

Measuring and allocating systemic risk*

Markus K. Brunnermeier
Department of Economics and
Bendheim Center for Finance
Princeton University
Princeton, NJ 08544, USA

Patrick Cheridito
Department of Mathematics
and RiskLab
ETH Zurich
8092 Zurich, Switzerland

April 2019

Abstract

In this paper we develop a framework for measuring, allocating and managing systemic risk. SystRisk, our measure of total systemic risk, captures the a priori cost to society for providing tail-risk insurance to the financial system. Our allocation principle distributes the total systemic risk among individual institutions according to their size-shifted marginal contributions. To describe economic shocks and systemic feedback effects we propose a reduced form stochastic model that can be calibrated to historical data. We also discuss systemic risk limits, systemic risk charges and a cap and trade system for systemic risk.

Keywords: Systemic risk measure, systemic risk allocation, feedback effects, shadow prices, systemic risk limits, systemic risk charges, cap and trade.

1 Introduction

The purpose of this paper is to develop a framework for measuring, allocating and managing systemic risk. Financial services play an important role in modern free market economies. Therefore, governments often do not have a choice but to provide support to failing financial institutions in order to protect the broader economy. We address this issue by studying the following two questions: (i) how to measure the total systemic risk generated by the financial sector and (ii) how to allocate the total systemic risk to individual financial institutions?

We view possible government support of financial institutions as an externality and measure total systemic risk by determining its a priori cost to society. To allocate the total systemic risk among individual institutions we propose to use their marginal contributions. To describe economic shocks, spillover, amplification and adverse feedback effects we develop a reduced form stochastic model that can be calibrated to historical data.

The most important features of our approach are:

*We thank Damir Filipović, Eric Jondeau, Anton Korinek, Yvan Lengwyl, Erwan Morellec, Angelo Ranaldo, Frank Riedel, Hyun Shin, Julius Vutz and Christian Wolf for helpful comments.

- It views financial institutions as parts of the financial system. For instance, a bank that behaves as part of a herd is allocated more systemic risk than a bank that acts independently of the rest of the financial sector.
- It relates the financial industry to the real economy. As a consequence, costs of externalities grow faster than proportionally if they become large compared to a country's GDP. Moreover, negative externalities that happen in states of the world where the overall economy is strong cost less than those that happen when the economy is weak.
- It satisfies the so called clone property. That is, if a financial institution is split into n equal parts, the total systemic risk does not change and the risk attributed to the institution also splits into n equal parts.
- It can detect risks that might not be fully reflected in market quotes of traded securities since it considers scenarios in which taxpayers provide support to financial institutions, resulting in distorted market prices.
- It can detect systemic risk in low volatility environments in which the risk of large economic shocks is low but the financial system is prone to negative feedback spirals in case a crisis erupts.
- It is based on a tolerance parameter that can be adjusted over time so as to implement countercyclical regulation.

During financial crises governments might have to support certain parts of the financial system to maintain a properly functioning economy. Whether this happens through loans, restructuring or nationalization, it amounts to costs that ultimately are borne by the taxpayers. SystRisk, our measure of total systemic risk, quantifies the a priori cost of this externality to society. It has the appealing feature that it grows faster than linearly as the exposures become large compared to the real economy. It also gives more weight to losses occurring in states of the world in which the overall economy is depressed. We distribute the total systemic risk among the components of the system based on their marginal contributions and provide a formula that expresses them as expectations with respect to a shadow pricing measure. Typically, they add up to more than the total systemic risk. We propose to reduce them proportionally to exogenous size parameters such as, for instance, the amount of corporate tax an institution payed in the previous year. This yields an allocation principle that attributes systemic risk to individual entities according to the role they play in the financial sector. If a financial firm is part of a herd or exposed to spillover effects, it is assigned more systemic risk than one that is more independent and better prepared to withstand a financial crisis.

Systemic risk can build up during economic booms but only materialize when a crisis erupts. If not taken into account by regulation, this can lead to the situation that a financial system is more vulnerable when observed volatility is low, a phenomenon coined "volatility paradox" in Brunnermeier and Sannikov (2014). The reason for this is that in a low volatility environment financial institutions might be induced to lever up and increase the liquidity mismatch of their balance sheets, that is, they are holding assets of low market liquidity which in a crisis can only be sold at a high discount, and a large part of their funding is short-term and hence, has to

be rolled over frequently.¹ We account for this by developing a simple two-stage reduced form model that can be calibrated to historical data. In a first step, individual firms receive shocks to their balance sheets. They can be the consequence of changes of macroeconomic variables or can originate within individual firms. In a second step, feedback effects play out between the institutions. While past crises have been triggered differently, systemic feedback mechanisms have exhibited similar patterns.² Our model differentiates between direct and indirect feedback effects. Direct effects happen when there exist contractual connections between companies. A defaulting firm directly impacts the balance sheets of its counterparties possibly triggering further defaults. Indirect feedback effects are caused by fire sales of illiquid assets and dry ups of funding liquidity. They are more pronounced if the liquidity mismatch of financial firms is high since this makes them more vulnerable to sudden declines in market liquidity and increases in short-term interest rates. By explicitly including systemic feedback effects, our approach addresses the volatility paradox. SystRisk is also less procyclical than standard risk measures, which typically do not pick up systemic risk in boom times and skyrocket when a crisis erupts, forcing institutions to delever, which in turn increases the risk measure further and worsens the crisis. In addition, we allow for a tolerance parameter that describes how much systemic risk a society is willing to bear. It can be adjusted so that more risk is tolerated during economic downturns to provide conditions for the economy to stabilize and return to a path of growth.

Any systemic risk measure naturally raises the question how it can be used to manage and regulate systemic risk. SystRisk and our allocation principle show where systemic risk is building up and provide guidance in which direction financial regulation is most effective in reducing it. However, financial institutions have their own objectives and will react to new rules. To find a socially optimal level of regulation, one can attempt to model the behavioral response of all market participants and develop a full equilibrium model. But doing this realistically is beyond the scope of this paper. The financial sector consists of many heterogeneous institutions. For example, banks will react differently than insurance companies, which in turn will have a different response from pension funds, sovereign wealth funds and hedge funds. In addition, many firms will try to circumvent regulatory measures. We adopt a more cautious *tatonnement* approach and view financial regulation as an ongoing process rather than the establishment of ultimate rules. In other words, regulation is introduced knowing that it might have to be reoptimized based on observed responses by financial institutions and updates of systemic risk measures. In this procedure the tolerance parameter has a second important role to play. It can be adjusted iteratively to find an optimal tradeoff between financial stability and enough leeway for financial firms to be able to provide the services a modern society depends on.

We discuss three different implementations of financial regulation: (i) setting individual risk limits, (ii) imposing risk charges, and (iii) implementing a cap and trade system. Setting individual risk limits is closest to current financial regulation in most developed countries. But in contrast to classical banking regulation, which views each firm as a separate entity, our method leads to risk limits for individual institutions that depend on the positions of other institutions and how the financial sector is related to the rest of the economy. Alternatively, in the spirit of Pigouvian taxes,

¹See Brunnermeier et al. (2014) and Bai et al. (2018) for a discussion of different types of liquidity and a definition of the liquidity mismatch index.

²We refer to Brunnermeier and Oehmke (2012) for a survey of financial crises and feedback effects that played an important role.

one can impose systemic risk charges depending on the positions the institutions are taking. This can be viewed as an extension of deposit insurance premiums. But again, the risk charges proposed here depend on a financial institution's contribution to systemic risk. Systemic risk limits can be complemented with a cap and trade scheme. For instance, an auction for systemic risk permits can be held, which then can be traded during a certain period of time. This has two advantages. First, it might allow the financial sector to mitigate systemic risk more efficiently, and second, it produces a market price for systemic risk.

Over the last few years, the literature on systemic risk has been growing fast. For a recent survey of different systemic risk measures we refer to Benoit et al. (2017). A large part of it aims at developing statistical measures through which systemic risk can be inferred directly from market quotes of financial securities; see e.g., Chan-Lau and Gravelle (2005), Lehar (2005), Avesani (2005), Avesani et al. (2006), Duan and Wei (2009), Giesecke and Kim (2011), Huang et al. (2012), Billio et al. (2012), Acharya et al. (2012), Allen et al. (2012), Giglio (2016), Adrian and Brunnermeier (2016). Brownlees and Engle (2017) use market as well as balance sheet information. However, market prices have to be interpreted with care since they can be artificially inflated if investors expect that the government will provide support in the event of a crisis. Since our approach is model-based, it allows us to identify the risk of scenarios in which financial institutions end up receiving support from taxpayers. Early models of financial stability put the main emphasis on networks describing direct domino effects caused by defaults on claims between financial institutions; see e.g., Allen and Gale (2000), Freixas et al. (2000), Eisenberg and Noe (2001), Boss et al. (2004), Leitner (2005), Elsinger et al. (2006), Gai and Kapadia (2010), Elliott et al. (2014), Gofman (2017). However, these models do not take all channels of contagion into account and may make the financial sector appear more robust than it is; see e.g., Upper (2007). In the financial crisis of 2007–2009 indirect spillover effects played an important role. In particular, informational contagion and common exposure (virtual links) amplified by fire sales and liquidity spirals exacerbated the crisis; see e.g., Aghion et al. (2000), Cifuentes et al. (2005), Diamond and Rajan (2005), Brunnermeier and Pedersen (2009), Cespa and Foucault (2014), Acharya and Thakor (2016). Different simulation models have been developed to capture adverse feedback loops, such as the RAMSI model built at the Bank of England (Burrows et al., 2012), the interaction model of the Austrian central bank (Puhr and Schmitz, 2014), the Bank of Canada Macro-Financial Risk Assessment Framework (Fique, 2017), or the agent based model of Bookstaber et al. (2018). Brunnermeier and Sannikov (2014), He and Krishnamurthy (2014) as well as Acharya et al. (2017) embed the financial sector in an equilibrium model and describe the behavioral response of financial institutions explicitly. But they assume all of them to be of the same type and therefore cannot capture the heterogeneity of the financial industry. Chen et al. (2013) put forward an axiomatic approach to measuring and allocating systemic risk, which was subsequently extended by Kromer et al. (2017). Methodological frameworks based on acceptable allocations have been proposed by Feinstein et al. (2017), Armenti et al. (2018) and Biagini et al. (2018). Our approach is different in that it relates externalities of the financial sector to the real economy and develops an allocation principle based on economic principles.

The rest of the paper is organized as follows: In Section 2 we define SystRisk, our measure of systemic risk of the financial sector. Section 3 introduces our systemic risk allocation principle. In Section 4 we propose a stochastic model for initial random shocks to the financial system and ensuing amplification mechanisms playing out within it. Section 5 discusses three different ways

for the regulator to manage systemic risk. Setting individual risk limits, imposing systemic risk charges and implementing a cap and trade scheme for systemic risk. All proofs are collected in the appendix.

2 Measuring total systemic risk

In this section we introduce our approach to measuring systemic risk in a general setting in which financial institutions cause externalities on society. A more specific model of financial losses and resulting externalities are discussed in Section 4 below.

We work with a finite set Ω of different states of the world. Then the space L of all mappings $X : \Omega \rightarrow \mathbb{R}$ can be identified with \mathbb{R}^d , where d is the number of elements in Ω . We write

- $X \geq X'$ if $X(\omega) \geq X'(\omega)$ for all $\omega \in \Omega$
- $X > X'$ if $X \geq X'$ and $X(\omega) > X'(\omega)$ for at least one $\omega \in \Omega$
- $X \gg X'$ if $X(\omega) > X'(\omega)$ for all $\omega \in \Omega$
- $L_+ := \{X \in L : X \geq 0\}$, $L_{++} := \{X \in L : X \gg 0\}$
- $\|X\|_\infty := \max_{\omega \in \Omega} |X(\omega)|$, $\|X\|_1 := \sum_{\omega \in \Omega} |X(\omega)|$
- $B_\infty^\varepsilon(X) := \{X' \in L : \|X - X'\|_\infty \leq \varepsilon\}$
- $\mathcal{P} :=$ the set of all probability measures on Ω
- $\mathcal{P}^f :=$ the set of all probability measures on Ω with full support.

2.1 The regulator

We assume there is a regulator overseeing a country's financial sector. It represents taxpayers, who in the event of a financial crisis, might have to prop up financial firms to prevent a collapse of the real economy. We fix a time period (e.g. a year or a quarter of a year) and model the country's real GDP generated over that period with a random variable $Y \in L$. We think of Y as the output of the real economy when it receives the financial services it needs to function properly. If essential parts of the financial system break down, the rest of the economy suffers severe consequences. To prevent this, the government is assumed to step in and provide the support necessary to keep the most important financial functions intact. This represents a negative externality caused by the financial sector. On the other hand, financial institutions can also have positive externalities if they perform better than expected.

We assume that the aggregate utility society is deriving from Y is given by $U(Y)$ for a preference functional $U : L \rightarrow \mathbb{R} \cup \{-\infty\}$. The effective domain of U ,

$$\text{dom } U := \{X \in L : U(X) \in \mathbb{R}\},$$

consists of all elements $X \in L$ for which $U(X)$ is real-valued. The interior of $\text{dom } U$ is given by

$$\text{int dom } U := \{X \in L : B_\infty^\varepsilon(X) \subseteq \text{dom } U \text{ for some } \varepsilon > 0\}.$$

The particular form of our systemic risk measure and allocation rule will depend on how the preference functional U is specified. But the approach works with any U satisfying the following two conditions. We assume them to hold throughout the paper.

(D) **Differentiability** $\text{int dom } U$ is non-empty and U is differentiable on $\text{int dom } U$

(S) **Strict monotonicity** $U(X) > U(X')$ for all $X, X' \in \text{dom } U$ such that $X > X'$.

If one identifies L with \mathbb{R}^d , $\text{int dom } U$ is an open subset of \mathbb{R}^d on which U is differentiable. We denote the gradient of U at X by $\nabla U(X)$.

A natural choice of U for our purposes is a CRRA expected utility

$$U(X) = \frac{1}{1-\gamma} \mathbb{E}_{\mathbb{P}} [X^{1-\gamma}] \quad (1)$$

for a relative risk aversions $\gamma > 1$ since it is the preference functional of a representative agent resulting from aggregating individual CRRA expected utility maximizers. If the probability measure \mathbb{P} has full support and (1) is understood to be $-\infty$ whenever $\mathbb{P}[X \leq 0] > 0$, then U is differentiable on $\text{int dom } U = \text{dom } U = L_{++}$. For $X \in L_{++}$, the components of the gradient $\nabla U(X)$ are given by $X(\omega)^{-\gamma} \mathbb{P}[\omega]$, $\omega \in \Omega$. So one can write

$$\nabla U(X) \cdot X' = \mathbb{E}_{\mathbb{P}} [X^{-\gamma} X'] \quad \text{for all } X' \in L,$$

where \cdot denotes the standard scalar product on \mathbb{R}^d . It is clear that (1) also satisfies condition (S) on L_{++} .

More generally, (D) and (S) are satisfied for any expected utility $U(X) = \mathbb{E}_{\mathbb{P}} [u(X)]$ as long as \mathbb{P} has full support and $u : \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty\}$ is a function with non-empty effective domain $\text{dom } u = \{x \in \mathbb{R} : u(x) \in \mathbb{R}\}$ that is strictly increasing and differentiable on the interior of $\text{dom } u$. For $X \in \text{int dom } U$, one has

$$\nabla U(X) \cdot X' = \mathbb{E}_{\mathbb{P}} [u'(X) X'] \quad \text{for all } X' \in L.$$

If in addition, $u : \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty\}$ is concave, then so is $U : L \rightarrow \mathbb{R} \cup \{-\infty\}$.

Going beyond expected utility, model uncertainty can be incorporated by introducing a non-empty closed subset $\mathcal{Q} \subseteq \mathcal{P}$ together with a lower semicontinuous mapping $c : \mathcal{Q} \rightarrow \mathbb{R}$ and defining

$$U(X) = \min_{\mathbb{Q} \in \mathcal{Q}} \{ \mathbb{E}_{\mathbb{Q}} [u(X)] + c(\mathbb{Q}) \} \quad (2)$$

for a function $u : \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty\}$. Preferences of this type, called variational preferences, were axiomatized by Maccheroni et al. (2006). If u is strictly increasing on $\text{int dom } u$ and \mathcal{Q} is contained in \mathcal{P}^f , U satisfies (S). Moreover, if $\text{dom } u$ is non-empty, u is differentiable on $\text{int dom } u$ and c is chosen appropriately, U also has the differentiability property (D). For example, it is well-known (see, e.g., Föllmer and Schied, 2004) that for $\mathcal{Q} = \mathcal{P}$, $\gamma > 0$, $\mathbb{P} \in \mathcal{P}^f$ and $c(\mathbb{Q}) = \frac{1}{\gamma} \mathbb{E} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}} \right]$, (2) becomes

$$U(X) = -\frac{1}{\gamma} \log \mathbb{E}_{\mathbb{P}} [\exp(-\gamma u(X))],$$

which satisfies (D) and (S) if u is strictly increasing and differentiable on $\text{int dom } u$.

2.2 The financial system

We suppose that the financial system consists of $I \in \mathbb{N}$ institutions, each of which causes an externality of the form

$$E_i = -\alpha_i V_i^- + \beta_i (V_i - v_i)^+,$$

where

- $V_i \in L$ is the net worth of institution i at the end of the measuring period, calculated e.g., as market value of assets minus book value of liabilities. V_i is assumed to already include non-linearities generated by negative feedback effects occurring during a crisis. A reduced form stochastic model for V_i is proposed in Section 4 below. V_i^- denotes the negative part $\max\{0, -V_i\}$ and $(V_i - v_i)^+$ the positive part $\max\{V_i - v_i, 0\}$.
- $\alpha_i \geq 0$ is a constant, and the term $\alpha_i V_i^-$ describes a potential cost to society if institution i 's net worth falls below zero. We assume α_i to be a number between 0 and 1. To avoid moral hazard, it is important that there be no bailout guarantee. But we assume that in case of a crisis, the government decides on a case by case basis which institutions have to be supported to prevent negative effects on the real economy. Companies could receive a mix of bailout payments and government loans at preferred conditions, or they could be nationalized. In all these cases some of the losses have to be borne by the taxpayer. We allow α_i to depend on the type of institution i since the cost caused by a bankruptcy depends on the structure of a financial company. For instance, a financial institution with several different business lines is more difficult to save or liquidate than one which concentrates only on a few activities. Furthermore, a typical bank relies heavily on short-term debt financing. So to protect the economy from an impending collapse, a relatively high fraction of its losses will have to be covered. Other components of the financial system, such as insurance companies or pension plans have more long-term liabilities and can continue operating even if their net worth falls below zero. As a consequence it might be possible to help them through a crisis with a minimal amount of funding.
- $\beta_i, v_i \geq 0$ are constants, and $\beta_i (V_i - v_i)^+$ models a possible positive externality. For instance, if institution i 's net worth exceeds the level v_i , the government may benefit from additional tax revenues of the form $\beta_i (V_i - v_i)^+$.

If the aggregate externality $E := \sum_{i \in I} E_i$ is absorbed by society, the real GDP changes from Y to $Y + E$. Our approach only needs the pair (Y, E) as input to determine the total systemic risk. To solve the systemic risk allocation problem we will have to specify the members of the financial system further. This will be done in Section 3 below.

2.3 Total systemic risk

We assume the regulator determines a tolerance level $e \in \mathbb{R}$ satisfying $U(Y + e) > -\infty$ and deems an externality E acceptable if $U(Y + E) \geq U(Y + e)$. For a given externality $E \in L$, we are asking how much money the regulator would have to receive to be compensated for absorbing E . This leads to the following definition of systemic risk.

Definition 2.1 We define *SystRisk* of $E \in L$ by

$$\rho(E) := \inf \{m \in \mathbb{R} : U(Y + E + m) \geq U(Y + e)\}. \quad (3)$$

Note that $\rho(E)$ is the cost of the externality E measured in currency units at the end of the measuring period. If used to determine systemic risk charges to be collected beforehand, it has to be discounted at the risk-free rate. For monitoring purposes, one can fix $e = 0$. But if *SystRisk* is used for regulation, the regulator can decide to set $e < 0$. Then negative externalities are tolerated as long as their impact is not too severe. Moreover, by adjusting the tolerance level e to the economic situation, the regulator can fine tune the stringency of financial regulation and implement countercyclical policies. For instance, in a recession financial regulation can be relaxed to permit the financial sector to operate with fewer constraints.

Under our assumptions on U , ρ has the following properties:

Lemma 2.2 ρ is a function from L to \mathbb{R} satisfying

(N) Normalization at the tolerance level: $\rho(e) = 0$

(M) Monotonicity: $\rho(E) \geq \rho(E')$ for all $E, E' \in L$ such that $E \leq E'$

(T) Translation property: $\rho(E + m) = \rho(E) - m$ for all $E \in L$ and $m \in \mathbb{R}$

(L) Lipschitz-continuity: $|\rho(E) - \rho(E')| \leq \|E - E'\|_\infty$ for all $E, E' \in L$.

In particular, ρ is a monetary risk measure on L in the sense of Föllmer and Schied (2004). But since it is of the special form (3), it has additional properties. The next result shows that ρ is differentiable and strictly decreasing at all $E \in L$ satisfying the condition

(I) $Y + E + \rho(E) \in \text{int dom } U$.

This will be important for the risk allocation method proposed in Section 3. Note that if E satisfies **(I)**, the gradient $\nabla U(Y + E + \rho(E))$ exists, and since U has the strict monotonicity property **(S)**, all its components are strictly positive. So

$$\mathbb{Q}^E := \frac{\nabla U(Y + E + \rho(E))}{\|\nabla U(Y + E + \rho(E))\|_1}$$

defines a probability measure with full support.

Theorem 2.3 Assume $E \in L$ satisfies condition **(I)**. Then ρ is differentiable at E with gradient

$$\nabla \rho(E) = -\frac{\nabla U(Y + E + \rho(E))}{\|\nabla U(Y + E + \rho(E))\|_1}. \quad (4)$$

In particular, all components of $\nabla \rho(E)$ are strictly negative,

$$\lim_{\varepsilon \rightarrow 0} \frac{\rho(E + \varepsilon E') - \rho(E)}{\varepsilon} = \mathbb{E}_{\mathbb{Q}^E}[-E'] \quad \text{for all } E' \in L, \quad (5)$$

and

$$\rho(E + E') < \rho(E) < \rho(E - E') \quad \text{for all } E' \in L \text{ satisfying } E' > 0. \quad (6)$$

\mathbb{Q}^E can be viewed as a shadow pricing measure for systemic risk, as $\mathbb{E}_{\mathbb{Q}^E}[-E']$ gives the marginal change in systemic risk stemming from a small deviation E' from the aggregate externality E .

It follows from Theorem 2.3 that in case of a CRRA expected utility

$$U(X) = \frac{1}{1-\gamma} \mathbb{E}_{\mathbb{P}} [X^{1-\gamma}] \quad \text{for a risk aversion parameter } \gamma > 1, \quad (7)$$

ρ is differentiable at every $E \in L$ satisfying (I) with gradient $\nabla \rho(E) = -\mathbb{Q}^E$, where the shadow pricing measure \mathbb{Q}^E is given by

$$\frac{d\mathbb{Q}^E}{d\mathbb{P}} = \frac{(Y + E + \rho(E))^{-\gamma}}{\mathbb{E}_{\mathbb{P}}[(Y + E + \rho(E))^{-\gamma}]}.$$

Note also that (7) is concave on L and strictly concave on $\text{int dom } U = L_{++}$; that is, for all $\lambda \in (0, 1)$, one has

$$U(\lambda X + (1-\lambda)X') \geq \lambda U(X) + (1-\lambda)U(X') \quad \text{for } X, X' \in L,$$

and the inequality is strict if $X, X' \in L_{++}$ such that $X \neq X'$. This implies that ρ is convex. More precisely, the following holds:

Proposition 2.4 *If $U : L \rightarrow \mathbb{R} \cup \{-\infty\}$ is concave, then ρ is convex and can be represented as*

$$\rho(E) = \max_{\mathbb{Q} \in \mathcal{P}} \left\{ \mathbb{E}_{\mathbb{Q}}[-E] - \rho^{\#}(\mathbb{Q}) \right\} \quad \text{for } \rho^{\#}(\mathbb{Q}) := \sup_{E \in L} \{ \mathbb{E}_{\mathbb{Q}}[-E] - \rho(E) \}. \quad (8)$$

Moreover, for all $E \in L$ satisfying (I), \mathbb{Q}^E is the unique maximizer in (8).

If furthermore, U is strictly concave on $\text{int dom } U$, then

$$\rho(\lambda E + (1-\lambda)E') < \lambda \rho(E) + (1-\lambda)\rho(E')$$

for all $0 < \lambda < 1$ and $E, E' \in L$ satisfying (I) such that $E - E'$ is non-deterministic.

(8) expresses ρ as a maximum of shifted expectations with respect to different probability measures. This is called robust representation in the risk measure literature; see e.g., Föllmer and Schied (2004) or Cheridito and Li (2008, 2009). Every real-valued convex function on a finite-dimensional vector space has a representation as a maximum of affine functions. That the maximum in (8) can be taken over the set of probability measures \mathcal{P} is due to the fact that ρ has the monotonicity property (M) and the translation property (T). That \mathbb{Q}^E is the unique maximizer in (8) follows from the differentiability assumption (D). Details are given in the appendix.

As a consequence of Proposition 2.4, one obtains that in case U is concave and strictly concave on $\text{int dom } U$, the systemic risk measure ρ scales superlinearly:

Corollary 2.5 *Assume U is concave on L and strictly concave on $\text{int dom } U$. Let $E \in L$ be a non-deterministic externality satisfying (I). Then*

$$\rho(\lambda E) > \lambda(\rho(E) - e) + e \quad \text{for all } \lambda > 1.$$

In particular, in the case $e \leq 0$, one has

$$\rho(\lambda E) > \lambda \rho(E) \quad \text{for } \lambda > 1. \quad (9)$$

Relation (9) means that if financial institutions decide to change their positions such that their aggregate externality on the rest of society increases from E to λE for a constant $\lambda > 1$, then SystRisk increases by more than λ . This stems from the fact that externalities are compared to the size of the real GDP Y and a strictly concave preference functional U corresponds to a risk averse representative taxpayer.

3 Systemic risk allocation

It is important to know how much each component of the financial system contributes to overall systemic risk. We propose an allocation rule based on externalities. They can result from macroeconomic and idiosyncratic shocks as well as feedback effects between financial institutions (a detailed model is given in Section 4 below). In particular, the externalities of one firm can be affected by the behavior of others. But it is each institution's own responsibility to manage its exposure to spillover risk.

A risk allocation is simply a vector $k = (k_1, \dots, k_I) \in \mathbb{R}^I$ whose components specify how much of the total risk is allocated to the i -th institution. If the allocation is used to implement capital requirements or systemic risk charges, it should satisfy

(FA) Full allocation: $\sum_{i=1}^I k_i = \rho(E)$.

But for monitoring purposes alone, (FA) is not necessary. For instance, the with-without allocation

$$WW_i = \rho(E) - \rho(E - E_i) \quad (10)$$

(see e.g. Merton and Perold, 1993, or Matten, 1996) does not satisfy (FA). But on the other hand, it has the following two properties:

(RA) Riskless allocation: $k_i = -E_i$ if E_i is deterministic

(CR) Causal responsibility: If the i -th externality changes from E_i to $E'_i = E_i + \Delta E_i$ for some $\Delta E_i \in L$ and $E'_j = E_j$ for $j \neq i$, then the adjusted allocation k' satisfies $k'_i - k_i = \rho(E + \Delta E_i) - \rho(E)$.

(RA) means that if it is clear in advance that firm i will cause an externality costing m units of currency, then its systemic risk allocation should be m . If an allocation principle satisfies (CR) and one of the firms decides to change its exposure, it has to bear the full cost of the resulting change in total systemic risk. Note that under the allocation rule (10), a change in one firm's externality also affects the others. Their allocations change from $k_j = \rho(E) - \rho(E - E_j)$ to

$$k'_j = \rho(E + \Delta E_i) - \rho(E + \Delta E_i - E_j).$$

But if ΔE_i is small compared to E , then k'_j is close to k_j . WW_i can be approximated with the marginal contributions

$$MC_i := \lim_{\varepsilon \rightarrow 0} \frac{\rho(E + \varepsilon E_i) - \rho(E)}{\varepsilon} = \mathbb{E}_{\mathbb{Q}^E}[-E_i]. \quad (11)$$

If E satisfies condition (I), it follows from Theorem 2.3 that the marginal contributions exist. They still satisfy (RA), and (CR) holds approximately. Moreover, they have the following property:

(AD) Additivity: If firms i and j are merged in such a way that the externality of the combined unit becomes $E_{i+j} = E_i + E_j$ and the externalities of the other firms stay the same, then the new allocation k' satisfies $k'_{i+j} = k_i + k_j$ as well as $k'_l = k_l$ for $l \neq i, j$.

The marginal contributions (11) still do not satisfy (FA). But the following proposition shows that if U is concave and $e \leq 0$, they sum up to at least the total systemic risk $\rho(E)$.

Proposition 3.1 *Assume ρ is convex, $E \in L$ satisfies (I) and $e \leq 0$. Then $\sum_{i=1}^I MC_i \geq \rho(E)$.*

So if the conditions of Proposition 3.1 hold, (FA) can be achieved by reducing the marginal contributions MC_i . To keep the correct marginal incentives, we shift them by exogenous size parameters $s_i \in \mathbb{R}_+$.

Definition 3.2 *Assume $E \in L$ satisfies condition (I) and $\sum_{i=1}^I s_i \neq 0$. Then the size-shifted marginal contributions are given by*

$$SMC_i := MC_i - \mu s_i, \tag{12}$$

where $\mu \in \mathbb{R}$ is chosen so that the full allocation principle (FA) holds.

The parameters s_i are meant to reflect the institutions' sizes and should not be easy to manipulate with the intent to avoid regulation. For instance, they can be chosen as the amount of corporate taxes paid last year. Then, if the assumptions of Proposition 3.1 hold, (12) reduces the marginal contribution of each firm by an amount that is proportional to the taxes it payed in the previous year. The size-shifted marginal contributions satisfy (FA) by construction. Technically, they violate (RA). But if a financial institution is liquidated immediately in case it has a deterministic negative externality and exempted from regulation if the externality is deterministic positive, the remaining externalities are non-deterministic, and (RA) holds trivially. Moreover, the size-shifted marginal contributions (12) satisfy (CR) approximately, and if the size parameters s_i add up in mergers, they inherit the additivity property (AD) from the marginal contributions (11). That is, the allocation rule (12) is additive in the externalities. However, while it is natural to assume that the size parameters of merging companies add up, their externalities can behave in a number of different ways. The precise behavior of externalities under mergers depends on the relation between the future net worths V_i . This issue is discussed in more detail in Subsection 4.4 below.

Note that the 2010 Dodd–Frank Act in the United States also acknowledges that some financial institutions are more systemically important than others. For this reason the Financial Stability Oversight Council was established, one of whose main tasks is to designate certain financial institutions as systemically important (SIFIs). The current designation procedure looks at a variety of variables, including size, leverage and stress-test performance. Our systemic risk allocation goes beyond a zero-one designation, as it also captures the degree to which an institution contributes to overall systemic risk.

4 Modeling losses in the financial sector

In this section we propose a reduced form model for losses in the financial sector and resulting externalities. The suggested model has the advantage that its coefficients can be estimated from

historical data. But our approach also works with a different underlying model, like for instance, a simulation model such as the RAMSI model of Burrows et al. (2012), the interaction model of of Pühr and Schmitz (2014), the Macro-Financial Risk Assessment Framework of Fique (2017), or the agent based model of Bookstaber et al. (2018).

We assume that losses in the financial system are the result of initial random shocks and negative feedback effects in the system.

4.1 Initial losses

Initial shocks can be caused by macroeconomic factors or firm specific events. We suppose that at the beginning of the measuring period every institution of the financial system starts with a deterministic net worth $x_i \in \mathbb{R}$. Then all of them receive a random shock of the form $z_i(F, W_i)$, where $F = (F_1, \dots, F_n) \in L^n$ is a vector of macroeconomic variables, $W_i \in L$ is a firm-specific shock³ and $z_i : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ a deterministic function. After the initial shocks the institutions' net worths are

$$X_i = x_i + z_i(F, W_i).$$

4.2 Feedback mechanisms

In a second step, spillover effects between financial institutions take place, potentially amplifying the crisis. We divide them into direct and indirect spillover effects.

- Direct spillovers are caused by contractual connections between firms.
- Indirect spillover effects that played an important role in the subprime mortgage crisis starting in 2007 were asset fire sales and the dry up of funding liquidity. Asset fire sales occur when firms sell illiquid assets. This creates downward pressure on their prices and affects other institutions that are holding them. Funding liquidity dries up when, due to distress in some parts of the financial sector, lenders become more risk averse and suddenly demand higher interest. This is especially problematic for firms with a lot of short term debt since they have to refinance more frequently.

To capture vulnerabilities to direct spillover effects, we introduce an $I \times I$ -matrix $c = (c_{ij})_{i \neq j}$ of interconnectedness (it does not need diagonal elements, or they can be set equal to zero). Component c_{ij} of the interconnectedness matrix c gives a measure of future cash flows to be paid by institution i to institution j . If the former is in distress, it might not be able to honor its commitments, creating a problem for the latter. Indirect spillover effects depend on liquidity mismatches. Denote by $l = (l_1, \dots, l_I)$ the vector of liquidity mismatch indexes of the firms $i \in I$. They measure the firms' debt structure compared to the market liquidity of their assets. If the majority of an institution's debt is long term and all its assets can easily be sold in a crisis, it will not be affected by indirect spillover effects. Consequently, its liquidity mismatch index is low. On the other hand, if it has a lot of short term debt and many illiquid assets, it is more prone to indirect spillovers and has a high liquidity mismatch index; see Brunnermeier et al. (2014).

³The shocks $F_1, \dots, F_n, W_1, \dots, W_I$ can be modeled to be dependent. In particular, firm-specific shocks can be correlated with the macroeconomic variables as well as other firm-specific shocks.

We describe negative feedback effects by assuming that after the arrival of the initial shocks, amplification effects move institution i 's net worth from X_i to

$$V_i = a_i(F, X, c, l),$$

where $a_i : \mathbb{R}^{n+I+I^2} \rightarrow \mathbb{R}$ is a deterministic function (F has n components, X and l each have I components, and the matrix c has $I^2 - I$ non-trivial entries).

The global financial crisis was triggered by the collapse of the US subprime mortgage market in 2007. This market was small relative to the overall US mortgage market and tiny compared to the size of the global financial markets. The reason that problems in the US subprime mortgage market could trigger a global crisis of a scale second only to the great depression of the 1930s, was the vulnerability of the financial system to spillover and amplification effects. These effects were caused by liquidity mismatch and leverage. A significant fraction of the liquidity mismatch was disguised in the shadow banking system; see e.g., Pozsar et al. (2012) and Bengtsson (2016). Our reduced form model is designed to capture direct contagion, amplification mechanisms as well as adverse feedback effects in the shadow banking system.

4.3 The regulator

We assume that the GDP of the real economy is of the form $Y = y(F)$, where the components of $F = (F_1, \dots, F_n)$ are the macroeconomic variables introduced in Subsection 4.1 and $y : \mathbb{R}^n \rightarrow \mathbb{R}$ is a deterministic function. Suppose that E satisfies condition (I). Then we obtain from Theorem 2.3 that the gradient of ρ at E acts on a random variable $E' \in L$ like

$$\nabla \rho(E) \cdot E' = \mathbb{E}_{\mathbb{Q}^E}[-E'],$$

where

$$\mathbb{Q}^E = \frac{\nabla U(Y + E + \rho(E))}{\|\nabla U(Y + E + \rho(E))\|_1}.$$

For instance, if U is a CRRA expected utility $\mathbb{E}_{\mathbb{P}}[u(x)]$ for a probability measure $\mathbb{P} \in \mathcal{P}^f$ and a function u of the form

$$u(x) = \begin{cases} \frac{1}{1-\gamma} x^{1-\gamma} & \text{if } x > 0 \\ -\infty & \text{if } x \leq 0 \end{cases} \quad \text{for some } \gamma > 1,$$

then U is strictly concave on $\text{int dom } U$, and

$$\frac{d\mathbb{Q}^E}{d\mathbb{P}} = \frac{(Y + E + \rho(E))^{-\gamma}}{\mathbb{E}_{\mathbb{P}}[(Y + E + \rho(E))^{-\gamma}]}.$$

So it follows from Corollary 2.5 that for $e \leq 0$, one has

$$\rho(\lambda E) > \lambda \rho(E) \quad \text{for } \lambda > 1.$$

This means that if the financial sector decides to behave in way such that the aggregate externality multiplies by a factor $\lambda > 1$, then total systemic risk increases more than proportionally. The size-shifted marginal contributions of Definition 3.2 take the form

$$SMC_i = \mathbb{E}_{\mathbb{Q}^E}[-E_i] - \mu s_i,$$

where $\mu \in \mathbb{R}_+$ has to be chosen so that the full allocation principle (FA) holds.

4.4 Mergers and spinoffs

We now discuss how systemic risk is affected by mergers and spinoffs. We consider m firms i_1, \dots, i_m in I and compare the case where they are separate entities to the one where they are merged into one big company. We assume that a merger does not affect the initial shocks $z_i(F, W_i)$ and the size parameters s_i add up. Then, without restructuring, the net worth of the combined unit after the occurrence of initial shocks is $\hat{X} = \sum_{j=1}^m X_{i_j}$ and its size parameter $\hat{s} = \sum_{j=1}^m s_{i_j}$. But after that there are two layers of non-linearities.

1. Amplification mechanisms are non-linear. So \hat{V} is not necessarily equal to the sum $\sum_{j=1}^m V_{i_j}$.
2. Externalities are non-linear in final net worths. So even if \hat{V} equals $\sum_{j=1}^m V_{i_j}$, \hat{E} is in general different from $\sum_{j=1}^m E_{i_j}$.

We recall that firm i 's externality is

$$E_i = -\alpha_i V_i^- + \beta_i (V_i - v_i)^+$$

for parameters $\alpha_i, \beta_i, v_i \in \mathbb{R}_+$, whereas the merged firm causes an externality of the form

$$\hat{E} = -\hat{\alpha} \hat{V}^- + \hat{\beta} (\hat{V} - \hat{v})^+$$

for parameters $\hat{\alpha}, \hat{\beta}, \hat{v} \in \mathbb{R}_+$.

Depending on the circumstances, mergers and spinoffs can have varying effects on total systemic risk. In the following we discuss three different scenarios:

4.4.1 Clone property

As a benchmark, we first consider the case where the firms i_1, \dots, i_m are clones in the sense that they start with the same initial net worths x_i , are of the same size s_i , have identical exposures and react in the same way to shocks. In particular, they receive the same initial shocks and respond identically. This leads to equal final net worths V_i . Moreover, since they all experience identical gains and losses in the same scenarios, negative feedback effects should not depend on whether the firms operate independently or as one combined company. So, in this special case, final net worths and externalities both add up; that is, $\hat{V} = \sum_{j=1}^m V_{i_j} = mV_{i_1}$ and $\hat{E} = \sum_{j=1}^m E_{i_j} = mE_{i_1}$. In particular, the total systemic risk $\rho(E)$ does not change, and since the size-shifted marginal contributions (12) have the additivity property (AD), SystRisk and our allocation principle satisfy the following

Clone property:

Under a merger of the clones i_1, \dots, i_m , the total systemic risk $\rho(E)$ as well as the size-shifted marginal contributions of the firms different from i_1, \dots, i_m stay the same, while the size-shifted marginal contribution of the new firm is $mSMC_{i_1}$.

4.4.2 Reducing systemic risk through spinoffs

From the perspective of systemic risk, the goal of spinoffs should be to separate large financial institutions into different business units so as to prevent the spread of problems and facilitate liquidations in case of insolvency. This will not reduce losses in all possible scenarios. For instance, in case $V_{i_1} < 0$ and $V_{i_2}, \dots, V_{i_m} \geq 0$, the first of the small firms goes bankrupt if they are separate entities while it could have been saved by the others, had they been part of the same company. That is, large institutions are more diversified and can compensate small losses in one line of business with gains from others. However, if done properly, spinoffs have two benefits: (i) They introduce fire walls and reduce negative feedback effect resulting in less severe aggregate losses during crises. So with non-linear amplification effects, large values of $\sum_{j=1}^m V_{i_j}^-$ are less likely than large values of \hat{V}^- . (ii) Small firms with few business lines are easier to support or wind down in times of turmoil than multinational financial conglomerates. This must be reflected by setting the cost coefficient $\hat{\alpha}$ for conglomerates higher than those for less complex institutions $\alpha_{i_1}, \dots, \alpha_{i_k}$ (by choosing, for instance, α_i as an increasing function of the number of business lines of firm i). All in all, if a complex financial institution is split into simpler parts, the likelihood that $\sum_{j=1}^m E_{j_i}$ takes large negative values decreases compared to \hat{E} . This reduces total systemic risk as well as the contribution of the firms i_1, \dots, i_m to the whole.

4.4.3 Cases where spinoffs increase systemic risk

If a large financial institution is divided into smaller parts in a way that does not reduce negative amplification effects or complexity, total systemic risk can go up. If for instance, $\hat{V} = \sum_{j=1}^m V_{i_j}$, $\hat{\alpha} = \alpha_{i_1} = \dots = \alpha_{i_m} = \alpha$, $\hat{\beta} = \beta_{i_1} = \dots = \beta_{i_m} = \beta$ and $\hat{v} = v_{i_1} = \dots = v_{i_m} = 0$ for numbers $\alpha \geq \beta \geq 0$, one has

$$\hat{E} = \varphi(\hat{V}) \quad \text{and} \quad E_{i_j} = \varphi(V_{i_j}) \quad \text{for all } j = 1, \dots, m,$$

for the super-additive function $\varphi(x) = -\alpha x^- + \beta x^+$, and it follows that $\hat{E} \geq \sum_{j=1}^m E_{i_j}$. That is, spinoffs have negative effects on externalities and increase total systemic risk. For example, if $\alpha > 0$ and $\beta = 0$, only negative externalities are taken into account. So if the combined firm does not cause more negative feedback effects and is not more complex than the sum of the small ones, it is safer because it can use profits of one unit to absorb losses in another one.

5 Managing systemic risk: a discussion

There are different ways of managing systemic risk. Generally, the regulator regards the system as safe if $\rho(E) \leq 0$. But if $\rho(E) > 0$, it poses an unacceptable risk that has to be mitigated. $-\nabla\rho(E)$ points in the direction in which total systemic risk is reduced most effectively.

Financial institutions have their own aims and will respond to new regulation. But the financial system is heterogenous. A bank functions differently from an insurance company, pension plan, sovereign wealth fund or hedge fund. We do not model their behavior. This has the advantage that our approach does not depend on behavioral assumptions. But if new regulation is implemented, responses in the financial sector and their ramifications for the real economy have to be monitored.

The stricter the regulation, the tighter the constraints under which the financial sector has to operate. Our systemic risk measure and allocation principle show where the risk is coming from. By adjusting the tolerance parameter, the regulator can decide how much overall systemic risk to accept depending on current economic conditions.

5.1 Systemic risk limits

One way to manage systemic risk is to force each bank's systemic risk contribution towards zero. The regulator can guarantee $\rho(E) \leq 0$ by requiring that all size-shifted marginal contributions satisfy $SMC_i \leq 0$, $i = 1, \dots, I$. If $SMC_i > 0$ for some i , then firm i most efficiently changes its externality in the direction $-\nabla\rho(E)$. Alternatively, one can determine the systemic capital surcharge for member i as the minimal amount of additional capital that makes SMC_i fall to zero. This amounts to setting systemic risk limits and is close to current regulation in most modern economies. But standard financial regulation views financial firms as separate entities and ignores their systemic interactions.

5.2 Systemic risk charges

Instead of setting risk limits, the regulator can let financial firms act freely and charge them a premium for their contribution to systemic risk. If collected at the beginning of the measuring period, they have to be in the amount

$$\tau_i = \frac{SMC_i^+}{1+r}, \quad (13)$$

where r is the risk-free interest rate, so that total systemic risk is compensated.

Compared to risk limits, systemic risk charges allow for the pooling of capital buffers. In case of a crisis they can be used to support failing components of the system. Moreover, if the abatement costs are only known to the financial institutions but not the regulator, imposing systemic risk charges has the advantage that firms can decide themselves how much risk is optimal for them to take on. Firms with low abatement costs will reduce their risk exposures to lower their charge. Others find it optimal to keep their positions and pay a higher premium. Systemic risk charges act like Pigouvian taxes. In contrast to systemic risk limits, they do not allow control of the aggregate risk in the financial system, but they require financial firms to pay compensation for it. To decide if it is preferable to regulate with systemic risk limits, charges or a mixture of the two is related to the classical question whether it is better to control an economic system through quantities or prices; see Weitzman (1974).

5.3 Cap and trade

Similarly to cap and trade schemes for the emission of pollutants, the regulator can distribute (for free or through an auction) a number $P \in \mathbb{R}_+$ of systemic risk permits and require that each member of the system holds them in an amount P_i such that $SMC_i - P_i \leq 0$. A possible implementation is as follows. At the beginning of each measuring period (e.g. quarterly) permits are auctioned. Then they can be traded for a few days. A cap and trade system allows the regulator to control the overall level of systemic risk but lets the market decide how systemic

risk is reduced most efficiently. Those financial institutions with the lowest abatement costs can reduce their contributions and sell off extra permits to others for which a reduction of systemic risk is more expensive. In addition, a cap and trade system can provide valuable information. If permits trade at prices that correctly reflect the underlying fundamentals, they reveal the system's marginal abatement costs, which is useful information for the implementation of socially optimal regulation⁴.

6 Conclusions

This paper proposes a practical way of measuring and allocating systemic risk. It views the possibility that financial institutions might have to be supported in a crisis to protect the economy from negative consequences as an externality on society. SystRisk, our measure of total systemic risk, quantifies the a priori cost of this externality. It relates the financial industry to the real economy. Therefore, costs of externalities grow faster than linearly if they become large compared to the output of the real economy. Our approach is based on reduced form model that is designed to capture direct spillovers caused by defaults on claims between financial institutions as well as indirect feedback effects such as informational contagion, fire sales and liquidity spirals. Since we take into account that the financial sector might receive government support, SystRisk can detect risks that might not be fully reflected in market prices of traded instruments such as equity shares, bonds, options or credit default swaps. It also addresses the volatility paradox, which refers to the phenomenon that financial systems tend to be more vulnerable to systemic feedback effects if volatility is low. This is related to the procyclicality problem. Standard risk measures typically do not pick up systemic risk in economic booms but spike up sharply when a crisis erupts. If this is directly translated into tighter regulation, financial institutions are forced to decrease their leverage at the beginning of economic downturns, which exacerbates the problem. By incorporating leverage and liquidity mismatch, which make a financial system vulnerable to systemic feedback effects, our approach is able to identify systemic risk in low volatility environments. Moreover, it includes a tolerance parameter that can be adjusted over time to implement countercyclical regulation.

Our systemic risk allocation principle distributes total systemic risk among individual firms according to their size-shifted marginal contributions. This ensures that financial institutions are allocated systemic risk according to the role they play in the financial system. A financial firm that is part of a herd or exposed to spillover effects is assigned more systemic risk than one that is more independent and better prepared to withstand a financial crisis. Moreover, our method satisfies the so called clone property, that is, if a financial institution is split into n equal clones, the total systemic risk does not change, and the institution's risk contribution splits into n equal parts.

The paper outlines three different ways how our method of measuring and allocating systemic risk can be translated into financial regulation: setting systemic risk limits, imposing systemic risk charges, and a cap and trade system for systemic risk. Implementation of financial regulation

⁴See also Stein (2012), who compares reserve requirements to a cap and trade system in which the central bank governs the overall supply of reserves, while the fed funds market determines the allocation of reserves within the banking system.

that is fully optimal for the whole society would require an equilibrium model that can predict the behavioral response of financial institutions to regulatory requirements and the consequences for the real economy. This paper takes a reduced form approach in that it focuses on quantifying the cost of financial crises to society. Our method detects where systemic risk is coming from. Generally, stricter regulation means that the financial sector is under tighter constraints and therefore, can offer fewer services. By adjusting the tolerance parameter, the regulator can find the right level of stringency for an optimal balance between financial stability and enough flexibility for the financial sector to operate efficiently. This has to be achieved in an iterative process in which the regulator fine tunes the tolerance parameter and monitors the response from the financial sector and the effects on the rest of the economy.

A Proofs

Proof of Lemma 2.2

It is easy to check that if U satisfies (S) and $U(Y + e) > -\infty$, ρ is a function from L to \mathbb{R} with the properties (N), (M) and (T). So for $E, E' \in L$, one has

$$\rho(E) - \rho(E') \leq \rho(E) - \rho(E + \|E - E'\|_\infty) = \|E - E'\|_\infty,$$

and by symmetry,

$$\rho(E') - \rho(E) \leq \|E - E'\|_\infty,$$

which shows that ρ satisfies (L). □

Proof of Theorem 2.3

First note that it follows from condition (I) that $U(Y + E + \rho(E)) = U(Y + e) \in \mathbb{R}$. Now let E_n , $n \in \mathbb{N}$, be a sequence in L converging to 0. Since ρ satisfies the Lipschitz condition (L), $Y + E + E_n + \rho(E + E_n)$ is in $\text{int dom } U$ for n large enough, and therefore,

$$\begin{aligned} U(Y + E + \rho(E)) &= U(Y + e) = U(Y + E + E_n + \rho(E + E_n)) \\ &= U(Y + E + \rho(E)) + \nabla U(Y + E + \rho(E)) \cdot (E_n + \rho(E + E_n) - \rho(E)) + o(\|E_n\|_\infty) \end{aligned}$$

as $n \rightarrow \infty$. It follows that

$$\rho(E + E_n) - \rho(E) = -\frac{\nabla U(Y + E + \rho(E))}{\|U(Y + E + \rho(E))\|_1} \cdot E_n + o(\|E_n\|_\infty) \quad \text{for } n \rightarrow \infty,$$

which shows that ρ is differentiable at E with gradient (4). In particular, one has (5), and all components of $\nabla \rho$ are negative. The latter implies the strict monotonicity (6) around E . □

Proof of Proposition 2.4

Let $E, E' \in L$ and $0 \leq \lambda \leq 1$. Choose $m, m' \in \mathbb{R}$ such that $\rho(Y + E + m) \geq U(Y + e)$ and $\rho(Y + E' + m') \geq U(Y + e)$. Then $U(Y + \lambda E + (1 - \lambda)E' + \lambda m + (1 - \lambda)m') \geq U(Y + e)$. So $\rho(\lambda E + (1 - \lambda)E') \leq \lambda m + (1 - \lambda)m'$ and therefore, $\rho(\lambda E + (1 - \lambda)E') \leq \lambda \rho(E) + (1 - \lambda)\rho(E')$. This shows that ρ is convex.

Now it follows from standard convex duality arguments that

$$\rho(E) = \max_{Z \in L} \{E \cdot Z - \rho^*(Z)\} \quad \text{for} \quad \rho^*(Z) := \sup_{E \in L} \{E \cdot Z - \rho(E)\}. \quad (\text{A.1})$$

Since by Lemma 2.2, ρ satisfies (M) and (T), one obtains $\rho^*(Z) = \infty$ for $-Z \notin \mathcal{P}$. So (A.1) can be written in the form (8) for $\mathbb{Q} = -X$.

By Theorem 2.3, ρ is differentiable at every $E \in L$ satisfying (I) with gradient $\nabla\rho(E) = -\mathbb{Q}^E$. Since ρ is convex, one has $\rho(E') - \rho(E) \geq \mathbb{E}_{\mathbb{Q}^E}[E] - \mathbb{E}_{\mathbb{Q}^E}[E']$ for all $E' \in L$. It follows that \mathbb{Q}^E is a maximizer in the representation (8), and it must be the unique maximizer because otherwise, ρ would not be differentiable at E .

If U is strictly concave on $\text{int dom } U$, choose E, E' satisfying (I) such that $E - E'$ is non-deterministic. Then $Y + E + \rho(E)$ and $Y + E' + \rho(E')$ are both in $\text{int dom } U$, and one has

$$U(Y + E + \rho(E)) = U(Y + E' + \rho(E')) = U(Y + e).$$

Moreover, $Y + E + \rho(E) \neq Y + E' + \rho(E')$. So one obtains from the strict concavity of U on $\text{int dom } U$ that

$$U(Y + \lambda E + (1 - \lambda)E' + \lambda\rho(E) + (1 - \lambda)\rho(E')) > U(Y + e) \quad \text{for } 0 < \lambda < 1. \quad (\text{A.2})$$

Since $\text{int dom } U$ is convex, one obtains

$$Y + \lambda E + (1 - \lambda)E' + \lambda\rho(E) + (1 - \lambda)\rho(E') \in \text{int dom } U,$$

which together with (A.2), implies $\rho(\lambda E + (1 - \lambda)E') < \lambda\rho(E) + (1 - \lambda)\rho(E')$. \square

Proof of Corollary 2.5

Assume U is concave and strictly concave on $\text{int dom } U$. Let $E \in L$ be a non-deterministic externality satisfying (I). Choose $\lambda > 1$ and set $E' = \lambda E$. Since U is concave, one obtains from Proposition 2.4 that

$$\rho(\mu E') \leq \mu\rho(E') + (1 - \mu)\rho(0) = \mu\rho(E') + (1 - \mu)e \quad \text{for all } 0 \leq \mu \leq 1.$$

Now let us assume the inequality is an equality for $\mu = 1/\lambda$. Then it has to be an equality for all $0 \leq \mu \leq 1$. Choose μ_1, μ_2 such that $0 < \mu_1 < 1 < \mu_2 < \lambda$ and set $F = \mu_1 E$, $G = \mu_2 E$. Then $F - G$ is non-deterministic and

$$\rho(\nu F + (1 - \nu)G) = \nu\rho(F) + (1 - \nu)\rho(G) \quad \text{for all } 0 \leq \nu \leq 1. \quad (\text{A.3})$$

However, by Lemma 2.2, ρ is Lipschitz-continuous, So if μ_1 and μ_2 are chosen sufficiently close to 1, then F and G satisfy condition (I). But then (A.3) contradicts the last part of Proposition 2.4. So one must have

$$\rho(E'/\lambda) < \rho(E')/\lambda + (1 - 1/\lambda)e,$$

which is equivalent to $\rho(\lambda E) > \lambda\rho(E) - (\lambda - 1)e$. \square

Proof of Proposition 3.1

If ρ is convex and $E \in L$ satisfies (I), it follows from the proof of Proposition 2.4 that

$$\rho(E) = \mathbb{E}_{\mathbb{Q}^E}[-E] - \rho^\#(\mathbb{Q}^E).$$

Since $\rho^\#(\mathbb{Q}^E) \geq \mathbb{E}_{\mathbb{Q}}[-e] - \rho(e) \geq -e \geq 0$, one has

$$\rho(E) \leq \mathbb{E}_{\mathbb{Q}^E}[-E] = \sum_{i=1}^I \mathbb{E}_{\mathbb{Q}^E}[-E_i] = \sum_{i=1}^I MC_i.$$

\square

References

- [1] V.V. Acharya, R.F. Engle and M.P. Richardson. Capital shortfall: a new approach to ranking and regulating systemic risks. *American Economic Review: Papers & Proceedings* 2012, 102(3), 59–64.
- [2] V.V. Acharya, L.H. Pedersen, T. Philippon and M.P. Richardson. Measuring systemic risk. *Rev. Fin. Studies* 2017, 30(1), 2–47.
- [3] V.V. Acharya and A.V. Thakor. The dark side of liquidity creation: Leverage and systemic risk. *J. Fin. Intermediation* 2016, 28, 4–21.
- [4] T. Adrian and M.K. Brunnermeier. CoVaR. *Am. Econ. Rev.* 2016, 106(7), 1705–1741.
- [5] P. Aghion, P. Bolton and M. Dewatripont. Contagious bank failures in a free banking system. *European Economic Review* 2000, 44, 713–718.
- [6] L. Allen, T.G. Bali and Y. Tang. Does systemic risk in the financial sector predict future economic downturns? *Rev. Fin. Studies* 2012, 25(10), 3000–3036.
- [7] F. Allen and D. Gale. Financial contagion. *Journal of Political Economy* 2000, 108, 1–33.
- [8] Y. Armenti, S. Crépey, S. Drapeau and A. Papapantoleon. Multivariate shortfall risk allocation and systemic risk. *SIAM J. Fin. Math.* 2018, 9(1), 90–126.
- [9] R. Avesani. FIRST: a market-based approach to evaluate financial system risk and stability. *IMF Working Paper* 232, 2005.
- [10] R. Avesani, A. Garcia Pascual and J. Li A new risk indicator and stress testing tool: a multi-factor nth-to-default CDS basket. *IMF Working Paper* 105, 2006.
- [11] J. Bai, A. Krishnamurthy and C.H. Weymuller. Measuring liquidity mismatch in the banking sector. *J. Fin.* 2018, 73(1), 51–93.
- [12] E. Bengtsson. Investment funds, shadow banking and systemic risk. *Journal of Financial Regulation and Compliance* 2016, 24(1), 60–73.
- [13] S. Benoit, J.E. Colliard, C. Hurlin and C. Pérignon. Where the risks lie: a survey on systemic risk. *Review of Finance* 2017, 21(1), 109–152.
- [14] F. Biagini, J.P. Fouque, M. Frittelli and T. Meyer-Brandis. A unified approach to systemic risk with acceptance sets. *Math. Finance* 2018, 29, 329–67
- [15] M. Billio, M. Getmansky, A.W. Lo and L. Pellizon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J. Fin. Econ.* 2012, 104(3), 535–559.
- [16] R. Bookstaber, M. Paddrik and B. Tivnan. An agent-based model for financial vulnerability. *J. Econ. Interact. Coord.* 2018, 13, 433–466.

- [17] M. Boss, H. Elsinger, M. Summer and S. Thurner. Network topology of the interbank market. *Quant. Finance* 2004, 4, 677–684.
- [18] C. Brownlees and R.F. Engle. SRISK: a conditional capital shortfall measure of systemic risk. *Rev. Fin. Studies* 2017, 30(1), 48–79.
- [19] M.K. Brunnermeier, G. Gorton and A. Krishnamurthy. Liquidity mismatch measurement. *Risk Topography: Systemic Risk and Macro Modeling* 2014, 99–112.
- [20] M.K. Brunnermeier and M. Oehmke. Bubbles, financial crises, and systemic risk. *Handbook of the Economics of Finance* 2013, Volume 2, Part B, 1221–1288.
- [21] M.K. Brunnermeier and L.H. Pedersen. Market liquidity and funding liquidity. *Rev. Fin. Studies* 2009, 22(6), 2201–2238.
- [22] M.K. Brunnermeier and Y. Sannikov. A macroeconomic model with a financial sector. *Amer. Econ. Rev.* 2014, 104(2), 379–421.
- [23] O. Burrows, D. Learmonth and J. McKeown. RAMSI: A top down stress testing model. *Financial Stability Paper* 17, 2012, Bank of England.
- [24] S. Calimani, G. Halaj and D. Żochowski. Simulating fire-sales in a banking and shadow banking system. *European Systemic Risk Board Working Paper* 46, 2017.
- [25] G. Cespa and T. Foucault. Illiquidity contagion and liquidity crashes. *Rev. Fin. Studies* 2014, 27, 1615–1660.
- [26] J.A. Chan-Lau and T. Gravelle. The END: A new indicator of financial and non-financial corporate sector vulnerability. *IMF Working Paper* 231, 2005.
- [27] C. Chen, G. Iyengar, G. and C.C. Moallemi. An axiomatic approach to systemic risk. *Manag. Science* 2013, 59(6), 1373–1388.
- [28] P. Cheridito and T. Li. Dual characterization of properties of risk measures on Orlicz hearts. *Math. Fin. Economics* 2008, 2(1), 29–55.
- [29] P. Cheridito and T. Li. Risk measures on Orlicz hearts. *Math. Finance* 2009, 19(2), 189–214.
- [30] R. Cifuentes, G. Ferrucci and H.S. Shin. Liquidity risk and contagion. *J. Europ. Econ. Association* 2005, 3, 556–566.
- [31] D.W. Diamond and R.G. Rajan. Liquidity shortages and banking crises. *J. Finance* 2005, 60, 615–647.
- [32] J.C. Duan and J. Wei. Systematic risk and the price structure of individual equity options. *Rev. Fin. Studies* 2009, 22(5), 1981–2006.
- [33] L. Eisenberg and T.H. Noe. Systemic risk in financial systems. *Manag. Science* 2001, 47(2), 236–249.

- [34] M. Elliott, B. Golub and M.O. Jackson. Financial networks and contagion. *Amer. Econ. Review* 2014, 104, 3115–3153.
- [35] H. Elsinger, A. Lehar and M. Summer. Risk assessment for banking systems. *Manag. Science* 2006, 52(9), 1301–1314.
- [36] Z. Feinstein, B. Rudloff and S. Weber. Measures of systemic risk. *SIAM J. Fin. Math.* 2017, 8, 672–708.
- [37] J. Figue. The macrofinancial risk assessment framework (MFRAF), Version 2.0. Bank of Canada 2017.
- [38] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. Second Edition. Walter de Gruyter Inc. 2004.
- [39] X. Freixas, B.M. Parigi and J.C. Rochet. Systemic risk, interbank relations, and liquidity provision by the Central Bank. *J. Money, Credit and Banking* 2000, 32, 611–638.
- [40] P. Gai and S. Kapadia. Contagion in financial networks. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 2010, 466, 2401–2423.
- [41] K. Giesecke and B. Kim. Systemic risk: what defaults are telling us. *Manag. Science* 2011, 57(8), 1387–1405.
- [42] S. Giglio. Credit default swap spreads and systemic financial risk. *European Systemic Risk Board Report Working Paper* 2016.
- [43] M. Gofman. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *J. Fin. Economics* 2017, 124(1), 113–146.
- [44] Z. He and A. Krishnamurthy. A macroeconomic framework for quantifying systemic risk. *NBER Working Paper* 2014.
- [45] X. Huang, H. Zhou and H. Zhu. Systemic risk contributions. *J. Fin. Serv. Research* 2012, 42, 55–83. *Amer. Econ. Review* **2016**, 106(3), 699–738.
- [46] E. Kromer, L. Overbeck and K. Zilch. Systemic risk measures on general measurable spaces. *Math. Methods Oper. Res.* 2017, 84, 323–357.
- [47] A. Lehar. Measuring systemic risk: a risk management approach. *J. Banking & Finance* 2005, 29(10), 2577–2603.
- [48] Y. Leitner. Financial networks: contagion, commitment, and private sector bailouts. *J. Finance* 2005, 60, 2925–2953.
- [49] F. Maccheroni, M. Marinacci and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* 2006, 74, 1447–1498.
- [50] C. Matten. *Managing Bank Capital*. Wiley, Chichester 1996.

- [51] R. Merton and A. Perold. Theory of risk capital in financial firms. *J. Appl. Corporate Finance* 1993, 6(3), 16–32.
- [52] Z. Pozsar, T. Adrian, A. Ashcraft and H. Boesky. Shadow banking. *Federal Reserve Bank of New York Staff Report* 458, 2012.
- [53] C. Pühr and S.W. Schmitz. A view from the top: The interaction between solvency and liquidity stress. *J. Risk Manag. in Institutions* 2014, 7(1), 38–51.
- [54] J.C. Stein. Monetary policy as financial stability regulation. *Quart. J. Economics* 2012, 127(1), 57–95.
- [55] C. Upper. Using counterfactual simulations to assess the danger of contagion in interbank markets. *BIS Working Paper* 234, 2007.
- [56] M.L. Weitzman. Prices vs. quantities. *Rev. Econ. Stud.* 1974, 41(4), 477–491.