A Theory of Credit Rating Criteria

Nan Guo
Research Division, China Bond Rating Co. Ltd., 28 Financial St., Xicheng Dist., Beijing, China., abraham.guo@qq.com

Steven Kou
Department of Finance, Questrom School of Business, Boston University, 595 Commonwealth Ave., Boston, MA 02215, U.S.A., kou@bu.edu

Bin Wang
RCSDS, NCMIS, AMSS, Chinese Academy of Sciences, 55 Zhongguancun East Rd., Haidian Dist., Beijing, China, wangbin@amss.ac.cn

Ruodu Wang
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave W., Waterloo ON, Canada, wang@uwaterloo.ca

We propose a theory for rating financial securities in the presence of structural maximization by the issuer in a market with investors who rely on credit rating. Two types of investors, simple investors who price tranches solely based on the ratings and model-based investors who use the rating information to calibrate models, are considered. Concepts of self-consistency and information gap are proposed to study different rating criteria. In particular, the expected loss criterion used by Moody's satisfies self-consistency but the probability of default criterion used by S&P does not. Moreover, the probability of default criterion typically has a higher information gap than the expected loss criterion. Empirical evidence in the post-Dodd-Frank period is consistent with our theoretical implications. We show that a set of axioms based on self-consistency leads to a tractable representation for all self-consistent rating criteria, which can also be extended to incorporate economic scenarios. New examples of self-consistent and scenario-based rating criteria are suggested.

Key words: credit ratings, structured finance, Dodd-Frank, axiomatic characterization

1. Introduction

Despite the ubiquitous appearance of credit rating in modern financial systems, there are various criticisms of credit rating especially after the 2008 financial crisis, resulting in some regulatory changes, such as the Dodd-Frank Act in the United States. Consequently, there is a growing literature on rating-seeking behaviors and competitions in the market for ratings, such as rating shopping, rating inflation and competitions, informed intermediation, and rating catering.

However, direct studies on suitable rating criteria and how to regulate them are largely missing. This paper intends to bridge this gap by studying rational credit rating criteria, via a decision-theoretic approach, and its impact on structural maximization of tranches.

Different rating agencies tend to use different rating criteria. For example, S&P and Fitch choose the probability of default as their primary criterion, henceforth referred to as the PD criterion, while Moody's uses the expected loss as a percentage of its nominal value, henceforth referred to as the EL criterion. Intuitively, different rating criteria may have desirable or undesirable pricing and investment implications. In particular, in this paper we attempt to answer the following two questions:

(I) Can an issuer[1] exploit credit rating criteria to enhance the total deal value by restructuring?

(II) If so, can we develop a rational framework of credit rating criteria that does not encourage such exploitation?

1.1. A Motivating Example and Our Contribution

Consider an issuer who restructures an A-rated risky debt into a senior debt and a subordinate debt. Under the PD criterion, the new subordinate debt would surely maintain its A rating while the senior debt may get a higher rating, e.g. AA, because the PD criterion is defined based on the probability of first dollar loss. Thus, one A-rated debt becomes an A-rated debt and an AA-rated debt. On the other hand, if an A-rated risky debt under the EL criterion is restructured into two pieces, the new subordinate debt might get a lower rating (e.g., B) due to increased expected loss per dollar although the new senior debt might get a higher rating. Hence, under the PD criterion, the issuer surely gets either improved or the same ratings, in contrast to the EL criterion which has a probability of getting a worse rating for the subordinate debt. Thus, intuitively, the PD rating criterion encourages the issuer to pursue a multiple-tier structure of debts by having more tranches, which could lead to potential financial gains if investors use credit ratings as a basis for pricing; this will be analyzed rigorously in the paper.

The concept of structural maximization, i.e., designing tranches optimally (including the number of tranches and the cutoff points for tranches) to take advantage of a particular group of investors who believe in rating criteria, is first suggested by Brennan et al. (2009) (although they did not use the same terminology). Because the marketing gains of structural maximization cannot be observed easily in a direct way, Brennan et al. (2009) give numerical examples to estimate the gains based on the classical Merton's structural model, which show that there are substantial gains under the PD criterion while in most cases the gain under the EL criterion is small (p. 916 therein). An interesting link of structural maximization to dominance is studied by Hull and White (2012), who also point out the possibility of the financial gains of issuers by designing more tranches to take advantage a particular group of investors who believe in certain rating criteria.

[1] For this paper, The terminology "issuer" is used in a broad sense, which means the counterpart of investors. Mainly we refer to the underwriter (structurer) who solicit credit ratings from the major credit rating agencies; however it could also refer to the collection of banks and other participants from the sell side.

Although the above motivating example is simple and illuminating, three main questions remain:

(1) The above example only gives a particular way to undertake structural maximization by having two tranches. How do we know which criterion, EL or PD, may lead to higher values for *all* structural maximization schemes? In fact, we shall prove that, in a market of simple investors, for two tranches and under a comparability condition, the optimal total deal values under the EL and the PD criteria are identical for the optimal structural maximization. However, for three or more tranches we show that the PD criterion always leads to strictly higher total deal values, mainly because the PD criterion encourages structural maximization but the EL criterion does not.

(2) How effective is the information passed from the rating agency to the investors in the presence of structural maximization? We prove that, in a market of model-based investors, who use rating information to calibrate their model-based deal prices, PD gives more advantage to the issuer than EL, as measured by the information gap. This phenomenon is more pronounced when the issuer has substantially more private information about the potential loss being very risky than the investors have.

(3) What is the representation of credit rating criteria (such as EL) that does not encourage structural maximization? To answer this question, we propose an axiom of self-consistency that does not allow issuers to gain, by tranching financial securities, from investors who rely on the rating criterion.[2] We give a representation of *all* self-consistent rating criteria via a Choquet integral. We show that the EL criterion used by Moody's satisfies the self-consistency, whereas the PD criterion used by S&P does not. To our best knowledge, this is the first paper on the axiomatic foundation for credit rating criteria. Our axiomatic framework suggests some new self-consistent and scenario-relevant rating criteria beyond the PD and EL criteria.

In summary, the contribution of the paper is three-fold. First, we give a comprehensive study of the impacts of exploring rating criteria to do structural maximization, and point out their empirical implications. Second, we study how effective different rating criteria pass information to the investors through letter-valued rating categories, and find some limitations of the PD criterion in this context. Third, we provide an axiomatic foundation for credit rating criteria that is self-consistent and does not encourage structural maximization.

1.2. Literature Review

Besides Brennan et al. (2009) and Hull and White (2012) this paper is related to two strands of literature. First, after the 2008 financial crisis there is a growing literature on rating seeking behaviors and competition in the market for ratings, such as rating shopping (Skreta and Veldkamp

[2] The focus of this paper is not to discuss how to obtain statistical quantities such as modeling or estimating default probability or loss distribution. Our focus is on a later (but also crucial) step: After all relevant statistical quantities are available, how does one assign a rating?

2009, Bolton et al. 2012, Sangiorgi and Spatt 2017), rating inflation and competitions (Frenkel 2015, Chu and Rysman 2019), informed intermediation (DeMarzo 2005), adjustments to credit rating agency models (Griffin and Tang 2012), rating catering (Griffin et al. 2013), and the regulatory advantage of highly rated securities (Opp et al. 2013). There is also a large literature studying the quality and suitability of credit rating results for investment and risk management purposes, especially after the 2008 financial crisis. For example, Cornaggia and Cornaggia (2013) compare the stability and timeliness of ratings in predicting default risk; Cornaggia et al. (2017) test whether ratings are comparable across asset classes; Giesecke et al. (2014) study the macroeconomic effects of bond market crises; Wojtowicz (2014) examines the sufficiency of credit ratings for pricing Collateralized Debt Obligations (CDOs); Hilscher and Wilson (2017) investigate the information in corporate ratings on default probability and systematic risk. Dimitrov et al. (2015) analyze the adverse effect of Dodd-Frank on credit ratings, suggesting the quality of corporate bond ratings becomes lower after the Dodd-Frank act. In a broader context, Azizpour et al. (2018) investigate the sources of corporate default clustering in U.S.; Keppo et al. (2010) study impact of the governmental market risk requirement on the default probability of the bank; Chen et al. (2017) show how the design and incentive effects of contingent convertible debt can reduce debt-induced collapse. We complement this strand of literature by focusing on the rating criteria directly.

Bolton et al. (2012) introduce two types of investors for rated securities: trusting and sophisticated investors. Their distinction is whether the investor takes the ratings at face value. Bolton et al. (2012) argue that both investors coexist in the market due to different incentives and roles. We also have two types of investors, both of which trust that the rating is provided truthfully; the simple investors use the rating directly for pricing, whereas the model-based investors use the rating to get information on the quality of the investment. Our model-based investors resemble the idea behind investors analyzed by DeMarzo and Duffie (1999) and DeMarzo (2005), who do not question the moral hazard of intermediaries (e.g., Malamud et al. 2013), but instead use information, such as retention levels in DeMarzo and Duffie's case, and ratings in our case, as a signal of the credit quality of the underlying assets. The pricing mechanism of such investors uses conditional expectation and is similar to that of Kyle (1985) with insider information. Consequently, we address the problem of designing securities with asymmetric information. Other equilibrium pricing models are considered by Bongaerts et al. (2011) and Toda (2015) without focusing on information asymmetry.

Second, there is a large stream of research on the characterization of decision criteria (e.g., Yaari (1987), Maccheroni et al. (2006)), which is closely connected to the axiomatic approaches for risk measures since the work of Artzner et al. (1999); see Cerreia-Vioglio et al. (2011), Chen et al. (2013), Wang and Zitikis (2021) and the references therein for more recent progress. In particular, our methodology is closely related to the characterization of Choquet integrals as studied by, e.g., Schmeidler (1986), Yaari (1987), Kou and Peng (2016). Standard textbooks on risk management with risk measures are McNeil et al. (2015) and Föllmer and Schied (2016).

We complement this stream of theoretical literature in several ways: (a) In the classic literature, Choquet integrals are typically characterized via so-called comonotonic additivity (Schmeidler 1986, Yaari 1987), which we do not assume. Instead, we show that self-consistency leads to a property on the corresponding rating measure which is weaker than comonotonic additivity, but is still sufficient to produce the representation. (b) Axioms for rating criteria are economically designed from the perspectives of rating agencies and investors rather than the perspective of the regulators, which is common in the literature of risk measures. (c) Unlike common risk measures such as the Value-at-Risk (VaR) or the Expected Shortfall (ES), credit ratings are categorical (e.g., letter-valued). This leads to substantial technical difficulties as well as novel mathematical treatment. (d) To be consistent with rating practice, we study the characterization of non-law-based rating criteria, which extends some recent results of scenario-based risk measures in Wang and Ziegel (2021). Scenario-relevant rating criteria are able to incorporate risk factors beyond loss distribution such as credit risk correlation, an important issue in credit rating; see Bai et al. (2015), Nickerson and Griffin (2017), Hilscher and Wilson (2017) and Monfort et al. (2021).

The rest of this paper is organized as follows. After some preliminary results in Section 2, Section 3 shows the consequence of structural maximization. In Section 4, we propose a model to analyze the efficiency of different rating criteria in terms of passing private information to the investor. An axiom of self-consistency is proposed in Section 5. We characterize all self-consistent rating criteria in Section 6. Extensions to law-invariant and scenario-relevant rating criteria are given in Section 7, with some examples given in Section 8. Section 9 concludes the paper with several remarks and potential future directions. Some preliminary empirical evidence and support to our theory are reported in Appendices A, B, and C in the online supplement. Although our paper primarily focuses on the tranching effects of rating criteria in the procedure of securitization, the pooling effects under the PD and EL criteria are discussed in Appendix D via a concept of pooling effect consistency. Some additional examples and all proofs are put in Appendices E and F, respectively.

2. Basic Settings and Some Preliminary Results

In this section, we shall present basic settings, definitions and examples for our theoretical framework. There are three groups of agents: rating agencies, security issuers and security investors. In particular, the rating criterion (e.g. PD or EL) used by a rating agency can affect both issuers and investors; see the following graph for an illustration, which will be explained in detail in this section and the following two sections.

Rating agencies (rating criteria) ↗ { Simple investors (rating-based prices); Model-based investors (rating-calibrated prices) }
↘ Issuers (tranching schemes)

We first introduce rating criteria used by the rating agencies, and then the problem of structural maximization of the issuers. The two types of investors will be studied separately in different markets, leading to different yet related implications. Markets with simple investors and model-based investors will be studied in Section 3 and Section 4, respectively.

2.1. Rating Criteria

Credit ratings are assigned to defaultable securities. Consider a one-period model for a defaultable security with a nominal value $M \in \mathbb{R}_+$ and a random loss $L \in \mathcal{L}$, where $\mathbb{R}_+ = (0, \infty)$ and $\mathcal{L}$ is the set of bounded random variables in a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For notational convenience, we denote the set of all defaultable securities by $\mathcal{X}$, namely,

$$\mathcal{X} = \{(L, M) \in \mathcal{L} \times \mathbb{R}_+ : 0 \leqslant L \leqslant M\},$$

and we denote by $\mathcal{L}_1$ the set of normalized losses, i.e., $[0,1]$-valued random variables.

In practice, credit ratings are categorical, i.e., letter-valued. To assign a credit rating to a defaultable security in $\mathcal{X}$, we categorize it into one of n classes $(1, \dots, n)$, where $n \geqslant 2$. Formally, a *rating criterion* is a mapping from $\mathcal{X}$ to $(1, \dots, n)$, denoted by I, satisfying *nominal-invariance*:

$$I(\lambda L, \lambda M) = I(L, M), \quad (L, M) \in \mathcal{X},\ \lambda > 0,$$

and *monotonicity*:

$$\text{For } (L_1, M), (L_2, M) \in \mathcal{X}, \quad L_1 \leqslant L_2 \implies I(L_1, M) \leqslant I(L_2, M).$$

The nominal-invariance requirement states that the rating of a defaultable security is independent of its nominal value or the amount of shares being rated. Consequently, to describe a rating criterion, we only need to consider defaultable securities with a nominal value 1, and we simply write this as $I(L) = I(L, 1)$ for $L \in \mathcal{L}_1$. The monotonicity requirement says that a defaultable security with a surely larger (per-unit) loss cannot have a better rating.

For a rating criterion I, we write $I_k = \{(L, M) \in \mathcal{X} : I(L, M) = k\}$, $k = 1, \dots, n$, which represents the set of securities rated k. Without loss of generality, we assume that the sets $I_1, \dots, I_n$ are non-empty, and they are ordered such that I_1 represents the best rated securities (e.g. AAA), and I_n represents the worst rated securities. In many structured finance securities, the last tranche is not rated, and we include such non-rated ones simply in the last category I_n.

A rating criterion is often generated by first giving a rating "score" $\rho(L,M)$ to a defaultable security (L,M) and then assign a categorical credit rating according to predetermined intervals the score falls in. Formally, let $\pi_n(\mathbb{R})$ be the set of n-partitions of $\mathbb{R}$. We say that the rating criterion I is generated by a rating measure $\rho:\mathcal{L}_1\to\mathbb{R}$ if, for some ordered partition[3] $(J_1,\dots,J_n)\in\pi_n(\mathbb{R})$,

$$I(L,M)=k \;\Leftrightarrow\; \rho(L/M)\in J_k, \quad k=1,\dots,n. \tag{1}$$

Note that (1) is equivalent to

$$I_k=\{(L,M)\in\mathcal{X}:\rho(L/M)\in J_k\}, \quad k=1,\dots,n. \tag{2}$$

Below are two major examples of rating criteria used throughout this paper, both of which are generated by rating measures.

Example 1 (The PD criterion). The PD criterion I satisfies that, for some numbers[4] $q_0<0<q_1<\cdots<q_n=1$,

$$I_k=\{(L,M)\in\mathcal{X}:\mathbb{P}(L>0)\in(q_{k-1},q_k]\}, \quad k=1,\dots,n.$$

In other words, the PD criterion is generated by $\rho(X)=\mathbb{P}(X>0)$.

Example 2 (The EL criterion). The EL criterion I satisfies that, for some numbers $q_0<0<q_1<\cdots<q_n=1$,

$$I_k=\{(L,M)\in\mathcal{X}:\mathbb{E}[L/M]\in(q_{k-1},q_k]\}, \quad k=1,\dots,n.$$

In other words, the EL criterion is generated by $\rho(X)=\mathbb{E}[X]$.

Since $n\geqslant 2$, in both examples $q_1\in(0,1)$; that is, there are at least two different rating categories. The above examples of PD and EL criteria are simplistic and hence very convenient for theoretical analysis. When combined with the concept of scenario relevance, they become the practical rating methods used by S&P and Moody's, respectively, which are formally introduced in Examples 3 and 4. The properties of the generalized rating criteria below are very similar to the PD and the EL criteria, and for this reason we took PD and EL as our primary examples.

Example 3 (S&P's rating criterion). Standard and Poor's (2016) specifies disjoint scenarios S_j, $j=1,\dots,s$ to reflect different economic situations (modeled by random events), from the most adverse (S_1) to the safest (S_s). A security (L,M) is given a rating $k\in\{1,\dots,s+1\}$ (i.e., $n=s+1$) if it can survive scenarios $S_k,\dots,S_s$ but not S_{k-1} (the latter is not needed if $k=1$).

[3] A partition of $\mathbb{R}$ is *ordered* if for each k, $x<y$ for all $x\in J_k$ and $y\in J_{k+1}$.

[4] In both Examples 1 and 2, we impose $q_0<0$ to interpret $(q_0,q_1]$ simply as $[0,q_1]$, noting that the loss L is non-negative.

For instance, an AAA-rated security should survive the most adverse scenario S_1, namely $\mathbb{P}(L > 0|S_1) = 0$.[5] Mathematically speaking, it is

$$I(L,M) = \max\{k \in \{1,\dots,s\} : \mathbb{P}(L > 0|S_k) > 0\} + 1, \quad (L,M) \in \mathcal{X}, \tag{3}$$

where we use the convention $\max\varnothing = 0$. Assume that, for each $k \in \{1,\dots,s-1\}$, L has a larger loss under scenario S_k than its loss under scenario S_{k+1}; this assumption is reasonable for most defaultable securities since S_k is more severe than S_{k+1}. Then, under this assumption, we can show that[6] (3) is precisely a PD criterion in Example 1, with $q_k = 1 - \sum_{j=k}^{s} \mathbb{P}(S_j)$, $k = 1,\dots,s$. Hence, the rating criterion used by S&P can be seen as a scenario-based version of the PD criterion. Later we shall show that this is a special case of generalized PD criteria.

Example 4 (Moody's rating criterion). Moody's (2017) considers three scenarios with different correlation matrices $\Sigma_1, \Sigma_2, \Sigma_3$ to represent three separate states of low, medium, and high correlations (p. 6 therein).[7] Mathematically, the three scenarios are chosen as $S_j = \{\Sigma = \Sigma_i\}$, $j = 1,2,3$, where Σ is the correlation matrix for the portfolio. Weights associated with each scenario are specified as $(\lambda_1, \lambda_2, \lambda_3) = (0.7, 0.2, 0.1)$. The rating measure of a security (L,M) is calculated as

$$\rho\left(\frac{L}{M}\right) = \sum_{j=1}^{3} \lambda_j \mathbb{E}\left[\frac{L}{M}|S_j\right], \tag{4}$$

and the rating is assigned according to the above quantity. Thus the rating criterion used by Moody's can be seen as a scenario-based version of EL criterion in Example 2.

We can check that the rating criterion in Example 4 is equal to EL with a different probability measure (representing weighted stressed scenarios) Q specified by $\mathrm{d}Q/\mathrm{d}\mathbb{P} = \sum_{i=1}^{n} \lambda_j/\mathbb{P}(S_j)\mathbb{1}_{S_j}$. Hence, this rating criterion shares the same theoretical properties as EL.

We define a class of rating criteria based on a property satisfied by the PD criterion. A rating criterion is called a *generalized PD criterion* if it satisfies the tranche-cut invariance property:

$$I(L \wedge K, K) = I(L,M) \text{ for all } (L,M) \in \mathcal{X} \text{ and } K \in (0,M]. \tag{5}$$

Property (5) is satisfied by the PD but not the EL criterion, which can be checked directly by definition. The scenario-based PD criterion in Example 3, which closely resembles the practice of S&P, also satisfies (5); see Lemma EC.1 in the online supplement for this assertion. The property (5) will be helpful to establish several results on the PD criterion.

[5] For simplicity, here we let the highest rating category contain only the security which survives the most adverse scenario, and thus 0 probability of default. In practice, the highest rated securities also have a very small probability of default.

[6] To show this, note that under this assumption for $k = 1,\dots,s$, the following two conditions (i) $\mathbb{P}(L > 0|S_k) > 0$ and $\mathbb{P}(L > 0|S_{k+1}) = 0$ (where we set $\mathbb{P}(L > 0|S_{s+1}) = 0$); (ii) $\mathbb{P}(L = 0) \geqslant \sum_{j=k+1}^{s} \mathbb{P}(S_j)$ (where we set the sum to 0 if $k = s$) and $\mathbb{P}(L = 0) < \sum_{j=k}^{s} \mathbb{P}(S_j)$, are equivalent. Let $q_k = 1 - \sum_{j=k}^{s} \mathbb{P}(S_j)$, $k = 1,\dots,s+1$. The claim follows because (i) is equivalent to $I(L,M) = k+1$, and (ii) is equivalent to $\mathbb{P}(L > 0) \in (q_k, q_{k+1}]$.

[7] Moody's states that the correlation setting "reflects the historical default patterns observed back to the 1920s and allows us to better match losses in the tail of the distribution" (p. 4).

2.2. Issuers: Choosing Tranching Schemes and Rating Criteria

A firm can issue different classes of debts with different payment priorities. For example, instead of a single debt (L,M), a firm can issue subordinated debts for a fixed level $K \in (0,M)$, e.g., a senior debt with nominal value $M-K$ and random loss $(L-K)_+$, and a junior debt with nominal value K and random loss $L \wedge K$.[8] More generally, in structured finance, an issuer can use multi-layer tranching schemes. Precisely, a *tranching scheme* of (L,M) is presented by a set $\mathbf{K} = \{K_1, ..., K_m\} \subseteq [0,M]$, where each K_j is a tranche level. We always use the order $M > K_1 > \cdots > K_{m-1} > K_m = 0$, and there are m tranches of $\mathbf{K}$ in total. The j-th tranche has unit loss

$$L_j = (L-K_j)_+ \wedge (K_{j-1} - K_j)/(K_{j-1} - K_j), \qquad j = 1, ..., m, \tag{6}$$

with L_1 being the loss of the most senior tranche, and $K_0 = M$.

For a defaultable security (L,M), the most important quantity for the issuer is the total deal value. A defaultable security (L,M), a tranching scheme $\mathbf{K}$, and a rating-implied price vector $\mathbf{P}$ together constitute a deal. Denoting by P_j the unit price of the j-th tranche for a tranching scheme $\mathbf{K}$, the total deal value is given by

$$\sum_{j=1}^{m} P_j (K_{j-1} - K_j). \tag{7}$$

The set of all tranching schemes for defaultable securities with nominal amount M is denoted by $\mathcal{T}_M$, and for simplicity write $\mathcal{T} = \mathcal{T}_1$. The prices $P_1, \ldots, P_m$ depend on both the rating criterion I and the tranching scheme $\mathbf{K}$, and they will be specified later in Sections 3 and 4.

For a given defaultable security (L,M), an issuer prefers a higher total deal value in (7). In other words, the issuer has incentives to maximize the deal value by choosing a tranching scheme $\mathbf{K}$ and a rating agency (thus a rating criterion).

We formally define structural maximization, as one of the key points of this paper.

Definition 1. *Structural maximization* is the optimization of (7) by an issuer via choosing (a) a tranching scheme (i.e., $\mathbf{K}$) or (b) a tranching scheme and a rating criterion (i.e., both $\mathbf{K}$ and I).

Note that in a problem of structural maximization, the tranching schemes may be chosen freely from $\mathcal{T}_M$, whereas there are only several rating criteria (e.g., those of S&P, Moody's and Fitch) to choose from. In the next two sections, we will study structural maximization in markets with different types of investors.

[8] In this paper, $x_+ = \max\{x, 0\}$ and $x \wedge y = \min\{x, y\}$.

3. Simple Investors and Structural Maximization

Each investor of defaultable securities has a subjective pricing scheme, representing the price at which they are willing to trade defaultable securities. An investor may or may not use the rating information in the subjective pricing. As we are studying rating in the paper, we shall focus on those investors who rely on the rating information. In this section, we consider a market with simple investors, who price each tranche purely on the rating given to the tranche without any other pricing models. In the next section we shall consider model-based investors, who use rating information to compute a conditional model to price securities.

We shall discuss the external and internal implications of structural maximization with a special focus on the PD and the EL criteria. The external implication refers to the issuer's choice of a rating agency (or a rating criterion). The internal implication refers to the issuer's design of the tranching scheme $\mathbf{K}$ for a given rating criterion, thus a known rating-implied subjective price $\mathbf{P}$.

3.1. Rating-Implied Prices

If an investor uses the rating information, a natural requirement is that the investor should agree that an AAA-rated security is worth more than an AA-rated security; this is the key connection between subjective prices and credit ratings. For a chosen rating criterion I, we say that a pricing scheme is *rating-implied* (or I*-implied*, to emphasize its reliance on I) if the unit price of any defaultable security is determined by its rating, and the price increases as the rating gets better.[9] In this case, two defaultable securities with the same rating have the same unit prices. A rating-implied price will be represented by a *price vector* $\mathbf{P} = (P_1, \ldots, P_n) \in [0, \infty)^n$ with components in a strictly decreasing order, which each P_i stands for the unit price of securities in the i-th rating category.[10] Using the price vector $\mathbf{P}$, the price of a defaultable security (L, M) is given by $MP_{I(L/M)}$. The investor who uses a rating-implied pricing scheme will be called a *simple investor*.

Putting the rating-implied price vector $\mathbf{P}$ into (7), the total deal value for the issuer in the market of simple investors is given by

$$V(\mathbf{K}, \mathbf{P}, I) = \sum_{j=1}^{m} P_{I(L_j)}(K_{j-1} - K_j). \tag{8}$$

The value in (8) will be maximized by the issuer in a problem of structural maximization.

[9] A large body of research suggests that there may be a significant proportion of investors who use rating-implied prices: (i) SEC (2014) describes the structured finance market as a "rated" market, in which "the valuations of the products depend significantly on credit ratings" (p.12). (ii) Cuchra (2005) reports that ratings explain 70-80% of launch spreads on structured products in Europe. (iii) Ashcraft et al. (2011) find that the nonprime MBS prices are excessively sensitive to credit ratings, relative to the informational content of ratings. (iv) Downing et al. (2009) provide empirical evidence for the lower quality ("lemons") of assets securitized into mortgage-backed securities. The asymmetry in information may promote the use of ratings as a basis of pricing for uninformed investors. (v) Partnoy (2010) points out investors tend to have overdependence on credit ratings.

[10] We allow $P_i > 1$ to accommodate possibly negative yields. If desirable, one may also assume $P_i \in (0, 1)$, which does not affect any of our discussions.

3.2. Dominance Relation of Generalized PD Criteria

Two different rating criteria may assign different ratings to the same defaultable security, and they may have different numbers of rating categories. For a given security (L,M), two rating criteria I and I' are called *comparable* if

$$I((L-K)_+, M-K) = I'((L-K)_+, M-K) \text{ for all } K \in [0, M). \tag{9}$$

In other words, two rating criteria are comparable if they give the same rating for the most senior tranche of (L,M), regardless of the tranche level. The assumption (9) on comparability is only used in this section. Note that two rating criteria I and I' do not need to agree on the rating of tranches other than the most senior one, or even the total number of rating categories.

The following proposition extends the dominance results in Brennan et al. (2009) and Hull and White (2012) from PD to the family of generalized PD criteria.

PROPOSITION 1. *Suppose that $(L,M) \in \mathcal{X}$, the rating criteria I and I' satisfy (9), and I satisfies property (5). For all price vectors $\mathbf{P}$ and tranching schemes $\mathbf{K}$, we have $V(\mathbf{K},\mathbf{P},I) \geqslant V(\mathbf{K},\mathbf{P},I')$.*

The next proposition gives a necessary and sufficient condition for the family of generalized PD criteria.

PROPOSITION 2. *A rating criterion I satisfies (5) if and only if*

$$V(\mathbf{K}',\mathbf{P},I) \geqslant V(\mathbf{K},\mathbf{P},I) \text{ for all } \mathbf{K} \subseteq \mathbf{K}', \text{ all price vectors } \mathbf{P}, \text{ and all securities } (L,M). \tag{10}$$

For such I, $V(\mathbf{K}',\mathbf{P},I) > V(\mathbf{K},\mathbf{P},I)$ as soon as tranches of $\mathbf{K}' \supseteq \mathbf{K}$ have more rating categories than those of $\mathbf{K}$.

Proposition 2 shows that for a generalized PD criterion the deal value is always improved or unchanged by further slicing tranches arbitrarily. The improvement is strict if a tranche with a new rating is created.

Note that the above two propositions do not imply that the issuer would prefer generalized PD criteria over other criteria, because the inequality in Proposition 1 may become an equality and the strict inequality in Proposition 2 is not about the comparison between two rating criteria I and I'. In the next subsection, we shall introduce structural maximization to study the strict dominance between two rating criteria.

3.3. External Implication: Issuer's Selection between Rating Criteria

We first focus on the choice for the issuer between two rating criteria for structural maximization in the sense of Definition 1 (b). For this, we need to establish a link between rating criteria and the rating-implied prices.

For a given security (L,M), we denote by N_I the maximal number of distinct rating categories over all tranching schemes of (L,M) under I.

THEOREM 1. *For $(L,M)\in\mathcal{X}$ consider all rating-implied subjective price vectors $\mathbf{P}$ and tranching schemes $\mathbf{K}$.*

(i) Suppose the rating criteria I and I' satisfy (9), and I is a generalized PD criterion, i.e., satisfying (5), and I' is the EL criterion, and $N_I\geqslant 3$. Then

$$\sup_{\mathbf{K}\in\mathcal{T}_M} V(\mathbf{K},\mathbf{P},I) > \sup_{\mathbf{K}\in\mathcal{T}_M} V(\mathbf{K},\mathbf{P},I'). \tag{11}$$

(ii) For any two rating criteria I, I' (not necessarily related to either EL or PD) satisfying (9), if $N_I, N_{I'}\leqslant 2$, then

$$\sup_{\mathbf{K}\in\mathcal{T}_M} V(\mathbf{K},\mathbf{P},I) = \sup_{\mathbf{K}\in\mathcal{T}_M} V(\mathbf{K},\mathbf{P},I'). \tag{12}$$

In particular, if there are only two possible rating categories or two differently rated tranches, then there is essentially no difference between the generalized PD and EL when V is maximized.

For a fixed defaultable security, the implication of Theorem 1 is that if a generalized PD criterion is comparable to an EL criterion, and the rating-implied price vector $\mathbf{P}$ is the same for each identical rating category, then PD is always favored by the issuer due to a higher deal value (8).

3.4. Internal Implication: Issuer's Maximization Strategy Given a Rating Criterion

For a given rating criterion, an issuer attempts to optimize the total tranche number and tranche thresholds in the sense of Definition 1 (a). The next theorem finds the optimal total tranche number n and the optimal tranche thresholds $K_1,\ldots,K_n$ for a given generalized PD criterion.

THEOREM 2. *Suppose that the rating criterion I is a generalized PD criterion, i.e., satisfying (5). For any rating-implied subjective price vector $\mathbf{P}$ and any security (L,M):*

(i) If a tranching scheme maximizes the deal value (8), then it has the maximum number (i.e., $n=N_I$) of distinct rating categories.

(ii) If a maximizer for (8) exists, then there is a unique maximizer with exactly N_I tranches, and it is given by $\mathbf{K}=\{k_1,\ldots,k_n\}$, where[11]

$$k_j = \inf\left\{k\geqslant 0 : I\left(\frac{(L-k)_+}{M-k}\right)\leqslant j\right\},\quad j=1,\ldots,n. \tag{13}$$

(iii) A maximizer for (8) always exists if I is the PD or the scenario-based PD criterion.

The internal implication of the above theorem is that if investors use a generalized PD criterion as the main source of their subjective pricing, then the issuer would have the incentives to create as many tranches as possible. As we will see later, this is not the case for the EL criterion; indeed a

[11] The set $\{k_1,\ldots,k_n\}$ has N_I elements as we count equal numbers in a set only once.

maximizer to (8) may have fewer tranches than N_I and it is not obtained by the way described in Theorem 2 (ii).[12] A more general result in Theorem 6 says that if $n \geqslant 3$, then for any price vector, there exists some defaultable securities such that a maximizer of (8) for the EL criterion has fewer than N_I thanches.

3.5. Empirical Implications of Using the PD Criterion

In summary, Theorems 1 and 2 lead to two empirical implications.

Implication (i), Bigger deal value than competitors: Issuers prefer a generalized PD criterion over an EL criterion to gain more total deal value using the same price vector, although the total deal of all tranches as a whole does not become safer by choosing a particular credit rating criterion. This result is shown in Theorem 1. This is consistent with the empirical evidence in both Appendix A.2 and Appendix C.

Implication (ii), Creating many tranches: Under a generalized PD criterion, an optimal (maximizing the deal value) tranching scheme of any security always has as many rating categories as possible. Thus, this may lead to excessive issuance of tranches and over-securitization of the financial asset. This result is shown in Theorem 2. This is consistent with the empirical evidence in Appendix A.3.

Some remarks are needed to clarify these implications. First, just like the definition of arbitrage, these hidden features do not imply something really harmful unless issuers can scale up the tranche offerings significantly. Second, the absence of Implications (i) and (ii), e.g., using the EL criterion, does not mean that there is no way for issuers to gain from structural maximization, particularly if issuers are willing to take some risk. Indeed, as shown in the Brennan et al. (2009) and also confirmed in our study in Appendix C using new data in the post-Dodd-Frank period, there are gains from structural maximization even under the EL rating criterion, although the gains are much smaller than that under the PD criterion, consistent to our theory.[13] What is implied here is that using the generalized PD criteria there is a riskless (easy) way for the issuer to take advantage of the credit rating to gain from a group of investors who believe in the rating.

[12] For a very simple example, take a defaultable security $(L, 1)$ with fixed loss $L = q_1$ where q_1 is the first threshold in Example 2. The trivial tranching scheme of only one tranche would be 1-rated, thus having the highest possible deal value in (8). By tranching differently (e.g., with $K = q_1$), one can create a junior tranche with sure full loss which will be rated the worst, and a senior tranche with no loss which will be rated the best, thus $N_I \geqslant 2$. This shows that a maximizer for (8) does not have N_I different tranches.

[13] The definition of structural maximization does not specify the magnitude of the gains from structural maximization. To evaluate the magnitude of the potential gains for the issuers, one has to assume a particular economic model, e.g., the classical Merton's structural model, and calibrate or estimate its parameters.

4. Model-Based Investors and Information Efficiency

Instead of the simple investors in Section 3 whose price for a security is purely determined by the rating of that security, we considered in this section *model-based investors* in the sense that they use rating information to compute a conditional model to price securities. In particular, we analyze the efficiency of a rating criterion in passing information to the investor, based on a different pricing scheme of the investor. More precisely, The main message from this section is threefold: (1) PD is generally less efficient in passing information from the rating agency to the investor. (2) PD typically has a larger information gap than EL. (3) Maximizing the number of tranches may no longer be optimal for PD in this setting, as the issuer may have incentives to pass less information to the investor by choosing a small number of tranches.

4.1. Information Model and Calibrated Prices by Investors

There are an issuer, a rater, and a (representative) investor, which we refer to as agents. The information available to the issuer and the rater is modeled by a random variable Z, which may represent some private information about the company's value or riskiness. We assume that Z takes values in an interval $D \subseteq \mathbb{R}$. The realized value of $Z = z$ is not available to the investor, but all agents know the joint distribution of L and Z.

This setting is similar to DeMarzo and Duffie (1999), where the investor cannot see the realized value of the hidden variable. The novelty of our model is the inclusion of the rating criteria, and the information of Z will be passed to the investor through the rating. More precisely, to model the information asymmetry between the issuer, the rater, and the investor, we shall apply rating criterion I to update the conditional distribution of the loss given a hidden variable. To do this, we assume the rating criterion is determined by the loss distribution:

[LI] *Law invariance*: $I(L_1, M) = I(L_2, M)$ for all $(L_1, M), (L_2, M) \in \mathcal{X}$ satisfying $L_1 \overset{d}{=} L_2$, where $\overset{d}{=}$ stands for equality in distribution.

Both PD and EL in Examples 1 and 2 are law-invariant. Without loss of generality, we take $M = 1$, and let $L \in \mathcal{L}_1$ be a random loss. Furthermore, since Z represents a hidden factor of riskiness, it is natural to impose the following assumption.

Assumption 1. *The function $z \mapsto \mathbb{P}(L > x \mid Z = z)$ is strictly increasing at each $x \in (0,1)$ satisfying $\mathbb{P}(L > x \mid Z = z) \in (0,1)$. In other words, a larger value of z makes L strictly riskier.*

The condition $\mathbb{P}(L > x \mid Z = z) \in (0,1)$ indicates that if $\mathbb{P}(L > x \mid Z = z) = 0$ or 1 then we do not require it to increase strictly in z, allowing for L that is bounded away from 0 or 1. The dependence between (L, Z) satisfies positive regression dependence in statistics; see Lehmann (1966) and Benjamini and Yekutieli (2001). Below are four simple examples of such a factor model. All of these four models satisfy Assumption 1 and will be used later in our numerical examples.

Example 5 (Conditional Beta loss). Suppose that $Z>0$ and L given $Z=z$ follows a Beta distribution Beta$(z,1)$. This model admits a simple representation $L=U^{1/Z}$ where U is uniformly distributed on $[0,1]$ independent of Z.

Example 6 (Additive factor model). In an additive factor model, L is given by $L=f(g(U)+h(Z))$ where Z and U are independent taking values in $(0,1)$ and f,g,h are increasing functions. We assume that U is continuously distributed. For instance, by choosing $f(x)=\exp(x)$, $g(x)=\log(x)$ and $h(x)=\theta\log(x)$ for some $\theta>0$, we get the model $L=UZ^{\theta}$.

Example 7 (Full information model). An extreme case of information asymmetry is that the issuer and the rater have full information about the loss L; that is, L is a function of Z. Without loss of generality, we can assume $L=Z$. This model may be unrealistic in practice, but it will be helpful later to illustrate some key differences between PD and EL.

Example 8 (Exceedance model). Let $L=(L'-t)_+$ where L' is a loss that is strictly increasing in Z (in the sense of Assumption 1) and $t\in(0,1)$ is a threshold. This model allows for L to have a point mass at 0. For instance, using the loss $L'=U^{1/Z}$ in Example 5, the exceedance model is $L=(U^{1/Z}-t)_+$.

Next we explain the pricing mechanism in the model. Let $L\in\mathcal{L}_1$ be a loss and Z be a hidden factor satisfying Assumption 1. The realization z of Z is known to the issuer, who can choose a tranching scheme $\mathbf{K}=\{K_1,...,K_m\}\in\mathcal{T}$ where $1=K_0>K_1>\cdots>K_{m-1}>K_m=0$ as in Section 2. The unit loss of tranche j is given by (6). Moreover, let $X_j=1-L_j$, which is the unit payoff of the j-th tranche.

The rating agency also sees z. Each tranche is given a discrete rating $R_j(z)=I(L_j|z)$, where $I(L_j|z)$ represents the rating applied to the conditional distribution of L_j given $Z=z$. The rating vector is denoted by $R_{\mathbf{K}}(z)=(R_1(z),\dots,R_m(z))$, which clearly depends on $\mathbf{K}$.

The investor does not observe z, but can observe the tranche levels $\mathbf{K}$ and the rating vector $R_{\mathbf{K}}(z)$, from which a range of z can be determined. For a unit of the j-th tranche, X_j, the investor is willing to pay

$$P_j=\mathbb{E}[X_j\mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)]. \tag{14}$$

The price in (14) is a *rating-calibrated price* by the investor, which is different from the rating-implied price in Section 3.[14]

[14] As in DeMarzo and Duffie (1999), the representative investor is risk-neutral and uses the conditional expectation to price securities; the risk-neutral assumption is imposed to focus on the rating effects rather than those of risk aversion. Note that under risk neutrality, investors may not break even if the prices are only computed by the expectation; indeed, sellers and buyers cannot be both profitable if they are risk neutral. Similar settings of risk neutrality can be found in DeMarzo (2005) and Malamud et al. (2013), where a conditional expectation is used to model unobserved information. Similarly, to focus on the rating effects, we omit other information implied by a tranching scheme and assume that the tranching scheme affects the pricing of X_j only through (14).

4.2. Information Gap

As in (7), the objective of the issuer is to maximize its total deal value[15]

$$S(\mathbf{K},z,I)=\sum_{j=1}^{m}P_j(K_{j-1}-K_j).$$

We denote by $S^*(z,I)$ the maximum value of this quantity, that is,

$$S^*(z,I)=\sup_{\mathbf{K}\in\mathcal{T}}S(\mathbf{K},z,I).$$

There are two main differences between $S(\mathbf{K},z,I)$ in this section and $V(\mathbf{K},\mathbf{P},I)$ in (8). First, the observation z in this section does not appear in the model in (8). Second, the price $(P_1,\dots,P_m)$ relies on rating information to compute the conditional expectation in (14), whereas the rating-implied price $\mathbf{P}$ in (8) is a fixed vector representing the unit price of each rating category.

For a realized $z\in\mathbb{R}$, define the information gap (IG) that can be exploited by the issuer as

$$\mathrm{IG}_I(z)=S^*(z,I)-\mathbb{E}[1-L\mid Z=z]=\sup_{\mathbf{K}\in\mathcal{T}}S(\mathbf{K},z,I)-\mathbb{E}[1-L\mid Z=z],$$

where $\mathbb{E}[1-L\mid Z=z]$ is the expected value of the whole asset given $Z=z$, known to the issuer. The larger IG is, the less efficiently the rating passes the information of Z to the investor. Using the fact that $\sum_{j=1}^{m}X_j(K_{j-1}-K_j)=1-L$, the IG can be rewritten as[16]

$$\begin{aligned}\mathrm{IG}_I(z)&=\sup_{\mathbf{K}\in\mathcal{T}}\mathbb{E}\left[(1-L)\mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)\right]-\mathbb{E}[1-L\mid Z=z]\\&=\mathbb{E}[L\mid Z=z]-\inf_{\mathbf{K}\in\mathcal{T}}\mathbb{E}\left[L\mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)\right].\end{aligned}\tag{15}$$

By (15), it is clear that IG measures the gap between the expected loss under the accurate information $Z=z$ and the rating information $R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)$. *If all information has been passed to the investor via rating, then IG is zero, and the issuer has no gain from information asymmetry.*

Also by (15), $\mathrm{IG}_I(z)\leqslant\mathbb{E}[L\mid Z=z]$. Under Assumption 1, $\mathbb{E}[L\mid Z=z]$ is small when z is small. Therefore, for small values of z, there is not much to gain for the issuer due to information asymmetry. This is intuitively clear: When z is small, the private information of the investor is that the security is very safe, and hence there is not much advantage to take.

[15] In case $\mathbb{E}[X_j\mid Z=z]>P_j$, the issuer may choose to not make the transaction and reserve X_j to himself. Nevertheless, the value P_j can be interpreted as the market price of the tranche X_j, and the issuer likes to maximize the total deal price marked to the market.

[16] Our setting in (14) can be extended to include the pricing model $P_j=\theta\mathbb{E}[X_j\mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)]$ where $\theta\in(0,1]$ is an exogenously specified parameter. If $\theta<1$, then the investor demands some excess return from the security, and this will lead to a positive expected gain. This is especially relevant for the very risky junior tranches because such a trade can still be attractive to the issuer, even if it has a negative expected value, as the issuer can reduce uncertainty by liquidating some very risky assets. Introducing θ does not affect the technical results on structural optimization and their implications below, because it results in the same optimization problem in (15), by replacing $\inf_{\mathbf{K}\in\mathcal{T}}\mathbb{E}\left[L\mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)\right]$ with $\theta\inf_{\mathbf{K}\in\mathcal{T}}\mathbb{E}\left[L\mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)\right]$.

Denote by $A_{\mathbf{K}} = \{z' \in D : R_{\mathbf{K}}(z') = R_{\mathbf{K}}(z)\}$, which is the set of possible values taken by Z which are consistent with the rating information. The next result shows that this set is an interval,[17] leading to an upper bound on $\mathrm{IG}_I(z)$.

PROPOSITION 3. *Suppose that Assumption 1 holds. For any $\mathbf{K} \in \mathcal{T}$ and $z \in D$, we have that the set $A_{\mathbf{K}} = \{z' \in D : R_{\mathbf{K}}(z') = R_{\mathbf{K}}(z)\}$ observed by the investor is an interval containing z. Moreover, an upper bound on $\mathrm{IG}_I(z)$ is given by*

$$\mathrm{IG}_I(z) \leqslant \mathbb{E}[L \mid Z = z] - \mathbb{E}[L \mid Z \leqslant z]. \tag{16}$$

The intuition behind the upper bound (16) is that the interval $A_{\mathbf{K}}$ passed to the investor from the issuer cannot be better than $\{z' \in D : z' \leqslant z\}$ from the issuer's perspective since $z \in A_{\mathbf{K}}$. Later we will see that the bound (16) is sharp for the PD criterion under some additional assumptions.

Proposition 3 also illustrates that by choosing a tranching scheme, the rating of tranche j will indicate an interval A_j for the investor. Thus, the issuer discloses an interval $A_{\mathbf{K}} = \bigcap_{j=1}^m A_j$ the investor.

4.3. Comparing the Information Gap of PD and EL

We next focus on the EL and PD criteria in Examples 1-2. The next result shows that the information gap is always nonnegative under mild conditions for PD and EL. For this result, we use a simple condition that is satisfied by all examples listed above.

ASSUMPTION 2. *L given $Z = z \in D$ is either continuously distributed on $(0,1)$ possibly with a point mass at 0 (as in Examples 5-6) or a strictly increasing function of z (as in Example 7). Example 8 also satisfies the first case of this assumption if L' given $Z = z$ is continuously distributed.*

PROPOSITION 4. *Suppose that I is PD or EL and Assumptions 1 and 2 hold. Then, $\mathrm{IG}_I(z) \geqslant 0$ for each z satisfying $I(\mathbb{1}_{\{L>0\}}|z) > 1$.*

The condition $I(\mathbb{1}_{\{L>0\}}|z) > 1$ holds true if a junior tranche from level 0 to level ε is not rated the highest for some $\varepsilon > 0$, because $\mathbb{1}_{\{L>0\}}$ dominates the loss from one unit of the junior tranche. This requirement is mild as the junior tranche is never rated the highest in practice.[18] Note that $I(\mathbb{1}_{\{L>0\}}|z) > 1$ holds if either $I(L|z) > 1$ or $\mathbb{P}(L = 0 \mid Z = z) = 0$.

The intuition behind Proposition 4 is that the private information $Z = z$ available to the issuer will give an advantage to the issuer to sell the asset for a price higher than its expected value

[17] This interval may be open, closed, or neither, depending on the rating criterion I.

[18] Hypothetically, if $\mathbb{P}(L = 0 \mid Z = z) = 1$, then $I(\mathbb{1}_{\{L>0\}}|z) = 1$. In this case, there is no information advantage for the issuer since $\mathbb{E}[1 - L \mid Z = z] = 1$. Indeed, even if every tranche has the highest rating, investors may still believe that there is some small risk, i.e., possibly $\mathbb{P}(L = 0 \mid Z \in A_{\mathbf{K}}) < 1$ and $S^*(z, I) < 1$, leading to $\mathrm{IG}_I(z) < 0$.

given $Z=z$. Although part of this information can be passed to the investor through the rating $R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)$, the issuer may still gain because of the rating does not reveal the full information to the investors.

For I being the PD or the EL criterion in Examples 1-2, denote by

$$\delta_I=\max\{q_k-q_{k-1}:k=1,\dots,n\}\quad\text{and}\quad\gamma_I=1-q_{n-1}.$$

In other words, δ_I is the largest gap between thresholds, and γ_I is the difference between the thresholds of the lowest rating category. By definition, $\delta_I,\gamma_I\in(0,1)$. The next result shows that, with private information, PD gives the issuer more advantage than EL.

Theorem 3. *(i) For any EL criterion I', $\mathrm{IG}_{I'}(z)\leqslant\delta_{I'}$ for all (L,Z) and z.*

(ii) Consider a PD criterion I. Let (L,Z) satisfy the full information model $L=Z$ in Example 7 for a continuously distributed Z with positive density on $(0,1)$. Then, $\mathrm{IG}_I(z)\to 1-\mathbb{E}[L]$ as $z\to 1$.

(iii) Suppose that Assumptions 1-2 hold and $\mathbb{P}(L=0\mid Z=z)<\gamma_I$ for all z. Then, the information gap under the PD criterion is always larger than that of any other rating criterion. More precisely, $\mathrm{IG}_I(z)=\mathbb{E}[L\mid Z=z]-\mathbb{E}[L\mid Z\leqslant z]$ and $\mathrm{IG}_I(z)\geqslant\mathrm{IG}_{I'}(z)$ for all z and any rating criterion I'.

Theorem 3 (i) shows that for the EL criterion I', the information gap $\mathrm{IG}_{I'}(z)$ is bounded by $\delta_{I'}$. In particular, $\mathrm{IG}_{I'}(z)$ vanishes if $\delta_{I'}$ is sufficiently small. This suggests that IG can be controlled by using an EL criterion with rich rating categories. In contrast, Theorem 3 (ii) shows that, when $\mathbb{E}[L]$ is small, even if the PD criterion has rich rating categories, its IG can be very large (close to 1) for large values of z (i.e., when the private information of the issuer indicates that the security is very risky.)

The conclusion in Theorem 3 (iii) is the most useful. It shows that the upper bound (16) on IG in Proposition 4 is attained by any PD criterion, and hence its IG dominates that of *any other* criterion, not only EL. In addition to Assumptions 1 and 2, this result further requires $\mathbb{P}(L=0\mid Z=z)<\gamma_I$, which holds if L given $Z=z$ is continuously distributed (as in Examples 5 and 6).[19] In practice, γ_I is typically larger than 0.5 (in Table 6 of Standard and Poor's (2016), the lowest rating CCC has one-year default probability exceeding 0.3, yielding $\gamma_I=0.7$). A small value of $\mathbb{P}(L=0\mid Z=z)$ is realistic if L represents the normalized loss from a large pool of assets, which is common for structured finance securities. As we will see below in our numerical results, IG of PD may be smaller than that of EL in case $\mathbb{P}(L=0\mid Z=z)$ is large, which only happens when z is small; that is, the private information of the issuer says that the security is very safe.

To illustrate IG for PD and EL numerically, we consider four specific models from Examples 5-8. In (a), (b) and (d) below, $U,Z\sim\mathrm{U}(0,1)$ are independent.

[19] For PD, the condition $\mathbb{P}(L=0\mid Z=z)<\gamma_I$ is slightly stronger than the condition $I(\mathbb{1}_{\{L>0\}}|z)>1$ in Proposition 3. The latter is equivalent to $\mathbb{P}(L=0\mid Z=z)<1-q_1$ for PD.

(a) $L=U^{1/Z}$, a special case of Example 5;

(b) $L=UZ$, a special case of Example 6;

(c) $L=Z$ where $Z\sim\text{Beta}(1,9)$;[20]

(d) $L=(U^{1/Z}-0.1)_+$, a special case of Example 8;

The numerical results are presented in Figure 1.[21] As we can see from panels (a)-(c), IG of PD is larger than IG of EL for models (a)-(c), regardless of the thresholds chosen for these criteria, illustrating Theorem 3 (iii). Indeed, in panels (a)-(c), the curves of IG of PD attain the upper bound (16) in Proposition 4. Moreover, when z is small, $\text{IG}_I(z)$ is also small, consistent with the upper bound in (16). Panel (c) illustrates Theorem 3 (i) and (ii), from which we can see that IG of PD is close to 1 if the observed z is large, whereas the IG of EL is very small for all values of z. Finally, in panel (d), when $\mathbb{P}(L=0\,|\,Z=z)$ is large, i.e., the condition in Theorem 3 (iii) does not hold, IG of PD and EL may not dominate each other, and EL may have a larger IG for small values of z (the safe scenario). However, for large values of z (the risky scenario), PD has a larger IG.

We draw the following observations from our theoretical and numerical results. First, under some natural conditions, PD has a higher IG than EL. This is shown in Theorem 3 (iii) and panels (a)-(c) of Figure 1. Second, the IG of EL decreases as the rating categories become rich. This is shown in Theorem 3 (i) and panel (c) of Figure 1. Third, more information advantage of the issuer leads to a larger IG of the PD criterion, for which IG can be close to 1. Moreover, the private information of a more risky scenario gives a bigger IG to PD than to EL. This is shown in Theorem 3 (ii) and panels (c)-(d) of Figure 1.

These observations lead to the implication that PD typically gives more advantage than EL to the issuer by yielding a larger total deal value, which is consistent with Implication (i) in Section 3.5. This implication is supported by the empirical observations presented in both Appendix A.2 and Appendix C.

In contrast to the model in Section 3, maximization of the number of tranches no longer applies to PD due to information asymmetry. This is different from Implication (ii) in Section 3.5. The intuition behind this difference is that, when the value of z is large, it is beneficial for the issuer to use a small number of tranches to minimize the information passed to the investor.[22]

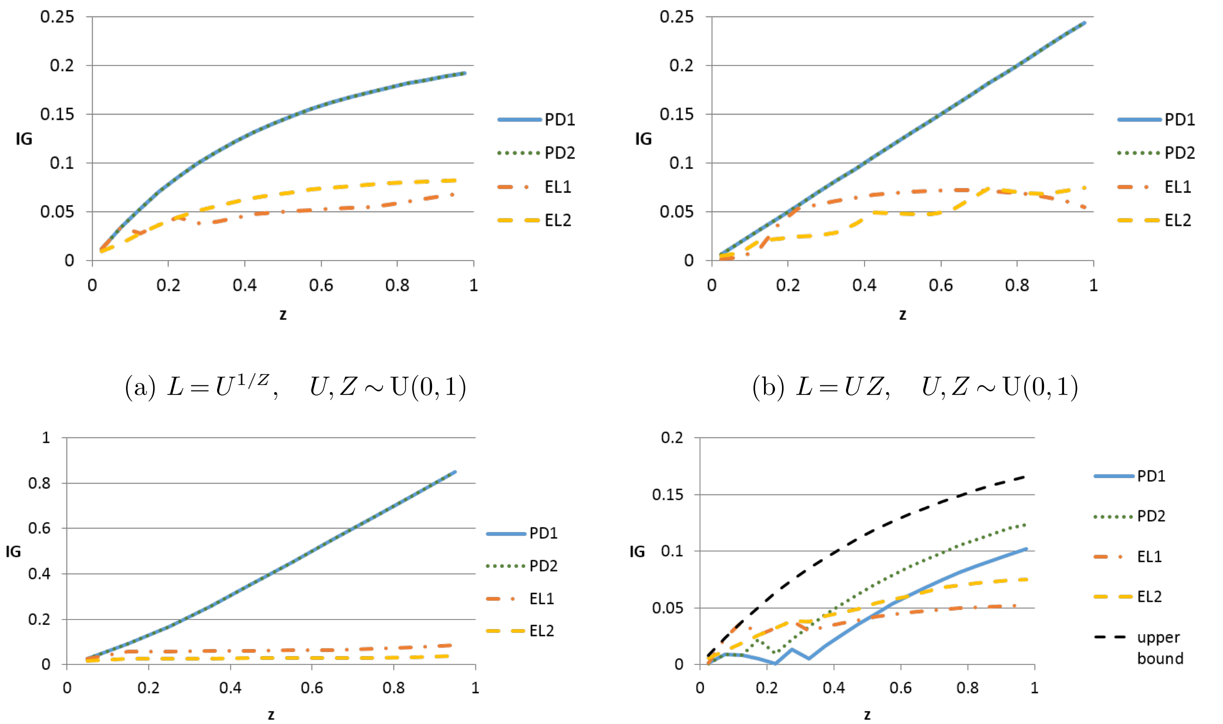
In a real-world market, it is reasonable to have both simple investors who rely solely on rating information and model-based investors. The empirical observations in Appendix A.3 and Appendix

[20] We choose a small $\mathbb{E}[L]$ to illustrate Theorem 3 (ii).

[21] All numerical results below are computed by the `optimization toolbox` of `MATLAB` among choices of $\mathbf{K}$.

[22] In the proof of Theorem 3, we show that for under some conditions, it suffices to use 2 tranches for PD to arrive at the theoretical upper bound (16) on $\text{IG}_I(z)$.

Figure 1 $\mathrm{IG}_I(z)$ **for PD and EL**



(a) $L = U^{1/Z}, \quad U, Z \sim \mathrm{U}(0,1)$

(b) $L = UZ, \quad U, Z \sim \mathrm{U}(0,1)$

(c) $L = Z, \quad Z \sim \mathrm{Beta}(1,9)$

(d) $L = (U^{1/Z} - 0.1)_+, \quad U, Z \sim \mathrm{U}(0,1)$

Notes. In panels (a), (b) and (d), both PD1 and EL1 have levels $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 1\}$, and both PD2 and EL2 have levels $\{0, 0.001, 0.01, 0.05, 0.1, 0.2, 0.35, 1\}$. In panel (c), both PD1 and EL1 have levels $\{k/10 : k = 1, \ldots, 10\}$ and both PD2 and EL2 have levels $\{k/20 : k = 1, \ldots, 20\}$. The upper bound (16) coincides with PD1 and PD2 in panels (a)-(c).

B suggest that there may be more investors that are similar to simple investors than to model-based investors.

5. Self-Consistent Rating Criteria

Using rating-implied pricing in Section 3 is convenient for discussing the behavior of the issuer, but may be overly simplistic, as not all securities with the same rating have the same yield. On the other hand, the model-implied subjective prices in Section 4 depend on the rating in an implicit way making it difficult to study. In this section, we will introduce a more flexible pricing scheme for general investors, not necessarily rating-implied or model-implied, based on a self-consistency axiom.

5.1. A Relaxation of Rating Consistent Price Schemes

The minimal requirement for a price to be consistent with a rating I is that a higher-rated security should have a higher unit price. To accommodate this, a subjective unit price is a functional $p : \mathcal{L}_1 \to [0, \infty)$ satisfying monotonicity:

$$p(L_1) \geqslant p(L_2), \quad \text{for all } L_1, L_2 \in \mathcal{L}_1 \text{ with } L_1 \leqslant L_2, \tag{17}$$

and p is *weakly rating-implied* if

$$I(L_1,1) < I(L_2,1) \;\Rightarrow\; p(L_1) > p(L_2), \quad \text{for all } L_1, L_2 \in \mathcal{X}. \tag{18}$$

In other words, a weakly rating-implied unit price p always gives a higher price to a security with surely less loss, and a higher price to a better-rated security.

Rating-implied prices in Section 3 are special cases of weakly rating-implied prices, in which the reverse direction of (18) also holds so that the unit price is solely determined by the rating and it can be represented by a price vector. In addition to simple investors, the condition (18) may also include some model-based investors using (14) in Section 4. For instance, when I is the EL criterion, a tranche with a lower rating always has a lower expected value, which is used for pricing in (14).

Using the subjective unit price p, the price of a defaultable security (L, M) is given by $Mp(L/M)$. Similarly to (8), for a defaultable security (L, M), a tranching scheme $\{K_1, ..., K_m\}$, and a unit price p, the total deal value is given by

$$\sum_{j=1}^{m} p(L_j)(K_{j-1} - K_j), \tag{19}$$

where $L_j = (L - K_j)_+ \wedge (K_{j-1} - K_j)/(K_{j-1} - K_j)$ is the unit loss of the j-th tranche. The difference between (19) and (8) is that the unit price $p(L_j)$ of a tranche is allowed to depend on the actual loss random variable L_j instead of its rating under I. The deal value in (19) depends on the rating criterion only implicitly through its restrictions on p.

5.2. An Axiom of Self-Consistent Rating Criteria

In the context of real-world markets, various tranching schemes can lead to different valuations of an asset pool. Nonetheless, in the spirit of the celebrated theorem of Modigliani and Miller (1958) (the MM theorem), in an idealized frictionless market, the act of dividing a financial security into tranches should not inherently result in financial gain. This principle seems to be pertinent in the realm of structured finance, where Special Purpose Vehicles (SPV) are common. An SPV, which holds a pool of assets separate from the originating firm's balance sheet, characterized by the absence of taxes, transparent capital structures, and standardized channels of financing, can be viewed as a setting of an ideal market for studying the tranche effects.

We say that a subjective price $p : \mathcal{L}_1 \to \mathbb{R}_+$ is *MM-balanced* if it satisfies the equation

$$p\left(\frac{(L-K)_+}{M-K}\right)(M-K) + p\left(\frac{L \wedge K}{K}\right)K = p\left(\frac{L}{M}\right)M$$

for all $(L, M) \in \mathcal{X}$ and $K \in [0, M]$. For an MM-balanced subjective price scheme, the value of a defaultable security should not depend on how it is structured into two tranches.

For an MM-balanced subjective price, we can easily check by mathematical induction that the value of any tranching scheme in (19) is a constant and is equal to $p(L/M)M$. Thus, MM-balanced subjective prices are invariant in structural maximization, as there is no difference in the total deal value between different tranching schemes, making them a nice building block for the theoretical consideration of suitable rating criteria.

An MM-balanced subjective price cannot be rating-implied (which are discrete); nevertheless, they can be weakly rating-implied and take continuous values.[23] This is the main reason why we introduce the definition of weakly rating-implied price. With the above two definitions, we propose the following axiom:

[SC] *Self-consistency*: A rating criterion I is said to be self-consistent if there exists one weakly I-implied MM-balanced subjective price.

Note that the axiom [SC] is quite mild, as it only requires the existence of just one weakly rating-implied MM-balanced subjective price. In a real-world market, the MM theorem typically does not hold due to market imperfections; however, the MM theorem calls for the study of how these imperfections affect corporate finance decisions. Similarly, many imperfections can lead to different total market values of different tranche splits of the same asset pool, including agency costs, information asymmetry, and unintended consequences of regulation. If a rating criterion does not satisfy [SC], then there is an intrinsic contradiction between an MM-balanced pricing scheme and any weakly rating-implied subjective price; in fact, any investor who uses an MM-balanced pricing scheme would find the rating criterion incompatible with her prices. Similarly to the MM theorem, this suggests that there is a strong connection between the rating criterion that does not satisfy [SC] and market imperfections, which calls for further investigation of the link between such rating criterion and imperfections. This view of imposing axioms in rational decisions (in the sense that rationality is desirable, and, if it fails, a further study of why it fails is of interest) is also consistent with the majority of literature on the axiomatization of decision criteria (e.g., Yaari 1987 and Maccheroni et al. 2006).

THEOREM 4. *Suppose $n \geqslant 3$. We have: (i) The EL criterion is self-consistent. (ii) The intersection of the set of self-consistent rating criteria and that of generalized PD criteria is empty.*

Note that the PD criterion is self-consistent if $n = 2$.[24] Since the case of two rating categories is unrealistic in practice, we shall generally say that the generalized PD criteria are not self-consistent.

[23] As a simple example, for the EL criterion, a linear pricing $p(L/M) = \mathbb{E}^Q[M - L]/M$ is weakly rating-implied and MM-balanced, where Q is any probability measure.

[24] See Example EC.1 in the online supplement for this statement.

6. An Axiomatic Approach to Credit Rating Criteria

Theorems 1-3 yield two interesting questions: First, in addition to the EL criterion, are there any other rating criteria that do not generate Implications (i) and (ii)? Second, if so, can we have an analytical representation for such rating criteria? In this section, we attempt to provide an axiomatic framework to answer the two questions. More precisely, We shall present two main theoretical results. The first theorem yields that a rating criterion satisfying the self-consistency admits a representation via Choquet integrals; we have shown in the previous section that EL satisfies the self-consistency, but PD does not. The second theorem shows that a self-consistent rating criterion (from a rating agency) rules out Implications (i) and (ii) in structural maximization (from an issuer) for rating-implied prices (from an investor).

6.1. A Representation Theorem for Self-Consistent Rating Criteria

THEOREM 5. *A rating criterion I satisfies [SC] if and only if it is generated by*

$$\rho(X) = \int_0^1 g(X > x)\mathrm{d}x, \ X \in \mathcal{L}_1,$$

for some increasing set function[25] $g : \mathcal{F} \to \mathbb{R}$ *with* $g(\varnothing) = 0$.

The integral $\int_0^\infty g(X > x)\mathrm{d}x$ for an increasing set function g is called a Choquet integral, and it is a popular tool in decision theory (e.g., Schmeidler 1989). The main technical difficulty here is that we do not have comonotonic additivity which is commonly used to characterize Choquet integrals in the literature (Schmeidler 1986, Yaari 1987). Instead, we first show in Section F.8 of the online supplement that a rating criterion I satisfies [SC] if and only if it is generated by a rating measure ρ satisfying some properties called [B1]-[B3], and then we rely on [B3], which is a weaker property than comonotonic additivity, together with some continuity arguments to get the representation.[26]

The representation in Theorem 5 suggests that there are many rating criteria, besides the EL criteria, that are self-consistent. We shall give examples in Section 8, such as the rating criteria based on the Value-at-Risk or the Expected Shortfall.

6.2. A Theorem Connecting Self-Consistency to Implications (i) and (ii)

We have seen in Section 3 that Implications (i) and (ii) are associated with the generalized PD criteria for any rating-implied price. Proposition 4 further suggests that a generalized PD criterion is generally not self-consistent whereas an EL criterion is. These two facts are not a coincidence. Indeed, we shall see in the next theorem that the axiom [SC] rules out Implications (i) and (ii) altogether, regardless of the choice of the rating-implied price.

[25] A set function $g : \mathcal{F} \to \mathbb{R}$ is increasing if $g(A) \leqslant g(B)$ for all $A \subseteq B$; note that the term "increasing" is in the non-strict sense in this paper.

[26] If we set $g(\cdot) = 0$ in Theorem 5, then we obtain a trivial rating criterion which only has one non-empty category $I_1 = \mathcal{X}$. We are mainly interested in rating criteria with more than two rating categories.

Theorem 6. *Assume $n \geqslant 3$. A self-consistent rating criterion leads to neither Implication (i) nor (ii) in Section 3.5 for any rating-implied price.*

The proof of Theorem 6 relies on the characterization of self-consistent rating criterion established in Theorem 5. To complete the discussion, we remark that a non-self-consistent rating criterion may or may not lead to Implication (i) and (ii). See Example EC.2 in the online supplement for a non-self-consistent rating criterion that does not lead to Implication (i) and (ii).

7. Extensions to Law-Invariant and Scenario-Relevant Rating Criteria and Their Related Representations

The representation result in Theorem 5 is given in terms of set functions, which are quite abstract objects. It is useful to link them to distribution functions for practical usage. To this end, we shall consider law-invariant and scenario-relevant rating criteria. Law invariance means that the risk assessment is based on the overall distribution of the losses, while scenario relevance is based on conditional distributions under some scenarios (e.g. stress scenarios). As we shall see, law invariance is a special case of scenario relevance. Scenario-relevant credit rating criteria are preferred to law-invariant rating criteria after the 2008 financial crisis, because they can incorporate systematic risk better.[27] In this section we introduce the axioms of law invariance and scenario relevance and study the related representations for rating criteria.

Let $S = (S_1, \ldots, S_s) \in \mathcal{F}^s$ be a *collection of economic scenarios* of rating relevance, where $S_1, \ldots, S_s$ are disjoint events with non-zero probability. For instance, S may represent economic regimes, different levels of strength of the current economic growth, or indices of financial stress.[28] Usually, the collection of scenarios satisfies $\bigcup_{j=1}^{s} S_j = \Omega$, i.e. the scenarios cover the sample space; however we do not need to make this assumption in the following Theorem 7.

For the collection of scenarios $S = (S_1, \ldots, S_s)$, we say that two losses $L_1, L_2 \in \mathcal{L}$ are equivalent under S, denoted by $L_1 \overset{S}{\sim} L_2$, if L_1 and L_2 are identically distributed conditional on S_j for each $j = 1, \ldots, s$. This leads to the following axiom of scenario relevance for a rating criterion I.

[SR] *Scenario relevance* (with respect to S): $I(L_1, M) = I(L_2, M)$ for all $(L_1, M), (L_2, M) \in \mathcal{X}$ satisfying $L_1 \overset{S}{\sim} L_2$.

To motivate [SR] from a slightly different angle, note that, if the losses from two securities have identical distributions under all economic scenarios, then they are indistinguishable statistically, and should be given the same rating.[29]

[27] Structured finance securities are described by Coval et al. (2009) as "economic catastrophe bonds" which tend to default in the worst economic states. Hilscher and Wilson (2017) concluded that a meaningful measure of credit risk should capture both raw loss probability and systematic risk.

[28] Here, s is a finite number. One can also think about the case of infinitely many scenarios. The results in that case are similar.

[29] See Wang and Ziegel (2021) for more discussion on scenario-based risk measures.

The axiom of scenario relevance is able to address correlated default risk. In particular, our framework can give different prices and ratings to losses that have different correlation structures, even if their loss distributions are the same. Consider two identically distributed losses L_1 and L_2, where L_1 is positively correlated with a systemic loss event S_1 and the L_2 is independent of S_1. Although they have the same distribution, L_1 incurs much more loss than L_2 conditionally on S_1, and therefore it may be reasonable to assign a lower rating to L_1. This can be achieved by a rating criterion satisfying scenario relevance with respect to (S_1, S_2), where S_2 is the complement of S_1.

In the special case of only one scenario, i.e., $S = \Omega$, [SR] becomes [LI] in Section 4. In general, if $\bigcup_{j=1}^{s} S_j = \Omega$, then [LI] implies [SR].[30] Thus, the requirement of [SR] is weaker than [LI] generally, allowing for more flexible choices of rating criteria. As a corollary, the EL and the PD criteria satisfy [SR] for any collections S with $\bigcup_{j=1}^{s} S_j = \Omega$.

As illustrated in Example 3 and 4, both S&P and Moody's have recently applied scenario-relevant criteria to rate defaultable securities such as CDOs/CLOs.

Next, we give analytical representations for a rating criterion that is self-consistent and scenario-relevant (or law-invariant).

Theorem 7. *Fix a collection of scenarios S. A rating criterion I satisfies [SC] and [SR] if and only if it is generated by*

$$\rho(X) = \int_0^1 h(\mathbb{P}(X > x|S_1), \ldots, \mathbb{P}(X > x|S_s))\mathrm{d}x, \quad X \in \mathcal{L}_1, \tag{20}$$

for some component-wise increasing function $h : [0,1]^s \to \mathbb{R}$ with $h(\mathbf{0}) = 0$.

To prove the representation (20) in Theorem 7, there are some extra technical challenges compared to Theorem 5, as explained in Section F.11 in the online supplement. A direct corollary of Theorem 7 is that if $S = \Omega$, i.e., with only one scenario, then ρ becomes a law-invariant Choquet integral, also known as a distortion risk measure if $\rho(1) = 1$.

Corollary 1. *A rating criterion I satisfies [SC] and [LI] if and only if it is generated by*

$$\rho(X) = \int_0^1 h(\mathbb{P}(X > x))\mathrm{d}x, \; X \in \mathcal{L}_1$$

for some increasing function $h : [0,1] \to \mathbb{R}$ with $h(0) = 0$.

[30] To see this, assume [LI]. For two defaultable securities (L_1, M) and (L_2, M), if $L_1 \overset{S}{\sim} L_2$, then $L_1 \overset{d}{=} L_2$. Hence, $I(L_1, M) = I(L_2, M)$. Thus I satisfies [SR].

8. Examples of Self-Consistent Rating Criteria Beyond EL

We now discuss some examples of new self-consistent and scenario-relevant rating criteria, in addition to the EL criteria. Fix the collection S of scenarios. A natural choice is the rating measure ρ of the form

$$\rho(X) = f(\rho_1(X), \dots, \rho_s(X)), \qquad X \in \mathcal{L}_1, \tag{21}$$

where $f: \mathbb{R}^s \to \mathbb{R}$ is an aggregation function, and for each $j = 1, \dots, s$, $\rho_j(X)$ is evaluated from the conditional distribution of X under S_j. In other words, ρ in (21) aggregates the information of $\rho_1(X), \dots, \rho_s(X)$ via the function f.[31] As characterized in Theorem 7, we consider a few particularly useful and simple special cases of (20), and most of them also belong to (21).

First, we consider the additive form of (21),

$$\rho(X) = \sum_{j=1}^{s} a_j \rho_j(X) = \sum_{j=1}^{s} a_j \int_0^1 h_j(\mathbb{P}(X > x | S_j)) \mathrm{d}x, \tag{22}$$

where $a_j \geqslant 0$ and h_j is an increasing function on $[0,1]$, $h_j(0) = 0$. Without loss of generality we can normalize ρ by requiring $\sum_{j=1}^{s} a_j h_j(1) = 1$, so that ρ is a weighted average with $\rho(1) = 1$. Examples 9–12 below are of this form. Second, we consider a maximum form as a special case of both (20) and (21), which we present in Example 13. Finally, we present Example 14, which is a special case of (20) but does not belong to (21).

Example 9 (Average EL). The simplest choice is to take a linear function h, that is, $h(u_1, \dots, u_s) = \sum_{j=1}^{s} a_j u_j$, $(u_1, \dots, u_s) \in [0,1]^s$. By simple calculation,

$$\rho(X) = \sum_{j=1}^{s} a_j \mathbb{E}[X | S_j]. \tag{23}$$

Note that $s = 3$ in (23) leads to the Moody's rating method for CDOs/CLOs in (4). If $a_j = \mathbb{P}(S_j)$, $j = 1, \dots, s$, then $\rho(X) = \sum_{j=1}^{s} \mathbb{P}(S_j) \mathbb{E}[X|S_j] = \mathbb{E}[X]$, which is the EL criterion in Example 2. In the simple case $s = 1$, we can assume that S_1 represents an adverse economic scenario, for instance, $S_1 = \{F > \mathrm{VaR}_p(F)\}$ indicating big losses in the system, where F is a systemic factor and $\mathrm{VaR}_p(F)$ is the Value-at-Risk (quantile) of F at level $p \in (0,1)$, given by $\mathrm{VaR}_p(F) = \inf\{x \in \mathbb{R} : \mathbb{P}(F \leqslant x) \geqslant p\}$. Then, one arrives at the so-called risk measure *Marginal Expected Shortfall* (see e.g. Acharya et al. 2012, 2017 and Brownlees and Engle 2016): $\rho(X) = \mathbb{E}[X | F > \mathrm{VaR}_p(F)]$.

Example 10 (Average VaR). The second example is to choose h as the sum of indicator functions. For $p \in (0,1)$, let $h(u_1, \dots, u_s) = \sum_{j=1}^{s} a_j \mathbb{1}_{\{u_j > 1-p\}}$, $(u_1, \dots, u_s) \in [0,1]^s$. With the above h, ρ in (20) has the form $\rho(X) = \sum_{j=1}^{s} a_j \mathrm{VaR}_p(X|S_j)$, where $\mathrm{VaR}_p(X|S_j)$ is the conditional p-quantile function of X given S_j.

[31] Functionals of the form (21) are studied in the context of risk measures; see Kou and Peng (2016) for axioms on the aggregation function f.

Example 11 (Average ES). For $p \in (0,1)$, let $h(u_1,\dots,u_s) = \frac{1}{1-p}\sum_{j=1}^s a_j(u_j \wedge (1-p))$, $(u_1,\dots,u_s) \in [0,1]^s$. With the above h, ρ in (20) has the form $\rho(X) = \sum_{j=1}^s a_j \mathrm{ES}_p(X|S_j)$, where, for $j = 1,\dots,s$, $\mathrm{ES}_p(X|S_j)$ is the Expected Shortfall of X under scenario S_j at level p, namely, $\mathrm{ES}_p(X|S_j) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_p(X|S_j)\mathrm{d}t$.

Example 12 (Average MAXVAR). For $\gamma \in (0,1)$, let $h(u_1,\dots,u_s) = \sum_{j=1}^s a_j u_j^\gamma$, $(u_1,\dots,u_s) \in [0,1]^s$. With the above h, ρ in (20) has the form $\rho(X) = \sum_{j=1}^s a_j \mathrm{MAXVAR}_\gamma(X|S_j)$, where, for $j = 1,\dots,s$, the MAXVAR risk measure (Cherny and Madan 2009) is defined as $\mathrm{MAXVAR}_p(X|S_j) = \int_0^1 (\mathbb{P}(X > x|S_j))^\gamma \mathrm{d}x$.

Example 13 (Max of VaR). For $p \in (0,1)$, let $h(u_1,\dots,u_s) = \bigvee_{j=1}^s \mathbb{1}_{\{u_j > 1-p\}}$, $(u_1,\dots,u_s) \in [0,1]^s$. With the above h, ρ in (20) has the form $\rho(X) = \bigvee_{j=1}^s \mathrm{VaR}_p(X|S_j)$.

Example 14 (Max of ES). For $p \in (0,1)$, let $h(u_1,\dots,u_s) = \frac{1}{1-p}\int_p^1 \left(\bigvee_{j=1}^s \mathbb{1}_{\{u_j > 1-p\}}\right)\mathrm{d}q$, $(u_1,\dots,u_s) \in [0,1]^s$. With the above h, ρ in (20) has the form

$$\rho(X) = \frac{1}{1-p}\int_p^1 \left(\bigvee_{j=1}^s \mathrm{VaR}_q(X|S_j)\right)\mathrm{d}q.$$

We note that ρ does not belong to (21) since it relies on the class $\mathrm{VaR}_q(X|S_j)$, $j = 1,\dots,s$, $q \in (p,1)$ instead of s individual risk measures.

9. Concluding Remarks

Now we are ready to answer the two questions raised at the beginning of the introduction section. (I) To answer the first question, we show theoretically that if a rating criterion does not satisfy self-consistency, then it encourages issuers to take advantage of generalized PD criteria through structural maximization. The gain from structural maximization is not consistent with the MM theorem. We also propose a model of information asymmetry, and find that PD gives more advantage to the issuer than EL when the issuer has private information and the investor tries to price securities using rating results. (II) To answer the second question, we propose a framework for rating financial securities based on an axiom of self-consistency. The axiom does not allow issuers to gain, by tranching financial securities, from investors who rely on the rating criterion. We show that the expected loss criterion used by Moody's does satisfy the self-consistency, whereas the default probability criterion used by S&P does not. A self-consistent rating measure admits a Choquet integral representation, and this representation becomes tractable if relevant economic scenarios are specified. We also suggest new examples of scenario-based self-consistent rating criteria, such as ones based on the VaR and the expected shortfall, which can be useful as alternatives to the scenario-based EL criterion.

Although our results favor the use of EL over PD theoretically, PD and EL co-exist in the real-world market. One possible reason for the co-existence may be the need for diversity in financial services. For example, two business factors may contribute to the existence of multiple rating criteria: First, as the two oldest rating agencies, Moody's and S&P have historically sought to differentiate themselves for competitive reasons. Using different rating criteria allowed them to stand out in the nascent days of their business. Second, the PD and EL criteria were initially developed for rating corporate bonds. As the structured finance market emerged, it made sense for the agencies to extend the existing rating criteria for corporate bonds to new asset classes to ensure uniform comparisons across different security classes.

We should point out that a significant drawback of credit rating is that it has discrete categories. In addition to causing significant technical issues in our axiomatic analysis, it would be better to provide other information besides the discrete categories. Indeed, as commented by Hilscher and Wilson (2017), "given the multidimensional nature of credit risk, it is not possible for one measure to capture all the relevant information" (p. 3414). However, currently there are no viable alternatives to the standard letter-based credit ratings in practice. Complementing ratings of credit quality, rating agencies also provide (discrete) information/ratings on the expected recovery rate for structured finance products, which are relevant for investors. A joint framework of credit rating and expected recovery rate may be useful to further study how information can be efficiently passed to the investor.

Our framework captures how investors rely on credit ratings to price securities under various rating criteria. The results in Section 4 show that compared to EL, PD is less effective in passing information to the investor. We do not distinguish the accuracy of the rating results from different rating criteria; instead, we focus on the usefulness of the observed ratings by the investors. Future research is needed to examine the extent to which investors update their beliefs about the quality of the rating (see Bolton et al. (2012), Baghai et al. (2014), and Chernenko et al. (2016)), and how these beliefs are affected by different rating criteria.

Our framework assumes that an issuer can arbitrarily choose a tranching scheme and receive ratings from the loss model of each tranche. In practice, there are other factors that may affect credit rating. Re-tranching will result in a different coupon stack for a structured finance product, and S&P's evaluation of structured finance products typically involves the modeling of cash flows. Including such features in the paper makes the model more complicated to analyze with current techniques, and they may be explored in the future.

We mention a few more possible extensions of our framework. First, the optimal tranching scheme for the PD criterion obtained in Theorem 2 has tranche levels at the edge of downgrading. Taking a dynamic view, a rating agency may be hesitant to rate a tranche in a category that will be downgraded with small fluctuation, because rating stability is also important rating agencies. Second, to analyze structural maximization, we take the perspective of the issuer and do not consider the reputation risk of the rating agency. Third, our analysis focuses on the use of rating on structured finance securities sold as tranches. In practice, ratings are used in many other domains, including rating stand-alone debt of companies and rating sovereign debt, and the pools of assets may be sold as piecemeal rather than as tranches. Our study of rating criteria does not apply to these cases.Finally, a potential limitation of the model-based investors in Section 4 is that, similar to the existing literature, the model does not allow for individual preferences of the investors other than risk neutrality. Investigating the rating stability, the reputation risk of rating agencies, other rating products, and the impact of investors' preferences on the information gap may be promising directions for study. Besides structural maximization, there could be other directions worth exploring. For example, Rauh and Sufi (2010) find empirically that, while higher rated firms rely almost exclusively on two tiers of capital (senior unsecured debt and equity), lower rated firms tend to use multiple tiers of debt; however, they cannot explain this with existing theoretical models. Given the similarity between the fact they describe and the structural maximization, it is an interesting open problem to see whether different rating criteria can give an explanation to this empirical puzzle.

References

Acharya, V. V., Engle, R., and Richardson, M. (2012), Capital shortfall: A new approach to ranking and regulating systemic risks, *American Economic Review*, **102**(3), 59–64.

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017), Measuring systemic risk, *Review of Financial Studies*, **30**(1), 2–47.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999), Coherent measures of risk, *Mathematical Finance*, **9**(3), 203–228.

Ashcraft, A., Goldsmith-Pinkham, P., Hull, P., and Vickery, J. (2011), Credit ratings and security prices in the subprime MBS market, *American Economic Review*, **101**(3), 115–119.

Azizpour, S., Giesecke, K., and Schwenkler, G. (2018), Exploring the sources of default clustering, *Journal of Financial Economics*, **129**(1), 154–183.

Baghai, R.P., Servaes, H. and Tamayo, A. (2014), Have rating agencies become more conservative? Implications for capital structure and debt pricing, *Journal of Finance*, **69**(5), 1961–2005.

Bai, J., Collin-Dufresne, P., Goldstein, R. S., and Helwege, J. (2015), On bounding credit-event risk premia, *Review of Financial Studies*, **28**(9), 2608–2642.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**(4), 1165–1188.

Brennan, M. J., Hein, J., and Poon, S.-H. (2009), Tranching and rating, *European Financial Management*, **15**(5), 891–922.

Bolton, P., Freixas, X., and Shapiro, J. (2012), The credit ratings game, *Journal of Finance*, **67**(1), 85–111.

Bongaerts, D., De Jong, F., and Driessen, J. (2011), Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market, *Journal of Finance*, **66**(1), 203–240.

Brownlees, C. and Engle, R. F. (2016), SRISK: A conditional capital shortfall measure of systemic risk, *Review of Financial Studies*, **30**(1), 48–79.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2011), Risk measures: Rationality and diversification, *Mathematical Finance*, **21**(4), 743–774.

Chen, C., Iyengar, G., and Moallemi, C. C. (2013), An axiomatic approach to systemic risk, *Management Science*, **59**(6), 1373–1388.

Chen, N., Glasserman, P., Nouri, B., and Pelger M. (2017), *Review of Financial Studies*, **30**(11), 3921-3969.

Chernenko, S., Hanson, S. G., and Sunderam, A. (2016), Who neglects risk? Investor experience and the credit boom, *Journal of Financial Economics*, **122**(2), 248–269.

Cherny, A. and Madan, D. (2009), New measures for performance evaluation, *Review of Financial Studies*, **22**(7), 2571–2606.

Chu, C. S., and Rysman, M. (2019), Competition and Strategic Incentives in the Market for Credit Ratings: Empirics of the Financial Crisis of 2007, *American Economic Review*, **109**(10), 3514–55.

Cornaggia, J. and Cornaggia, K. J. (2013), Estimating the costs of issuer-paid credit ratings, *Review of Financial Studies*, **26**(9), 2229–2269.

Cornaggia, J. N., Cornaggia, K. J., and Hund, J. E. (2017), Credit ratings across asset classes: A long-term perspective, *Review of Finance*, **21**(2), 465–509.

Coval, J. D., Jurek, J. W., and Stafford, E. (2009), Economic catastrophe bonds, *American Economic Review*, **99**(3), 628–666.

Cuchra, F.-M. (2005), Explaining launch spreads on structured bonds, working paper, University of Oxford.

DeMarzo, P. M. (2005), The pooling and tranching of securities: A model of informed intermediation, *Review of Financial Studies*, **18**(1), 1–35.

DeMarzo, P., and Duffie, D. (1999), A liquidity-based model of security design, *Econometrica*, **67**(1), 65–99.

Dimitrov, V., Palia, D., and Tang, L. (2015), Impact of the Dodd-Frank Act on credit ratings, *Journal of Financial Economics*, **115**(3), 505–520.

Downing, C., Jaffee, D., and Wallace, N. (2009), Is the market for mortgage-backed securities a market for lemons? *Review of Financial Studies*, **22**(7), 2457–2494.

Föllmer, H. and Schied, A. (2016), *Stochastic Finance. An Introduction in Discrete Time.* Walter de Gruyter, Berlin, Fourth Edition.

Frenkel, S. (2015), Repeated interaction and rating inflation: A model of double reputation, *American Economic Journal: Microeconomics*, **7**(1), 250-80.

Giesecke, K., Longstaff, F. A., Schaefer, S., and Strebulaev, I. A. (2014), Macroeconomic effects of corporate default crisis: A long-term perspective, *Journal of Financial Economics*, **111**(2), 297–310.

Griffin, J. M. and Tang, D. Y. (2012), Did subjectivity play a role in CDO credit ratings? *Journal of Finance*, **67**(4), 1293–1328.

Griffin, J. M., Nickerson, J., and Tang, D. Y. (2013), Rating shopping or catering? An examination of the response to competitive pressure for CDO credit ratings, *Review of Financial Studies*, **26**(9), 2270–2310.

Hilscher, J. and Wilson, M. (2017), Credit ratings and credit risk: Is one measure enough? *Management Science*, **63**(10), 3414–3437.

Hull, J. and White, A. (2012), Ratings arbitrage and structured products, *Journal of Derivatives*, **20**(1), 80–86.

Keppo, J., Kofman, L., and Meng X. (2010), Unintended consequences of the market risk requirement in banking regulation, *Journal of Economic Dynamics and Control*, **34**(10), 2192–2214.

Kou, S. and Peng, X. (2016), On the measurement of economic tail risk, *Operations Research*, **64**(5), 1056–1072.

Kyle, A. S. (1985), Continuous auctions and insider trading. *Econometrica*, **53**(6), 1315–1335.

Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, **37**(5), 1137–1153.

Maccheroni, F., Marinacci, M. and Rustichini, A. (2006), Ambiguity aversion, robustness, and the variational representation of preferences, *Econometrica*, **74**(6), 1447–1498.

Malamud, S., Rui, H. and Whinston, A. (2013), Optimal incentives and securitization of defaultable assets, *Journal of Financial Economics*, **107**(1), 111–135.

McNeil, A. J., Frey, R. and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.

Modigliani, F. and Miller, M. H. (1958), The cost of capital, corporation finance and the theory of investment, *American Economic Review*, **48**(3), 261–297.

Monfort, A., Pegoraro, F., Renne, J. P., and Roussellet, G. (2021), Affine modeling of credit risk, pricing of credit events, and contagion, *Management Science*, **67**(6), 3674–3693.

Moody's (2017), Approach to rating corporate synthetic collateralized debt obligations. Available on `https://www.moodys.com/research` (accessed December 2020).

Nickerson, J., and Griffin, J. M. (2017), Debt correlations in the wake of the financial crisis: What are appropriate default correlations for structured products? *Journal of Financial Economics*, **125**(3), 454–474.

Opp, C. C, Opp, M. M, and Harris, M. (2013), Rating agencies in the face of regulation, *Journal of Financial Economics*, **108**, 46–61.

Partnoy, F. (2010), Overdependence on credit ratings was a primary cause of the crisis, *The First Credit Market Turmoil of the 21st Century*, Edward Elgar Press.

Rauh, J. D., and Sufi, A. (2010), Capital structure and debt structure, *Review of Financial Studies*, **23**(12), 4242–4280.

Sangiorgi, F., and Spatt, C. (2017), Opacity, credit rating shopping, and bias, *Management Science*, **63**(12), 4016–4036.

Schmeidler, D. (1986), Integral representation without additivity, *Proceedings of the American Mathematical Society*, **97**(2), 255–261.

Schmeidler, D. (1989), Subject probability and expected utility without additivity, *Econometrica*, **57**(3), 571–587.

SEC (U.S. Securities and Exchange Commission) (2014), Final rule: Nationally Recognized Statistical Rating Organizations. Washington, DC.

Skreta, V., and Veldkamp, L. (2009), Ratings shopping and asset complexity: A theory of ratings inflation, *Journal of Monetary Economics*, **56**(5), 678-695.

Standard and Poor's (2016), Global methodologies and assumptions for corporate cash flow and synthetic CDOs. Available on `https://www.standardandpoors.com` (accessed December 2020).

Toda, A. A. (2015), Securitization and leverage in general equilibrium, Available at SSRN 2336582.

Wang, R., and Ziegel, J. F. (2021), Scenario-based risk evaluation, *Finance and Stochastics*, **25**, 725–756.

Wang, R. and Zitikis, R. (2021), An axiomatic foundation for the Expected Shortfall, *Management Science*, **67**, 1413–1429.

Wojtowicz, M. (2014), CDOs and the financial crisis: Credit ratings and fair premia, *Journal of Banking and Finance*, **39**, 1–13.

Yaari, M. E. (1987), The dual theory of choice under risk, *Econometrica*, **55**(1), 95–115.

Electronic Companion

A Theory of Credit Rating Criteria

A. Preliminary Empirical Evidence

In this appendix, we provide some preliminary empirical evidence to illustrate our results on structural maximization in Sections 3 (both Implications (i) and (ii)) and 4 (only Implication (i)).

So far there is no full empirical study of structural maximization. A significant challenge in conducting such a full empirical investigation with counterfactual analysis is that the total market deal values are not observable, as issuers typically keep the lowest (equity) tranches themselves instead of selling them in the market. Furthermore, one may argue that the same rating (such as AA for S&P and Aa for Moody's) may be perceived differently by investors. The fact that most non-AAA tranches were rated by several agencies before the Dodd-Frank Act further complicated the task.

However, even without observing the total market deal values, Theorems 1 and 2 can still lead to two empirical implications, Implications (i) and (ii), for the generalized PD criteria. In contrast, Theorems 4 and 6 together imply that the EL criterion leads to neither Implication (i) nor (ii).

The striking difference in terms of implications gives us an opportunity to conduct an empirical study in the structured finance market for the post-Dodd-Frank period. More precisely: (a) The Dodd-Frank Act reduces the regulatory reliance on credit ratings, which leads to a phenomenon that many, if not most, non-AAA tranches in a structured finance deal are rated solely by one rating agency, thus enabling us to separate the effect of PD and EL rating criteria. (b) The act mandates a study on the feasibility and desirability of credit rating standardizations, reflecting the concern that the same rating category should have the same meaning to investors in terms of subjective pricing implications. Appendix B shows that we do observe that investors do not price differently between S&P and Moody's ratings in the post-Dodd-Frank period[32].

A.1. An Overview of the Data and the CDOs/CLOs Markets

The CDO, whose volume peaks at 2007 and almost vanishes in the post-Dodd-Frank period. However, as a variation of CDO, the Collateralized Loan Obligation (CLO) market quickly recovers after the 2007-2009 crisis, and the issuance volume hits a new high in 2014 and then again in 2018. Indeed Table EC.1 shows that, perhaps contrary to some people's intuition, the CDO/CLO

[32] There are some debates on whether investors differentiate between S&P and Moody's ratings. For example, Billingsley et al. (1985) show that primary issue bonds that receive different (split) ratings from S&P and Moody's are sold with reoffering yields more similar to the lower of the two ratings. Livingston et al. (2010) show that, at issuance, yields on split rated bonds with Moody's ratings are about 8 basis points lower than yields on split rated bonds with superior S&P ratings. Appendix B shows that we do not observe these phenomena in our data.

business in the post-Dodd-Frank period (about $927.4 billions from 2010 to 2019) *is already bigger than that* in the pre-Dodd-Frank period (about $912.4 billions). The key difference in the data is that in the post-Dodd-Frank period, more than 97% of CDOs/CLOs are CLOs, while only 37% of CDOs/CLOs are CLOs in the pre-Dodd-Frank period.

Table EC.1 CDO/CLO tranche-level rating coverage

	Before Dodd-Frank				After Dodd-Frank			
	# tranche	%	Capital ($B)	%	# tranche	%	Capital ($B)	%
AAA tranche	3,611	34.7	723.3	79.3	3,191	25.4	609.1	65.7
Solo S&P	121	1.2	14.6	1.6	234	1.9	25.4	2.7
Solo Moody's	160	1.5	26.2	2.9	372	3.0	46.2	5.0
Solo Fitch	9	0.1	3.0	0.3	135	1.1	9.7	1.1
SP & Moody's	2,496	24.0	532.4	58.4	966	7.7	209.3	22.6
SP & Fitch	138	1.3	15.6	1.7	313	2.5	71.1	7.7
Moody's & Fitch	89	0.9	12.3	1.3	1,165	9.3	245.1	26.4
All three	598	5.7	119.1	13.1	6	0.0	2.3	0.2
Non-AAA tranche	6,802	65.3	189.1	20.7	9,371	74.6	318.3	34.3
Solo S&P	188	1.8	7.0	0.8	3,979	31.7	127.7	13.8
Solo Moody's	742	7.1	20.1	2.2	4,156	33.1	153.7	16.6
Solo Fitch	188	1.8	8.0	0.9	70	0.6	2.4	0.3
SP & Moody's	4,100	39.4	107.2	11.7	401	3.2	11.7	1.3
SP & Fitch	377	3.6	10.7	1.2	55	0.4	2.0	0.2
Moody's & Fitch	321	3.1	9.9	1.1	710	5.7	20.7	2.2
All three	886	8.5	26.2	2.9	0	0.0	0.0	0.0
Total	10,413	100.0	912.4	100.0	12,562	100.0	927.4	100.0

Notes. This table reports the rating coverage of S&P, Moody's and Fitch for CDOs/CLOs denominated in US dollars from January 1997 to December 2019 listed in the Bloomberg Database. Dodd-Frank is enacted on July 21, 2010. Following the convention of Griffin et al. (2013), we name AAA tranches as the ones received an AAA rating by at least one rating agency.

We extract the information of all Collateralized Debt Obligations/Collateralized Loan Obligations (CDOs/CLOs) listed on Bloomberg denominated in US dollars from January 1997 to December 2019. More precisely, we collect the data related to (1) cash CDO/CLO; (2) synthetic CDO/CLO; (3) hybrid CDO/CLO; (4) market value CDO/CLO.

Table EC.1 presents the rating coverage by either Moody's, S&P or Fitch on the tranche level. The pre-Dodd-Frank period includes 10,413 tranches with total value of $912.4 billion. The largest part of the deals (63.4% by numbers, and 70.1% by capital) is dual rated by S&P and Moody's. The second largest part of the tranches (14.2% by numbers, and 16.0% by capital) is triple rated by S&P, Moody's and Fitch. The pattern is similar if we classify tranches into AAA rated and non-AAA rated. (Following the convention of Griffin et al. 2013, we name AAA tranches as the ones received an AAA rating by at least one rating agency.)

In the post-Dodd-Frank period the rating mode shifts dramatically. First, among 12,562 tranches with a total value of $927.4 billion, there is a dramatic shrink of the portion of triple-rated tranches (0.0% by numbers, and 0.2% by capital). Second, more interestingly, while a substantial portion of AAA-rated tranches are dual-rated, most of the non-AAA tranches are solo-rated. At the same time, the portion of non-AAA tranches grows bigger. While the capital amount of AAA tranches is 4 times of that of non-AAA tranches before Dodd-Frank, the ratio becomes less than 2 after Dodd-Frank. Finally, although Fitch's market share grows fast, it still plays a supplement role in the CDO/CLO market, with the observation that most of the tranches rated by Fitch are in conjunction with S&P or Moody's.[33]

A.2. Implication (i)

To study Implication (i), we focus on deals rated by both S&P (using a generalized PD criterion) and Moody's (using the EL criterion), the biggest two players in the market. To meet the requirement (9) in Theorem 1, we select the deals with AAA tranches rated by both of them and agreed on AAA rating. We then classify deals into three mutually exclusive groups: the deals with non-AAA tranches solo rated by S&P (called *S&P non-AAA*), solo rated by Moody's (called *Moody's non-AAA*), and rated by both of them (called *Both non-AAA*).

Table EC.2 shows a clear difference before and after Dodd-Frank. While issuers choose both agencies to rate non-AAA tranches before 2010, issuers tend to choose one rating agency after that. The surprising thing is that the number of deals with non-AAA tranches rated by S&P is persistently much higher than that by Moody's in every single year after 2010, given the fact that Moody's takes the dominant overall market share shown in Table EC.1. This observation is consistent with Theorem 1, that for deal value maximizers, PD (associated with S&P) strictly dominates EL (associated with Moody's) if there are at least three possible rating categories of (L, M). The observation is also consistent with Theorem 3, that PD gives more advantages to the issuer for model-based investors.

As results in Section 4 show that PD is less effective in disclosing information, the observation that more PD is used after Dodd-Frank is consistent with that of Dimitrov et al. (2015) on credit rating being less informative after Dodd-Frank.

A.3. Implication (ii)

To investigate Implication (ii), Panel A of Table EC.3 presents the number of deals grouped by the number of differently rated[34] tranches using the same sample as that of Table EC.2. We see that, during the post-Dodd-Frank period, the number of tranches for the deals with non-AAA tranches solo rated by S&P are generally larger than those with non-AAA tranches solo rated

[33] Using a sample from 2000 to 2008, Bongaerts et al. (2012) also document that almost all large, liquid U.S. corporate bond issues are rated by both S&P and Moody's. Fitch typically plays the role of a "third opinion" for large bond issues (p. 114).

[34] As explained in Appendix B, we use the letter rating category (i.e., AAA, AA, A, BBB, BB and B).

Table EC.2 The number of deals grouped by rating coverage for non-AAA tranches

	Year	S&P non-AAA	Moody's non-AAA	Both non-AAA
Before Dodd-Frank	1997	0	0	3
	1998	6	1	3
	1999	13	0	8
	2000	2	10	19
	2001	0	14	42
	2002	1	2	89
	2003	0	5	78
	2004	4	7	117
	2005	0	2	212
	2006	0	9	471
	2007	4	13	407
	2008	0	1	43
	2009	0	1	5
	2010	0	0	6
After Dodd-Frank	2011	20	0	7
	2012	74	2	16
	2013	117	13	13
	2014	79	20	9
	2015	30	17	12
	2016	29	25	4
	2017	37	18	41
	2018	37	6	23
	2019	39	1	17

Notes. This table reports the ratings coverage of deals with AAA tranches rated by both S&P and Moody's and non-AAA tranches rated by either S&P or Moody's. The table shows that during the post Dodd-Frank period, S&P tends to rate non-AAA tranches much more than Moody's does. In column headings, "S&P (Moody's) non-AAA" means the deal with the non-AAA tranches rated only by S&P (Moody's); "Both non-AAA" means the deal with the non-AAA tranches rated both by S&P and Moody's.

by Moody's. More precisely, the mean tranche number is 5.272 for *S&P non-AAA* and 5.039 for *Moody's non-AAA*.

To check if the difference of 0.233 is statistically significant, we conduct a regression in Panel B of Table EC.3. To control the influence of other factors, we include fixed year, collateral or size effect. We measure the deal size with the logarithm of the total tranche face value and the non-AAA tranche face value. The explanatory variables are different rating modes as dummy variables. Panel B of Table EC.3 shows a significantly lower number of tranches for the deals when Moody's solo rates the non-AAA tranches than those when S&P solo rates non-AAA tranches, after Dodd-Frank enacted.

In summary, the empirical evidence seems to suggest that, in the post-Dodd-Frank period, even if overall Moody's is the dominant player in the rating market for CDOs/CLOs, S&P tends to rate more non-AAA tranches when the AAA tranche in the same deal is rated by both S&P and Moody's, consistent with Implication (i), and S&P tends to create more tranches in a deal, consistent with Implication (ii).

Table EC.3 The number of differently rated tranches

Panel A: Number of deals grouped by the number of tranches

Number of tranches	Before Dodd-Frank S&P non-AAA	Before Dodd-Frank Moody's non-AAA	Before Dodd-Frank Both non-AAA	**After Dodd-Frank** S&P non-AAA	**After Dodd-Frank** Moody's non-AAA	**After Dodd-Frank** Both non-AAA
2	17	19	39	3	2	1
3	6	32	158	6	6	4
4	5	12	477	9	7	3
5	1	2	560	289	58	62
6	1	0	35	156	29	72
Total	30	65	1,269	463	102	142
Mean	2.767	2.954	4.310	5.272	5.039	5.408
Std.	1.073	0.779	0.839	0.627	0.878	0.736

Panel B: Regression on rating modes

	Before Dodd-Frank		**After Dodd-Frank**	
	(1)	(2)	(3)	(4)
S&P non-AAA	-1.041***	-0.903***	0.022	-0.022
	(0.100)	(0.242)	(0.059)	(0.063)
Moody's non-AAA	-1.097***	0.004	-0.180**	-0.179**
	(0.191)	(0.164)	(0.081)	(0.089)
Controls:				
Log total tranche face value	-0.039	-0.094*	-1.940***	-1.830***
	(0.035)	(0.048)	(0.319)	(0.349)
Log non-AAA tranche face value	0.633***	0.603***	1.971***	1.846***
	(0.037)	(0.062)	(0.306)	(0.343)
Collateral fixed effect	y	y	n	n
Year fixed effect	n	y	n	y
No. Obs.	1,364	1,364	707	707
R^2	0.379	0.437	0.311	0.344

Notes. Panel A reports the number of deals grouped by the number of differently rated tranches and summary statistics. Panel B reports the results of OLS regression. The dependent variable is the number of differently rated tranches for each deal. *S&P (Moody's) non-AAA* is a dummy variable that takes on a value of one when the non-AAA tranche of the dual rated deal is rated exclusively by S&P (Moody's). *Collateral (Year) controls* indicates specifications when collateral type (year) fixed effects were used. In the post-Dodd-Frank period, we do not control collateral types since more than 97% of collaterals belong to CLOs. Standard errors (reported in parentheses) are robust (columns (1) and (3)) and clustered by years (columns (2) and (4)). The symbols *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

A.4. Three Reverse Causality Tests

The regression result in Table EC.3 may be subject to some endogeneity issues. For example, the dependent variable (the number of tranches) may have some additional information that is currently not used by investors, The information may make it more suitable for a deal associated with a particular total tranche number to be rated by a specific rating criterion, thus creating a case of reverse causality. Can we still detect the effect of rating criteria on the number of tranches in the presence of reverse causality? This is impossible unless one gets panel data.

Fortunately, we do get panel data for both y and X. There are three popular ways to run a regression using panel data in the presence of the reverse causality, namely the time-lagged regression, the first difference model, and the autoregressive time series regression model; for a survey, see, e.g., Leszczensky and Wolbring (2022). Subsequently, we employ all three methods.

In our regression settings, y_{it} represents the number of differently rated tranches for each issuer i in period t, while X_{it} includes rating modes as the primary explanatory variables and size-related variables as control variables. The time index t tracks the chronological sequence of deals issued. Table EC.4 presents the results of three reverse causality tests on the effect of rating criteria on the number of tranches in the post-Dodd-Frank period.

The first and perhaps most widely utilized approach is the time-lagged regression. This method is based on the premise that $y_{i,t}$ should not exert a causal effect on $X_{i,t-1}$; see Calvet and Sodini (2014), Fagereng et al. (2009). Column (2) presents random effect estimates derived from the lagged model $y_{i,t} = \beta X_{i,t-1} + \alpha_i + \varepsilon_{i,t}$. The second approach involves the application of the first-difference model, as detailed in Allison (2009). The estimates for this model are presented in Column (3), where the model is specified as $\Delta y_{i,t} = \Delta X_{i,t} + \Delta \varepsilon_{i,t}$. The third strategy employs an autoregressive time series regression model, formulated as $y_{i,t} = \beta_1 y_{i,t-1} + \beta_2 X_{i,t} + \alpha_i + \varepsilon_{i,t}$, with further insights provided in Arellano and Bond (1991). The results of employing this third method are shown in Column (4). Additionally, as another robustness check, Column (1) presents estimates from the standard random effect model. This model, contrasting with the fixed effect model in Table EC.3, is specified as $y_{i,t} = \beta X_{i,t} + \alpha_i + \varepsilon_{i,t}$.

For all three different econometrics models, we obtain either a significantly lower number of tranches when Moody's solo rates the non-AAA tranches (the negative coefficients in Models (3) and (4)), or a significantly higher number of tranches when S&P solo rates (the positive coefficient in Model (2)), demonstrating the robustness of our empirical conclusion in presence of the reverse causality.

In summary, all three reverse causality econometric models indicate a significant effect of rating criteria on the number of tranches, thus providing robustness tests that strengthen the evidence for our hypothesized causal relationship.

B. Validation for the Assumption of Identical Rating-Implied Subjective Pricing in the Post-Dodd-Frank Period

Theorem 1 assumes that the investors do not differentiate between the same rating category from different agencies in terms of subjective pricing. Here we validate this in two steps using the rating sample from S&P and Moody's.

First, we show that S&P and Moody's generally agree on tranche ratings in Table EC.5, which reports that the rating results become more equivalent after Dodd-Frank for tranches rated both

Table EC.4 Three reverse causality tests and the random effect test for the impact on the number of differently rated tranches after Dodd-Frank

	(1)	(2)	(3)	(4)
S&P non-AAA	0.053	0.134**	0.103	0.07
	(0.059)	(0.057)	(0.066)	(0.067)
Moody's non-AAA	-0.152**	0.133	-0.202**	-0.187**
	(0.075)	(0.113)	(0.098)	(0.092)
Lagged number of tranches				-0.196***
				(0.046)
Size controls:				
Log total tranche face value	-1.707***	-0.449***	0.313**	0.436***
	(0.160)	(0.133)	(0.153)	(0.148)
Log non-AAA tranche face value	1.810***	0.395***	0.561***	0.300**
	(0.122)	(0.134)	(0.142)	(0.146)
Constant	5.388***	6.803***		-8.122***
	(1.656)	(1.300)		(1.570)
No. Obs.	691	519	519	417
No. Groups	199	123	123	96
R^2	0.311	0.036	0.219	–

Notes. This table reports the estimation outcomes derived from four different panel models. The dependent variable is the number of differently rated tranches for each deal. *S&P (Moody's) non-AAA* is a dummy variable that takes on a value of one when the non-AAA tranche of the dual rated deal is rated exclusively by S&P (Moody's). Column (1) shows estimates using the random effect model. Column (2) presents lagged independent variable random effects estimates. Column (3) contains first-difference (FD) estimates. Column (4) presents the result using the Arellano-Bond model. For brevity, *S&P non-AAA*, *Moody's non-AAA*, and size control variables are denoted as such regardless of lag or first-difference transformations. Standard errors are reported in parentheses. The symbols *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. The R^2 is not reported for column (4) because it is a time series regression.

by S&P and Moody's.[35] Before Dodd-Frank, 3.4% of the CDO/CLO tranches are with different ratings. In contrast, after Dodd-Frank, the portion of disagreement drops to merely 0.1%. One may notice that there should be three subcategories (called notch ratings) in each letter rating category from AA to B, i.e., AA+, AA, AA-, ..., B+, B, B-. By observing that the notch ratings are concentrated in one certain subcategory in each letter rating for either S&P or Moody's, the letter rating category is used for analysis.

Second, to test if investors price differently between S&P and Moody's ratings, we regress tranche-level yield spread on rating agencies in Table EC.6. The sample includes 5,870 floating-rate notes exclusively rated by S&P or Moody's. These tranches are issued at par and have the Libor as base rate. Thus the demanded yield spread is simply equal to the coupon spread. To capture the effect of rating agencies for non-AAA tranches, we use the test variable *Solo Moody's*, which is equal to one when the tranche is rated by Moody's exclusively and zero otherwise. The control variables include the weighted average life of the tranche at issuance as well as its squared value, and log tranche face value as well as its squared value.

[35] Although S&P and Moody's use slightly different rating notation, they are generally thought to be equivalent. For example, S&P's BBB corresponds to Moody's Baa.

Table EC.5 CDO/CLO tranche rating agreement between S&P and Moody's

S&P \ Moody's	Aaa	Aa	A	Baa	Ba	B	Total	% split
Panel A: Before Dodd-Frank								
AAA	2,895	152	4	3	0	0	3,054	6.1
AA	19	1,348	11	0	0	0	1,378	13.1
A	4	19	1,411	9	1	1	1,445	4.8
BBB	2	3	20	1,408	9	4	1,446	3.9
BB	4	0	1	7	705	1	718	3.4
B	0	0	1	0	2	21	24	30.0
Total	2,924	1,522	1,448	1,427	717	27	8,065	3.4
Panel B: After Dodd-Frank								
AAA	971	1	0	0	0	0	972	0.1
AA	0	137	0	0	0	0	137	0.7
A	0	0	73	0	0	0	73	0.0
BBB	0	0	0	65	0	0	65	0.0
BB	0	0	0	0	70	1	71	1.4
B	0	0	0	0	0	55	55	1.8
Total	971	138	73	65	70	56	1373	0.1

Notes. This table compares tranche ratings between S&P and Moody's, the largest two players in the CDO/CLO market. S&P and Moody's use slightly different rating notation but they are generally thought to be equivalent. For example, S&P's BBB corresponds to Moody's Baa. *% split* reports the degree of rating disagreement for each rating category. For example, the number 13.1% is calculated based on the sample of tranches rated AA by S&P or Aa by Moody's.

Columns (1)-(5) of Table EC.6 shows the test results for the sample of tranches rated AA, A, BBB, BB and B, respectively. We find that the coefficient for *Solo Moody's* is not statistically significant for AA, A, and BBB rating categories. The coefficient is positive and significant at the 5% level for BB, and negative and significant at the 10% level for B. However, given that the mean values of yield spread for BB and B rated tranches are 8.56% and 8.86%, the coefficients around 10 basis points are not economically significant. Thus, it seems that the investors do not price differently between S&P and Moody's ratings.

C. Quantifying the Potential Gains

To explore the magnitude of marketing gains in the post-Dodd-Frank period, we shall redo the examples shown in Table 4 and 5 of Brennan et al. (2009) by updating the estimated parameters. According to Kenneth R. French Data Library, from 2010 to 2018, the US equity market risk premium and the risk-free rate has averaged about 12.1% and 0.3%, and the annualized monthly standard deviation of the market factor is 12.9%. We use the target default probabilities and expected losses according to Standard and Poor's (2016) and Moody's (2017).

Table EC.7 reports the potential financial gains from tranched debt issued by a single firm with a total value of 100. The findings of Brennan et al. (2009) (as shown in Column "Gain before Dodd-Frank") become even more significant in the post-Dodd-Frank period.

D. Pooling Effect

This paper focuses on the tranching effect, i.e., splitting the security pool into parts. In this section, we shall give a preliminary study of the effect of asset pooling.

Table EC.6 Spread regressions for either S&P or Moody's solo rated CDO/CLO tranches after Dodd-Frank

	(1) AA	(2) A	(3) BBB	(4) BB	(5) B
Solo Moody's	0.059	-0.006	0.06	0.136**	-0.074*
	(0.052)	(0.049)	(0.069)	(0.051)	(0.038)
Controls:					
Weighted avg. life	-0.08	0.052	0.093	-0.121	-0.525
	(0.091)	(0.111)	(0.189)	(0.415)	(0.313)
Weighted avg. life squared	-0.002	-0.006	-0.011	0	0.028
	(0.003)	(0.003)	(0.007)	(0.024)	(0.025)
Log tranche face value	3.081***	4.858***	1.391	-0.486	15.752***
	(0.853)	(0.826)	(1.757)	(1.991)	(1.102)
Log tranche face value squared	-0.093***	-0.151***	-0.048	0.006	-0.505***
	(0.025)	(0.025)	(0.051)	(0.060)	(0.031)
Constant	-21.522**	-35.039***	-4.657	13.675	-113.084***
	(7.145)	(7.169)	(15.716)	(16.112)	(9.526)
Year fixed effect	Yes	Yes	Yes	Yes	Yes
No. Obs.	1,376	1,381	1,349	1,333	431
R^2	0.630	0.518	0.475	0.683	0.768

Notes. We regress tranche-level yield spread on rating agencies and various control variables in the post-Dodd-Frank period. The spread is in %. We include tranches rated solely by S&P or Moody's. *Solo Moody's* is the test variable, equal to one when the tranche is rated by Moody's solely and zero otherwise. Columns (1)-(5) include the sample of tranches rated AA, A, BBB, BB and B, respectively. Standard errors (in parentheses) are clustered by years. The symbols *,**, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table EC.7 The marketing gains by tranching

	S&P			Moody's		
Tranche rating	Default probability	Gain before Dodd-Frank	**Gain after Dodd-Frank**	Expected Loss	Gain before Dodd-Frank	**Gain after Dodd-Frank**
AAA	0.150%	0.00	0.00	0.002%	0.00	0.00
AA	0.514%	0.01	0.11	0.037%	0.00	0.00
A	1.622%	0.03	0.36	0.257%	0.01	0.03
BBB	3.995%	0.23	0.84	0.869%	0.02	0.07
BB	13.587%	1.39	3.56	4.626%	0.24	0.79
B	31.246%	3.78	8.37	11.390%	0.20	0.57
Total	-	5.45	13.25	-	0.47	1.47

Notes. This table reports marketing gains from tranched debt issued by a single firm with a total value of 100. under the PD (S&P) and EL (Moody's) rating criteria. The values in column "Gain before Dodd-Frank" is taken from column "Gain $S_{i,i_k} - W_{i,i_k}$" in Table 4 and 5 of Brennan et al. (2009). Parameter assumption updatings for the post-Dodd-Frank period: $r_f = 0.3\%, r_m - r_f = 12.1\%, \sigma_m = 0.129$. The default probabilities and expected losses are taken from Standard and Poor's (2016) and Moody's (2017).

Suppose that there is a pool of loans. Due to information asymmetry, each loan may be subject to a wide information gap between the investor and the issuer. The issuer can tranche the pool to create a relatively safe security (senior tranche) and a relatively unsafe one (junior tranche). Thanks to the aggregation of assets and the protection from the junior tranche, intuitively the senior tranche should have less information gap compared to the individual loans. However, this cannot be proved easily using our model in Section 4, which concerns the tranching phase of securitization. Nevertheless, we can address another equally important question: What is a good rating criterion for the pooling of loans? We shall consider a desirable property of rating criteria within the pooling phase of securitization, through a concept of *pooling effect consistency*. This concept builds on the simple intuition that, for assets with similar losses, the senior tranche should become safer if the number of pooled assets grows, due to the diversification.

We will illustrate this idea by considering a simple one-factor model for the credit portfolio, in line with Coval et al. (2009), which is also widely applied in the industry (e.g., Gordy 2003). We assume that the losses $L_1, L_2, \dots \in \mathcal{L}_{[0,1]}$ from the underlying assets are conditionally i.i.d. given a common factor Z. Note that here Z no longer represents private information as in Section 4.1, but a common factor in the pool of assets. We normalize the portfolio losses with the total nominal value, and this is denoted by $L^{(\ell)} = (L_1 + \dots + L_\ell)/\ell$ for a positive integer ℓ. The most senior tranche after structuring is $((L^{(\ell)} - K)_+, 1-K)$ for some $K \in (0,1)$. In practice, the level K for the most senior tranche should be chosen as relatively large (compared to the mean of $L^{(\ell)}$) so that the senior tranche has a desirable credit rating. This motivates the introduction of the following requirement.

[PE] *Pooling effect consistency*: As the number of assets in the pool increases, the most senior tranche should be safer. More precisely, for all conditionally iid sequences $L_1, L_2, \dots \in \mathcal{L}_{[0,1]}$, $\ell \leqslant \ell'$, and $K \in (0,1)$,

$$I((L^{(\ell)} - K)_+, 1-K) \geqslant I((L^{(\ell')} - K)_+, 1-K).$$

PROPOSITION EC.1. *The EL criterion satisfies [PE], whereas for $n \geqslant 2$, the PD criterion does not.*

By Proposition EC.1, among pools $L^{(\ell)}$ with different sizes ℓ, a higher rating for the most senior tranche is directly linked to a larger ℓ via the EL criterion, but this does not hold for the PD criterion. Therefore, the EL criterion can more consistently reflect the diversification of assets in a pool, assigning a higher rating for the most senior tranche to a more diversified pool.

Proof of Proposition EC.1.

(i) A random variable $X \in \mathcal{L}^\infty$ is said to be smaller than $Y \in \mathcal{L}^\infty$ in increasing convex order, denoted by $X \leqslant_{\text{icx}} Y$, if $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for all increasing convex $\phi: \mathbb{R} \to \mathbb{R}$. Increasing convex order is a commonly used partial order among losses, and it is equivalent to second-order stochastic dominance up to a sign change.

For $\ell \leqslant \ell'$, with the conditional i.i.d. assumption, it follows from Example 3.A.29 of Shaked and Shanthikumar (2007) that $L^{(\ell')} \leqslant_{\text{icx}} L^{(\ell)}$. Note that the function $x \mapsto (x-K)_+/(1-K)$ is increasing and convex. Therefore,

$$\mathbb{E}\left[\frac{(L^{(\ell')} - K)_+}{1-K}\right] \leqslant \mathbb{E}\left[\frac{(L^{(\ell)} - K)_+}{1-K}\right].$$

Since the EL criterion I is generated by the expectation, we have $I((L^{(\ell')} - K)_+, 1-K) \leqslant I((L^{(\ell)} - K)_+, 1-K)$. Hence I satisfies pooling effect consistency.

(ii) It suffices to provide a counter-example to [PE] for the PD criterion. Take a sequence of i.i.d. Bernoulli random variables $L_1, L_2, \ldots$ with $\mathbb{P}(L_1 = 1) = 1 - \mathbb{P}(L_1 = 0) = p_1$. Take $K = 1/3$. We have $\mathbb{P}((L^{(1)} - K)_+ > 0) = \mathbb{P}(L^{(1)} > 0) = p_1$ and

$$\mathbb{P}((L^{(2)} - K)_+ > 0) = \mathbb{P}(L^{(2)} > 0) = 2p_1 - p_1^2 > p_1.$$

Therefore, by definition of the PD criterion I, we have

$$I((L^{(1)} - K)_+, 1 - K) = 1 \quad \text{and} \quad I((L^{(2)} - K)_+, 1 - K) > 1.$$

Hence, I does not satisfy [PE]. Q.E.D.

E. Additional Examples

In this section, we collect a few examples mentioned in the paper which are used to discuss some properties of rating criteria.

Example EC.1 (Self-consistency of the PD criterion for $n = 2$). In this example, we give an MM-balanced price functional weakly implied by the PD criterion, showing that the PD criterion is self-consistent for $n = 2$, as claimed in Section 5.

Let the functional $p : \mathcal{L}_1 \to \mathbb{R}$ be given by $p(L) = 1 - \mathrm{VaR}_{1-q_1}(L)$ where VaR_t for $t \in (0, 1)$ is the VaR of a random variable X, given by $\mathrm{VaR}_t(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant t\}$. We next verify that p is indeed an MM-balanced subjective price weakly implied by I, the PD criterion in Example 1. For $(L_1, M), (L_2, M) \in \mathcal{X}$, if $I(L_1, M) > I(L_2, M)$, then $\mathbb{P}(L_1 > 0) > q_1$ and $\mathbb{P}(L_2 > 0) \leqslant q_1$, implying $\mathrm{VaR}_{1-q_1}(L_1) > 0$ and $\mathrm{VaR}_{1-q_1}(L_2) = 0$. Therefore, $p(L_1) = 1 - \mathrm{VaR}_{1-q_1}(L_1) < 1 = 1 - \mathrm{VaR}_{1-q_1}(L_2) = p(L_2)$, and (18) is satisfied. Hence p is weakly implied by I. The unit price p is MM-balanced because VaR is well known to be comonotonic additive (e.g., McNeil et al. (2015, Proposition 7.20)) and admits a Choquet representation.

Example EC.2 (A modification of the PD criterion). In this example we give a rating criterion, which is not self-consistent, and yet it does not yield Implications (i) and (ii) for any price vectors.

Let $n \geqslant 3$ and I be a rating criterion that satisfies that, for some numbers $q_0 < 0 < q_1 < \cdots < q_n = 1$,

$$I_k = \{(L, M) \in \mathcal{X} : \mathbb{P}(L/M = 1) \in (q_{k-1}, q_k]\}, \quad k = 1, \ldots, n.$$

In other words, I is generated by $X \mapsto \mathbb{P}(X = 1)$.

For a normalized security $(L, 1)$ and any tranching scheme $(K_1, ..., K_m)$, we note that $L = 1$ implies $L - K_j \geqslant K_{j-1} - K_j$ for each j, and hence

$$\mathbb{P}\left(\frac{(L - K_j)_+ \wedge (K_{j-1} - K_j)}{K_{j-1} - K_j} = 1\right) \geqslant \mathbb{P}(L = 1).$$

Thus, by definition, each tranche will have a worse or equal rating as the original security $(L,1)$. Therefore, to maximize the total portfolio value (19) for any price vector, the optimal strategy is to maintain the original bond with no further tranching.

Next, we show that I is not self-consistent, in a way similar to the PD criterion. For a constant $\varepsilon \in (0,1)$, it is clear that $(\varepsilon,1)\in\mathcal{X}$ has the highest rating according to I since $\mathbb{P}(\varepsilon=1)=0$. Suppose for the purpose of contradiction that there exists an MM-balanced subjective price p weakly implied by I. We have

$$(1-\varepsilon)p(0)+\varepsilon p(1)=p(\varepsilon).$$

Taking $\varepsilon\uparrow 1$ on both sides of the above equation, we have

$$\lim_{\varepsilon\uparrow 1} p(\varepsilon)=p(1).$$

Take $(L,1)\in\mathcal{X}$ such that $\mathbb{P}(L=1)\in(q_1,q_2]$, or equivalently, $I(L,1)=2$. Note that $(L,1)$ has a better rating than $(1,1)$ and a worse rating than $(\varepsilon,1)$. By (18), we have $p(\varepsilon)>p(L)>p(1)$. Therefore,

$$p(1)<p(L)\leqslant \lim_{\varepsilon\uparrow 0} p(\varepsilon)=p(1),$$

leading to a contradiction. We conclude that there does not exist an MM-balanced subjective price p weakly implied by I.

F. Proofs and Some Technical Discussions

F.1. Proof of Propositions 1 and 2

Proof of Proposition 1 Take any $\mathbf{K}\in\mathcal{T}_M$ and denote by L_j the unit loss of the j-th tranche. Note that

$$\frac{(L-K_j)_+}{K_{j-1}-K_j}\wedge 1 \geqslant \frac{(L-K_j)_+}{M-K_j}\wedge 1.$$

By monotonicity and nominal-invariance of I and I' and the fact that they are comparable, we have

$$I'(L_j)=I'\left(\frac{(L-K_j)_+}{K_{j-1}-K_j}\wedge 1\right)\geqslant I'\left(\frac{(L-K_j)_+}{M-K_j}\wedge 1\right)=I\left(\frac{(L-K_j)_+}{M-K_j}\wedge 1\right).$$

Moreover, since I satisfies (5), we have

$$I\left(\frac{(L-K_j)_+}{M-K_j}\wedge 1\right)=I\left(\frac{(L-K_j)_+}{K_{j-1}-K_j}\wedge 1\right)=I(L_j).$$

This shows that $I'(L_j)\geqslant I(L_j)$ for each $j=1,\dots,m$. Hence, each tranche in the portfolio is either better rated by I than by I' or equally rated by both. This implies $V(\mathbf{K},\mathbf{P},I)\geqslant V(\mathbf{K},\mathbf{P},I')$. Q.E.D.

Proof of Proposition 2 The "only-if" direction: Since $\mathbf{K} \subseteq \mathbf{K}'$, some tranches in $\mathbf{K}$ are further divided into several subtranches in $\mathbf{K}'$. By (5), the lowest rated subtranche from such a division has the same rating as the original tranche, and all other subtranches have at least the same rating as the original tranche. Therefore, the total deal value cannot decrease, showing $V(\mathbf{K}, \mathbf{P}, I) \leqslant V(\mathbf{K}', \mathbf{P}, I)$.

The "if" direction: Suppose for the purpose of contradiction that I does not satisfy (5). Then there exist $(L, M) \in \mathcal{X}$ and $K \in (0, M)$ such that $I(L \wedge K, K) \neq I(L, M)$. Note that $(L \wedge K)/K \geqslant L/M$, and by monotonicity of the rating criterion I, we know $I(L \wedge K, K) > I(L, M)$. Denote by $k = I(L \wedge K, K)$ and $\ell = I(L, M)$. Take tranching schemes $\mathbf{K} = \{0\}$ and $\mathbf{K}' = \{K, 0\}$, which clearly satisfies $\mathbf{K} \subseteq \mathbf{K}'$. Take a price vector $\mathbf{P} = (P_1, \ldots, P_n)$ satisfying $P_1(M - K) + P_k K < P_\ell M$, which is always true if $\ell = 1$, and in case $\ell > 1$ we can easily find such $\mathbf{P}$ because the only constraint is $P_1 > P_\ell > P_k > 0$. It then follows that

$$V(\mathbf{K}', \mathbf{P}, I) \leqslant P_1(M - K) + P_k K < P_\ell M = V(\mathbf{K}, \mathbf{P}, I),$$

which contradicts the $V(\mathbf{K}', \mathbf{P}, I) \geqslant V(\mathbf{K}, \mathbf{P}, I)$. Therefore, I satisfies (5).

For the last statement of the proposition, it suffices to note that if one of the subtranches obtained from $\mathbf{K}'$ has a higher rating than the original tranche in $\mathbf{K}$ then there is an increase in the unit price of this subtranche. Since there is no decrease of unit price in all subtranches, we know that $V(\mathbf{K}', \mathbf{P}, I) > V(\mathbf{K}, \mathbf{P}, I)$ holds. Q.E.D.

F.2. Additional Notation and Some Lemmas

To prepare for the proofs of our main theorems, we first collect some notation used throughout our proofs. Recall that $\mathcal{X}$ is the set of defaultable securities, defined as

$$\mathcal{X} = \{(L, M) \in \mathcal{L} \times \mathbb{R}_+ : 0 \leqslant L \leqslant M\},$$

and $\mathcal{L}_1$ is the set of normalized losses, defined as

$$\mathcal{L}_1 = \{X \in \mathcal{L} : 0 \leqslant X \leqslant 1\}.$$

A rating criterion maps $\mathcal{X}$ to $\{1, \ldots, n\}$, and we also write $I(L) = I(L, 1)$ for $L \in \mathcal{L}_1$. A tranching scheme of (L, M) with m tranches is finite subset $\{K_1, ..., K_m\}$ of $[0, M]$. We always use the order $M > K_1 > \cdots > K_{m-1} > K_m = 0$. For a defaultable security (L, M), a tranching scheme $\mathbf{K} = \{K_1, ..., K_m\}$, and a rating-implied price vector $\mathbf{P}$, the total deal value is given by

$$V(\mathbf{K}) := V(\mathbf{K}, \mathbf{P}, I) = \sum_{j=1}^{m} P_{I(L_j)}(K_{j-1} - K_j), \tag{EC.1}$$

where $L_j = (L - K_j)_+ \wedge (K_{j-1} - K_j)/(K_{j-1} - K_j)$ is the unit loss of the j-th tranche with L_1 being the most senior, and $K_0 = M$. We say that K_j is the level of the j-th tranche. The set $\mathcal{T}_M$ is the set of all tranching schemes for bonds with nominal amount M. We will use $V(\mathbf{K})$ if the price vector $\mathbf{P}$ and the rating criterion I are fixed and clear from the context.

As in Examples 1 and 2, the PD criterion and the EL criterion are defined respectively through equations, for some numbers $q_0 < 0 < q_1 < \cdots < q_n = 1$,

$$I(L,M) = k \iff \mathbb{P}(L>0) \in (q_{k-1}, q_k], \quad k = 1, \ldots, n, \tag{EC.2}$$

and

$$I(L,M) = k \iff \mathbb{E}[L/M] \in (q_{k-1}, q_k], \quad k = 1, \ldots, n. \tag{EC.3}$$

We will constantly refer to these two rating criteria.

To prove our main results, an important property is the tranche-cut invariance property (5) in Section 3:

$$I(L \wedge K, K) = I(L,M) \text{ for all } (L,M) \in \mathcal{X} \text{ and } K \in (0,M],$$

which is satisfied by the PD criterion but not the EL criterion. In addition, we will also use a continuity condition

$$\lim_{K \downarrow 0} I((L-K)_+, M-K) = I(L,M) \text{ for all } (L,M) \in \mathcal{X} \tag{EC.4}$$

which is satisfied by both the PD and the EL criteria. Similar statements hold for the scenario-based PD criterion used by S&P in Example 3 and the scenario-based EL criterion used by Moody's in Example 4.

A subjective unit price is a functional $p : \mathcal{L}_1 \to [0,\infty)$ satisfying monotonicity

$$p(L_1) \geqslant p(L_2), \quad \text{for all } L_1, L_2 \in \mathcal{L}_1 \text{ with } L_1 \leqslant L_2. \tag{EC.5}$$

Moreover, p is weakly rating-implied if

$$I(L_1,1) < I(L_2,1) \Rightarrow p(L_1) > p(L_2), \quad \text{for all } L_1, L_2 \in \mathcal{X} \tag{EC.6}$$

and p is *MM-balanced* if it satisfies

$$(M-K)p\left(\frac{(L-K)_+}{M-K}\right) + Kp\left(\frac{L \wedge K}{K}\right) = Mp\left(\frac{L}{M}\right) \tag{EC.7}$$

for all $(L,M) \in \mathcal{X}$ and $K \in [0,M]$. For a defaultable security (L,M), a tranching scheme $\mathbf{K} = \{K_1, ..., K_m\}$, and a subjetive unit price p, the total deal value is given by

$$V(\mathbf{K}) := V(\mathbf{K}, p) = \sum_{j=1}^{m} p(L_j)(K_{j-1} - K_j), \tag{EC.8}$$

where $L_j = (L-K_j)_+ \wedge (K_{j-1}-K_j)/(K_{j-1}-K_j)$ is the unit loss of the j-th tranche. Note that (EC.8) is more general than (EC.1) based on a rating-implied price vector, and the dependence via I is implicit through p. If p is rating-implied (thus represented by a price vector $\mathbf{P}$), then (EC.8) is the same as (EC.1). Obviously, if p is MM-balanced, then (EC.8) is a constant cross different choices of $\mathbf{K}$.

For given (L,M) and I, we say that the tranching scheme $\mathbf{K}$ is maximal if it has the largest possible number N_I of distinct rating categories.

LEMMA EC.1. *The PD and the scenario-based PD criteria both satisfy* (5) *and* (EC.4)*. The EL and the scenario-based EL criteria satisfy* (EC.4)*, but not* (5) *for* $n \geqslant 3$*.*

Proof. We first check that the PD criterion I satisfies (5) and (EC.4).

1. Because $\mathbb{P}(L>0)=\mathbb{P}(L\wedge K>0)$ for all $(L,M)\in\mathcal{X}$ and $K\in(0,M]$, by (EC.2), we have $I(L\wedge K,K)=I(L,M)$, and thus (5) is satisfied.
2. It is well known that the survival function $x\mapsto\mathbb{P}(L>x)$ of any random variable L is right-continuous in x. Hence,
$$\mathbb{P}(L>K)\uparrow\mathbb{P}(L>0)\quad\text{as } K\downarrow 0.$$
Using this fact and (EC.2), PD satisfies (EC.4).

Next, we check the two conditions for I being the scenario-based PD criterion in Example 3.

1. Note that for $(L,M)\in\mathcal{X}$ and $K\in(0,M]$,
$$\begin{aligned} I(L\wedge K,K) &= \max\{k\in\{1,\dots,m\}:\mathbb{P}(L\wedge K>0|S_k)>0\}+1\\ &= \max\{k\in\{1,\dots,m\}:\mathbb{P}(L>0|S_k)>0\}+1\\ &= I(L,M).\end{aligned}$$
Therefore, I satisfies (5).
2. Note that for any $k=1,\dots,m$ and $K>0$,
$$\mathbb{P}((L-K)_+>0|S_k)=\mathbb{P}(L>K|S_k).$$
Note that if $\mathbb{P}(L>0|S_k)=0$ then $\mathbb{P}(L>K|S_k)=0$. Moreover, if $\mathbb{P}(L>0|S_k)>0$, then $\mathbb{P}(L>K|S_k)>0$ for K small enough, since the conditional survival function $K\mapsto\mathbb{P}(L>K|S_k)$ is right-continuous. Running this argument through all finitely many k, we obtain that for $K>0$ small enough, for all $k=1,\dots,m$,
$$\mathbb{P}(L>0|S_k)>0 \iff \mathbb{P}((L-K)_+>0|S_k)>0. \tag{EC.9}$$

As a consequence of (EC.9), we have, for K small enough,

$$
\begin{aligned}
& I((L-K)_+, M-K) \\
& = \max\{k \in \{1,\ldots,m\} : \mathbb{P}((L-K)_+ > 0 | S_k) > 0\} + 1 \\
& = \max\{k \in \{1,\ldots,m\} : \mathbb{P}(L > 0 | S_k) > 0\} + 1 \\
& = I(L, M).
\end{aligned}
$$

Therefore, $\lim_{K \downarrow 0} I((L-K)_+, M-K) = I(L, M)$, and (EC.4) is satisfied.

Finally, we check the conditions for I being the EL criterion. Since the scenario-based EL in Example 4 is simply EL with a different probability measure, its properties are similar.

1. Since $n \geqslant 3$, there exists $1 > \varepsilon > 0$ such that $I(\varepsilon) = 2$ since $I(0) = 1$ and $I(1) \geqslant 3$ by definition. It is clear that $I(\varepsilon \wedge \varepsilon, \varepsilon) = I(1) \geqslant 3 > I(\varepsilon, 1)$, and hence (5) does not hold for EL.
2. Note that $(L-K)_+/(M-K)$ is increasing as K decreases, and
$$\mathbb{E}\left[\frac{(L-K)_+}{M-K}\right] \uparrow \mathbb{E}\left[\frac{L}{K}\right] \quad \text{as } K \downarrow 0.$$
Hence, if $\mathbb{E}[L/M] \in (q_{k-1}, q_k]$, then so is $\mathbb{E}[(L-K)_+/(M-K)]$ for K close to 0. Using this fact and (EC.3), EL satisfies (EC.4).

Hence, PD and generalized PD satisfy both conditions whereas EL and generalized EL only satisfy (EC.4). Q.E.D.

Next lemma gives an important consequence of the tranche-cut invariance property (5). A tranching scheme of (L, M) is said to be maximal if it has the largest possible number (i.e., N_I) of distinct rating categories among all tranching schemes of (L, M).

Lemma EC.2. *Suppose that the rating criterion I satisfies (5). For a non-maximal tranching scheme $\mathbf{K}$ and a rating implied price vector $\mathbf{P}$, there exists a maximal tranching scheme $\mathbf{K}^+$ such that $V(\mathbf{K}) < V(\mathbf{K}^+)$.*

Proof. Let $\mathbf{K} = \{K_1, \ldots, K_{m_1}\}$ be a non-maximal tranching scheme and $\mathbf{K}' = \{K'_1, \ldots, K'_{m_2}\}$ be a maximal tranching scheme. Consider the tranching scheme $\mathbf{K}^+$ constructed from including all tranching levels of $\mathbf{K}$ and $\mathbf{K}'$; that is,

$$\mathbf{K}^+ = \{K_1^+, \ldots, K_m^+\} = \{K_1, \ldots, K_{m_1}, K'_1, \ldots, K'_{m_2}\}. \tag{EC.10}$$

where $K_1^+, \ldots, K_m^+$ are ordered in decreasing order. In other words, some tranches in $\mathbf{K}$ are further divided into several subtranches in $\mathbf{K}^+$. Note that by (5), the lowest rated subtranche from such a division has the same rating as the original tranche, and all other subtranches have at least the same rating as the original tranche. Therefore, all rating categories that exist in either $\mathbf{K}$ or $\mathbf{K}'$

also exist in $\mathbf{K}^+$, and hence $\mathbf{K}^+$ is maximal. Moreover, since $\mathbf{K}$ is not maximal, at least one of the new subtranche has a higher rating than the original tranche; no subtranche has lower rating than the original tranche. Therefore, $V(\mathbf{K}^+) > V(\mathbf{K})$. Q.E.D.

Lemma EC.2 shows that a non-maximal tranching scheme is always strictly dominated by a maximal one, for any I satisfying (5). Lemma EC.2 will be useful in several places below as we analyze structural maximization.

F.3. Proof of Theorem 1

In what follows, let I' be EL which has $n \geqslant 3$ rating categories, and we show (11). We note that the portfolio value $V(\mathbf{K},\mathbf{P},I')$ admits a maximizer $\mathbf{K}^* \in \mathcal{T}_M$ such that

$$V(\mathbf{K}^*,\mathbf{P},I') = \max_{\mathbf{K}\in\mathcal{T}_M} V(\mathbf{K},\mathbf{P},I').$$

This assertion is true because (a) it suffices to consider tranching schemes $\mathbf{K}$ of dimension at most n as combining tranches with the same rating does not change the deal value, and (b) for $\mathbf{K}$ with the fixed dimension n, the value $V(\mathbf{K},\mathbf{P},I')$ is upper semi-continuous in each component of $\mathbf{K}$. Therefore, a maximizer $\mathbf{K}^* = \{K_1^*,\dots,K_n^*\}$ exists.

Let L_j^* be the unit loss of the j-th tranche of the scheme $\mathbf{K}^*$, $j = 1,\dots,n$. If there exists j such that $I(L_j^*) < I'(L_j^*)$, then using the above result, we have

$$\sup_{\mathbf{K}\in\mathcal{T}_M} V(\mathbf{K},\mathbf{P},I) \geqslant V(\mathbf{K}^*,\mathbf{P},I) > V(\mathbf{K}^*,\mathbf{P},I') = \max_{\mathbf{K}\in\mathcal{T}_M} V(\mathbf{K},\mathbf{P},I'),$$

and hence (11) holds. If $I(L_j^*) = I'(L_j^*)$ for all j, we consider two cases:

(a) $\mathbf{K}^*$ has at least 3 distinct rating categories under I'. In this case, there exists j such that $I'(L_1^*) < I'(L_j^*) < I'(L_{j+1}^*)$. By definition of EL, we can check

$$\inf\{k > 0 : I'((L-k)_+ \wedge (M-k), M-k) \leqslant I'(L_j^*)\} < K_j.$$

Since I and I' are comparable and I satisfies (i), we know that there exists $k' < K_j$ such that the tranche $(L-k')_+ \wedge (M-k'), M-k')$ is rated at least as good as $I'(L_j^*)$. Hence, we can replace K_j^* by k' and maintain all other levels in $\mathbf{K}^*$, and denote the resulting tranching scheme by $\mathbf{K}'$. This leads to $V(\mathbf{K}',\mathbf{P},I) > V(\mathbf{K}^*,\mathbf{P},I')$, and hence (11) holds.

(b) $\mathbf{K}^*$ has only 2 distinct rating categories under I'. Since $I(L_j^*) = I'(L_j^*)$ for all j and $N_I \geqslant 3$, we know that $\mathbf{K}^*$ is not maximal. By Lemma EC.2, there exists a maximal tranching scheme $\mathbf{K}^+$ such that $V(\mathbf{K}^+,\mathbf{P},I) > V(\mathbf{K}^*,\mathbf{P},I')$, and hence (11) holds.

Finally we only need show (12) for $N_I = N_{I'} = 2$, which means that there are (and can only be) two rating categories. It is easy to see that the maximizer $\mathbf{K}^*$ for $V(\mathbf{K},\mathbf{P},I)$ can be chosen as $\mathbf{K}^* = \{K, 0\}$ where K is the smallest number such that the senior tranche has the better rating, making the senior tranche as large as possible. Since I' and I are comparable, $\mathbf{K}^*$ is also a maximizer of $V(\mathbf{K},\mathbf{P},I')$ by the same logic, and we have $V(\mathbf{K},\mathbf{P},I) = V(\mathbf{K},\mathbf{P},I')$. Q.E.D.

F.4. Proof of Theorem 2

(i) This part directly follows from Lemma EC.2, which says that any non-maximal tranching scheme cannot maximize (EC.1).

(ii) First, we note that there is no further improvement of having more than N_I tranches, since combining two tranches with the same rating will not increase or decrease the rating. Therefore, if a maximizer for (EC.1) exists, then there exists a maximizer $\mathbf{K}'$ with exactly N_I tranches.

Suppose that $\mathbf{K}'$ is not equal to $\mathbf{K}$. Since both $\mathbf{K}'$ and $\mathbf{K}$ have N_I tranches, there exists j and d such that $K'_d < K_j < K'_{d-1}$ where $K'_0 = M$. Let $K^*_j = K_j + \varepsilon$ where $\varepsilon \in (0, K'_{d-1} - K_j)$. Note that by the definition of K_j via (13) and $K'_d < K_j$, this adjustment ensures that

$$I\left(\frac{(L-(K_j+\varepsilon))_+}{M-(K_j+\varepsilon)}\right) < I\left(\frac{(L-K'_d)_+}{M-K'_d}\right),$$

which implies that adding K^*_j as a new tranche level to $\mathbf{K}'$ will create a higher rated subtranche of the original d-th tranche. Since the rating of the other resulting subtranche is unaffected (similarly to the situation in the proof of Lemma EC.2), this leads to an increase of the total deal value. As a consequence, (EC.1) is not maximized by $\mathbf{K}'$. Hence, $\mathbf{K}$ is the unique maximizer to (EC.1) with N_I tranches.

(iii) For the PD criterion or the scenario-based PD criterion I, it suffices to show that $\mathbf{K}$ defined via (13) is indeed a maximizer of the deal value. Such a result actually holds for all rating criterion that satisfy both (5) and (EC.4). By (13), we know that for any tranching scheme, its size of all tranches that are rated better than or equal to j cannot be more than $M - k_j$. Moreover, the continuity property (EC.4) implies that $I((L-k_j)_+, M-k_j) \leqslant j$ for each $j = 1, \ldots, n$. Further, using (5), all tranches with level higher than or equal to k_j are at least j-rated. Therefore, the tranching scheme $\mathbf{K}$ has the largest possible size of at least j-rated tranches among all tranching schemes, for each j. Note that for any tranching scheme $\mathbf{K}'$, its deal value can be rewritten as

$$V(\mathbf{K}') = \sum_{j=1}^{n} (P_j - P_{j+1}) z_j$$

where z_j is the size of at least j-rated tranches, and $p_{n+1} = 0$. Since $\mathbf{K}$ has the largest $z_1, \ldots, z_n$ among all tranching schemes, we know that $\mathbf{K}$ is a maximizer for the deal value. Q.E.D.

F.5. Proof of Proposition 3

For a given tranche, Assumption 1 and monotonicity and law invariance of the rating criterion imply that the rating of the tranche is monotonically increasing in z; i.e., a larger value of z leads to a worse or equal rating. Therefore, for each $j = 1, \ldots, m$ where m is the number of tranches in

$\mathbf{K}$, the set $\{z' \in D : R_j(z') = R_j(z)\}$ is an interval. Since $A_{\mathbf{K}} = \bigcap_{j=1}^{m} \{z' \in D : R_j(z') = R_j(z)\}$, we know that $A_{\mathbf{K}}$ is also an interval. It is straightforward that $z \in A_{\mathbf{K}}$.

Next, we show (16). By the above result on $A_{\mathbf{K}}$ and Assumption 1, we have $\mathbb{E}[L \mid Z \in A_{\mathbf{K}}] \geqslant \mathbb{E}[L \mid Z \leqslant z]$. Using (15), we get

$$\mathrm{IG}_I(z) \leqslant \mathbb{E}[L \mid Z = z] - \mathbb{E}[L \mid Z \in A_{\mathbf{K}}] \leqslant \mathbb{E}[L \mid Z = z] - \mathbb{E}[L \mid Z \leqslant z],$$

which shows (16). Q.E.D.

F.6. Proof of Proposition 4

Fix $z \in D$. Let us consider a tranching scheme with $n = 2$, that is, $\mathbf{K} = \{K_1, K_2\} = \{x, 0\}$ where $x \in (0,1)$. We will show that it is possible to choose x such that $A_{\{x,0\}}$ is an interval with right end-point z.

Under Assumptions 1 and 2 and $I(\mathbb{1}_{\{L>0\}}|z) > 1$, for both PD and EL, by choosing x close to 1, we can make the senior tranche rated the highest, and by choosing x close to 0, we can make the junior tranche rated not the highest. Since I has at least two rating categories, we can either create a senior tranche that is rated better than the original L without tranching (i.e., $I(L|z) > 1$), or we can create a junior tranche that is rated worse than L (i.e., $I(L|z) = 1$; this can only happen for EL). We will assume the first situation, and the second situation is analogous and omitted. Let q_1 be the upper threshold for the first rating category of PD or EL.

We first note the useful fact that, if $A_{\{x^*,0\}}$ has right end-point z, then, by Assumption 1, $\mathbb{E}[L \mid Z \in A_{\{x^*,0\}}] \leqslant \mathbb{E}[L \mid Z = z]$. In this case, by choosing the tranching scheme $\mathbf{K} = \{x^*, 0\}$, we get $\mathrm{IG}_I(z) \geqslant \mathbb{E}[L \mid Z = z] - \mathbb{E}[L \mid Z \in A_{\mathbf{K}}] \geqslant 0$. Next, we will show that $A_{\{x^*,0\}}$ has right end-point z.

We first analyze the case of PD. Let $x^* = \inf\{x \in (0,1) : \mathbb{P}(L > x \mid Z = z) \leqslant q_1\}$. In other words, x^* is the smallest value of K_1 such that the senior tranche from level K_1 to level 1 is rated as 1 by PD. We use Assumption 2, which has two possibilities.

First, suppose that $x \mapsto \mathbb{P}(L > x \mid Z = z)$ is continuous on $(0,1)$. Since the condition $I(L|z) = I(\mathbb{1}_{\{L>0\}}|z) > 1$ yields $\mathbb{P}(L > 0 \mid Z = z) > q_1$, and $\mathbb{P}(L > x \mid Z = z) \to 0$ as $x \uparrow 1$, we know that $x^* \in (0,1)$ and $\mathbb{P}(L > x^* \mid Z = z) = q_1$. Using Assumption 1, for any $z' > z$, we have $\mathbb{P}(L > x^* \mid Z = z') > \mathbb{P}(L > x^* \mid Z = z) = q_1$. Hence, for any $z' > z$, the senior tranche from level x^* to level 1 will be rated worse than 1. This shows that $A_{\{x^*,0\}}$ has right end-point z.

Second, suppose that L given $Z = z$ is a strictly increasing function of z, denoted by $f(z)$. In this case, $x^* = f(z)$, and for any $z' > z$ $\mathbb{P}(L > x^* \mid Z = z') = 1$. Therefore, $A_{\{x^*,0\}}$ has right end-point z.

Next, we consider the case of EL, which does not require Assumption 2. Let $x^* = \inf\{x \in (0,1) : \mathbb{E}[(L - x)_+/(1 - x) \mid Z = z) \leqslant q_1\}$. The condition $I(L|z) > 1$ yields $\mathbb{E}[L \mid Z = z] > q_1$. Since $x \mapsto \mathbb{E}[(L - x)_+/(1 - x) \mid Z = z]$ is continuous on $(0,1)$ with limit 0 as $x \uparrow 1$, we know that $x^* \in (0,1)$ and

$\mathbb{E}[(L-x^*)_+/(1-x^*) \mid Z=z]=q_1$. Using Assumption 1, for any $z'>z$, we have $\mathbb{E}[(L-x^*)_+/(1-x^*) \mid Z=z'] > \mathbb{E}[(L-x^*)_+/(1-x^*) \mid Z=z]=q_1$. Hence, for any $z'>z$, the senior tranche from level x^* to level 1 will be rated worse than 1. Therefore, $A_{\{x^*,0\}}$ has right end-point z. Q.E.D.

F.7. Proof of Theorem 3

(i) We can rewrite, for a fixed $\mathbf{K}=\{K_1,\dots,K_m\}$,

$$S(\mathbf{K},z,I')-\mathbb{E}[1-L \mid Z=z]=\sum_{j=1}^{m}\left(\mathbb{E}\left[X_j \mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)\right]-\mathbb{E}\left[X_j \mid Z=z\right]\right)(K_{j-1}-K_j).$$

Let $k_j=I(X_j)$ for each j. It follows from the definition of the EL criterion that

$$q_{k_j}\leqslant \mathbb{E}[X_j \mid Z=z]<q_{k_j+1} \quad \text{and} \quad q_{k_j}\leqslant \mathbb{E}[X_j \mid R_{\mathbf{K}}(Z)=R_{\mathbf{K}}(z)]<q_{k_j+1}.$$

Therefore,

$$|\mathrm{IG}_I(z)|\leqslant \sup_{\mathbf{K}\in\mathcal{T}}\left|S(\mathbf{K},z,I')-\mathbb{E}[1-L \mid Z=z]\right|\leqslant \delta_I\sum_{j=1}^{m}(K_{j-1}-K_j)=\delta_I.$$

This shows the desired statement for EL.

(ii) To prove this statement, we will use the result in (iii). By (EC.11), we get $\mathrm{IG}_I(z)=z-\mathbb{E}[L \mid L\leqslant z]$. Taking $z\to 1$ and noting that L has positive density on $(0,1)$ gives $\mathrm{IG}_I(z)\to 1-\mathbb{E}[L]$.

(iii) Let $x^*=\inf\{x\in(0,1):\mathbb{P}(L>x \mid Z=z)\leqslant q_1\}$. Note that the conditions here are stronger than those in Proposition 3. Hence, from the proof of Proposition 3, $A_{\mathbf{K}}$ has right end-point z, where $\mathbf{K}=\{x^*,0\}$.

Next, for any $z'\leqslant z$ we have $\mathbb{P}(L>0 \mid Z=z')>1-\gamma_I$ and $\mathbb{P}(L>x^* \mid Z=z')\leqslant q_1$. This means that the senior tranche is rated as the highest and the junior tranche is rated as the lowest. Therefore, $R_{\mathbf{K}}(z')=R_{\mathbf{K}}(z)$. This implies that $A_{\mathbf{K}}\subseteq\{z'\in D: z'\leqslant z\}$. Together with the fact that the right end-point of $A_{\mathbf{K}}$ is z, we get $\mathbb{E}[L \mid Z\in A_{\mathbf{K}}]=\mathbb{E}[L \mid Z\leqslant z]$, which gives

$$\mathrm{IG}_I(z)\geqslant \mathbb{E}[L \mid Z=z]-\mathbb{E}[L \mid Z\in A_{\mathbf{K}}]=\mathbb{E}[L \mid Z=z]-\mathbb{E}[L \mid Z\leqslant z].$$

Using the upper bound (16) in Proposition 4, we get

$$\mathrm{IG}_I(z)=\mathbb{E}[L \mid Z=z]-\mathbb{E}[L \mid Z\leqslant z]. \tag{EC.11}$$

Since $\mathrm{IG}_I(z)$ attains the upper bound (16) for each z, we know that it dominates IG of any EL criterion (or any other criterion). Q.E.D.

F.8. Proof of Theorem 4

To show that the EL criterion satisfies [SC], it suffices to choose $p(L)=\mathbb{E}[1-L]$, $L\in\mathcal{L}_1$. Note that p is an MM-balanced subjective price since it is linear. Moreover, p is compatible with the EL criterion by definition in (EC.3).

To show the intersection of the set of self-consistent rating criteria and that of generalized PD criteria is empty, by Proposition 2, it is enough to show that the inequality (10) does not hold for a self-consistent rating criterion I. We consider two cases.

Case 1. $I_1=\{(L,M)\in\mathcal{X}:\rho(L/M)=0\}$. In this case, we take a constant $x\in(0,1)$ such that $(x,1)\in I_2$.

Case 2. $I_1=\{(L,M)\in\mathcal{X}:0\leqslant\rho(L/M)<a\}$ or $I_1=\{(L,M)\in\mathcal{X}:0\leqslant\rho(L/M)\leqslant a\}$ for some $a\in(0,1]$. we take a constant $x\in(0,a)$ such that $(x,1)\in I_1$.

In either case, we use a tranching scheme $\mathbf{K}=\{x/2,0\}$ to tranche the security $(x,1)$ into two tranches L_1 and L_2. In the first case, we have $L_2=(\varepsilon,\varepsilon)\in I_n$ and $L_1=(x-\varepsilon,1-\varepsilon)\in I_1$, and in the second case, we have $L_2=(\varepsilon,\varepsilon)\in I_n$ and $L_1=(x-\varepsilon,1-\varepsilon)\in I_2$. Since $n\geqslant 3$, we know that the above tranching scheme $\mathbf{K}$ reduces the deal value in either case, showing that (10) does not hold. Q.E.D.

F.9. Proof of Theorem 5

Below we prove Theorem 5 via three technical lemmas. We first list a few properties of a rating measure ρ which will be useful in the characterization of a rating criterion. Recall that $\mathcal{L}_1=\{X\in\mathcal{L}:0\leqslant X\leqslant 1\}$.

[B1] $\rho(0)=0$.

[B2] $\rho(X_1)\leqslant\rho(X_2)$ for $X_1,X_2\in\mathcal{L}_1$ with $X_1\leqslant X_2$;

[B3] For all $X\in\mathcal{L}_1$ and $\lambda\in(0,1)$,

$$(1-\lambda)\rho\left(\frac{(X-\lambda)_+}{1-\lambda}\right)+\lambda\rho\left(\frac{X\wedge\lambda}{\lambda}\right)=\rho\left(X\right). \tag{EC.12}$$

The next lemma gives a characterization of rating criteria by their rating measures.

Lemma EC.3. *A rating criterion I satisfies [SC] if and only if it is generated by a rating measure ρ satisfying [B1]-[B3].*

Proof. The "if" statement. Suppose that I is generated by ρ satisfying [B1]-[B3]. Define a functional p by

$$p(L)=\rho(1)-\rho(L),\quad L\in\mathcal{L}_1. \tag{EC.13}$$

Below we verify that p is a weakly I-implied subjective price. First, by [B2] of ρ, $\rho(X)$ is monotonically increasing in $X \in \mathcal{L}_1$. Hence, $p(L_1) \geqslant p(L_2)$ if $L_1 \leqslant L_2$. For $(L,M) \in \mathcal{X}$ and $K \in (0,M)$, let $\lambda = K/M$, and using [B3] of ρ, we have

$$\begin{aligned}
&p\left(\frac{(L-K)_+}{M-K}\right) + Kp\left(\frac{L\wedge K}{K}\right)\\
&= (M-K)\left(\rho(1) - \rho\left(\frac{(L-K)_+}{M-K}\right)\right) + K\left(\rho(1) - \rho\left(\frac{L\wedge K}{K}\right)\right)\\
&= M\left(\rho(1) - (1-\lambda)\rho\left(\frac{(L/M-\lambda)_+}{1-\lambda}\right) - \lambda\rho\left(\frac{L/M\wedge\lambda}{\lambda}\right)\right)\\
&= M\left(\rho(1) - \rho(L/M)\right) = Mp(L/M).
\end{aligned}$$

This shows that p is an MM-balanced subjective price. Moreover, p is weakly I-implied because of (1) and (EC.13). Hence, we can conclude that I satisfies [SC].

The “only-if” statement. Suppose that I satisfies [SC], and hence there exists an MM-balanced subjective price p weakly implied by I. Define

$$\rho(X) = p(0) - p(X), \quad X \in \mathcal{L}_1. \tag{EC.14}$$

We note that ρ satisfies [B1] by plugging in $X = 0$ in (EC.14) and it satisfies [B2] by (EC.5). For [B3], we can check, using the fact that p is MM-balanced,

$$\begin{aligned}
&(1-\lambda)\rho\left(\frac{(X-\lambda)_+}{1-\lambda}\right) + \lambda\rho\left(\frac{X\wedge\lambda}{\lambda}\right)\\
&= p(0) - (1-\lambda)p\left(\frac{(X-\lambda)_+}{1-\lambda}\right) - \lambda p\left(\frac{X\wedge\lambda}{\lambda}\right)\\
&= p(0) - p(X) = \rho\left(X\right).
\end{aligned}$$

Therefore, ρ satisfies [B1]-[B3]. Since p is weakly implied by I, using (EC.6) and (EC.14), we conclude that I is generated by ρ. Q.E.D.

To arrive at the statement in Theorem 5, using Lemma EC.3, it suffices to identify functionals $\rho : \mathcal{L}_1 \to \mathbb{R}$ satisfying [B1]-[B3]. Next, we provide another technical lemma on the property [B3].

Lemma EC.4. *Suppose that ρ satisfies [B3] and $\rho(0) = 0$. For any positive integer n, a decreasing sequence of sets $E_1, \dots, E_n \in \mathcal{F}$ and $a_1, \dots, a_n > 0$ with $\sum_{i=1}^n a_i \leqslant 1$, we have $\rho(\sum_{i=1}^n a_i \mathbb{1}_{E_i}) = \sum_{i=1}^n a_i \rho(\mathbb{1}_{E_i})$.*

Proof. For each $j = 1, \dots, n$, denote by

$$m_j = \rho\left(\frac{\sum_{i=j}^n a_i \mathbb{1}_{E_i}}{1 - \sum_{i=1}^{j-1} a_i}\right),$$

and $m_{n+1} = 0$. For $j = 1, \dots, n$, by choosing

$$X = \frac{\sum_{i=j}^{n} a_i \mathbb{1}_{E_i}}{1 - \sum_{i=1}^{j-1} a_i} \quad \text{and} \quad \lambda = \frac{a_j}{1 - \sum_{i=1}^{j-1} a_i}$$

in (EC.12), we obtain

$$\begin{aligned} m_j &= \rho(X) \\ &= \lambda \rho\left(\frac{X \wedge \lambda}{\lambda}\right) + (1-\lambda)\rho\left(\frac{(X-\lambda)_+}{1-\lambda}\right) \\ &= \frac{a_j}{1 - \sum_{i=1}^{j-1} a_i} \rho\left(\mathbb{1}_{E_j}\right) + \frac{1 - \sum_{i=1}^{j} a_i}{1 - \sum_{i=1}^{j-1} a_i} \rho\left(\frac{(\sum_{i=j}^{n} a_i \mathbb{1}_{E_i} - a_j)_+}{1 - \sum_{i=1}^{j} a_i}\right) \\ &= \frac{a_j}{1 - \sum_{i=1}^{j-1} a_i} \rho\left(\mathbb{1}_{E_j}\right) + \frac{1 - \sum_{i=1}^{j} a_i}{1 - \sum_{i=1}^{j-1} a_i} m_{j+1}. \end{aligned}$$

Using the above relation iteratively, we have

$$\rho\left(\sum_{i=1}^{n} a_i \mathbb{1}_{E_i}\right) = m_1 = a_1 \rho(\mathbb{1}_{E_1}) + (1-a_1) m_2 = \cdots = \sum_{i=1}^{n} a_i \rho(\mathbb{1}_{E_i}),$$

thus the desired result. Q.E.D.

The next lemma provides the final ingredients to show Theorem 5.

LEMMA EC.5. *A functional $\rho : \mathcal{L}_1 \to \mathbb{R}$ satisfies [B1]-[B3] if and only if*

$$\rho(X) = \int_0^1 g(X > x) \mathrm{d}x, \quad X \in \mathcal{L}_1 \tag{EC.15}$$

for some increasing set function $g : \mathcal{F} \to \mathbb{R}$ with $g(\varnothing) = 0$.

Proof. The “if” statement. The functional ρ defined by (EC.15) is a Choquet integral. It is well known that a Choquet integral is positively homogeneous, and additive for comonotonic random variables (see e.g. Schmeidler (1989)). Positive homogeneity implies

$$(1-\lambda)\rho\left(\frac{(X-\lambda)_+}{1-\lambda}\right) + \lambda \rho\left(\frac{X \wedge \lambda}{\lambda}\right) = \rho\left((X-\lambda)_+\right) + \rho\left(X \wedge \lambda\right),$$

and by noting that $(X-\lambda)_+$ and $X \wedge \lambda$ are comonotonic, we have

$$\rho\left((X-\lambda)_+\right) + \rho(X \wedge \lambda) = \rho((X-\lambda)_+ + X \wedge \lambda) = \rho(X),$$

and hence ρ satisfies [B3]. Further, ρ satisfies [B1] by definition, and it satisfies [B2] by noting that $g(X > x) \leqslant g(Y > x)$ if $X \leqslant Y$.

The "only-if" statement. Let $X=\sum_{i=1}^n a_i \mathbb{1}_{E_i}$ as in Lemma EC.4 where $\sum_{i=1}^n a_i \leqslant 1$. Using the conclusion of Lemma EC.4, by defining $g:\mathcal{F}\to\mathbb{R}$, $E\mapsto \rho(\mathbb{1}_E)$, we arrive at

$$\begin{aligned}\rho\left(\sum_{i=1}^n a_i \mathbb{1}_{E_i}\right) &= \sum_{i=1}^n a_i \rho(\mathbb{1}_{E_i})\\ &= \sum_{i=1}^n a_i g(E_i)\\ &= \sum_{i=1}^n \int_{\sum_{j=1}^{i-1} a_j}^{\sum_{j=1}^{i} a_j} g(E_i)\mathrm{d}x\\ &= \sum_{i=1}^n \int_{\sum_{j=1}^{i-1} a_j}^{\sum_{j=1}^{i} a_j} g\left(\sum_{i=1}^n a_i \mathbb{1}_{E_i} > x\right)\mathrm{d}x\\ &= \int_0^1 g\left(\sum_{i=1}^n a_i \mathbb{1}_{E_i} > x\right)\mathrm{d}x.\end{aligned}$$

Therefore, if $X\in\mathcal{L}_1$ takes finitely many discrete values, then $\rho(X)=\int_0^1 g(X>x)\mathrm{d}x$ for some increasing set function g with $g(\varnothing)=\rho(0)=0$.

Let $\mathcal{L}_D$ be the set of random variables in $\mathcal{L}_1$ that take finitely many discrete values. Let the Choquet integral $\rho_g:\mathcal{L}_1\to\mathbb{R}$ be given by $\rho_g(X)=\int_0^1 g(X>x)\mathrm{d}x$. Thus, ρ and ρ_g coincide on $\mathcal{L}_D$. Note that ρ_g is Lipchitz-continuous with respect to the supremum-norm, which gives (see e.g. Proposition 4.11 of Marinacci and Montrucchio (2004)), for all $\varepsilon>0$,

$$|\rho_g(Y)-\rho_g(X)|\leqslant \varepsilon\rho(1) \quad \text{for } X,Y\in\mathcal{L}_1,\ |Y-X|\leqslant\varepsilon. \tag{EC.16}$$

With the help of (EC.16), we can approximate a general $X\in\mathcal{L}_1$ by a discrete approximation. Define two sequences of discrete approximations $\{\overline{X}_n\}_{n\in\mathbb{N}}\subset\mathcal{L}_D$ and $\{\underline{X}_n\}_{n\in\mathbb{N}}\subset\mathcal{L}_D$ by

$$\overline{X}_n=\frac{1}{n}\lceil nX\rceil \quad \text{and} \quad \underline{X}_n=\frac{1}{n}\lfloor nX\rfloor,$$

where for $x\in\mathbb{R}$, $\lceil x\rceil$ is the smallest integer greater than or equal to x and $\lfloor x\rfloor$ is the greatest integer smaller than or equal to x. Note that by definition, $\underline{X}_n\leqslant X\leqslant\overline{X}_n$ and $\overline{X}_n-\underline{X}_n\leqslant 1/n$. Using monotonicity of ρ in [B2] and the fact that $\rho=\rho_g$ on $\mathcal{L}_D$, we have

$$\rho_g(\underline{X}_n)=\rho(\underline{X}_n)\leqslant\rho(X)\leqslant\rho(\overline{X}_n)=\rho_g(\overline{X}_n). \tag{EC.17}$$

Using (EC.16) and letting $n\to\infty$, we obtain

$$\rho_g(\underline{X}_n)\to\rho_g(X) \quad \text{and} \quad \rho_g(\overline{X}_n)\to\rho_g(X).$$

Therefore, by (EC.17), we have $\rho(X)=\rho_g(X)$, showing that (EC.15) holds on $\mathcal{L}_1$. Q.E.D.

The statement in Theorem 5 follows directly from a combination of the results in Lemmas EC.3 and EC.5. Q.E.D.

F.10. Proof of Theorem 6

Take any price vector $\mathbf{P} = (P_1, \ldots, P_n)$. First we shall prove that Implication (i) in Section 3.5 does not hold for a self-consistent rating criterion I. Take a security $(X, 1)$ such that X has a probability mass at 0 and has a continuous density over $[0,1]$. Note that $\mathbb{P}(X > x)$ is a continuous function of $x \in (0,1)$. We design I' via, for all $x \in [0,1)$,

$$I'((X-x)_+, 1-x) = k \iff I((X-x)_+, 1-x) = k, \quad k = 1, \ldots, n, \tag{EC.18}$$

and letting $I'(L, M) = I'((X-x)_+, 1-x)$ where x is such that $\mathbb{P}(L > 0) = \mathbb{P}(X > x)$. If $\mathbb{P}(L > 0) > \mathbb{P}(X > 0)$, then we let $I'(L, M) = I(X, 1)$. Note that I' satisfies (5) and (9), and thus I' is similar to the PD criterion, and I and I' are comparable. Because of (5), for any $\mathbf{K} \subseteq \mathbf{K}'$, we have $V(\mathbf{K}, \mathbf{P}, I') \leqslant V(\mathbf{K}', \mathbf{P}, I')$ by Theorem 2. If we can find $(X, 1)$ satisfying the above condition such that there exists a tranching scheme $\mathbf{K}$ with $V(\mathbf{K}, \mathbf{P}, I) < V(\{0\}, \mathbf{P}, I)$, then we have $V(\mathbf{K}, \mathbf{P}, I) < V(\mathbf{K}, \mathbf{P}, I')$, showing that Implication (i) does not hold. Thus, it suffices to find $(X, 1)$ such that $\{0\}$ is not maximizing the deal value for I, which is an elementary exercise.

Next, we show that Implication (ii) in Section 3.5 does not hold for a self-consistent rating criterion I, which is main task in the proof. Let I be a self-consistent rating criterion of $n \geqslant 3$ rating categories. Using Theorem 5, I is generated by a rating measure

$$\rho(X) = \int_0^1 g(X > x)\mathrm{d}x, \ X \in \mathcal{L}_1, \tag{EC.19}$$

for some increasing set function $g : \mathcal{F} \to \mathbb{R}$ with $g(\varnothing) = 0$. Since I has at least three categories, g is not constantly 0. Without loss of generality, we can set $\rho(1) = g(\Omega) = 1$. In this case, for any constant $x \in [0,1]$, $\rho(x) = x$.

We look at the highest rated category,

$$I_1 = \{(L, M) \in \mathcal{X} : I(L, M) = 1\}.$$

Since I is generated by ρ, I_1 either has the form, for some $a \in [0,1)$,

$$I_1 = \{(L, M) \in \mathcal{X} : 0 \leqslant \rho(L/M) \leqslant a\}$$

or the form, for some $a \in (0,1]$,

$$I_1 = \{(L, M) \in \mathcal{X} : 0 \leqslant \rho(L/M) < a\}.$$

We investigate these cases separately.

Case 1. $I_1 = \{(L, M) \in \mathcal{X} : 0 \leqslant \rho(L/M) < 1\}$. This case is not possible since

$$\mathcal{X} \setminus I_1 = \{(L, M) \in \mathcal{X} : \rho(L/M) = 1\},$$

which cannot be further divided to two distinct rating categories generated by ρ. This contracts the fact that I has at least three rating categories.

Case 2. $I_1 = \{(L,M) \in \mathcal{X} : 0 \leqslant \rho(L/M) \leqslant a\}$ or $I_1 = \{(L,M) \in \mathcal{X} : 0 \leqslant \rho(L/M) < a\}$ for some $a \in (0,1)$. Take a security $B = (a/2, 1)$ of constant loss. By definition, $B \in I_1$.

Note that securities in I_1 has the highest unit price. Hence, it is not optimal to create more tranches with lower ratings from B. Thus, the trivial tranching scheme $\mathbf{K}_0 = \{0\}$ of B (i.e. no tranching) is not dominated by any other tranching schemes of the security.

On the other hand, using the tranching scheme $\{a/4, 0\}$ of B, we can create a tranche $(a/4, a/4) \in \mathcal{X}$ with full loss (which is rated the worst) and a tranche $(a/4, 1 - a/4) \in I_1$ which has the best rating. In other words, a maximal tranching scheme of B has at least two rating categories. Hence, $\mathbf{K}_0$ is not maximal, and thus I does not lead to Implication (ii).

Case 3. $I_1 = \{(L,M) \in \mathcal{X} : \rho(L/M) = 0\}$. In this case, either

$$I_2 = \{(L,M) \in \mathcal{X} : 0 < \rho(L/M) \leqslant a\}$$

or

$$I_2 = \{(L,M) \in \mathcal{X} : 0 < \rho(L/M) < a\}$$

for some $a \in (0,1]$. Take a security $B = (x, 1)$ of constant loss where $x \in (0, a)$. By definition, $B \in I_2$.

Since B has a constant loss, no matter how many pieces of tranches it is divided into, all tranches except one are either a sure loss or a zero loss. Therefore, B can be divided into at most three tranches of different rating categories. On the other hand, by choosing $b_1 > x$ and $b_2 < x$, the tranching scheme $\{b_1, b_2, 0\}$ tranches B into three different rating categories. Hence, maximal tranching schemes of B have three different rating categories.

For security B, a maximal tranching scheme of more than 3 levels is equivalent to one with exactly 3 levels, and this is implied by the fact that combining tranches with the same rating into one tranche does not change the portfolio value.

Below we take an arbitrary maximal tranching scheme $\mathbf{K} = \{b_1, b_2, 0\}$, $b_1 > x > b_2$, which has three tranches by the arguments above. Suppose that the three tranches of $\mathbf{K}$ are rated as 1, j and n, respectively.

Define a constant $c = (x - b_2)/(b_1 - b_2) \in (0,1)$. For some $d \in \mathbb{R}$, consider

$$\mathbf{K}_d = \{b_1 + (1-c)d, b_2 - cd, 0\},$$

which is a valid tranching scheme if $|d|$ is small enough. Note that the second tranche of $\mathbf{K}_d$ is $(x - b_2 + cd, b_1 - b_2 + d)$ which satisfies

$$\frac{x - b_2 + cd}{b_1 - b_2 + d} = \frac{(x - b_2)\left(1 + \frac{d}{b_1 - b_2}\right)}{(b_1 - b_2)\left(1 + \frac{d}{b_1 - b_2}\right)} = \frac{x - b_2}{b_1 - b_2} = c.$$

Hence, $I(x-b_2+cd, b_1-b_2+d) = I(x-b_2, b_1-b_2) = j$. Thus, the tranching scheme $\mathbf{K}_d$ also has tranches rated as 1, j and n.

In the following table we summarize the sizes of tranches in $\mathbf{K}$ and $\mathbf{K}_d$, as well as their difference.

	$\mathbf{K}$	$\mathbf{K}_d$	difference $(\mathbf{K}_d - \mathbf{K})$
1-rated tranche	$1-b_1$	$1-b_1-(1-c)d$	$-(1-c)d$
j-rated tranche	b_1-b_2	b_1-b_2+d	d
n-rated tranche	b_2	b_2-cd	$-cd$

Using the above table, we can calculate the difference between the portfolio values of $\mathbf{K}$ and $\mathbf{K}_d$. Let $\delta_1 = P_1 - P_j > 0$ and $\delta_2 = P_j - P_n > 0$. We have

$$\begin{aligned} V(\mathbf{K}_d) - V(\mathbf{K}) &= -(1-c)dP_1 + dP_j - cdP_n \\ &= -d\delta_1 + cd(\delta_1+\delta_2) \\ &= d(c\delta_1 + c\delta_2 - \delta_1). \end{aligned} \tag{EC.20}$$

If $c\delta_1 + c\delta_2 - \delta_1 \geqslant 0$, then by (EC.20), the price $V(\mathbf{K}_d)$ is at least the same as $V(\mathbf{K})$ for $d \geqslant 0$. Let

$$d^* = \frac{1-b_1}{1-c} \wedge \frac{b_2}{c}.$$

We design a new tranching scheme, as the limit of d going to d^*, in the following way.

1. If $\frac{1-b_1}{1-c} < \frac{b_2}{c}$, then let $\mathbf{K}^* = \{b_2 - cd^*, 0\}$.
2. If $\frac{1-b_1}{1-c} > \frac{b_2}{c}$, then let $\mathbf{K}^* = \{b_1 + (1-c)d^*, 0\}$.
3. If $\frac{1-b_1}{1-c} = \frac{b_2}{c}$, then let $\mathbf{K}^* = \{0\}$.

In all cases,

$$V(\mathbf{K}^*) = \lim_{d \uparrow d^*} V(\mathbf{K}_d) \geqslant V(\mathbf{K}).$$

Hence, $\mathbf{K}^*$ has less rating categories than $\mathbf{K}$, and its value is not strictly dominated by $\mathbf{K}$.

If $c\delta_1 + c\delta_2 - \delta_1 < 0$, then using (EC.20) again, we can take a negative value of d, and an analogous argument leads to a tranching scheme $\mathbf{K}^*$ with less rating categories than T, whose value is not strictly dominated by $\mathbf{K}$. Therefore, no maximal tranching scheme $\mathbf{K}$ strictly dominates all non-maximal tranching schemes.

Summarizing the above three cases, we conclude that, for any price vector $\mathbf{P}$, a maximizer to the deal value does not necessarily have the maximal number of rating categories, and thus Implication (ii) in Section 3.5 do not hold for a self-consistent I. Q.E.D.

F.11. Proof of Theorem 7

The main idea to prove Theorem 7 is similar to the one we used in the proof of Theorem 5. However, there is a technical difference between the theorems. In the proof of Theorem 5 (in particular, Lemma EC.3), the compatible subjective price is used to construct the corresponding rating measure. With the scenario relevance property [SR], the compatible subjective price is not necessarily scenario-relevant even if the rating measure is, thus making the construction of a scenario-relevant rating measure much more complicated. Below we address this technical issue.

In addition to the properties listed in Section F.8, we further introduce, for a rating measure ρ, the property [B4],

[B4] $\rho(X_1)=\rho(X_2)$ for all $X_1,X_2\in\mathcal{L}_1$ satisfying $X_1\overset{S}{\sim}X_2$.

The next lemma, which can be seen as a scenario-relevant version of Lemma EC.3, shows that [B4] corresponds to [SR] of the rating criterion I.

Lemma EC.6. *Fix a collection of scenarios S. A rating criterion I satisfies [SC] and [SR] if and only if it is generated by a rating measure ρ satisfying [B1]-[B4].*

Proof. The "if" statement. Suppose that I is generated by ρ satisfying [B1]-[B4]. By Lemma EC.3, I satisfies [SC]. Moreover, using (1), [B4] of ρ implies [SR] of I.

The "only-if" statement. Suppose that I satisfies [SC] and [SR]. By [SC], there exists an MM-balanced subjective price p weakly implied by I. Throughout the proof, we introduce the following notation: for any $X\in\mathcal{L}_1$, and fixed $j\in\{1,\dots,s\}$, let $Q_X(\cdot,j)$ be the conditional quantile function of X under S_j. In addition, let $Q_X(t,0)=0$ for $t\in(0,1)$.

Take a random variable U that is uniformly distributed on $[0,1]$ under each scenario $S_1,\dots,S_s$. To show the existence of such a random variable U, we note that for each $j=1,\dots,s$, a uniform random variable U_j on $[0,1]$ exists since the probability space $(S_j,\mathcal{F}|_{S_j},\mathbb{P}(\cdot|S_j))$ is atomless, where $\mathcal{F}|_{S_j}$ is the σ-algebra generated by the intersections of S_j and sets in $\mathcal{F}$. Then we can construct U by

$$U=\sum_{j=1}^{s}U_j\mathbb{1}_{S_j}.$$

Let T be a random variable defined as $T=\sum_{j=1}^{s}j\mathbb{1}_{S_j}$. Note that $Q_X(U,T)$ is a random variable depending on U and T, and $Q_X(U,T)\overset{S}{\sim}X$. For all $X\in\mathcal{L}_1$, define

$$\rho(X)=p(0)-p(Q_X(U,T)). \tag{EC.21}$$

We observe that ρ satisfies [B4] since $Q_X(U,T)$ is determined by the conditional distribution of X under $S_1,\dots,S_s$. Moreover, for two random variables X and Y, if $X\leqslant Y$, then $Q_X(u,j)\leqslant Q_Y(u,j)$ for all $(u,j)\in[0,1]\times\{1,\dots,s\}$. By the fact that p, as a subjective price, is decreasing in its first argument, we obtain [B2] of ρ.

Next we check [B3]. Note that, by basic properties of the quantile function, for $\lambda \in (0,1)$,

$$Q_{(X-\lambda)_+}(U,T) = (Q_X(U,T) - \lambda)_+, \tag{EC.22}$$

and

$$Q_{X\wedge\lambda}(U,T) = Q_X(U,T) \wedge \lambda. \tag{EC.23}$$

Using (EC.21)-(EC.23), for $X \in \mathcal{L}_1$, we have

$$\begin{aligned}
&(1-\lambda)\rho\left(\frac{(X-\lambda)_+}{1-\lambda}\right) + \lambda\rho\left(\frac{X\wedge\lambda}{\lambda}\right) \\
&= p(0) - (1-\lambda)p\left(\frac{Q_{(X-\lambda)_+}(U,T)}{1-\lambda}\right) - \lambda p\left(\frac{Q_{X\wedge\lambda}(U,T)}{\lambda}\right) \\
&= p(0) - (1-\lambda)p\left(\frac{(Q_X(U,T)-\lambda)_+}{1-\lambda}\right) - \lambda p\left(\frac{Q_X(U,T)\wedge\lambda}{\lambda}\right) \\
&= p(0) - (1-\lambda)p\left((Q_X(U,T)-\lambda)_+\right) - \lambda p\left(Q_X(U,T)\wedge\lambda\right) \\
&= p(0) - p(Q_X(U,T)) \\
&= \rho(X).
\end{aligned}$$

Hence, ρ satisfies [B3]. Finally, we show that I is generated by ρ. Let

$$J_k = \{\rho(L/M) : (L,M) \in I_k\}, \quad k = 1,\dots,n.$$

It suffices to show that $(J_1,\dots,J_n)$ is ordered. Let $(L_1,M),(L_2,M) \in \mathcal{X}$ be such that $I(L_1,M) > I(L_2,M)$. Using $Q_X(U,T) \overset{S}{\sim} X$ for $X \in \mathcal{L}$ and [SR], we have

$$I(Q_{L_1}(U,T),M) > I(Q_{L_2}(U,T),M).$$

Since p is weakly implied by I, using (18), we obtain

$$p(Q_{L_1}(U,T),M) < p(Q_{L_2}(U,T),M),$$

and equivalently, $\rho(L_1/M) > \rho(L_2/M)$. Therefore, $(J_1,\dots,J_n)$ is ordered. We can expand $(J_1,\dots,J_n)$ to include points outside the range of ρ on $\mathcal{L}_1$, so that $(J_1,\dots,J_n) \in \pi_n(\mathbb{R})$ and (1) holds. Therefore, I is generated by ρ. Q.E.D.

The next lemma is a scenario-relevant version of Lemma EC.5, which characterizes functionals satisfying [B1]-[B4].

Lemma EC.7. *Fix a collection of scenarios S. A functional $\rho : \mathcal{L}_1 \to \mathbb{R}$ satisfies [B1]-[B4] if and only if*

$$\rho(X) = \int_0^1 h(\mathbb{P}(X > x|S_1),\dots,\mathbb{P}(X > x|S_s))\mathrm{d}x, \quad X \in \mathcal{L}_1, \tag{EC.24}$$

for some component-wise increasing function $h : [0,1]^s \to \mathbb{R}$ with $h(\mathbf{0}) = 0$.

Proof. The "if" statement. Note that (EC.24) is a special case of the Choquet integral in (EC.15), with the set function g given by

$$g(E) = h(\mathbb{P}(E|S_1), \dots, \mathbb{P}(E|S_s)), \quad E \in \mathcal{F}.$$

Therefore, using Lemma EC.5 we know that ρ defined by (EC.24) satisfies [B1]-[B3]. To show that ρ also satisfies [B4], it suffices to notice that if $X \overset{S}{\sim} Y$, then $\mathbb{P}(X > x|S_j) = \mathbb{P}(Y > x|S_j)$ for all $x \in \mathbb{R}$ and $j = 1, \dots, s$, which implies $\rho(X) = \rho(Y)$. Hence, ρ satisfies [B4].

The "only-if" statement. By Lemma EC.5, we know that ρ has the form in (EC.15). Hence, ρ is a Choquet integral, thus comonotonic-additive. By Theorem 3 of Wang and Ziegel (2021), a comonotonic-additive functional satisfying [B4] admits the representation in (EC.24), and Proposition 2 of Wang and Ziegel (2021) verifies that h is component-wise increasing. Q.E.D.

Finally, Theorem 7 follows directly from combining the results in Lemmas EC.6 and EC.7. Q.E.D.

References

Allison, P., (2009), *Fixed Effects Regression Models.* London, England: Sage

Arellano, M., and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, **58**(2), 277–297.

Billingsley, R. S., Lamy, R. E., Marr, M. W., and Thompson, G. R. (1985), Split ratings and bond reoffering yields. *Financial Management*, **14**:59–65.

Bongaerts, D., Cremers, M., and Goetzmann, W. (2012), Tiebreaker: Certification and multiple credit ratings, *Journal of Finance*, **67**(1), 113–152.

Brennan, M. J., Hein, J., and Poon, S.-H. (2009), Tranching and rating, *European Financial Management*, **15**(5), 891–922.

Calvet, L., and Sodini, P. (2014), Twin picks: Disentangling the determinants of risk-taking in household portfolios, *Journal of Finance*, **69**(2), 867–906.

Coval, J. D., Jurek, J. W., and Stafford, E. (2009), Economic catastrophe bonds, *American Economic Review*, **99**(3), 628–666.

Fagereng, A., Guiso, L., Malacrino, D., and Pistaferri, L. (2009), Heterogeneity and persistence in returns to wealth, *Econometrica*, **88**(1), 115–170.

Gordy, M. B. (2003), A risk-factor model foundation for ratings-based bank capital rules, *Journal of Financial Intermediation*, **12**(3), 199–232.

Griffin, J. M., Nickerson, J., and Tang, D. Y. (2013), Rating shopping or catering? An examination of the response to competitive pressure for CDO credit ratings, *Review of Financial Studies*, **26**(9), 2270–2310.

Leszczensky, L., and Wolbring, T. (2022), How to deal with reverse causality using panel data? Recommendations for researchers based on a simulation study, *Sociological Methods and Research*, **51**(2), 837–865.

Livingston, M., Wei, J., and Zhou, L. (2010), Moody's and S&P ratings: Are they equivalent? Conservative ratings and split rated bond yields, *Journal of Money, Credit and Banking*, **42**(7), 1267–1293.

Marinacci, M. and Montrucchio L. (2004), Introduction to the mathematics of ambiguity, In *Uncertainty in Economic Theory*, I. Gilboa, Ed. Routledge, New York, NY, USA. 46–107.

McNeil, A. J., Frey, R. and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.

Moody's (2017), Approach to rating corporate synthetic collateralized debt obligations. Available on `https://www.moodys.com/research` (accessed December 2020).

Schmeidler, D. (1989), Subject probability and expected utility without additivity, *Econometrica*, **57**(3), 571–587.

Shaked, M. and Shanthikumar, J. G. (2007), *Stochastic orders*. Springer Science and Business Media.

Standard and Poor's (2016), Global methodologies and assumptions for corporate cash flow and synthetic CDOs. Available on `https://www.standardandpoors.com` (accessed December 2020).

Wang, R., and Ziegel, J. F. (2021), Scenario-based risk evaluation, *Finance and Stochastics*, **25**, 725–756.