



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

The Marginal Cost of Risk, Risk Measures, and Capital Allocation

Daniel Bauer, George Zanjani

To cite this article:

Daniel Bauer, George Zanjani (2016) The Marginal Cost of Risk, Risk Measures, and Capital Allocation. Management Science 62(5):1431-1457. <http://dx.doi.org/10.1287/mnsc.2015.2190>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2015, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

The Marginal Cost of Risk, Risk Measures, and Capital Allocation

Daniel Bauer, George Zanjani

Department of Risk Management and Insurance, Georgia State University, Atlanta, Georgia 30303
{dbauer@gsu.edu, gzanjani@gsu.edu}

Financial institutions use risk measures to calculate the marginal capital cost when expanding the exposure to a certain risk within their portfolio. We reverse this approach by calculating the marginal cost based on economic fundamentals for a profit-maximizing firm and then by identifying the risk measure delivering the correct marginal cost. The resulting measure depends on context. Whereas familiar measures can be recovered in some circumstances, other circumstances yield unfamiliar forms. In all cases, the risk preferences of the institution's claimants determine how the correct risk measure must weight various default states. Our results demonstrate that risk measures used for pricing and performance measurement should be chosen based on economic fundamentals and may not necessarily adhere to the mathematical properties typically imposed in the literature.

Keywords: capital allocation; risk measure; profit maximization; counterparty risk

History: Received May 15, 2013; accepted January 16, 2015, by Jerome B. Detemple, finance. Published online in *Articles in Advance* October 15, 2015.

1. Introduction

Risk measures are widely used by financial institutions for pricing and performance measurement.¹ The question of which measure to use is thus of great practical importance and has attracted much scholarly attention in recent years. Among the most influential contributions on this topic are papers from the mathematical finance literature, the most well known of which (Artzner et al. 1999) introduced axioms defining a *coherent* risk measure. The axiomatic approach in general, and *coherence* in particular, is widely embraced by both academics and practitioners.

In this paper, we argue that the properties of a risk measure should flow from the economic context of the problem rather than arbitrarily imposed axioms. We demonstrate that the most widely accepted axioms may fail when used for risk pricing and performance measurement. More precisely, financial institutions use the gradients of risk measures to allocate the firm's capital to the various risks within its portfolio—a process that effectively determines the marginal cost of

risk. We reverse the sequence of this approach. Instead of taking the gradient of an arbitrary risk measure to identify the marginal cost of risk, we start with an economic model of a financial institution with risk-averse counterparties in an incomplete market with frictional capital costs and show that profit maximization yields an endogenous expression for marginal cost that can be used for capital allocation. We then derive the risk measure that gives the correct capital allocations.

The form of the derived measure depends crucially on institutional specifics and the economic environment. In some cases, familiar measures are obtained. In other cases, however, unfamiliar measures emerge that do not adhere to the mathematical axioms of coherence. In particular, cases where risks are associated with liabilities rather than assets lead to new risk measures.

We develop the latter finding in §2 with a one-period model in which a firm sells insurance to a set of counterparties. The firm's risk taking is constrained in this case solely by the preferences of counterparties, who of course want to be paid in full. When the firm sells additional insurance to one counterparty so as to make insolvency more likely or more severe, all counterparties are affected and are thus willing to pay less for the firm's contracts. The marginal cost of contractual risk exposure in this environment has two components. The first reflects the marginal increase in payments made under the contract in question, and the second reflects a risk penalty for the marginal impact of that contract's exposure on the entire set of counterparties. We show the latter

¹ For instance, a Baer et al. (2011) survey among a “diverse group of 11 leading banks” (p. 1) revealed that the “vast majority of respondents use economic-capital [EC] models” (p. 2), mostly for “tracking performance of individual business units or portfolios” (p. 4), although more sophisticated applications such as pricing or risk-based strategic decision making also appear in the sample. Similarly, according to Farr et al. (2008), more than 80% of all insurance companies calculated EC or considered the implementation of an EC framework in 2006, where “allocation of capital” and “measure of risk-adjusted performance” (p. 5) were listed as the two primary drivers.

penalty can be reinterpreted as an allocation of the firm's capital times the cost of that capital. We then "reverse-engineer" the risk measure whose gradient yields the economically correct capital allocations. The resulting risk measure takes an unfamiliar form—the exponential of a weighted average of the logarithm of portfolio outcomes in states of default, with the weights being determined by the relative values the firm's counterparties place on recoveries in the various states of default.

In §3, we contrast the new risk measure with the popular measures value-at-risk (VaR) and expected shortfall (ES) in two example settings: (i) homogeneous exponential losses and (ii) heterogeneous Pareto losses. In particular, we compare the axiomatic properties and capital allocations associated with the measures. We show that, because of the logarithm transformation, the functional form of the new risk measure does not adhere to the coherence properties. Specifically, the measure fails to be invariant to constant shifts of the risk and it does not always reward diversification. For instance, in analogy to VaR, this is the case in the context of the Pareto example for heavy tails with a tail index of less than 1. We also find that VaR and ES capital allocations generally fail to weight default outcomes properly. Specifically, whereas VaR generally underweights large losses in our examples, ES-based allocations may penalize large losses too much or too little, depending on the underlying parameters. For instance, in cases where counterparties are strongly risk averse or where potential losses are large relative to counterparty wealth, ES-based capital allocations tend to underweight severe outcomes; in cases where counterparties are only weakly risk averse or where potential losses are relatively small, ES-based allocations tend to overweight severe outcomes.

These discrepancies flow from a fundamental difference in the basis for capital allocation under ES versus the basis implied by the economics. The economic basis for allocation is *recoveries*: risks are penalized according to how they affect claimant recoveries in states of default. The basis for allocating capital to a risk under ES, on the other hand, is simply that risk's share of *losses* in excess of a given threshold. The approaches coincide if there is a dollar-for-dollar relationship between losses and claimant recoveries, as is the case when the risks being measured are those associated with assets: a dollar reduction in the value of an asset obviously translates into a dollar less of recoveries when the firm is liquidated. However, liability risks such as those of §2 affect recoveries in a different way: a liability risk generates an additional competing claimant on the firm without affecting the value of assets. Thus, liability risks (from sources that are *pari passu* with other claimants) affect the distribution of recoveries among claimants rather than the

total amount recovered, a difference that leads to a fundamentally different risk measure.

In §4 and Appendix B, we report the effects of various extensions and modifications of the model from §2. We first show that introducing risk aversion into the firm's objective does not change the form of the allocation result, although the definition of the capital cost changes relative to §2. We also verify that the results continue to apply when securities markets are introduced, so that both the firm and its counterparties can hedge by investing in other financial instruments.

We continue extending the model by introducing (i) an external risk measure constraint, as might be imposed by a regulator or a rating agency, and (ii) multiple periods. These extensions introduce additional influences on the marginal cost of risk. Now, in addition to its impact on current counterparties, risk also affects the regulatory constraint as well as profit flows in future periods. The optimal capital allocation rule in this environment ends up being a weighted average of rules emerging from each of these three influences, and one obtains different risk measures corresponding to each of them. Of particular interest here is the measure introduced by multiperiodicity: in our setting, future profit flows are forfeited if losses exceed available assets, a risk that is captured naturally by VaR.

In §5, we shift the focus to the setting of a retail bank, where the risks emanate from investment choices rather than uncertain obligations to counterparties. Here, the risk measure takes a different form—a weighted average of the asset returns in states of default, with the weights being determined by the depositor's relative valuation of recoveries in these states. The form matches that of a "spectral" version of ES introduced by Acerbi (2002). The resurrection of ES in this case has to do with the shift of the impact of risk from the liability side to the asset side of the balance sheet. Now bad realizations reduce total recoveries dollar for dollar in default states, so the foundation of ES—the expected value of asset losses in bad states of the world—directly aligns with depositor concerns.

Additional extensions are possible, and we discuss some in more detail in the final section of the paper. However, these extensions do not affect the key point: the true marginal cost of risk and the associated allocation of capital should flow from the economic context of the problem. Different model setups will yield different risk measures, but a risk measure chosen for its technical properties such as coherence, rather than for the specific economic circumstances, will generally fail to yield correct pricing and efficient allocation of capital from the perspective of its user.

1.1. Relationship to the Literature

Formal analysis of the problem of capital allocation based on the gradient of a risk measure appeared in the

banking and insurance literatures around the turn of the millennium and was subsequently generalized (see Schmock and Straumann 1999, Denault 2001, Tasche 2004, Kalkbrenner 2005, and Powers 2007, among others). Broadly speaking, these papers start with a differentiable risk measure and end up allocating capital by computing the marginal capital increase required to maintain the risk measure at a threshold value as a particular risk exposure within the portfolio is expanded, an approach referred to as *gradient* allocation or *Euler* allocation.

The technique thus neatly defines the marginal cost of risk if the risk measure is—in one way or another—embedded in the institution's optimization problem (Myers and Read 2001, Zanjani 2002, Stoughton and Zechner 2007, Erel et al. 2015). Unfortunately, excepting highly specialized circumstances,² economic theory offers little guidance on the choice of measure. Perhaps as a consequence, the debate on risk measure selection has largely centered on mathematical properties of the measures (Artzner et al. 1999, Föllmer and Schied 2002, Kou et al. 2013). Yet the choice obviously has profound economic consequences, because it ultimately determines how the institution perceives risk.

Other papers derive the marginal cost of risk and capital allocations from the fundamentals of the institution's profit maximization problem without imposing a risk measure. The ensuing results are transparent if complete and frictionless markets are assumed (Phillips et al. 1998, Ibragimov et al. 2010), although this setting begs the question of why intermediaries would hold capital in the first place.³ Others have studied incomplete market settings. In particular, Froot and Stein (1998) and Froot (2007) introduce the frictions suggested by Froot et al. (1993) to motivate capital holding and risk management. Their models generate a marginal cost of risk determined by an institution's portfolio and effective risk aversion (as implied by a concave payoff function and a convex external financing cost). Institution-specific risk pricing and capital allocation is also found by Zanjani (2010)—although in the context of a social planning problem—where

risk management is motivated by counterparty risk aversion.

Our paper builds on the incomplete market approaches. The key motivation for risk management and key determinant of the marginal cost of risk is counterparty risk aversion, although our general setting also features regulatory constraints and costly bankruptcy.⁴ We recover familiar results in certain cases,⁵ but the general form of capital allocation is multifaceted. The complexity echoes Froot and Stein's (1998) criticism that allocating capital via arbitrary risk measures is problematic because it is “not derived from first principles to address the objective of shareholder value maximization” (p. 57). Froot and Stein, however, did not attempt to reconcile risk-measure-based approaches with those based on “first principles.” This leaves a gap between financial theory and practice that we close here by extracting capital allocations from marginal cost calculations and then deriving risk measures consistent with the extracted allocations. This connection between the marginal cost obtained from the fundamentals of the institution's problem and that obtained from approaches based on risk measures has, to our knowledge, never been explored.

2. Profit Maximization and Capital Allocation

We frame our approach for the case of an insurance company, and our language reflects this in that we refer to the financial contracts as “insurance coverage” and the counterparties of the institution as “consumers.” The setup fits other institutions providing similar contracts, such as reinsurance companies and private pension plan sponsors—or even to an institution selling credit default swaps: the unifying feature of these contexts is that the main uncertainties emanate from obligations to counterparties. The approach can also be adapted to fit other institutions where capital allocation is relevant (such as commercial banks) but where the key risks emanate from the asset side of the balance sheet. We consider this adaptation in §5. The key assumption in all settings, however, is that the stakeholders are exposed to the failure of the institution—and their preferences for solvency drive the motivation for risk management.

² If *institutional* preferences are defined by a risk-averse utility function of outcomes, a particular risk measure may be implied (see, e.g., Föllmer and Schied 2010). Alternatively, Adrian and Shin (2014) justify using VaR in a model with limited commitment and a specialized risk structure.

³ Rampini and Viswanathan (2014) provide a rationale for financial intermediary capital in a complete market environment. In their model, it is opportune for intermediaries to hold capital because of different collateral requirements for households and intermediaries arising, e.g., from limited enforcement as in Rampini and Viswanathan (2010)—i.e., intermediaries are “collateralization specialists.” By contrast, within our incomplete market setting, holding capital is a risk management device that averts default in adverse states of the world.

⁴ Bankruptcy costs originate from shareholders not having access to future profits in default states as in Smith and Stulz (1985) and Smith et al. (1990). As noted by Froot et al. (1993), this produces “similar mechanics” to those obtained when considering a convex cost of external finance as in Froot and Stein (1998).

⁵ For instance, in the limiting case of a complete market, our allocation rule reduces to the allocation derived in Ibragimov et al. (2010), whereas it coincides with the allocation from Zanjani (2010) for the specialized risk structure considered there.

We start by considering a greatly simplified one-period model. In §4, we generalize the results to the case where both the firm and its consumers have access to securities markets and to multiple periods.

Formally, we consider an insurance company that has N consumers, with consumer i facing a loss L_i modeled as a continuous, nonnegative random variable with (joint) density function f_{L_1, L_2, \dots, L_N} . The firm determines the optimal level of assets a for the company, as well as levels of insurance coverage for the consumers, with the coverage indemnification level for consumer i denoted as a function of the loss experienced and a parameter q_i . For tractability, we focus on a proportional arrangement, i.e., a linear contract, where the insurer agrees to reimburse q_i per dollar of loss:

$$I_i = I_i(L_i, q_i) = q_i \times L_i.$$

However, generalizations are possible.

If a consumer experiences a loss, she claims to the extent of the promised indemnification. If total claims are less than company assets, all are paid in full. If not, all claimants are paid at the same rate per dollar of coverage. The total claims submitted are

$$I = I(L_1, L_2, \dots, L_N, q_1, q_2, \dots, q_N) = \sum_{j=1}^N I_j(L_j, q_j),$$

and we define the consumer's recovery as

$$R_i = \min \left\{ I_i(L_i, q_i), \frac{a}{I} I_i(L_i, q_i) \right\}. \quad (1)$$

Accordingly, let $\{I > a\}$ denote the states in which the company defaults, and let $\{I \leq a\}$ denote the solvent states. The expected value of recoveries for the i th consumer is hence given by

$$e_i = \mathbb{E}[R_i] = \underbrace{\mathbb{E}[R_i \mathbf{1}_{\{I \leq a\}}]}_{=e_i^Z} + \underbrace{\mathbb{E}[R_i \mathbf{1}_{\{I > a\}}]}_{=e_i^D}.$$

There is a frictional cost—including taxes, agency, and monitoring costs—associated with holding assets in the company. In the spirit of Froot and Stein (1998), we represent the cost as a tax on assets:

$$\tau \times a,$$

although it is also possible to represent frictional costs as a tax on equity capital, as in

$$\tau \times \left(a - \mathbb{E} \left[\sum_{i=1}^N \min \left\{ I_i(L_i, q_i), \frac{a}{I} I_i(L_i, q_i) \right\} \right] \right),$$

and this does not change the ensuing allocation result. Without a frictional cost, the problem of default would be trivially solved because there would be no point in burdening the consumers with default risk.

We denote the premium charged to consumer i as p_i , and consumer utility may be expressed as

$$v_i(a, w_i - p_i, q_1, \dots, q_N) = \mathbb{E}[U_i(w_i - p_i - L_i + R_i)],$$

where w_i denotes consumer i 's wealth, and we write $v'_i(\cdot) = (\partial/\partial w_i)v_i(\cdot)$.

The firm then solves

$$\max_{a, \{q_i\}, \{p_i\}} \left\{ \sum p_i - \sum e_i - \tau a, \right\} \quad (2)$$

subject to participation constraints for the consumers:

$$v_i(a, w_i - p_i, q_1, \dots, q_N) \geq \gamma_i \quad \forall i. \quad (3)$$

The participation constraints contain the parameters γ_i , which can be used to incorporate different degrees of competition. For example, γ_i set to reflect uninsured consumer utilities would correspond to a case of pure monopoly, where the monopolist could practice first-degree price discrimination and extract all consumer surplus associated with insurance. At the other end of the spectrum, γ_i could be set so high as to simulate (close to) perfect competition.

Let λ_k be the Lagrange multiplier associated with the participation constraint for consumer k . Then the first-order conditions of the optimization problem are

$$\begin{aligned} [a] \quad & -\sum_k \frac{\partial e_k}{\partial a} - \tau + \sum_k \lambda_k \frac{\partial v_k}{\partial a} = 0, \\ [q_i] \quad & \lambda_i \frac{\partial v_i}{\partial q_i} - \sum_k \frac{\partial e_k}{\partial q_i} + \sum_{k \neq i} \lambda_k \frac{\partial v_k}{\partial q_i} = 0, \\ [p_i] \quad & 1 - \lambda_i v'_i = 0. \end{aligned}$$

The first-order condition for $[a]$ reveals the balancing of the marginal cost of holding capital (the first two terms reflecting the increase in actuarial cost and the frictional cost) with the marginal benefits enjoyed by the consumers as a result of the improved security (the third term). The first-order condition for $[q_i]$ balances the marginal benefit associated with improving consumer i 's utility (the first term) with the marginal increase in the actuarial cost (the second term) and the externalities on other consumers produced by the marginal increase in consumer i 's exposure (the third term). Hence, since costly capital can make up for the externalities caused by expanding i 's coverage, taken together, these conditions for the optimal policy balance insurance benefits to the consumer with insurance and capital costs. In particular, their combination yields an allocation of capital costs to consumers.

To see this, consider the firm implementing its optimum by giving each consumer a premium function $p_i^*(\cdot)$ and allowing each to freely choose the level of

coverage. The consumer's optimal choice, given the premium function, will then satisfy

$$\frac{\partial}{\partial q_i} v_i(a, w_i - p_i^*(q_i), q_1, q_2, \dots, q_N) = 0 \Leftrightarrow \frac{\partial p_i^*}{\partial q_i} v_i' = \frac{\partial v_i}{\partial q_i}. \quad (4)$$

To make this consistent with the firm's profit maximization goal, the marginal premium income for the risk associated with consumer i , $\partial p_i^*/\partial q_i$, must equal the marginal cost of risk i . Using $[q_i]$ and $[p_i]$,

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \lambda_i \frac{\partial v_i'}{\partial q_i} = \sum_k \frac{\partial e_k}{\partial q_i} - \sum_{k \neq i} \frac{\partial v_k/\partial q_i}{v_k'} \\ &= \mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U_i'}{v_i'} \frac{a I_i}{I} \frac{\partial I_i/\partial q_i}{I} \right] \\ &\quad + \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \sum_k \frac{U_k'}{v_k'} \frac{a I_k}{I} \frac{\partial I_i/\partial q_i}{I} \right]. \end{aligned} \quad (5)$$

The first two terms on the right-hand side relate to changes in consumer i 's recoveries, whereas the origin of the last term lies in the negative effects that an increase in consumer i 's exposure has on all consumers of the firm. Such effects, of course, materialize only in states where the insurer is defaulting, in which case the other consumers experience a reduction in recoveries because of larger claims from consumer i . However, as indicated above, holding costly capital can be used to offset this reduction of recoveries in default states, and the optimal policy balances benefits and costs. More precisely, $[a]$ gives $\sum_k \lambda_k (\partial v_k/\partial a) = \mathbb{E}[\mathbf{1}_{\{I > a\}} \sum_k (U_k'/v_k')(I_k/I)] = \tau + \mathbb{P}(I > a)$, so that we obtain for the marginal cost of risk

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U_i'}{v_i'} \frac{a I_i}{I} \frac{\partial I_i/\partial q_i}{I} \right] \\ &\quad + \underbrace{\frac{\mathbb{E}[\mathbf{1}_{\{I > a\}} ((\partial I_i/\partial q_i)/I) \sum_k (U_k'/v_k')(I_k/I)]}{\mathbb{E}[\mathbf{1}_{\{I > a\}} \sum_k (U_k'/v_k')(I_k/I)]}}_{=\tilde{\phi}_i} \\ &\quad \cdot [\tau a + \mathbb{P}(I > a)a]. \end{aligned} \quad (6)$$

Since $\sum_i q_i \tilde{\phi}_i = 1$, we can interpret the last term in the marginal cost as an allocation of capital costs that "adds up" to the firm's total capital cost:

$$\sum_i q_i \tilde{\phi}_i a [\tau + \mathbb{P}(I > a)] = a [\tau + \mathbb{P}(I > a)].$$

Each consumer's share of the capital cost derives from the *value* of the recovered fraction that she obtains from the firm in various states of default, where the *value* of recoveries is determined by the weighted-average marginal utilities of affected consumers in states of default.⁶

⁶ We obtain the identical allocation result when we allow for risk aversion at the firm level, although in this case the company's preferences affect capital costs and the valuation of indemnities in solvent states. We refer to Appendix B.1 for details.

REMARK 1. Although the aggregated marginal costs $\sum_i q_i (\partial p_i^*/\partial q_i)$ recover the full capital cost $[\tau a + \mathbb{P}(I > a)a]$, they do not recoup the full cost in solvent states $\mathbb{E}[\mathbf{1}_{\{I \leq a\}} I]$ due to the second term on the right-hand side of Equation (6). This observation that aggregated marginal costs do not equal total cost is not surprising because of the nonlinear nature of the problem. However, following related literature, it is possible to align the two under slight modifications of the setting with the capital allocation result remaining intact. More precisely, a common assumption in the transportation economics literature, where a similar problem arises in the context of congestion externalities (Keeler and Small 1977), asserts that the consumer ignores her own contribution to the overall outcome when choosing the participation level. If we adopt this assumption, the second term in (6) will vanish, and we obtain

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \underbrace{\mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right]}_{(I)} + \underbrace{\tilde{\phi}_i a [\tau + \mathbb{P}(I > a)]}_{(II)} \\ \Rightarrow \sum_i q_i (\partial p_i^*/\partial q_i) &= \sum_i e_i^Z + \mathbb{P}(I > a)a + \tau a; \end{aligned} \quad (7)$$

i.e., aggregated marginal costs fully recover all of the firm's costs. In particular, in this case we can decompose the marginal cost of risk into actuarial cost (I) and capital cost (II) for consumer i . Alternatively, it is possible to derive (7) as the limit of (6) when the portfolio size increases under certain assumptions on the distribution of the losses and the preference specification. We discuss both situations in more detail in Appendix A.1.

Given the capital allocation implicit in (6) and (7), we now turn to the question of what risk measure delivers this allocation, as the typical industry practice is to allocate capital via the gradients of risk measures. In general, a risk measure ρ is a mapping that assigns a real number $\rho(X)$ to each random variable X . Under the assumption of *positive homogeneity* ($\rho(\alpha X) = \alpha \rho(X)$ for $\alpha > 0$)—which is satisfied for popular risk measures such as standard deviation, VaR, and ES and is one of the defining properties of coherence—it is possible to decompose the measure of the aggregate risk via Euler's homogeneous function theorem:

$$\rho(I) = \rho \left(\sum_{i=1}^N I_i \right) = \sum_{i=1}^N q_i \frac{\partial}{\partial q_i} \rho(I).$$

One can then allocate capital to consumers according to their contribution to the aggregate risk, resulting in so-called *gradient* or *Euler allocations*:⁷

$$q_i \frac{(\partial/\partial q_i) \rho(I)}{\rho(I)}.$$

⁷ Formally, this method of defining marginal capital costs arises from the maximization of profits subject to a risk measure constraint

To align the risk-measure-based approach with our economically derived allocations, we introduce the probability measure $\tilde{\mathbb{P}}$ via its likelihood ratio:

$$\frac{\partial \tilde{\mathbb{P}}}{\partial \mathbb{P}} = \frac{\mathbf{1}_{\{I > a\}} \sum_k (U'_k/v'_k)(I_k/I)}{\mathbb{E}[\mathbf{1}_{\{I > a\}} \sum_k (U'_k/v'_k)(I_k/I)]'}$$

where I , U , etc., are evaluated at the optimum. Under this new probability measure, the probability of an event A is evaluated as

$$\tilde{\mathbb{P}}(A) = \mathbb{E}^{\tilde{\mathbb{P}}}[\mathbf{1}_A] = \mathbb{E}\left[\frac{\partial \tilde{\mathbb{P}}}{\partial \mathbb{P}} \mathbf{1}_A\right] = \int_A \frac{\partial \tilde{\mathbb{P}}}{\partial \mathbb{P}} d\mathbb{P}.$$

Hence, in addition to the statistical probability ($d\mathbb{P}$), the new probability is affected by the value placed on recoveries by the various consumers ($\sum_k (U'_k/v'_k)(I_k/I)$) in states of default ($\mathbf{1}_{\{I > a\}}$). This is akin to asset pricing theory, where *risk-neutral probabilities* used for pricing are weighted by the marginal utilities of (risky) consumption of the agents in the economy (e.g., Duffie 2010). In particular, under $\tilde{\mathbb{P}}$ all the probability mass is concentrated in default states, since this is where the consumers in our model are exposed to the risks in the company.⁸

For random variables that are positive in default states, we define the risk measure:

$$\tilde{\rho}(X) = \exp\{\mathbb{E}^{\tilde{\mathbb{P}}}[\log\{X\}]\}.$$

Then, $\tilde{\rho}(X)$ is exactly the risk measure that gives our counterparty-driven capital allocations:

$$q_i \frac{(\partial/\partial q_i)\tilde{\rho}(I)}{\tilde{\rho}(I)} = q_i \mathbb{E}^{\tilde{\mathbb{P}}}\left[\frac{\partial}{\partial q_i} \log\{I\}\right] = q_i \mathbb{E}^{\tilde{\mathbb{P}}}\left[\frac{\partial I_i/\partial q_i}{I}\right] = q_i \tilde{\phi}_i.$$

But what are the mathematical properties of this measure? In what proved to be a hugely influential paper, Artzner et al. (1999) define a risk measure ρ as *coherent* if, in addition to *positive homogeneity*, the measure satisfies the axioms of *monotonicity* ($\rho(X) \leq \rho(Y)$ for $X \leq Y$), *subadditivity* ($\rho(X + Y) \leq \rho(X) + \rho(Y)$), and *translation invariance* ($\rho(X + c) = \rho(X) + c$). The importance of adherence to these axioms, or to closely related sets of axioms that define alternative classes of risk measures (such as *convex* risk measures, where the *homogeneity* and *subadditivity* axioms are replaced with a *convexity* axiom; see Föllmer and Schied 2002), is frequently stressed by both scholars and practitioners.

(Schmock and Straumann 1999, Tasche 2004, McNeil et al. 2005) or game-theoretic considerations (Denault 2001, Powers 2007), as well as from other approaches (Kalkbrenner 2005, Myers and Read 2001). We refer to Bauer and Zanjani (2013) for a detailed overview of the different approaches:

⁸ This will still be the case in a setting with securities markets under the assumption that the insurance market is incomplete, although here, statistical probabilities ($d\mathbb{P}$) are replaced by risk-neutral probabilities ($d\mathbb{Q}$) accounting for aggregate economic risk. We refer to §4.1 and Appendix B.2 for details.

By construction, the new measure $\tilde{\rho}$ obviously is positive homogeneous, the property that allows the convenience of “adding up” for gradient-based allocations. The property that higher losses yield a higher risk causes the risk measure to be monotonic. However, as we will demonstrate in the context of examples in the next section, because of the log transformation, the functional form generally is not subadditive; i.e., it does not always reward diversification. Moreover, it is not translation invariant, because it is readily verified that $\exp\{\mathbb{E}^{\tilde{\mathbb{P}}}[\log\{X + c\}]\} \neq \exp\{\mathbb{E}^{\tilde{\mathbb{P}}}[\log\{X\}]\} + c$, so the measure is neither coherent nor convex.

The log transformation apparently yields unconventional mathematical properties, so it is worth exploring why the underlying economics have produced such a measure. Although higher aggregate loss realizations I are worse for individual consumers, the marginal impact in monetary terms decreases as I increases. The reason lies in the definition of individual recoveries (1), which are a function of one over I in default states—and recoveries in default states are ultimately what consumers care about. For instance, if the aggregate loss increases from a realization exactly at the asset level, $I = a$, to twice the asset level, $I = 2a$, according to Equation (1), a consumer would go from being paid in full to receiving 50 cents on the dollar—a loss of 50% in relation to the promised indemnity. However, when the aggregate loss level increases by the same amount but from $2a$ to $3a$, the same consumer goes from being paid at a rate of 50 cents on the dollar to 33.33 cents—a loss of only 16.7% in relation to the promised indemnity. Similarly, when the loss level increases from $9a$ to $10a$, the impact in relation to the promised indemnity will merely be a single percent. Hence, raising the loss level has a progressively smaller impact on recovery rates, with the consequence that an increase in a particular consumer’s loss will generate smaller absolute externalities on other consumers when the aggregate loss level is higher.

The log transformation reflects this, and it depresses the impact of large loss states (relative to their actual size). Offsetting this, however, is the influence of risk aversion through the probability transform: recoveries are more highly valued by consumers when their wealth levels are low as a result of large losses, meaning that high loss states may be associated with high marginal utilities. It is ultimately the interaction of these two influences that jointly determines the risk measure and the resulting capital allocations. We explore this interplay in more detail in the next section.

3. Comparison of Allocations and Risk Measures

We examine the properties of the risk measure and the associated capital allocations in the context of

two example settings. First, we consider an arbitrary number of homogeneous (independent and identically distributed (iid)) exponential risks, and then we study allocations to two heterogeneous Pareto risks. In particular, we compare the results to those obtained from the most popular risk measures, VaR and ES.

3.1. Homogeneous Exponential Losses

Assume that there are N identical, independent consumers with wealth level w that face independent, exponentially distributed losses: $L_i \sim \text{Exp}(\nu)$, $1 \leq i \leq N$. Assume further that all consumers exhibit constant absolute risk aversion (CARA) utility, with coefficient of absolute risk aversion $\alpha < \nu$, and that the participation constraint of each is given by the autarky level:

$$\gamma = \gamma_i = \mathbb{E}[U(w - L_i)] = -e^{-\alpha w} \frac{\nu}{\nu - \alpha}.$$

As detailed in Appendix A.2, the expected values in the firm's profit maximization problem can be evaluated in closed form in this setting. Since consumers and losses are symmetric, optimal premium and coverage amounts are the same across consumers, so the optimization problem reduces to three parameters— a , p , and q .

3.1.1. Capital Allocations. Any reasonable allocation rule will yield identical allocations of $1/N$ of the firm's capital to each identical consumer. However, we may analyze *how* different allocation methods arrive at this result by comparing the implicit weight each puts on different (aggregate) loss states. As discussed in the previous section, there are two aspects that influence the economic capital allocations: recoveries and marginal utilities in default states. Since there is no asset risk, the expected value of recoveries for each consumer, $\mathbb{E}[a(L^i/L)] = a/N$, is simply the assets divided by the number of consumers and is constant across aggregate loss states. However, risk aversion will cause the consumers to value recoveries more highly in severe states of default.

Formally, we obtain

$$q_i \tilde{\phi}_i = \mathbb{E}^{\tilde{\mathbb{P}}} \left[\frac{L_i}{L} \right] = \frac{1}{N} \mathbb{E} \left[\underbrace{\mathbb{E} \left[\frac{\partial \tilde{\mathbb{P}}}{\partial \mathbb{P}} \middle| L \right]}_{\tilde{\psi}(L)} \right] = \frac{1}{N},$$

and we show in Appendix A.2 that $\tilde{\psi}$, which reflects the relative values that consumers place on recoveries in default states, can be represented as

$$\tilde{\psi}(L) = \mathbf{1}_{\{qL > a\}} \hat{c}_{N, \nu, \alpha, a, q} \sum_{k=0}^{\infty} \frac{(k+1)(\alpha(L-a))^k}{(N+k)!},$$

where $\hat{c}_{N, \nu, \alpha, a, q}$ is a constant ensuring that $\mathbb{E}[\tilde{\psi}(L)] = 1$. Thus, in the (limiting) case of risk-neutral consumers ($\alpha = 0$), $\tilde{\psi}(L)$ will be constant across default states,

Table 1 Parametrizations of the Exponential Losses Model

Exponential losses: Parametrizations								
No.	N	ν	τ	α	w	a	p	q
1	5	2.0	0.050	0.25	3.0	1.4663	0.2598	0.5713
2	5	2.0	0.050	1.25	3.0	4.0036	0.7401	0.9494

and all default states, no matter how severe, will be weighted evenly. In the general case when consumers are risk averse ($\alpha > 0$), however, larger consumer losses associated with larger aggregate loss states yield higher marginal utilities. Thus, under risk aversion, $\tilde{\psi}(L)$ is increasing in the aggregate loss state L so that recoveries in severe states of default are weighted more heavily than those in mild ones.

For gradient allocations based on VaR or ES with confidence level ε , it is well known that (e.g., McNeil et al. 2005)

$$\begin{aligned} q_i \frac{\partial \text{VaR}_\varepsilon(I)/\partial q_i}{\text{VaR}_\varepsilon(I)} &= \frac{\mathbb{E}[L_i | I = \text{VaR}_\varepsilon(I)]}{\text{VaR}_\varepsilon(I)} \\ &= \frac{1}{N} \underbrace{\mathbb{E}[\text{const} \times L | qL = a]}_{=1}, \end{aligned}$$

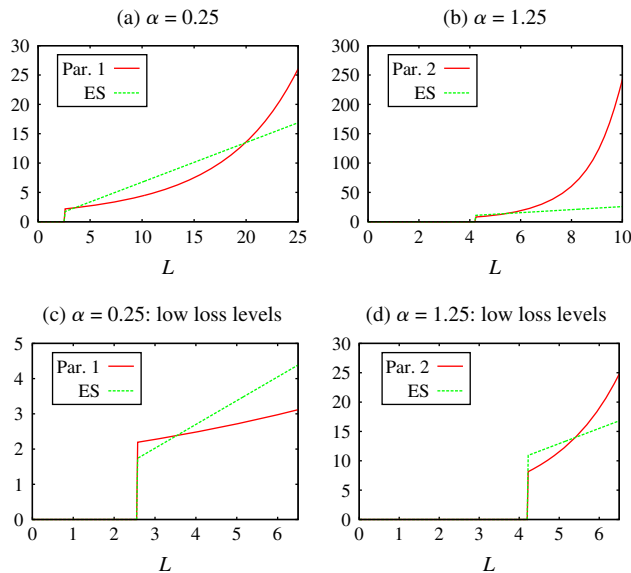
and

$$\begin{aligned} q_i \frac{\partial \text{ES}_\varepsilon(I)/\partial q_i}{\text{ES}_\varepsilon(I)} &= \frac{\mathbb{E}[L_i | I \geq \text{VaR}_\varepsilon(I)]}{\mathbb{E}[I | I \geq \text{VaR}_\varepsilon(I)]} \\ &= \frac{1}{N} \underbrace{\mathbb{E}[\text{const} \times L | qL \geq a]}_{=1}, \end{aligned}$$

where we choose $\varepsilon = \mathbb{P}(I > a)$ as the default probability. As is evident from the former equation for VaR, only the aggregate loss level at the cutoff state is material for the allocations: the size of the loss in states of default is irrelevant. This is a familiar feature of VaR observed in the literature before (e.g., Basak and Shapiro 2001). By contrast, ES corresponds to a linear weighting of the aggregate loss in default states so that severe loss states are weighted more heavily than mild ones. Comparing ES to the economic allocations, the weighting function $\tilde{\psi}$ associated with the economic allocations in this case is convex when consumers are risk averse, and there will always be a threshold loss level after which point the economic weighting exceeds the ES weighting.⁹

To compare the two over smaller loss states, in Table 1 we present two parametrizations of the setup with different risk-aversion levels and the corresponding optimal parameters a , p , and q . The properties are as expected: a , p , and q all are increasing in risk aversion.

⁹ The convexity of $\tilde{\psi}$ is not a general feature but depends on preference specifications and the underlying loss distributions. In fact, it is possible to construct examples where the shape is linear or even strictly concave, so that ES overweights rather than underweights severe loss states.

Figure 1 (Color online) Weighting Function $\tilde{\psi}$ Within the Economic Allocations $q_i \tilde{\phi}_i = \mathbb{E}[\tilde{\psi}(L)]/N$ as a Function of the Aggregate Loss State L for the Two Parametrizations

Note. Shown are the two parametrizations with different risk aversion parameters α (red solid curves). For comparison, we also plot the linear weighting functions associated with ES allocations (green dashed curves).

Figure 1 plots the corresponding weighting functions $\tilde{\psi}$ against the linear weighting function associated with ES. Panel (a) shows results for parametrization 1, and panel (b) shows results for parametrization 2. Panels (c) and (d) depict the same graphs focusing on relatively small loss states.

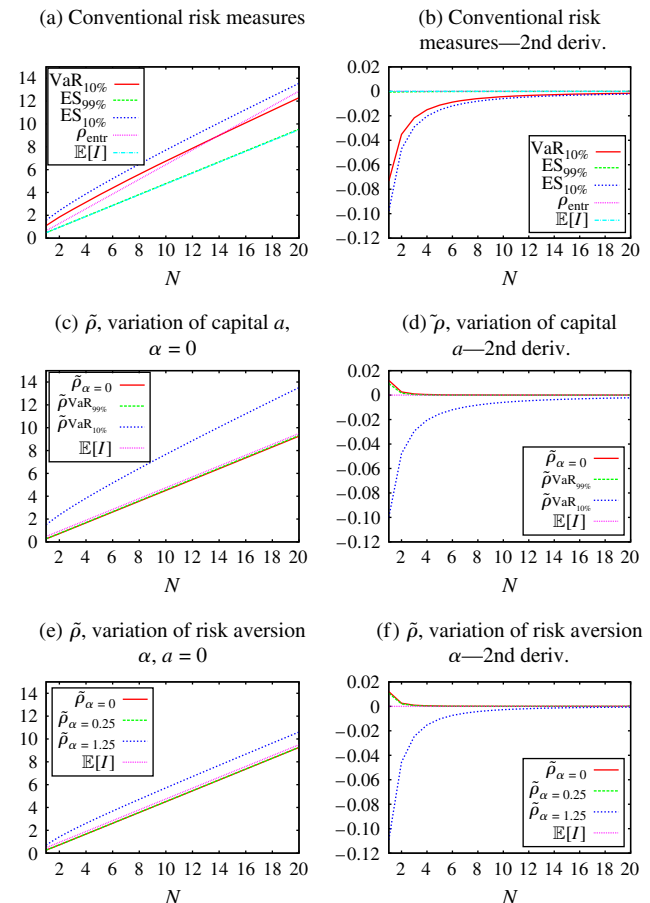
We find two qualitatively different shapes:¹⁰ For the high-risk aversion level, $\tilde{\psi}$ crosses the linear weighting function once from below; thus, in this case, relatively lower loss states are weighted more heavily by the ES allocation, whereas the weighting is higher for the economic allocation in high loss states. For the low-risk aversion level, $\tilde{\psi}$ crosses the linear weighting function twice. In this case, the weighting function within the economic allocation puts more mass on low and extremely high loss states, whereas the weights are smaller for intermediate to high loss states.

3.1.2. Risk Measures. For the risk measure $\tilde{\rho}$ supporting the economic allocations, the weighting function $\tilde{\psi}$ again plays an important role. In particular, when evaluating $\tilde{\rho}$ at the aggregate loss I , we obtain

$$\tilde{\rho}(I) = \exp\{\mathbb{E}^{\tilde{\mathbb{P}}}[\log\{qL\}]\} = \exp\{\mathbb{E}[\tilde{\psi}(L) \log\{qL\}]\}.$$

Thus, since $\tilde{\psi}$ places all probability mass on states of default, $\tilde{\rho}$ in this sense is, similar to ES, a *tail risk measure*. The form is similar to so-called *spectral risk measures* (Acerbi 2002), where higher weights are placed on bad outcomes via a weighting function referred to

¹⁰ Analyses with respect to other parameters, particularly expected loss size $1/\nu$, show similar results.

Figure 2 (Color online) Comparison of Risk Measures ρ of the Aggregate Loss $I_N = \sum_{i=1}^N I_i$ as a Function of the Portfolio Size N 

Notes. The panels on the left-hand side depict $\rho(I_N)$ for conventional risk measures (panel (a)) as well as for the new risk measure given different capital levels under risk neutrality (panel (c)) and given different risk aversion levels with a capital level of zero (panel (e)). The panels on the right-hand side give approximations of the corresponding second derivatives, where a concave shape (negative second derivative) corresponds to subadditivity, whereas a convex shape (positive second derivative) corresponds to superadditivity of the portfolios.

as the *risk spectrum*—which is supposed to encode the “subjective risk aversion of an investor” (p. 1510).

When analyzing the properties of $\tilde{\rho}$ and when comparing it to conventional risk measures, it is necessary to evaluate the same measure for different, *given* random variables.¹¹

In Figure 2, we present comparisons of risk measures by evaluating them at the aggregate loss $I_N = \sum_{i=1}^N qL_i$

¹¹ This, in some sense, is in contrast to the spirit of its construction, where the parameters, and especially the aggregate loss $I = qL$, arise *endogenously* as the solution of the firm’s profit maximization problem. Nonetheless, a comparison is feasible when fixing the parameters, and we will present corresponding results in what follows. In particular, this level of complication is also present for other risk measures used in practice since parameters, such as the confidence level for VaR or ES, or the risk spectrum within a spectral risk measure, have to be chosen exogenously.

for different portfolio sizes N . Panel (a) shows corresponding results for five well-known measures (see Appendix A.2 for their derivation): the expected value $\mathbb{E}[I]$, VaR at the 10% level, ES at the 99% level, ES at the 10% level, and the entropic risk measure $\rho(X) = \log\{\mathbb{E}[\exp\{X\}]\}$ —as an example of a convex risk measure that is not coherent (Föllmer and Schied 2002, 2010). We assume throughout that the participation level is set at $q = 0.94$ and $\nu = 2.0$ as in parametrization 2 from Table 1. Panel (b) presents second differences,

$$[\rho(I_{N+2}) - \rho(I_{N+1})] - [\rho(I_{N+1}) - \rho(I_N)],$$

for each risk measure as an approximation of the second derivative. This expression is closely linked to the measure's diversification properties and thus to subadditivity. The first term corresponds to the additional risk when moving from a portfolio size of $N + 1$ to $N + 2$, whereas the second term corresponds to the additional risk when moving from a portfolio size of N to $N + 1$. When the first term is less than the second term—so in case the second difference is negative—the measure rewards diversification since the same risk is perceived less precarious in the context of a larger portfolio. In particular, concavity of the measure as a function of N —i.e., a negative second derivative—implies subadditivity for the portfolios.

As can be seen from panels (a) and (b) of Figure 2, VaR and ES result in a concave shape, although the second difference is barely negative for ES at 99% because of the high confidence level. This is not surprising since both measures are generally subadditive, excepting “extreme” situations in the case of VaR (Dánielsson et al. 2013). By contrast, the other two measures yield a linear progression. This is obviously the case for the expected value, and also for the entropic risk measure, it is well known that it is additive for independent risks (Föllmer and Knispel 2011).

Panels (c)–(f) of Figure 2 provide results for the new risk measure $\tilde{\rho}$ under different parameter choices. Appendix A.2 shows that it is possible to derive closed-form expressions. For instance, in the case of risk neutrality ($\alpha = 0$) and zero assets ($a = 0$), such that the probabilities \mathbb{P} and $\tilde{\mathbb{P}}$ coincide, the measure can be evaluated as

$$\tilde{\rho}(I_N) = \exp\{\mathbb{E}[\log\{I_N\}]\} = \frac{q}{\nu} \exp\{\Psi(N)\},$$

where $\Psi(\cdot)$ is the digamma function. The measure and the corresponding second difference in this case are illustrated as the solid lines in panels (c) and (e) and (d) and (f), respectively. The second difference is positive, meaning that $\tilde{\rho}(I_N)$ is a convex function in N so that here we have superadditivity for the portfolios (i.e., diversification is penalized). This shows again that, in general, the new risk measure does not satisfy the axioms of coherence and convexity.

However, as noted earlier, the parameters of the risk measure flow from the solution of an optimization problem and thus differ for different portfolio sizes. For instance, a larger portfolio will typically be associated with more assets a held in the firm. Panels (c) and (d) of Figure 2 show results for the new risk measure when the asset level a is cast by VaR with different confidence levels, so that similarly to ES, only loss realizations beyond that level are relevant. The risk aversion parameter is still set to risk neutrality ($\alpha = 0$). We find that for the high confidence level (99%), the second difference is still positive for small values of N , whereas it becomes slightly negative for large portfolios. Hence, the shape for $\tilde{\rho}_{\text{VaR}_{99\%}}(I_N)$ changes from convex to concave—or superadditive to subadditive—as N increases. By contrast, the graph for the low confidence level (10%) is concave throughout, so that the measure is subadditive and diversification is rewarded.

In Figure 2, panels (e) and (f), we vary the risk aversion coefficient α . Note that then for levels $\alpha > 0$, the portfolio size N also enters the weighting function $\tilde{\psi}$ in the definition of $\tilde{\rho}$. The asset level a is again set to zero. For a relatively small risk aversion coefficient ($\alpha = 0.25$), the second difference remains slightly positive and $\tilde{\rho}(I_N)$ is still convex in N , so that the risk measure is still superadditive. However, for a larger risk aversion level ($\alpha = 1.25$), the second difference is negative and the risk measure is subadditive.

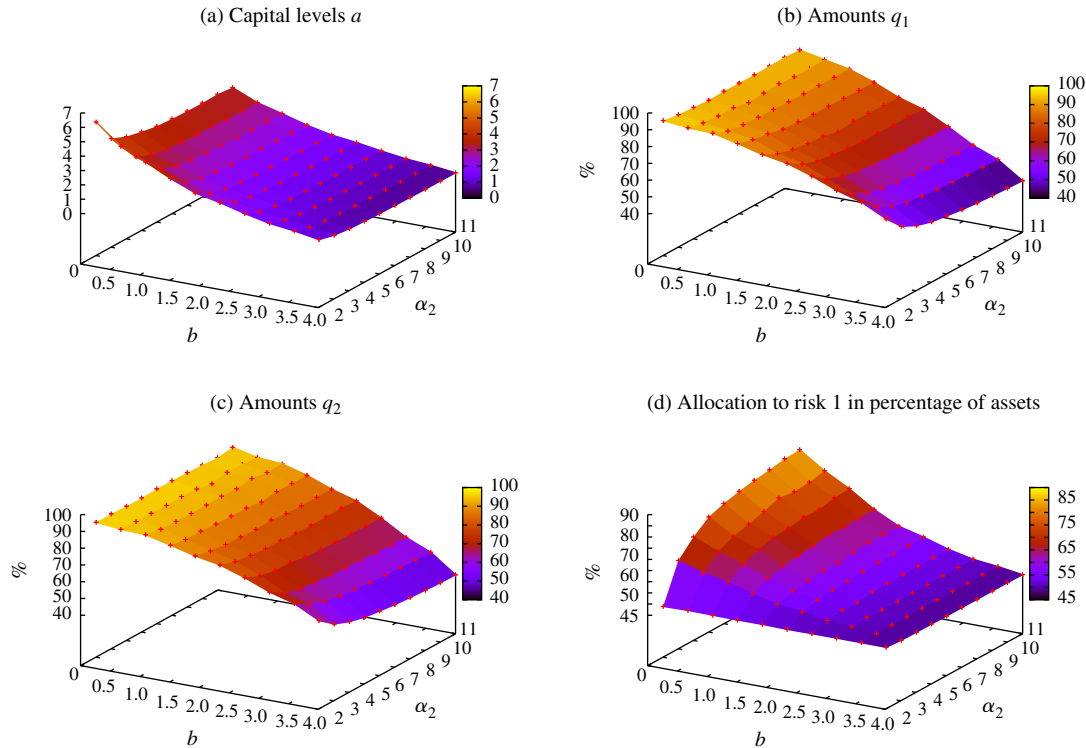
Hence, counterparty risk aversion affects the measure by motivating the company to hold more capital to avert default and by increasing the relative weighting of severe loss states. Both aspects push the measure toward subadditivity so that diversification is increasingly rewarded under $\tilde{\rho}$.

3.2. Heterogeneous Pareto Losses

In this section, we consider two consumers facing different (independent) Pareto II (Lomax) distributed losses with scale parameters σ_1 and σ_2 and tail indices α_1 and α_2 , respectively. This example is relevant in at least two regards. First, in contrast to the previous example where homogeneous risks led to identical allocations, a setting with heterogeneous risks allows for a direct analysis of differing allocations. Second, heavy-tailed power law distributions have an important role in describing natural and economic phenomena, particularly in insurance (Embrechts et al. 1997, Gabaix 2009, Ibragimov 2009a, among others), and in some cases heavy tails can lead to qualitatively different results for portfolio diversification and for the properties of risk measures such as VaR (Ibragimov 2009b, Ibragimov et al. 2009, Dánielsson et al. 2013, among others).

We assume that the consumers have quadratic utility functions

$$U_i(x) = -x^2 + 2b_i x, \quad b_i \geq 0,$$

Figure 3 (Color online) Optimal Capital Level a , Optimal Coverage Levels q_1 and q_2 , and Economic Capital Allocations to Risk 1 as a Function of the Tail Index of Risk 2 α_2 and the Risk Aversion Parameter b (Identical for Both)

Note. The tail index of risk 1 is fixed at 2.05, and expected values for both risks are 0.5.

and normalize wealth w_i to zero so that the utility is formed over negative arguments. Absolute risk aversion,

$$RA(x) = -U_i''(x)/U_i'(x) = \frac{1}{b_i - x},$$

is decreasing in the risk aversion parameter b_i . Again, we assume that the participation constraint for each consumer is given by the autarky level:

$$\gamma_i = \mathbb{E}[U(-L_i)] = -\frac{2\sigma_i^2}{(\alpha_i - 2)(\alpha_i - 1)} - \frac{2b_i\sigma_i}{\alpha_i - 1}.$$

Note that for utilities to be defined, we require tail indices to be greater than 2.

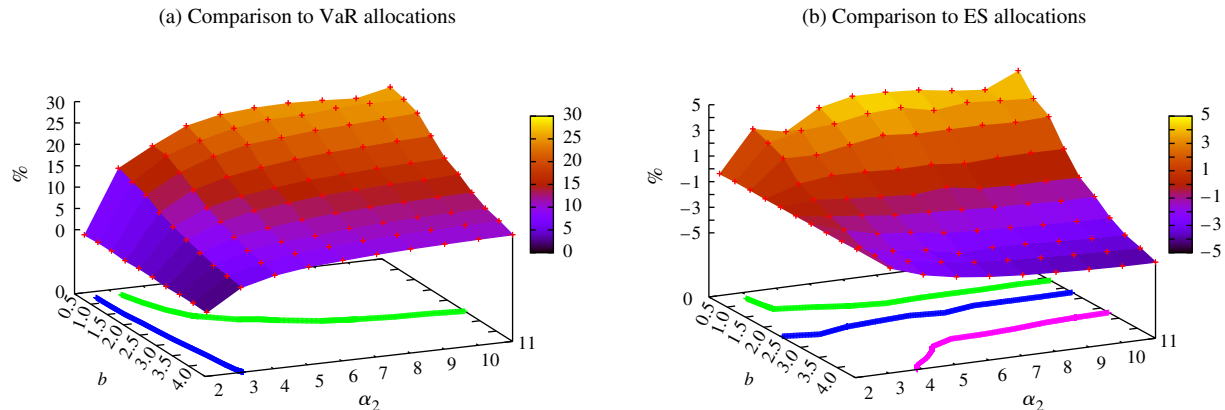
We parametrize the model by fixing $\tau = 0.05$, the scale parameter $\sigma_1 = 0.5025$, and the tail index $\alpha_1 = 2.05$, so that the expected value of risk 1 is $\mathbb{E}[L_1] = \sigma_1/(\alpha_1 - 1) = 0.5$. For risk 2, we vary the tail index α_2 between 2.05 and 11, where we set the corresponding scale parameters to $\sigma_2 = 0.5(\alpha_2 - 1)$. Thus, the resulting expected value is also 0.5 so that allocations will not be driven by differences in expected values. We assume both consumers have identical risk aversion parameters $b = b_1 = b_2$, which we vary between 0.25 and 4. Appendix A.3 provides details on the solution approach.

Figure 3 shows resulting optimal asset levels a , optimal participation levels q_1 and q_2 for the consumers,

and economic capital allocations to risk 1 in percentage of the assets as functions of the risk aversion parameter b and the tail index for risk 2 α_2 . A higher tail index for risk 2 corresponds to less overall risk in the portfolio requiring less capital, and a larger b means less risk-averse consumers that, ceteris paribus, are less concerned with the nonperformance of their insurance contract. Thus, we find that the optimal asset level is decreasing in the tail index α_2 and decreasing in the risk aversion parameter b (panel (a)). Similarly, participations q_1 and q_2 decrease in the risk aversion parameter—i.e., higher risk aversion corresponds to more insurance coverage (panels (b) and (c)). Moreover, optimal participation levels are also decreasing in the tail index α_2 because of a diminishing willingness of consumer 2 to bear capital costs.

3.2.1. Capital Allocations. Figure 3, panel (d) plots economic capital allocations to risk 1 in percentage of the assets. For most considered parameter combinations, the allocations exceed half of the assets, which is not surprising since consumer 1 faces more tail risk. The allocations to risk 1 typically increase in α_2 —i.e., as risk 2 becomes less heavy tailed—whereas, obviously, allocations are 50% for $\alpha_2 = 2.05$ since both consumers are identical in this case. Exceptions occur for relatively large b , corresponding to a low level of risk aversion. Although here risk 1 is more heavy-tailed, consumer 1's participation level decreases by

Figure 4 (Color online) Difference in Capital Allocations to Risk 1 as a Percentage of Assets Between the Economic Allocation Rule and the VaR/ES Gradient Allocation Rules as a Function of the Tail Index of Risk 2 α_2 and the Risk Aversion Parameter b (Identical for Both)



Notes. The tail index of risk 1 is fixed at 2.05. Contour lines in (a) are fixed at 5% (dark line) and 15% (light line). Contour lines in (b) are fixed at -2.5% (medium line), 0% (dark line), and 2.5% (light line).

more than consumer 2's participation—i.e., consumer 1 purchases less insurance—leading to capital allocations for consumer 1 of less than 50%. Nonetheless, the riskier consumer 1 is still allocated a larger share of capital relative to participation. For instance, for a risk aversion parameter of $b = 4$ and a tail index $\alpha_2 = 11$, the ratio of participation levels q_1/q_2 is about 90% whereas the ratio of capital allocations $(q_1\phi_1)/(q_2\phi_2)$ is approximately 96%, so that ϕ_1/ϕ_2 is greater than 1.

Gradient capital allocations associated with VaR or ES generally exhibit similar characteristics. In particular, allocations to risk 1 increase in the tail index α_2 relative to participation. Moreover, the allocations are increasing in risk aversion, although here, the effect is *indirect* in the sense that it stems solely from changes in the participation level: in contrast to the economic risk measure, VaR and ES themselves are unaffected by changes in b .

We compare the economic capital allocations to those obtained from VaR and ES. Figure 4 plots the difference between the percentage of capital allocated to risk 1 under the economic allocation and under the VaR allocation in panel (a) and the difference for ES in panel (b). As in the previous section, we choose the confidence level as the default probability $\mathbb{P}(I > a)$ for each parameter combination to facilitate comparison.

Economic allocations to risk 1 always exceed VaR allocations with a difference as large as 26.4% in the considered parameter range. The differences are increasing in α_2 (i.e., as risk 2 gets lighter tailed) whereas, obviously, the differences are zero for $\alpha_2 = 2.05$, at which point the risks are identical. The differences are decreasing in the risk aversion parameter, so the allocations are closer for low risk aversion levels although differences are still substantial (e.g., close to 10% for $b = 4$ and $\alpha_2 = 11$). The reason, as discussed in the previous section, is that for VaR only the cutoff loss states at the confidence level—which is fixed at the

default probability for Figure 4—are material, whereas economic allocations weight larger loss states more heavily, particularly for high risk aversion levels. Even when we decrease the confidence level to half of the default probability, the economic allocations to risk 1 still always exceed the corresponding VaR allocations in our parameter range, although the differences are smaller (between 0 and 16.6%). This is due to the heavy tails of the loss distributions: the higher weight put on large loss states, combined with the significant probability of achieving these large loss states, severely penalizes the loss with the fatter tails even when risk aversion is relatively low.

The situation changes considerably for ES, which places greater weights on larger loss states. Here, we are able to see the two opposing influences on the economic allocations. For low risk aversion levels (large b), the weighting function is less important, and the use of ES results in higher allocations to the heavier-tailed risk 1 as a result of the focus on loss states rather than recoveries. For high risk aversion levels, on the other hand, the heavy economic weighting of severe loss states causes the economic allocations to exceed the corresponding ES allocations. For the considered parameters, the difference can be as small as -4.05% for $\alpha_2 = 11/b = 4$ (low risk aversion) and as large as 4.98% for $\alpha_2 = 11/b = 0.25$ (high risk aversion). By contrast, these competing effects offset each other for medium risk aversion parameters, as can be seen by the 0% contour line (dark line in Figure 4).¹² Hence, we find that ES-based allocations can put too much weight on the heavy-tailed risk 1 if consumers are not very risk averse. On the other hand, if consumers are

¹² Obviously, parameters on the b axis (where $\alpha_2 = \alpha_1 = 2.05$) yield 50% for both allocations so that the difference is zero here as well, but for simplicity, we did not include this segment in the contour line.

Table 2 Comparison of Risk Measures for Pareto Losses as a Function of the Tail Index (α , Identical for the Two Risks) and the Scale Parameter (σ , Identical for the Two Risks)

Pareto losses: Risk measures													
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
α	σ	VaR _{5%} One	Sum	Frac.	ES _{5%} One	Sum	Frac.	$\tilde{\rho}_{a=0}$ One	Sum	Frac.	$\tilde{\rho}_{VaR_{5\%}}$ One	Sum	Frac.
0.50	0.14	55.09	220.5	4.00	NA	NA	NA	0.55	2.65	4.80	407.8	1,649	4.04
0.75	0.27	14.52	38.20	2.63	NA	NA	NA	0.45	1.73	3.81	55.69	143.4	2.58
1.00	0.41	7.87	17.17	2.18	NA	NA	NA	0.41	1.42	3.43	21.95	46.13	2.10
1.25	0.56	5.58	11.04	1.98	30.14	54.06	1.79	0.39	1.28	3.24	12.97	24.40	1.88
1.50	0.71	4.49	8.38	1.87	14.88	25.07	1.68	0.38	1.20	3.13	9.29	16.36	1.76
1.75	0.85	3.87	6.96	1.80	10.16	16.47	1.62	0.37	1.14	3.06	7.39	12.48	1.69
2.00	1.00	3.47	6.10	1.76	7.94	12.55	1.58	0.37	1.11	3.01	6.26	10.28	1.64
2.25	1.15	3.20	5.52	1.73	6.68	10.37	1.55	0.36	1.08	2.97	5.53	8.90	1.61
2.50	1.30	3.00	5.11	1.70	5.87	9.00	1.53	0.36	1.06	2.94	5.01	7.95	1.59
2.75	1.44	2.85	4.80	1.69	5.30	8.07	1.52	0.36	1.04	2.91	4.64	7.28	1.57
3.00	1.59	2.73	4.57	1.67	4.89	7.40	1.51	0.36	1.03	2.90	4.35	6.77	1.56

α	σ	VaR _{1%} One	Sum	Frac.	ES _{1%} One	Sum	Frac.	$\tilde{\rho}_{a=0}$ One	Sum	Frac.	$\tilde{\rho}_{VaR_{1\%}}$ One	Sum	Frac.
0.50	0.14	1,381	5,528	4.00	NA	NA	NA	0.55	2.65	4.80	10,203	39,483	3.87
0.75	0.27	126.2	321.3	2.55	NA	NA	NA	0.45	1.73	3.81	479.5	1,209	2.52
1.00	0.41	41.01	84.17	2.05	NA	NA	NA	0.41	1.42	3.43	112.0	226.4	2.02
1.25	0.56	21.69	39.52	1.82	110.7	194.7	1.76	0.39	1.28	3.24	48.83	87.10	1.78
1.50	0.71	14.49	24.57	1.70	44.87	72.89	1.62	0.38	1.20	3.13	28.77	47.48	1.65
1.75	0.85	10.99	17.80	1.62	26.78	41.33	1.54	0.37	1.14	3.06	20.01	31.38	1.57
2.00	1.00	9.00	14.14	1.57	19.00	28.33	1.49	0.37	1.11	3.01	15.38	23.32	1.52
2.25	1.15	7.74	11.90	1.54	14.85	21.62	1.46	0.36	1.08	2.97	12.62	18.67	1.48
2.50	1.30	6.88	10.43	1.51	12.34	17.65	1.43	0.36	1.06	2.94	10.81	15.72	1.45
2.75	1.44	6.27	9.39	1.50	10.67	15.09	1.41	0.36	1.04	2.91	9.56	13.72	1.43
3.00	1.59	5.80	8.62	1.49	9.50	13.31	1.40	0.36	1.03	2.90	8.65	12.29	1.42

Note. The three columns for each risk measure ρ display $\rho(I_1)$ (the “One” columns), $\rho(I_1 + I_2)$ (the “Sum” columns), and $\rho(I_1 + I_2)/\rho(I_1)$ (the “Frac.” columns).

very risk averse, the economic approach allocates more to the high-risk group than does ES.

3.2.2. Risk Measures. When comparing risk measures, the same caveat applies as in the previous example: we have to choose and hold fixed parameters that are determined endogenously in the firm’s profit maximization problem. In what follows, we analyze the influence of the tail index. In particular, we also examine extreme levels where the variance or the expected value of the losses may not exist.

Table 2 displays risk measures for two risks with identical tail indices $\alpha = \alpha_1 = \alpha_2$ varying between 0.5 and 3. Here, we set the scale $\sigma = \sigma_1 = \sigma_2$ for the tail index $\alpha = 2$ to one in order to obtain an expected loss of 0.5 for ease of comparison with the allocations in the previous part. Then, we choose the scale parameters for all other tail indices to match the median value for $\alpha = 2/\sigma = 1$, since the expected value may not exist. The participation level $q = q_1 = q_2$ is set to 1 for simplicity. For each risk measure ρ , we calculate the measure $\rho(I_1) = \rho(I_2)$ on a stand-alone basis (the “One” columns), the risk measure of the sum of the risks $\rho(I_1 + I_2)$ (the “Sum” columns), and their quotient $\rho(I_1 + I_2)/\rho(I_1)$ (the “Frac.” columns). Obviously, this quotient will always be greater than 1 for a monotonic

risk measure, and it will be smaller (greater) than 2 if the risk measure is subadditive (superadditive)—since under subadditivity, the risk of the diversified portfolio $\rho(I_1 + I_2)$ should be less than twice the risk for each loss on a stand-alone basis $2\rho(I_1)$.

Columns (3)–(5) for the top and bottom parts of Table 2 display results for VaR at confidence levels 5% and 1%, respectively. Columns (6)–(8) show corresponding results for ES. All risk measures increase as the risks become heavier tailed, i.e., as α becomes smaller. The well-known subadditivity of ES is evident, as the quotient $\rho(I_1 + I_2)/\rho(I_1)$ is always less than 2 for ES. However, ES is not defined for tail indices less than or equal to 1, since here, the expected value does not exist. By contrast, as pointed out by Ibragimov (2009b, Remark 4.4), VaR_ε is subadditive for tail indices $\alpha > 1$ and sufficiently small confidence levels ε , whereas it is superadditive for $\alpha < 1$ and sufficiently small ε . Thus, we observe quotients of less than 2 for $\alpha > 1$ —just as for ES—and we only have quotients greater than 2 for $\alpha \leq 1$ —where ES is not defined.¹³

¹³ In our setting, for $\alpha = 1$, we have superadditivity, although the fraction converges to 2 for $\varepsilon \rightarrow 0$. For instance, for $\varepsilon = 0.1\%$, we obtain $\rho(I_1 + I_2)/\rho(I_1) = 2.0090$, and for $\varepsilon = 0.01\%$, it is 2.0032.

Columns (9)–(11) of Table 2 show results for the new risk measure $\tilde{\rho}$, where we set the asset level $a = 0$ and assume risk neutrality so that the probabilities \mathbb{P} and $\tilde{\mathbb{P}}$ coincide (note that the results in the top half and bottom half of the table are identical). The risk measures increase as the risks become heavier tailed on a stand-alone basis as well as for the portfolio, but the quotient $\rho(I_1 + I_2)/\rho(I_1)$ now is greater than 2 for all considered parameters. This is similar to the previous example (in §3.1.2), where we also obtained superadditivity in the case $\mathbb{P} = \tilde{\mathbb{P}}$.

This is no longer the case when we cast the asset level a via VaR—accounting for the idea that a riskier portfolio will typically be associated with a larger amount of assets held in the firm (see also Figure 3, panel (a)). Corresponding results are displayed in the last three columns of the top and bottom parts of Table 2 for confidence levels 5% and 1%, respectively. Again, risk levels are higher for lower tail indices. Moreover, in analogy to VaR, we observe quotients less than 2—associated with subadditivity—for tail indices greater than 1, whereas we have quotients greater than 2—associated with superadditivity—for tail indices less than or equal to 1. It is worth noting that the quotients tend to fall between those for VaR and ES, suggesting that the risk measure $\tilde{\rho}$ is *less subadditive* than ES and *more subadditive* than VaR (or *less superadditive* than VaR, depending on the tail index). Here, as with ES, the size of the loss beyond the cutoff state is material—although, in contrast to ES, the log transformation compresses high loss states and leads the risk measure to exist even for extreme tail indices. Therefore, when endowing the risk measure with an exogenous asset level originating from VaR and under risk neutrality, the measure appears to fall in between VaR and ES with respect to diversification properties.

4. Extensions

In this section, we discuss various extensions to the basic model setup presented in §2. In particular, we discuss the impact of securities markets (§4.1) as well as regulatory constraints and multiple periods (§4.2) on marginal-cost allocations and the corresponding risk measure.¹⁴

4.1. Allocation in a Securities Market Equilibrium

To this point, we have ignored hedging opportunities associated with securities markets. Such opportunities could be important for the insurer in the course of managing the risks in its liability portfolio and for consumers in managing the risks associated with

insurer default and any residual uninsured loss risk. We now allow both the insurer and consumers to invest in (risky) capital market assets with the goals of maximizing insurer value and consumer utility, respectively.

Of course, integrating a securities market requires a considerable extension of the basic setup introduced in §2. We refer to Appendix B.2 for details on corresponding assumptions and derivations. However, despite these differences, the resulting allocations to consumer i and the supporting risk measure closely resemble the previous results:

$$\begin{aligned} q_i \tilde{\phi}_i &= q_i \frac{\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{I > A\}} ((\partial I_i / \partial q_i) / I) \sum_k (U'_k / \mathbb{E}[U'_k | S]) (I_k / I) A]}{\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{I > A\}} \sum_k (U'_k / \mathbb{E}[U'_k | S]) (I_k / I) A]} \\ &= q_i \frac{(\partial / \partial q_i) \tilde{\rho}(I)}{\tilde{\rho}(I)}, \end{aligned} \quad (8)$$

where

$$\begin{aligned} \tilde{\rho}(X) &= \exp\{\mathbb{E}^{\tilde{\mathbb{P}}}[\log\{X\}]\} \quad \text{and} \\ \frac{\partial \tilde{\rho}}{\partial \mathbb{P}} &= \frac{(\partial \mathbb{Q} / \partial \mathbb{P}) \mathbf{1}_{\{I > A\}} \sum_k (U'_k / \mathbb{E}[U'_k | S]) (I_k / I) A}{\mathbb{E}[(\partial \mathbb{Q} / \partial \mathbb{P}) \mathbf{1}_{\{I > A\}} \sum_k (U'_k / \mathbb{E}[U'_k | S]) (I_k / I) A]}. \end{aligned}$$

In comparison with the capital allocations $q_i \tilde{\phi}_i$ in the basic setup, the company can now allocate its asset portfolio to securities market states, so that the asset level is now a securities market state-dependent random variable (denoted by A instead of the fixed a). Similarly, consumer marginal utilities will reflect optimal wealth allocations, and thus consumers will measure their marginal utilities U'_k of insurance payments relative to expected marginal utilities given the corresponding payoffs by marketed securities S , $\mathbb{E}[U'_k | S]$. In turn, capital allocations as well as the supporting risk measure now account for the aggregate economic (market) risk by assessing securities market states by their risk-neutral probabilities \mathbb{Q} . However, the form of the capital allocation result does not change: the basis for the allocations is still consumer recoveries measured relative to state weights associated with counterparty marginal utilities in insurer default states, since this is where the consumers face *nonhedgable* risk.

In the limiting case of a complete market, the result reduces to allocating capital cost to consumers in proportion to their share of the total market value of recoveries, which is exactly the allocation result in Ibragimov et al. (2010). It is important to note, however, that in the complete market case, purchasing protection from an insurance company with costly capital is inefficient since consumers can hedge insurance risk themselves.¹⁵

¹⁴ The extension when the company exhibits risk aversion is treated in Appendix B.1. Since the allocation result in this case is identical to §2, we do not discuss it separately in the main text. See also Footnote 6.

¹⁵ Ibragimov et al. (2010) deal with this by assuming that “the insureds do not have direct access to the market for risk” (p. 556), whereas the insurer faces a “friction-free complete market for risk” (p. 555).

4.2. Multiple Periods and Regulatory Constraint

In this section, we generalize the original setup to include multiple periods and an external regulatory constraint. We refer to Appendix B.3 for details and derivations but lay out the basic structure here. To keep things tractable, we retain the flat cost of holding capital but do not introduce any new costs of raising capital and assume further that the firm defaults if its liabilities exceed its assets. With these and other simplifying assumptions, the problem becomes stationary, and the complication introduced by multiple periods is simply that if the firm defaults, it loses future profit opportunities. Defining the value of the firm as V , we can then write the Bellman equation:

$$V = \max_{a, \{q_i\}, \{p_i\}} \left\{ \sum_i p_i - \sum_i \mathbb{E}[R_i] - \tau a + \mathbb{P}(I \leq a) V \right\}$$

subject to the (period) participation constraint

$$v_i(a, w_i - p_i, q_1, \dots, q_N) \geq \gamma_i \quad \forall i$$

and a new regulatory constraint

$$s(q_1, \dots, q_N) \leq a,$$

where s is an externally supplied risk measure with a set threshold dictating the requisite capitalization in each period for the firm as a function of the risk it has taken.

With respect to allocating capital costs, multiple sources of discipline factor into the firm's capitalization decision in this extended setting. In addition to the concerns about counterparties that drove the analysis of §2, capital is desired to satisfy the external regulatory constraint and to protect the shareholder's access to future profit flows. Capital allocations will reflect all three influences.

We obtain the following marginal cost of risk for consumer i :

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} = & \underbrace{\mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U'_i}{v'_i} \frac{a I_i}{I} \frac{\partial I_i / \partial q_i}{I} \right]}_{(I)} \\ & + \tilde{\theta}_i [V f_I(a)] + \frac{\partial s}{\partial q_i} \left[\mathbb{P}(I > a) + \tau - \sum_k \frac{\partial v_k / \partial a}{v'_k} - V f_I(a) \right] \\ & + \underbrace{\tilde{\phi}_i a \left[\sum_k \frac{\partial v_k / \partial a}{v'_k} \right]}_{(II)}, \end{aligned} \quad (9)$$

where

$$\begin{aligned} \tilde{\theta}_i &= \mathbb{E} \left[\frac{\partial I_i}{\partial q_i} \mid I = a \right], \\ \tilde{\phi}_i &= \frac{\mathbb{E}[\mathbf{1}_{\{I > a\}} ((\partial I_i / \partial q_i) / I) \sum_k (U'_k / v'_k) (I_k / I)]}{\mathbb{E}[\mathbf{1}_{\{I > a\}} \sum_k (U'_k / v'_k) (I_k / I)]}, \end{aligned}$$

and f_I is the probability density function of I . Therefore, just as in the single-period model, the marginal cost decomposes into one part relating to consumer period recoveries (I) and another part relating to period capital costs (II). However, the latter part now consists of three components. The third and last cost term is familiar from §2 and arises as a result of the impact of expanding coverage for consumer i on all consumers. The first term originates from the multiperiodicity or, more precisely, from the effect of the expansion of consumer i 's coverage on the likelihood of default. In particular, we can interpret it as a risk penalty reflecting the loss of future profit flows for the shareholders should the firm become insolvent, with the marginal default probability being given by the probability density of the aggregate loss at the current capitalization $f_I(a)$. Finally, the second term comes from the regulatory capital constraint, which requires the company to hold capital given by the externally supplied risk measure s .

Despite these differences, the interpretation of the latter part (II) in its entirety remains the same as in §2. It again adds up to the total (period) capital cost:

$$\begin{aligned} \sum_i q_i \left(\tilde{\theta}_i [V f_I(a)] + \frac{\partial s}{\partial q_i} \left[\mathbb{P}(I > a) + \tau - \sum_k \frac{\partial v_k / \partial a}{v'_k} - V f_I(a) \right] \right. \\ \left. + \tilde{\phi}_i a \left[\sum_k \frac{\partial v_k / \partial a}{v'_k} \right] \right) = [\mathbb{P}(I > a) a + \tau a] \end{aligned} \quad (10)$$

and thus again can be interpreted as an allocation of capital according to marginal costs. In fact, the adding up property holds for each of the individual allocation terms:

$$\sum_i q_i \tilde{\theta}_i = \sum_i q_i \frac{\partial s}{\partial q_i} = \sum_i q_i \tilde{\phi}_i a = a,$$

whereas the three weights $[V f_I(a)]$, $[\sum_k ((\partial v_k / \partial a) / v'_k)]$, and $[\mathbb{P}(I > a) + \tau - \sum_k ((\partial v_k / \partial a) / v'_k) - V f_I(a)]$ then add up to the cost fraction $[\mathbb{P}(I > a) + \tau]$.

These weights, in turn, reflect an allocation of the total cost of capital to the three sources of risk penalties, with allocation following the benefit that a marginal unit of capital brings to each of the components. To elaborate, $\sum_k ((\partial v_k / \partial a) / v'_k)$ reflects the marginal benefits enjoyed by counterparties, $V f_I(a)$ is simply the marginal increase in the expected continuation value of the firm, and the remainder $(\mathbb{P}(I > a) + \tau - \sum_k ((\partial v_k / \partial a) / v'_k) - V f_I(a))$ reflects the extent to which the firm was pushed to capitalize beyond the point it would have chosen in the absence of regulation.

The weights can vary substantially according to economic circumstances. For instance, in a situation with close to perfect competition, the firm value V approaches 0 and thus so does the weight associated with the allocation $\tilde{\theta}_i$. Hence, in this case counterparties and regulators will be the primary drivers for capital costs. By contrast, in the situation where consumers are

fully insured by a guaranty fund scheme and where the regulatory constraint is nonbinding, the weights associated with the allocations $\tilde{\phi}_i$ and $\partial s/\partial q_i$ will be 0, and the level of firm capital, as well as its allocation, is solely determined by $\tilde{\theta}_i$ associated with the firm's value as a going concern (again we refer to Appendix B.3 for details).

Allocating capital based on the external risk measure s will only be optimal in specific circumstances. Indeed, it is interesting to note that, even if the regulatory constraint binds, it may not necessarily be an important determinant of capital allocation. If consumer and shareholder considerations in a private (unregulated) market yield a similar capitalization as imposed by regulation,

$$\mathbb{P}(I > a) + \tau \approx \sum_k \frac{\partial v_k / \partial a}{v'_k} + Vf_I(a),$$

then internal counterparty and continuation value concerns will dominate, as the weight on the regulatory allocation will be very small.

As also detailed in Appendix B.3, the three allocation components are supported by three different risk measures—in the sense that the correct marginal economic allocations (10) can be derived as the weighted average of their gradients. Obviously, counterparty-based allocation is again associated with the new risk measure $\tilde{\rho}$ introduced in §2, whereas allocation according to the regulatory constraint directly emanates from the external risk measure s . The allocation piece associated with the firm's value as a going concern, $\tilde{\theta}_i$, on the other hand, can be derived as the gradient of VaR when fixing the confidence level at the firm's default probability—so that VaR arises endogenously in our setting (see Garman 1997 for gradient allocations based on VaR). The intuition is that under our model specification, the value of the company V is constant so that the owners solely care about the probability of (non)default, but unlike the counterparties that collect recoveries in default states, they are not concerned with the size of the shortfall. This leads to VaR, which is exactly the measure that has “its focus on the probability of a loss regardless of the magnitude” (Basak and Shapiro 2001, p. 373). In this setting, the correct marginal cost allocation is delivered by a weighted average of three risk measures, at least two of which (VaR and $\tilde{\rho}$) generally do not adhere to the axioms of coherence or convexity.

5. Liability vs. Asset Risk: Adaptation to Retail Banking

To this point, the analysis has focused exclusively on evaluating liability risks. We now shift our attention to asset risks by adapting the model to retail banking. As

in §2 for an insurance company, we consider a very simple one-period setup.

Specifically, we assume that a bank with deposits d invests in N risky projects with random returns S_i at amounts q_i . Funds beyond the deposits carry a cost exceeding the expected project returns, so that raising capital and investing by itself is not opportune for the bank. Hence, the bank's function is to intermediate access to return opportunities for the single depositor. It pays an interest rate r in nondefault states, whereas the depositor seizes all the assets in states of default. Formally, we study the following optimization problem:

$$\max_{\{q_i\}} \mathbb{E} \left[\sum_{i=1}^N q_i S_i \right] - \mathbb{E} \left[\min \left\{ d(1+r), \sum_{i=1}^N q_i S_i \right\} \right] - \tau \left(\sum_{i=1}^N q_i - d \right)$$

subject to the participation of the depositor:

$$\mathbb{E} \left[U \left(w - d + \min \left\{ d(1+r), \sum_{i=1}^N q_i S_i \right\} \right) \right] \geq \gamma.$$

In view of allocating capital, the key difference is that in this case, risky return opportunities emanate from investment choices rather than exposures to counterparties. The depositor prefers a high level of capital and also investments in safe projects so that bankruptcy is a low probability event and the portfolio value is high even in default states. The bank prefers a low capital level invested in projects with a high return in nondefault states. The first-order condition for investment i thus balances the returns enjoyed by the bank in solvent states (the left-hand side) against the cost of capital, net of the benefits the investment confers on the depositor in states of default (the right-hand side):

$$\mathbb{E}[\mathbf{1}_{\{S \geq d(1+r)\}} S_i] = \tau - \mu \mathbb{E}[\mathbf{1}_{\{S < d(1+r)\}} U'(S_i)],$$

where $S = \sum_i q_i S_i$ and μ is the Lagrange multiplier associated with the constraint. Assuming for expositional purposes that one of the investment opportunities generates a risk-free return r_f , we can write for the return on project i to the bank in excess of a risk-free investment, which, because of the possibility of default, differs from the unconditional expected return:

$$\mathbb{E}[\mathbf{1}_{\{S \geq d(1+r)\}} [S_i - r_f]] = \mu \mathbb{E}[\mathbf{1}_{\{S < d(1+r)\}} U'(r_f - S_i)].$$

Paralleling the usual asset pricing logic, projects with high expected excess returns in solvent states will thus have significant negative excess returns in costly default states, whereas projects with low excess returns in solvent states will generally impose less costs on the depositor as they enjoy better returns in default states.

This condition can also be reinterpreted as an allocation:

$$\mathbb{E}[\mathbf{1}_{\{S \geq d(1+r)\}}[S_i - r_f]] = \bar{\phi}_i \mathbb{E}[\mathbf{1}_{\{S \geq d(1+r)\}}[\bar{S} - r_f] \sum_i q_i], \quad (11)$$

where $\bar{S} = \sum_i q_i S_i / \sum_i q_i$ and

$$\bar{\phi}_i = \frac{\mathbb{E}[\mathbf{1}_{\{S < d(1+r)\}} U'[r_f - S_i]]}{\mathbb{E}[\mathbf{1}_{\{S < d(1+r)\}} U'[\sum_i q_i r_f - S]]}.$$

We again have $\sum_i q_i \bar{\phi}_i = 1$, i.e., the allocation adds up, and again, claimant valuations of recoveries in default states are the primary drivers of the allocation. Projects with poor returns in default states (i.e., where $r_f - S_i$ is large) evidently receive higher allocations because they are perceived as being more risky. What is being allocated in this case is the total cost of diverting safe investments into the risky portfolio (which yields subpar returns in states of default). This happens to equate with the expected excess return of the total portfolio since the marginal cost of risk taking is being equated with its marginal benefit,

$$\begin{aligned} & \mathbb{E}[\mathbf{1}_{\{S \geq d(1+r)\}}[\bar{S} - r_f] \sum_i q_i] \\ &= (\mu \mathbb{E}[\mathbf{1}_{\{S < d(1+r)\}} U' r_f] - \mu \mathbb{E}[\mathbf{1}_{\{S < d(1+r)\}} U' \bar{S}]) \sum_i q_i. \end{aligned}$$

However, the allocations in this model $\bar{\phi}_i$ now have a different form than the allocations $\tilde{\phi}_i$ in the insurance setting of §2. In particular, we do not have the fractional form, which in §2 arises as a result of the dependence of recoveries on incurred losses *relative* to the total loss. By contrast, here total recoveries are simply the portfolio value. This difference leads to a different risk measure. Specifically, we can represent

$$\bar{\phi}_i = \frac{(\partial/\partial q_i) \mathbb{E}^{\bar{\mathbb{P}}}[\sum_i q_i r_f - S]}{\mathbb{E}^{\bar{\mathbb{P}}}[\sum_i q_i r_f - S]} = \frac{(\partial/\partial q_i) \bar{\rho}(S - \sum_i q_i r_f)}{\bar{\rho}(S - \sum_i q_i r_f)}, \quad (12)$$

where $\bar{\rho}(X) = \mathbb{E}^{\bar{\mathbb{P}}}[-X]$, and the probability $\bar{\mathbb{P}}$ is given by

$$\frac{\partial \bar{\mathbb{P}}}{\partial \mathbb{P}} = \frac{\mathbf{1}_{\{S < d(1+r)\}} U'}{\mathbb{E}[\mathbf{1}_{\{S < d(1+r)\}} U']}.$$

Hence, the correct risk measure here is a weighted version of the expected portfolio value,¹⁶ where the weights originate from marginal utilities in default

states, since company default states is where the depositor faces risk. Since the weights are positive only in default states, the risk measure supporting the correct economic allocations therefore can be interpreted as a spectral version of ES (Acerbi 2002), which is obtained by starting with ES (with a threshold based on the default point) and weighting the outcomes in excess of the threshold according to depositor utilities.

This risk measure can then be used for return on risk-adjusted capital calculations popular in the banking literature (Tasche 2004, Stoughton and Zechner 2007). Using Equations (11) and (12), we can write

$$\frac{\mathbb{E}[\mathbf{1}_{\{S \geq d(1+r)\}}[S_i - r_f]]}{(\partial/\partial q_i) \bar{\rho}(S - \sum_i q_i r_f)} = \frac{\mathbb{E}[\mathbf{1}_{\{S \geq d(1+r)\}}[\bar{S} - r_f] \sum_i q_i]}{\bar{\rho}(S - \sum_i q_i r_f)}.$$

Thus, the return for each investment opportunity needs to be measured relative to its risk contribution. For instance, a high return investment for the bank with a high risk allocation $(\partial/\partial q_i) \bar{\rho}(S - \sum_i q_i r_f)$ may be less opportune than a low return investment with a relatively small contribution. Weighing return against risk will result in an optimal portfolio.

The fundamental difference in the source of the risk, viewed from the perspective of the creditor, accounts for the change in the basic form of the measure. In this simple banking model, the risks are associated with investment returns. Bad return realizations in default states translate dollar for dollar into less recoveries for the depositor, so the starting point for risk penalties under ES—a simple probability weighting of losses in default states—is an appropriate basis for capital allocation. The situation is different when the risks in question are not affecting the total recoverable amount but rather the total amount of claims on the institution—as for the case of an insurance company, or if the bank had swap or other exposures that generate liabilities. When liability risks are involved, ES allocations are no longer consistent with the underlying economics.

6. Conclusion

The gradient allocation principle prescribes an allocation of capital to risks in the company's portfolio according to their marginal costs as defined via a *given* risk measure. However, risk measure selection is a thorny issue that can be resolved only through careful consideration of institutional context—a problem implicitly recognized in the early literature on capital allocation (see, e.g., Myers and Read 2001, Tasche 2004, both of whom ultimately fall back on regulation as motivating the choice of risk measure) but since overlooked in favor of a focus on mathematical properties of allocation rules and risk measures.

Instead of starting with a risk measure, this paper starts with primitive assumptions and calculates the

¹⁶ It is important to note that unlike the insurance setting, here higher risk is associated with a lower portfolio return in default states, so that the risk measure is equipped with a negative sign—as is also conventional for ES when focusing on asset risk. This dissonance between insurance and financial formulations of risk measures is well known and particularly results in differences in the definitions of common risk measure axioms.

marginal cost of risk from the perspective of a profit-maximizing firm with risk-averse counterparties in an incomplete market setting with frictional capital costs. In this setting, and in various extensions, we then derive the risk measures whose gradients are consistent with marginal cost from the perspective of the firm.

The results of this reconciliation show that, for purposes of pricing and performance measurement, the appropriate risk measure depends crucially on economic context. The economic drivers of risk penalties flow from the costs of default imposed on counterparties, regulators, and shareholders. These influences can be incorporated into risk measures, but the process yields measures with properties viewed as unappealing by many. In particular, we identify several circumstances where the appropriate risk measure is neither coherent nor convex.

We compare the allocations obtained in a single-period setting without regulation to those obtained from the gradients of VaR and ES—the coherent risk measure currently favored by many academics and regulators. We show that ES may underweight or overweight severe states of default, depending on the nature of customer risk aversion. This raises the interesting possibility that a transition away from a system of regulation relying on risk measure-based solvency assessment to one relying on market (counterparty) discipline will not necessarily mitigate the oft-lamented failure of financial institutions to penalize “tail” risk.

Numerous extensions are possible. Our setting is but one possible specification of the profit maximization problem. Others could include informational frictions with regard to consumer endowments and preferences, explicit consideration of managerial incentives in a decentralized organization of the multiline business, a more detailed modeling of the firm’s capital raising and capital structure decision, or any number of other complications. All of these may, of course, lead to different risk measures, underscoring the point that risk measures chosen for their technical properties generally fail to yield correct pricing and efficient allocation of capital.

One can also contemplate changes of perspective: the calculations in this paper are done from the perspective of a profit-maximizing firm, but one could also consider the calculus of a regulator or social planner. In specific cases, the calculus will be similar. For example, a regulator without responsibility for unpaid losses (i.e., if no guaranty fund or deposit insurance scheme exists) but in a context where counterparties are uninformed may view risk similarly to a (competitive) profit-maximizing firm. However, a regulator responsible for unpaid losses would presumably have to consider their value in selecting a risk measure as well as other issues—such as bankruptcy costs not internalized by private

firms and the production cost associated with deposit insurance—that would determine the optimal level of capitalization for financial institutions as well as the social cost of risk.¹⁷

In general, the particular specifications of the economic environment will lead to different risk measures. What is appropriate for one type of institution and set of circumstances is not likely to suit another. Going forward, the challenge for companies and regulators will be to choose risk measures based on consistency with their own economic objectives and constraints, as opposed to consistency with an arbitrary set of axioms.

Acknowledgments

The authors thank Jerome Detemple (the editor), an anonymous associate editor, two anonymous referees, Enrico Biffis, Keith Crocker, Richard Derrig, Anastasia Kartasheva, John Major, Stephen Mildenhall, Giacomo Nocera, Richard Phillips, Markus Stricker, and Ajay Subramanian; and seminar participants at the 2012 Allied Social Science Association Meetings, the 2011 National Bureau of Economic Research Insurance Project Workshop, the 2012 Casualty Actuarial Society Spring Meeting, the International Conference on Mathematical Finance and Economics 2011, the Actuarial Research Conference 2012, the American Risk and Insurance Association 2012 Meeting, ETH Zurich, Georgia State University, Goethe University Frankfurt, Imperial College London, the University of Cincinnati, the University of Copenhagen, the University of Iowa, the University of South Carolina, Ludwig Maximilian University, National Chengchi University, Ulm University, and Pennsylvania State University for helpful comments. Ongoing research support from the Casualty Actuarial Society and travel support from the Center for the Economic Analysis of Risk are gratefully acknowledged.

Appendix A. Technical Appendices

A.1. Technical Appendix to §2: Details on Remark 1

Modification of the Setting. Consider the same setup as in the main text, but following the transportation economics literature on congestion pricing (Keeler and Small 1977), we additionally assume that the consumer ignores her own contribution to the overall risk in the company’s portfolio when making a decision about her coverage level. This assumption could be motivated by behavioral arguments alluding to consumer myopia, e.g., that a policyholder makes a coverage decision by consulting company ratings but not considering the impact of her coverage on the ratings. However, the primary motivation here is of technical nature—namely, that this assumption delivers aggregate marginal costs equating with total company cost in (7). And, in analogy to the transportation literature, this assumption appears acceptable when the consumer is small in relation to the pool (see also Zanjani 2010).

Mathematically, the assumption that the consumer ignores her influence on company risk means that she ignores the

¹⁷ One such exercise is performed by Acharya et al. (2010), who study a specific environment that leads to a new risk measure, dubbed *systemic expected shortfall*.

influence of her (marginal) coverage decision on recoveries, since company risk only materializes in default states. To formally introduce this into the problem, we thus modify the original utility function to

$$\tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) = \mathbb{E}[U_i(w_i - p_i^*(q_i) - L_i + \tilde{R}_i)],$$

where

$$\begin{aligned} \tilde{R}_i &= \tilde{R}_i(q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) \\ &= \min \left\{ L_i(L_i, q_i), \frac{\tilde{a}}{\sum_{j=1}^N L_j(L_j, \tilde{q}_j)} L_i(L_i, q_i) \right\}, \end{aligned}$$

and $p_i^*(\cdot)$ is a premium function as before. The idea here is to fix recovery rates by fixing the quantities \tilde{a} and $\{\tilde{q}_i\}$, leaving the consumer with the free choice of q_i —but with the caveat that this choice does not influence recovery rates.¹⁸

The firm's objective function remains the same, except that the pricing function now accounts for the skew in the consumer's perception in that her decision does not impact recovery rates. In particular, the company solves

$$\max_{\tilde{a}, \{p_i^*(\cdot)\}, \{\tilde{q}_i\}} \left\{ \sum p_i^*(\tilde{q}_i) - \sum e_i - \tau \tilde{a} \right\}$$

subject to

$$v_i(\tilde{a}, w_i - p_i^*(\tilde{q}_i), \tilde{q}_1, \dots, \tilde{q}_N) \geq \gamma_i,$$

and, in addition, the new constraint:

$$\tilde{q}_i \in \arg \max_{q_i} \tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N), \quad \forall i. \quad (\text{A1})$$

Equation (A1) is an incentive compatibility constraint requiring the choice of coverage level to be consistent with the consumer optimizing, given her perception of own utility (which ignores her own impact on recovery rates) and the selected pricing function.

It is evident that the firm's profits under this maximization can be no better than those achieved under the original program (2) and (3), since we have simply added another constraint, and choosing the premium schedule at different points than \tilde{q}_i is immaterial to the company's profits. It is therefore clear that, given optimal choices \tilde{a} , $\{\tilde{q}_i\}$, and $\{\tilde{p}_i\}$ to the original program, the firm will maximize profits under the new setup if it can choose those same asset and coverage levels and find a pricing function $p_i^*(\cdot)$ that both satisfies $p_i^*(\tilde{q}_i) = \tilde{p}_i$ and induces consumers to choose the original solution:

$$\tilde{q}_i \in \arg \max_{q_i} \tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) \quad \forall i.$$

The following lemma shows that such a function exists.

¹⁸ Alternatively, we could also specify $\tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) = \tilde{v}_i(w_i - p_i^*(q_i), q_i; (q_i/\tilde{q}_i)(\tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N))$, where the consumer is cognizant of her own coverage but "scales" the company according to her own coverage level; this would not change the ensuing result. Again, the important point is that the consumer expects the same recovery per dollar of coverage in default states independent of her choice of q_i .

LEMMA 1. Suppose \tilde{a} , $\{\tilde{q}_i\}$, and $\{\tilde{p}_i\}$ are the optimal choices maximizing (2) subject to (3). Then, for each i , there exists a smooth, monotonically increasing function $p_i^*(\cdot)$ satisfying

1. $p_i^*(\tilde{q}_i) = \tilde{p}_i$ and
2. $\tilde{q}_i \in \arg \max_{q_i} \tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N)$.

PROOF. Start by noting that it is evident that the constraint (3) binds. Note further that the function of x

$$g(x) = \tilde{v}_i(w_i - x, 0; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N)$$

is monotonically decreasing and, hence, invertible, so that we may uniquely define

$$p_i^*(0) = g^{-1}(\gamma_i), \quad (\text{A2})$$

which obviously satisfies

$$\tilde{v}_i(w_i - p_i^*(0), 0; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) = \gamma_i.$$

Furthermore, let $p_i^*(\cdot)$ be a solution to the initial value problem¹⁹

$$\begin{aligned} \frac{\partial p_i^*(x)}{\partial x} &= \frac{(\partial/\partial q_i) \tilde{v}_i(w_i - p_i^*(x), x; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N)}{(\partial/\partial w) \tilde{v}_i(w_i - p_i^*(x), x; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N)}, \\ p_i^*(0) &= g^{-1}(\gamma_i), \quad (\text{A3}) \end{aligned}$$

on the compact choice set for q_i . Because of Peano's theorem, we are guaranteed existence of such a function and that it is smooth. Moreover, since $\partial \tilde{v}_i/\partial w, \partial \tilde{v}_i/\partial q_i > 0$, we know that the function is monotonically increasing.

Moving on, by construction we know that

$$\begin{aligned} &\tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) \\ &= \gamma_i + \int_0^{q_i} \left[\frac{\partial}{\partial q_i} \tilde{v}_i(w_i - p_i^*(x), x; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) \right. \\ &\quad \left. - \frac{\partial}{\partial w} \tilde{v}_i(w_i - p_i^*(x), x; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) \cdot \frac{\partial p_i^*(x)}{\partial x} \right] dx \\ &= \gamma_i + 0, \quad q_i > 0. \quad (\text{A4}) \end{aligned}$$

In particular,

$$\tilde{v}_i(w_i - p_i^*(\tilde{q}_i), \tilde{q}_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) = \gamma_i,$$

which, since it is evident that the constraint (3) binds in the original optimization, can be true if and only if

$$p_i^*(\tilde{q}_i) = \tilde{p}_i,$$

proving the first part of the lemma. Moreover, (A4) directly implies that

$$\tilde{q}_i \in \arg \max_{q_i} \tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N),$$

proving the second part. \square

¹⁹ Here, $\partial \tilde{v}_i/\partial w$ and $\partial \tilde{v}_i/\partial q_i$ denote the derivatives with respect to the first and second arguments of \tilde{v}_i , respectively.

Now at the optimum, we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial q_i} \tilde{v}_i(w_i - p_i^*(q_i), q_i; \tilde{a}, \tilde{q}_1, \dots, \tilde{q}_N) \Big|_{\tilde{q}_i=q_i} \\ \Leftrightarrow 0 &= \mathbb{E} \left[\frac{\partial \tilde{R}_i}{\partial q_i} U'_i \right] - \frac{\partial p_i^*}{\partial q_i} \mathbb{E}[U'_i] \\ \Leftrightarrow 0 &= \underbrace{\mathbb{E} \left[\frac{\partial \tilde{R}_i}{\partial q_i} U'_i \right] - \mathbb{E} \left[\mathbf{1}_{\{I>a\}} U'_i \frac{\partial I_i / \partial q_i}{I} \frac{a I_i}{I} \right]}_{=\partial v_i / \partial q_i} \\ &\quad + \mathbb{E} \left[\mathbf{1}_{\{I>a\}} U'_i \frac{\partial I_i / \partial q_i}{I} \frac{a I_i}{I} \right] - \frac{\partial p_i^*}{\partial q_i} v'_i. \end{aligned}$$

The term in the outer brackets represents how the consumer perceives the marginal benefit of additional coverage, which, because of the additional assumption, differs from the true impact in (4) by $\mathbb{E}[\mathbf{1}_{\{I>a\}} U'_i ((\partial I_i / \partial q_i) / I) (a I_i / I)]$. Working with this modified marginal cost condition in (5) and (6), the second term in (6) vanishes, and we obtain (7); i.e., aggregate marginal costs fully recover all of the firm's costs.

Limiting Behavior. To analyze the limiting behavior as the portfolio size goes to infinity, consider the aggregated marginal costs relative to company assets:

$$\begin{aligned} \frac{\sum_{i=1}^N q_i (\partial p_i^* / \partial q_i)}{a} &= \frac{\mathbb{E}[\mathbf{1}_{\{I \leq a\}} I]}{a} + [\tau + \mathbb{P}(I > a)] \\ &\quad - \mathbb{E} \left[\mathbf{1}_{\{I>a\}} \sum_{i=1}^N \frac{U'_i}{v'_i} \frac{I_i}{I} \right]. \end{aligned}$$

Thus, the difference between the aggregated marginal costs and the company's total cost in (7) is the last term. To analyze the limit as N goes to infinity, we represent it as

$$\mathbb{E} \left[\mathbf{1}_{\{I>a\}} \sum_{i=1}^N \frac{U'_i}{v'_i} \frac{I_i}{I} \right] = \frac{1}{N} \mathbb{E} \left[\mathbf{1}_{\{I>a\}} \frac{(1/N) \sum_{i=1}^N (U'_i / v'_i) I_i^2}{((1/N) \sum_{i=1}^N I_i)^2} \right]. \quad (\text{A5})$$

Now in the case of iid losses with finite means, it can be readily seen that the denominator converges almost surely to the square of the expected indemnity payment by the law of large numbers. Similarly, in the (limiting) case of risk neutrality such that $U'_i / v'_i = 1$ and if the second moment exists, the numerator will converge to the expected value of the squared indemnity, and the expectation on the right-hand side of (A5) will thus be finite as $N \rightarrow \infty$. Thus, the $1/N$ in front of the expected value will yield this term to vanish as the portfolio size is extended, and we obtain the adding up result in (7).

The result will also hold in far more general circumstances, in view of preferences, loss distributions, and their dependence structure. In various settings, the Birkhoff–Khinchin ergodicity theorem can be applied to verify that (A5) goes to zero. However, since the formulation in the main text is general, the limit has to be verified on a case-by-case basis. Situations where the zero limit result does not hold essentially only arise when the (preference-adjusted) loss for one risk dominates the portfolio outcome.

A.2. Technical Appendix to §3.1: Homogeneous Exponential Losses

As indicated in §3.1, we consider N identical consumers facing iid, exponentially distributed losses $L_i \sim \text{Exp}(\nu)$, $1 \leq i \leq N$. As also detailed there, we assume that consumers have CARA utilities and that γ is given by the autarky level; i.e.,

$$\gamma = e^{-\alpha w} \frac{\nu}{\nu - \alpha}.$$

LEMMA 2. The firm's optimization problem (2) and (3) takes the form

$$\begin{cases} \max_{a, q, p} \left\{ Np - Nq \left[\frac{1}{\nu} \Gamma_{N-1, \nu} \left(\frac{a}{q} \right) - \frac{\nu^{N-1}}{(N-1)!} e^{-\nu(a/q)} \left(\frac{a}{q} \right)^{N-1} \right. \right. \\ \quad \left. \left. \cdot \left(\frac{1}{\nu} + \frac{1}{N} \frac{a}{q} \right) \right] - a \bar{\Gamma}_{N, \nu} \left(\frac{a}{q} \right) - \tau a \right\} \\ \text{subject to} \\ \gamma \leq e^{-\alpha(w-p)} \left\{ \frac{\nu}{\nu - (1-q)\alpha} \left[\Gamma_{N-1, \nu} \left(\frac{a}{q} \right) - \frac{e^{-(a/q)(\nu - (1-q)\alpha)} \nu^{N-1}}{((1-q)\alpha)^{N-1}} \right. \right. \\ \quad \cdot \Gamma_{N-1, (1-q)\alpha} \left(\frac{a}{q} \right) \right] - \sum_{k=0}^{\infty} \left(\frac{\alpha}{\nu} \right)^k \frac{(N-1)!}{(N-1+k)!} e^{-\nu a} \\ \quad \left. \cdot \sum_{j=0}^{N-1} \frac{(a\nu)^j}{j!} \frac{(N-j+k-1)!}{(N-j-1)!} \bar{\Gamma}_{N-j+k, \nu} \left(a \left(\frac{1-q}{q} \right) \right) \right\}, \end{cases} \quad (\text{A6})$$

where $\bar{\Gamma}_{m, b}(x) = 1 - \Gamma_{m, b}(x)$ and $\Gamma_{m, b}(\cdot)$ denotes the cumulative distribution function of the Gamma distribution with parameters m and b .

PROOF. For consumer N , $L_N \sim \text{Exp}(\nu)$, and the loss incurred by the other consumers is $L_{-N} = \sum_{i=1}^{N-1} L_i \sim \text{Gamma}(N-1, \nu)$. Then,

$$e = e_N = \underbrace{\mathbb{E}[q L_N \mathbf{1}_{\{q(L_{-N} + L_N) < a\}}]}_{\text{part i}} + a \underbrace{\mathbb{E} \left[\frac{q L_N}{q(L_{-N} + L_N)} \mathbf{1}_{\{q(L_{-N} + L_N) \geq a\}} \right]}_{\text{part ii}}.$$

For part ii, note that $L_N / (L_{-N} + L_N)$ is Beta(1, $N-1$) distributed independent of $L_{-N} + L_N \sim \text{Gamma}(N, \nu)$. Hence, part ii can be written as

$$a \mathbb{P} \left(L_{-N} + L_N \geq \frac{a}{q} \right) \mathbb{E} \left[\frac{L_N}{L_{-N} + L_N} \right] = a \bar{\Gamma}_{N, \nu} \left(\frac{a}{q} \right) N^{-1}.$$

For part i, we have

$$\begin{aligned} &q \mathbb{E}[L_N \mathbf{1}_{\{q(L_{-N} + L_N) < a\}}] \\ &= q \int_0^\infty \int_0^\infty \mathbf{1}_{\{i+l < a/q\}} l \nu e^{-\nu l} \frac{\nu^{N-1}}{(N-2)!} i^{N-2} e^{-\nu i} dl di \\ &= q \frac{\nu^N}{(N-2)!} \int_0^{a/q} \int_0^{a/q-i} l e^{-\nu l} dl i^{N-2} e^{-\nu i} di \\ &= q \frac{\nu^N}{(N-2)!} \int_0^{a/q} \left[\frac{1}{\nu^2} - \frac{1}{\nu} \left(\frac{a}{q} + \frac{1}{\nu} \right) e^{-\nu a/q} e^{\nu i} + i \frac{1}{\nu} e^{-\nu a/q} e^{\nu i} \right] \\ &\quad \cdot i^{N-2} e^{-\nu i} di \\ &= q \frac{\nu^{N-2}}{(N-2)!} \int_0^{a/q} i^{N-2} e^{-\nu i} di - q \frac{\nu^{N-1}}{(N-2)!} e^{-\nu a/q} \\ &\quad \cdot \left[\left(\frac{a}{q} + \frac{1}{\nu} \right) \int_0^{a/q} i^{N-2} di - \int_0^{a/q} i^{N-1} di \right] \\ &= \frac{q}{\nu} \Gamma_{N-1, \nu}(a/q) - q \frac{\nu^{N-1}}{(N-1)!} e^{-\nu a/q} \left(\frac{a}{q} \right)^{N-1} \left[\frac{1}{\nu} + \frac{1}{N} \frac{a}{q} \right]. \end{aligned}$$

Therefore, since all consumers are identical, the objective function (2) takes the form displayed in (A6).

For condition (3), on the other hand, we have

$$\begin{aligned} V &= V_N = \mathbb{E}[U(w - p - L_N + R_N)] \\ &= \underbrace{\mathbb{E}[U(w - p - (1 - q)L_N) \mathbf{1}_{\{q(L_{-N} + L_N) < a\}}]}_{\text{part i}} \\ &\quad + \underbrace{\mathbb{E}\left[U\left(w - p - L_N + a \frac{L_N}{L_{-N} + L_N}\right) \mathbf{1}_{\{q(L_{-N} + L_N) \geq a\}}\right]}_{\text{part ii}}. \end{aligned}$$

For part i, we obtain

$$\begin{aligned} &\mathbb{E}[U(w - p - (1 - q)L_N) \mathbf{1}_{\{q(L_{-N} + L_N) < a\}}] \\ &= - \int_0^\infty \int_0^\infty \mathbf{1}_{\{i+l < a/q\}} e^{-\alpha(w-p-(1-q)l)} \nu e^{-\nu l} \\ &\quad \cdot \frac{\nu^{N-1}}{(N-2)!} i^{N-2} e^{-\nu i} dl di \\ &= -e^{-\alpha(w-p)} \int_0^{a/q} \frac{1}{\nu - \alpha(1-q)} [1 - e^{-a/q(\nu-(1-q)\alpha)-i(1-q)\nu+iv}] \\ &\quad \cdot \frac{\nu^{N-1}}{(N-2)!} i^{N-2} e^{-\nu i} di \\ &= e^{-\alpha(w-p)} \left[\frac{1}{\nu - \alpha(1-q)} \Gamma_{N-1, \nu}(a/q) \right. \\ &\quad \left. - \frac{e^{-(a/q)(\nu-(1-q)\alpha)} \nu^{N-1}}{((1-q)\alpha)^{N-1}} \Gamma_{N-1, (1-q)\alpha}\left(\frac{a}{q}\right) \right]. \end{aligned}$$

For part ii, note that

$$\begin{aligned} &\mathbb{E}\left[U\left(w - p - ((L_{-N} + L_N) - a) \frac{L_N}{L_{-N} + L_N}\right) \mathbf{1}_{\{q(L_{-N} + L_N) \geq a\}}\right] \\ &= \int_0^1 \int_0^\infty e^{-\alpha(w-p-(l-a)y)} \mathbf{1}_{\{l \geq a/q\}} \frac{\nu^N}{(N-1)!} \\ &\quad \cdot l^{N-1} e^{-\nu l} (N-1)(1-y)^{N-2} dl dy \\ &= -e^{-\alpha(w-p)} \int_{a/q}^\infty \frac{\nu^N}{(N-1)!} e^{-\nu l} l^{N-1} \\ &\quad \cdot \underbrace{\int_0^1 e^{(\alpha(l-a))y} (N-1)(1-y)^{N-2} dy}_{\text{mgf}_{\text{Beta}(1, N-1)}(\alpha(l-a))} dl, \end{aligned}$$

where mgf_{Beta} denotes the moment-generating function of the Beta distribution. Hence,

$$\begin{aligned} &\mathbb{E}\left[U\left(w - p - ((L_{-N} + L_N) - a) \frac{L_N}{L_{-N} + L_N}\right) \mathbf{1}_{\{q(L_{-N} + L_N) \geq a\}}\right] \\ &= -e^{-\alpha(w-p)} \int_{a/q}^\infty \frac{\nu^N}{(N-1)!} e^{-\nu l} l^{N-1} \sum_{k=0}^\infty \frac{(\alpha(l-a))^k}{(N-1+k)!} dl \\ &= -e^{-\alpha(w-p)} \sum_{k=0}^\infty \sum_{j=0}^{N-1} \binom{N-1}{j} a^j e^{-\nu a} \frac{\alpha^k}{(N-1+k)!} \\ &\quad \cdot \frac{(N+k-j-1)!}{\nu^{-j+k}} \\ &\quad \cdot \int_{a/q}^\infty \frac{\nu^{N+k-j}}{(N+k-j-1)!} e^{-\nu(l-a)} (l-a)^{N+k-j-1} dl \end{aligned}$$

$$\begin{aligned} &= -e^{-\alpha(w-p)} \sum_{k=0}^\infty \sum_{j=0}^{N-1} \frac{(N-1)!}{(N-1+k)!} \frac{(a\nu)^j}{j!} e^{-\nu a} \left(\frac{\alpha}{\nu}\right)^k \\ &\quad \cdot \frac{(N+k-j-1)!}{(N-1-j)!} \bar{\Gamma}_{N-j+k, \nu}\left(a\left(\frac{1-q}{q}\right)\right). \quad \square \end{aligned}$$

Note that the constraint will bind at the optimum, so we can solve for p :

$$\begin{aligned} p(a, q) &= \frac{1}{\alpha} \left(\log \left\{ \frac{\nu}{\nu - \alpha} \right\} - \log \left\{ \frac{\nu}{\nu - (1-q)\alpha} \right\} \right. \\ &\quad \cdot \left[\frac{e^{-(a/q)(\nu-(1-q)\alpha)} \nu^{N-1}}{((1-q)\alpha)^{N-1}} \Gamma_{N-1, (1-q)\alpha}\left(\frac{a}{q}\right) - \Gamma_{N-1, \nu}\left(\frac{a}{q}\right) \right] \\ &\quad + \sum_{k=0}^\infty \left(\frac{\alpha}{\nu}\right)^k \frac{(N-1)!}{(N-1+k)!} e^{-\nu a} \sum_{j=0}^{N-1} \frac{(a\nu)^j}{j!} \frac{(N-j+k-1)!}{(N-j-1)!} \\ &\quad \cdot \bar{\Gamma}_{N-j+k, \nu}\left(a\left(\frac{1-q}{q}\right)\right) \Bigg). \end{aligned}$$

Hence, by inserting $p(a, q)$ into the objective function, we can solve for the optimum of a simple bivariate optimization problem, which can be solved using numerical methods or by (numerically) solving the first-order conditions.

For the counterparty-based allocations, we get the following, in accordance with the main text:

$$\begin{aligned} \frac{1}{N} &= \frac{\mathbb{E}[\mathbf{1}_{\{qL \geq a\}} \sum_{j=1}^N U'(w - p - L_j + a(L_j/L))(L_j/L)](L_i/L)}{\mathbb{E}[\mathbf{1}_{\{qL \geq a\}} \sum_{j=1}^N U'(w - p - L_j + a(L_j/L))(L_j/L)]} \\ &= \frac{1}{N} \mathbb{E} \left[\underbrace{\left[\frac{\partial \tilde{\mathbb{P}}}{\partial \mathbb{P}} \right] L}_{\tilde{\psi}(L)} \right]. \end{aligned}$$

We obtain the following result for the weighting function $\tilde{\psi}$.

LEMMA 3. We have

$$\tilde{\psi}(L) = \mathbf{1}_{\{qL \geq a\}} \hat{c}_{N, \nu, \alpha, a, q} \sum_{k=0}^\infty \frac{(k+1)(\alpha(L-a))^k}{(N+k)!}.$$

PROOF. Similar to the proof of Lemma 2, for consumer N with $L = \sum_{i=1}^N L_i$,

$$\begin{aligned} &\mathbb{E} \left[\sum_{j=1}^N U' \left(w - p - L_j + a \frac{L_j}{L} \right) \frac{L_j}{L} \frac{L_N}{L} \middle| L \right] \\ &= \sum_{j=1}^{N-1} \underbrace{\mathbb{E} \left[U' \left(w - p - (L-a) \frac{L_j}{L} \right) \frac{L_j}{L} \frac{L_N}{L} \middle| L \right]}_{\text{part i}} \\ &\quad + \underbrace{\mathbb{E} \left[U' \left(w - p - (L-a) \frac{L_N}{L} \right) \left(\frac{L_N}{L} \right)^2 \middle| L \right]}_{\text{part ii}}. \end{aligned}$$

Note that $L_j/L, L_N/L \sim \text{Beta}(1, N-1)$, and for the joint distribution,

$$f_{L_j/L, L_N/L}(x, y) = (1-x-y)^{N-3} (N-2)(N-1) \mathbf{1}_{\{x, y \geq 0, x+y \leq 1\}}, \quad j \neq N.$$

Hence, for part i,

$$\begin{aligned} & \mathbb{E} \left[U' \left(w - p - (L - a) \frac{L_{N-1}}{L} \right) \frac{L_{N-1}}{L} \frac{L_N}{L} \middle| L \right] \\ &= \alpha e^{-\alpha(w-p)} \int_0^1 \int_0^{1-x} e^{\alpha(L-a)x} xy(N-1)(N-2) \\ & \quad \cdot (1-x-y)^{N-3} dy dx \\ &= \alpha e^{-\alpha(w-p)} \int_0^1 e^{\alpha(L-a)x} x \\ & \quad \cdot \underbrace{\int_0^{1-x} y(N-1)(N-2)(1-x-y)^{N-3} dy dx}_{=(1-x)^{N-1}} \\ &= \alpha e^{-\alpha(w-p)} \beta(2, N) \int_0^1 e^{\alpha(L-a)x} \frac{1}{\beta(2, N)} x(1-x)^{N-1} dx \\ & \quad \underbrace{= \text{mgf}_{\text{Beta}(2, N)}(\alpha(L-a))}_{= \text{mgf}_{\text{Beta}(2, N)}(\alpha(L-a))} \\ &= \alpha e^{-\alpha(w-p)} \frac{1}{N(N+1)} (N+1)! \sum_{k=0}^{\infty} \frac{(k+1)(\alpha(L-a))^k}{(N+k+1)!}, \end{aligned}$$

whereas for part ii,

$$\begin{aligned} & \mathbb{E} \left[U' \left(w - p - (L - a) \frac{L_N}{L} \right) \left(\frac{L_N}{L} \right)^2 \middle| L \right] \\ &= \alpha e^{-\alpha(w-p)} \mathbb{E} \left[\exp \left\{ \alpha(L-a) \frac{L_N}{L} \right\} \left(\frac{L_N}{L} \right)^2 \middle| L \right] \\ & \quad \underbrace{= (\partial^2 / \partial t^2) \text{mgf}_{\text{Beta}(1, N-1)}(t) |_{t=\alpha(L-a)}}_{= (\partial^2 / \partial t^2) \text{mgf}_{\text{Beta}(1, N-1)}(t) |_{t=\alpha(L-a)}} \\ &= \alpha e^{-\alpha(w-p)} (N-1)! \sum_{k=0}^{\infty} \frac{(k+1)(k+2)(\alpha(L-a))^k}{(N+k+1)!}, \end{aligned}$$

so that

$$\begin{aligned} & \mathbb{E} \left[\sum_{j=1}^N U' \left(w - p - L_j + a \frac{L_j}{L} \right) \frac{L_j}{L} \frac{L_N}{L} \middle| L \right] \\ &= \alpha e^{-\alpha(w-p)} (N-1)! \\ & \quad \cdot \sum_{k=0}^{\infty} \frac{(N-1)(k+1)(\alpha(L-a))^k + (k+1)(k+2)(\alpha(L-a))^k}{(N+k+1)!} \\ &= \alpha e^{-\alpha(w-p)} (N-1)! \sum_{k=0}^{\infty} \frac{(k+1)(N+1+k)(\alpha(L-a))^k}{(N+k+1)!} \\ &= \alpha e^{-\alpha(w-p)} (N-1)! \sum_{k=0}^{\infty} \frac{(k+1)(\alpha(L-a))^k}{(N+k)!}. \end{aligned}$$

For the denominator,

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}_{\{qL \geq a\}} \sum_{j=1}^N U' \left(w - p - L_j + a \frac{L_j}{L} \right) \frac{L_j}{L} \right] \\ &= \mathbb{E} \left[\mathbf{1}_{\{qL \geq a\}} N \alpha e^{-\alpha(w-p)} (N-1)! \sum_{k=0}^{\infty} \frac{(k+1)(\alpha(L-a))^k}{(N+k)!} \right] \\ &= N! \alpha e^{-\alpha(w-p)} \sum_{k=0}^{\infty} \frac{k+1}{(N+k)!} \alpha^k \\ & \quad \cdot \int_{a/q}^{\infty} \frac{\nu^N}{(N-1)!} (l-a)^k l^{N-1} e^{-\nu l} dl \\ &= N! \alpha e^{-\alpha(w-p)} \sum_{k=0}^{\infty} \left(\frac{\alpha}{\nu} \right)^k \frac{k+1}{(N+k)!} e^{-\nu a} \end{aligned}$$

$$\begin{aligned} & \cdot \sum_{j=0}^{N-1} \binom{N-1}{j} (a\nu)^j \frac{(N+k-j-1)!}{(N-1)!} \\ & \cdot \int_{a/q}^{\infty} \frac{\nu^{N+k-j}}{(N+k-j-1)!} e^{-\nu(l-a)} (l-a)^{N+k-j-1} dl \\ &= N! \alpha e^{-\alpha(w-p)} \sum_{k=0}^{\infty} \left(\frac{\alpha}{\nu} \right)^k \frac{k+1}{(N+k)!} e^{-\nu a} \\ & \cdot \sum_{j=0}^{N-1} \frac{(a\nu)^j}{j!} \frac{(N+k-j-1)!}{(N-j-1)!} \bar{\Gamma}_{N+k-j, \nu} \left(a \left(\frac{1-q}{q} \right) \right). \end{aligned}$$

Hence,

$$\begin{aligned} q\tilde{\varphi}_i &= \frac{1}{N} \mathbb{E} \left[\mathbf{1}_{\{qL \geq a\}} \right. \\ & \cdot \left(\sum_{k=0}^{\infty} \frac{(k+1)(\alpha(L-a))^k}{(N+k)!} \right) \cdot \left(\sum_{k=0}^{\infty} \left(\frac{\alpha}{\nu} \right)^k \frac{k+1}{(N+k)!} e^{-\nu a} \right. \\ & \cdot \underbrace{\sum_{j=0}^{N-1} \frac{(a\nu)^j}{j!} \frac{(N+k-j-1)!}{(N-j-1)!} \bar{\Gamma}_{N+k-j, \nu} \left(a \left(\frac{1-q}{q} \right) \right)}_{=\tilde{\psi}(L)} \left. \right)^{-1} \left. \right]. \quad \square \end{aligned}$$

For implementation purposes, the numerator can be expressed as

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{(k+1)t^k}{(N+k)!} \bigg|_{t=\alpha(L-a)} &= \frac{\partial}{\partial t} \left[\sum_{k=0}^{\infty} \frac{t^{k+1}}{(N+k)!} \right] \bigg|_{t=\alpha(L-a)} \\ &= \frac{\partial}{\partial t} \left[t^{-(N-1)} \sum_{k=N}^{\infty} \frac{t^k}{k!} \right] \bigg|_{t=\alpha(L-a)} \\ &= \frac{\partial}{\partial t} \left[t^{-(N-1)} \left(e^t - \sum_{k=0}^{N-1} \frac{t^k}{k!} \right) \right] \bigg|_{t=\alpha(L-a)} \\ &= -(N-1)t^{-N} \left(e^t - \sum_{k=0}^{N-1} \frac{t^k}{k!} \right) \\ & \quad + t^{-(N-1)} \left(e^t - \sum_{k=0}^{N-1} \frac{kt^{k-1}}{k!} \right) \bigg|_{t=\alpha(L-a)} \\ &= \left(e^t - \sum_{k=0}^{N-2} \frac{t^k}{k!} \right) \times (t^{-N+1} - t^{-N}(N-1)) \\ & \quad + \frac{(N-1)}{(N-1)!} t^{-1} \bigg|_{t=\alpha(L-a)}. \end{aligned}$$

We obtain for the risk measure supporting the counterparty-based allocations when evaluated at the aggregate loss

$$\tilde{\rho}(I) = \tilde{\rho}(qL) = \exp\{\mathbb{E}[\tilde{\psi}(L) \log(qL)]\},$$

and

LEMMA 4.

$$\begin{aligned} \tilde{\rho}(I) &= q \exp \left\{ \frac{\hat{c}_{N, \nu, \alpha, a, q}}{(N-1)!} \sum_{k=0}^{\infty} \frac{\alpha^k}{(N+k)!} \right. \\ & \cdot \sum_{j=0}^k \frac{(k+1)!}{j!(k-j)!} (-a)^{k-j} \left(\frac{1}{\nu} \right)^j \left(\log \left[\frac{a}{q} \right] \Gamma \left(N+j, \frac{a\nu}{q} \right) \right. \\ & \quad \left. \left. + \frac{a\nu}{q} G_{2,3}^{3,0} \left[\begin{matrix} 0, 0, \dots, 0 \\ N+j-1, -1, \dots, -1 \end{matrix} \middle| \frac{a\nu}{q} \right] \right) \right\}, \end{aligned}$$

where G denotes the Meijer G -function.

PROOF. Inserting $\tilde{\psi}$ gives

$$\begin{aligned}\tilde{\rho}(I) &= q \exp \left\{ \mathbb{E} \left[\tilde{\psi}(L) \log\{L\} \right] \right\} \\ &= q \exp \left\{ \mathbb{E} \left[\mathbf{1}_{\{qL \geq a\}} \hat{c}_{N, \nu, \alpha, a, q} \sum_{k=0}^{\infty} \frac{(k+1)(\alpha(L-a))^k}{(N+k)!} \log\{L\} \right] \right\} \\ &= q \exp \left\{ \hat{c} \int_{a/q}^{\infty} \sum_{k=0}^{\infty} \frac{\alpha^k}{(N+k)!} (k+1) \sum_{j=0}^k \binom{k}{j} y^j (-a)^{k-j} \right. \\ &\quad \cdot \log\{y\} \frac{\nu^N}{(N-1)!} y^{N-1} e^{-\nu y} dy \left. \right\} \\ &= q \exp \left\{ \frac{\hat{c} \nu^N}{(N-1)!} \sum_{k=0}^{\infty} \frac{\alpha^k}{(N+k)!} (k+1) \sum_{j=0}^k \binom{k}{j} (-a)^{k-j} \right. \\ &\quad \cdot \int_{a/q}^{\infty} \log\{y\} y^{N+j-1} e^{-\nu y} dy \left. \right\}.\end{aligned}$$

Substituting $\nu y = u$,

$$\begin{aligned}\tilde{\rho}(I) &= q \exp \left\{ \frac{\hat{c}}{(N-1)!} \sum_{k=0}^{\infty} \frac{\alpha^k}{(N+k)!} (k+1) \sum_{j=0}^k \binom{k}{j} (-a)^{k-j} \right. \\ &\quad \cdot \left(\frac{1}{\nu} \right)^j \left(-\log\{\nu\} \underbrace{\int_{a\nu/q}^{\infty} u^{N+j-1} e^{-u} du}_{=\Gamma(N+j, a\nu/q)} \right. \\ &\quad \left. \left. + \underbrace{\int_{a\nu/q}^{\infty} \log\{u\} u^{N+j-1} e^{-u} du}_{=\partial/\partial(N+j)\Gamma(N+j, a\nu/q)} \right) \right\},\end{aligned}$$

where $\Gamma(s, x)$ is the upper incomplete Gamma function. Since

$$\frac{\partial \Gamma(s, x)}{\partial s} = \log\{x\} \Gamma(s, x) + x G_{2,3}^{3,0} \left[\begin{matrix} 0, 0, \dots, 0 \\ s-1, -1, \dots, -1 \end{matrix} \middle| x \right],$$

we obtain the result. \square

We have the following special cases.

COROLLARY 1. 1. For risk neutrality $\alpha = 0$, we obtain

$$\begin{aligned}\tilde{\rho}(I) &= \exp \left\{ \log\{a\} e^{-(a\nu)/q} \sum_{k=0}^{N-1} \frac{(a\nu/q)^k}{k!} + \frac{1}{(N-1)!} \frac{a\nu}{q} \right. \\ &\quad \cdot G_{2,3}^{3,0} \left[\begin{matrix} 0, 0, \dots, 0 \\ N-1, -1, \dots, -1 \end{matrix} \middle| \frac{a\nu}{q} \right] \left. \right\}.\end{aligned}$$

2. For the asset level $a = 0$, we obtain

$$\tilde{\rho}(I) = \frac{q}{\nu} \exp \left\{ \frac{\sum_{k=0}^{\infty} ((k+1)/(N+k)) (\alpha/\nu)^k \Psi(N+k)}{\sum_{k=0}^{\infty} ((k+1)/(N+k)) (\alpha/\nu)^k} \right\},$$

where $\Psi(\cdot)$ is the digamma function.

3. For asset level $a = 0$ and risk neutrality $\alpha = 0$, we obtain

$$\tilde{\rho}(I) = \frac{q}{\nu} \exp\{\Psi(N)\}.$$

For comparison, conventional risk measures can be evaluated as

$$\text{VaR}_\varepsilon = q \Gamma_{N, \nu}^{-1}(1 - \varepsilon),$$

$$\begin{aligned}\text{ES}_\varepsilon &= \frac{qN}{\nu} \frac{[1 - \Gamma_{N+1, \nu}(\Gamma_{N, \nu}^{-1}(1 - \varepsilon))]}{\varepsilon}, \\ \text{and } \log\{\mathbb{E}[\exp\{I\}]\} &= N \log\left\{ \frac{\nu}{\nu - q} \right\}.\end{aligned}$$

A.3. Technical Appendix to §3.2: Heterogeneous Pareto Losses

As introduced in §3.2, we assume a quadratic utility function,

$$U_i(x) = -x^2 + 2b_i x, \quad b_i \geq 0,$$

and we normalize the wealth w_i to zero, since adjusting w_i is equivalent to adjusting b_i . Hence,

$$\begin{aligned}\mathbb{E}[U_i] &= \mathbb{E} \left[U_i \left(-p_i - L_i(1 - q_i) \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}} \right. \right. \\ &\quad \left. \left. - L_i \left(1 - \frac{q_i a}{q_1 L_1 + q_2 L_2} \right) \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right) \right] \\ &= p_i^2 + 2A_i p_i - B_i,\end{aligned}$$

where

$$\begin{aligned}A_i &= b_i + (1 - q_i) \mathbb{E}[L_i \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}}] + \mathbb{E}[L_i \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}}] \\ &\quad - q_i a \mathbb{E} \left[\frac{L_i}{q_1 L_1 + q_2 L_2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right],\end{aligned}$$

and

$$\begin{aligned}B_i &= \frac{2\sigma_i^2}{(\alpha_i - 2)(\alpha_i - 1)} + \frac{2b_i \sigma_i}{(\alpha_i - 1)} - (1 - q_i)^2 \mathbb{E}[L_i^2 \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}}] \\ &\quad - \mathbb{E}[L_i^2 \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}}] + 2q_i a \mathbb{E} \left[\frac{L_i^2}{q_1 L_1 + q_2 L_2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right] \\ &\quad - q_i^2 a^2 \mathbb{E} \left[\frac{L_i^2}{(q_1 L_1 + q_2 L_2)^2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right] \\ &\quad - 2b_i(1 - q_i) \mathbb{E}[L_i \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}}] - 2b_i \mathbb{E}[L_i \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}}] \\ &\quad + 2b_i q_i a \mathbb{E} \left[\frac{L_i}{q_1 L_1 + q_2 L_2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right].\end{aligned}$$

Thus, since the constraint (3) binds at the optimum,

$$p_i = p_i(q_1, q_2, a) = \sqrt{B_i + A_i^2} - A_i,$$

and we can represent the optimization problem as

$$\begin{aligned}\max_{q_1, q_2, a} \{ &p_1(q_1, q_2, a) + p_2(q_1, q_2, a) - q_1 \mathbb{E}[L_1 \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}}] \\ &- q_2 \mathbb{E}[L_2 \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}}] - a[\mathbb{P}(q_1 L_1 + q_2 L_2 > a) + \tau] \},\end{aligned}$$

which can be solved using numerical methods or by (numerically) solving the first-order conditions. The integrals corresponding to the expected values must be solved numerically, although in most cases the (improper) double integrals can be reduced to definite one-dimensional integrals via standard integration by parts and substitution techniques.

For the allocations, we obtain

$$\begin{aligned}\tilde{\phi}_i &= \frac{1}{D} \mathbb{E} \left[\mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \frac{L_i}{q_1 L_1 + q_2 L_2} \right. \\ &\quad \cdot \left. \frac{q_1 L_1}{q_1 L_1 + q_2 L_2} \frac{2(p_1 + L_1(1 - (q_1 a)/(q_1 L_1 + q_2 L_2)) + b_1)}{v_1'} \right]\end{aligned}$$

$$\begin{aligned} & + \frac{q_2 L_2}{q_1 L_1 + q_2 L_2} \frac{2(p_2 + L_2(1 - (q_2 a)/(q_1 L_1 + q_2 L_2)) + b_2)}{v'_2} \Bigg\} \\ & = \frac{1}{D} \sum_{i=1}^2 \frac{2(p_i + b_i)q_i}{v'_i} \mathbb{E} \left[\frac{L_i^2}{(q_1 L_1 + q_2 L_2)^2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right] \\ & + \frac{2q_i}{v'_i} \mathbb{E} \left[\frac{L_i^3}{(q_1 L_1 + q_2 L_2)^2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right] \\ & - \frac{2q_i^2 a}{v'_i} \mathbb{E} \left[\frac{L_i^3}{(q_1 L_1 + q_2 L_2)^3} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right], \end{aligned}$$

where

$$\begin{aligned} v'_i & = 2b_i + \mathbb{E} \left[U'_i \left(-p_i - L_i(1 - q_i) \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}} \right. \right. \\ & \quad \left. \left. - L_i \left(1 - \frac{q_i a}{q_1 L_1 + q_2 L_2} \right) \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right) \right] \\ & = 2p_i + 2(1 - q_i) \mathbb{E}[L_i \mathbf{1}_{\{q_1 L_1 + q_2 L_2 \leq a\}}] + 2 \mathbb{E}[L_i \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}}] \\ & \quad - 2q_i a, \mathbb{E} \left[\frac{L_i}{q_1 L_1 + q_2 L_2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right], \end{aligned}$$

and the denominator

$$\begin{aligned} D & = \mathbb{E} \left[\mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \sum_{i=1}^2 \frac{q_i L_i}{q_1 L_1 + q_2 L_2} \right. \\ & \quad \left. \cdot \frac{2(p_i + L_i(1 - (q_i a)/(q_1 L_1 + q_2 L_2)) + b_i)}{v'_i} \right] \\ & = \sum_{i=1}^2 \frac{2(p_i + b_i)q_i}{v'_i} \mathbb{E} \left[\frac{L_i}{(q_1 L_1 + q_2 L_2)^2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right] \\ & + \frac{2q_i}{v'_i} \mathbb{E} \left[\frac{L_i}{(q_1 L_1 + q_2 L_2)^2} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right] \\ & - \frac{2q_i^2 a}{v'_i} \mathbb{E} \left[\frac{L_i^2}{(q_1 L_1 + q_2 L_2)^3} \mathbf{1}_{\{q_1 L_1 + q_2 L_2 > a\}} \right]. \end{aligned}$$

For risk measures of one risk, we obtain

$$\begin{aligned} \text{VaR}_\varepsilon(L_i) & = \sigma_i \left[\left(\frac{1}{\varepsilon} \right)^{1/\alpha_i} - 1 \right], \\ \text{ES}_\varepsilon(L_i) & = \frac{1}{\varepsilon} \left[\text{VaR}_\varepsilon(L_i) \left(1 + \frac{\text{VaR}_\varepsilon(L_i)}{\sigma_i} \right)^{-\alpha_i} \right. \\ & \quad \left. + \frac{\sigma_i}{\alpha_i - 1} \left(1 + \frac{\text{VaR}_\varepsilon(L_i)}{\sigma_i} \right)^{-\alpha_i + 1} \right]. \end{aligned}$$

For the new risk measure as well as for bivariate risk measures and corresponding allocations, it is necessary to evaluate integrals numerically although simplifications again are possible.

Appendix B. Extensions of the Model

B.1. Allocation for a Risk-Averse Company

Consider the same setup as in §2, but now assume the company is risk averse and maximizes the expected utility of its payoff based on an increasing concave utility function $V(\cdot)$. Thus, the firm's optimization problem takes the form

$$\max_{a, \{q_i\}, \{p_i\}} \mathbb{E} \left[V \left(\sum_{i=1}^N p_i - \sum_{i=1}^N R_i - \tau a \right) \right]$$

subject to $v_i \geq \gamma_i \forall i$. Then, the utility function V enters the first-order conditions in the following way:

$$\begin{aligned} [a] \quad & -\mathbb{E}[V'(\mathbf{1}_{\{I > a\}} + \tau)] + \sum_k \lambda_k \frac{\partial v_k}{\partial a} = 0, \\ [q_i] \quad & -\mathbb{E} \left[V' \frac{\partial I_i}{\partial q_i} \mathbf{1}_{\{I \leq a\}} \right] + \sum_k \lambda_k \frac{\partial v_k}{\partial q_i} = 0, \\ [p_i] \quad & \mathbb{E}[V'] - \lambda_i v'_i = 0. \end{aligned}$$

Hence, again using $[p_i]$ in $[q_i]$, the marginal cost condition (4) gives

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} & = \frac{\lambda_i}{\mathbb{E}[V']} \frac{\partial v_i}{\partial q_i} = \frac{1}{\mathbb{E}[V']} \left[-\sum_{k \neq i} \frac{\mathbb{E}[V']}{v'_k} \frac{\partial v_k}{\partial q_i} + \mathbb{E} \left[V' \frac{\partial I_i}{\partial q_i} \mathbf{1}_{\{I \leq a\}} \right] \right] \\ & = \mathbb{E} \left[V' / \mathbb{E}[V'] \mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U'_i}{v'_i} \frac{a I_i}{I} \frac{\partial I_i / \partial q_i}{I} \right] \\ & \quad + \underbrace{\frac{\mathbb{E}[\mathbf{1}_{\{I > a\}} ((\partial I_i / \partial q_i) / I) \sum_k (U'_k / v'_k) (I_k / I)]}{\mathbb{E}[\mathbf{1}_{\{I > a\}} \sum_k (U'_k / v'_k) (I_k / I)]}}_{=\hat{\phi}_i} \\ & \quad \cdot \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \sum_k \frac{U'_k}{v'_k} \frac{I_k}{I} \right] a \\ & \stackrel{(I)}{=} \mathbb{E} \left[V' / \mathbb{E}[V'] \mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U'_i}{v'_i} \frac{a I_i}{I} \frac{\partial I_i / \partial q_i}{I} \right] \\ & \quad + \underbrace{\hat{\phi}_i [\mathbb{E}[V' / \mathbb{E}[V'] \mathbf{1}_{\{I > a\}}] a + \tau a]}_{(II)}. \end{aligned}$$

Therefore, we obtain exactly the same allocation fraction $q_i \hat{\phi}_i$ as in the basic setup from §2. However, the company's marginal utilities now enter the valuation of indemnities in solvent states (I) as well as the capital costs in (II). In fact, defining the firm's valuation measure $\hat{\mathbb{P}}$ that accounts for the company's preferences for risk taking via the Radon–Nikodym derivative $\partial \hat{\mathbb{P}} / \partial \mathbb{P} = V' / \mathbb{E}[V']$, we can interpret (II) as the default probability under $\hat{\mathbb{P}}$ and (I) as the expected value under $\hat{\mathbb{P}}$ of indemnities in solvent states, since it is solvent states where the company in our model faces risk (the payoff in default states will always be a).

B.2. Allocation in a Securities Market Equilibrium

To keep the setup simple, we limit our considerations to a one-period market with a finite number of securities (M), each security with potentially distinct payoffs in X states, and we assume that the risk-free rate is zero. More specifically, let $\Omega^{(S)} = \{\omega_1^{(S)}, \dots, \omega_X^{(S)}\}$ be the set of these states with associated sigma-algebra $\mathcal{F}^{(S)}$ given by its power set, and let $p_j^{(S)} = \mathbb{P}(\{\omega_j^{(S)}\})$ denote the associated (physical) probabilities. Let D be the $M \times X$ matrix, with D_{ij} describing the payoff of the i th security in state $\omega_j^{(S)}$, where we assume

$$\text{span}(D) = \mathbb{R}^X.$$

This condition allows us to define unique state prices, consistent with the absence of arbitrage within the securities market, denoted by π_j , $j = 1, \dots, X$ (e.g., Duffie 2010). Thus, any arbitrary menu of securities market substate-contingent consumption can be purchased at time zero. However, it would

be misleading to characterize markets as complete, since $\Omega^{(S)}$ does not provide a full description of the “states of the world.” We also have loss random variables L_i , $1 \leq i \leq N$, that exist on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Thus, we characterize the full probability space as $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$, with

$$\bar{\Omega} = \Omega^{(S)} \times \Omega = \{\bar{\omega} = (\omega^{(S)}, \omega) \mid \omega^{(S)} \in \Omega^{(S)}, \omega \in \Omega\},$$

$$\bar{\mathcal{F}} = \mathcal{F}^{(S)} \vee \mathcal{F}, \quad \text{and}$$

$$\bar{\mathbb{P}}(\bar{A}) = \sum_{j \in \mathcal{T}_A} p_j^{(S)} \times \mathbb{P}(A_j \mid \{\omega_j^{(S)}\})$$

for $\bar{A} = \bigcup_{j \in \mathcal{T}_A} \{\omega_j^{(S)}\} \times A_j \in \bar{F}$ with $A_j \in \mathcal{F}$, $j \in \mathcal{T}_A \subseteq \{1, 2, \dots, X\}$.

Our problem now, however, is that the market is no longer complete, so we need a notion of what insurance liabilities are “worth” to the insurer when they cannot be hedged completely. We make the assumption that the insurance market is “small” relative to the securities market and, for purposes of valuing insurance liabilities, employ the so-called *minimal martingale measure*.²⁰

$$\mathbb{Q}(\bar{A}) = \sum_{j \in \mathcal{T}_A} \pi_j \times \mathbb{P}(A_j \mid \{\omega_j^{(S)}\}), \quad \bar{A} \subseteq \bar{\Omega};$$

i.e., \mathbb{Q} is defined by the Radon–Nikodym derivative $\partial \mathbb{Q} / \partial \mathbb{P}((\omega_j^{(S)}, \omega)) = \pi_j / p_j^{(S)}$.

Consumer utility now depends on the individual’s chosen security market allocation,

$$v_i = \mathbb{E}^{\mathbb{P}}[U_i(W_i - p_i - L_i + R_i)]$$

$$\text{with } v'_i = \mathbb{E}^{\mathbb{P}}[U'_i(W_i - p_i - L_i + R_i)],$$

where W_i is $\mathcal{F}^{(S)}$ -measurable with $w_{ij} = W_i(\omega_j^{(S)})$ and $\sum_j \pi_j w_{ij} = w_i$, whereas L_i is $\bar{\mathcal{F}}$ -measurable. Similarly, the recovery R_i now depends both on insurance loss activity as well as on portfolio decisions made within the insurance company. To elaborate on this, the budget constraint of the insurance company may be expressed as

$$a = \sum_j \pi_j K_j a \Rightarrow 1 = \sum_j \pi_j K_j,$$

where $K_j a$ reflects consumption purchased in the securities market state $\omega_j^{(S)}$ or, more precisely, in the states of the world $\bar{\Omega}_j = \{\bar{\omega} = (\omega^{(S)}, \omega) \mid \omega^{(S)} = \omega_j^{(S)}\}$. We write K to denote the corresponding $\mathcal{F}^{(S)}$ -measurable random variable. Consumer i ’s recovery can then be expressed as

$$R_i = \min \left\{ I_i, \frac{K a}{I} I_i \right\},$$

²⁰ As indicated by Björk and Slinko (2006), the minimal martingale measure provides us with a canonical benchmark for pricing in incomplete markets by valuing risk that is not spanned by traded securities in a risk-neutral way. It emerges as the optimal martingale measure given various criteria proposed in the mathematical finance literature if the market for insurance risk is small relative to financial markets (Bauer et al. 2013), and it also appears in other settings throughout the finance literature. For instance, the minimal martingale measure coincides with the “hedge-neutral measure” in Basak and Chabakauri (2010), and it arises as the limit of Cochrane and Saá-Requejo (2000) price bounds for Itô price processes as shown by Černý (2003).

and the fair valuation of claims is thus

$$e_i = \mathbb{E}^{\mathbb{Q}}[R_i] = \mathbb{E}^{\mathbb{Q}}[R_i \mathbf{1}_{\{I \leq K a\}}] + \mathbb{E}^{\mathbb{Q}}[R_i \mathbf{1}_{\{I > K a\}}].$$

As before, we can now derive the capital allocation according to the company’s marginal cost by working through its optimization problem. More precisely, in this setting, the firm’s problem becomes

$$\max_{a, \{q_i\}, \{p_i\}, \{K_j\}, \{w_{ij}\}} \left\{ \sum p_i - \sum e_i - \tau a \right\},$$

subject to $v_i \geq \gamma_i$, $\sum_j \pi_j K_j = 1$, and $\sum_j \pi_j w_{ij} = w_i$.

In addition to a new set of optimality conditions connected with $\{K_j\}$ and $\{w_{ij}\}$, we have the same set of first-order conditions:

$$[a] \quad - \sum_k \frac{\partial e_k}{\partial a} - \tau + \sum_k \lambda_k \frac{\partial v_k}{\partial a} = 0,$$

$$[q_i] \quad - \sum_k \frac{\partial e_k}{\partial q_i} + \sum_k \lambda_k \frac{\partial v_k}{\partial q_i} = 0,$$

$$[p_i] \quad 1 - \lambda_i \frac{\partial v_i}{\partial w} = 0.$$

The first-order condition for $\{w_{ij}\}$ is

$$[w_{ij}] \quad \lambda_i \frac{\partial v_i}{\partial w_{ij}} - \eta_i \pi_j = 0$$

$$\Leftrightarrow \lambda_i p_j^{(S)} \mathbb{E}^{\mathbb{P}}[U'_i(W_i - p_i - L_i + R_i) \mid \omega_j^{(S)}] - \eta_i \pi_j = 0,$$

where $\{\eta_i\}$ are the Lagrange multipliers for the individual wealth constraints. Since

$$0 = \sum_j (\lambda_i p_j^{(S)} \mathbb{E}^{\mathbb{P}}[U'_i(W_i - p_i - L_i + R_i) \mid \omega_j^{(S)}] - \eta_i \pi_j) \\ = \lambda_i (\partial v_i / \partial w) - \eta_i,$$

with $[p_i]$ we obtain $\eta_i \equiv 1$. Thus, we also have

$$\frac{\pi_j}{p_j^{(S)}} = \frac{\mathbb{E}^{\mathbb{P}}[U'_i(W_i - p_i - L_i + R_i) \mid \omega_j^{(S)}]}{\partial v_i / \partial w}. \quad (\text{B1})$$

As before, we seek a pricing function satisfying

$$\frac{\partial p_i^*}{\partial q_i} v'_i = \frac{\partial v_i}{\partial q_i}.$$

Using $[q_i]$ and $[p_i]$, we obtain

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \sum_k \frac{\partial e_k}{\partial q_i} - \sum_{k \neq i} \frac{\partial v_k / \partial q_i}{v'_k} \\ &= \mathbb{E}^{\mathbb{Q}} \left[\mathbf{1}_{\{I \leq K a\}} \frac{\partial I_i}{\partial q_i} \right] - \frac{\mathbb{E}^{\mathbb{P}}[\mathbf{1}_{\{I > K a\}} U'_i((K a) / I^2) I_i (\partial I_i / \partial q_i)]}{v'_i} \\ &\quad + \mathbb{E}^{\mathbb{P}} \left[\mathbf{1}_{\{I > K a\}} \sum_k \frac{U'_k K a}{v'_k I^2} I_k \frac{\partial I_i}{\partial q_i} \right] \\ &= \mathbb{E}^{\mathbb{Q}} \left[\mathbf{1}_{\{I \leq K a\}} \frac{\partial I_i}{\partial q_i} \right] - \frac{\mathbb{E}^{\mathbb{P}}[\mathbf{1}_{\{I > K a\}} U'_i((K a) / I^2) I_i (\partial I_i / \partial q_i)]}{v'_i} \\ &\quad + \frac{\mathbb{E}^{\mathbb{P}}[\mathbf{1}_{\{I > K a\}} \sum_k (U'_k / v'_k) (K / I^2) I_k (\partial I_i / \partial q_i)]}{\sum_k (\partial v_k / \partial a) / v'_k} \sum_k \frac{\partial v_k / \partial a}{v'_k} a \\ &\stackrel{[a]}{=} \mathbb{E}^{\mathbb{Q}} \left[\mathbf{1}_{\{I \leq K a\}} \frac{\partial I_i}{\partial q_i} \right] - \frac{\mathbb{E}^{\mathbb{P}}[\mathbf{1}_{\{I > K a\}} U'_i((K a) / I^2) I_i (\partial I_i / \partial q_i)]}{v'_i} \\ &\quad + \tilde{\phi}_i a [\mathbb{E}^{\mathbb{Q}}[K \mathbf{1}_{\{I > K a\}}] + \tau], \end{aligned}$$

where

$$\begin{aligned}\tilde{\phi}_i &= \frac{\mathbb{E}^{\mathbb{P}}[\mathbf{1}_{\{I > Ka\}} \sum_k (U'_k/v'_k)(K/I^2)I_k(\partial I_i/\partial q_i)]}{\mathbb{E}^{\mathbb{P}}[\mathbf{1}_{\{I > Ka\}} \sum_k (U'_k/v'_k)K(I_k/I)]} \\ &\stackrel{(B1)}{=} \frac{\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{I > Ka\}} \sum_k (U'_k/(\mathbb{E}^{\mathbb{P}}[U'_k | \omega^{(S)}]))(K/I^2)I_k(\partial I_i/\partial q_i)]}{\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{I > Ka\}} \sum_k (U'_k/(\mathbb{E}^{\mathbb{P}}[U'_k | \omega^{(S)}]))K(I_k/I)]}.\end{aligned}$$

This is exactly Equation (8) in the main text, where we write $A = Ka$ for the state-dependent assets and $\mathbb{E}^{\mathbb{P}}[U'_k | S] = \mathbb{E}^{\mathbb{P}}[U'_k | \omega^{(S)}]$ for the marginal utilities given the security prices $S(\omega_j^{(S)}) = D_j$.

As pointed out there, we obtain an allocation result similar to that of the previous section. The cost of capital,

$$[\mathbb{E}^{\mathbb{Q}}[Ka\mathbf{1}_{\{I > Ka\}}] + \tau a],$$

which now reflects state prices and the company's asset allocation, can be decomposed according to the marginal cost for each of the individual exposures as

$$[\mathbb{E}^{\mathbb{Q}}[Ka\mathbf{1}_{\{I > Ka\}}] + \tau a] = \sum_i \tilde{\phi}_i q_i [\mathbb{E}^{\mathbb{Q}}[Ka\mathbf{1}_{\{I > Ka\}}] + \tau a]. \quad (B2)$$

The key difference is that the marginal pricing rule now only applies in every *branch* of the securities market where the incompleteness becomes material. In particular, after adjusting for state prices by conditioning on each branch, capital allocation weights are still determined by consumer marginal utility.

In the limiting case of a complete market (i.e., the case when L_i and R_i are $\mathcal{F}^{(S)}$ -measurable so that we can write $l_{ij} = L_i(\omega_j^{(S)})$ and $r_{ij} = R_i(\omega_j^{(S)})$, $l_{ij}, r_{ij} \in \mathbb{R}$), clearly $U'_i(w_{ij} - p_i - l_{ij} + r_{ij})$ is $\mathcal{F}^{(S)}$ -measurable, so that we obtain

$$\begin{aligned}\tilde{\phi}_i &= \frac{\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{I > Ka\}} \sum_k (K/I^2)I_k(\partial I_i/\partial q_i)]}{\mathbb{E}^{\mathbb{Q}}[K\mathbf{1}_{\{I > Ka\}}]} \\ &= \frac{\mathbb{E}^{\mathbb{Q}}[K((\partial I_i/\partial q_i)/I) | I > Ka]}{\mathbb{E}^{\mathbb{Q}}[K | I > Ka]},\end{aligned} \quad (B3)$$

and thus

$$q_i \tilde{\phi}_i \mathbb{E}^{\mathbb{Q}}[Ka\mathbf{1}_{\{I > Ka\}}] = \mathbb{E}^{\mathbb{Q}}\left[\frac{Ka}{I} \mathbf{1}_{\{I > Ka\}}\right]$$

is the fair (risk-neutral) price of the recovery. This result—where capital is allocated to consumers in proportion to their share of the total market value of recoveries—is the same allocation result as in Ibragimov et al. (2010).

B.3. Multiple Periods and Regulatory Constraint

To account for multiple periods, let now L_i^t denote the loss incurred by consumer i , $i \in \{1, 2, \dots, N\}$, in period t , $t \in \{1, 2, \dots\}$. Here we assume that L_i^t , $t > 0$ —for fixed i —are independent and identically distributed, and we define the relevant filtration $\mathbf{F} = (\mathcal{F}_t)_{t \geq 0}$ via $\mathcal{F}_t = \sigma(L_i^s, i \in \{1, 2, \dots, N\}, s \leq t)$. The firm determines the optimal level of assets, a_t , at the beginning of each period (i.e., (a_t) is \mathbf{F} -predictable) for a period cost of τa_t . Similarly to before, the company chooses \mathbf{F} -predictable amounts q_i^t in $I_i^t = I(L_i^t, q_i^t) = q_i^t L_i^t$ and prices p_i^t at the beginning of the period, and we denote the total claims by $I^t = \sum_{j=1}^N I_j^t$.

Now the company defaults if $I^t > a_t$, so that the recovery paid to each consumer is $R_i^t = \min\{I_i^t, (a_t/I^t)I_i^t\}$ and the

company shuts down in case of default, i.e., shareholders do not have access to future profit flows.²¹ The consumer's utility in period t is given by

$$v_i^t(a_t, w_i^t - p_i^t, q_i^t, \dots, q_N^t) = \mathbb{E}_{t-1}[U_i(w_i^t - p_i^t - L_i^t + R_i^t)],$$

where, for simplicity, we assume that wealth is homogeneous across periods, i.e., $w_i^t \equiv w_i$.²² Note that the consumer purchases one-period contracts at each point in time t . Hence, the participation constraint must be satisfied period by period, and we obtain for the company's profit maximization problem

$$\begin{aligned}\max_{\{a_t\}, \{q_i^t\}, \{p_i^t\}} V_0 &= \sum_{t=1}^{\infty} \mathbb{E}\left[\mathbf{1}_{\{I^1 \leq a_1, \dots, I^{t-1} \leq a_{t-1}\}}\right. \\ &\quad \times \left(\sum_i p_i^t - \sum_i \mathbb{E}_{t-1}[R_i^t] - \tau a_t\right)\end{aligned}$$

with constraints

$$v_i^t(a_t, w_i^t - p_i^t, q_i^t, \dots, q_N^t) \geq \gamma_i \quad \forall i, \forall t, \quad (B4)$$

$$s(q_i^t, \dots, q_N^t) \leq a_t \quad \forall t, \quad (B5)$$

and where $\{a_t\}$, $\{q_i^t\}$, and $\{p_i^t\}$ are predictable.

In addition to the participation constraints (B4), we consider a (period) solvency constraint imposed by the regulator (B5). Here, s is assumed to be an externally supplied risk measure with a set threshold dictating the requisite capitalization in each period for the firm as a function of the risk it has taken, where we assume the risk measure to be differentiable and positive homogeneous.

Under the assumptions above, it is clear that there exists an optimal stationary policy $(a, \{q_i\}, \{p_i\})$ that solves the Bellman equation:

$$V = \max_{a, \{q_i\}, \{p_i\}} \left\{ \sum_i p_i - \sum_i \underbrace{\mathbb{E}[R_i^t]}_{e_i} - \tau a + \mathbb{P}[I^t \leq a] \times V \right\} \quad (B6)$$

under conditions (B4) and (B5). Hence, we have a similar program as in the basic setup from §2, where the primary difference is the last term in (B6) involving the value of the company.

The first-order conditions of program (B6) are²³

$$\begin{aligned}[a] \quad & -\sum_k \frac{\partial e_k}{\partial a} - \tau + V f_I(a) + \sum_k \lambda_k \frac{\partial v_k}{\partial a} + \xi = 0, \\ [q_i] \quad & -\sum_k \frac{\partial e_k}{\partial q_i} - V \mathbb{E}\left[\frac{\partial I_i(L_i, q_i)}{\partial q_i} \frac{\mathbf{1}_{\{I=a\}}}{dt}\right] \\ & + \sum_k \lambda_k \frac{\partial v_k}{\partial q_i} - \xi \frac{\partial s}{\partial q_i} = 0, \\ [p_i] \quad & 1 - \lambda_i \frac{\partial v_i}{\partial w} = 0,\end{aligned} \quad (B7)$$

²¹ Alternatively, it is possible to allow the distressed company to raise additional funds in the case of default at a higher (or even increasing) cost akin to Froot et al. (1993) and Froot and Stein (1998). Here, we limit our considerations to this simple case and leave the further exploration of alternative settings for future research.

²² Typically consumers will form utilities over consumption in multiple periods. In particular, future (random) losses will also be material. Thus, here U should rather be interpreted as a value function (of end-of-period wealth) than as a utility function (of end-of-period consumption).

²³ For simplicity, we omit the “ t ” superscripts and subscripts in cases where no ambiguity arises.

where λ_k and ξ denote the Lagrange multipliers for the first set of conditions and the second condition, respectively, and f_I denotes the probability density function of I .

The period marginal pricing condition (4) still reads

$$\frac{\partial p_i^*}{\partial q_i} \frac{\partial v_i}{\partial w} = \frac{\partial v_i}{\partial q_i},$$

so that with $[q_i]$ and $[p_i]$ as before

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \frac{1}{\partial v_i / \partial w} \left[\frac{\partial v_i}{\partial q_i} \right] = \sum_k \frac{\partial e_k}{\partial q_i} + \xi \frac{\partial s}{\partial q_i} + V \mathbb{E} \left[\frac{\partial I_i(L_i, q_i)}{\partial q_i} \frac{\mathbf{1}_{\{I=a\}}}{dt} \right] \\ &\quad - \sum_{k \neq i} \frac{\partial v_k / \partial q_i}{v'_k} \\ &\stackrel{[a]}{=} \mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U'_i a I_i}{v'_i I} \frac{\partial I_i / \partial q_i}{I} \right] \\ &\quad + \left[\mathbb{P}(I > a) + \tau - \sum_k \frac{\partial v_k / \partial a}{v'_k} - V f_I(a) \right] \frac{\partial s}{\partial q_i} \\ &\quad + V f_I(a) \mathbb{E} \left[\frac{\partial I_i}{\partial q_i} \middle| I = a \right] + \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{\partial I_i / \partial q_i}{I} \sum_k \frac{U'_k a I_k}{v'_k I} \right]. \end{aligned}$$

Since

$$\sum_k \frac{\partial v_k / \partial a}{v'_k} = \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \sum_k \frac{U'_k I_k}{v'_k I} \right],$$

we obtain

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i(L_i, q_i)}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U'_i a I_i}{v'_i I} \frac{\partial I_i / \partial q_i}{I} \right] \\ &\quad + \tilde{\theta}_i [V f_I(a)] + \frac{\partial s}{\partial q_i} \left[\mathbb{P}(I > a) + \tau - \sum_k \frac{\partial v_k / \partial a}{v'_k} - V f_I(a) \right] \\ &\quad + \tilde{\phi}_i a \left[\sum_k \frac{\partial v_k / \partial a}{v'_k} \right], \end{aligned} \quad (\text{B8})$$

where

$$\begin{aligned} \tilde{\theta}_i &= \mathbb{E} \left[\frac{\partial I_i(L_i, q_i)}{\partial q_i} \middle| I = a \right] \quad \text{and} \\ \tilde{\phi}_i &= \frac{\mathbb{E}[\mathbf{1}_{\{I > a\}} ((\partial I_i / \partial q_i) / I) \sum_k (U'_k / v'_k) (I_k / I)]}{\mathbb{E}[\mathbf{1}_{\{I > a\}} \sum_k (U'_k / v'_k) (I_k / I)]}, \end{aligned}$$

which is exactly Equation (9) in the main text.

Two special institutional situations were discussed in more detail:

- In the limiting case of perfect competition, the firm value V approaches zero, as does the weight associated with the allocation $\tilde{\theta}_i$. Hence, the limiting allocation ends up being a weighted average of that dictated by internal counterparties and that dictated by regulators:

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] - \mathbb{E} \left[\mathbf{1}_{\{I > a\}} \frac{U'_i a I_i}{v'_i I} \frac{\partial I_i / \partial q_i}{I} \right] + \frac{\partial s}{\partial q_i} \\ &\quad \cdot \left[\mathbb{P}(I > a) + \tau - \sum_k \frac{\partial v_k / \partial a}{v'_k} \right] + \tilde{\phi}_i \times a \times \left[\sum_k \frac{\partial v_k / \partial a}{v'_k} \right]. \end{aligned}$$

- In case of imperfect competition (so that $V > 0$), but where consumers are fully insured by a guaranty fund scheme and where any regulatory constraint is nonbinding, consumers are indifferent to the capitalization of the firm, and the firm is not affected by the solvency constraint at the

margin. Mathematically, this means that $\sum_k ((\partial v_k / \partial a) / v'_k) = 0$ and $\mathbb{P}(I > a) + \tau - \sum_k ((\partial v_k / \partial a) / v'_k) - V f_I(a) = 0$, so that

$$\begin{aligned} \frac{\partial p_i^*}{\partial q_i} &= \mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] + V f_I(a) \mathbb{E} \left[\frac{\partial I_i}{\partial q_i} \middle| I = a \right] \\ &= \mathbb{E} \left[\mathbf{1}_{\{I \leq a\}} \frac{\partial I_i}{\partial q_i} \right] + \left[\mathbb{P}(I > a) + \tau \right] \mathbb{E} \left[\frac{\partial I_i}{\partial q_i} \middle| I = a \right], \end{aligned} \quad (\text{B9})$$

where the latter equality follows from the first-order condition for a (see Equation (B7)).

As before, we can recover the risk measure whose gradient yields the correct capital allocations. However, the more complicated form of (9) and (B8) suggests a more complicated risk measure. Indeed, the marginal cost allocation ends up being determined by the weighted average of three component risk measures, each corresponding to one of the three components of the risk penalties (see the main text for details).

Given previous arguments, two of the component risk measures are relatively straightforward. The component measure emanating from counterparty preferences was derived in §2, and the component measure corresponding to the regulatory constraint is simply the regulatory risk measure s . To isolate the third, consider the environment underlying (B9), where risk taking is only hindered by shareholder concerns about profit flows in future periods. Taking the confidence level $\varepsilon = \mathbb{P}(I > a)$, it turns out that in this case we can derive the capital allocation as

$$q_i \tilde{\theta}_i = q_i \mathbb{E} \left[\frac{\partial I_i}{\partial q_i} \middle| I = a \right] = q_i \frac{\partial}{\partial q_i} \text{VaR}_\varepsilon(I),$$

i.e., the supporting risk measure is VaR, which thus arises endogenously within our framework (see Garman 1997 for VaR-based allocations).

General institutional circumstances yield an allocation that combines the three foregoing risk measures. Specifically, we have the following expression for the capital cost allocated to consumer i in (9) and (B8):

$$\begin{aligned} &\tilde{\theta}_i [V f_I(a)] + \tilde{\phi}_i a \left[\sum_k \frac{\partial v_k / \partial a}{v'_k} \right] \\ &\quad + \frac{\partial s}{\partial q_i} \left[\mathbb{P}(I > a) + \tau - \sum_k \frac{\partial v_k / \partial a}{v'_k} - V f_I(a) \right] \\ &= \frac{\partial \text{VaR}_\varepsilon}{\partial q_i} [V f_I(a)] + \frac{(\partial / \partial q_i) \tilde{\rho}}{\tilde{\rho}} a \left[\sum_k \frac{\partial v_k / \partial a}{v'_k} \right] \\ &\quad + \frac{\partial s}{\partial q_i} \left[\mathbb{P}(I > a) + \tau - \sum_k \frac{\partial v_k / \partial a}{v'_k} - V f_I(a) \right] \\ &= \left(\zeta_1 \frac{\partial \text{VaR}_\varepsilon}{\partial q_i} + \zeta_2 \frac{(\partial / \partial q_i) \tilde{\rho}}{\tilde{\rho}} a + (1 - \zeta_1 - \zeta_2) \frac{\partial s}{\partial q_i} \right) \\ &\quad \cdot [\mathbb{P}(I > a) + \tau], \end{aligned}$$

where

$$\zeta_1 = \frac{V f_I(a)}{\mathbb{P}(I > a) + \tau} \quad \text{and} \quad \zeta_2 = \frac{\sum_k (\partial v_k / \partial a) / v'_k}{\mathbb{P}(I > a) + \tau}.$$

As indicated in Footnote 7, the gradient allocation principle can formally be derived as the marginal cost allocation when maximizing profits subject to a risk-measure constraint (Bauer

and Zanjani 2013). In this setting with multiple sources of discipline associated with different risk measures, we can similarly contemplate profit maximization subject to three different constraints associated with the three risk measures. In the course of solving the constrained optimization problem, the weights will then be determined by the corresponding Lagrange multipliers—although, of course, it is necessary to adequately parameterize the risk measures, i.e., to choose ε , a , etc., at their optimum values based on the firm's optimization problem. This complication, however, is always present when working with risk measures, since it is necessary to choose the underlying parameters in accordance with the firm's objectives.

References

- Acerbi C (2002) Spectral measures of risk: A coherent representation of subjective risk aversion. *J. Banking Finance* 26:1505–1518.
- Acharya VV, Pedersen LH, Phillipon T, Richardson M (2010) Measuring systemic risk. Working paper, New York University, New York.
- Adrian T, Shin H (2014) Procyclical leverage and value-at-risk. *Rev. Financial Stud.* 27:373–403.
- Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Math. Finance* 9:203–228.
- Baer T, Mehta A, Samandari H (2011) The use of economic capital in performance management for banks: A perspective. McKinsey Working Papers on Risk 24, McKinsey & Company, Washington, DC.
- Basak S, Chabakauri S (2010) Dynamic mean-variance asset allocation. *Rev. Financial Stud.* 23:2970–3016.
- Basak S, Shapiro A (2001) Value-at-risk-based risk management: Optimal policies and asset prices. *Rev. Financial Stud.* 14:371–405.
- Bauer D, Zanjani G (2013) Capital allocation and its discontents. Dionne G, ed. *Handbook of Insurance*, 2nd ed. (Springer, New York), 863–880.
- Bauer D, Phillips RD, Zanjani G (2013) Financial pricing of insurance. Dionne G, ed. *Handbook of Insurance*, 2nd ed. (Springer, New York), 627–645.
- Björk T, Slinko I (2006) Towards a general theory of good-deal bounds. *Rev. Finance* 10:221–260.
- Černý A (2003) Generalised Sharpe ratios and asset pricing in incomplete markets. *Rev. Finance* 7:191–233.
- Cochrane JH, Saá-Queijo J (2000) Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *J. Political Econom.* 108:79–119.
- Daniellson J, Jorgensen BN, Samorodnitsky G, Sarma M, de Vries CG (2013) Fat tails, VaR and subadditivity. *J. Econometrics* 172:283–291.
- Denault M (2001) Coherent allocation of risk capital. *J. Risk* 4:7–21.
- Duffie D (2010) *Dynamic Asset Pricing Theory* (Princeton University Press, Princeton, NJ).
- Embrechts P, Klüppelberg C, Mikosch T (1997) *Modelling Extremal Events for Insurance and Finance* (Springer, New York).
- Erel I, Myers SC, Read JA Jr (2015) A theory of risk capital. *J. Financial Econom.* 118:620–635.
- Farr I, Mueller H, Scanlon M, Stronkhorst S (2008) Economic capital for life insurance companies. Report, Society of Actuaries, Schaumburg, IL. <http://www.soa.org/files/pdf/research-ec-report.pdf>.
- Föllmer H, Knispel T (2011) Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. *Stochastic Dynam.* 11:333–351.
- Föllmer H, Schied A (2002) Convex measures of risk and trading constraints. *Finance Stochastics* 6:429–447.
- Föllmer H, Schied A (2010) Convex risk measures. Cont R, ed. *Encyclopedia of Quantitative Finance* (John Wiley & Sons, Hoboken, NJ), 355–363.
- Froot KA (2007) Risk management, capital budgeting, and capital structure policy for insurers and reinsurers. *J. Risk Insurance* 74:273–299.
- Froot KA, Stein JC (1998) Risk management, capital budgeting, and capital structure policy for financial institutions: An integrated approach. *J. Financial Econom.* 47:55–82.
- Froot KA, Scharfstein DS, Stein JC (1993) Risk management: Coordinating corporate investment and financing policies. *J. Finance* 48:1629–1658.
- Gabaix X (2009) Power laws in economics and finance. *Annual Rev. Econom.* 1:255–294.
- Garman M (1997) Taking VaR to pieces: Presenting a method to express a portfolio's VaR as the sum of its parts. *Risk* 10:70–71.
- Ibragimov R (2009a) Heavy-tailed densities. Durlauf SN, Blume LE, eds. *The New Palgrave Dictionary of Economics Online* (Palgrave Macmillan, London). http://www.dictionaryofeconomics.com/article?id=pde2009_H000191.
- Ibragimov R (2009b) Portfolio diversification and value at risk under thick-tailedness. *Quant. Finance* 9:565–580.
- Ibragimov R, Jaffee D, Walden J (2009) Nondiversification traps in catastrophe insurance markets. *Rev. Financial Stud.* 22:959–993.
- Ibragimov R, Jaffee D, Walden J (2010) Pricing and capital allocation for multiline insurance firms. *J. Risk Insurance* 77:551–578.
- Kalkbrener M (2005) An axiomatic approach to capital allocation. *Math. Finance* 15:425–437.
- Keeler TE, Small KA (1977) Optimal peak load pricing, investment, and service levels on urban expressways. *J. Political Econom.* 85:1–25.
- Kou SG, Peng XH, Heyde CC (2013) External risk measures and Basel Accords. *Math. Oper. Res.* 38:393–417.
- McNeil AJ, Frey R, Embrechts P (2005) *Quantitative Risk Management: Concepts, Techniques, and Tools* (Princeton University Press, Princeton, NJ).
- Myers SC, Read JA (2001) Capital allocation for insurance companies. *J. Risk Insurance* 68:545–580.
- Phillips RD, Cummins JD, Allen F (1998) Financial pricing of insurance in the multiple line insurance company. *J. Risk Insurance* 65:597–636.
- Powers MR (2007) Using Aumann-Shapley values to allocate insurance risk: The case of inhomogeneous losses. *North Amer. Actuarial J.* 11:113–127.
- Rampini AA, Viswanathan S (2010) Collateral, risk management, and the distribution of debt capacity. *J. Finance* 65:2293–2322.
- Rampini AA, Viswanathan S (2014) Financial intermediary capital. Working paper, Duke University, Durham, NC.
- Schmock U, Straumann D (1999) Allocation of risk capital and performance measurement. Presentation at Risk Day 1999, October 22, ETH Zürich, Zürich. <http://www.fam.tuwien.ac.at/~schmock/slides/AllocationSlidesOct1999.pdf>.
- Smith CW, Stulz RM (1985) The determinants of firms' hedging policies. *J. Financial Quant. Anal.* 20:391–405.
- Smith CW, Smithson CW, Wilford SD (1990) *Managing Financial Risk* (Harper Business, New York).
- Stoughton NM, Zechner J (2007) Optimal capital allocation using RAROC and EVA. *J. Financial Intermediation* 16:312–342.
- Tasche D (2004) Allocating portfolio economic capital to sub-portfolios. Dev A, ed. *Economic Capital: A Practitioner's Guide* (Risk Books, London), 275–302.
- Zanjani G (2002) Pricing and capital allocation in catastrophe insurance. *J. Financial Econom.* 65:283–305.
- Zanjani G (2010) An economic approach to capital allocation. *J. Risk Insurance* 77:523–549.