

Loerke, Petra; Niedermayer, Andreas

**Working Paper**

## Crises and rating agencies: On the effect of aggregate uncertainty on rating agencies' incentives to distort ratings

SFB/TR 15 Discussion Paper, No. 521

**Provided in Cooperation with:**

Free University of Berlin, Humboldt University of Berlin, University of Bonn, University of Mannheim, University of Munich, Collaborative Research Center Transregio 15: Governance and the Efficiency of Economic Systems

*Suggested Citation:* Loerke, Petra; Niedermayer, Andreas (2015) : Crises and rating agencies: On the effect of aggregate uncertainty on rating agencies' incentives to distort ratings, SFB/TR 15 Discussion Paper, No. 521, Sonderforschungsbereich/Transregio 15 - Governance and the Efficiency of Economic Systems (GESY), München

This Version is available at:

<https://hdl.handle.net/10419/121381>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



GOVERNANCE AND THE EFFICIENCY  
OF ECONOMIC SYSTEMS  
**GESY**

Discussion Paper No. 518

**Crises and Rating Agencies:  
On the Effect of Aggregate Uncertainty  
on Rating Agencies' Incentives to  
Distort Ratings**

Petra Loerke\*  
Andreas Niedermayer\*

\* University of Mannheim

September 2015

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · [www.sfbtr15.de](http://www.sfbtr15.de)  
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München  
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

---

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,  
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

# Crises and Rating Agencies: On the Effect of Aggregate Uncertainty on Rating Agencies' Incentives to Distort Ratings\*

Petra Loeke<sup>†</sup>      Andras Niedermayer<sup>‡</sup>

February 16, 2015

## Abstract

We analyze a rating agency's incentives to distort ratings in a model with a monopolistic profit maximizing rating agency, a continuum of heterogeneous firms, and a competitive market of risk-neutral investors. Firms sell bonds, the value of a firm's bond is known to the firm and observable by the agency, but not by buyers. Firms can choose to get a rating. The rating agency can reveal a signal of arbitrary precision about the quality of the bond. In contrast to the existing literature, we allow aggregate uncertainty. As in the existing literature, one rating class is optimal. However, the rating agency does not choose a socially optimal cutoff: the agency is more likely to be too lenient if the distribution of aggregate uncertainty has a lower mean, a higher variance, and is more left skewed. It is more likely to be too strict if the opposite holds.

**Keywords:** Rating agencies, certification, aggregate uncertainty

**JEL-Classification:** C72, D42, D82, G20

---

\*We want to thank Dirk Hoffmann, Raphaël Levy, Leonard Nakumura, Martin Peitz, André Stenzel, Joel Shapiro, Ernst-Ludwig von Thadden, Thomas Tröger, and participants of IIOC 2011 in Boston, EARIE 2011 in Stockholm, the SFB TR 15 Meeting 2011 in Tutzing, the MaCCI Competition and Regulation Day 2014 in Mannheim and seminars at the University of Mannheim and at Concordia University for their helpful comments. The authors acknowledge financial support from the Deutsche Forschungsgemeinschaft through SFB-TR 15.

<sup>†</sup>Economics Department, University of Mannheim, L7, 3-5, D-68131 Mannheim, Germany.  
Email: ploerke@rumms.uni-mannheim.de

<sup>‡</sup>Economics Department, University of Mannheim, L7, 3-5, D-68131 Mannheim, Germany.  
Email: aniederm@rumms.uni-mannheim.de

# 1 Introduction

Ratings and other quality certifications by third parties play an important role in today's economy. For instance, the volume of rated debt issues was over \$8,000 billion in 2006. Ratings are used by investors to guide their investment decisions. They are also crucial for financial regulation: Basel III includes ratings as one criterion for the calculation of the capital adequacy requirements for banks. So does the Solvency II Directive of the European Union, passed on March 11, 2014, which harmonizes insurance regulation in the European Union and is scheduled to come into effect on January 1, 2016.

However, ratings as a basis of regulation have been viewed controversially, especially after the financial crisis. The major concern is that the ratings used for regulation are given by rating agencies, which may have an incentive to distort ratings in order to maximize profit. These concerns have been expressed both in public policy debates and in investigations and lawsuits against rating agencies.<sup>1</sup> As a reaction to this concern, Section 939A of the Dodd-Frank Act (effective since 2010) requires that all federal agencies “must remove any reference to or requirement of reliance on credit ratings”.

The current article addresses the question of incentives to distort ratings by a profit maximizing rating agency under particular consideration of aggregate uncertainty. Aggregate uncertainty plays a major role in many markets. As an example, for subprime mortgages the question was not only how good the subprime mortgages were that one particular financial institution invested in. The question was whether subprime mortgages as a whole were a sufficiently safe investment.

To investigate the effect of aggregate uncertainty on incentives to distort, we consider a model in which all other possible incentives to distort are shut down. In particular, we consider a monopolistic rating agency that can credibly commit

---

<sup>1</sup>As the most recent example, see the settlement between Standard's and Poors, the U.S. Department of Justice, and 19 states attorneys general that includes a payment of \$1.5 billion by S&P in the context of the rating of securities backed by subprime mortgages. While “S&P did not admit to any violations of law”, the rating agency “did sign a statement of facts acknowledging that its executives in 2005 delayed implementing new models that produced more negative ratings” (Viswanatha and Freifeld, 2015).

to a rating strategy in a one period model. This shuts down effects such as forum shopping, renegeing on the ratings strategy, or reputational cycles.

Besides the rating agency there is a continuum of sellers selling bonds. There is a continuum of investors seeking to buy bonds. The mass of investors is larger than the mass of sellers, so that competition leads to prices being bid up to the expected value of a bond. The quality of a seller's bond is perfectly known to the seller, but unknown to investors. The rating agency has a technology to perfectly observe the seller's quality. Sellers can decide whether they want to be rated. The aggregate distribution of sellers' types is initially unknown to all market participants, except for a common prior about the distribution of the aggregate states of the world. The states of the world differ by a different aggregate distribution of sellers' types. After sellers get rated, the aggregate state of the world is revealed to all market participants and investors buy the bonds. The price depends on the expected quality in a rating class for the realized aggregate state of the world.

We show that in accordance to the existing literature, a profit maximizing rating agency will choose a coarse binary rating: either investment grade or junk bonds. However, in sharp contrast to the existing literature, aggregate uncertainty leads to the cutoff not being at the socially optimal level. Whether the rating agency has an incentive to be too lenient (a negative cutoff) or too strict (a positive cutoff) is pinned down by three moments of the aggregate belief distribution. The aggregate belief distribution is defined as follows: Take for every state of the world the mean quality of bonds that would be bought if quality were publicly known. Market participants' belief distribution of these means is the aggregate belief distribution. The rating agency has more of an incentive to be too lenient if the distribution has a low mean, a high variance, and a low higher order skewness (defined as the sum of the third and higher moments). A low higher order skewness can be thought of as a left skewed distribution, i.e. with a high probability bonds have a mean quality above average, but the distribution has a fat tail at the bottom which implies that with a small probability bonds have a very low mean quality. The opposite result holds for a larger mean, lower variance, and a larger higher order skewness. These

results can be interpreted as two opposite effects on the rating agency's incentive to distort ratings. One effect is procyclical: they have an incentive to be too lenient before the outbreak of a crisis (interpreting this period as a period with a large variance and left skewness of aggregate uncertainty) and an incentive to be too strict after the outbreak of the crisis. The other effect is countercyclical: a higher mean in market beliefs about aggregate uncertainty (likely to occur before a crisis) gives the rating agency an incentive to be too strict and a lower mean (after a crisis) to be too lenient. While anecdotal evidence suggests that the procyclical effect is stronger,<sup>2</sup> it is ultimately an empirical question, which effect dominates.

This sheds light on a disturbing aspect of using credit ratings for capital adequacy regulation: they may introduce procyclicality into the system. Capital adequacy requirements based on ratings (such as in Basel III and the Solvency II Directive) may be too lenient before and too strict after the crisis. Our theory can be seen to justify two possible policies to deal with this problem. One policy, as in Section 939A of the Frank-Dodd Act, is to remove any reference to or requirement of reliance on credit ratings from regulation. This approach has the advantage of having a clear unambiguous rule. However, this is also viewed controversially, since it may be too costly for smaller banks to replace external credit ratings with internal credit rating systems.<sup>3</sup> An alternative policy would be to use credit ratings, but take into account their cyclicity in regulation. In particular, if one believes that the procyclical element dominates, capital adequacy requirements based on ratings should include countercyclical elements to counterbalance procyclicality.

We provide two extensions of our main result. First, we outline an empirical strategy to determine whether the procyclical or the countercyclical effect dominates. While an empirical analysis is beyond of the scope of this paper, we show how the moments of the distribution of aggregate uncertainty can be identified from the prices of financial derivatives.

---

<sup>2</sup>In hindsight, observers of financial markets considered the ratings of agencies to have been too lenient before and too strict after the crisis.

<sup>3</sup>See, for example, <http://www.americanbanker.com/bankthink/an-easy-fix-to-dodd-franks-credit-ratings-rule-1063396-1.html>.

Second, we extend the model to a setup with risk aversion. A model with risk aversion explains why there are multiple rating categories and not just one (i.e. investment grade, and possibly a second, speculative grade). The reason is that with risk aversion, investors value more precise information about the quality of an asset to reduce risk. We provide numerical examples to illustrate that a hybrid model of risk aversion and aggregate uncertainty preserves the key insights about the rating agency being too lenient or too strict, but additionally predicts multiple rating categories.

Our paper relates to a large literature on rating agencies, experts, and reputation. We differ from all papers mentioned below by having market participants' uncertainty about the aggregate distribution of qualities as the driving force that determines the rating strategy.

If one were to remove aggregate uncertainty from our model, it would reduce to the model in Lizzeri (1999)'s seminal contribution on certification intermediaries. Lizzeri (1999) shows the by now well known result that certification intermediaries choose two categories (corresponding to investment grade and junk bonds) and set a cutoff at 0 which is the first-best level. (Note that this result can also be viewed as only one rating category being chosen – investment grade – and other assets not being rated.) Lizzeri (1999)'s work has been extended in a number of directions, including Doherty, Kartasheva, and Phillips (2012)'s work on risk-averse buyers. With risk-averse buyers, it can be optimal to have more than two categories.

Two papers allow for changes in the economic environment in a dynamic model. In Bolton, Freixas, and Shapiro (2012) the rating agency trades off short term profits from consumers taking the rating at face value and long term reputational concerns. They assume that in a boom the fraction of naive consumers is high and, together with a low default risk, this gives the agency an incentive to inflate ratings during booms. Bar-Isaac and Shapiro (2013) investigate the quality of ratings when accuracy is costly for the agency. They combine reputational concerns with the change of economic fundamentals which affect, e.g., the costs for accuracy, possible profits and the default probability. They find that the rating quality is lower in booms than

in recessions. Our analysis is complementary to these articles, since we show that a rating agency has an incentive to distort ratings even if all investors are rational and it is costless for the rating agency to assess the quality of the rating. Our results rely on the joint distribution of aggregate and idiosyncratic uncertainty.

In a wider sense, our paper also relates to the literature on experts and reputation. Reputation gives an incentive to report truthfully. Strausz (2005) shows that reputation leads to monopolization and that honest certification may require a price above that of a monopolist. Nevertheless, reputation is often not enough to ensure accurate information transmission (see Ottaviani and Sørensen, 2006; Bouvard and Levy, 2009; Mariano, 2008; Pollrich and Wagner, 2013). Mathis, McAndrews, and Rochet (2009) show that reputation and confidence cycles may exist, because the certifier likes to build up reputation so as to later inflate the grades and make larger profits.

The paper is structured as follows. Section 2 describes the model. Section 3 shows that it is optimal to rate according to a simple cutoff rule and Section 4 derives conditions under which this cutoff is positive or negative. Section 5 describes a stylized empirical identification strategy. Section 6 shows that with risk-averse investors several rating classes can be optimal but that the effects of aggregate uncertainty on the optimal cutoff remain. Section 7 concludes.

## 2 Model

There is one rating agency, a continuum of firms, and a continuum of possible investors. Each firm sells a bond of quality  $t$ , where  $t$  is a random variable with support  $[\underline{t}, \bar{t}]$  with  $\underline{t} < 0 < \bar{t}$ . The firm has private information about the quality. Investors are risk neutral and an investor's expected gross utility from buying the good is equal to the quality  $t$ . A firm's utility of retaining the bond is normalized to zero. The quality  $t$  of the bond can be thought of as capturing the probability of default and the loss given default.

There are  $N$  different states of the world. The probability of the world being in



state  $i$  is  $\epsilon_i$ . Having a two dimensional distribution (different states of the world, different distributions of qualities in each state of the world) adds a considerable amount of complexity. To still have a tractable model, we impose a restriction on this two dimensional distribution. We assume that there is a mass  $\kappa_i$  of sellers whose quality  $\underline{t}$  is so low that one would never want to rate them (we will formalize this later on in Assumption 2). There is a mass  $\mu_i$  of sellers whose quality  $\bar{t}$  is so high that one would always want to rate them. And then there is a mass  $\lambda_i$  of sellers with intermediate qualities  $t \in (\underline{t}, \bar{t})$ . We allow for arbitrary distributions of  $\kappa_i$ ,  $\lambda_i$ ,  $\mu_i$  (with the only restrictions that the sum  $\kappa_i + \lambda_i + \mu_i$  is constant and Assumption 2), but restrict the distribution conditional on being in  $(\underline{t}, \bar{t})$  to be a distribution  $F$  which is the same for all states. We assume that  $F$  is continuously differentiable with density  $f(t) > 0$  for all  $t$  in  $(\underline{t}, \bar{t})$ .

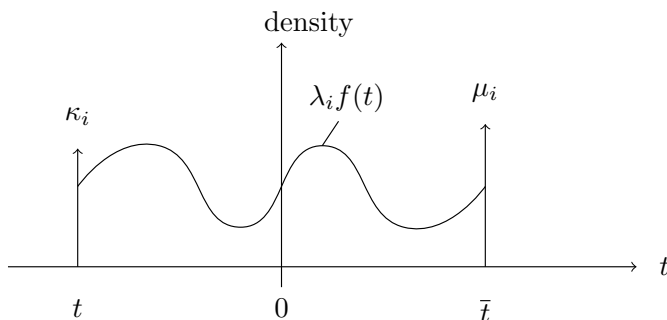


Figure 1:  $\kappa_i$  and  $\mu_i$  are the mass points at  $\underline{t}$  and  $\bar{t}$  in state  $i$ .  $\lambda_i$  is the mass in state  $i$  that is allotted to the types  $t \in (\underline{t}, \bar{t})$  with the distribution  $F$ .

Further, define the expected masses on  $(\underline{t}, \bar{t})$  as  $\tilde{\mu} := \sum_i \epsilon_i \mu_i$  and on  $\bar{t}$  as  $\tilde{\lambda} := \sum_i \epsilon_i \lambda_i$ . Normalize  $\tilde{\lambda}$  to 1. The probabilities  $\epsilon_i$  and the distributions of quality are known to all players.

A firm can choose to pay an upfront fee  $P$  to the rating agency in order to get rated before the state of the world becomes known to market participants. The agency rates firms that paid for a rating according to a precommitted rating strategy.<sup>4</sup>

The timing of moves is as follows:

---

<sup>4</sup>It does not matter in equilibrium whether the strategy is known at the beginning or not.

- The agency sets the rating fee  $P$  and commits to a rating strategy  $s$ ,  $s(t) = r$ ,  $s : \mathbb{R} \rightarrow \mathbb{R} \cup \{\emptyset\}$ .
- Nature draws the state of the world  $i$  and quality  $t$  of each firm.
- The firms observe their own qualities, but not the state of the world, and decide whether to go to the agency to ask for a rating or not. This decision depends on the own type  $t$ , the strategy of the agency  $s$  and the price  $P$ .
- The agency observes the quality of the firms asking for ratings and gives ratings according to its strategy. The ratings are publicly observable. However, investors do not observe whether a firm went to the rating agency if the firm gets no rating ( $\emptyset$ ).
- Observing the state of the world, the buyers decide how much to bid in a second price auction for a good. Since it is a second price auction, buyers bid their own expected valuation which depends on their belief about the expected quality given the information  $(s, P, r, i)$ . Assuming that there are more investors than firms, investors will pay exactly the expected quality in equilibrium.

To solve the setup for equilibria we use Perfect Bayesian Equilibrium. We restrict the strategy of the firms to pure strategies and look at symmetric equilibria.

The profits of the agency in one state of the world is the rating fee  $P$  times the mass of firms asking for a rating. This mass depends on  $P$  and the rating strategy  $s$ . The agency is risk neutral and chooses  $s$  and  $P$  to maximize expected profits before knowing the state of the world.

The rating agency's rating strategy  $s$  partitions the set  $[\underline{t}, \bar{t}]$  into  $M$  subsets, with each subset  $m = 1, \dots, M$  being the set of types  $T_m = \{t | s(t) = r_m\}$  with  $M$  distinct  $r_m$ .<sup>5</sup> We will call these subsets rating classes in the following. Since in the end only

---

<sup>5</sup>Technically speaking, there are  $M + 1$  subsets because there can be types which do not receive any rating,  $s(t) = \emptyset$ . We will show in the following of this paper that it cannot be optimal to have more than two rating categories. Therefore, for the sake of notational simplicity, we do not consider an uncountable infinity of rating classes. To take into account the possibility of an uncountable infinity of rating classes, e.g. full disclosure, one could use the correspondence  $T(r) = \{t | s(t) = r\}$  with  $r \in \mathbb{R} \cup \{\emptyset\}$  instead of the sets  $\{T_m\}_{m=1}^M$ .

the  $M$  distinguishable classes  $\{T_m\}_{m=1}^M$  matter and not the labels  $\{r_m\}_{m=1}^M$  attached to them, the following analysis will focus on  $\{T_m\}$ .

It is useful to define the expected quality in state  $i$  conditional on  $t$  being above a threshold  $x > \underline{t}$  as

$$E_i(x) := \frac{\lambda_i \int_x^{\bar{t}} t dF + \mu_i \bar{t}}{\lambda_i \int_x^{\bar{t}} dF + \mu_i}.$$

A firm in  $(\underline{t}, \bar{t})$  attaches probability  $\hat{\epsilon}_i := \epsilon_i \lambda_i / \tilde{\lambda}$  to being in state  $i$ . Consequently, from a  $(\underline{t}, \bar{t})$  firm's perspective, the expected quality above a threshold  $x$  over all states is

$$\tilde{E}(x) := \sum_i \hat{\epsilon}_i E_i(x).$$

In the following, we will assume that the virtual valuation function attached to  $\tilde{E}(x)$  is monotone in  $x$  for  $x \in (\underline{t}, \bar{t})$ .

**Assumption 1.**  $\tilde{E}(x) - \tilde{E}'(x) \frac{1-F(x)+\bar{\mu}}{f(x)}$  is monotone in  $x$  for  $x \in (\underline{t}, \bar{t})$ .

This assumption basically ensures that the second-order condition is fulfilled whenever the first-order condition is fulfilled and it excludes the corner solution that it is optimal to only rate  $\bar{t}$ . Monotonicity of the virtual valuation function is a standard assumption, see e.g. Myerson (1981).

We further assume that  $\underline{t}$  is sufficiently small:

**Assumption 2.**

$$\underline{t} < -\frac{\lambda_i \int_0^{\bar{t}} t dF(t) + \mu_i \bar{t}}{\kappa_i}, \quad \forall i = 1, \dots, N$$

Assumption 2 makes sure that we do not have to deal with the uninteresting corner solution in which the rating agency wants to rate all firms, including  $\underline{t}$  firms (see Lemma 1 below).<sup>6</sup>

To avoid analytical intractability even for simple aggregate distributions we have described above an assumption on the idiosyncratic distribution, namely that there

---

<sup>6</sup>Without Assumption 2, our results for the interior solutions given by the first-order conditions would still hold, but we would have to distinguish between the cases with an interior solution and a corner solution.

are two mass points at  $\underline{t}$  and  $\bar{t}$ , respectively, and that the density on  $(\underline{t}, \bar{t})$  moves proportionally in different states. This simplification buys us a surprising level of generality in terms of the distribution of aggregate uncertainty: we do not need to make any assumptions about a first-order stochastic dominance ranking of different states of the world and do not need to assume the monotone likelihood ratio property. Our assumption can be viewed as the distribution above a threshold being collapsed into the mass point  $\bar{t}$  and the distribution below a threshold being collapsed into  $\underline{t}$ .

### 3 Optimality of Threshold Rating Strategy

In the following, we will show that it is optimal to rate all firms in an interval  $[x, \bar{t}]$  in one rating class and not to give a rating to all firms with  $t < x$ . Formally,  $s(t) = 1$  for all  $t \geq x$  and  $s(t) = \emptyset$  for all  $t < x$ .<sup>7</sup> We will show this in four steps. First, we show that it cannot be optimal to exclude type  $\bar{t}$ . Second, we show that the price of a rating is determined by firms with  $t < \bar{t}$ . Third, given that  $\bar{t}$  is included, it is optimal to have only one rating class rather than multiple classes. Fourth, given that there is only one rating class, the set of types belonging to this class has to be convex.

While the following Lemmas are intuitive, their proofs are surprisingly long. We therefore provide an intuition in the main text and relegate the proofs to the Appendix.

**Lemma 1.** *(i) It cannot be optimal that  $\underline{t} \in \cup_{m=1}^M T_m$ . (ii) It cannot be optimal that  $\bar{t} \notin \cup_{m=1}^{\tilde{M}} \tilde{T}_m$ .*

Part (i) of the Lemma holds by Assumption 1. The intuition for part (ii) of the Lemma is that  $\bar{t}$  should be included in the rating because it increases the mass of rated firms as well as, due to its high type, other firms' willingness to pay for a rating.

Next, we state a lemma which will be useful throughout our analysis. The lemma states that if both firms with  $t \in (\underline{t}, \bar{t})$  and with  $t = \bar{t}$  are in the same rating class,

---

<sup>7</sup>This is equivalent to  $s(t) = 1$  for all  $t \geq x$  and  $s(t) = 0$  for all  $t < x$  because firms with  $t < x$  are not rated in equilibrium.

then firms with  $t \in (\underline{t}, \bar{t})$  have a lower willingness to pay for a rating than firms with  $t = \bar{t}$ .

**Lemma 2.** *Take an arbitrary rating class  $T$  that includes firms with  $t \in (\underline{t}, \bar{t})$  and  $t = \bar{t}$ . The willingness to pay for a rating is higher for  $\bar{t}$  than for  $t \in (\underline{t}, \bar{t})$ .*

The reason is that firms update  $\hat{\epsilon}_i$  differently and we show that firms with a type  $\bar{t}$  assign a higher probability to states with higher expected quality than firms with  $t \in (\underline{t}, \bar{t})$ . Lemma 2 can be used to prove the next lemma, which states that if there are multiple rating classes and the highest type  $\bar{t}$  is included, then it is better to merge all rating classes to one single class.

**Lemma 3.**  *$M = 1$  with  $T_1 = \cup_{m=1}^{\tilde{M}} \tilde{T}_m$  is better than  $\left\{ \tilde{T}_m \right\}_{m=1}^{\tilde{M}}$  with  $\tilde{M} > 1$  if  $\bar{t} \in \cup_{m=1}^{\tilde{M}} \tilde{T}_m$ .*

Considering types that the agency intends to attract, the rating fee is always determined by the type with the lowest willingness to pay for a rating. Merging the rating class with a lowest willingness to pay with classes with a higher willingness to pay, the expected quality and thus, also the minimum willingness to pay increase.

The next lemma states that all firms above a threshold are rated which means that no types in between are excluded.

**Lemma 4.** *If  $M = 1$  and  $\bar{t} \in T_1$ , then  $T_1$  has to be convex.*

If the set were not convex, there would be at least one unrated hole in the middle and the agency could rate firms in the hole instead of rating some other types below with the same mass. This would increase the expected type in every state and, therefore, also the rating fee the agency could charge from the firms. Therefore, a non-convex  $T_1$  cannot be optimal.

Lemmas 1, 3, and 4 together lead to the following proposition.

**Proposition 1.** *It is optimal to choose  $M = 1$  with  $T_1 = [x, \bar{t}]$  for some  $x$ .*

Proposition 1 shows that the best equilibrium for the rating agency is such that the agency offers the following ratings strategy:

$$s(t) = \begin{cases} 1 & \text{if } t \geq x, \\ \emptyset & \text{otherwise,} \end{cases}$$

that is, all firms above some cutoff  $x$  get a positive rating. Subsequently, all firms with  $t \in [x, \bar{t}]$  get rated and investors pay the expected quality over  $[x, \bar{t}]$ .

As usual in such models, there is a multiplicity of equilibria in the subgame following the ratings agency's announcement of its price  $P$  and rating strategy  $s$ . For example, there is the trivial equilibrium in which no firm applies for a rating and investors have the off-equilibrium belief that firms that do get a rating are of the worst possible rated quality  $x$ . Since  $x$  is less than the price of a rating  $P$ , it is a best response for firms to stay unrated.

The usual arguments for selecting the equilibrium we described apply: The rating agency has a first-mover advantage, hence, it is reasonable that the equilibrium most favorable to the rating agency will be selected. Further, by a small perturbation of its strategy, the rating agency can get rid of undesired equilibria. For example, if no firm gets a rating, the agency might incentivize the first firms who apply for a rating in order to jump-start the market.<sup>8</sup>

## 4 Optimal Threshold

By Proposition 1 we can restrict our attention to threshold rules which consist of all types above a cutoff  $x$  being pooled in one class and all types below not being rated. If there were only one state of the world, the optimal threshold would be  $x = 0$ . To see this, take a model with only one state of the world, e.g. by setting  $\mu_i = \tilde{\mu}$  and

---

<sup>8</sup>A simple, albeit extreme example is the following: As long as not all firms with a quality  $t \in [x, \bar{t}]$  enter, firms get their rating fees refunded and get an additional small compensation. This makes sure that any equilibrium in which not all firms in  $[x, \bar{t}]$  get rated breaks down, so that the refund never has to be paid in equilibrium.

$\lambda_i = \tilde{\lambda} = 1$  for all  $i$ . Then the agency's profit is

$$\Pi = (1 - F(x) + \tilde{\mu}) \frac{\int_x^{\bar{t}} t dF(t) + \tilde{\mu} \bar{t}}{1 - F(x) + \tilde{\mu}} = \int_x^{\bar{t}} t dF(t) + \tilde{\mu} \bar{t}.$$

which is equal to welfare. The first derivative is  $\frac{\partial \Pi}{\partial x} = -xf(x)$ , which is equal 0 if  $x = 0$ . Therefore, the optimal threshold for the agency is  $x = 0$ .<sup>9</sup> This special case of our model corresponds to Lizzeri (1999)'s results.

If there are  $N$  states of the world, the rating agency's profit is

$$\begin{aligned} \Pi(x) &:= \left( \sum_{i=1}^N (\lambda_i(1 - F(x)) + \mu_i) \epsilon_i \right) \left( \sum_{j=1}^N E_j(x) \hat{\epsilon}_j \right) \\ &= (1 - F(x) + \tilde{\mu}) \tilde{E}(x) \end{aligned}$$

where  $\tilde{E}(x)$  is the expected value of a rating from a firm's perspective which assigns the probabilities  $\hat{\epsilon}_i$  to different states.

The welfare with  $N$  states of the world is

$$\begin{aligned} W(x) &:= \sum_{i=1}^N \left( \lambda_i \int_x^{\bar{t}} t dF(t) + \mu_i \bar{t} \right) \epsilon_i \\ &= \sum_{i=1}^N E_i(x) (\lambda_i(1 - F(x)) + \mu_i) \epsilon_i. \end{aligned}$$

Define the expected type on  $[x, \bar{t})$  as

$$E_0(x) := \frac{\int_x^{\bar{t}} t dF(t)}{1 - F(x)}.$$

Rearrange the expression for the welfare to

$$\begin{aligned} W(x) &= \sum_i (\lambda_i(1 - F(x)) E_0(x) + \mu_i \bar{t}) \epsilon_i \\ &= (1 - F(x) + \tilde{\mu}) \hat{E}(x) \end{aligned}$$

with

$$\hat{E}(x) := \frac{(1 - F(x)) E_0(x) + \tilde{\mu} \bar{t}}{1 - F(x) + \tilde{\mu}},$$

---

<sup>9</sup>It is easy to check that the second-order condition is also satisfied at  $x = 0$ .

which can also be written as

$$\hat{E}(x) = \frac{\sum_i \epsilon_i (\lambda_i (1 - F(x)) + \mu_i) E_i(x)}{1 - F(x) + \tilde{\mu}}.$$

$\hat{E}(x)$  is the expected value of a rating from a welfare perspective which takes into account that the quantity of firms being rated  $(\lambda_i(1 - F(x)) + \mu_i)$  is different in every state. In the following, we will drop the argument  $x$  in  $E_i(x)$ ,  $E_0(x)$ ,  $\tilde{E}(x)$ ,  $\hat{E}(x)$  when it is unambiguous in order to simplify notation.  $\hat{E}$  and  $\tilde{E}$  compare in the following way.

**Lemma 5.** *The value of a rating is larger from a welfare then from a firm's perspective;  $\hat{E} \geq \tilde{E}$  for all  $x$ .*

This implies that  $W(x) \geq \Pi(x)$ . For non-degenerate distributions of the state of the world, the inequality is strict and in contrast to a one-state-of-the-world setup the rating agency cannot extract the whole surplus,  $W(x) > \Pi(x)$ .<sup>10</sup>

The derivative of the profit with respect to the cutoff is

$$\begin{aligned} \Pi'(x) &= (1 - F(x) + \tilde{\mu}) \sum_i \hat{\epsilon}_i \frac{\partial E_i}{\partial x} - f(x) \tilde{E}(x) \\ &= -f(x) \left[ \underbrace{\tilde{E}(x)}_{\text{marginal effect}} - \underbrace{\frac{1 - F(x) + \tilde{\mu}}{f(x)} \frac{\partial \tilde{E}}{\partial x}}_{\text{inframarginal effect}} \right]. \end{aligned} \quad (1)$$

and we will show later that the first order condition is sufficient for profit maximization. Thus, the profit maximizing cutoff is given by  $\Pi'(x) = 0$ . Changing the cutoff has two opposite effects on the agency's profit; increasing the cutoff decreases the mass of firms asking to be rated (marginal effect) but it also increases the expected quality of firms being rated and by this it increases a firm's willingness to pay for being rated (inframarginal effect).

We call the expression in the squared brackets in (1) the virtual valuation function

---

<sup>10</sup>Even if  $\hat{\epsilon}_i = \epsilon_i$  for all  $i$ , the inequality is strict for non-degenerated distributions. Besides by the updating of  $\hat{\epsilon}_i$ , the difference between  $\hat{E}$  and  $\tilde{E}$  is caused by the different mass of firms being rated in different states of the world.



for  $\tilde{E}$ .<sup>11</sup> By Assumption 1 it is monotone and, thus, the first order condition is sufficient to find an optimum.<sup>12</sup> This also implies that there is a unique solution of the first order condition.

We are interested in comparing the profit maximizing cutoff with the welfare maximizing cutoff. Thus, we also have to determine the socially optimal cutoff. The derivative of welfare with respect to the threshold is

$$W'(x) = -f(x) \left( \underbrace{\hat{E}(x)}_{\text{marginal effect}} - \underbrace{\frac{1 - F(x) + \tilde{\mu}}{f(x)} \frac{\partial \hat{E}}{\partial x}}_{\text{inframarginal effect}} \right). \quad (2)$$

or written in a simpler way

$$W'(x) = - \sum_i \hat{\epsilon}_i \lambda_i x f(x) = -x f(x)$$

which is the same as for one state of the world. The derivative is 0 if  $x = 0$  and thus, the welfare maximizing cutoff is at 0.

To derive the difference between the profit of an agency and the welfare, we write the profit as

$$\begin{aligned} \Pi(x) &= \left( \sum_i (\lambda_i (1 - F(x)) + \mu_i) \epsilon_i \right) \left( \sum_j E_j(x) \hat{\epsilon}_j \right) \\ &= \sum_j \left( E_j(x) \hat{\epsilon}_j \left( \sum_i (\lambda_i (1 - F(x)) + \mu_i) \epsilon_i \right) \right) \\ &= \sum_j E_j(x) (1 - F(x) + \mu_j) \hat{\epsilon}_j - \sum_j E_j(x) \hat{\epsilon}_j \left( \mu_j - \sum_i \mu_i \epsilon_i \right) \\ &= W(x) + L(x) \end{aligned}$$

where  $L(x) := - \sum_j E_j \hat{\epsilon}_j (\mu_j - E[\mu])$  is the non-extractable part of the surplus ("loss" compared to extracting total surplus). Remember that  $W'(0) = 0$  which implies that

---

<sup>11</sup>We can rewrite the virtual valuation in terms of  $E_i$  as  $\sum_i \epsilon_i \lambda_i \left( E_i - E_i \frac{1 - F + \tilde{\mu}}{f} \right)$ .

<sup>12</sup>The second order condition follows directly from Assumption 1. That we do not have a corner solution at  $x = \bar{t}$  can be seen by observing that the profit  $\Pi(x)$  is continuous at  $x = \bar{t}$  and  $\lim_{x \rightarrow \bar{t}} \Pi'(x) < 0$ . Assumption 2 implies that there is no corner solution at  $x = \underline{t}$  (see proof of Lemma 1).

$L'(0) = \Pi'(0)$ . Thus, the incentive for the agency to distort the rating compared to the welfare maximizing rating is given by the sign of  $L'(0)$ . The optimal cutoff is positive if  $L'(0) > 0$  and it is negative if  $L'(0) < 0$ .

**Proposition 2.** *The derivative  $L'(0)$  is given by*

$$L'(0) = \frac{f(0)\tilde{E}}{\bar{t} - \hat{E}} \left( \hat{E} - \tilde{E} - \frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}} \right). \quad (3)$$

The proof of Proposition 2 is provided in the Appendix.

Since the expression before the parenthesis is always positive, the sign of  $L'(0)$  and, therefore, the sign of the profit maximizing cutoff depends on the sign of  $\left( \hat{E} - \tilde{E} - \frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}} \right)$ .  $\hat{E} - \tilde{E}$  is positive and can be interpreted as the difference of the expected value of a rating from a social and a firm's perspective. An intuition for  $\frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}}$  is that it is the variance divided by the mean of the posterior distribution of  $E_i$  and it reflects the uncertainty about the state of the world: if this uncertainty is sufficiently large, the cutoff is negative. The reason for this is that firms care less about the effect of the cutoff  $x$  on the expected quality of a rated firm if the expected quality is to a large extent driven by uncertainty about the state of the world. Thus, the sign of  $L'(0)$  is determined by the difference of the expected value of a rating and the ratio of variance to mean of the posterior distribution of  $E_i$ .

While the above expression for  $L'(0)$  provides some insights on the determinants of the optimal cutoff, it is difficult to use it for comparative statics, since a change of the mean and variance of  $E_i$  will also change  $\hat{E}$ . Therefore, in the following, we will express  $L'(0)$  in terms of the moments of the posterior distribution of  $E_i$ . Because of  $\mu_i = \lambda_i(1 - F(x)) \frac{E_i - E_0}{\bar{t} - E_i}$ , we can write

$$\begin{aligned} \tilde{\mu} &= \sum_i \epsilon_i \lambda_i (1 - F(x)) \frac{E_i - E_0}{\bar{t} - E_i} \\ &= (1 - F(x)) \sum_i \hat{\epsilon}_i \frac{E_i - \bar{t} + \bar{t} - E_0}{\bar{t} - E_i} \\ &= (1 - F(x)) \left( -1 + (\bar{t} - E_0) \sum_i \hat{\epsilon}_i \frac{1}{\bar{t} - E_i} \right). \end{aligned} \quad (4)$$

Plugging (4) into the definition of  $\hat{E}$  we get

$$\begin{aligned}\hat{E} &= \frac{(1 - F(x))E_0 + \tilde{\mu}\bar{t}}{1 - F(x) + \tilde{\mu}} \\ &= \frac{E_0 + \left(-1 + (\bar{t} - E_0) \sum_i \hat{e}_i \frac{1}{\bar{t} - E_i}\right) \bar{t}}{1 + \left(-1 + (\bar{t} - E_0) \sum_i \hat{e}_i \frac{1}{\bar{t} - E_i}\right)} \\ &= \bar{t} - \frac{1}{\sum_i \hat{e}_i \frac{1}{\bar{t} - E_i}}.\end{aligned}$$

Define the scaled value of a rating as  $e_i := \frac{E_i}{\bar{t}}$ . Then  $\sum_i \hat{e}_i \frac{1}{\bar{t} - E_i} = \sum_i \hat{e}_i \frac{1}{\bar{t}(1 - e_i)}$ . The  $k$ th derivative of  $\frac{1}{\bar{t}(1 - e_i)}$  with respect to  $e_i$  is

$$\frac{\partial^k}{\partial e_i^k} \left[ \frac{1}{\bar{t}(1 - e_i)} \right] = k! \frac{1}{\bar{t}(1 - e_i)^{k+1}}.$$

Using these derivatives one can construct a Taylor series of  $\frac{1}{\bar{t}(1 - e_i)}$  with respect to  $e_i$  around  $e_i = 0$ . This yields

$$\begin{aligned}\frac{1}{\bar{t}(1 - e_i)} &= \sum_{k=0}^{\infty} \frac{e_i^k}{k!} \frac{\partial^k}{\partial e_i^k} \left[ \frac{1}{\bar{t}(1 - e_i)} \right] \Bigg|_{e_i=0} \\ &= \sum_{k=0}^{\infty} \frac{e_i^k}{\bar{t}}.\end{aligned}$$

Taking expectations over the state of the world yields

$$\sum_i \hat{e}_i \left[ \frac{1}{\bar{t}(1 - e_i)} \right] = \frac{1}{\bar{t}} \left( 1 + m_1 + m_2 + \sum_{k=3}^{\infty} m_k \right),$$

where  $m_k := \sum_i \hat{e}_i e_i^k$  is the  $k$ th moment of the posterior distribution of  $e_i$ . This implies that we can write

$$\hat{E} = \bar{t} - \frac{\bar{t}}{1 + m_1 + m_2 + \sum_{k=3}^{\infty} m_k}. \quad (5)$$

Define  $m_{3+} := \sum_{k=3}^{\infty} m_k$ . Observe that (3) simplifies to

$$L'(0) = \frac{f\tilde{E}}{\bar{t} - \hat{E}} \left( \hat{E} - \frac{\sum_i \hat{e}_i E_i^2}{\tilde{E}} \right). \quad (6)$$

Plugging (5),  $\tilde{E} = m_1 \bar{t}$  and  $\sum_i \hat{e}_i E_i^2 = \bar{t}^2 m_2$  into (6) yields

$$\begin{aligned} L'(0) &= \frac{f m_1 \bar{t}}{\frac{\bar{t}}{1+m_1+m_2+m_{3+}}} \left( \bar{t} - \frac{\bar{t}}{1+m_1+m_2+m_{3+}} - \frac{\bar{t}^2 m_2}{m_1 \bar{t}} \right) \\ &= f m_1 \bar{t} \left( \sum_{k=0}^{\infty} m_k \right) \left[ 1 - \frac{1}{1+m_1+m_2+m_{3+}} - \frac{m_2}{m_1} \right]. \end{aligned}$$

Observing that everything outside of the square brackets is positive and that the optimal cutoff has the same sign as  $L'(0)$  yields the main proposition of our paper:

**Proposition 3.** *Define*

$$S := 1 - \frac{1}{1+m_1+m_2+m_{3+}} - \frac{m_2}{m_1}.$$

*The optimal cutoff is positive if  $S > 0$ , negative if  $S < 0$ , and zero if  $S = 0$ .*

Note that  $S$  only depends on the moments of  $e_i$ , more precisely, it depends only on the mean  $m_1$ , the second moment  $m_2$  and the sum of all higher moments  $m_{3+}$ . For example, assume that  $L'(0) < 0$ . Keeping the mean and the second moment constant and increasing the sum of higher moments,  $S$  increases and  $L'(0)$  can switch signs from negative to positive.

We can calculate the threshold  $\bar{m}_{3+}$  for which  $L'(0)$  is 0. Set

$$1 - \frac{1}{1+m_1+m_2+\bar{m}_{3+}} - \frac{m_2}{m_1} = 0$$

which is equivalent to

$$\bar{m}_{3+} = \frac{m_2^2 + m_2 - m_1^2}{m_1 - m_2}.$$

Observe that  $\bar{m}_{3+}$  is always positive because  $m_1 > m_2$  and  $m_2 - m_1^2$  is the variance of  $e_i$ . This implies that for  $m_{3+} < \bar{m}_{3+}$  the expression  $S$  is negative and thus  $L'(0) = \Pi'(0) < 0$ .

**Proposition 4.** *The optimal cutoff for the rating agency is negative if  $m_{3+} < \bar{m}_{3+}$  and positive if  $m_{3+} > \bar{m}_{3+}$ .*

We also derive thresholds for  $m_1$  and  $m_2$ . First, observe that  $S$  is increasing in  $m_1$  and decreasing in  $m_2$  given that  $m_1, m_2, m_{3+} > 0$ . Second, by setting the

expression in square brackets to zero and solving for  $m_1$  and  $m_2$ , respectively, one gets thresholds for  $m_1$  and  $m_2$  that determine whether the cutoff of the rating agency is positive or negative. The thresholds are stated in the following two Propositions.

**Proposition 5.** *The optimal cutoff for the rating agency is negative if  $m_1 < \bar{m}_1$  and positive if  $m_1 > \bar{m}_1$ , where*

$$\bar{m}_1 := \frac{1}{2} \left( -m_{3+} + \sqrt{4m_2 + (2m_2 + m_{3+})^2} \right)$$

**Proposition 6.** *The optimal cutoff for the rating agency is negative if  $m_2 > \underline{m}_2$  and positive if  $m_2 < \underline{m}_2$ , where*

$$\underline{m}_2 := \frac{1}{2} \left( -1 - m_{3+} + \sqrt{2m_{3+} + 1 + (2m_1 + m_{3+})^2 + 2m_{3+} + 1} \right)$$

Both thresholds,  $\bar{m}_1$  and  $\underline{m}_2$ , are positive given that  $m_1, m_2, m_{3+} > 0$ .

An intuition for these results is that with a high variance and a small higher order skewness (i.e left-skewed distribution), the value of a rating is highly volatile, not because of the rating strategy, but because of aggregate uncertainty. Hence the rating agency worries less about keeping the value of the rating high by being restrictive on which firms get a rating.

Propositions 4, 5, and 6 have a striking implication: the rating agency has more of an incentive to be too lenient if the distribution of aggregate uncertainty is more left skewed (in the sense of a smaller higher order skewness, i.e. a lower  $m_{3+}$ ), the mean is smaller, or the variance is larger. Left skewness and a high variance can be reasonably considered as being associated with a period preceding the beginning of a crisis. For moments that can be reasonably associated with a period shortly after a crisis (right skewness, low variance), incentive of the rating agency move in the opposite direction: the rating agency has an increasing incentive to be too strict. This gives the rating agency an incentive to rate procyclically: excessively lenient

ratings expand investments during booms, excessively restrictive ratings restrict investments during recessions. Observe that the mean of aggregate uncertainty has a counter cyclical effect; a small expected average, which can be associated with a period shortly after a crisis, gives the rating agency an incentive to be too lenient. The opposite holds for a high expected average.

## 4.1 Example of Beta Distributions

It is illustrative to parametrize the posterior distribution of  $E_i$  as a Beta distribution with support  $[E_0, \bar{t}]$ , i.e.  $E_i$  having a density  $h(y) \propto y^{\alpha-1}(1-y)^{\beta-1}$ , where  $y = (E_i - E_0)/(\bar{t} - E_0)$ . The distribution of  $E_i/\bar{t}$  is determined by the three parameters  $\alpha$ ,  $\beta$ , and  $e_0 := E_0/\bar{t}$ . (The upper bound of the support of  $E_i/\bar{t}$  is 1.) These three parameters pin down  $m_1$ ,  $m_2$ , and  $m_{3+}$ :

$$\begin{aligned} m_1 &= \frac{\alpha + \beta e_0}{\alpha + \beta}, \\ m_2 &= \frac{(1 - e_0)^2 \alpha \beta}{(\alpha + \beta)^2 (1 + \alpha + \beta)} + m_1^2 \\ m_{3+} &= \frac{\alpha + \beta - 1}{(1 - e_0)(\beta - 1)} - 1 - m_1 - m_2 \end{aligned}$$

It can be shown that this is a one-to-one mapping from  $(\alpha, \beta, e_0)$  to  $(m_1, m_2, m_{3+})$ .<sup>13</sup>

One can use this one-to-one mapping for comparative statics with respect to say  $m_{3+}$

---

<sup>13</sup>The mapping in the opposite direction can be derived in closed form, but the resulting expressions are rather long and uninformative and therefore omitted.  $m_1$  and  $m_2$  are the well-known first two moments of the Beta distribution.  $m_{3+}$  can be derived by observing that  $E[(1-y)^{-1}(1-e_0)^{-1}] = E[\sum_{k=0}^{\infty} (e_0 + (1-e_0)y)^k] = E[\sum_{k=0}^{\infty} e^k] = 1 + m_1 + m_2 + m_{3+}$ , where  $e = E_i/\bar{t} = e_0 + (1-e_0)y$ . For a Beta distribution with density  $h(y) = y^{\alpha-1}(1-y)^{\beta-1}/B(\alpha, \beta)$  the expected value is

$$E \left[ \frac{1}{1-y} \right] = \int_0^1 \frac{y^{\alpha-1}(1-y)^{\beta-2}}{B(\alpha, \beta)} dy = \frac{B(\alpha, \beta-1)}{B(\alpha, \beta)} = \frac{\alpha + \beta - 1}{\beta - 1},$$

where the last equality follows from the relation of the Beta to the Gamma function

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)},$$

and the property  $\Gamma(x+1) = x\Gamma(x)$  of the Gamma function which imply

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta-1)(\beta-1)}{\Gamma(\alpha + \beta - 1)(\alpha + \beta - 1)} = \frac{\beta-1}{\alpha + \beta - 1} B(\alpha, \beta-1).$$

Putting this together yields the expression for  $m_{3+}$ .

while keeping  $m_1$  and  $m_2$  constant. Figure 2 shows a Beta distribution with  $\alpha = 3$ ,  $\beta = 5$  and  $e_0 = 0.1$  (dashed line). For this distribution,  $L'(0) = 0$ , i.e. the rating agencies sets the cutoff at exactly the socially optimal level  $x = 0$ . For the dotted line,  $m_1$  and  $m_2$  are kept constant and  $m_{3+}$  is reduced by 0.01. The dotted line has a fatter lower tail which means that it has a higher mass at the bottom of the distribution. The mean and variance remain the same, but if a crisis hits, it is more likely to be severe. For the dotted distribution  $L'(0) < 0$  and hence the cutoff is negative,  $x < 0$ , which means that the rating criteria are too loose compared to the socially optimal ones. For the solid line,  $m_{3+}$  is increased by 0.01 while keeping  $m_1$  and  $m_2$  constant. For this distribution  $L'(0) > 0$  and hence  $x > 0$ , that is, the rating is too strict compared to the socially optimal one.

Figures 3, 4, and 5 illustrate the change of  $L'(0)$  as  $m_{3+}$ ,  $m_1$ , and  $m_2$  are changed, respectively, while keeping the other parameters constant. The optimal cutoff for example can switch from a negative to a positive cutoff if the mean or the higher order skewness increase or if the variance decreases. For all values of  $m_1$ ,  $m_2$ , and  $m_{3+}$ , the parameters  $\alpha$ ,  $\beta$ , and  $e_0$  are in permissible ranges.<sup>14</sup>

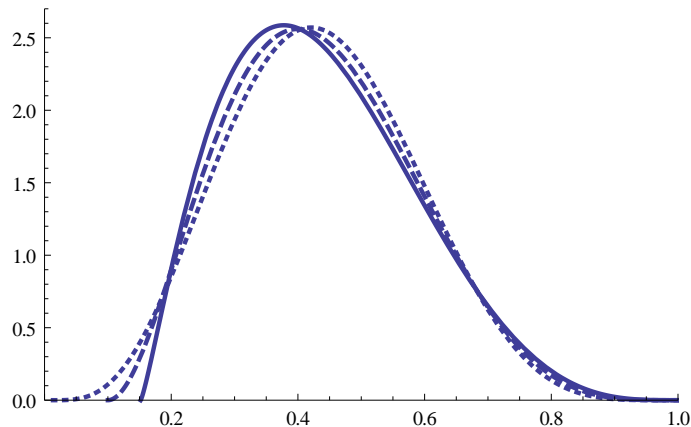


Figure 2: Density of  $E_i/\bar{t}$  for  $\alpha = 3$ ,  $\beta = 5$ ,  $e_0 = 0.1$  (dashed line). For the dotted line,  $m_{3+}$  is reduced by 0.01, for the solid line,  $m_{3+}$  is increased by 0.01, while  $m_1$  and  $m_2$  are kept constant. (The corresponding parameters are  $\alpha = 4.4322$ ,  $\beta = 5.8781$ ,  $e_0 = 0.013363$  for the dotted and  $\alpha = 2.23$ ,  $\beta = 4.38985$ ,  $e_0 = 0.151755$  for the solid distribution.)

<sup>14</sup>The permissible ranges are  $\alpha > 0$ ,  $\beta > 0$ , and  $e_0 \in (0, 1)$ .

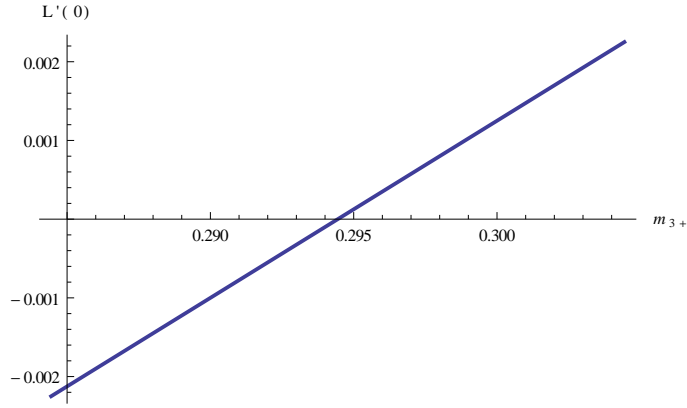


Figure 3: Values of  $L'(0)$  as  $m_{3+}$  is changed and  $m_1$  and  $m_2$  are kept constant. Starting point is  $\alpha = 3$ ,  $\beta = 5$ ,  $e_0 = 0.1$  (which corresponds to  $m_1 = 0.4375$ ,  $m_2 = 0.2125$ , and  $m_{3+} = 0.294444$ ) for which  $L'(0) = 0$ . Further parameters are normalized to  $\bar{t} = 1$  and  $f(0) = 1$ .

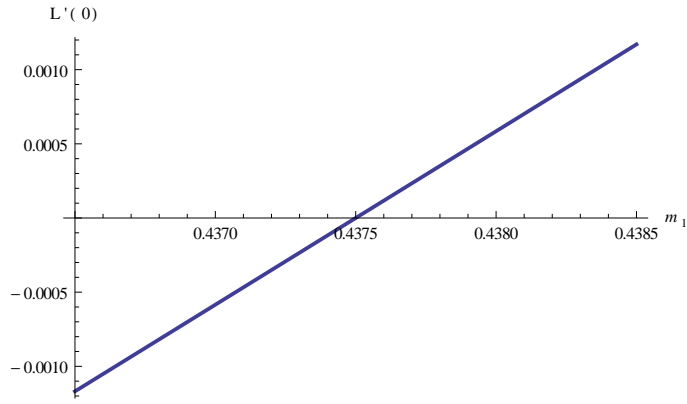


Figure 4: Values of  $L'(0)$  as  $m_1$  is changed and  $m_2$  and  $m_{3+}$  are kept constant. Starting point is  $\alpha = 3$ ,  $\beta = 5$ ,  $e_0 = 0.1$  (which corresponds to  $m_1 = 0.4375$ ,  $m_2 = 0.2125$ , and  $m_{3+} = 0.294444$ ) for which  $L'(0) = 0$ . Further parameters are normalized to  $\bar{t} = 1$  and  $f(0) = 1$ .

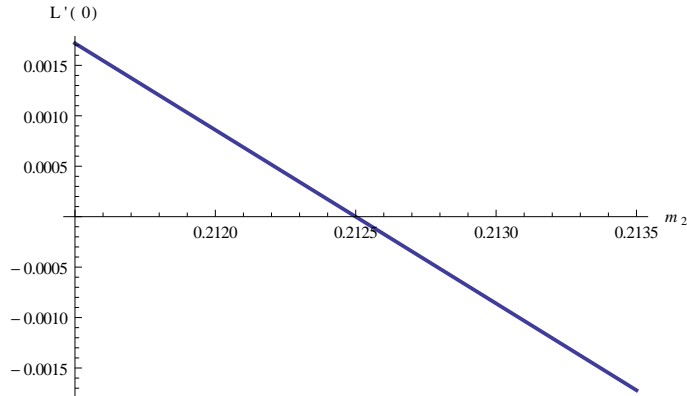


Figure 5: Values of  $L'(0)$  as  $m_2$  is changed and  $m_1$  and  $m_{3+}$  are kept constant. Starting point is  $\alpha = 3$ ,  $\beta = 5$ ,  $e_0 = 0.1$  (which corresponds to  $m_1 = 0.4375$ ,  $m_2 = 0.2125$ , and  $m_{3+} = 0.294444$ ) for which  $L'(0) = 0$ . Further parameters are normalized to  $\bar{t} = 1$  and  $f(0) = 1$ .



## 5 Empirical Implications

Our model shows how the rating agency's incentive to be too lenient or too strict depends on the moments of aggregate uncertainty. Since these moments cannot be observed directly, one may wonder about the empirical content of our model.

First, it should be noted that an empirical estimate of the distribution of aggregate uncertainty is non-trivial, especially if the main concern is about the distribution of aggregate uncertainty shortly before a crisis. The reason is that only few crises occur, so it is difficult to have larger amounts of data.

However, an empirical estimate of market participants' beliefs about the distribution of aggregate uncertainty can be obtained. We illustrate the basic idea of how to estimate these moments in a strongly stylized setup containing the core idea of the empirical strategy.

Consider the following stylized setup. There is an index for the bonds being sold by the firms in the market. As an example, one can think of a subprime mortgages-backed securities index. Further, there is a market for financial derivatives based on this index. Call options on the index can be bought in the first period of the model, before aggregate uncertainty is realized.<sup>15</sup> The options expire in the second period after aggregate uncertainty has realized. Time is discrete and the options are European options.<sup>16</sup> Further, aggregate uncertainty is such that the mid-quality firms' beliefs are the same as the general market beliefs, formally,  $\hat{\epsilon}_i = \epsilon_i$  for all  $i$ .<sup>17</sup> Suppose that the cut-off of the agency is close to 0 ( $x \approx 0$ ), so that the value of the index  $E_i(x)$  is well approximated by  $E_i(0)$ .

Further, assume that there exists a call option with strike price  $y_i = E_i$  for each state of the world  $i$ . Without loss of generality, order the states of the world increasingly, i.e.  $E_j > E_i$  if  $j > i$ . The second-period value of a call option with

---

<sup>15</sup>For the sake of simplicity, we focus on call options on an index. One could also think of put options on an index or e.g. mortgage credit default swap ABX indices.

<sup>16</sup>In a discrete two-period model, it does not matter whether the option is European or American. In a continuous time model, calculations for American options are somewhat more complex, but standard and well known in the literature.

<sup>17</sup>A sufficient condition is that  $\lambda_i = \tilde{\lambda}$  for all  $i$ , that is, aggregate uncertainty enters through changes of  $\kappa_i$  and  $\mu_i$  for different states of the world  $i$ .

strike price  $y_j$  in state  $i$  is  $E_i - y_j$  if  $E_i > y_j$  and 0 if  $E_i \leq y_j$ . Denote the first-period price of option  $j$  with strike price  $y_j$  as  $O_j$ .  $O_j$  is given by the market's expected value of the second period value (ignoring discounting):

$$O_j = \sum_{i=1}^N \epsilon_i \max\{E_i - y_j, 0\} = \sum_{i=j+1}^N \epsilon_i (E_i - E_j) \quad (7)$$

where the second equality follows from  $y_j = E_j$ . (For  $i = N$ ,  $O_j = 0$ .)

The next proposition shows that given a set of call options, the information on their strike prices  $y_j$  and first-period prices  $O_j$  identifies the market's beliefs about the distribution of aggregate uncertainty; it identifies the probability  $\epsilon_i$  for the expected quality  $E_i$ . The proof is provided in the Appendix.

**Proposition 7.** *Given strike prices and first period prices  $\{(y_j, O_j)\}_{j=1}^N$ , the probability mass function of the distribution of aggregate uncertainty is given by*

$$\epsilon_j = \frac{O_j - O_{j+1}}{y_{j+1} - y_j} - \frac{O_{j-1} - O_j}{y_j - y_{j-1}}$$

for  $1 < j < N$  and

$$\begin{aligned} \epsilon_N &= \frac{O_{N-1}}{y_N - y_{N-1}}, \\ \epsilon_1 &= 1 - \sum_{i=2}^N \epsilon_i. \end{aligned}$$

A similar result can be obtained for a continuous distribution of  $E_i$ . For the continuous distribution version, drop the index in  $E_i$  and denote the distribution of  $E$  as  $G$ . Assume that prices  $O(y)$  for call options with a continuum of strike prices  $y \in [\underline{t}, \bar{t}]$  are observed. Then  $O(y)$  is given by

$$O(y) = \int_{\underline{t}}^{\bar{t}} \max\{E - y, 0\} dG(E) = \int_y^{\bar{t}} (E - y) dG(E).$$

The first derivative is

$$O'(y) = \int_y^{\bar{t}} (-1) dG(c) - [(y - y)g(y)] = -(1 - G(y)),$$

and the second

$$O''(y) = g(y).$$

This is analogous to the discrete distribution result and the distribution  $G$  is non-parametrically identifiable given data on call option prices.

In practice, one expects to observe less options than there are states of the world, so parametric assumptions are required to be able to estimate the distribution of  $E$ .

In the following, we make the parametric assumption that the distribution  $G$  is a polynomial with lower bound of support  $E_0$  and upper bound  $\bar{t}$ . As an example, consider a cubic function

$$G(E) = a_1 + 2a_2E + 3a_3E^2 + 4a_4E^3.$$

The price of a call option will also be a polynomial function of the strike price  $y$ , since

$$O(y) = \int_y^{\bar{t}} (1 - G(E))dE = a_0 + a_1y + a_2y^2 + a_3y^3 + a_4y^4,$$

where

$$a_0 = - \sum_{i=1}^4 a_i \bar{t}^i.$$

Suppose we observe data for five call options with strike prices  $\{y_j\}_{j=1}^5$  and option prices  $\{O(y_j)\}_{j=1}^5$ . In this case, the parameters  $\{a_i\}_{i=0}^4$  are given by the linear equation system

$$O(y_j) = a_0 + a_1y_j + a_2y_j^2 + a_3y_j^3 + a_4y_j^4, \quad j = 1, \dots, 5. \quad (8)$$

As long as the matrix  $[y_j^i]_{j=1, \dots, 5; i=0, \dots, 4}$  is non-singular, the equation system (8) yields a unique solution for the variables  $\{a_i\}_{i=0}^4$ . Note that  $E_0$  and  $\bar{t}$  are uniquely pinned down by the parameters  $\{a_i\}_{i=0}^4$  and by the equations  $G(E_0) = 0$  and  $G(\bar{t}) = 0$ .<sup>18</sup>

Given the distribution  $G$  of  $E$ , we can obtain the distribution of  $e = E/\bar{t}$  and

---

<sup>18</sup>While the  $G(E)$  has multiple roots due to  $G$  being a polynomial, the solution of  $G(E_0) = 0$  is unique nonetheless. This is because of the constraints  $G'(E) > 0$  for  $E \in [E_0, \bar{t}]$  and  $y_j \in [E_0, \bar{t}]$  for all  $j$ . By the same reasoning, there is a unique solution of  $G(\bar{t}) = 1$ .

the moments  $m_1, m_2, m_{3+}$  of  $e$ . This in turn yields

$$S = 1 - \frac{1}{1 + m_1 + m_2 + m_{3+}} - \frac{m_2}{m_1}$$

from expression Proposition 3 and determines the sign of the marginal profit  $\Pi'(0)$  at  $x = 0$ . Table 1 provides examples of hypothetical observed prices of call options and corresponding estimated parameters, moments, and  $S$ . For the first set of observations (first line), the rating agency has an incentive to choose the cutoff at the first best level  $x = 0$ . For the second line and third line, the agency has an incentive to choose a negative and a positive cutoff, respectively.

observed prices					estimated parameters				moments			$S$
$O_1$	$O_2$	$O_3$	$O_4$	$O_5$	$a_1$	$a_2$	$a_3$	$a_4$	$m_1$	$m_2$	$m_{3+}$	
11	8.0	5.7	3.9	2.6	-0.37	0.0096	-0.000030	$3.4 \times 10^{-8}$	0.35	0.15	0.30	0.0
5.9	4.1	2.7	1.8	1.1	0.0	0.0077	-0.000026	$3.4 \times 10^{-8}$	0.24	0.095	0.12	-0.073
14	10	7.2	4.9	3.2	-0.89	0.014	-0.000043	$5.1 \times 10^{-8}$	0.41	0.20	0.37	0.019

Table 1: Example of parameter estimates for data on call option prices  $O_j = O(y_j)$  for strike prices  $(y_1, y_2, y_3, y_4, y_5) = (80, 90, 100, 110, 120)$ .

We have illustrated the basic idea behind an empirical strategy to estimate the moments of aggregate uncertainty. To practically apply this strategy, several additional steps are required, which are beyond the scope of this article. First, one needs to construct synthetic call options for the index of the bonds being rated. Second, the pricing of options in a multi-period environment is more complicated than the simple two-period setup we used for illustrative purposes. These problems are far from trivial, but well studied in Finance, see e.g. Hull (2009). Additionally, one could use a different parametrization for  $G$  instead of the polynomial parametrization or, if sufficiently many observations are available, one could possibly even use a non-parametric estimate of the function  $O(y)$  given the observations  $\{y_j, O(y_j)\}_j$ . Further, one would also want to estimate the confidence interval for  $S$ .

## 6 Risk-aversion

In the main part of this paper we have assumed that investors are risk neutral and we have shown that it is optimal for the agency to pool all types above a cutoff in one rating class. Doherty, Kartasheva, and Phillips (2012) extend the model of Lizzeri (1999) by allowing investors to be risk-averse and they show that, if the level of risk aversion is sufficiently high, the rating agency rates types above a cutoff in several rating classes.

First, following the paper of Doherty et al. (2012) we provide a simplified hybrid model incorporating risk aversion and aggregate uncertainty. We show that introducing risk aversion in a model with several states of the world can also yield several rating classes. In this case our previous analysis can be interpreted as determining the optimal cutoff of the lowest investment grade rating class (e.g. BBB). Second, we provide a numerical example to show that the effects of the moments of the expected quality distribution on the optimal cutoff have the same sign as before even with risk aversion and several rating classes.

We provide the simplest possible setup which is rich enough to illustrate the idea. Assume that buyers are risk-averse. Their utility of an asset is equal to  $t$  but their expected utility depends on both the mean and the variance of the quality of the asset they buy. We include a second mass point at  $\bar{t}_2$ ,  $\bar{t}_2 \geq \bar{t}$ , with mass  $\gamma_i$  in state  $i$ . To avoid confusion denote  $\bar{t}$  by  $\bar{t}_1$ . See Fig. 6.

If buyers are risk-averse, a welfare maximizing rating strategy needs to perfectly disclose the type of all assets with a positive value because any kind of pooling and being vague about a firm's quality leads to a welfare loss. However, such a strategy cannot be optimal for the rating agency.<sup>19</sup> To analyze a general model with risk-aversion is beyond the scope of this article. In the following, we compare two rating strategies: (i) pooling all types above a cutoff in one rating class, which is the optimal strategy without risk-aversion, and (ii) a strategy in which the agency only pools low types and rates high types separately. Doherty et al. (2012) show

---

<sup>19</sup>To ensure that all firms with  $t \geq 0$  are willing to pay the rating fee under full disclosure, the rating fee has to be 0.

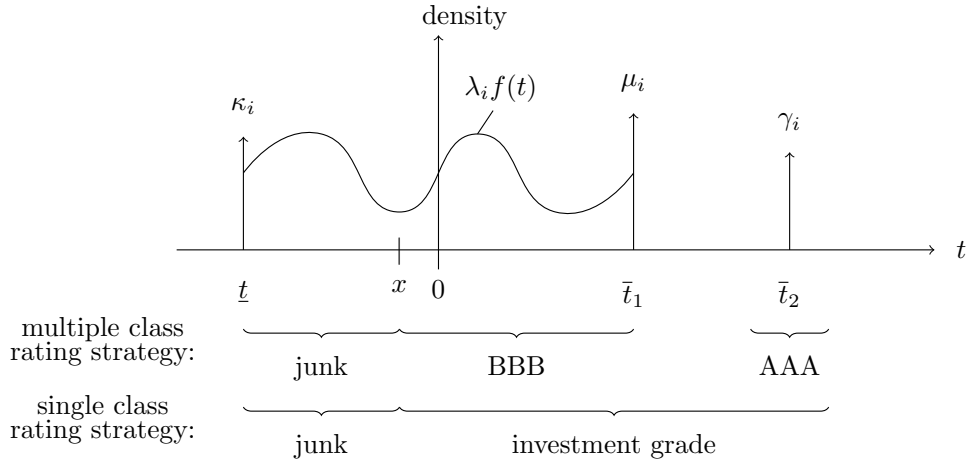


Figure 6:  $\kappa_i$ ,  $\mu_i$  and  $\gamma_i$  are the mass points at  $\underline{t}$ ,  $\bar{t}_1$  and  $\bar{t}_2$  in state  $i$ .  $\lambda_i$  is the mass in state  $i$  that is allotted to the types  $t \in (\underline{t}, \bar{t})$  with the distribution  $F$ .

that strategy (ii) is optimal in a model with one state of the world if the level of risk-aversion is sufficiently high.

Analogously to the case without risk-aversion, we derive the profit of the agency if it pools all types above a cutoff  $x$  in one class. The expected type above a cutoff  $x$  is

$$Q_i(x) := E[t|t \geq x]$$

and the variance is

$$\sigma_i(x) := \text{Var}[t|t \geq x].$$

The buyer's valuation for the asset of a seller in this rating class is

$$Q_i(x) - a\sigma_i(x)$$

where  $a$  is a measure for risk-aversion. If  $a = 0$ , the buyers are risk-neutral and the model is equivalent to the one before.

The profit of the agency if it pools all types is

$$\begin{aligned} \hat{\Pi}(x) &:= \left( \sum_i (\lambda_i(1 - F(x)) + \mu_i + \gamma_i)\epsilon_i \right) \left( \sum_i \hat{\epsilon}_i(Q_i(x) - a\sigma_i(x)) \right) \\ &= (1 - F(x) + \tilde{\mu} + \tilde{\gamma}) \sum_i \hat{\epsilon}_i(Q_i(x) - a\sigma_i(x)) \end{aligned}$$

where  $\tilde{\gamma}$  is the expected value of  $\gamma$ ,  $\tilde{\gamma} = \sum_i \epsilon_i \gamma_i$ .

Alternatively, the rating agency can pool  $t \in [x, \bar{t}_1]$  and rate  $\bar{t}_2$  separately as shown in Fig. 6. If the agency rates types  $\bar{t}_2$  in a separate class, these sellers are willing to pay a high rating fee (up to  $\bar{t}_2$ ) and therefore the rating fee is determined by sellers in the class  $t \in [x, \bar{t}_1]$ . Keeping the cutoff  $x$  constant, the mass of rated firms is the same for both strategies and the rating fee decides which rating strategy yields higher profits. If the agency pools types  $t \in [x, \bar{t}_1]$ , the expected type in this rating class is smaller than  $Q_i(x)$  but the variance is also smaller than  $\sigma_i(x)$ . Thus, it is not straight forward to see under which strategy the rating fee can be higher.

Now, we derive sufficient conditions such that the agency prefers to rate  $\bar{t}_2$  separately instead of pooling all types above  $x$  in one rating class. Define  $z_i := \gamma_i \bar{t}_2$  and  $\tilde{z}$  as the expected value of  $z_i$ ,  $\tilde{z} := \sum_i \epsilon_i z_i$ . Rewrite  $\tilde{z}$  as  $\tilde{z} = \bar{t}_2 \tilde{\gamma}$ , which can be interpreted as the agency's profit if it charges a rating fee of  $\bar{t}_2$  and rates only firms with type  $\bar{t}_2$ . Remember that  $\Pi(x)$  is defined as the profit if the agency rates only  $t \in [x, \bar{t}_1]$  and pools them all in one class. The sufficient condition is given in the following proposition, the proof being provided in the Appendix.

**Proposition 8.** *Take an arbitrary cutoff  $x$ . For any  $\tilde{z}$  with  $\tilde{z} \leq \Pi(x)$  there exists a  $\bar{T}_2$  such that for all  $\bar{t}_2 \geq \bar{T}_2$  the rating agency is better off pooling  $t \in [x, \bar{t}_1]$  and rating  $\bar{t}_2$  in a separate class than pooling all types above  $x$  in one rating class.*

Since investors are risk-averse, their expected utility buying an asset in a given rating class decreases if the variance inside this rating class becomes larger. If the variance is sufficiently large, investors are not willing to pay any positive price for an asset even if the expected quality is positive. Thus, if the variance is large, the agency is better off splitting the types in several rating classes in order to reduce the variance inside one class and to increase investors' willingness to pay for an asset. The condition that  $\tilde{z} \leq \Pi(x)$  ensures that the agency does not prefer to charge a rating fee of  $\bar{t}_2$  and to exclude firms with  $t < \bar{t}_2$  from the rating.

Risk aversion does not only have the effect of multiple rating classes becoming optimal, but it also has an additional effect on the optimal cutoff. Increasing the

cutoff reduces the variance in a rating class and this can give additional incentives to increase the cutoff.<sup>20</sup> In the following we provide numerical examples in which we show that the effects of the first, second, and higher moments are similar to our analysis without risk aversion.<sup>21</sup> In the numerical example we have four states of the world. We take the Generalized Pareto distribution  $F(t) = 1 - ((1 - t)/2)^3$  for  $t \in (-1, 1)$  and fix  $\bar{t}_1 = 1$ . This gives us  $E_0 = 1/4$ . We fix  $\lambda = 5$ ,  $\bar{t}_2 = 110$ , and  $\nu_i = 0.0001$  for all  $i$ . The states only differ in the weights  $\mu_i$  at the mass point at  $\bar{t}_1$ , with  $\mu_1 = 0.03$ ,  $\mu_2 = 0.2$ ,  $\mu_3 = 0.4$ , and  $\mu_4 = 0.7$ . Changing the moments of the aggregate distribution, we keep the distribution inside a state constant (and therefore also the expected type) and only vary the probabilities for the different states. There is a one-to-one mapping from  $(\epsilon_1, \epsilon_2, \epsilon_3)$  to  $(m_1, m_2, m_{3+})$  and the fourth probability is pinned down by  $\epsilon_4 = 1 - \epsilon_1 - \epsilon_2 - \epsilon_3$ . For all values of the example, the probabilities are in  $[0, 1]$ .

Figures 7, 8, and 9 illustrate the change of the optimal cutoff as  $m_{3+}$ ,  $m_1$  and  $m_2$  are changed while keeping the other moments constant. The solid line is the optimal cutoff for  $a = 0$ , the dashed line for  $a = 0.01$  and the dotted-dashed line for  $a = 0.02$ . If investors are risk neutral,  $a = 0$ , the agency pools all types above the cutoff in one class. For  $a = 0.01$  and  $a = 0.02$ , investors are risk averse and the agency prefers to pool all types  $t \in [x, \bar{t}_1]$  in one class and to rate  $\bar{t}_2$  separately. Note that increasing the level of risk aversion leads to an increase in the optimal cutoff  $x^*$ . The figures show that our results from the main part of the paper carry over to a setup including risk-aversion: Keeping the other moments constant, a higher mean, a lower variance, or an increase in the higher order skewness lead to an increase in the optimal cutoff. For changes with the opposite sign, the optimal cutoff decreases.

Hence, results from the risk-neutral setup carry over qualitatively to a setup with

---

<sup>20</sup>Doherty et al. (2012) show that the optimal cutoff can be positive even with only one state of the world if the level of risk aversion is sufficiently high.

<sup>21</sup>In the main part of the paper the moments were defined for the distribution of the expected type in  $[0, \bar{t}]$  (scaled by  $\bar{t}$ ). For the sake of comparison, in the numerical examples the moments are defined for the distribution of the expected type in  $[x, \bar{t}_1]$  and thus, the expected type is not influenced by changes in the mass on  $\bar{t}_2$ . We deviate from our previous analysis by taking the threshold  $x$  as the lower bound of the interval. In this way we can determine the optimal cutoff explicitly and not only its sign.



risk aversion. A setup with risk aversion comes at the cost of being analytically intractable, but has the desirable feature that it makes the more realistic prediction of multiple rating classes.

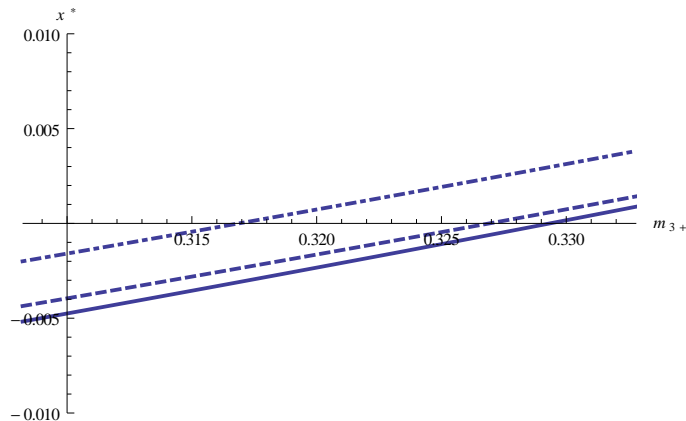


Figure 7: Values of the optimal threshold  $x^*$  as  $m_{3+}$  is changed and  $m_1$  and  $m_2$  are kept constant. For the solid line  $a = 0$ , for the dashed line  $a = 0.01$  and for the dotted-dashed line  $a = 0.02$ . The rating strategy for the solid line is to pool all types above  $x$ . For the dashed and dotted-dashed line all types in  $[x, \bar{t}_1]$  are pooled and  $\bar{t}_2$  is rated separately. (The starting values are  $\epsilon_i = 1/4$  for all  $i$ . This implies  $m_1 = 0.47627$ ,  $m_2 = 0.244859$  and as a starting value  $m_{3+} = 0.321538$ .)

## 7 Conclusions

We have considered the profit maximizing rating strategy of a rating agency in the face of aggregate uncertainty. We have shown that with risk neutral investors it is still optimal for the rating agency, as in a setup without aggregate uncertainty, to choose only one rating class for rated firms and to not rate the remaining firms.

The model's predictions about the cutoff for the rating class strikingly differ from the predictions of a model without aggregate uncertainty: the rating agency has more of an incentive to be too lenient if the expected average quality is small, the variance large, and the higher order skewness small. For larger averages, smaller variances, and larger higher order skewness the opposite holds: the rating agency has more of an incentive to be too strict. These results can be interpreted as ratings having either a procyclical or an countercyclical effect. We outline an empirical strategy to estimate the moments of aggregate uncertainty which can be used to

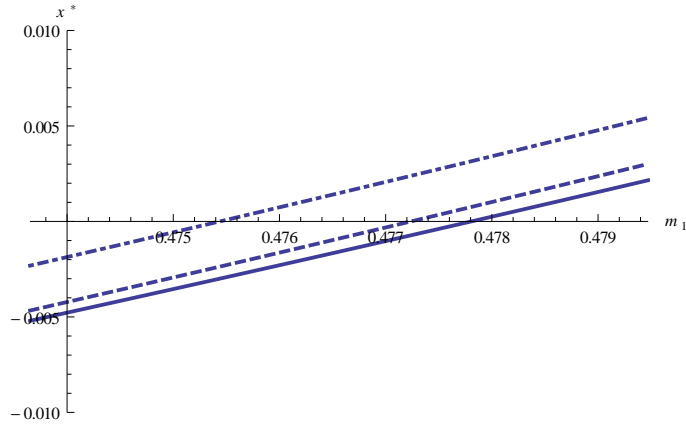


Figure 8: Values of the optimal threshold  $x^*$  as  $m_1$  is changed and  $m_2$  and  $m_{3+}$  are kept constant. For the solid line  $a = 0$ , for the dashed line  $a = 0.01$  and for the dotted-dashed line  $a = 0.02$ . The rating strategy for the solid line is to pool all types above  $x$ . For the dashed and dotted-dashed line all types in  $[x, \bar{t}_1]$  are pooled and  $\bar{t}_2$  is rated separately. (The starting values are  $\epsilon_i = 1/4$  for all  $i$ . This implies  $m_2 = 0.244859$ ,  $m_{3+} = 0.321538$  and as a starting value  $m_1 = 0.47627$ .)

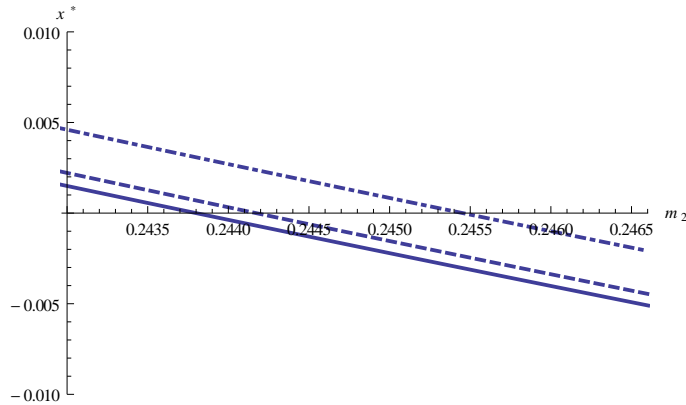


Figure 9: Values of the optimal threshold  $x^*$  as  $m_2$  is changed and  $m_1$  and  $m_{3+}$  are kept constant. For the solid line  $a = 0$ , for the dashed line  $a = 0.01$  and for the dotted-dashed line  $a = 0.02$ . The rating strategy for the solid line is to pool all types above  $x$ . For the dashed and dotted-dashed line all types in  $[x, \bar{t}_1]$  are pooled and  $\bar{t}_2$  is rated separately. (The starting values are  $\epsilon_i = 1/4$  for all  $i$ . This implies  $m_1 = 0.47627$ ,  $m_{3+} = 0.321538$  and as a starting value  $m_2 = 0.244859$ .)

determine which effect dominates.

Our analysis identifies one up to now unconsidered factor that affects the rating strategy of an agency – aggregate uncertainty – and thereby sheds further light in understanding the behavior of rating agencies. In line with our model, one disturbing effect of using ratings as the basis for financial regulation is that a possible procyclicality of ratings leads to a procyclicality of capital adequacy requirements for banks, and hence to a procyclicality of lending. One solution is to avoid using ratings as the basis for financial regulation. Another is to counterbalance the procyclicality of ratings by adding countercyclicality to capital adequacy requirements that are based on ratings.

The usual disclaimer for the policy implications holds. This article is about a thorough analysis of the effects of aggregate uncertainty, shutting down other effects such as like reputation cycles, imperfect rating technology, and competition between agencies. Further, the implications of the theory depend on the empirical moments of the distribution of aggregate uncertainty. Hence, an empirical analysis is needed to estimate these moments and the relative magnitude of the different effects. Our article provides a starting point for such an empirical analysis. This paper also serves as a word of caution: using a distribution which is pinned down by its mean and variance (e.g. the often used normal distribution) for an empirical analysis will neglect the impact of the higher order skewness. However, the skewness is crucial for the incentive of the rating agency to distort ratings.

## References

- BAR-ISAAC, H. AND J. SHAPIRO (2013): “Ratings quality over the business cycle,” *Journal of Financial Economics*, 108, 62–78.
- BOLTON, P., X. FREIXAS, AND J. SHAPIRO (2012): “The credit ratings game,” *The Journal of Finance*, 67, 85–111.
- BOUVARD, M. AND R. LEVY (2009): “Humouring both parties: a model of two-sided reputation,” Tech. rep., working paper.
- DOHERTY, N. A., A. V. KARTASHEVA, AND R. D. PHILLIPS (2012): “Information

- effect of entry into credit ratings market: The case of insurers' ratings," *Journal of Financial Economics*, 106, 308–330.
- HULL, J. (2009): *Options, futures and other derivatives*, Pearson education.
- LIZZERI, A. (1999): "Information revelation and certification intermediaries," *The RAND Journal of Economics*, 30, 214–231.
- MARIANO, B. (2008): "Do reputational concerns lead to reliable ratings?" Tech. rep., Financial Markets Group, London School of Economics and Political Science.
- MATHIS, J., J. MCANDREWS, AND J. ROCHET (2009): "Rating the raters: are reputation concerns powerful enough to discipline rating agencies?" *Journal of Monetary Economics*, 56, 657–674.
- MYERSON, R. B. (1981): "Optimal auction design," *Mathematics of operations research*, 6, 58–73.
- OTTAVIANI, M. AND P. SØRENSEN (2006): "Reputational cheap talk," *The Rand journal of economics*, 37, 155–175.
- POLLRICH, M. AND L. WAGNER (2013): "Informational opacity and honest certification," *SFB/TR 15 Discussion Paper No. 481*.
- STRAUSZ, R. (2005): "Honest certification and the threat of capture," *International Journal of Industrial Organization*, 23, 45–62.
- VISWANATHA, A. AND K. FREIFELD (2015): "S&P reaches \$1.5 billion deal with U.S., states over crisis-era ratings," *Reuters*, February 3, 2015.

## Appendix

### Proof of Lemma 1

*Proof.* (i) Denote the class containing  $\underline{t}$  as  $T_n$ . Define an alternative rating class  $T_n^* = \{\underline{t}\} \cup [0, \bar{t}]$ . Observe  $\kappa_i \underline{t} + \lambda_i \int_0^{\bar{t}} t dF(t) + \mu_i \bar{t}$  is the expected quality in  $T_n^*$  in state  $i$ ,  $E[T_{n,i}^*]$ , and by Assumption 2 this is smaller than zero. The expected quality in class  $T_n$  in state  $i$ ,  $E[T_{n,i}]$ , can be larger or smaller than  $E[T_{n,i}^*]$ . If  $E[T_{n,i}]$  is smaller than  $E[T_{n,i}^*]$ , it follows directly that  $E[T_{n,i}] < 0$ .

If  $E[T_{n,i}]$  is larger than  $E[T_{n,i}^*]$ ,  $T_n$  must include types  $t \in [E[T_{n,i}^*], 0]$  to raise the expected quality. Including negative qualities in  $T_n$  can increase the expected type in comparison to  $T_n^*$  but the expected type  $E[T_{n,i}]$  stays negative. Therefore, the willingness to pay for a rating in category  $T_n$  is negative and the rating agency prefers not to have category  $T_n$ .

(ii) Take a rating strategy  $\{\tilde{T}_m\}_{m=1}^{\tilde{M}}$ . Assume that  $\bar{t}$  is not in any  $\tilde{T}_m$ . Define for all rating classes  $\tilde{T}_m$  the expected value

$$E_m^* = \frac{\int_{t \in \tilde{T}_m} t dF(t)}{\int_{t \in \tilde{T}_m} dF(t)} \quad (9)$$

which is constant over all states of the world. The price is determined by the lowest willingness to pay,  $\min_m E_m^*$ , and the expected mass of rated firms  $\sum_i \epsilon_i \sum_m \int_{t \in \tilde{T}_m} \lambda_i dF(t)$ . Using  $\sum_i \epsilon_i \lambda_i = 1$ , we get for profits

$$\tilde{\Pi} = \left[ \min_m E_m^* \right] \left\{ \sum_{m=1}^{\tilde{M}} \int_{t \in \tilde{T}_m} dF(t) \right\} \quad (10)$$

Now take a rating strategy with  $M = \tilde{M} + 1$ ,  $T_m = \tilde{T}_m$  for  $m \leq \tilde{M}$  and  $T_M = \{\bar{t}\}$ . Including the  $\bar{t}$  types adds expected mass  $\tilde{\mu}$  to the mass of rated firms. Hence, expected profits are

$$\Pi = \left[ \min \left( \{\bar{t}\} \cup \{E_m^*\}_{m=1}^{\tilde{M}} \right) \right] \left\{ \tilde{\mu} + \sum_{m=1}^{\tilde{M}} \int_{t \in T_m} dF(t) \right\} \quad (11)$$

Since in (11) the expression in square brackets is weakly greater than in (10) and the expression in curly braces is strictly greater in (11) than in (10), we have  $\Pi > \tilde{\Pi}$ .  $\square$

## Proof of Lemma 2

*Proof.* It holds that

$$\sum_i \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) = 0$$

because of  $\sum_i \epsilon_i \mu_i / \tilde{\mu} = \sum_i \epsilon_i \lambda_i / \tilde{\lambda} = 1$  Define two sets of states of the world;  $i \in A$  if  $\frac{\mu_i \epsilon_i}{\lambda_i} > \tilde{\mu}$  and  $i \in B$  if  $\frac{\mu_i \epsilon_i}{\lambda_i} \leq \tilde{\mu}$ . Thus

$$\sum_{i \in A} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) + \sum_{i \in B} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) = 0$$

and multiply by a constant  $c$

$$\sum_{i \in A} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c + \sum_{i \in B} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c = 0. \quad (12)$$

The expected quality in state  $i$  is

$$E_i = \frac{\lambda_i \int_{t \in T} tdF + \mu_i \bar{t}}{\lambda_i \int_{t \in T} dF + \mu_i} = \frac{\int_{t \in T} tdF + \mu_i / \lambda_i \bar{t}}{\int_{t \in T} dF + \mu_i / \lambda_i}$$

and is increasing in  $\mu_i / \lambda_i$ . Define  $c$  as  $\frac{\int_{t \in T} tdF + \tilde{\mu} \bar{t}}{\int_{t \in T} dF + \tilde{\mu}}$ . The expected quality  $E_i$  for  $i \in A$  is larger than  $c$  and  $E_i < c$  for  $i \in B$ . It follows that

$$\sum_{i \in A} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c < \sum_{i \in A} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i \quad (13)$$

and

$$\sum_{i \in B} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c < \sum_{i \in B} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i. \quad (14)$$

Using inequalities (13) and (14) gives us

$$\begin{aligned} & \sum_{i \in A} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i + \sum_{i \in B} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i \\ & > \sum_{i \in A} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c + \sum_{i \in B} \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c \end{aligned}$$

which is equal 0 by equation (12). Therefore, it holds that

$$\sum_i \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} E_i \right) - \sum_i \left( \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} E_i \right) > 0.$$

Since  $\sum_i \left( \frac{\epsilon_i \mu_i}{\tilde{\mu}} E_i \right)$  is the willingness to pay for a rating for type  $\bar{t}$  and  $\sum_i \left( \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} E_i \right)$  for type  $t \in (\underline{t}, \bar{t})$ , the lemma follows.  $\square$

### Proof of Lemma 3

*Proof.* Label the rating class that includes  $\bar{t}$  as  $\tilde{T}_1$  and the remaining rating classes as  $\tilde{T}_{-1} = \cup_{m \neq 1} \tilde{T}_m$ . Denote the expected type of  $\tilde{T}_1$  conditional on being in state  $i$  as  $\tilde{E}_i = [\int_{t \in \tilde{T}_1} t dF(t) + \mu_i \bar{t}] / [\int_{t \in \tilde{T}_1} dF(t) + \mu_i]$ . Denote the mass of all other classes as  $\mu^* = \int_{t \in \tilde{T}_{-1}} dF(t)$  and the expected type as  $t^* = [\int_{t \in \tilde{T}_{-1}} t dF(t)] / [\int_{t \in \tilde{T}_{-1}} dF(t)]$ .

Profits for only one rating class  $T_1 = \cup_m \tilde{T}_m$  are

$$\Pi = \left[ \sum_i \hat{e}_i E_i \right] (\mu^* + \tilde{\mu})$$

where

$$E_i = \frac{\lambda_i (\int_{t \in \tilde{T}_{-1}} t dF(t) + \int_{t \in \tilde{T}_1} t dF(t) + \mu_i \bar{t})}{\lambda_i (\int_{t \in \tilde{T}_{-1}} dF(t) + \int_{t \in \tilde{T}_1} dF(t) + \mu_i)}$$

is the expected type in state  $i$  if there is only one rating class. Profits for separate rating classes  $\{\tilde{T}_m\}$  are

$$\tilde{\Pi} = \left[ \min \left( \{E_m^*\}_{m=1}^{\tilde{M}} \cup \left\{ \sum_i \hat{e}_i \tilde{E}_i \right\} \right) \right] (\mu^* + \tilde{\mu}),$$

where  $E_m^*$  is defined as in (9). Further, define the profit in case all rating classes  $m \neq 1$  were merged, such that one had two rating classes  $\tilde{T}_1$  and  $\cup_{m=2}^{\tilde{M}} \tilde{T}_m$ , as

$$\hat{\Pi} = \left[ \min \left\{ t^*, \sum_i \hat{e}_i \tilde{E}_i \right\} \right] (\mu^* + \tilde{\mu}).$$

Since  $t^*$  is a weighted average of  $\{\tilde{E}_m\}_{m=1}^{\tilde{M}}$ , we have  $t^* \geq \min\{\tilde{E}_m\}_{m=1}^{\tilde{M}}$  and therefore  $\hat{\Pi} \geq \tilde{\Pi}$ . (Note that  $\Pi$ ,  $\hat{\Pi}$ , and  $\tilde{\Pi}$  only differ in the expressions in square brackets.)

We will prove the lemma by contradiction. Assume to the contrary that separate classes are desirable, i.e.  $\tilde{\Pi} > \Pi$ . This implies  $\hat{\Pi} > \Pi$ , which is equivalent to

$$\min \left\{ t^*, \sum_i \hat{e}_i \tilde{E}_i \right\} > \sum_i \hat{e}_i E_i,$$

by comparison of the expressions in square brackets. This condition is equivalent to

both

$$t^* > \sum_i \hat{\epsilon}_i E_i \quad (15)$$

and

$$\sum_i \hat{\epsilon}_i \tilde{E}_i > \sum_i \hat{\epsilon}_i E_i \quad (16)$$

being satisfied at the same time.

The expected value  $E_i$  can be written as weighted average of  $t^*$  and  $\tilde{E}_i$  for every state  $i$

$$\begin{aligned} E_i &= \frac{\lambda_i (\int_{t \in \tilde{T}_{-1}} t dF(t) + \int_{t \in \tilde{T}_1} t dF(t)) + \mu_i \bar{t}}{\lambda_i (\int_{t \in \tilde{T}_{-1}} dF(t) + \int_{t \in \tilde{T}_1} dF(t)) + \mu_i} \\ &= \frac{\lambda_i t^* \int_{t \in \tilde{T}_{-1}} dF(t) + \tilde{E}_i (\int_{t \in \tilde{T}_1} \lambda_i dF(t) + \mu_i)}{\lambda_i (\int_{t \in \tilde{T}_{-1}} dF(t) + \int_{t \in \tilde{T}_1} dF(t)) + \mu_i}. \end{aligned}$$

Solving for  $\tilde{E}_i$ , we get

$$\tilde{E}_i = E_i + \frac{\lambda_i \int_{t \in \tilde{T}_{-1}} dF(t)}{\lambda_i \int_{t \in \tilde{T}_1} dF(t) + \mu_i} (E_i - t^*)$$

Plugging  $\tilde{E}_i$  into (16), we get

$$\sum_i \hat{\epsilon}_i \left( E_i + \frac{\lambda_i \int_{t \in \tilde{T}_{-1}} dF(t)}{\lambda_i \int_{t \in \tilde{T}_1} dF(t) + \mu_i} (E_i - t^*) \right) > \sum_i \hat{\epsilon}_i E_i$$

or equivalently

$$\sum_i \hat{\epsilon}_i \left( \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + \mu_i / \lambda_i} (E_i - t^*) \right) > 0. \quad (17)$$

Define two sets of states of the world;  $i \in A$  if  $E_i \geq t^*$  and  $i \in B$  if  $E_i < t^*$ . It holds that  $\mu_i / \lambda_i > \mu_j / \lambda_j$  for all  $i \in A$  and  $j \in B$ . This can be seen by checking that

$$E_i = \frac{(\int_{t \in \tilde{T}_{-1}} t dF(t) + \int_{t \in \tilde{T}_1} t dF(t)) + \mu_i / \lambda_i \bar{t}}{(\int_{t \in \tilde{T}_{-1}} dF(t) + \int_{t \in \tilde{T}_1} dF(t)) + \mu_i / \lambda_i}$$

is increasing in  $\mu_i / \lambda_i$ . Denote  $c_A = \min \{\mu_i / \lambda_i | i \in A\}$  and  $c_B = \max \{\mu_i / \lambda_i | i \in B\}$ .

Note that  $c_A > c_B$ .



Then (15) can be rewritten as

$$\sum_i \hat{\epsilon}_i(E_i - t^*) < 0$$

which is equivalent to

$$\sum_{i \in A} \hat{\epsilon}_i(E_i - t^*) + \sum_{i \in B} \hat{\epsilon}_i(E_i - t^*) < 0.$$

This implies

$$\left[ \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + c_A} \sum_{i \in A} \hat{\epsilon}_i(E_i - t^*) \right] + \left[ \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + c_B} \sum_{i \in B} \hat{\epsilon}_i(E_i - t^*) \right] < 0 \quad (18)$$

since

$$\frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + c_A} < \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + c_B}$$

and the sum over  $i \in A$  is positive and the sum over  $i \in B$  is negative. Since  $\mu_i/\lambda_i \geq c_A$  for all  $i \in A$  and  $\sum_{i \in A} \hat{\epsilon}_i(E_i - t^*)$  positive, the first expression in square brackets in (18) is bounded from below by

$$\frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + c_A} \sum_{i \in A} \hat{\epsilon}_i(E_i - t^*) \geq \sum_{i \in A} \hat{\epsilon}_i \left( \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + \mu_i/\lambda_i} (E_i - t^*) \right). \quad (19)$$

The second expression in square brackets is bounded from below by

$$\sum_{i \in B} \hat{\epsilon}_i(E_i - t^*) \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + c_B} \geq \sum_{i \in B} \hat{\epsilon}_i \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + \mu_i/\lambda_i} (E_i - t^*). \quad (20)$$

because of  $\mu_i/\lambda_i \leq c_B$  for all  $i \in B$  and the negativity of  $\sum_{i \in B} \hat{\epsilon}_i(E_i - t^*)$ .

(18),(19) and (20) imply

$$\sum_{i \in A} \hat{\epsilon}_i \left( \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + \mu_i/\lambda_i} (E_i - t^*) \right) + \sum_{i \in B} \hat{\epsilon}_i \frac{\int_{t \in \tilde{T}_{-1}} dF(t)}{\int_{t \in \tilde{T}_1} dF(t) + \mu_i/\lambda_i} (E_i - t^*) < 0$$

which contradicts (17). □

## Proof of Lemma 4

*Proof.* Assume that  $\hat{T}$  is not convex. Take a convex set  $T'$  such that it has the same expected mass of rated firms ( $\int_{t \in \hat{T}} dF(t) = \int_{t \in T'} dF(t)$ ) and both sets include  $\bar{t}$ . Remember that the profit is

$$\Pi(T) = \left[ \sum_i \hat{\epsilon}_i \frac{\lambda_i \int_{t \in T} t dF(t) + \mu_i \bar{t}}{\lambda_i \int_{t \in T} dF(t) + \mu_i} \right] \left( \int_{t \in T} dF(t) + \tilde{\mu} \right).$$

Since  $\int_{t \in \hat{T}} dF(t) = \int_{t \in T'} dF(t)$ , comparing the profits  $\Pi(\hat{T})$  and  $\Pi(T')$  boils down to comparing the willingness to pay for  $\hat{T}$  and  $T'$ , which is given in square brackets. Since  $\hat{T}$  is not convex, there is at least one unrated hole in the middle and it is possible to rate the mass in the holes instead of rating some types below with the same mass. This increases  $\int_{t \in T} t dF(t)$ , while the mass of rated types stays constant. It follows that  $\frac{\lambda_i \int_{t \in T} t dF(t) + \mu_i \bar{t}}{\lambda_i \int_{t \in T} dF(t) + \mu_i}$  is greater for  $T'$  than for  $\hat{T}$  and hence  $\Pi(T') > \Pi(\hat{T})$ . Therefore, it is optimal to rate a set  $T$  which is convex and includes  $\bar{t}$ .  $\square$

## Proof of Lemma 5

*Proof.* Analogously to the proof of Lemma 2 define two sets of states of the world;  $i \in A$  if  $\frac{\mu_i}{\lambda_i} > \frac{\tilde{\mu}}{\lambda}$  and  $i \in B$  if  $\frac{\mu_i}{\lambda_i} \leq \frac{\tilde{\mu}}{\lambda}$ . It holds that  $\sum_i \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) = 0$  which we can write as  $\sum_A \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) + \sum_B \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) = 0$ . Multiplied by a constant  $c = \frac{1}{(1-F)+\tilde{\mu}}$  the expression is still equal to 0. For  $i \in A$ ,  $\frac{1}{(1-F)+\mu_i/\lambda_i}$  is smaller than  $\frac{1}{(1-F)+\tilde{\mu}}$  and for  $i \in B$  it is the other way round. It follows that

$$\sum_A \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) \frac{1}{(1-F) + \mu_i/\lambda_i} + \sum_B \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) \frac{1}{(1-F) + \mu_i/\lambda_i} < 0$$

because  $\mu_i - \lambda_i \tilde{\mu}$  is positive for  $i \in A$  and negative for  $i \in B$ . This is equivalent to  $\frac{(\bar{t}-E_0)(1-F)}{\tilde{\lambda}(1-F)+\tilde{\mu}} \sum_i \hat{\epsilon}_i \frac{\tilde{\lambda}\mu_i - \tilde{\mu}\lambda_i}{\lambda_i(1-F)+\mu_i} < 0$  and thus,  $\sum_i \hat{\epsilon}_i \frac{\lambda_i(1-F)E_0 + \mu_i \bar{t}}{\lambda_i(1-F)+\mu_i} < \frac{\tilde{\lambda}(1-F)E_0 + \tilde{\mu}\bar{t}}{\tilde{\lambda}(1-F)+\tilde{\mu}}$ .  $\square$

## Proof of Proposition 2

*Proof.*

$$\begin{aligned}
L'(x) &= \Pi'(x) - W'(x) \\
&= f(x) \left( \left( \hat{E} - \frac{1 - F(x) + \tilde{\mu}}{f} \hat{E}' \right) - \left( \tilde{E} - \frac{1 - F(x) + \tilde{\mu}}{f} \tilde{E}' \right) \right) \\
&= f(x) \left( \hat{E} - \tilde{E} + \frac{1 - F(x) + \tilde{\mu}}{f} (\tilde{E}' - \hat{E}') \right). \tag{21}
\end{aligned}$$

We know that  $\hat{E} \geq \tilde{E}$  but the sign of  $\tilde{E}' - \hat{E}'$  can go in both directions.

Next, we rewrite (21) such that we can express  $L'(x)$  only in terms of  $E_i$ ,  $\tilde{E}$ , and  $\hat{E}$ . The derivative of  $\tilde{E}$  with respect to  $x$  is

$$\sum_i \hat{\epsilon}_i \frac{\partial E_i}{\partial x} = \sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1 - F(x)) + \mu_i} f(x) \lambda_i.$$

and analogously it can be shown that

$$\frac{\partial \hat{E}}{\partial x} = \frac{\hat{E} - x}{1 - F(x) + \tilde{\mu}} f(x).$$

Using these two expressions in (21), we can write

$$\begin{aligned}
L'(x) &= f(x) \left( \hat{E} - \tilde{E} + \frac{1 - F(x) + \tilde{\mu}}{f} \left( \sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1 - F(x)) + \mu_i} \lambda_i f - \frac{\hat{E} - x}{1 - F(x) + \tilde{\mu}} f \right) \right) \\
&= f(x) \left( \hat{E} - \tilde{E} + (1 - F(x) + \tilde{\mu}) \sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1 - F(x)) + \mu_i} \lambda_i - (\hat{E} - x) \right).
\end{aligned}$$

From the definitions of  $E_i$  and  $\hat{E}$  we derive  $\mu_i = \lambda_i(1 - F(x)) \frac{E_i - E_0}{\bar{t} - E_i}$  and  $\tilde{\mu} = (1 - F(x)) \frac{\hat{E} - E_0}{\bar{t} - \hat{E}}$  which leads to

$$\begin{aligned}
L'(x) &= f(x) \left( x - \tilde{E} + (1 - F(x) + (1 - F(x)) \frac{\hat{E} - E_0}{\bar{t} - \hat{E}}) \sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1 - F(x)) + \lambda_i(1 - F(x)) \frac{E_i - E_0}{\bar{t} - E_i}} \right) \\
&= f(x) \left( x - \tilde{E} + \left( 1 + \frac{\hat{E} - E_0}{\bar{t} - \hat{E}} \right) \sum_i \hat{\epsilon}_i \frac{(E_i - x)(\bar{t} - E_i)}{\bar{t} - E_0} \right) \\
&= f(x) \left( x - \tilde{E} + \sum_i \hat{\epsilon}_i \frac{(E_i - x)(\bar{t} - E_i)}{\bar{t} - \hat{E}} \right).
\end{aligned}$$

Remember that  $W'(0) = 0$  which implies that  $L'(0) = \Pi'(0)$ . Thus, the sign of  $L'(0)$  determines the sign of the profit maximizing cutoff. To determine the sign of  $L'(0)$  we set  $x = 0$  in the above expression and

$$\begin{aligned}
L'(0) &= f(0) \left( -\tilde{E} + \frac{\sum_i \hat{\epsilon}_i E_i (\bar{t} - E_i)}{\bar{t} - \hat{E}} \right) \\
&= f(0) \left( -\tilde{E} + \frac{\bar{t}\tilde{E} - \sum_i \hat{\epsilon}_i E_i^2 - \tilde{E}\hat{E} + \tilde{E}\hat{E}}{\bar{t} - \hat{E}} \right) \\
&= f(0) \left( -\tilde{E} + \frac{\tilde{E}\hat{E} - \sum_i \hat{\epsilon}_i E_i^2}{\bar{t} - \hat{E}} + \frac{(\bar{t} - \hat{E})\tilde{E}}{\bar{t} - \hat{E}} \right) \\
&= f(0) \left( \underbrace{-\tilde{E}}_{\text{marginal effect}} + \frac{\tilde{E}}{\bar{t} - \hat{E}} \left( \hat{E} - \frac{\sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}} \right) + \tilde{E} \right).
\end{aligned}$$

This expression gives us another way to write the inframarginal effect of a change of the threshold at  $x = 0$  on the profit  $\Pi$ . We can simplify  $L'(0)$  to

$$\begin{aligned}
L'(0) &= \frac{f(0)\tilde{E}}{\bar{t} - \hat{E}} \left( \hat{E} - \frac{\sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}} \right) \\
&= \frac{f(0)\tilde{E}}{\bar{t} - \hat{E}} \left( \hat{E} - \tilde{E} - \frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}} \right).
\end{aligned}$$

□

## Proof of Proposition 7

*Proof.* The expression for  $\epsilon_N$  can be derived by observing that

$$O_{N-1} = \sum_{i=N}^N \epsilon_i (E_i - y_{N-1}) = \epsilon_n (y_N - y_{N-1}).$$

The expression for  $\epsilon_j$  for  $1 < j < N$  can be obtained by first observing that

$$O_{j-1} - O_j = \sum_{i=j}^N \epsilon_i (E_i - E_{j-1}) - \sum_{i=j+1}^N \epsilon_i (E_i - E_j) = \sum_{i=j}^N \epsilon_i (E_j - E_{j-1}),$$

where the first equality follows from (7) and the second equality can be obtained by rearranging the sums. Dividing by  $E_j - E_{j-1}$  yields

$$\frac{O_{j-1} - O_j}{E_j - E_{j-1}} = \sum_{i=j}^N \epsilon_i,$$

and taking differences

$$\frac{O_j - O_{j+1}}{y_{j+1} - y_j} - \frac{O_{j-1} - O_j}{y_j - y_{j-1}} = \sum_{i=j+1}^N \epsilon_i - \sum_{i=j}^N \epsilon_i = \epsilon_j,$$

that is, the expression for  $\epsilon_j$  for  $1 < j < N$  in the proposition. The expression for  $\epsilon_1$  simply follows from that fact that probabilities add up to one.  $\square$

## Proof of Proposition 8

*Proof.* We want to analyze the effect of a change of  $\bar{t}_2$ , while keeping  $z_i = \gamma_i \bar{t}_2$  constant. For this purpose the variance  $\sigma_i(x)$  can be rewritten as

$$\begin{aligned} \sigma_i(x) &= \frac{\lambda_i \int_x^{\bar{t}_1} t^2 dF(t) + \mu_i \bar{t}_1^2 + \gamma_i \bar{t}_2^2}{\lambda_i(1 - F(x)) + \mu_i + \gamma_i} - \left( \frac{\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + \gamma_i \bar{t}_2}{\lambda_i(1 - F(x)) + \mu_i + \gamma_i} \right)^2 \\ &= \frac{(\lambda_i(1 - F(x)) + \mu_i + \gamma_i)(\lambda_i \int_x^{\bar{t}_1} t^2 dF(t) + \mu_i \bar{t}_1^2 + \gamma_i \bar{t}_2^2) - (\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + \gamma_i \bar{t}_2)^2}{(\lambda_i(1 - F(x)) + \mu_i + \gamma_i)^2} \\ &= \frac{(\lambda_i(1 - F(x)) + \mu_i + z_i/\bar{t}_2)(\lambda_i \int_x^{\bar{t}_1} t^2 dF(t) + \mu_i \bar{t}_1^2 + z_i \bar{t}_2) - (\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + z_i)^2}{(\lambda_i(1 - F(x)) + \mu_i + z_i/\bar{t}_2)^2} \end{aligned}$$

For  $\bar{t}_2 \rightarrow \infty$  we get that the variance  $\sigma_i(x)$  goes to infinity

$$\lim_{\bar{t}_2 \rightarrow \infty} \sigma_i(x) = \infty$$

and the expected type  $Q_i(x)$  converges to

$$\lim_{\bar{t}_2 \rightarrow \infty} Q_i(x) = \frac{\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + z_i}{\lambda_i(1 - F(x)) + \mu_i} < \infty.$$

It follows that for  $a > 0$ , the utility in state  $i$ ,  $Q_i(x) - a\sigma_i(x)$ , becomes negative if  $\bar{t}_2$  is large enough. This implies that buyers are not willing to pay a positive price for a rated firm if the variance of types in one rating class is too high. Thus, for every

cutoff  $x$  there is a  $\bar{t}_2$  large enough such that the rating agency is better off pooling  $t \in [x, \bar{t}_1]$  and rating  $\bar{t}_2$  in a separate class than pooling all types above  $x$ .

A further condition that needs to be satisfied is that the agency does not prefer to charge a rating fee of  $\bar{t}_2$  and to rate only firms with type  $\bar{t}_2$  which yields profits of  $\bar{t}_2 \tilde{\gamma}$ . Since we keep  $\gamma_i \bar{t}_2$  constant when we increase  $\bar{t}_2$ , the profit of only rating types  $\bar{t}_2$  stays constant. A sufficient condition for the agency to prefer to rate  $[x, \bar{t}_1]$  and  $\bar{t}_2$  is that the profit from pooling  $t \in [x, \bar{t}_1]$  and not rating  $\bar{t}_2$  is larger than the profit from only rating  $\bar{t}_2$ . □