Equilibrium Risk Pools in a Regulated Market With Costly Capital

Stephen J. Mildenhall

Created 2020-10-21 19:07:32.129465

Abstract

This paper investigates equilibrium risk pools in a market with risk-based solvency regulation and costly capital. It considers a market with two classes of risk, each having different aggregate volatility characteristics, such as personal auto and catastrophe exposed property. It identifies three possible equilibrium solutions: a single multiline pool, a multiline pool and a monoline pool, and two monoline pools. It determines conditions under which each of the three solutions occurs. The requirements are sensitive to the relative tail risk of the two classes and the capital standard. The model shows that the more volatile risk class bears a higher proportion of the capital cost. The results help explain various features seen in insurance markets, including the structure of the Florida homeowners market and the US medical malpractice market. It can be applied more broadly to any regulated risk market.

JEL Codes: G22, G10

1 Introduction

Risk is better shared. The famous mutuality principle states that diversifiable risk should be removed by pooling. Remaining non-diversifiable risk is borne by each agent in proportion to their share of aggregate risk tolerance, Eeckhoudt, Gollier, and Schlesinger (2011). These results produce a world where risk-averse agents optimally pool all their risk.

In the real world, there is generally no way to pool risks in the manner envisioned by theory, Arrow (1996). Practical insurance involves limited liability insurance pools whose contractual (promised) payouts are state-independent. The pools must attract capital in a competitive market where it has a non-zero opportunity cost. As a result, risk pooling is expensive and has a cost above actuarial estimates.

Costly capital and the benefits of diversification lead us to expect that a few large multiline companies should dominate the insurance market. Indeed, large multiline carriers do write a substantial proportion of business. But looking more closely, we see many deviations from the pooling-is-best hypothesis. There is incomplete pooling between personal lines and commercial lines. Commercial, which is a more volatile business, having higher limits and a more severe pricing cycle, tends to be written by smaller companies. And over 2,600 property-casualty insurers operate in the US. Monoline companies write a disproportionate share of volatile business, such as Florida homeowners business, California earthquake, and medical malpractice[1]. What explains this structure?

Cost-based factors, economies or dis-economies of scale, and management span and control provide one explanation, Cummins et al. (2010). These are undoubtedly important. But, since this is a question about risk, it should have a risk-based answer and that is what this paper seeks to provide.

In recent work, Ibragimov, Jaffee, and Walden (2018) state that basic structure questions in a risk market with one-sided protection remain unanswered. Working in a perfect market with frictional costs of capital, they show that monoline solutions are more likely when risks are asymmetric or correlated and that multiline pooling solutions are more likely for well-behaved and independent risks. This paper comes to qualitatively similar results but uses a entirely different market model.

We show that when two risk classes buy insurance from capital-regulated, limited liability insurance pools facing costly capital, there are three possible market outcomes. There can be one insurer pooling all risks, two monoline insurers, or the insureds are split between one pool insurer and one monoline. One pool is the usual solution. Two monoline pools occur when the regulator capital function is super-additive, generally when insurance risk is very thick-tailed, Ibragimov and Walden (2007). In the split case, the higher volatility class splits between the pool and the monoline, and the lower risk class appropriates all of the diversification benefits. The higher risk class pays its monoline rate. This result is likely counter to insurance affordability and availability goals for the high risk class. These effects are more pronounced for weaker capital regulation and cheaper insurance.

The market structure solutions emerge from the interaction between the investor's view of risk and the regulator's. Even when the investor and the regulator risk measures are both coherent, their combination, which determines market prices, can fail to be subadditive. This failure results in monoline solutions. Our results extend Ibragimov, Jaffee, and Walden (2010), which had the critical insight that there is essentially only one marginal cost capital allocation consistent with fair pricing. Their marginal cost view becomes our marginal regulatory capital, and we replace their perfect market pricing model with an imperfect model.

A primary goal for this work is to answer market structure questions in a framework consistent with the real insurance market. Our title flags the two realities with which we are most concerned: regulation and costly capital. Regulation enters because we assume a mandatory insurance requirement. We assume capital is costly because investors are ambiguity averse. As a result, pricing is governed by a non-additive pricing rule. The rationale is laid out in section 2.

An extensive literature considers optimal risk-sharing between two risk classes. In contrast, we assume that insurance transfers risk to a separate investor group, independent of the

[1] Monoline companies write US mortgage guaranty for different legal reasons.

risk owners, what Ibragimov, Jaffee, and Walden (2018) call one-sided protection. We ask whether it is most efficient for all insureds to combine into one multiline pool when purchasing one-sided protection. Broadly, the global insurance market provides one-sided protection rather than inter-insured pooling. For example, Canadian pension funds are significant investors in US catastrophe insurance-linked securities: investors form a distinct group from risk owners.

We can consider risk pooling as a reinsurance question. In our model, insurance pools are transparent pass-through entities, but they have a legitimate interest in the shape of risk because it drives their cost of capital. Therefore, there is a potential role for reinsurance. We find that reinsurance enables a pooled premium rate rather than pooling risk. If it were optimal to pool the risks, then they would be pooled directly in the first place. Differential premium rates can be an impediment to this pooling that reinsurance can help overcome. Reinsurance does not lead to a new equilibrium solution because the possibility of direct contracting makes reinsurance pools unstable and, absent some enforcement mechanism, they unravel. Surprisingly pools with reinsurance have less risk pooling than those without it.

This paper offers the following contributions.

We contribute to the problem of the equilibrium industry structure. We show that a market with two classes of risk will consist of one multiline carrier, two monolines, or a multiline and a monoline. This last split result generally obtains when one class is more risky than the other. In that case, the less risky class appropriates all the diversification benefits. These results depend on the capital standard and highlight that regulation can have unintended consequences.

We provide new insight into how limited liability and diversification interact. Limited liability produces subsidies between lines, which offset the benefits of diversification. These effects are more pronounced when the capital standard is weaker. In a dynamic model, they can introduce subsidies that disadvantage low-risk insureds. Regulation should consider these impacts when calibrating a capital standard.

Our methods can help insurers answer strategic questions related to geographic or class expansion. Regulation introduces a rigidity in rates that can discourage diversification and lock-in subsidies between classes, and our analysis makes it clear where and how such rigidities will manifest themselves.

Finally, we show that it is practical to work in a realistic, imperfect market and obtain explicit premium, loss, and capital allocation results.

The rest of the paper is structured as follows. Section 2 describes the market participants and their interactions. Section 3 recalls results that determine the fair market value of insurance cash flows for each class in a multiple-class insurance pool. Section 4 derives the market outcomes that can occur. Section 5 gives some examples, and Section 6 concludes and suggests further research.

1.1 Notation and Conventions

The terminology describing risk measures is standard, and follows Föllmer and Schied (2011). We work on a standard probability space, Svindland (2009). The remaining notation is consistent with Major and Mildenhall (2020).

Total insured loss, or total risk, is described by a random variable X. X reflects policy limits but is not limited by insurer assets. $X = \sum_i X_i$ describes the split of losses by class. F, S, f, and q are the distribution, survival, density, and quantile function of X. Subscripts are used to clarify the random variable. $X \wedge a$ denotes $\min(X, a)$ and $X^+ = \max(X, 0)$. 1_A is the indicator function on a set A.

We use the actuarial sign convention: losses are positive and large positive values are bad.

2 Market Participants and Market Assumption

We work in a market with four participants: insureds, a regulator, insurance pools and investors, as shown in Figure 1.

We assume a one period model, with no expenses, no taxes, and a zero risk-free rate of interest. These are standard simplifying assumptions, e.g. Ibragimov, Jaffee, and Walden (2010).

At time $t = 0$ insureds form into limited liability insurance pools. The policies issued to pool participants aggregate to a total exposure $X = \sum_i X_i$. The regulator capital measure determines the amount of assets the pool needs to hold, $a = a(X)$. The pool raises a from a combination of premium and by selling its $t = 1$ residual value to investors to raise equity. At $t = 1$ claims become known and are paid. If losses $X \le a$ all insureds are paid in full and the residual value $a - X$ is distributed to investors. If $X > a$ the pool defaults and pays insureds on an equal priority, pro rata basis. The investors receive nothing. In both cases the pool is wound-up at $t = 1$. There are no transactions between 0 and 1 and hence no distinction between capital and assets.

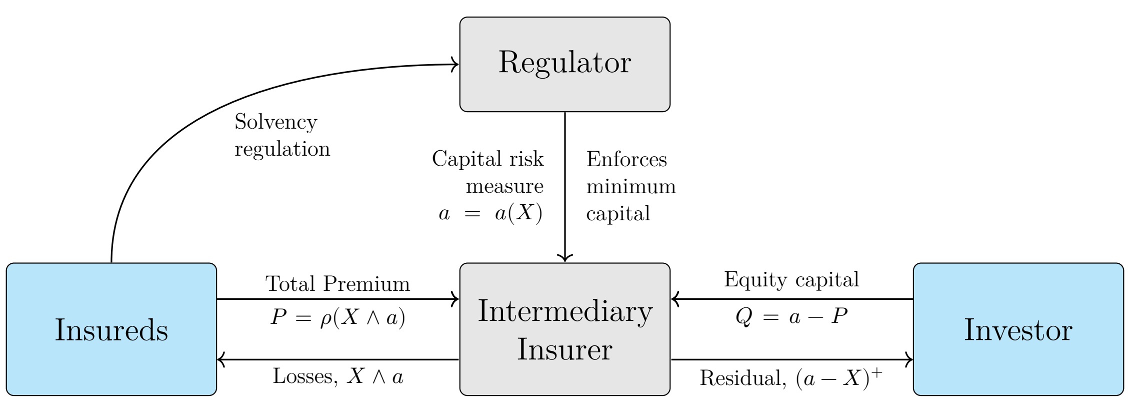


Figure 1: The four market actors and their interactions.

The investor prices with a functional ρ. The insurer purchases a quota share, with limit a, from the investors for a premium

$$\rho^a(X) := \rho(X \wedge a(X)). \tag{1}$$

Thus the market pricing functional combines investor and regulator risk functionals in an intricate and subtle way. All the effects we discuss are related to the properties of ρ^a. It is easy to see that if both a and ρ are positive homogeneous, monotone, translation invariant, law invariant, or comonotonic additive then ρ^a will have the same property too. But, critically, ρ^a can fail to be subadditive even when both a and ρ are subadditive. It is this fact that makes the pooling problem interesting and difficult.

We now specify the assumptions and behavior of each actor, starting with buying motivation.

2.1 Compulsory Insurance

Our model assumes that insurance is compulsory and is bought for the protection of others.

Aon Benfield (2015) estimate that around 60 percent of global property-casualty premium is compulsory or quasi-compulsory. Broadly interpreted, compulsory insurance includes auto liability, much property insurance providing collateral protection, contractual general liability and surety, and workers compensation for employers, amongst other lines. Flood insurance, but curiously not earthquake, is required to obtain a mortgage in the US on an at-risk property. We use compulsory as a catch-phrase to include any legal, contractual, or doing-business requirements.

Since insurance is bought for the protection of others, the insured has no interest in its quality, and so solvency regulation is necessary to ensure the insurance requirement functions effectively, Cummins (1988). As predicted by theory, compulsory insurance laws usually require insurance be provided by a licensed insurer, operating under risk-based capital regulation. US conforming mortgages require property insurance from an insurer with an adequate rating from a nationally recognized statistical rating organization, for example.

Insureds have no interest in the quality of their insurance either because they are judgment proof or because residual claims are covered by a guaranty fund or both. As a result, insureds are pure price buyers, concerned only that the insurance satisfies their compulsory insurance requirement.

2.2 Insureds

There are two classes of insureds. Throughout, class 0 is a lower risk class and class 1 higher risk. Risk is relative. Insureds within a class can be considered identical, with the same insurance contract, the particulars of which areirrelevant. As a result, adverse selection is not a problem.

We are only concerned with class risk and not with individual insured characteristics within the class. Correlations and dependencies between insureds determine class risk. It cannot be discerned insured by insured. Florida homeowners insurance is a high volatility class. Low limit, high frequency US-style personal automobile is a low volatility class. An *individual*

Florida homeowners policy may have the same range of outcomes as a home in France or Germany or Illinois. Still, the risk of a *portfolio* of Florida homes will be higher because of the possibility of hurricane events.

Aggregate losses are homogeneous within each class, meaning that losses from a proportion x_i of class i has distribution $x_i X_i$ for fixed distributions X_i, $i = 0, 1$. A homogeneous loss model is a common approximation borrowed from finance, where x_i is a position size and X_i is a security price. It is used by Myers and Read Jr. (2001). X_i can be considered as the loss ratio for the class. Boonen, Tsanakas, and Wüthrich (2017) compares the homogeneous assumption to a more realistic compound Poisson model, and Mildenhall (2017) shows that a homogeneous assumption is not unrealistic for larger portfolios.

All market participants agree best-estimate subjective probabilities that underlie X_i. They are recognized as ambiguous and uncertain. Commercial catastrophe models provide such probabilities for the property catastrophe market.

Individual risk aversion is not assumed in the model and is not necessary for the results.

2.3 Regulation

The solvency of insurance pools must regulated for compulsory insurance to be effective and provide the desired third-party protection.

Regulation takes the form of a regulatory capital risk measure a. We assume that a is law invariant, positive homogeneous, monotone, and translation invariant. In most cases a is given by value at risk (VaR) or tail value at risk (TVaR). US NAIC RBC and Solvency II are VaR-based; the Swiss Solvency Test is TVaR-based.

The regulatory risk measure is not assumed to be subadditive, and therefore may not be coherent.

There is no other regulation in the market. Rate regulation is not needed because we assume actors are well-informed. And there is no restriction on insurance pools, other than that they meet the capital requirement. A pool can consist of a single risk, or one pool could contain all risks, or any subset in-between. This flexibility means insureds have option of being written in a pool of one, which sets an upper bound on the premium they will pay.

2.4 Insurer Intermediaries

Insurer intermediaries are transparent pass-through entities that provide compulsory insurance meeting the regulator's capital standard. They manage a pool of insureds and arrange for the risk to be transferred to investors. They are like a smart-contract and operate costlessly. They have no employees or management, and so there are no principle-agent problems.

A *pool* refers to a group of insureds. There are two types of pool. A monoline pool contains insureds from one class. A multiline pool contains insureds from more than one class.

An *insurance* pool is a limited liability insurance company. In theory, pools could contract directly with insurers and obtain unlimited cover. This will not occur in our model because insureds do not consider insurance quality, only price.

Although insurers are incorporeal, they have a real economic impact because they enable limited liability, which changes the cash flows paid to insureds. Limited liability is vital to ambiguity averse investors: it truncates the tail, and it lowers the ambiguity of the losses they assume.

We could argue that since investors ultimately bear all the risk they should be indifferent to how it is packaged into pools. This is not correct. The pools alter the actual payments made by investors because they impact the incidence of default. A multiline insurer structure will default in different states than two monoline insurers.

Limited liability requires a legal framework to function; this is why the market uses insurer legal entities rather than third-party pool managers who only act as brokers between insureds and investors but assume no risk. The limited liability contract binds the participants to equal priority in default. Marshall (2018) provides an illuminating discussion of the importance of a legal entity to implement limited liability in the context of the California Earthquake Authority.

Since it is so important, we now describe in detail how equal priority in limited liability functions. It shares assets in proportion to unlimited claims. Individual insurance contracts promise to pay X_i to class i. In the aggregate, promised payments are split into realized payments and insurer default as

$$X = X \wedge a + (X - a)^+,$$

where a equals insurer assets. Multiplying both sides by X_i/X shows

$$\begin{aligned} X_i &= X_i \frac{X \wedge a}{X} + X_i \frac{(X-a)^+}{X} \\ &= (\text{payments to class } i) + (\text{default to class } i). \end{aligned}$$

The payments to class i under **equal priority** are defined as

$$X_i(a) := X_i \frac{X \wedge a}{X} = \begin{cases} X_i & X \le a \\ a\dfrac{X_i}{X} & X > a. \end{cases} \tag{2}$$

Under equal priority class i is paid in full when total losses are less than or equal to assets and payments are pro-rated down by X_i/X in default states. These payments are contractual and are legally specified. Note that $\sum_i X_i(a) = X \wedge a$.

Contractually specified payments to i are a function of X_i alone. These are the amounts promised to i by their insurance contract. $X_i(a)$ is the amount actually paid to i; it is a function of losses for all classes, $X_1, \dots, X_n$ as well as the total assets a held by the insurer. This distinction between promised and delivered payments is critical. It is responsible for almost all the complexity of insurance pricing. In economics, the benefit of a good to the buyer is usually independent of who else purchases it. With insurance risk pools that is not the case: the value to one insured is almost always changed in some way if the pool takes on other risks. To completely specify the payments to i it is necessary to know about possible payments on all other contracts. These facts have profound implications for pooling.

The insurance market is competitive. Since actors agree on probabilities, insureds only pay the fair price for insurance, and neither pools nor insurers make a profit.

2.5 Investor Risk Bearers

Investors bear insurance risk. Risk is transferred from the insureds to a distinct group of investors providing one-sided protection. It is not being pooled within insureds.

The theory of multiline pricing typically assumes a perfect complete market, but with a frictional cost of holding capital in an insurance company, Phillips, Cummins, and Allen (1998), Myers and Read Jr. (2001), Sherris (2006), Ibragimov, Jaffee, and Walden (2010), and Cummins (2000). These assumptions imply that diversifiable insurance risk is not priced. The catastrophe bond market is inconsistent with these implications. A catastrophe bond is deliberately structured to minimize frictional costs: it has no management, it is domiciled in a tax-free jurisdiction, and is very lightly regulated. And yet catastrophe bonds are typically priced at two to ten times best-estimate losses (see www.artemis.bm for examples of catastrophe bond pricing). Investors charge for bearing risk beyond the cost of holding capital in insurance entities.

Which risks do investors price? All theories price systematic, non-diversifiable risk. And while hurricane or earthquake are conceivably systematic, the idea becomes less credible for minor perils covered by catastrophe bonds at substantial margins. On the other hand, it is hard to argue that investors price pure roulette lottery risk since the mechanisms to pool and manage it are so evident. But this is a false dichotomy. Insurance is not a roulette lottery. In the phrase of Anscombe and Aumann (1963) it is an uncertain horse lottery, governed by unknown and ambiguous subjective probabilities rather than unique objective ones. There is extensive evidence that investors are ambiguity averse, Ellsberg (1961). Underwriters consider a one in one hundred year event risky primarily because they do not have hundreds of years of experience. Its objective probability is uncertain. Data is more abundant for more frequent events, and uncertainty is lower. Zhang (2002) and Klibanoff, Marinacci, and Mukerji (2005) describe ambiguity relevant to insurance pricing. The latter paper has been applied in an insurance context by Robert and Therond (2014), Dietz and Walker (2017) and Jiang, Escobar-Anel, and Ren (2020). Note that it is hard to distinguish risk aversion from ambiguity aversion because risky thick-tailed distributions are also more ambiguous.

We assume that the insurance market is imperfect. Investors are ambiguity averse but not necessarily risk-averse. Market prices incorporate investor ambiguity aversion using a non-additive pricing functional. Wang (1996) and Wang, Young, and Panjer (1997) apply a non-additive functional using distorted probabilities to insurance pricing, leveraging diverse theoretical underpinnings including Huber (1981), Schmeidler (1986), Schmeidler (1989), Yaari (1987), and Denneberg (1994). Within this theory, distortion risk measures (DRM) occur repeatedly and in many guises. Kusuoka (2001) characterize DRMs as coherent, law invariant, and comonotonic additive functionals. We assume investors price using a DRM. DRMs are also known as spectral risk measures, Acerbi (2002), because of their weighted-VaR representation.

DRMs have many appealing properties. However, they are not additive and include transaction costs, via an implied bid-ask spread, Castagnoli, Maccheroni, and Marinacci (2004). Important results from Chateauneuf, Kast, and Lapied (1996), De Waegenaere (2000), and especially Castagnoli, Maccheroni, and Marinacci (2002) and De Waegenaere, Kast, and Lapied (2003), show DRMs are consistent with general equilibrium models, even though they are non-additive. DRMs are consistent with arbitrage-free pricing: the presence of transaction costs neutralizes apparent arbitrage opportunities caused by non-additivity.

DRMs are practical and easy to work with. Most importantly, they have a natural allocation of a pool's diversification benefit back to its members, which we outline in Section 3. The natural allocation relies on the fact that DRMs are coherent, law invariant and comonotonic additive.

DRMs respect diversification because they are coherent and hence subadditive. As a result, pooling two risks with no default will always be advantageous, provided both premiums are finite. Therefore the failure of pooling must rely on something more than a DRM pricing functional. The extra ingredient is the interaction between the pricing functional and solvency regulation.

Finally, we assume there are are no frictional costs for holding capital in an insurance entity. Cummins (2000) explains that agency conflict, tax, and regulation are the primary causes of frictional costs. Our model is structured to remove these costs: there are no taxes, there is no agency conflict because insurers operate like an autonomous contract with no managerial discretion once they have been set up (like a catastrophe bond), and regulation is limited to solvency requirements. There is no economic reason why investors have to transfer assets into the insurer entity at all; it could operate like an old-style Lloyds syndicate provided the regulator had sufficiently vicious rights to enforce payment.

Castagnoli, Maccheroni, and Marinacci (2004) shows DRMs always have a bid-ask spread. We do not count the spread as an expense because insurance positions are always long. Insurable interest laws make it impossible to short insurance.

We assume that a competitive market produces prices that are modeled by a DRM. Kusuoka (2001) and Föllmer and Schied (2011, chap. 4) show that a DRM is entirely characterized by a distortion function. A distortion function is an increasing concave function $g : [0,1] \to [0,1]$ satisfying $g(0) = 0$ and $g(1) = 1$. The DRM ρ_g associated with a distortion g acts on a non-negative random variable X as

$$\rho_g(X) := \int_0^\infty g(S(x))dx. \tag{3}$$

When ρ is combined with a regulatory capital function using eq. (1) we get the market pricing functional

$$\rho_g^a(X) := \rho_g(X \wedge a(X)). \tag{4}$$

Thus $\rho_g^a(X)$ is the premium charged for the insurance pool with total risk X. Since the survival functions S_X and $S_{X\wedge a}$ agree on $[0, a)$, $S_{X\wedge a}(x) = 0$ for $x > a$ and $g(0) = 0$, we get

$$\rho_g^a(X) = \int_0^\infty g(S_{X\wedge a(X)}(x))dx = \int_0^{a(X)} g(S_X(x))dx. \tag{5}$$

We drop the subscript g from the notation below since it is fixed.

Providing a clear motivation for the existence of insurance pools is a significant advantage of our framework Under perfect markets, there is no diversification benefit because the pricing rule is linear and hence additive, which means they do not allow a diversification benefit. As a result, there is no need for a pool. Justifying insurers usually involves an appeal to market access or transaction expenses, Ibragimov, Jaffee, and Walden (2010) (especially footnote 9 and surrounding discussion). Insurers in our model exist to allow limited liability and to economize on the use of costly capital.

2.6 Market Assumption Summary

Here is a summary of our insurance market assumptions.

1. Insureds are required to purchase insurance. They are pure price buyers.
2. Insurance must be purchased from a one period, limited liability insurance pool. Pools can consist of one or more policies. There is no restriction on the size or composition of pools. Single policy pools are allowed.
3. Insurance pools are required by regulation to capitalize according to a risk-based capital formula. It is law invariant, positive homogeneous, monotonic, translation invariant, but not necessarily subadditive. It is taken to be p-VaR for p close to 1. VaR is the most common measure used in practice.
4. Insurance losses are homogeneous within each pool.
5. There is equal priority in claim payments if the pool defaults because it has insufficient assets to pay its obligations.
6. Investors provide equity to insurance pools by buying its residual value. Investors price using a DRM ρ associated with a distortion g. Equation (5) computes the pool premium.
7. Insurance pools are costless to form and operate, other than the cost of risk transfer to investors. There are no taxes.
8. All market participants agree on a set of subjective probabilities. Insurance pools are transparent, and individuals know the fair market value of their insured claims. The market is perfectly competitive and acts to remove any excess profits.

Under these assumptions, we now investigate what types of equilibrium insurance pools will form. To do that, we start by deriving the fair market value of pool insurance to pool participants.

3 Market Pricing and the Natural Allocation

This subsection recalls the main results from Major and Mildenhall (2020), which gives a natural allocation of pool premium, under the assumptions set out in section 2. The natural allocation transfers perfect market pricing theory to a non-additive, imperfect market setting.

A perfect market pricing functional has the form $\mathsf{E}_{\mathsf{Q}}[\cdot]$, where Q is a risk-adjusted measure. The Radon Nikodym derivative of Q defines the state price density. Sherris (2006) and Ibragimov, Jaffee, and Walden (2010) show that the fair premium for risk i is simply $\mathsf{E}_{\mathsf{Q}}[X_i(a)]$. It is an allocation of total premium $\mathsf{E}_{\mathsf{Q}}[X \wedge a]$ because expectations are linear.

Under a DRM, the state price density Q becomes a function of the risk being priced; it is the measure solving $\rho(X) = \mathsf{E}_{\mathsf{Q}}[X]$, subject to certain other conditions. But we can still use the perfect market approach to allocate premium, even though Q is no longer fixed. Specifically, if $X = \sum_i X_i$ then the fair premium for X_i is $\mathsf{E}_{\mathsf{Q}}[X_i]$. We call this the natural allocation premium and use the notation $\rho_X(X_i)$. We will apply the natural allocation to $X \wedge a = \sum_i X_i(a)$. Since DRMs are comonotonic additive, and since X, $X \wedge a$ and $(X-a)^+$ are all comonotonic, it follows that $\rho(X) = \rho(X \wedge a) + \rho(X-a)^+)$ and hence that $\rho(X \wedge a) = \mathsf{E}_{\mathsf{Q}}[X \wedge a]$ for the same measure Q solving $\rho(X) = \mathsf{E}_{\mathsf{Q}}[X]$, Shapiro (2012). Thus the natural allocation to $X_i(a)$ is given by $\mathsf{E}_{\mathsf{Q}}[X_i(a)] = \mathsf{E}_{\mathsf{Q}}[X_i(X \wedge a)/X]$.

Expected losses for class i, accounting for possible default, can be computed by conditioning on X as

$$\mathsf{E}[X_i(a)] = \mathsf{E}[\mathsf{E}[X_i(a) \mid X]] = \mathsf{E}[X_i \mid X \le a]F(a) + a\mathsf{E}[X_i/X \mid X > a]S(a). \tag{6}$$

Total premium under ρ is given by eq. (1). We now compute the natural allocation premium and premium density for each class. Using integration by parts the price of an unlimited cover on X is

$$\rho(X) = \int_0^\infty g(S(x))\,dx = \int_0^\infty x g'(S(x))f(x)\,dx = \mathsf{E}[Xg'(S(X)))]. \tag{7}$$

It is important that this integral is over all $x \ge 0$ so the $xg(S(x))|_0^a$ term disappears. This shows that $Z = g'(S(X))$ is the Radon Nikodym derivative of the measure Q discussed above. Thus the natural allocation premium is $\mathsf{E}[X_i g'(S(X))]$. The choice $g'(S(X))$ is economically meaningful because it weights the largest outcomes of X the most, which is appropriate from a social, regulatory and investor perspective. Since DRMs are comonotonic additive the same density can be used for $\rho(X \wedge a)$. This will not be true for non-comonotonic additive risk measures.

Applying the same argument to $X \wedge a$ gives the following, essentially unique, expression for the natural allocation of eq. (1)

$$\rho_{X\wedge a}(X_i(a)) = \mathsf{E}_{\mathsf{Q}}[X_i \mid X \le a](1 - g(S(a))) + a\mathsf{E}_{\mathsf{Q}}[X_i/X \mid X > a]g(S(a)) \tag{8}$$

$$= \mathsf{E}[X_i 1_{\{X_t \le a_t\}} g' S_t(X_t)] + a_t \mathsf{E}\left[\frac{X_i}{X_t} 1_{\{X_t > a_t\}} g' S_t(X_t)\right]. \tag{9}$$

The premium formula eq. (8) only assumes that capital is provided at a cost g and there is equal priority by class. There is no need to assume the X_i are independent. The formula is computationally tractable.

As a result of the general theory, we have the following behaviors.

1. In total, the margin is always non-negative for all asset levels.
2. For full coverage, $a = \infty$, the natural allocation premium is always less than the monoline (stand-alone) premium and contains a non-negative margin.

See Major and Mildenhall (2020) for a complete derivation of these claims.

4 Market Outcomes

This section proves the main market structure results of the paper. We show how premium varies as a function of the pool mix between two classes with a homogeneous loss model. Then, we use these results to characterize all possible market structures.

The eight assumptions from section 2.6 hold throughout the section. In particular, the regulatory capital standard is given by p-VaR.

4.1 How Pricing Varies with Pool Mix

There are two classes of insured, with independent aggregate loss random variables X_0 and X_1. Losses from a homogeneous pool, with a proportion t, $0 \le t \le 1$, of class 1 risks and $1-t$ of class 0, are given by

$$X_t := (1-t)X_0 + tX_1. \tag{10}$$

To avoid trivially different equations for class 0 and 1, we will write out expressions for class 1 in full, with the understanding an analogous statement holds for class 0 on replacing t with $1-t$.

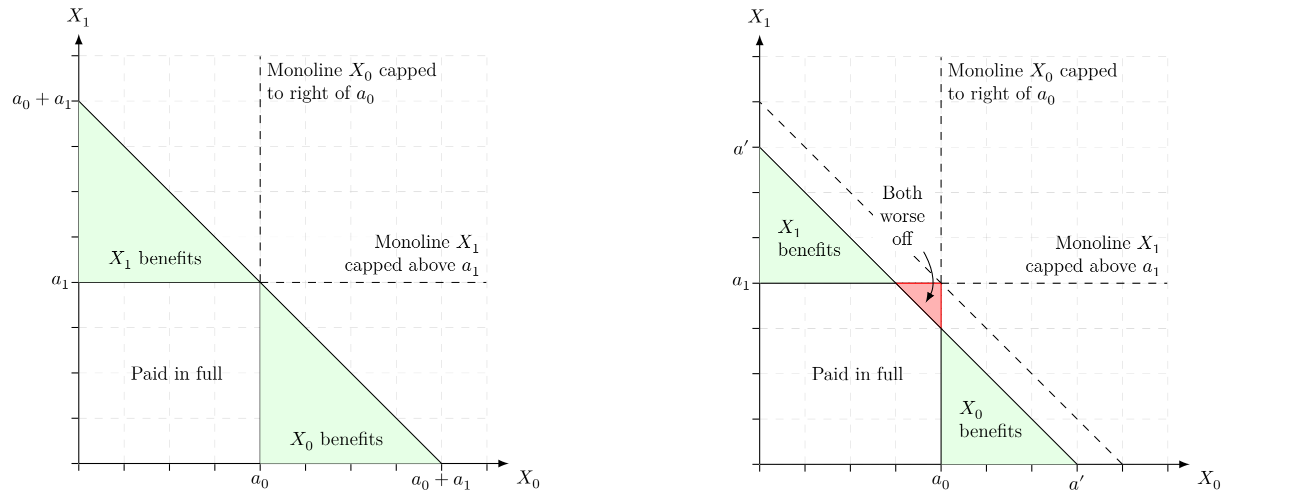
$P(t) := \rho(X_t \wedge a_t)$ denotes the premium for the pool X_t with capital $a_t := \mathsf{VaR}_p(X_t)$. The natural premium allocation to class 1 is

$$P_1(t) := \rho_{X_t \wedge a_t}((tX_1)(a_t)) = \mathsf{E}\left[tX_1 \, \frac{X_t \wedge a(X_t)}{X_t} \, g'(S_t(X_t)) \right]. \tag{11}$$

Clearly $P(t) = P_0(t) + P_1(t)$. The expression for $P_1(t)$ combines three elements: class losses, the common default pro rata factor, and the probability adjustment reflecting investor ambiguity aversion. The corresponding premium rate per unit of exposure is

$$R_1(t) := \frac{P_1(t)}{t}$$

for $t \in (0,1)$. Clearly $P(t) = (1-t)R_0(t) + tR_1(t)$. Recall that $g'(S_t(X_t))$ is the Radon Nikodym derivative of a measure Q_t that solves $\rho(X_t) = \mathsf{E}_{\mathsf{Q}_t}[X_t]$. These equations show the rate is determined as a weighted average of *solvent* and *default* parts. When $t = 1$,



(a) Diversification benefit, with no reduction in capital.

(b) Reducing capital lowers diversification benefit.

Figure 2: A subadditive capital standard reduces the diversification benefit.

$P_1(1) = P(1)$ equals $\rho(X_1 \wedge a_1)$, the monoline premium for all of class 1, and similarly when $t = 0$, $P_0(0) = P(0)$.

We expect $P(t)$ to be non-constant because the composition and diversification of the pool varies with t. Diversification creates two offsetting effects, illustrated in fig. 2. The left figure illustrates combining two classes and adding supporting assets. Provided losses are not perfectly dependent, there will be events where one class has a large loss and the other a small loss, and where both can be paid in full from a pool while a monoline would cap the large loss (shaded green triangles). A pool with no reduction in assets is a Pareto improvement for both participants. Offsetting this unequivocal benefit, the pool, or its regulator, can use diversification to justify holding fewer assets $\max(a_0, a_1) < a' < a$. Pool insureds are still better off when one or other has a large loss, but the improvement is smaller. It is possible for this benefit to go too far, increasing risk in very adverse scenarios, Dhaene et al. (2008). With less capital, pooling is not necessarily a Pareto improvement: there are fewer assets available when both classes have a large loss simultaneously (shaded red triangle). Diversification with no reduction in assets improves the quality of insurance and we therefore expect a higher premium, but the net effect is unclear when asset are reduced.

There is a third dynamic that further complicates the picture. Since a pool will generally have more a predictable, less ambiguous outcome distribution it will be more appealing to investors. As a result, the fair premium will include a lower margin. Note this does not create an economic benefit for insureds or a more competitive product: all premiums are fair relative to the protection they provide. These off-setting factors make the impact of pooling on premium P indeterminate.

Figure 3 illustrates how premium and rate typically varies with mix. After explaining the figure, we will characterize the range behaviors more precisely. The left plot shows rate and the right plot shows both rate and premium. The horizontal axis shows pool mix t. X_0 and

X_1 are independent gamma variables with $\mathsf{E}[X_i] = 100$ and coefficient of variation 0.25 and 0.30 respectively. Both lines are thin tailed and have log concave densities. To show various behaviors more clearly, the model uses a weak 0.90 VaR capital standard and an expensive insurance, priced using a proportional hazard distortion $g(s) = s^{0.3}$.

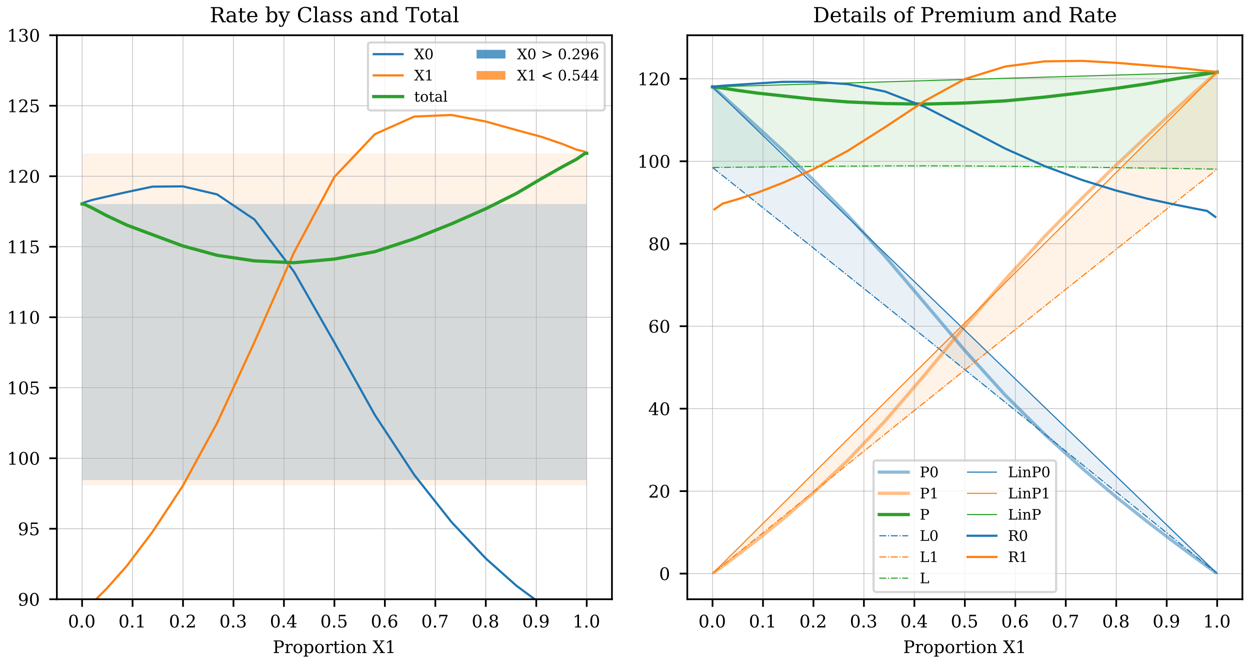


Figure 3: Typical behavior of pricing variables. Horizontal axis t determines the pool mix as $(1-t)X_0 + tX_1$.

The figure illustrates several behaviors.

- The premium rate $R_i(t)$ for each class (blue and orange) and total premium $P(t)$ (green). The three lines intersect at a point around $t = 0.4$. They will always meet because the total premium is a $(1-t, t)$ weighted average of the by-line rates. They generally do not intersect at the minimum premium.
- On the left, the horizontal shaded bands (which overlap in the grey) show the range from monoline expected loss $\mathsf{E}[X_i \wedge a_i]$ up to the monoline premium $\rho(X_i \wedge a_i)$ for each class. Because the capital standard is weak monoline expected losses are $< \mathsf{E}[X] = 100$ by limited liability. Since X_1 is more risky than X_0 it has a lower expected loss and higher monoline premium.
- $R_1(1) = P(1)$ and $R(t)$ bows up above $P(1)$ and is greater than $P(1)$ for $t > t^* = 0.544$. This reflects the fact X_1 benefits from pooling with X_0. Similar comments apply to R_0 near $t = 0$; its rate is above the monoline rate for $t < t_* = 0.296$. The critical t values are shown in the legend.
- The total premium line only gives the premium for the $(1-t, t)$ weighted pool (it is the rate for the pool); it does not give the total market premium because the blended portfolio can only include all risks in the special case $t = 0.5$. When $t \neq 0.5$ the multiline pool with include all of one class, but the other will be split and partially written in a monoline pool.

- On the right, the triangles show the premium as opposed to rate. The lines `Lin` show $tP(1)$, etc. and P show the premium.

For the purposes of understanding premiums and rates it is instructive to start by considering the case where one line is a constant. Suppose that class 0 has certain losses, $X_0 = x_0$, and hence $a_0 = x_0$ and $P(0) = R_0(0) = x_0$ (because ρ is coherent and hence translation invariant, meaning $\rho(X + c) = \rho(X) + c$ for constant c).

Starting from a monoline X_1 pool consider the impact of adding a small amount of X_0, corresponding to t decreasing from 1. Pool assets are $a_t = (1-t)x_0 + ta_1$. Both lines are paid in full in non-default states. In default states, class 1 will recover $a_t \mathsf{E}[tX_1/((1-t)x_0 + tX_1) \mid X_1 > a_1] \ge ta_1$[2] and hence class 0 will recover less than $(1-t)x_0$. In a sense, default is *caused* by class 1 having a large loss. By the action of equal priority, it recovers more than it would from a monoline pool. As a result, assets are transfered from class 0 to class 1 in default ex post. To compensate for this, class 1 must pay a higher ex ante premium. Thus $R_1(t) > P(1)$ ($R_0(t) < P(0)$) for t close to 1. The amount by which the pooled rate exceeds (falls below) the monoline rate will increase as the capital standard becomes weaker because default states become more important.

On the other hand, starting from a monoline X_0 pool, still with certain losses, consider the impact of adding a small amount of X_1, corresponding to t close to 0. For class 0 we have

$$\begin{aligned} R_0(t) &= \mathsf{E}[X_0\, g'(S_1(X_1))1_{\{X_1 \le a_1\}}] + a_t \mathsf{E}[X_0/X_t\, g'(S_1(X_1))1_{\{X_1 > a_1\}}] \\ &= x_0(1 - g(1-p)) + x_0 \mathsf{E}[\frac{(1-t)x_0 + ta_1}{(1-t)x_0 + tX_1}\, g'(S_1(X_1))1_{\{X_1 > a_1\}}] \\ &\le x_0(1 - g(1-p)) + x_0 g(1-p) \\ &= x_0 \end{aligned}$$

since $\dfrac{(1-t)x_0 + ta_1}{(1-t)x_0 + tX_1} < 1$ when $X_1 > a_1$. For class 1

$$\begin{aligned} R_1(t) &= \mathsf{E}[(X_1\, g'(S_1(X_1))1_{\{X_1 \le a_1\}}] + a_t \mathsf{E}[X_1/X_t\, g'(S_1(X_1))1_{\{X_1 > a_1\}}] \\ &\gtreqless \mathsf{E}[(X_1\, g'(S_1(X_1))1_{\{X_1 \le a_1\}}] + a_1 \mathsf{E}[X_1/X_t\, g'(S_1(X_1))1_{\{X_1 > a_1\}}] \\ &= P(1) \end{aligned}$$

according to $a_t \gtreqless a_1$, i.e., $x_0 \gtreqless a_1$. This makes sense: if $x_0 > a_1$, then by pooling with X_0 class 1 obtains access to relatively more capital per unit of loss, which improves quality and raises price. Notice that S_1 is used throughout; the ordering of outcomes is determined by X_1 alone.

Figure 3 illustrates these behaviors. For t near 1, $R_1(t) > P(1)$. In this case $x_0 < a_1$ and so $R_1(t) < P(1)$. This corresponds to the orange rate line bowing up above the level $P(1)$ for large t and then decreasing below $P(1)$ for small t. There is a similar, but less pronounced, effect for class 0 near $t = 0$.

[2] The inequality holds if $\dfrac{X_1}{a_1}\dfrac{ta_1 + (1-t)x0}{tX_1 + (1-t)x0} \ge 1$ which is true iff $X_1 > a_1$.

4.2 Basic Results

The next theorem summarizes properties of premiums and rates by class. We say a function of t is differentiable if it is differentiable on $(0,1)$, right differentiable at 0 and left differentiable at 1.

Theorem 1. *Suppose the distributions of X_i are independent with a jointly continuous density and that the distortion g is differentiable on $[0,1]$. Then*

1. *S_t, a_t, $P(t)$, α_i, and β_i are differentiable.*
2. *In the case of full coverage, $p=1$, the rate for each class lies between the expected loss and the stand-alone premium*

$$\mathsf{E}[X_1] \le R_1(t) \le \rho(X_1 \wedge a_1) \tag{12}$$

   for $t \in (0,1)$.

3. *For all p and all $t \in (0,1]$, $R_1(t) < \rho(X_1)$.*
4. *$P_1(t)$ is differentiable and $R_1(t)$ is differentiable on $(0,1]$.*
5. *If $\rho(X_i) < \infty$, then for all $p < 1$, $R_1(t) > tP(1)$ for t sufficiently close to 1 provided*

$$\frac{a_0}{\mathsf{E}[X_0]} \ge \mathsf{E}_{\mathsf{Q}_1}\left[\frac{a_1}{X_1} \mid X_1 > a_1\right]. \tag{13}$$

6. *$\lim_{t\downarrow 0} R_1(t) \ge \mathsf{E}[X_1](1-g(1-p))$.*

Comments. The assumption on g rules out $\lim_{s\downarrow 0} g(s) > 0$. Since g is concave it is differentiable almost everywhere. Since the distributions are continuous $S_t(a_t) = 1-p$ for all t. $\rho(X_1)$ is the price for unlimited cover, which is generally $> P(1)$. The first assumption is not overly restrictive because by Lusin's theorem we can approximate any measurable function closely by a continuous one. The assumption eq. (13) in Part (5) is always true for reasonable parameters because $a_0 > \mathsf{E}[X_0]$ (capital is greater than expected losses) whereas the right-hand side is manifestly < 1. Equation (13) can fail if X_0 is very thick tailed and the VaR capital is below the mean. This is the well known failure of VaR for very skewed risks. The condition also fails if $\mathsf{E}[X_0] = \infty$ and the right hand size is positive.

Proof:

1. Differentiability of S_t and a_t under these assumptions is proved in Tasche (2001). See **??** for the exact derivatives. Differentiability of P follows from

$$\begin{aligned}\frac{dP}{dt} &= \frac{d}{dt}\int_0^{a_t} g(S_t(x))\,dx \\ &= \int_0^{a_t} g'(S_t(x))\frac{dS_t}{dt}(x)\,dx + g(S_t(a_t))\frac{da}{dt}\end{aligned}$$

   using McShane p.218 to justify differentiating through the integral. Similarly, we can differentiate α and β through the integral.

2. Follows from Major and Mildenhall (2020), Proposition 1 on multiplying by t. The no undercut does not require independence, but positive margin does.

3. $R_1(t) < \rho(X_1)$ iff $P_1(t) < \rho(tX_1)$. Let $\tilde{X}_1$ be the random variable equal to payments allocated to class 1 under equal priority. Then $\tilde{X}_1 \le X_1$ for all outcomes. ρ is monotonic because it is coherent and therefore $\rho(\tilde{X}_1) \le \rho(X_1)$. Finally, $P_1(t) = \mathsf{E}[\tilde{X}_1 g' S_t] \le \sup_{\mathsf{Q}} \mathsf{E}_Q[\tilde{X}_1] = \rho(\tilde{X}_1)$ since $g'S_t$ is a legitimate test measure for computing ρ.

4. $P_1(t)$ is differentiable because Part (1) shows all its constituent parts are differentiable. It follows that $R_1(t) = P_1(t)/t$ is differentiable for $t > 0$.

5. To simplify the formulas, write $g'S_t = g'(S_t(X_t))$ and $\tau = (1-t)/t$. Then, for t sufficiently close to 1:

$$\begin{aligned} R_1(t) &= \mathsf{E}[X_1 1_{\{X_t \le a_t\}} g' S_t] + a_t \mathsf{E}\left[\frac{X_1}{X_t} 1_{\{X_t > a_t\}} g' S_t\right] \\ &\approx \mathsf{E}[X_1 1_{\{X_1 \le a_1\}} g' S_1] + (ta_1 + (1-t)a_0)\mathsf{E}\left[\frac{X_1}{X_t} 1_{\{X_1 > a_1\}} g' S_1\right]. \end{aligned}$$

   This implies

$$\begin{aligned} R_1(t) - P(1) &= (ta_1 + (1-t)a_0)\mathsf{E}[\frac{X_1}{X_t} 1_{\{X_1 > a_1\}} g' S_1] - a_1 \mathsf{E}[1_{\{X_1 > a_1\}} g' S_1] \\ &= \mathsf{E}\left[\frac{tX_1}{tX_1 + (1-t)X_0} \frac{ta_1 + (1-t)a_0}{t} - a_1 \cdot 1_{\{X_1 > a_1\}} g' S_1\right] \\ &= \mathsf{E}\left[\frac{1}{1 + \tau\frac{X_0}{X_1}}(a_1 + \tau a_0) - a_1 \cdot 1_{\{X_1 > a_1\}} g' S_1\right] \\ &\approx \mathsf{E}\left[(1 - \tau\frac{X_0}{X_1})(a_1 + \tau a_0) - a_1 \cdot 1_{\{X_1 > a_1\}} g' S_1\right] \\ &= \tau \mathsf{E}\left[(a_0 - a_1\frac{X_0}{X_1}) \cdot 1_{\{X_1 > a_1\}} g' S_1\right] \\ &= \tau\left(a_0 \mathsf{E}[1_{\{X_1 > a_1\}} g' S_1] - a_1 \mathsf{E}[X_0]\mathsf{E}\left[\frac{1}{X_1} \cdot 1_{\{X_1 > a_1\}} g' S_1\right]\right) \end{aligned}$$

   where the last line follows from the eventual independence of X_0. The final expression is clearly positive iff eq. (13) holds.

6. This follows because

$$
\begin{aligned}
\lim_{t\downarrow 0} R_1(t) &\ge \liminf_{n\to\infty} R_1(1/n) \\
&= \liminf \mathsf{E}[X_1 g' S_{1/n} 1_{\{X_{1/n}\le a_{1/n}\}}] + a_{1/n}\mathsf{E}[\frac{X_1}{X_{1/n}} g' S_{1/n} 1_{\{X_{1/n}>a_{1/n}\}}] \\
&\ge \liminf \mathsf{E}[X_1 g' S_{1/n} 1_{\{X_{1/n}\le a_{1/n}\}}] \\
&\ge \mathsf{E}[X_1 \liminf g' S_{1/n} 1_{\{X_{1/n}\le a_{1/n}\}}] \\
&\ge \mathsf{E}[X_1 g' S_0 1_{\{X_0<a_0\}}] \\
&= \mathsf{E}[X_1](1-g(1-p))
\end{aligned}
$$

by Fatou's lemma and independence.

□

Examples show that a condition stronger than Part (5) holds, however, we have been unable to find a general proof. It is therefore cast as an assumption and we explain why it is plausible. It is called Condition **M** for monotone.

Condition M. If $\rho(X_i) < \infty$, then for all $p < 1$, $R_1(t)$ has at most one turning point in $(0, 1)$, which is a maximum.

When there is a turning point, Assumption M means that $R_1(t)$ increases monotonically to a maximum $> P(1)$ and then decreases to $R_1(1) = P(1)$. It can also decrease monotonically to $P(1)$, when $\rho(X_1)$ is infinite, or increase monotonically to $P(1)$, in the case of unlimited cover. Assumption M implies there exits t^* so that $R_1(t) > tP(1)$ for all $t > t^*$, and $R_1(t) < tP(1)$ and $R_1(t)$ is non-decreasing for $t < t^*$. Similarly, there exists t_* so that $R_0(t) > (1-t)P(0)$ for all $t < t_*$, and $R_0(t) < (1-t)P(0)$ and $R_0(t)$ is non-increasing for $t > t_*$. These behaviors are evident in fig. 3, with $t_* = 0.296$ and $t^* = 0.544$.

Why is condition M reasonable? The slope of $R_1(t)$ is determined by the amount of assets available per unit of loss for class 1, the share of those losses class 1 captures in default, and the impact of class 1 on the risk load. The first effect is generally decreasing in t at a decreasing rate: $a_t \approx (1-t)a_0 + ta_t$ so relative assets are $a_t/t = \tau a_0 + a_1$ and $t \mapsto \tau$ is decreasing in t at a decreasing rate. The second effect is increasing in t at a decreasing rate: as class 1 becomes larger it becomes more correlated to total losses, $\beta_{1,t}(x)$ is increasing in t for all x. And the third effect is also increasing in t at a decreasing rate for the same reason. Until the first effect dominates and $R_1(t)$ is increasing, thereafter it decreases.

Going forward we assume Condition M holds.

It is possible for $R_1'(1)$ to be so negative that $P_1'(1) < 0$, i.e., premium increases absolutely for line X_1 as its proportion in the pool decreases. This occurs when class 1 is very volatile and class 0 is very large and very thin tailed.

4.3 Characterization of Possible Market Structures

Different pools, with different mixes by class, provide different covers. Although the premiums vary by class for pools with different mixes, all the premiums are fair values. Premium differences reflect different payments each pool will make in default states. Insureds select between pools solely on the basis of price; they do not consider the quality (payment) differences. In this section we explain why we can assume there is at most one monoline pool for each class, and one multiline pool. We then show how to determine which outcome occurs.

The homogeneous loss model, combined with positive homogeneous risk-based capital and investor pricing formula functionals, implies the economics of any pool only depend on its mix of business and not on its size. Premium, loss, risk, and capital all scale with volume; loss ratio and return on equity are size-independent.

Trivially, monoline pools within a class all have the same mix, and therefore their economics are independent of size. Hence, we can merge all the monoline pools by class into one, and assume that the merged pool includes all the risks not in multiline pools. This is an important simplification. As a result, our homogeneous loss model is the perfect laboratory in which to study the pooling problem because it puts all the focus on the composition of pools.

Next we explain why there can only be one multiline pool. If $P(t) > (1-t)P(0) + tP(1)$ then one or both allocated rate is greater than the monoline rate: $R_0(1-t) > P(0)$ or $P_1(t) > P(1)$. If this is true then there is no pooled solution at t—insureds will never pay more than the monoline rate. Therefore, any pool occurs with mix t satisfying $t_* \le t \le t^*$. By Condition M, the natural premium rate for each risk in a multiline pool varies monotonically with the proportion of each class in the pool for $t_* \le t \le t^*$. Therefore, two pools with different proportions of each class have different rates by class. Each class has a definite preference for the cheaper of the two pools. As a result, a market with two multiline pools with different mixes cannot be in equilibrium. When the mixes are the same, the two pools can be combined by homogeneity. Therefore we can assume there is only one multiline pool.

These two observations show the market structure is determined by a single decision variable: the mix by class in the multiline pool.

If a multiline pool has $t \le 0.5$ then it can be scaled up to include all the class 0. If $t \ge 0.5$ it can scale to include all the class 1 risks. In both cases, homogeneity implies the economics are independent of pool scale. Thus there are just three structures that can occur:

1. $t = 0.5$: **complete pooling** in a single multiline pool;
2. $t = 0, 1$: **no pooling** with two disjoint monoline pools; or
3. $0 < t < 1$: $t \ne 0.5$: **partial pooling** with a multiline pool and a monoline pool.

In structure 3, all of class 0 is pooled for $0 < t < 0.5$ and all of class 1 for $0.5 < t < 1$. The remaining risks are monoline. Note that $t = 0$ and $t = 1$ define the same outcome. Outcomes are topologically a circle.

Our central question is to determine conditions under which each structure occurs. And when structure 3 occurs, what determines the proportions, whether an individual is pooled or monoline, and how is the diversification benefit distributed? Building on the observations above, the following Theorem presents the answer.

Theorem 2. *Under the standing assumptions from section 2.6, the assumptions of the previous theorem, amd Assumption M we have*

1. *If $P(t)$ is concave then there is no pooling.*
2. *If $t_* < 0.5 < t^*$ there is complete pooling.*
3. *If $t_* > t^*$ there is no pooling.*
4. *If $t_* < t^* < 0.5$ ($0.5 < t_* < t^*$) then the equilibrium is at $t = t^*$ (resp. $t = t_*$) and there is partial pooling.*

Proof:

By market assumption 1, that insureds are pure price buyers. A market with prices greater than monoline rates will not be in equilibrium because risks will defect into a cheaper monoline pool. This has two implications. First, if $P(t)$ is concave then for each t either $R_0(t) > P(0)$ or $R_1(t) > P(1)$. Thus there is no viable solution other than $t = 0$ or $t = 1$, giving Part (1). Second, the feasible range for a solution is $t_* < t < t^*$. For $t < t_*$ the pool rate for class 0 is greater than the monoline rate, so it will not pool. And for $t > t^*$ the same holds for class 1.

We can now determine the market equilibrium, determined by t.

- If $t_* < 0.5 < t^*$ and $t \neq 0.5$ then risks left in the monoline pool, paying a higher rate, will offer to pool at a slightly more advantageous proportion, i.e., lower rate, for the other class and the pool will unravel. The pool is only in equilibrium when there are no risks left in monoline pools, i.e., when $t = 0.5$ and we have structure 1, showing Part (2).
- If $t_* > t^*$ there is no overlap of feasible solutions and we are in structure 2 with no pooling and two monoline pools. The feasible region only contains the point $t = 0 = 1$, remember 0 and 1 define the same outcome. This shows Part (3).
- If $t_* < t^* < 0.5$ then the equilibrium is at $t = t^*$ and we have partial pooling, structure 3. The multiline pool contains all of class 0 and a proportion of class 1. Class 0 gets the benefits of diversification and pays less than its monoline rate. Class 1 pays its monoline rate. Class 0 has the negotiating power: it only needs to attract a portion of the class 1 risks and it knows there are always class 1 risks paying the monoline rate, so that acts as a bogey. It can offer to pool at a proportion $t^* - \epsilon$ for small $\epsilon > 0$. This produces a price $R_1(t^* - \epsilon) < R_1(t^*) = R_2(1)$ below the monoline price, which will be enough to attract as many class 1 risks as needed. If $\epsilon > 0$ then the class 1 risks who remain in a monoline pool, paying a higher rate, have an incentive to offer to pool at a share $t^* - \epsilon/2$ to class 0. The original pool will unravel as this will be a lower price for class 0. Thus the equilibrium is at t^*. Class 1 has no negotiating power because there will always be risks in its monoline pool, whereas all class 0 is entirely in the multiline pool. Similarly if $0.5 < t_* < t^*$ then the multiline pool contains all class 1 and some of class 0. Now class 0 pays the monoline rate and class 1 benefits from the pooling with

a lower rate. The equilibrium is at $t = t_*$ showing Part (4). □

When partial pooling, structure 3, applies, the fully pooled class captures all of the diversification benefit and the split class pays its monoline rate. Since the degree to which the rate function bows up from the monoline rate is drive by how volatile it is, the pooled class will be lower volatility and the split class higher. Thus the capital standard will increase the cost for the more volatile class. An attempt to lower costs by decreasing the capital standard could back-fire because a split solution is more likely with a lower capital standard.

Examples show complete pooling generally applies, especially when the two classes are of comparable volatility or the capital standard is very strict. Complete failure, with two monoline pools occurs for example when X_0, X_1 are very thick tailed or when VaR is not subadditive.

All the equilibriums are Pareto optimal because of the shape of $R_0(t)$ and $R_1(t)$ at t: one is increasing and one decreasing so it is not possible to make both classes better off. Changing t will increase the rate for at least one class.

Returning to fig. 3 we see that its equilibrium is complete pooling, structure 1, since $t_* = 0.206 < 0.5 < 0.544 = t^*$.

It is highly unlikely the solution is at the minimum of $P(t)$. That solution only includes all risks when the minimum occurs at $t = 0.5$.

A reinsurance can be used in structure 3 to provide a blended rate to the split class. Suppose class 1 is split. The reinsurer enables a solution with proportion t where $t_* < t < t^*$ by assuming class 1 from the monoline and the pool in return for a blended rate. However, without an enforcement mechanism this is not a stable solution. Members of class 0 have an incentive to offer reinsurance pool participants the true, lower, multiline rate unless $t = t^*$. In that case the pool would unravel as before.

As the capital standard strengthens the rates bow up less and the feasible region gets broader. With unlimited liability, i.e., $a = \infty$, the market pricing functional reduces to the investor functional, which is coherent. In that case pooling is always optimal provided both premiums are finite. This follows from Proposition 2: for a coherent risk measure the natural allocation always contains a non-negative margin and is less than the monoline rate (no-undercut). Thus the natural allocation premium $P_1(t)$ satisfies $\mathsf{E}[tX_1] \le P_1(t) \le \rho(tX_1)$. Since ρ is positive homogeneous, dividing by t gives $\mathsf{E}[X_1] \le R_1(t) \le \rho(X_1)$. Note $\rho(X_1)$ is the monoline premium for unlimited cover.

The right hand plot of fig. 3 illustrates these points. The two shaded triangles show the segments bounded by $\mathsf{E}[X_i]$ and $\rho(X_i)$. The thicker premium line is within the segment precisely when the rate is between monoline loss cost and premium. The green band shows the corresponding bounds for total premium.

The following behaviors are evident.

1. Full pooling is more likely with a stronger capital standard. In the limit of full coverage, we always get full poling provided the full coverage premiums are both finite. If one or other premium is infinite full pooling will not be an equilibrium solution. Coherent risk

measures on unbounded variables must sometimes assume infinite values by Delbaen (2002), Theorem 5.1.

2. For thin tailed lines with log concave densities full pooling is more likely with more expensive pricing. (Expensive pricing means a greater markup over expected loss.) The opposite is true for thick tailed lines. Thin tailed lines are concentrated near the mean and large losses occur when each component has a somewhat above average loss. In this situation pooling is not greatly beneficial. thick tailed lines are concentrated away from the mean. A large loss occurs from a large loss on one component and a small loss on the other. This situation benefits from pooling and results in superior pooled coverage, which therefore costs more. Thus the market pricing operator fails to be subadditive.
3. Classes with unbalanced tails are more likely to result in solution 3 than balanced classes. This follow for the same reason as the second case of 2.

5 Examples

5.1 Structure 2: No Pooling

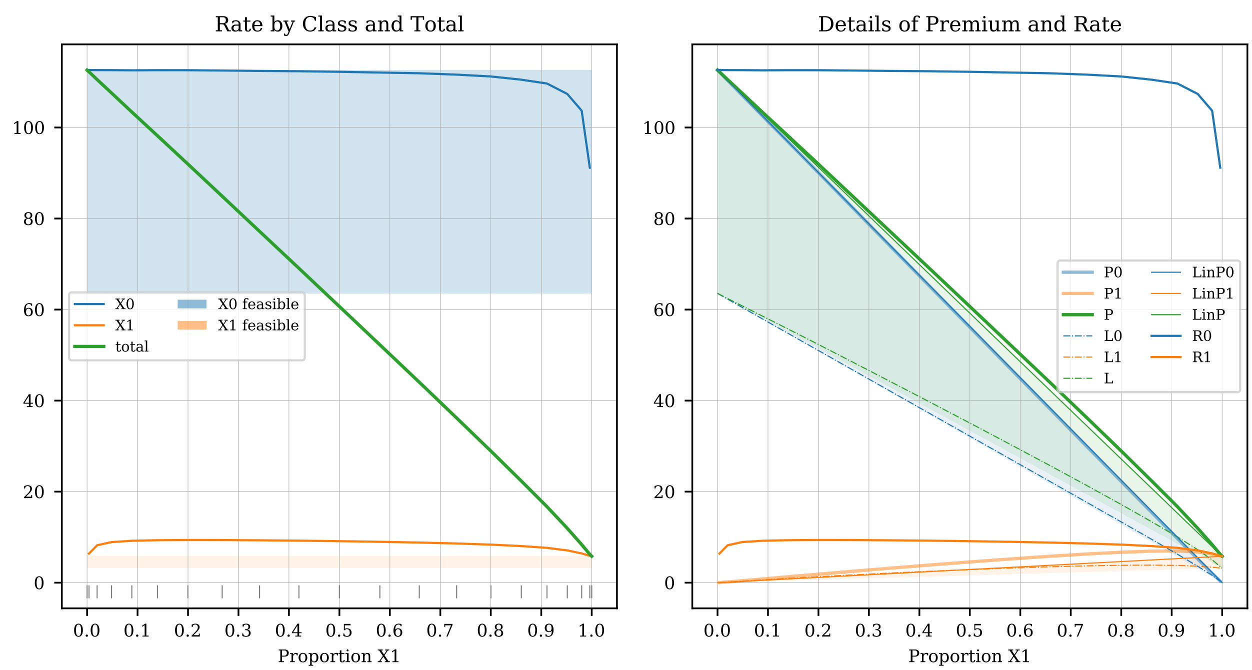


Figure 4: Market equilibrium structure 2: two monoline pools.

In fig. 4 X_0 is a lognormal with unlimited mean 100 and $\sigma = 1.5$ and X_1 is a very heavy tailed Pareto with unlimited mean 10 and $\alpha = 1.2$. The right hand plot shows the (green) premium line bows up from the line connecting the two monoline rates, meaning there can be no pooled solution. The weighted average of two rates lower than the monoline rates has to lie below the line connecting them.

In this example $P(t)$ has a maximum at $t < 1$ and is actually decreasing for t close to 1.

Pooling a small amount of (the much larger) X_0 greatly increases the assets available and X_1 captures an outsize proportion of them in default because of its thick tail. As a result its absolute premium increases as t decreases from 1, not just its premium rate.

Since X_0 is actually quite thick tailed $R_0(t) > \mathsf{E}[X_0]$ for all t: the blue rate line stays in the shaded blue area on the left.

5.2 Structure 3: Partial Pooling

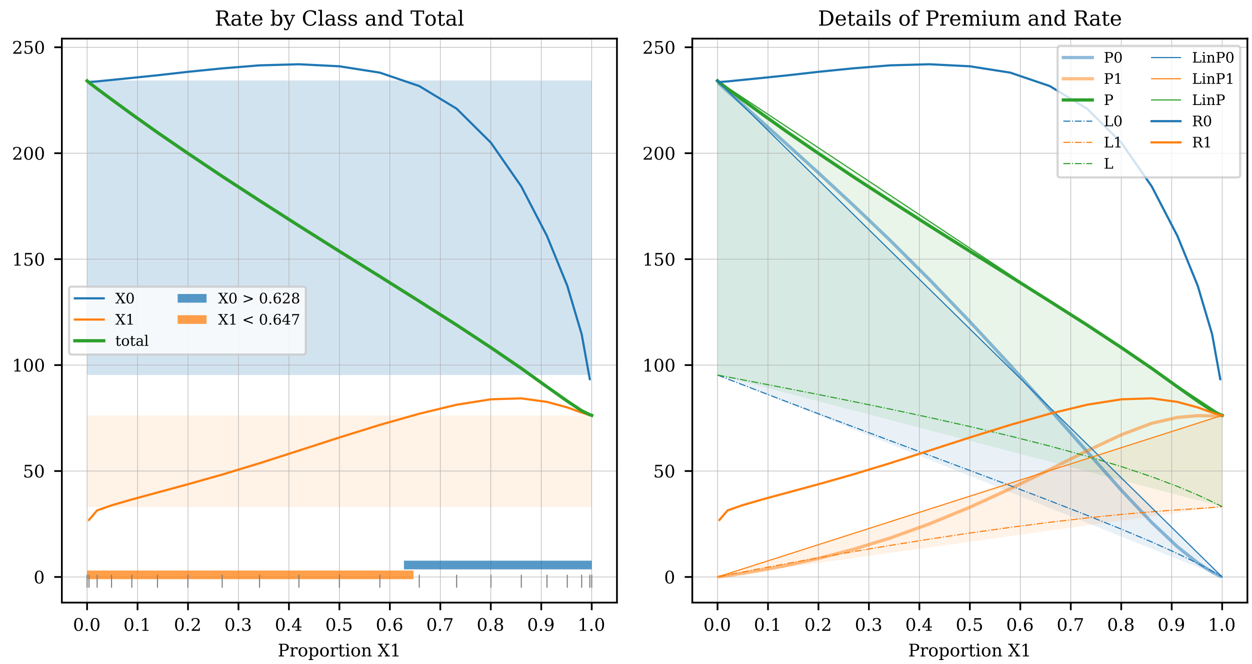


Figure 5: Market equilibrium structure 3: partial pooling.

In fig. 6 illustrates several interesting features. X_0 is a lognormal with $\sigma = 1.5$ and unlimited mean 150; X_1 is Pareto with unlimited mean 50 and $\alpha = 1.9$, so both lines have very thick tails. The capital standard is VaR 0.9 and pricing uses a proportional hazard $g(s) = s^{1/4}$.

The figure shows the following features.

1. $P_1(t) > P(1)$ for t close to 1.
2. P has an inflexion point and is sub-additive for $t \in [0, 0.7]$ and super-additive for $t \in [0.7, 1.0]$.
3. The feasible region is very small, approximately $[0.628, 0, 647]$, which does not include $t = 0.5$ and therefore we have partial pooling, structure 2.

The drop in R_1 for t close to zero is a numerical artifact.

5.3 A Realistic Example

To draw out different behaviors, the two previous examples use more extreme parameters that would normally apply to an insurance book. Our final example shows more typical parameter values to illustrate a generic view. It results in a full pooling structure. Class 0 is gamma, mean 150 and coefficient of variation 0.15 typical for a low limit liability book. Class 1 has a lognormal distribution with a mean of 100 and $\sigma = 0.3$. It represents a higher limit book. The distortion is $g(s) = s^{0.8}$, which is in-line with distortions calibrated to market pricing, and the capital standard is Sovlvency II, $p = 0.995$.

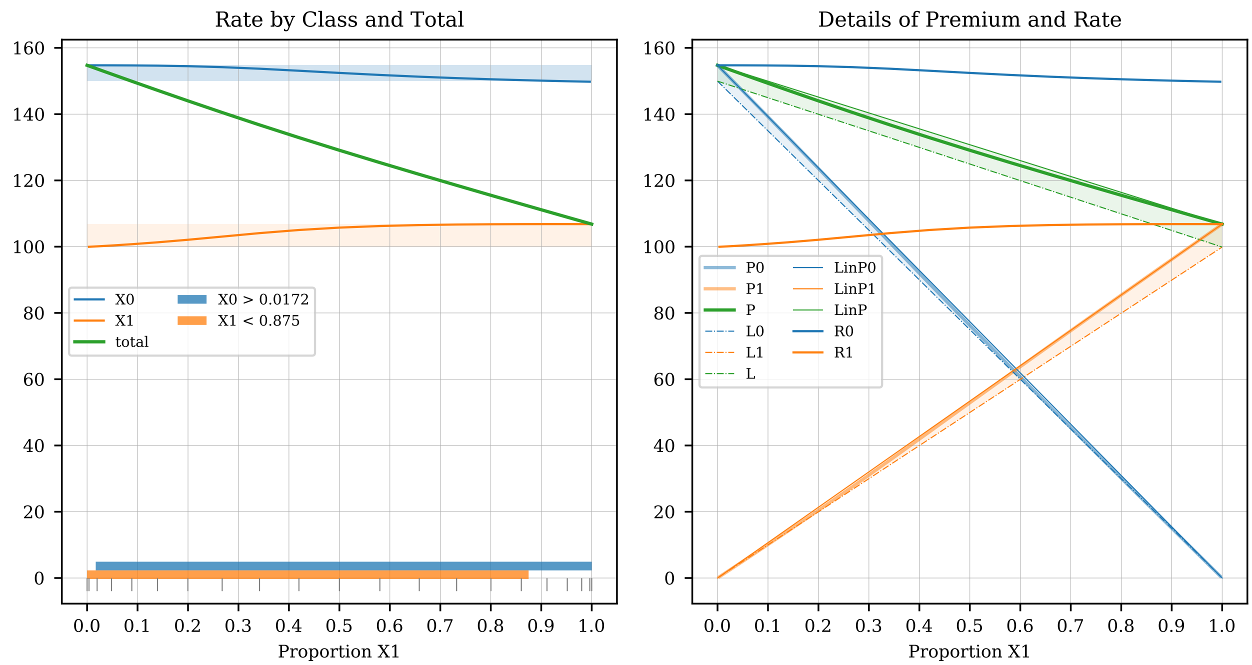


Figure 6: Market equilibrium structure 1: full pooling with parameters reflecting a liability and property insurance portfolio.

It may appear that fig. 6 shows only small differences between classes. Low volatility insurance is a very competitive business, written with very thin margins. Figure 7 shows the implied loss ratios by class. These are comparable with combined ratios, since our model excludes expenses. The loss ratio differences are material. The equilibrium, $t = 0.5$ shows class 0 written at 98.4 percent vs. its monoline rate of 97 percent. Class 1 is written at 94.5 percent vs. 93.5 percent. Thus class 0 achieves a 1.4 percentage point decrease in loss ratio from pooling vs. 1 point for class 1.

5.4 Note on the Computations

The computations underlying each figure were performed using discrete approximations with 2^{16} equally sized buckets and a sample of 21 values of t in $[0, 1]$ inclusive. Convolutions are performed using Fast Fourier Transforms (FFT), Grubel and Hermesmeier (1999), Mildenhall (2005). The calculations are essentially exact other than a minor discretization error. The conditional expectations needed for κ_i are also be performed using a FFT

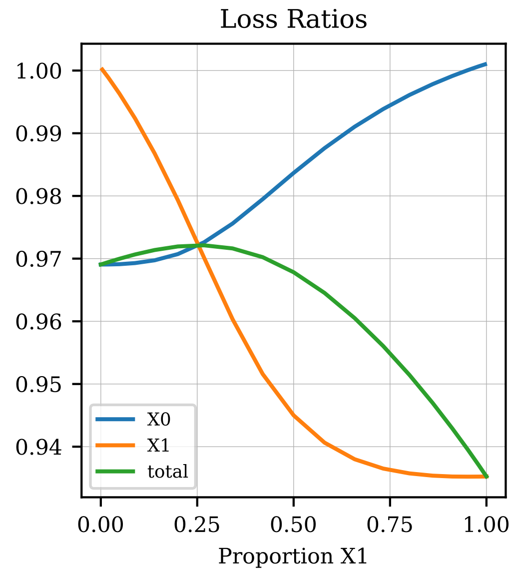


Figure 7: Loss ratios by class across different portfolio corresponding to fig. 6.

6 Conclusions

We have presented a novel but realistic model of a two class insurance market. The market includes a compulsory insurance requirement, capital regulation, and costly capital but otherwise is efficient. Depending on the aggregate loss characteristics of the two risk classes the Pareto optimal market equilibrium can be two monoline pools, a monoline and a multiline pool, or one multiline pool. In general when the classes are comparably risky there is one multiline pool. When a more risky class is combined with a less risky one, the less risky class often gets the benefit of pooling and pays a rate below its monoline premium while the more risky class pays its monoline premium. Stricter capital standards make more complete pooling more likely because it increases the importance of economizing on capital. There is no pooling when the risk have extremely thick tails and the capital standard is not subadditive.

Reinsurance can be used to pool premium rates, but without an enforcement mechanism it does not provide a stable solution.

The results are consistent with observed market structure in US property-casualty insurance, where more volatile lines are often written by monoline companies. Florida homeowners and medical malpractice liability are two examples. It is also consistent with the existence of highly leveraged, low risk pools, such as monoline auto writers.

References

Acerbi, Carlo, 2002, Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking & Finance* 26, 1505–1518.

Anscombe, F. J., and R. J. Aumann, 1963, A Definition of Subjective Probability, *The Annals of Mathematical Statistics* 34, 199–205.

Aon Benfield, 2015, *Insurance Risk Study.* 10th ed.

Arrow, Kenneth J., 1996, The theory of risk-bearing: Small and great risks, *Journal of Risk and Uncertainty* 12, 103–111.

Boonen, Tim J., Andreas Tsanakas, and Mario V. Wüthrich, 2017, Capital allocation for portfolios with non-linear risk aggregation, *Insurance: Mathematics and Economics* 72, 95–106.

Castagnoli, Erio, Fabio Maccheroni, and Massimo Marinacci, 2002, Insurance premia consistent with the market, *Insurance: Mathematics and Economics* 31, 267–284.

Castagnoli, Erio, Fabio Maccheroni, and Massimo Marinacci, 2004, Choquet insurance pricing: A caveat, *Mathematical Finance* 14, 481–485.

Chateauneuf, A., R. Kast, and A. Lapied, 1996, Choquet pricing for financial markets with frictions, *Mathematical Finance* 6, 323–330.

Cummins, J. David, 1988, Risk-Based Premiums for Insurance Guaranty Funds, *Journal of Finance* 43, 823–839.

Cummins, J. David, 2000, Allocation of capital in the insurance industry, *Risk Management and Insurance Review*, 7–28.

Cummins, J. David, Mary A. Weiss, Xiaoying Xie, and Hongmin Zi, 2010, Economies of scope in financial services: A DEA efficiency analysis of the US insurance industry, *Journal of Banking and Finance* 34, 1525–1539.

Delbaen, Freddy, 2002, Coherent risk measures on general probability spaces, *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, 1–37.

Denneberg, Dieter, 1994, *Non-additive measure and integral* (Kluwer Academic, Dordrecht).

De Waegenaere, Anja, 2000, Arbitrage and Viability in Insurance Markets, *GENEVA Papers on Risk and Insurance Theory* 25, 81–99.

De Waegenaere, Anja, Robert Kast, and Andre Lapied, 2003, Choquet pricing and equilibrium, *Insurance: Mathematics and Economics* 32, 359–370.

Dhaene, J, RJA J A Laeven, S Vanduffel, G. Darkiewicz, and M. J. Goovaerts, 2008, Can a coherent risk measure be too subadditive?, *Journal of Risk and Insurance* 75, 365–386.

Dietz, Simon, and Oliver Walker, 2017, Ambiguity and Insurance: Capital Requirements and Premiums, *Journal of Risk and Insurance* 86, 213–235.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger, 2011, *Economic and financial decisions under risk* (Princeton University Press).

Ellsberg, Daniel, 1961, Risk, Ambiguity, and the Savage Axioms, *The Quarterly Journal of Economics* 75, 643–669.

Föllmer, Hans, and Alexander Schied, 2011, *Stochastic finance: an introduction in discrete time.* Third Edit. (Walter de Gruyter).

Grubel, Rudolf, and Renate Hermesmeier, 1999, Computation of Compound Distributions I: Aliasing Errors and Exponential Tilting, *Astin Bulletin* 29, 197–214.

Huber, Peter J., 1981, *Robust statistics* (John Wiley & Sons, Inc.).

Ibragimov, Rustam, Dwight Jaffee, and Johan Walden, 2018, Equilibrium with monoline and multiline structures, *Review of Finance* 22, 595–632.

Ibragimov, Rustam, Dwight Jaffee, and Johan Walden, 2010, Pricing and Capital Allocation for Multiline Insurance Firms, *Journal of Risk and Insurance* 77, 551–578.

Ibragimov, Rustam, and Johan Walden, 2007, The limits of diversification when losses may be large, *Journal of Banking and Finance* 31, 2551–2569.

Jiang, Wenjun, Marcos Escobar-Anel, and Jiandong Ren, 2020, Optimal insurance contracts under distortion risk measures with ambiguity aversion, *ASTIN Bulletin* 50, 1–28.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji, 2005, A smooth model of decision making under ambiguity, *Econometrica* 73, 1849–1892.

Kusuoka, Shigeo, 2001, On law invariant coherent risk measures, *Advances in Mathematical Economics* 3, 83–95.

Major, John A., and Stephen J. Mildenhall, 2020, Pricing and Capital Allocation for Multiline Insurance Firms With Finite Assets in an Imperfect Market, *Arxiv*, 1–33.

Marshall, Daniel, 2018, An Overview of the California Earthquake Authority, *Risk Management and Insurance Review* 21, 73–116.

Mildenhall, Stephen J, 2005, Correlation and Aggregate Loss Distributions With An Emphasis On The Iman-Conover Method, *Casualty Actuarial Society Forum, WinterSociety.*

Mildenhall, Stephen J, 2017, The Coming Revolution in Actuarial Modeling—Election Day Lessons for the Predictive Data Analyst, *Actuarial Review*, 38——43.

Myers, Stewart C, and James A Read Jr., 2001, Capital allocation for insurance companies, *Journal of Risk and Insurance* 68, 545–580.

Phillips, Richard D., J. David Cummins, and Franklin Allen, 1998, Financial Pricing of Insurance in the Multiple-Line Insurance Company, *Journal of Risk and Insurance* 65, 597–636.

Robert, Christian Y., and Pierre-E. Therond, 2014, Distortion Risk Measures, Ambiguity Aversion and Optimal Effort, *ASTIN Bulletin* 44, 277–302.

Schmeidler, David, 1986, Integral representation without additivity, *Proceedings of the American Mathematical Society* 97, 255–255.

Schmeidler, David, 1989, Subjective Probability and Expected Utility without Additivity, *Econometrica* 57, 571–587.

Shapiro, Alexander, 2012, On Kusuoka Representation of Law Invariant Risk Measures, *Mathematics of Operations Research* 38, 142–152.

Sherris, Michael, 2006, Solvency, capital allocation, and fair rate of return in insurance, *Journal of Risk and Insurance* 73, 71–96.

Svindland, Gregor, 2009, Continuity properties of law-invariant (quasi-)convex risk functions on L^∞, *Mathematics and Financial Economics* 3, 39–43.

Tasche, Dirk, 2001, Conditional expectation as quantile derivative, *Arxiv preprint math/0104190*, 1–12.

Wang, Shaun, 1996, Premium Calculation by Transforming the Layer Premium Density, *ASTIN Bulletin* 26, 71–92.

Wang, Shaun, Virginia Young, and Harry Panjer, 1997, Axiomatic characterization of insurance prices, *Insurance: Mathematics and Economics* 21, 173–183.

Yaari, Menahem E., 1987, The Dual Theory of Choice under Risk, *The Econometric Society* 55, 95–115.

Zhang, Jiankang, 2002, Subjective Ambiguity, Expected Utility and Choquet Expected Utility, *Economic Theory* 20, 159–181.