments, which in themselves can influence credit risk and liquidation policy (Titman, 1984).

The most obvious case is that of firms which have entered into agreements with rating triggers, i.e., clauses that imply additional obligations in the event

---

\(^{10}\)Standard and Poor’s actually denotes unsolicited ratings by a subscript “pi”, standing for “public information”.

of a downgrade. Such ‘credit puts’ can require significant amounts retired or posted as collateral upon a rating event; according to S&P (2004) they were, for instance, involved in the infamous demises of General American Life Insurance and Enron Corporation. See Bhanot and Mello (2006) for a theoretical exposition when and which triggers can be optimal.

Note that the feedback effect is agnostic about information content: it is consistent with both highly informative and purely uninformative announcements (e.g., a trigger on weather conditions). The point is that rating events per se impact an obligor’s credit risk, and not exclusively the other way around. This strengthens the results of this paper: I can demonstrate the incentive problem irrespective of whether agencies in fact do have an informational advantage.

In light of the literature, the main contribution of this paper is thus twofold: First, to model the feedback effect in order to illustrate that its presence can explain several puzzles in the credit-rating literature, first and foremost downgrade hesitation. I denote as such the frequently voiced concern that agencies tend to delay downgrades too long or potentially even omit warranted downgrades from reporting; phenomena also empirically recorded, for example by Altman and Rijken (2004), Johnson (2004), Löffler (2005) or Odders-White and Ready (2006). The second aim of this paper is to develop testable predictions which can be examined empirically in future research.

In contrast to prior literature, which acknowledges that it cannot differentiate dishonest behavior from conservatism (e.g., Bannier et al., 2010, footnote 3), the tests presented are designed to disentangle dishonesty from both honest mistakes and unobjectionable strategic conservatism.

3. The model

For clarity, I develop a simple 2-period, 2-state model to provide the basic economic intuition. In principle, the model can be extended to continuous states on the basis of a Merton (1974) model, keeping the assumption that bondholders act only at few discrete points in time for tractability. Subsequently, it is possible to further extend the model to cover also continuous time, for instance along the lines of Duffie and Lando (2001). Since the basic trade-off of the model and the major empirical predictions are already visible in the simpler case, I do not pursue that avenue in this paper.

3.1. Agents and timeline

In any setting, the following agents are relevant in a model of rating announcements: First, a rating agency (which maximizes profits), second, the bondholders (who value accurate information about credit risk), and third, a rated firm, run by equityholders.

While the firm owners buy the rating and choose to default or pay debtors as well as the agency, I assume they have no room for strategic interaction in order to abstract from the aspects of running the firm. This is analogous to the common assumption in models of capital structure that debt is competitively priced and in inelastic supply, where it serves to concentrate on equityholders.
Since ratings are of highest interest to creditors, I do the opposite: I pin equityholders to their reservation value and model the strategic choices of debtholders. The game thus reduces to two players: the agency and the bondholders.

The key idea of the paper is that in the presence of the feedback effect the agency faces an incentive to not report a downgrade in cases when it knows that it were warranted. Consequently, three points in time are required, as depicted in Figure 1: At initial $t_0$, the obligor is rated well. At intermediate $t_1$, credit risk has evolved and a downgrade might or might not be warranted, and might or might not be announced, even if warranted (i.e., the agency can implicitly pretend the good rating would still apply). If the agency acts upon its incentives, it will hesitate to announce downgrades. Bondholders observe the rating event, or its absence. They then decide whether to trust or distrust the agency. If they trust, and if there was no downgrade, they acquire additional debt of the firm. If they distrust, they do not increase their holding but rather monitor the firm in order to determine whether the agency has reported accurately. Finally, at $t_2$, uncertainty is resolved and payoffs are realized.

Note that to strengthen the focus on the problem of interest I let all information be common knowledge at $t_0$. This puts emphasis on the point that even with fully accurate ratings initially, a deterioration of the credit condition might go unannounced and therefore (for some time) unnoticed. For the same reason, I let all uncertainty be resolved at $t_2$. It strengthens the conclusions of the paper that they also hold when it is certain that all information will be revealed eventually.

3.2. Rating classes and the feedback effect

Again for simplicity, there are two rating classes: a good rating $A$, corresponding to high creditworthiness, and a worse rating $B$. Assume it is common knowledge that creditworthiness is measured in terms of default probability (PD). Put differently, rating $B$ refers to a higher PD than rating $A$.

Let the firm have low credit risk at $t_0$. This makes no statement about prevalence of credit quality, but otherwise a (further) worsening of default likelihood is not possible without additional rating classes. Since at this point in
time all information is common knowledge, the agency reports truthfully an $A$. Clearly, any leeway for agencies to misrepresent ratings from inception will, if they succumb to incentives to deviate from truthful reporting, only exacerbate the problem presented.

At time $t_1$, the credit quality of the firm may have deteriorated to a degree which would warrant an announcement of a downgrade to $B$, or alternatively have remained stable, further justifying the $A$. Clearly, the PD in the former state will be higher than in the latter; however, due to the feedback effect it will also depend on the agency’s announcement. Let $q$ be the commonly known probability of a deterioration from $t_0$ to $t_1$. Denote by $p^{AB}$ the probability of survival (one minus the PD) at $t_2$ of a firm in credit condition $A$ where the agency has announced $B$; and by $p^{AA}$, $p^{BA}$, and $p^{BB}$ the analogous survival probabilities.

Clearly, $p^{AA} > p^{BA}$ and $p^{AB} > p^{BB}$ by definition of credit risk; and moreover $p^{AA} > p^{AB}$ and $p^{BA} > p^{BB}$ due to the feedback effect. These relations are summarized in Table 1. Any default is assumed to occur later than $t_2$.

### Table 1: Survival rates of the firm depending on whether the credit condition has remained stable ($A$) or deteriorated ($B$) and on whether the agency has announced a downgrade ($B$) or refrained ($A$).

<table>
<thead>
<tr>
<th>Credit Condition</th>
<th>Agency Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$p^{AA}$</td>
</tr>
<tr>
<td>$B$</td>
<td>$p^{BA}$</td>
</tr>
<tr>
<td></td>
<td>$p^{AB}$</td>
</tr>
<tr>
<td></td>
<td>$p^{BB}$</td>
</tr>
</tbody>
</table>

3.3. The agency’s problem

As has been implicit so far, I consider a single rating agency.$^{11}$ The agency is endowed with a technology to observe the true asset value of the firm, so that it knows the true PD exactly. This assumption serves to preclude the agency from erring in good faith. Noise in the agency’s assessment would needlessly complicate the exposition while change none of the qualitative results as long as the agency obtains sufficiently valuable private information. However, when it assigns one of the two rating classes, the CRA can report a different creditworthiness than it observes. The users of ratings therefore take into account that despite the agency’s perfect technology, ratings may still be inaccurate.

$^{11}$Evidence of fierce competition in the rating market appears limited. This is not to imply that competition is irrelevant in this market. Competitive effects, however, are unclear and depend on detailed model assumptions. See Bolton et al. (2007) for a general model of financial intermediation where price competition interacts with the degree of truthfulness in information revelation. Doherty et al. (2011) present a model of credit-rating agencies where more competition increases the amount of information revealed. In contrast, Bolton et al. (2011) explicitly model competition of rating agencies and conclude that it may undermine market efficiency. Empirically, Becker and Milbourn (2010) study the entry of Fitch as a third accredited rating agency and find it coincides with lower-quality ratings from the incumbents and a deteriorated ability of ratings to predict default.
In principle, the agency has a simple business model: observe the PD, and provide this information by announcing a downgrade if the credit condition has worsened sufficiently, or (implicitly) confirm the good rating by making no announcement. The strategy set of the agency thus consists of the two actions *downgrade* and *refrain* from an announcement. Since this choice depends on the creditworthiness observed, there are two types of agency, depending on what the truly fair rating $F \in \{A, B\}$ would be: one has observed an obligor with risk corresponding to rating $A$, $F = A$; for the other one $F = B$. Denote by $R_t = \{A, B\}$ the rating subsequently “announced” explicitly or implicitly by a $t$-type agency, $t \in \{A, B\}$. Let $\rho_t = P(R_t = t)$ denote the probability with which the $t$-type agency reports truthfully rating $t$. Correct ratings are thus equivalent to $\rho_A = \rho_B = 1$.

The agency charges a fee for its service, which is flat across rating classes: Unless it defaults, a firm pays the same fee with no direct dependence on the rating assigned (a good rating is not more costly than a bad one).\(^{12}\)

Then why would the agency want to omit a downgrade? Due to the feedback effect a downgrade further increases the default probability; this is common knowledge, too. Thus a conflict of interest arises, since the agency maximizes profit. While the fee is constant across rating classes to preclude direct monetary incentives for rating inflation, it can only be charged if the obligor has not defaulted.

The value of the fee to the agency arises from two parts. On the one hand there may be a potential payment concurrent with the first acquisition of the rating. This payment is irrelevant for decisions thereafter, so I normalize it to zero. On the other hand, the agency profits from the income stream of maintenance fees and of continuation business, as for instance advisory services. Let $\phi_0$ denote the continuation value of the fee for the rating acquired at $t_0$, conditional upon the firm’s survival, as capitalized at $t_2$.

Moreover, in line with current market practice, an additional fee is due for the rating of additional securities. If the firm buys another rating for a second bond at $t_1$, then $\phi_1$ denotes the agency’s additional fee income, conditional upon survival, again capitalized at $t_2$.

From this contract, which I take as a given and reasonably commensurate description of such contracts in practice, there stems an indirect incentive to bias ratings: Due to the feedback effect, *not* reporting a warranted downgrade decreases the likelihood of a loss of fee income.

Finally, if the agency dishonestly reports an inaccurate rating and this fact is discovered, it incurs a reputation cost amounting to a payoff of $-c$. The threat of this cost is supposed to discipline the agency to honest reporting.

\(^{12}\)There are frequent claims to the contrary and the well-known conspicuous rating history of Hannover Re as recounted by Klein (2004) and debated in S.Hrg. 109-465 (2005), but the point is that even if it holds perfectly, there still remains a reason for the agency to prefer obligors’ survival, namely limited liability. For a view of how price discrimination across credit-rating classes could increase efficiency, see Jeon and Lovo (2010).
3.4. The bondholders’ problem

In contrast to the agency, bondholders\footnote{While I use the plural term, bondholders are assumed to act homogeneously to abstract from coordination problems.} do not possess the means to assess changes in credit quality of the firm during its operation in a timely manner. While at $t_0$ debt- and equityholders (as well as the rating agency) share perfect informational symmetry, bondholders cannot observe how the asset-value process of the firm evolves from there, and thus cannot assess the firm’s creditworthiness at $t_1$. This is commonly justified by assuming that equityholders have no way to credibly publish the relevant information without impairing the competitive position of the firm.

As a result of not knowing whether a given signal $R$ was sent by an $A$-type or a $B$-type agency, debtholders form contingent beliefs: They expect with probability $\alpha = \mathbb{P}(R_A = A)$ that the absence of a downgrade rightly indicates lower risk, and with $\beta = \mathbb{P}(R_B = B)$ a downgrade to be truthful.

In the long run, however, bondholders do learn the financial condition of the obligor. For simplicity, I model the uncertainty as fully resolved at $t_2$. The interpretation is a time lag in bondholders’ assessment of creditworthiness. If no information were ever to become verifiable, the common argument about reputation costs would be inapplicable since an indication to impose them would never arise. Moreover, financial health becomes evident at the latest when obligations come due and are met or defaulted on. Therefore, a role of information transmission for rating agencies must be inter-temporal in nature. While in an ongoing long-term relationship the CRA can possess at each point in time information that bondholders will learn only later, a two-period model requires the value of information to stem from a reduction of interim uncertainty. Assuming symmetry of information at $t_2$ is therefore the most clear-cut case.

Without loss of generality, assume bondholders find it incentive compatible to provide a notional of 1 to the firm at $t_0$. Then they ascribe value to the rating at $t_1$ because they can take it into account in their decision whether to provide additional capital to the firm at $t_1$. Let a correct low rating $B$, as well as the absence of credit information (i.e., the unconditional expectation) correspond to levels of credit risk at which bondholders do not wish to increase their exposure, while with a correct $A$ rating they do. Then the notional amount of the follow-up bond is inconsequential and thus also set to a payoff of 1 at $t_2$, conditional upon the firm’s survival. Denote the price for this second bond as $\eta p^{AA}$ with $\eta < 1$ and $\eta p^{AA} > p^{BA}$\footnote{Note that the price can be at most $p^{AA}$ with fully credible ratings, and needs to be larger than $p^{BA}$ to preclude the follow-up bond from being a bargain even with fully uninformative ratings.}

However, bondholders are aware of the CRA’s incentive problem and consequently doubtful of ratings. They face the strategy choice to trust or distrust them, $S_t = \{T, D\}$, depending on the observed rating $t$. If they trust, they will finance the second bond as long as no downgrade occurred, otherwise not. They will, however, not be able to verify the CRA’s announcement and hence be unable to impose costs on the agency. On the other hand, if they distrust the agency, they will not fund an increase in the firm’s debt level, but spend a
small cost $\epsilon$ to verify the announcement by $t_2$ in order to be able to impose the reputation cost $c$ on the agency if it mis-reported. This mechanism should induce the agency to resist the conflict of interest. More inaccurate ratings discipline the agency in two ways: first, they are more likely to be distrusted, and second, once distrusted, lead to higher expected reputation losses.

The assumption of an informationally relevant rating agency then amounts to the following.

**Assumption 1 (Valuable Ratings).** There exists a potential informational role for ratings, since a credible good rating can produce a gain of $\eta p^{AA}$ from the follow-up bond for debtholders, while its unconditional expectation is negative and worth less than paying the verification cost:

$$
(1 - q)(1 - \eta)p^{AA} + q (p^{BA} - \eta p^{AA}) < -\epsilon < 0. \tag{1}
$$

This assumption ensures that the verification cost is small enough to not incentivize bondholders to buy the follow-up bond just in order to avoid verifying. Put differently, an uninformative rating implies a too large risk of overpaying for a bad-quality bond, so that it is preferable to monitor the rating agency.

The second technical condition necessary to preclude degenerate outcomes is that the reputation costs must be sufficient to not become irrelevant to the rating agency. After all, due to the feedback effect an honest downgrade induces a loss in expected fee income, so the agency is willing to accept some losses to avoid this reduction. The following assumption rules out that even certain reputation costs are irrelevant.

**Assumption 2 (Sufficiency of Reputation Costs).** The potential reputation loss exceeds the reduction in the present value of initial fees which is due to the feedback effect,

$$
c > (p^{BA} - p^{BB}) \phi_0. \tag{2}
$$

The assumption implies that an agency which knows it will be verified finds it optimal to rate truthfully.

Denote by $\delta_t = \mathbb{P}(S_t = D)$ the probability with which the bondholders choose to distrust given they observe a rating of $t$, and by $\gamma_t$ the agency’s belief about $\delta_t$ for $t \in \{A, B\}$.

### 3.5 Payoffs

Denote by $\Pi_{ag}(F, R_F, S_{R_F})$ and by $\Pi_{bh}(F, R_F, S_{R_F})$ the payoffs to the agency respectively the bondholders if the warranted, fair rating is $F$, the credit-rating agency reports it as $R_F$, and the bondholders subsequently choose the strategy $S_{R_F}$.

---

15 The cost of verification $\epsilon$ can be interpreted as a percentage of the initial bond’s notional, since that is normalized to 1.
Then the agency’s conditional payoffs amount to
\[
\Pi_{ag}(F, R_F, S_{RF}) = p^{FR_F} \left[ \phi_0 + \phi_1 \mathbb{I}_{\{R_F = A \land S_{RF} = T\}} \right] - c \mathbb{I}_{\{F \neq R_F \land S_{RF} = D\}},
\] (3)
where the first term stems from the fees, initial and potential follow-up ones, and the second term denotes the reputations costs; \(\mathbb{I}_{\{E\}}\) denotes the indicator function for a random event \(E\) which equals 1 if \(E\) occurs and 0 otherwise.

The bondholders’ ultimate payoffs add up to
\[
\Pi_{bh}(F, R_F, S_{RF}) = p^{FR_F} + \left( p^{FR_F} - \eta p^{AA} \right) \mathbb{I}_{\{R_F = A \land S_{RF} = T\}} - \epsilon \mathbb{I}_{\{S_{RF} = D\}},
\] (4)
with the first term capturing the value of the initial bond, the second term that of a follow-up purchase, and the final third one the verification costs, if chosen to be incurred.

4. Equilibrium

In principle, given two states of nature (the obligor’s condition deteriorates or remains stable) and two strategy choices by each player (downgrade or refrain for the agency, trust or distrust for bondholders), the game has eight final nodes. However, it can be proven that a downgrade announcement is fully credible and thus the game simplifies to six feasible outcomes. Subsequently, I show that the agency will not report fully truthfully for any level of reputation costs, as the respective pure strategies are not incentive compatible. Finally, I derive and characterize the unique Perfect Bayesian Equilibrium of the game.

**Lemma 1** (Trusted Downgrades). If the agency reports a downgrade, bondholders trust with certainty, \(R_t = B \Rightarrow \delta_B = 0\) for \(t \in \{A, B\}\).

**Proof.** See Appendix.

The intuition is that after a downgrade the additional investment will not be undertaken, regardless of trust or distrust, and the obligor’s probability of default is the same for both strategies. Without upside potential, the verification cost \(\epsilon\) renders distrust a dominated strategy, regardless of the degree of truthful reporting by the agency.

**Lemma 2** (Truthful Downgrades). If the agency observes a good credit condition, no downgrade is announced, \(F = A \Rightarrow \rho_A = 1\).

**Proof.** See Appendix.

There is no incentive for the agency to deflate ratings, since it would be worse off on two fronts: First, a downgrade precludes with certainty income from the rating of the second \(t_1\)-bond. Second, at the same time it decreases the obligor’s survival probability due to the feedback effect, irrespective of bondholders’ strategy choice. On the other hand, if no announcement is made the
additional $\phi_1$ can be earned with probability $1 - \delta_A \geq 0$, and even if distrusted no loss of reputation will arise.

With these two lemmas, the game tree is reduced, since the $A$-type agency always plays the pure strategy $\rho_A = 1$, independently of the bondholders’ strategy. The extensive-form representation of the reduced game is depicted in Figure 2.

The question of accurate ratings simplifies to the choice of $\rho_B$. For brevity, define $\rho = \rho_B$.

In this setup, rating accuracy corresponds to a separating equilibrium where the $B$-type agency plays the pure strategy to always downgrade. However, the following theorem shows that no separating equilibrium exists.

**Theorem 1** (Impossibility of Fully Accurate Ratings). In equilibrium, the agency does not report all warranted downgrades, $\rho < 1$.

**Proof.** See Appendix.

This major result follows from a simple intuition: If ratings were announced perfectly accurately, bondholders would find it optimal to never distrust and hence never verify their accuracy. But then reputation costs would never be imposed—and without any disciplining threat to counterweigh the feedback effect, the agency strictly prefers to omit at least some fraction of downgrades.

**Corollary 1.** Reputation costs can exceed the total sum of potential fee income, and participation of the agency is guaranteed, as long as it is so without dishonest reporting.

**Proof.** See Appendix.
Since reputation costs are incurred probabilistically, an upper limit of their likelihood is endogenous to the rating agency. Therefore, if positive expected agency profits motivate the credit-rating agency to enter the market in the absence of dishonest reporting, the agency can contain the risk of reputation costs by adjusting the truthfulness of reporting. Consequently, if the rating business is at all viable with honest reporting, reputation costs cannot induce the agency to exit the market.

**Corollary 2.** For any finite amount of reputation costs, the agency will maintain a strictly positive probability of omitting warranted downgrades.

*Proof. See Appendix.*

In extension of the reasoning above, the agency can accommodate ex-ante any amount of reputation costs by limiting the risk of having to bear them. It can effectively endogenize the expected burden via its choice of reporting strategy: For given costs \( c \), as \( \rho \to 1 \) their expected value drops to zero.

**Theorem 2** (Impossibility of Fully Uninformative Ratings). In equilibrium, the agency does report warranted downgrades with strictly positive probability, \( \rho > 0 \).

*Proof. See Appendix.*

While Theorem 1 in conjunction with Lemma 2 proved that no separating equilibrium exists, Theorem 2 together with the same lemma shows that no pooling equilibrium exists, either. While the \( A \)-type agency plays the pure strategy of always refraining, the \( B \)-type agency plays a mixed strategy. And since any pure strategy by the bondholders would induce a pure strategy also by the \( B \)-type agency, it follows that in equilibrium bondholders choose a mixed strategy \( 0 < \delta_B < 1 \).

**Theorem 3** (Perfect Bayesian Equilibrium). The unique Perfect Bayesian Equilibrium consists of bondholders’ strictly mixed strategy \( (\delta_A, \delta_B) = (\delta^*, 0) \) and their assessment of \( (\alpha, \beta) = (\alpha^*, 1) \) with the agency’s strategy \( (\rho_A, \rho_B) = (1, \rho^*) \) and assessment \( (\gamma_A, \gamma_B) = (\gamma^*, 0) \), where

\[
\delta^* = \frac{(p^{BA} - p^{BB})\phi_0 + p^{BA}\phi_1}{c + p^{BA}\phi_1},
\]

\[
\alpha^* = \frac{1 - q}{1 - q\rho^*},
\]

\[
\rho^* = \frac{q p^{AA} - \epsilon}{q(p^{AA} - \epsilon)},
\]

\[
\gamma^* = \delta^*,
\]

with the agency’s equilibrium payoff of refraining to announce a warranted downgrade amounting to

\[
\pi^* = \delta^*\Pi_{ag}(B, A, D) + (1 - \delta^*)\Pi_{ag}(B, A, T)
\]

\[
= p^{BA}\phi_0 + (1 - \delta^*)\phi_1 - \delta^*c.
\]
Proof. See Appendix.

5. Empirical Implications

While the model presented shows that the feedback effect is sufficient to induce a conflict of interest within credit-rating agencies by providing an incentive to announce warranted downgrades only hesitatingly, and that reputation costs cannot be relied upon to prevent mis-reporting no matter how high they could realize, the question whether agencies succumb to biased reporting is ultimately an empirical one.

To assess it, the model provides several testable implications to be addressed in econometric analysis, which are detailed next. The design of the empirical strategy is elaborated on subsequently.

5.1. Empirical Predictions

The following relations can be derived from Theorem 3.

**Prediction 1.** A stronger feedback effect increases the extent of hesitation.

The main implication clearly relates the extent of the downgrade-induced default risk to the prevalence of delayed or omitted downgrades.

Note that while in a two-period model only omission of downgrades can be captured, for empirical tests I contend it is reasonable to interpret hesitation both as the failure to report a downgrade which is known to be warranted, as well as an undue delay in a rating event which eventually is announced, yet only after a period longer than would be justified by the objective of rating stability alone.

This prediction also implies how to achieve identification in disentangling downgrade hesitation from a pure desire for rating stability: It appears hard to argue that the latter should exhibit a systematic relation to the sensitivity to rating-induced default risk.

**Prediction 2.** Increases in reputation costs reduce downgrade hesitation.

As $c \to \infty$ both $\delta \to 0$ and $\rho \to 1$. While for any finite $c$ some hesitation prevails, time periods of heightened scrutiny of agency behavior, as for instance after crises or rating scandals, are predicted to reduce the prevalence of strategically delayed downgrades.

**Prediction 3.** More profitable continuation business with the ratee induces stronger downgrade hesitation.

A higher present value of the continued business relationship with the ratee increases the agency’s interest in the firm’s survival via two distinct channels: First, a lower default probability will increase the expected value of maintenance fees and complementary deals stemming from the original contract. At the same time, it also increases the likelihood of additional rating deals, which require that the firm survives.
Prediction 4. *Opaque firms are more likely to be subject to omitted or postponed downgrades.*

Since the bondholders credible threat of verifying the agency’s announcement in order to be able to detect mis-reporting and subsequently impose the reputation costs $c$ is discouraged with increasing $\epsilon$, the agency will become bolder in omitting or delaying rating events with higher cost of credit-condition verification, since it takes into account the decreased likelihood of being discovered.

Prediction 5. *Firms with higher risk of deteriorating credit condition obtain less informative ratings.*

The parameter $q$ enters considerations of bondholders since it determines the likelihood that an observed absence of a downgrade is due to a stable financial condition, and thus correct reporting by the agency. The prediction follows from the fact that the probability $\alpha$ can also be interpreted as a measure of information content of an (implicit) A rating.

5.2. *Identification Strategy*

To assess these predictions based on empirical data, the following approach is devised: First, estimate an obligor-specific measure of downgrade hesitation using the time dimension of rating histories. Second, exploit the cross-section of ratees to relate the hesitation estimates to firm characteristics proxying for the variables in the predictions.

The strength of this design lies in the fact that it accommodates the possibility of a time lag in downgrades due to credit-rating agencies’ desire for through-the-cycle ratings. Due to this stated objective, the magnitude of any hesitation measure per se is uninformative about incentive problems, but may merely reflect the importance ascribed to stable ratings.

On the other hand, it is also necessary to allow for agency incompetence. If ratings are inaccurate, this in itself also does not imply any misdemeanor, as it could simply be due to the fact that the technology employed does not allow for a better assessment of credit risk. Any judgment about a conflict of interest requires to assess not the information content of ratings, but existence or inexistence of resulting biases from incentive problems.

Relating the hesitation measure to the cross section of firm characteristics tackles both of these two challenges: If the time lag in announcements is due to noise, it should not be related to proxies for incentive problems. Similarly, if it is due to a general tendency to not react too fast to changes in creditworthiness, this delay again should be unrelated to the extent of downgrade-induced default risk. A significantly longer delay for firms that are relatively more profitable to the agency as compared to firms giving rise to a smaller incentive problem would mark a regularity that is hard to reconcile with objectivity of agencies and sufficient discipline induced by reputation costs.

To identify hesitation, it is necessary to relate daily rating histories to an alternative measure of credit risk which is free from an incentive to delay recognition of firms’ worsened financial health. I propose using CDS spreads as a benchmark of default risk against which to measure timeliness of rating events.
Figure 3: Actual agency rating and market-implied rating for one obligor (CTX). The volatile solid line depicts the running CDS spread for the debtor as quoted by Markit, while the solid step function shows its actual rating history by Standard & Poor’s (on a numerical scale where AAA is 1 and C is 21). From a panel of CDS data, market-implied ratings can be calibrated for each reference name; the corresponding time series for CTX is indicated by the dashed step function. As known from the literature, changes of CDS-implied ratings tend to lead agency re-ratings: this time gap can be taken as a proxy for downgrade hesitation.

Alternatively, also bond prices are well suited to the analysis, but it appears particularly reasonable to assume that CDS-market participants have an interest in a timely account of PD changes, independently of the extent of the feedback effect, the continuation value of an obligor’s business to rating agencies, or their prospective reputation costs.

Consequently, downgrade hesitation can be estimated via market-implied ratings (Breger et al., 2003). Deriving ratings from CDS spreads is known to yield rating changes which tend to anticipate changes by agency ratings well in advance, as shown by Kou and Varotto (2008) or Dilly and Mählmann (2010). Figure 3 illustrates how this time lag provides a natural and obligor-specific proxy for downgrade hesitation.

The second component is the strength of the feedback effect, which can be proxied for empirically by measures of sensitivity to rating announcements. These include the percentage of short-term debt, the reliance on debt financing, access to public debt markets, or in particular a credit rating at the boundary between investment and speculative grade.

Finally, the hesitation estimates, which are derived from the time-series di-
mension of the CDS panel data, can be related to the cross-sectional dimension of the proxies for incentive problems. Any systematic relation between those would be consistent with claims about the conflict of interest causing a bias in the timing of downgrade decisions or alternatively require a story going beyond rating stability (through-the-cycle ratings) as well as beyond good-faith errors and unpredictability.

6. Conclusion

Credit-rating agencies have been accused both of biased reporting and of incompetence due to ratings that are perceived as faulty too often—not only since the global financial crisis, but for decades. Due to the limited information content in purely ordinal ratings and the declared intent to keep them stable through the cycle, this claim is non-trivial to judge. The common justification to date has relied on reputation costs to discipline rating agencies’ reporting.

This paper presents a model that shows that reputation costs are insufficient to ensure accurate ratings as soon as the credit-rating agency has an interest in the ratee’s financial condition. The reason is that the risk of incurring reputation costs is endogenous to agency choice, and thus—while indeed incentivizing more accurate rating reports—will be traded off against the agency’s financial interest in the obligor. Such an interest arises from the feedback effect: the fact that downgrades not only reflect, but also impact credit risk. Since continued fee service is lost when the obligor defaults, a strategic agency will act upon its incentive to hesitate with announcements of deteriorated financial health.

An important insight is that the amount of reputation costs is inconsequential for the result on inaccurate ratings. Even if they exceed the present value of the entire rating business, the agency will find it optimal to risk incurring them with a strictly positive probability in equilibrium.

The main contribution of the paper, however, is to derive a strategy for empirically assessing the claims of misdemeanor of agencies. Downgrade hesitation provides testable implications that can be exploited to shed light on how far CRAs succumb to incentives for dishonest (omission of) downgrading.

The basic idea of the tests is to first use the time-series dimension to assess obligor-specific hesitation estimates. These can be obtained by considering rating events in relation to a benchmark of creditworthiness which is free from incentives to delay concession of worsening conditions. In a second step the hesitation measure can be related cross-sectionally to obligor characteristics which emerge from the theoretical model as related to the incentive problem.

Providing empirical evidence of how far agency ratings indeed exhibit the corresponding patterns clearly calls for further research currently under way. Moreover, the magnitude and properties of the feedback effect itself are both of high academic and practical relevance. Finally, the feedback effect and its model in this paper entail a number of further implications for empirical tests, and thus provide an avenue to answer several long-standing questions in the CRA literature—most prominently, the information content of credit ratings.
Appendix A. Proofs of Lemmas

Lemma 1 (Trusted Downgrades). If the agency reports a downgrade, bondholders trust with certainty, \( R_t = B \Rightarrow \delta_B = 0 \) for \( t \in \{ A, B \} \).

Proof. Since both
\[
\Pi_{bh}(A, B, D) = p^{AB} - \epsilon < p^{AB} = \Pi_{bh}(A, B, T)
\]
and
\[
\Pi_{bh}(B, B, D) = p^{BB} - \epsilon < p^{BB} = \Pi_{bh}(B, B, T),
\]
as long as there exist some verification costs \( \epsilon > 0 \), the bondholders will find it optimal, for any \( \beta \in [0, 1] \), to choose \( S_t = T \) for \( t \in \{ A, B \} \).

Lemma 2 (Truthful Downgrades). If the agency observes a good credit condition, no downgrade is announced, \( F = A \Rightarrow \rho_A = 1 \).

Proof. Given that \( \delta_B = 0 \), since both
\[
\Pi_{ag}(A, B, T) = p^{AB} \phi_0 < p^{AA} (\phi_0 + \phi_1) = \Pi_{ag}(A, A, T)
\]
and
\[
\Pi_{ag}(A, B, T) = p^{AB} \phi_0 < p^{AA} \phi_0 = \Pi_{ag}(A, A, D)
\]
the agency will, for any \( \delta_A \in [0, 1] \), refrain from announcing a downgrade, as that would provide no benefit but both hurt the existing claim on fees via the feedback effect and potentially lose additional business.
Appendix B. Proofs of Theorems and Corollaries

Theorem 1 (Impossibility of Fully Accurate Ratings). *In equilibrium, the agency does not report all warranted downgrades, \( \rho < 1 \).

*Proof.* By contradiction: Note that if the theorem were not true, all warranted downgrades would truthfully be announced, and since due to Lemma 2 no unwarranted downgrades are released, the situation would amount to a separating equilibrium. Assume thus that a separating equilibrium exists. Then \( \rho = 1 \), and hence the consistent belief of bondholders \( \beta = 1 \). Their best response is the pure strategy \( S = T \). Under the agency’s consistent belief of \( \gamma_B = 0 \), however, this condition will lead to deviating behavior, since

\[
\Pi_{ag}(B, B, T) = p^{BB}\phi_0 < p^{BA}\phi_0 < p^{BA}(\phi_0 + \phi_1) = \Pi_{ag}(B, A, T),
\]

where the first inequality follows from the feedback effect and the second from the increase in rating demand. This prevents \( \rho = 1 \) from being incentive compatible, and thus there can be no separating equilibrium. \( \square \)

Corollary 1. *Reputation costs can exceed the total sum of potential fee income, and participation of the agency is guaranteed, as long as it is so without dishonest reporting.*

*Proof.* Fix some costs \( \bar{\bar{c}} > \phi_0 + \phi_1 \). For the agency type observing the good condition \( A \), \( \bar{\bar{c}} \) is irrelevant, since by Lemma 1 they will never be incurred. The agency type observing a deteriorated condition \( B \) faces a choice between the uncertain payoff \( (p^{BA}(\phi_0 + \phi_1), p^{BA}\phi_0 - \bar{\bar{c}}) \), and the certain \( p^{BB}\phi_0 \), so its payoff is bounded below by \( p^{BB}\phi_0 \), which is its payoff in the case of perfect honesty. Hence, if the rating industry is viable at all with honest ratings, reputation costs cannot alter the agency’s participation decision. \( \square \)

Corollary 2. *For any finite amount of reputation costs, the agency will maintain a strictly positive probability of omitting warranted downgrades.*

*Proof.* Fix any finite amount of reputation costs \( \bar{\bar{c}} \) and apply the proof of Theorem 1. The reason why it still goes through is that in the counterfactual case of perfect honest, the extent of these costs is irrelevant, as they never need to be borne. Consequently, they do not figure in Equation (15) and thus cannot prevent the separating equilibrium from failing incentive compatibility. Note from Theorem 3, however, that as \( c \to \infty \), it does hold that \( \delta^* \to 0 \); so while reputation costs are insufficient to ensure honesty, they are not ineffective. \( \square \)

Theorem 2 (Impossibility of Fully Uninformative Ratings). *In equilibrium, the agency does report warranted downgrades with strictly positive probability, \( \rho > 0 \).*
Proof. By contradiction: Assume ρ = 0, so bondholders face a fully uninformative rating which is never downgraded. Note that uninformative ratings are equivalent to a pooling equilibrium, and since it was shown in Lemma 2 that the agency will not deflate ratings by announcing downgrades for firms in good condition, we only need to check the one where both agency types refrain.

Assume thus that such a pooling equilibrium exists. As a consequence, the bondholders’ consistent beliefs must be α = (1 − q), and they cannot answer with trust, as their expected payoff when trusting

\[(1-q)\Pi_{bh}(A, A, T) + q\Pi_{bh}(B, A, T) = (1-q)p^{AA} + qp^{BA} = (1-q)(1-\eta)p^{AA} + q(p^{BA} - \eta p^{AA}) < (1-q)p^{AA} + qp^{BA} - \epsilon = (1-q)\Pi_{bh}(A, A, D) + q\Pi_{bh}(B, A, D),\]

where the inequality follows from Assumption 1, so that they are strictly better off if they distrust whenever no downgrade is announced by an uninformative agency.

Given, however, bondholder’s distrust, the B-type agency will be found out and have to bear reputation costs if it mis-reports, and since

\[\Pi_{ag}(B, B, D) = p^{BB}\phi_0 > p^{BA}\phi_0 - c = \Pi_{ag}(B, A, D),\]

where the inequality is ensured by Assumption 2, the agency strictly prefers to truthfully downgrade the deteriorated rating—contradicting the assumption and completing the proof.

**Theorem 3 (Perfect Bayesian Equilibrium).** The unique Perfect Bayesian Equilibrium consists of bondholders’ strictly mixed strategy \((\delta_A, \delta_B) = (\delta^*, 0)\) and their assessment of \((\alpha, \beta) = (\alpha^*, 1)\) with the agency’s strategy \((\rho_A, \rho_B) = (1, \rho^*)\) and assessment \((\gamma_A, \gamma_B) = (\gamma^*, 0)\), where

\[\delta^* = \frac{(p^{BA} - p^{BB})\phi_0 + p^{BA}\phi_1}{c + p^{BA}\phi_1},\]

\[\alpha^* = \frac{1 - q}{1 - qp^*},\]

\[\rho^* = \frac{\eta p^{AA} - \epsilon - (1-q)(p^{AA} - p^{BA})}{q(\eta p^{AA} - \epsilon)},\]

\[\gamma^* = \delta^*,\]

with the agency’s equilibrium payoff of refraining to announce a warranted downgrade amounting to

\[\pi^* = \delta^*\Pi_{ag}(B, A, D) + (1 - \delta^*)\Pi_{ag}(B, A, T) = p^{BA}\phi_0 + (1 - \delta^*)\phi_1 - \delta^* c.\]
Proof. Lemmata 1 and 2 have already established that $\delta_B = 0$ and $\rho_A = 1$, and consistency of beliefs requires $\beta = 1$ and $\gamma_B = 0$ in any equilibrium.

Theorems 1 and 2 have then shown that no equilibrium in pure strategies exists, neither pure nor separating. By the existence theorem of Kreps and Wilson, however, a sequential equilibrium, and hence a PBE, must exist.

To find the equilibrium in mixed strategies, note that the $B$-type agency will only play a mixed strategy if, given the equilibrium strategy $\delta^*$ of bondholders, it is indifferent between downgrading and refraining, i.e.,

$$\Pi_{ag}(B, B, T) = \delta^* \Pi_{ag}(B, A, D) + (1 - \delta^*) \Pi_{ag}(B, A, T).$$

Inserting the payoffs and re-arranging leads to Equation (18), and Equation (21) follows immediately from consistency of beliefs.

Furthermore, the bondholders’ belief $\alpha^*$ must follow from Bayes’ law given the probability of deterioration $q$ and the agency’s equilibrium strategy, so that

$$\alpha^* = \frac{(1-q)1}{(1-q)1 + q(1-\rho^*)},$$

yielding Equation (19).

Finally, the bondholders will find a mixed-strategy $\delta^*$ optimal as long as they are indifferent between trusting and distrusting, conditional upon observing no downgrade and the agency’s equilibrium $\rho^*$. This implies

$$\alpha \Pi_{bh}(A, A, T) + (1 - \alpha) \Pi_{bh}(B, A, T) =$$
$$\alpha \Pi_{bh}(A, A, D) + (1 - \alpha) \Pi_{bh}(B, A, D),$$

which is equivalent to

$$\alpha ((2 - \eta)p^{AA}) + (1 - \alpha)(2p^{BA} - \eta p^{AA}) =$$
$$\alpha(p^{AA} - \epsilon) + (1 - \alpha)(p^{BA} - \epsilon),$$

from where it follows by re-arranging terms that

$$\alpha = \frac{\eta p^{AA} - \epsilon}{p^{AA} - p^{BA}}.$$

Now we can invoke Equation (19) to obtain that in equilibrium

$$(1 - q)(p^{AA} - p^{BA}) = (1 - q \rho^*)(\eta p^{AA} - \epsilon),$$

and after re-arranging find Equation (20).

The agency’s equilibrium payoff then results from simple substitution. \qed
References


