On the cost-of-capital rate under incomplete market valuation

Albrecher, Hansjörg; Eisele, Karl Theodor; Steffensen, Mogens; Wüthrich, Mario V.

Published in:
Journal of Risk and Insurance

DOI:
10.1111/jori.12406

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
On the cost-of-capital rate under incomplete market valuation

Hansjörg Albrecher¹,² | Karl-Theodor Eisele³ | Mogens Steffensen⁴ | Mario V. Wüthrich⁵

Abstract

In this paper we discuss the concept of the cost-of-capital (CoC) rate for an insurance company as an equilibrium in the economic triangle of policyholders, shareholders, and the regulator. This provides a possible rationalization and an economic foundation for a quantity that is widely used in practice but whose value is typically neither technically nor economically well justified. We show how it can be well founded in such a triangular equilibrium. Under a simple one-period model and a valuation procedure of a two-price economy for illiquid assets we provide a corresponding economic-theoretical quantification for the CoC rate. The resulting rates are illustrated by a number of concrete numerical examples.

KEYWORDS

cost of capital, insurance, risk margin, solvency capital requirements, valuation

1 | INTRODUCTION AND MOTIVATION

Insurance claims constitute a downside risk for shareholders of an insurance company because they need to finance possible shortfalls in insurance claims payments. The regulator decides up to which threshold the shareholders have to come up for shortfall payments through the

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.


J Risk Insur. 2022;89:1139–1158. wileyonlinelibrary.com/journal/JORI
regulatory capital requirement, and, in this sense, the shareholders’ liabilities have the features of a limited-liability option. Since the shareholders’ capital is at risk, they expect a certain return for providing this regulatory risk-bearing capital for potential shortfall payments. Our aim is to rationalize an appropriate expected rate of return on the risk-bearing capital provided by the shareholders. The expected rate of return is split into two parts: (i) the risk-free return, and (ii) the spread above the risk-free return for compensating the bearing of insurance risk. The present paper focuses on this spread, which is called the cost-of-capital (CoC) spread. We investigate its relation to the risk margin paid by the policyholders, for a given degree of diversification of the insured risks. This relation is known as the CoC approach to risk margin calculations.

The CoC approach to calculating the risk margin came from the Swiss Solvency Test into Solvency 2 by the general acceptance of chief financial and risk officers and other practitioners in the European insurance market. Although the idea of a CoC margin is a classical economic pattern of thinking and was stated by law, it was not shaped towards modern valuation methods in mathematical finance, including incomplete market valuation. The CoC approach is just one approach to the valuation of nonhedgeable risk. This approach is considered as an alternative to other approaches conforming, by the formalization of preferences of market participants, with incomplete market valuation methods. To the knowledge of the authors only a few papers have been written on an integrate connection between the CoC approach and established valuation methods. The purpose of the present paper is to add to this literature, we provide the crucial clarification about the risk measure, the risk margin, the limited-liability option, and the CoC rate, and we give further insight.

When relating our viewpoint with the ideas that have been proposed by others, a couple of decisive aspects must be clarified. One aspect is the financial market. One can choose a general market framework and add investment decision making as an integral part of the problem (see, e.g., Barigou et al., 2019, 2021; Deelstra et al., 2020; Dhaene et al., 2017, for discrete time; Delong et al., 2019a, 2019b; Pelsser & Stadje, 2014, for continuous-time extensions; Albrecher et al., 2018, for an overview). This generality comes at the cost of a rather technical presentation with heavy notation, and investment strategies as a side ingredient or result, secondary to the fundamental perception of the CoC idea, and not leading to an intuitive understanding of suitable sizes of margins.

A second aspect is the question of (partial) hedging of insurance risks. The regulator looks at both financial and insurance risks. On the basis of the company’s proposed financial strategy and the insurance risks, he fixes an aggregate capital requirement. Even when certain strategies increase, for example, the return (stocks instead of obligations), they may lead to a higher regulator’s requirement. Thus, market hedging has to be considered in the context of the regulator’s risk assessment (denoted $\rho$ below). The hedging strategy with a minimal value of $\rho$ is called the *optimal replicating portfolio*. A different hedging strategy with a higher requirement may also lead to additional financial risks which have to be included in the acceptance condition given later in (13). Therefore, we disregard hedgeable claims and concentrate, in the first place, on establishing a theoretical substantiation of a CoC pricing of nonhedgeable claims. As such, nonhedgeable claims are assumed to be independent from financial market risk drivers, and therefore in all what follows we assume that claims are (simply) discounted or, in other words, represented by a zero-coupon bond numeraire.

Another aspect is the time horizon. One can work in a multiperiod framework such that all claims, values and decisions become (stochastic) processes and a recursive solution structure emerges (like in many of the above-mentioned references). Also, one has to be careful that the
limited-liability option is actually of American type. Such generality comes at the cost of technicality and notation concerning existence and measurability of values and strategies. Instead, we work in a one-period model such that the business is all about realizing a risk after one period, contracting its distribution, and evaluating it before the period based on appropriate notions of valuation.

A different crucial assumption is the very definition of the risk margin. The convention is to decompose the value of the insurance claim into a best-estimate part and a risk margin. This decomposition depends on whether the best-estimate is a best-estimate of the claim excluding or including the limited-liability option of the shareholders. We take the best-estimate to be defined as excluding the limited-liability option. This conforms with the perception of the best-estimate in solvency rules but it does not necessarily conform with the perception of the best-estimate in market accounting rules (because the limited-liability option typically has a value different from zero). This distinction is absolutely crucial. In accounting, the true market value of the true (net) claim appears in the balance sheet. In solvency, the ultimate result is not the balance sheet but the capital requirement, and solvency balance sheet entries can take into account artificial valuation of artificial (or gross) claims. The standard solvency approach is to work with true market values of the no-limited-liability-claim which is artificial exactly because it disregards the limited-liability option. The brief logic behind the no-limited-liability-claim is that defining bankruptcy as a situation where the assets are smaller than the liabilities and the liabilities are based on the limited-liability option, then no firm ever goes bankrupt since the value of the insureds' part of the assets is never larger than the assets in total. Skipping the limited-liability option in the liability valuation makes both insolvency and the liability well defined.

Finally, the approach to valuation in incomplete markets is crucial. One idea is to let the capital provider assume a utility function based on which valuation and financial decision making takes place (see, e.g., Delbaen, 2012; Malamud et al., 2008). Alternatively, we assume here the existence of a set of valuation measures under which valuation is performed by taking expectations. In a bid–ask spread approach to valuation, a convex set of probability measures occurs. A stricter idea would be to simply assume a single valuation measure.

The presentation of these aspects and assumptions allows for a brief systematic review of the sparse literature on the subject matter of the actual choice of the CoC rate. Engsner et al. (2017) work with a general financial market in a multiperiod framework in contrast to us. The valuation by the capital provider is based on a utility whereas our valuation is motivated by a bid–ask spread approach to incomplete market valuation. Also, their ultimate objective is not to establish a connection between the solvency requirement, the CoC, and the risk margin as ours is, although they do calculate a risk margin based on the CoC approach to risk margin calculation. Möhr (2011) works in a framework similar to Engsner et al. (2017) with the main difference being that the investment strategy which is an integral part of the multiperiod problem in a given financial market, is not allowed to be dynamic. Engsner et al. (2020) look for an integrate connection between the solvency capital requirement (SCR) and the CoC in a way similar to ours. They work with a general financial market in a multiperiod framework in contrast to us. Their incomplete market valuation is based on a given valuation measure not inferred from the market. Moreover, and most crucially, their risk margin is defined on top of a best-estimate including the limited-liability option. Similarly, but being even more general by modeling not only multiperiod but also multirisk problems, Bauer and Zanjani (2021) work in the same direction as we do. On the other hand, their examples of incomplete market valuation methods are, fundamentally different from our change-of-measure approach, based on more classical actuarial premium principles. Even further away from our approach but still pursuing,
directly or indirectly, refinements in the quantification of the CoC, recent contributions include Niehaus (2022) who examines the impact from taxation on the CoC, Chiang et al. (2022) who study the impact from opaqueness of liabilities on the CoC, and Huggenberger and Albrecht (2022) who investigate the impact of risk pooling on the capital requirements and, thereby, indirectly on the CoC.

The rest of this paper is organized as follows. Section 2 describes the general setup and regulatory concepts relevant for the subsequent analysis. Section 3 develops the main ideas of the present contribution and embeds them into the context of the field. In Section 4 we work out concrete examples for Gaussian, log-normal, and Pareto risks for both the Value-at-Risk (VaR) and the Expected Shortfall as regulatory risk measures. Under specific numerical specifications of the underlying ingredients this leads to explicit CoC rates resulting from our equilibrium approach, and it allows for quantitative insights into the sensitivities of these rates with respect to the chosen specifications. Section 5 concludes.

2 | STAKEHOLDERS AND THE CoC VIEW

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and introduce an (aggregate) insurance claim \(Y\) on this probability space. We think of a one-period model where the claim \(Y\) is paid to the policyholders at the end of the period, and the management of the insurance company has to make their business decisions at the beginning of that period, say, at time 0 with respect to managing capital. That is, the management decides on how to distribute the costs of the insurance activity to policyholders and shareholders before that insurance period starts in which claim \(Y\) manifests. If there is access to a financial market, there is a second layer of business decisions to be made with respect to managing the assets. This dimension is beyond the scope of this presentation where we focus on the pure capital management problem.

Since we disconnect our consideration from any financial market risk drivers, we may and will assume that \(Y\) describes a discounted quantity throughout this paper. We assume that the regulatory framework prescribes a risk (assessment) measure \(\rho\) as the regulatory capital requirement. That is, the regulatory capital requirement \(C\) of insurance activity \(Y\) is defined by

\[
Y \mapsto C = \rho(Y). \tag{1}
\]

**Example 2.1.** If the VaR at security level \(p \in (0, 1)\) is used as risk measure, we have

\[
C = \text{VaR}_p(Y) = \inf \{\gamma \in \mathbb{R} : \mathbb{P}[Y \leq \gamma] \geq p\} = F^{-1}(p), \tag{2}
\]

where \(F\) is the distribution function of \(Y\) and \(F^{-1}\) its left-continuous generalized inverse.

If the Expected Shortfall/Tail-Value-at-Risk (TVaR) at security level \(p \in (0, 1)\) is used as risk measure, we have

\[
C = \text{TVaR}_p(Y) = \frac{1}{1 - p} \int_p^1 \text{VaR}_\alpha(Y) d\alpha. \tag{3}
\]

Note that if the distribution function \(F\) of \(Y\) is continuous, the TVaR is equal to the conditional tail expectation (CTE) given by \(\text{CTE}_p(Y) = \mathbb{E}[Y|Y > \text{VaR}_p(Y)]\) on the same security level \(p \in (0, 1)\).
We assume that the insurance company needs to fulfill the solvency requirement (1) at the beginning of the period, otherwise it is not allowed to run its insurance business, that is, it is declared to be insolvent if it does not hold at least capital $C$.

We disentangle the capital requirement given in (1) to see the contributions provided by the insured policyholders (through the insurance premium) and by the shareholders (through investments followed by dividend payments). Let $\mu = \mathbb{E}[Y]$ be the discounted mean, also called best-estimate or pure risk premium. Of course, from a practical point of view, the regulatory capital $C$ must be greater than this mean. This is actually a critical point, since for VaR (for any security level $p < 1$) one can find bounded random variables which do not satisfy this condition. This is one of the reasons why VaR may have a rather poor behavior as a risk measure. From a mathematical point of view, the condition $C > \mu$ is a condition on the relation between the risk measure $\rho$, the security level $p$ (in case of VaR or TVaR), and the distribution of $Y$.

In general, the (total/aggregate) premium $P$ (without administrative costs and taxes) paid by the policyholders exceeds the mean of the claim: $P \geq \mu$. In this case, the mean gain above the pure risk premium $\mu$ defines the risk margin $\text{RM}$,

$$\text{RM} := P - \mu \geq 0. \quad (4)$$

It forms the safety loading in the premium $P$ above the pure risk premium $\mu$ (best-estimate). This expected gain is paid to the shareholders for risk bearing. However, this argument needs a deeper analysis because (4) only gives an expected value view, and more explanation follows in the sequel.

Modern solvency and accounting rules allow for a profit margin $\text{PM}$ paid on top of the expected claim and the risk margin such that the premium actually exceeds $\mu + \text{RM}$. If we still denote this modified premium by $P$, the profit margin $\text{PM}$ is given by

$$\text{PM} := P - \mu - \text{RM} \geq 0. \quad (5)$$

This profit margin is here, for the moment, taken to be zero, conforming with definition (4), and making the risk margin $\text{RM}$ well defined. A vanishing profit margin should also be the result of a fully competitive insurance market with similar competitors.

The positive risk margin $\text{RM}$ is in general not sufficient to meet the regulatory requirement $C$, that is, $P < C$. Therefore, the shareholders—driven by the expectation to gain the risk margin $\text{RM}$—must provide an additional capital buffer. More precisely, they must supply/invest at least the difference between the regulatory capital requirement $C$ and the premium $P$. The minimum investment from the shareholder is called the solvency capital requirement $\text{SCR}$ and is defined by

$$\text{SCR} := C - P \geq 0. \quad (6)$$

In the Swiss Solvency Test, the risk margin is called the market value margin, the quantity

$$q := C - \mu > 0 \quad (7)$$

is called the target capital, and the amount $C$ is called the supervisory provision.

Let us briefly comment on the two cases where the inequalities in (4) and (6) do not hold. It can actually be the case that $P < \mu$ under an economic equilibrium and we also discuss this case briefly below. If this is the case, the shareholders make heavy use of their limited-liability option and ruin is not an unlikely event. Also, if the security level reflected by $\rho$ is low and/or the risk margin $\text{RM}$
provided by the policyholders is large, it may happen that \( C < P \). In that case the shareholders do not need to supply any capital SCR because it is already provided by the policyholders through the (higher) margin \( RM \). One may interpret that this is the situation of captives and mutuals, where the members are simultaneously policyholders and shareholders; the recently increasingly popular P2P insurance networks might also be seen in that category, see, for example, Denuit (2019, 2020). The possibly higher premium at the beginning can then be offset by a later profit at the end of the period due to the (expected) distributions of possible gains. However, with a proper distinction between the needed \( RM \) and \( SCR \), such simultaneous roles should eventually not influence the analysis. We will assume \( SCR \geq 0 \) in the sequel.

Most of the time, the shareholders’ part on the liability side of the solvency balance sheet is much higher than \( SCR \), which is only the minimal amount for regulatory acceptance (solvency). The percentage of the shareholders’ part with respect to the SCR is called the solvency ratio. However, for the analysis of the economic triangle between shareholders, policyholders, and the regulator, it is meaningful to consider a company providing exactly the minimal amount of capital, that is, the equity equals \( SCR = C - P \).

Note there is an economic balance between the risk margin \( RM \) provided by the policyholders as “Fremd”-capital and the SCR provided by the shareholders as equity (“Eigen”-capital, or “own funds”).\(^1\) We assume that this balance is in a market equilibrium which we want to analyze further.

As mentioned before, for providing the SCR, shareholders expect in the mean the risk margin \( RM \) as gain. The CoC rate \( R_{CoC} \) is defined as the ratio between the risk margin \( RM \), the expected gain, and the SCR, the shareholders’ equity:

\[
R_{CoC} := \frac{RM}{SCR}.
\] (8)

Read as \( RM = R_{CoC} \cdot SCR \), the \( R_{CoC} \) can also be defined as the amount of “Fremd”-capital needed to fund 1 unit of investment capital (\( SCR = 1 \)). If we had worked with investment opportunities, the expected gain \( RM \) would have come on top of the return from investments in the market. Thus, \( R_{CoC} \) would have been interpreted as the rate obtained from investment in the insurance risk, exclusively.

As defined in (8), \( R_{CoC} \) is a nice pattern of thinking, yielding also intuitive economic interpretations, but it is not based on a more profound economic analysis. There is no help from definition (8) itself in determining the size of any of its ingredients. The objective of Section 3 is to establish an equilibrium condition that determines the size of the ingredients in (8) and to derive the size of \( R_{CoC} \) from that equilibrium condition based on (8).

### 3 | EQUILIBRIUM PRICE AND EQUILIBRIUM CoC RATE

We are now going to consider how the capital is distributed at the end of the period. We make three fundamental assumptions: (1) the policyholders receive no more than \( Y \), that is, they do not participate in any way in profits, (2) the policyholders have priority in compensation, that is, the available capital is first used to compensate the policyholders, and (3) the shareholders

\(^1\)For a detailed discussion of the terms “Fremd”-capital versus “Eigen”-capital, see Eisele and Artzner (2011).
do not receive less than 0, that is, the limited-liability option applies. The profit-and-loss result of the insurance company is given by \( P + \text{SCR} - Y = C - Y \) (we neglect state taxes and other administrative expenses). Under the given conditions, the capital distribution to shareholders—either as capital value to support future business activities or as a dividend payment—is

\[
(C - Y)^+.
\]  

(9)

The compensation to the policyholders is the minimum of \( C \) and \( Y \), that is, the censored claim

\[
Y \land C.
\]  

(10)

The two options of course sum up to \( C \), the available capital to be distributed, with the repartition depending on the outcome of \( Y \).

Option (9) is called the insurance (limited-liability) option, sometimes it is also called exit option or insolvency option. It has the same structure as a classical financial put option. The difference is, however, that the underlying \( Y \) is not traded in a liquid financial market. We therefore refer to the theory of bid and ask prices in incomplete markets as presented, for example, in Madan and Cherny (2010), Madan and Schoutens (2010) and Eberlein et al. (2014). According to this approach, the bid price of the insurance option (9) at the beginning of the period is given by an operator \( \Psi \) and can be written as

\[
\Psi((C - Y)^+) = \inf_{Q \in \mathcal{M}} \mathbb{E}_Q[(C - Y)^+] \geq 0,
\]  

(11)

where \( \mathcal{M} \) is a nonempty convex set of probability measures that are absolutely continuous w.r.t. the real-world measure \( \mathbb{P} \). According to Artzner et al. (1999), \( \Psi \) is a coherent risk assessment (i.e., the negative of a coherent risk measure) satisfying the Fatou-property (see also Föllmer & Schied, 2010).

Only if the bid price of the insurance option offered by the company is large enough to cover the solvency requirement from the regulator, the risk trading actually takes place. Otherwise the company is not willing to offer the regulatory capital needed for the business to be approved by the regulator, with the consequence that the insurance portfolio cannot be offered to the policyholders. Thus, we have the inequality

\[
\text{SCR} \leq \inf_{Q \in \mathcal{M}} \mathbb{E}_Q[(C - Y)^+].
\]  

(12)

By (6) it follows that

\[
P = C - \text{SCR} \geq C - \inf_{Q \in \mathcal{M}} \mathbb{E}_Q[(C - Y)^+] = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[Y \land C],
\]  

(13)

where we have used that \( C \) is deterministic, and we use the decomposition

\[
C = Y \land C + (C - Y)^+.
\]
In a competitive insurance market, the premium $P$ attains the lower bound (13), since this is sufficient for the necessary counterpart SCR to be found at the financial market. Therefore, we can state the main economic equilibrium equations for the premium, the risk margin, the SCR, and the CoC rate:

**Theorem 3.1.** Inequality (13) yields the economic premium principle

$$P = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q [Y \land C].$$

(14)

Consequently, we get for RM and SCR the equilibrium equations

$$\text{RM} = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q [Y \land C] - \mathbb{E}[Y],$$

(15)

$$\text{SCR} = \inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+].$$

(16)

Finally, the CoC rate is given by the equality

$$R_{\text{CoC}} = \frac{C - \mathbb{E}[Y]}{\inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+]} - 1.$$  

(17)

**Proof.** Relation (14) follows by replacing the inequality in (13) by an equality in equilibrium. Relation (15) follows from the definition of the risk margin in relation to premiums (4) and (14). Relation (16) follows by replacing the inequality in (12) by an equality in equilibrium. Relation (17) follows from the definition of the CoC rate (8), (15), (16), and equality in (13), that is,

$$R_{\text{CoC}} = \frac{\text{RM}}{\text{SCR}} = \frac{\sup_{Q \in \mathcal{M}} \mathbb{E}_Q [Y \land C] - \mathbb{E}[Y]}{\inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+]} = \frac{C - \inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+] - \mathbb{E}[Y]}{\inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+]} = \frac{C - \mathbb{E}[Y]}{\inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+]} - 1.$$

This proves the results.

A series of discussions are appropriate at this point.

- Engsner et al. (2017) have established similar objects in a multiperiod framework. In their particular case of a one-period model, they obtain the formulas (14) and (16) in our special
case where the set $\mathcal{M}$ is a singleton. Engsner et al. (2021) then extend the results from Engsner et al. (2017) to the case of a more general set $\mathcal{M}$ such that results in the one-period special case of Engsner et al. (2021) essentially coincide with (14) and (16). A delicate issue studied in Engsner et al. (2021) is to structure the set $\mathcal{M}$ such that the multiperiod problem is well-posed. A key difference to these works, though, is that we define the risk margin as the value in excess of the expected no-limited-liability-claim, see also the next bullet item.

- In the general setting of Theorem 3.1, it may well happen that the risk margin $\text{RM}$ and, thus, also the CoC rate is negative. This reflects a situation where the risk aversion of the regulator, specified through the chosen risk measure $\rho$ and forming the value of $C$, is in a sense smaller (leading to a small $C$) than the risk aversion of the equity providers, specified through $\mathcal{M}$ and, for a given $C$, forming the values of RM and SCR. This is a consequence of defining the risk margin as the safety loading on top of the expected no-limited-liability-claim $\mu = \mathbb{E}[Y]$, see Section 1, that is, excluding own credit risk. Basically, this means that the policyholders pay for claim compensations $Y$, although they will only receive $Y \wedge C$. An alternative (more fair) definition of the risk margin would be the safety loading on top of the expected claim including own credit risk. With the capital allocation specified in (10) we redefine $\mu = \mathbb{E}[Y \wedge C]$ instead of $\mathbb{E}[Y]$. This redefinition of $\mu$ (limited-liability pure risk premium) changes the risk margin in (15)

$$\text{RM} = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q [Y \wedge C] - \mathbb{E} [Y \wedge C],$$

and, consequently,

$$R_{\text{CoC}}^* = \sup_{Q \in \mathcal{M}} \frac{\mathbb{E}_Q [Y \wedge C] - \mathbb{E} [Y \wedge C]}{\inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+]}
= \sup_{Q \in \mathcal{M}} \frac{\mathbb{E} [(C - Y)^+]}{\inf_{Q \in \mathcal{M}} \mathbb{E}_Q [(C - Y)^+]} - 1. \quad (18)$$

Engsner et al. (2020) define the risk margin as the price in excess of the expected limited-liability-claim. So, in their particular case of a one-period model they obtain a CoC rate similar to $R_{\text{CoC}}^*$ defined above when $\mathcal{M}$ is a singleton. In a market of risk-averse investors the set $\mathcal{M}$ ends up such that

$$\sup_{Q \in \mathcal{M}} \mathbb{E}_Q [Y \wedge C] > \mathbb{E} [Y \wedge C],$$

leading to a strictly positive $\text{RM}$ and a strictly positive $R_{\text{CoC}}^* > 0$. If risk-neutrality of investors is included such that $P \in \mathcal{M}$, we can only conclude a nonnegative $\text{RM}$ and a nonnegative $R_{\text{CoC}}^*$. The distinction between the results in the theorem and the alternative specification in this paragraph may be considered as a matter of convention. It is important, however, to note that the size of the risk margin and the size of the CoC rate, for a given set $(\rho, \mathcal{M})$
depend on the convention under which we work, and it is relevant for a transparent communication with the policyholders. There is a tendency to speak about the best-estimate $\mu$ as excluding own credit risk when working in solvency (accounting for the regulators) whereas the best-estimate includes own credit risk when working in accounting (accounting for the financial market). The results of the theorem and the results in this paragraph expose the crucial importance of this distinction, and also the fairness in communication towards policyholders what they will be compensated for.

- We stress that the only external specifications in Equations (14)–(17) are the choices of the risk measure $\rho$ of the regulatory capital requirement, and the set $\mathcal{M}$ which determines the bid price in an incomplete market. In the calculations above we think of $\rho$ and $\mathcal{M}$ as being settled independently of each other. The regulators have an aversion towards risk, or rather aversion towards insolvency, that is reflected in the choice of $\rho$. Separately from that, the market participants, the investors, and the policyholders, in equilibrium “agree” on a set $\mathcal{M}$ used for the settlement of the price. Then $\rho$ and $\mathcal{M}$ are plugged into (17) to reach an equilibrium price $P$, to which the policyholders eventually need to agree. If this is the case, all stakeholders do agree on this risk exchange, which is admitted by the regulator since the capitalization needs to fulfill the solvency requirements. This is the situation that we call the economic triangle between policyholders, shareholders, and the regulator. Note that, for a fixed $\mathcal{M}$, $P$ is increasing in $C$. The interpretation is that with a given risk aversion among the investors, the regulators can increase the protection of the policyholders by increasing $C$, but this comes at the cost of a higher insurance premium. Investors with large claims win from such an increased protection whereas investors with small claims lose. Before the claim realization, obviously, we do not know who wins and who loses. However, the triangular pattern of thinking may be questioned by different arguments. A different idea is that the regulator is essentially just an ambassador for the policyholders and, thus, these two stakeholders should not be separated but be seen as one. Then the risk aversion of the policyholders is expressed simultaneously through both $\rho$ and $\mathcal{M}$, and these two quantities cannot be settled one at a time. However, even in that case the calculations above are valid. We have then an economic agreement between only two stakeholders, the policyholders and the investors, which is, again, reflected in (17). It is a strong feature of our approach that it covers both the triangular and the two-participants interpretation. In the numerical examples, below, we do not pay any attention to the possible simultaneous settlement of $\rho$ and $\mathcal{M}$. We simply take one as given and vary over the other, or vice versa, corresponding to marginal changes of the interests of the regulators and the investors, respectively. On a political level, the distinction between the triangular and two-participant interpretation relates to whether the regulation (formulated by some political organizational level, be that a nation or a unity of nations) is thought of by the policyholders as external to themselves or as part of themselves.

- Equalities in the theorem arose from a series of equilibrium and competition arguments. First the profit margin in (5) was set to zero meaning that only the risk margin is charged on top of the expected claim. Practically, one may charge an even higher premium if the market conditions allow for. Second, we replace the inequality in (13) by an equality to form (14), also by arguments about the competition and equilibrium. Actually, these two competition arguments are equivalent. Finally, we let the investment of the investor be just sufficient to run the business, that is, no overcapitalization is assumed as discussed above. This meant that exactly the amount $C - Y$, and not more than that, is to be distributed at the end of the period. Of course, one may play around with the formula to obtain versions outside this equilibrium, for example, where the policyholders and/or the shareholders inject more capital into the system than they really have to. The convention is to work with a reference entity business
(basic principles of the solvency setup) and then define the risk margin and the CoC rate for this reference entity business which uses the optimal replicating portfolio (see above) and as such is independent of entity-specific pricing and capitalization outside the equilibrium. Thus, we conclude that the idea of a reference entity business conforms with the equilibrium-competition arguments advocated in this paper. Entity-specific profit margins that may arise outside equilibrium, for example, in the case of overcapitalization, should be defined in excess of such an equilibrium-based or reference entity-based risk margin.

4 | EXAMPLES

We present three distributional examples. For each of them we assume that the risk-free return is zero. The first example is a Gaussian one, because it is convenient to calculate with Gaussian distributions, and because for a large business, with independent and light-tailed individual risks, the central limit theorem allows for a Gaussian approximation for the aggregate claim. The second example is a log-normal one. This is one of the distributional approximations that is currently used in the Swiss Solvency Test. Our third and last example considers the heavy-tailed case of a Pareto distribution.

Example 4.1 (Gaussian distribution, VaR). Let $Y$ be Gaussian distributed with mean $\mu$ and variance $\sigma^2 > 0$; neglecting the fact that $Y$ can also take negative values, as for typical parameter values this happens with very small probability. Set $\rho(Y) = \text{VaR}_p(Y)$ for security level $p \in (1/2, 1)$. We have

$$
C = F^{-1}(p) = \mu + \sigma \Phi^{-1}(p) > \mu = \mathbb{E}[Y], \quad \text{and therefore},
q = \sigma \Phi^{-1}(p),
$$

(19)

where $\Phi$ denotes the standard normal cumulative distribution function.

For the set $\mathcal{M}$ of test probabilities, determining in (16) the bid price of the insurance option, we take the exponential family

$$
\mathcal{M} := \{Q_y | y \leq y_0\},
$$

(20)

where

$$
Q_y(dy) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu - \gamma\sigma)^2}{2\sigma^2}\right)dy.
$$

(21)

To get the economic premium $P$ in (14), we calculate

$$
\mathbb{E}_{Q_y}[Y \wedge F^{-1}(p)] = F^{-1}(p) - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{F^{-1}(p)} (F^{-1}(p) - y) \exp\left(-\frac{(y - \mu - \gamma\sigma)^2}{2\sigma^2}\right)dy
$$

(22)

$$
= \mu + \sigma \chi(p, \gamma),
$$

where $\chi$ denotes the characteristic function of the heavy-tailed distribution.
where
\[
\chi(p, \gamma) := \Phi^{-1}(p) - (\Phi^{-1}(p) - \gamma)\Phi(\Phi^{-1}(p) - \gamma) - \varphi(\Phi^{-1}(p) - \gamma),
\] (23)

and \(\varphi = \Phi'\) is the density of a standard normal random variable. Note that \(\chi(p, \gamma)\) can be interpreted as the loading coefficient \(\alpha_S\) in a standard deviation premium principle \(P = E(Y) + \alpha_S\sqrt{\text{Var}(Y)}\) in actuarial language (see, e.g., Albrecher et al., 2017, Chap. 7.1).

The function \(z \mapsto z\Phi(z) + \varphi(z) = E[(z - Z^+)\] (where \(Z\) is standard normal) is strictly increasing in \(z\). Therefore, the function \(\chi(p, \gamma)\) is strictly increasing in \(\gamma\) and \[\max_{\gamma \leq \gamma_0} \chi(p, \gamma) = \chi(p, \gamma_0)\].

For \(\gamma_0 = 0.15\), we get the following table for \(\chi(p, 0.15)\):

The fact that for security levels \(p < 0.795\) the condition (4) prevails over condition (13) is due to the limited-liability and the fact that claims may take negative values in the Gaussian case. From (14)–(17), we find for \(p > 0.5\)

\[
\begin{align*}
P &= \mu + \sigma\chi(p, \gamma_0), \\
\text{RM} &= \sigma\chi(p, \gamma_0), \\
\text{SCR} &= \sigma(\Phi^{-1}(p) - \chi(p, \gamma_0)), \\
R_{\text{CoC}} &= \frac{\chi(p, \gamma_0)}{\Phi^{-1}(p) - \chi(p, \gamma_0)}. \\
\end{align*}
\] (24)

Note that the CoC rate does not depend on \(\mu\) and \(\sigma\) in this parametrization. For the security level \(p = 0.995\) of Solvency II and \(\gamma_0 = 0.15\) we find from Table 1 that the CoC rate is

\[
R_{\text{CoC}} = \frac{0.1475}{2.5758 - 0.1475} = 6.07\%.
\] (25)

Recall that the CoC spread is interpreted as the spread over the return of the default-free asset. Thus, the last result corresponds well to the wide-spread opinion in insurance business that this spread should be about 6%; see, for example, Federal Office of Private Insurance (2006), Dhaene et al. (2017, Sec. 5.2), Deelstra et al. (2020, Sec. 7.2), and the discussion in Albrecher et al. (2018, Sec. 2.2). But of course, in our example the result depends heavily on the choice of the set \(\mathcal{M}\), and in particular on the value of \(\gamma_0 = 0.15\).

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Risk margin in percentage of the standard deviation for the Gaussian VaR case</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p)</td>
<td>(\chi(p, 0.15)) (%)</td>
</tr>
<tr>
<td>0.75</td>
<td>-4.03</td>
</tr>
<tr>
<td>0.95</td>
<td>+12.03</td>
</tr>
<tr>
<td>0.99</td>
<td>+14.48</td>
</tr>
<tr>
<td>0.995</td>
<td>+14.75</td>
</tr>
</tbody>
</table>

Abbreviation: VaR, Value-at-Risk.
Recall from Section 3 that the value $R^*_{\text{CoC}}$ of the CoC rate when including own credit risk in its definition is higher. For the concrete example of a Gaussian $Y$, with some elementary calculations, one gets from (18)

$$R^*_{\text{CoC}} = \frac{p\Phi^{-1}(p) + \varphi(\Phi^{-1}(p))}{\Phi^{-1}(p) - \chi(p, \gamma)} - 1,$$

see Figure 1 for a numerical comparison.

**Example 4.2** (Gaussian distribution, TVaR). In the setting of Example 4.1, let us now consider $\rho(Y) = \text{TVaR}_p(Y)$ for some security level $p \in (1/2, 1)$. Then

$$C = \mu + \sigma \frac{\varphi(\Phi^{-1}(p))}{1 - p} > \mu = \mathbb{E}[Y], \quad \text{and therefore},$$

$$q = \sigma \frac{\varphi(\Phi^{-1}(p))}{1 - p}.$$  

(26)

For the same set $\mathcal{M}$ of test probabilities as in Example 4.1, we get

$$\mathbb{E}_{\mathcal{Q}_p}[Y \wedge C] = \mu + \sigma \chi_e(p, \gamma),$$

(27)

with

$$\chi_e(p, \gamma) := \frac{\varphi(\Phi^{-1}(p))}{1 - p} - (\varphi(\Phi^{-1}(p))/(1 - p) - \gamma)\Phi(\varphi(\Phi^{-1}(p))/(1 - p) - \gamma) - \varphi(\varphi(\Phi^{-1}(p))/(1 - p) - \gamma).$$

(28)

Like in Example 4.1, this expression is maximized for $\gamma = \gamma_0$, and for $\gamma_0 = 0.15$ we get the values depicted in Table 2.
Example 4.3 (Log-normal distribution, VaR). Let $Y$ be log-normally distributed such that $\log Y$ is normally distributed with mean $\mu_0$ and variance $\sigma^2$ or, equivalently, $Y = e^{\mu_0 + \sigma Z}$, where $Z$ is standard normal. We have pure risk premium $E[Y] = \mu = e^{\mu_0 + \frac{1}{2}\sigma^2}$. We are interested in calculating

$$\inf_{Q \in \mathcal{M}} E_Q[(C - Y)^+] \tag{29}$$

Note that apart from the infimum over measures in $\mathcal{M}$ this formula reminds of evaluating a put option in a Black-Scholes market. A crucial difference is, of course, that there is no market that teaches us a unique $Q$ to use for valuation. Instead we take the set $\mathcal{M}$ to be the set under which the mean of $\log Y$ equals $\mu_0(1 + \gamma)$ with $|\gamma| \leq \gamma_0$ and denote the measure $Q_\gamma$ correspondingly. Now for each $\gamma$ we can calculate (29) in the following way:

$$\begin{align*}
E_{Q_\gamma}[(C - Y)^+] &= E_{Q_\gamma}[(C - Y) 1_{Y \leq C}] \\
&= CQ_\gamma[Y \leq C] - \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{C} y \exp\left(-\frac{(\log y - \mu_0(1 + \gamma))^2}{2\sigma^2}\right) dy \\
&= C\Phi\left(\frac{\log C - \mu_0(1 + \gamma)}{\sigma}\right) - e^{\mu_0(1+\gamma)+\frac{1}{2}\sigma^2}\Phi\left(\frac{\log C - \mu_0(1 + \gamma)}{\sigma}\right) - \sigma.
\end{align*}$$

To calculate $C$ we assume the VaR risk measure such that

$$C = F^{-1}(p) = \exp[\mu_0 + \sigma \Phi^{-1}(p)],$$

where $F$ is the log-normal distribution function with parameters $\mu_0$ and $\sigma^2$. The expectation $E_{Q_\gamma}[(C - Y)^+]$ reaches its minimum for $\gamma = \gamma_0$. Thus, we can express the $R_{CoC}$ as

$$R_{CoC} = \frac{C - E[Y]}{E_{Q_{\gamma_0}}[(C - Y)^+]} - 1 = \frac{C - \mu}{E_{Q_{\gamma_0}}[(C - Y)^+]} - 1.$$
Note that in contrast to the Gaussian case, this CoC rate depends on $\mu_0$ and $\sigma$. For parameters $\mu_0 = 0.1$ and $\sigma = 0.1$ (signifying a coefficient of variation for $Y$ of 0.1, which is a magnitude of practical relevance) and $\gamma_0 = 0.15$, we get the following table:

<table>
<thead>
<tr>
<th>$p$</th>
<th>$R_{\text{CoC}}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>-8.9</td>
</tr>
<tr>
<td>0.95</td>
<td>7.1</td>
</tr>
<tr>
<td>0.99</td>
<td>6.0</td>
</tr>
<tr>
<td>0.995</td>
<td>5.4</td>
</tr>
</tbody>
</table>

The table illustrates for $p = 0.75$ again the possibility of a negative CoC rate for low $p$. For the parameters $\mu_0 = 0.1$, $\sigma = 0.1$, and $p = 0.995$, when varying the parameter $\gamma_0$, we get the following table:

<table>
<thead>
<tr>
<th>$\gamma_0$</th>
<th>$R_{\text{CoC}}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.7</td>
</tr>
<tr>
<td>0.10</td>
<td>3.5</td>
</tr>
<tr>
<td>0.15</td>
<td>5.4</td>
</tr>
<tr>
<td>0.20</td>
<td>7.4</td>
</tr>
</tbody>
</table>

The table illustrates the increasing CoC rate in response to the increasing risk aversion by market participants for a given level of risk tolerance of the regulator.

**Example 4.4** (Log-normal distribution, TVaR). In the log-normal setting of Example 4.3 let us now consider $\rho(Y) = \text{TVaR}_p(Y)$ for some security level $p \in (1/2, 1)$. The only change is that we have to replace the VaR capital requirement by the corresponding TVaR requirement

$$C = \frac{\mu}{1 - p} (1 - \Phi[\Phi^{-1}(p) - \sigma]).$$

For parameters $\mu_0 = 0.1$ and $\sigma = 0.1$ (signifying a coefficient of variation for $Y$ of 0.1, which is a magnitude of practical relevance) and $\gamma_0 = 0.15$, we get the following table:

<table>
<thead>
<tr>
<th>$p$</th>
<th>$R_{\text{CoC}}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>6.1</td>
</tr>
<tr>
<td>0.95</td>
<td>6.6</td>
</tr>
<tr>
<td>0.99</td>
<td>5.2</td>
</tr>
<tr>
<td>0.995</td>
<td>4.8</td>
</tr>
</tbody>
</table>
For the parameters $\mu_0 = 0.1, \sigma = 0.1$, and $p = 0.99$, when varying the parameter $\gamma_0$, we get the following table:

<table>
<thead>
<tr>
<th>$\gamma_0$</th>
<th>$R_{CoC}$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.6</td>
</tr>
<tr>
<td>0.10</td>
<td>3.4</td>
</tr>
<tr>
<td>0.15</td>
<td>5.2</td>
</tr>
<tr>
<td>0.20</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Thus, these values are rather similar between the VaR and the TVaR case. Note that we use identical parameters in the two cases, and we only vary the security level from 0.995 in the VaR case to 0.99 in the TVaR case.

**Example 4.5** (Pareto distribution, VaR). Assume now that $Y$ is Pareto distributed with threshold $y_0 > 0$ and tail parameter $\alpha > 1$. We then have the pure risk premium $\mu = \mathbb{E}[Y] = y_0 \frac{\alpha}{\alpha - 1}$. We choose security level $p \in (1 - (1 - 1/\alpha)^\alpha, 1)$ to ensure for the VaR that $F^{-1}(p) = C = y_0(1 - p)^{-1/\alpha} > \mu$. Let us choose the set $\mathcal{M}$ of test probabilities as

$$M := \{Q_\gamma||\gamma|\leq\gamma_0\}, \quad \text{with}$$

$$Q_\gamma(dy) := 1_{[y_0, \infty)}(y) \frac{(1 + \gamma)\alpha y_0^{(1+\gamma)\alpha}}{y^{(1+\gamma)\alpha+1}}dy,$$

and $0 < \gamma_0 < 1 - 1/\alpha$, that is, the distortion here is with respect to the tail parameter. Straightforward calculations lead to

$$\sup_{Q \in \mathcal{M}} \mathbb{E}_Q[Y \wedge F^{-1}(p)] = \sup_{\gamma|\gamma|\leq\gamma_0} \frac{y_0}{1 + \gamma} (1 + \gamma)\alpha - (1 - p)^{1-1/(\alpha(1+\gamma))}).$$

The latter expression turns out to be decreasing in $\gamma$ in the specified range, so that the maximum is attained for $\gamma = -\gamma_0$. Correspondingly, the economic premium $P$ is given by

$$P = \frac{y_0}{1 - \gamma_0} (1 - (1 - \gamma_0)\alpha - (1 - p)^{1-1/(\alpha(1-\gamma_0))}),$$

the risk margin is given by

$$RM = \frac{y_0}{1 - \gamma_0} (1 - (1 - \gamma_0)\alpha - (1 - p)^{1-1/(\alpha(1-\gamma_0))}) - \frac{\alpha}{\alpha - 1} y_0,$$

and $SCR = y_0(1 - p)^{-1/\alpha} - P$. The following table illustrates the resulting CoC rates for $p = 0.995, \alpha = 2$, and $y_0 = 0.55$ (which leads to $\mu = 1.11$ as in the log-normal case, but here the variance is not finite).
Figure 2 depicts $R_{CoC}$ for this case as a function of $p$, together with $R_{CoC}^*$ (the analogous quantity from Equation 18 when own credit risk is included in the definition of $\mu$). One sees that in the heavy-tailed case the distinction between $R_{CoC}$ and $R_{CoC}^*$ is even more important than for the Gaussian case.

**Example 4.6** (Pareto distribution, TVaR). In the setting of Example 4.5, consider now $\rho(Y) = TVaR_p(Y)$. Using the fact that conditional on being larger than $VaR_p(Y)$, the random variable $Y$ is Pareto distributed with tail parameter $\alpha$ and new threshold $y_1 = VaR_p(Y) = y_0(1 - p)^{-1/\alpha}$, one immediately arrives at

$$C = \rho(Y) = \frac{\alpha}{\alpha - 1} \frac{y_0}{(1 - p)^{1/\alpha}},$$

which is larger than $\mu$ for every $p \in (0, 1)$. One obtains

$$\sup_{Q \in \mathcal{M}} E_Q[Y \wedge C] = \sup_{y \leq y_0} \frac{y_0 \alpha_y}{(\alpha_y - 1)^2 (\alpha_y - 1 - (1 - p)^{1-1/\alpha_y} (1 - 1/\alpha_y)^{\alpha_y})},$$

with $\alpha_y = (1 + \gamma)\alpha$. Since the above expression is decreasing in $\alpha_y$, we get

**FIGURE 2** $R_{CoC}$ (below) and $R_{CoC}^*$ (above) as a function of $p$ for the Pareto case with VaR and $\gamma_0 = 0.2$ as a function of $p$. CoC, cost-of-capital; VaR, Value-at-Risk. [Color figure can be viewed at wileyonlinelibrary.com]
\[ P = \frac{y_0(1-y_0)\alpha}{((1-y_0)\alpha - 1)^2} ((1-y_0)\alpha - 1 - (1-p)^{1-1/((1-y_0)\alpha)}) \]

\[ (1 - 1/((1-y_0)\alpha))^{(1-y_0)\alpha}) , \]

together with \( RM = P - \frac{\alpha}{\alpha - 1}y_0 \) and \( SCR = \frac{\alpha}{\alpha - 1(y_0 - 1)^{1/\alpha}} - P. \)

For \( \alpha = 2, y_0 = 0.55, \) and \( p = 0.99 \) (which is the \( p \) value of interest in the Swiss Solvency Test), this leads to a CoC rate as given in the following table:

<table>
<thead>
<tr>
<th>( y_0 )</th>
<th>( R_{CoC}(%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.92</td>
</tr>
<tr>
<td>0.15</td>
<td>1.77</td>
</tr>
<tr>
<td>0.20</td>
<td>2.86</td>
</tr>
<tr>
<td>0.25</td>
<td>4.36</td>
</tr>
<tr>
<td>0.30</td>
<td>6.46</td>
</tr>
</tbody>
</table>

We observe that in the Pareto case the TVaR on the security level 0.99 is less conservative than the VaR case on the security level 0.995. In fact, the differences are much bigger than in the log-normal case.

Figure 3 depicts the situation of Figure 2 for the case of TVaR as a function of \( p \), which shows less sensitivity with respect to \( p \) and similar absolute values of \( R_{CoC} \) in the region of interest.

5 | CONCLUSION

To counterbalance risk in the risk-based approach for insurance regulation, a high level of capital above the “best-estimate” is required by the regulator and provided by shareholders. The price for this amount of capital eventually has to be paid by the policyholders, and the
CoC rate is this price per money unit. Thus the CoC has to be interpreted in an equilibrium between shareholders, policyholders, and the regulator. While the implicit limited-liability option in current regulatory regimes has been widely discussed (see, e.g., Filipovic et al., 2015), the above triangular constellation seems not having been made explicit so far, and this is what we pursue in this paper.

We deliberately keep the underlying model simple, to facilitate a transparent discussion of the approach. This includes the restriction to a one-period model, a focus on the liability aspect only, and for the valuation of the illiquid liabilities we rely on the approach of bid and ask prices in incomplete markets from Madan and Cherny (2010) and Madan and Schoutens (2010); for alternatives, see, for instance, Barigou et al. (2021) and Dhaene et al. (2017). Under these assumptions, the perspective proposed in this paper leads to a way to challenge and/or justify explicit specifications of this rate, like the often used 6% (see, e.g., the specifications in the IAA position paper; International Actuarial Association [IAA], 2009, p. 79), or the 4% often assumed in practice in recent years (according to personal communications with some practitioners).

For the numerical examples some assumptions have to be imposed, especially concerning the degree of risk aversion of the shareholders. A sensitivity analysis of the chosen quantities will lead to modified figures in an analogous way. Inversely, every concrete use of a CoC rate can receive a specific calibration and corresponding economic interpretation within the preferred model setup. We worked out resulting CoC rates for various underlying distributions and regulatory risk measures for varying security levels, providing particular justifications for the magnitudes of these rates.

In some situations negative CoC rates may appear (reminding of negative interest rates). As discussed in the paper, these might occur for several reasons, but can in particular point out an insufficient requirement from the regulator, and a corresponding potential for abuse of the shareholders’ limited-liability.

For future work, it can be interesting to extend the approach introduced in this paper to the situation of run-off patterns across several accounting years and multiperiod models in general (where other pricing methods are frequently benchmarked to the CoC approach, see, e.g., Zeddouk & Devolder, 2019). Finally, a more detailed joint view on the asset and liability side of the solvency balance sheet of an insurance company can provide a further customized interpretation of a concrete value of the CoC rate for a particular purpose.

ACKNOWLEDGMENT

Open access funding provided by INSTITUTION.

REFERENCES


