Journal of Financial Economics 99 (2011) 333–348

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Diversification disasters ☆

Rustam Ibragimov [a,1], Dwight Jaffee [b,2], Johan Walden [b,*]

[a] Department of Economics, Harvard University, Littauer Center, 1875 Cambridge St., Cambridge, MA 02138, USA
[b] Haas School of Business, University of California at Berkeley, 545 Student Services Building #1900, CA 94720-1900, USA

ARTICLE INFO

Article history:
Received 20 November 2009
Received in revised form
8 February 2010
Accepted 10 March 2010
Available online 17 September 2010

JEL classification:
G20
G21
G28

Keywords:
Financial crisis
Financial institutions
Systemic risk
Limits of diversification

ABSTRACT

The recent financial crisis has revealed significant externalities and systemic risks that arise from the interconnectedness of financial intermediaries' risk portfolios. We develop a model in which the negative externality arises because intermediaries' actions to diversify that are optimal for individual intermediaries may prove to be suboptimal for society. We show that the externality depends critically on the distributional properties of the risks. The optimal social outcome involves less risk-sharing, but also a lower probability for massive collapses of intermediaries. We derive the exact conditions under which risk-sharing restrictions create a socially preferable outcome. Our analysis has implications for regulation of financial institutions and risk management.

1. Introduction

It is a common view that the interdependence of financial institutions, generated by new derivative products, had a large role in the recent financial meltdown. A primary mechanism is that by entering into sophisticated derivative contracts, e.g., credit default swaps (CDS) and collateralized debt obligations (CDO), financial institutions became so heavily exposed to each others' risks that when a shock eventually hit, it immediately spread through the whole system, bringing down critical parts of the financial sector. Since it created this systemic failure, the interdependence was extremely costly from a social standpoint.[1]

There is an opposite viewpoint, however, namely that such interdependence plays a much more functional role—that of diversification. To motivate this in a simple way, start with a case in which each financial firm holds a particular risk class with its unique idiosyncratic risk. Now allow the firms to form a joint mutual market portfolio, with each firm contributing its risky portfolio to

☆ We thank Greg Duffee, Paul Embrechts, Vito Gala, Todd Keister, seminar participants at the Wharton Symposium on the Measurement of Low Probability Events in the context of Financial Risk Management, April 16–17, 2009, the Second Annual NHH Symposium on Extreme Events, Bergen Norway, May 16th 2009, the 2010 FIRS Conference on Banking, Insurance and Intermediation, and at the 2010 meetings of the Western Finance Association. Ibragimov and Walden thank the NUS Risk Management Institute for support. Ibragimov also gratefully acknowledges partial support provided by the National Science Foundation grant SES-0820124 and a Harvard Academy Junior Faculty Development grant.

* Corresponding author. Tel.: +1 510 643 0547; fax: +1 510 643 1420.
E-mail addresses: ribragim@fas.harvard.edu (R. Ibragimov), jaffee@haas.berkeley.edu (D. Jaffee), walden@haas.berkeley.edu (J. Walden).
[1] Tel.: +1 203 887 1631; fax: +1 617 495 7730.
[2] Tel.: +1 510 642 1273; fax: +1 510 643 7441.

0304-405X/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2010.08.015

[1] This view is, e.g., expressed in Warren Buffet's 2009 letter to the shareholders of Berkshire Hathaway in which Mr. Buffet argues that the result is that "a frightening web of mutual dependence develops among huge financial institutions." A similar view is expressed in the academic literature in Jaffee (2009).

the total and receiving back its proportional share of the total. Given a sufficient variety and number of risk classes, the firms may succeed in eliminating the idiosyncratic risks embedded in their individual portfolios. This is in line with the classical view in finance, that risk-sharing—i.e., diversification—is always valuable (see, e.g., Samuelson, 1967). Therefore, interdependence is valuable and, indeed, what we should expect. In practice, we may expect both effects to be present, i.e., by sharing risks, intermediaries decrease the risk of individual failure, but increase the risk of massive, systemic failure.[2]

Which factors determine the risks of systemic failures of financial institutions and the benefits of diversification? When do the risks outweigh the benefits? What are the policy implications of such a trade-off? In this paper, we analyze these questions by introducing a parsimonious model that combines the two aspects of diversification in an integrated analysis. Along the lines of the previous viewpoints, while individual institutions may have an incentive to diversify their risks, diversification creates a negative externality in the form of systemic risk. If all intermediaries are essentially holding the same diversified portfolio, a shock may disrupt all the institutions simultaneously, which is costly to society, since it may take time for the financial system, and thereby the economy, to recover. Specifically, the slow recovery time creates a significant and continuing social cost because the unique market-making and information analysis provided by banks and other intermediaries[3] is lost until they recover; see Bernanke (1983). Indeed, Bernanke's concern with the social cost created by bank failures appears to have motivated many of the government bank bailouts.

We show that the costs and benefits of risk-sharing are functions of five properties of the economy. First, the number of asset classes is crucial: The fewer the number of distinct asset classes that are present, the weaker the case for risk-sharing. Second and third, the correlation between risks within an asset class, and the heavy-tailedness of the risks are important. The higher the correlation and the heavier the tails of the risk distribution, the less beneficial risk-sharing is. Fourth, the longer it takes for the economy to recover after a systemic failure, the more costly risk-sharing is and, fifth, lower discount rates also work against risk-sharing. We define the *diversification threshold* to be the threshold at which the cost to society of systemic failure begins to exceed the private benefits of diversification, and we derive a formula for the threshold as a function of these five properties.

The distributions of the risks that intermediaries take on are key to our results. When these risks are thin-tailed, risk-sharing is always optimal for both individual intermediaries and society. But, with moderately heavy-tailed risks, risk-sharing may be suboptimal for society, although individual intermediaries still benefit from it. In this case, the interests of society and intermediaries are unaligned. For extremely heavy-tailed risks, intermediaries and society once again agree, this time that risk-sharing is suboptimal.

One can argue that the focus of our study, moderately heavy-tailed distributions, is the empirically interesting case to study. It is well-known that diversification may be suboptimal in the extremely heavy-tailed case,[4] and some risks may indeed have extremely heavy tails (e.g., catastrophic losses, in which case individual insurers may withdraw from the market precisely because the benefits of diversification are unavailable; see Ibragimov, Jaffee, and Walden, 2009). However, arguably most financial risks are moderately heavy-tailed, as shown by several empirical studies in recent years. Specifically, the rate at which a distribution decreases for large values, the so-called *tail exponent*, α (which we rigorously define in the paper), provides a useful classification of heavy-tailedness. Risks with $\alpha < 1$ are extremely heavy-tailed, whereas risks with $1 < \alpha < \infty$ are moderately heavy-tailed, and when $\alpha = \infty$, they are thin-tailed. Many recent studies argue that the tail exponents in heavy-tailed models typically lie in the interval $2 < \alpha < 5$ for financial returns on various stocks and stock indices (see, among others, Jansen and de Vries, 1991; Loretan and Phillips, 1994; Gabaix, Gopikrishnan, Plerou, and Stanley, 2006; Gabaix, 2009). Among other results, Gabaix, Gopikrishnan, Plerou, and Stanley (2006) and Gabaix (2009) provide theoretical results and empirical estimates that support heavy-tailed distributions with tail exponents $\alpha \approx 3$ for financial returns on many stocks and stock indices in different markets.[5]

Our analysis has implications for risk management and policies to mitigate systemic externalities. We show that value at risk (VaR) considerations lead individual intermediaries to diversify, as per incentives similar to those in the Basel bank capital requirements. Within our framework, however, the diversification actions may lead to suboptimal behavior from a societal viewpoint. It then becomes natural to look for devices that would allow individual firms to obtain the benefits of diversification, but without creating a systemic risk that could topple the entire financial system. In Section 4, we provide a framework to develop such solutions and provide specific proposals.

Our paper is related to the recent, rapidly expanding literature on systemic risk and market crashes. The closest paper is Acharya (2009). Our definition of systemic risk is similar to Acharya's, as are the negative externalities of joint failures of intermediaries. The first and foremost

[2] That the risk of massive failure increases when risks are shared was noted within a finance context by Shaffer (1994).

[3] Our analysis applies to banks, but more broadly to general financial intermediaries, like pension funds, insurance companies, and hedge funds.

[4] See Mandelbrot (1997), Fama (1965), Ross (1976), Ibragimov and Walden (2007), Ibragimov (2009b), the review in Ibragimov (2009a), and references therein.

[5] As discussed in, e.g., Lux (1998), Guillaume, Dacorogna, Davé, Müuller, Olsen, and Pictet (1997), and Gabaix, Gopikrishnan, Plerou, and Stanley (2006), tail exponents are similar for financial and economic time series in different countries. For a general discussion of heavy-tailedness in financial time series, see the discussion in Loretan and Phillips (1994), Rachev, Menn, and Fabozzi (2005), Embrechts, Klüuppelberg, and Mikosch (1997), Gabaix, Gopikrishnan, Plerou, and Stanley (2006), Ibragimov (2009a), and references therein.

difference between the two papers is our focus on the distributional properties of risks and the number of risk classes in the economy, which is not part of the analysis in Acharya (2009). Moreover, the mechanisms that generate the systemic risks are different in the two papers. Whereas the systemic risk in Acharya (2009) arises when individual intermediaries choose correlated real investments, in our model the systemic risk is introduced when intermediaries with limited liability become interdependent when they hedge their idiosyncratic risks by taking positions in what is in effect each others' risk portfolios. Such interdependence may have been especially important for systemic risk in the recent financial crisis. This leads to a distinctive set of policy implications, as we develop in Section 4.

Wagner (2010) independently develops a model of financial institutions in which there are negative externalities of systemic failures, and diversification therefore may be suboptimal from society's perspective. Wagner's analysis, however, focuses on the effects of conglomerate institutions created through mergers and acquisitions and the effects of contagion. Furthermore, the intermediary size and investment decisions are exogenously given in Wagner's study, and only a uniform distribution of asset returns is considered. Our model, in contrast, emphasizes the importance of alternative risk distributions and the number of risks in determining the possibly negative externality of diversification. The two studies therefore complement each other.

A related literature models market crashes based on contagion between individual institutions or markets. A concise survey is available in Brunnermeier (2009). Various mechanisms to propagate the contagion have been used. Typically it propagates through an externality, in which the failure of some institutions triggers the failure of others. Rochet and Tirole (1996) model an interbank lending market, which intrinsically propagates a shock in one bank across the banking system. Allen and Gale (2000) extend the Diamond and Dybvig (1983) bank run liquidity risk model, such that geographic or industry connections between individual banks, together with incomplete markets, allows for shocks to some banks to generate industry-wide collapse. Kyle and Xiong (2001) focus on cumulative price declines that are propagated by wealth effects from losses on trader portfolios. Kodes and Pritsker (2002) use informational shocks to trigger a sequence of synchronized portfolio rebalancing actions, which can depress market prices in a cumulative fashion. Caballero and Krishnamurthy (2008) focus on Knightian uncertainty and ambiguity aversion as the common factor that triggers a flight to safety and a market crash. Most recently, Brunnermeier and Pedersen (2009) model a cumulative collapse created by margin requirements and a string of margin calls.

The key commonalities between our paper and this literature is the possibility of an outcome that allows a systemic market crash, with many firms failing at the same time. Moreover, as in many other papers, in our model there is an externality of the default of an intermediary—in our case, the extra time it takes to recover when many defaults occur at the same time. The key distinction between our paper and this literature is, again, our focus on the importance of risk distributions and number of asset classes in an economy. Thus, a unique feature of our model is that the divergence between private and social welfare arises from the statistical features of the loss distributions for the underlying loans alone. This leads to strong, testable implications and to distinctive policy implications. In our model, these effects arise even without additional assumptions about agency problems (e.g., asymmetric information) or third-party subsidies (e.g., government bailouts). No doubt, such frictions and distortions would make the incentives of intermediaries and society even less aligned.

The paper is organized as follows. In the next section we introduce some notation. In Section 3, we introduce the model, and in Section 4 we discuss its potential implications for risk management and policy making. Finally, some concluding remarks are made in Section 5. Proofs are left to a separate appendix. To simplify the reading, we provide a list of commonly used variables at the end of the paper.

2. Notation

We use the following conventions: lower case thin letters represent scalars, upper case thin letters represent sets and functions, lower case bold letters represent vectors, and upper case bold letters represent matrices. The ith element of the vector $\mathbf{v}$ is denoted $(\mathbf{v})_i$, or $\mathbf{v}_i$ if this does not lead to confusion, and the n scalars $v_i, i=1,\ldots,n$ form the vector $[v_i]_i$. We use T to denote the transpose of vectors and matrices. One specific vector is $\mathbf{1}_n = (\underbrace{1,1,\ldots,1}_{n})^T$, (or just $\mathbf{1}$ when n is obvious). Similarly, we define $\mathbf{0}_n = (\underbrace{0,0,\ldots,0}_{n})^T$.

The expectation and variance of a random variable, ξ, are denoted by $E[\xi]$ (or simply $E\xi$) and $var(\xi)$ (or $\sigma^2(\xi)$), respectively, provided they are finite. The correlation and covariance between two random variables ξ_1 and ξ_2 with $E\xi_1^2 < \infty$, $E\xi_2^2 < \infty$, are denoted by $cov(\xi_1,\xi_2)$ and $corr(\xi_1,\xi_2)$, respectively. We will make significant use, in particular, of the Pareto distribution. A random variable, ξ with cumulative distribution function (c.d.f.) $F(x)$ is said to be of Pareto-type (in the left tail), if

$$F(-x) = \frac{c+o(1)}{x^{\alpha}}\ell(x), \quad x \to +\infty, \tag{1}$$

where c,α are some positive constants and $\ell(x)$ is a slowly varying function at infinity:

$$\frac{\ell(\lambda x)}{\ell(x)} \to 1$$

as $x \to +\infty$ for all $\lambda > 0$. Here, $f(x) = o(1)$ as $x \to +\infty$ means that $\lim_{x\to+\infty} f(x) = 0$. The parameter α in (1) is referred to as the tail index or tail exponent of the c.d.f. F (or the random variable ξ). It characterizes the heaviness (the rate of decay) of the tail of F. We define a distribution to be *heavy-tailed* (in the loss domain) if it satisfies (1) with $\alpha \le 1$. It is *moderately heavy-tailed* if $1 < \alpha < \infty$, and it is *thin-tailed* if $\alpha=\infty$, i.e., if the distribution decreases faster than any algebraic power.

3. Model

Consider an infinite horizon economy, $t\in\{0,1,2,\ldots\}$, in which there are M different risk classes. Time value of money is represented by a discount factor $\delta<1$ so that the present value of one dollar at $t=1$ is δ. There is a bond market in perfectly elastic supply, so that at t, a risk-free bond that pays off one dollar at $t+1$ costs δ.

There are M risk-neutral trading units, each trading in a separate risk class. We may think of unit m as a representative trading unit for risk class m. Henceforth, we shall call these trading units *intermediaries*, capturing a large number of financial institutions, like banks, pension funds, insurance companies, and hedge funds. We thus assume that each trading unit, or intermediary, specializes in one risk class. Of course, in reality, intermediaries hold a variety, perhaps a wide variety of risks. The key point here is that the intermediaries are not initially holding the market portfolio of risks, so that they may have an incentive to share risks with each other.

We think of the M risk classes as different risk lines or "industries," e.g., representing real estate, publicly traded stocks, private equity, etc. Within each risk class, in each time period t, there is a large number, N, of individual multivariate normally distributed risks, $x_n^{t,m}$, $1\le n\le N$. We will subsequently let N tend to infinity, whereas M will be a small constant, typically less than 50, as is typical in financial and insurance applications (the results in the paper also hold in the case when the number of risks, N, in the mth class depends on m, as long as we let the number of risks in each risk class tend to infinity). For simplicity, we assume that risks belonging to different risk classes are independent, across time t, and risk class, n i.e., $x_n^{t,m}$ is independent of $x_{n'}^{t',m'}$ if $m\neq m'$ or $t'\neq t$. This is not a crucial assumption; similar results would arise with correlated risk classes.

For the time being, we focus on the first risk class, in time period zero. We therefore drop the m and t superscripts. Per assumption, the individual risks have multivariate normal distributions, related by

$$x_{i+1}=\rho x_i+w_{i+1},\quad i=1,\ldots,N-1, \tag{2}$$

for some $\rho\in[0,1)$.[6]

Here, w_i are independent and identically distributed (i.i.d.) normally distributed random variables with zero mean and variance $\sigma^2(w_i)=1-\rho^2$. Each x_i represents cross-sectional risk, with local dependence in the sense that $cov(x_i,x_j)$ quickly approaches zero when $|i-j|$ grows, i.e., the decay is exponential. The risks could, for example, represent individual mortgages and the total risk class would then represent all the mortgages in the economy. For low ρ, the risks of these mortgages are effectively uncorrelated, except for risks that are very close. "Close" here could, for example, represent mortgages on houses in the same geographical area. If ρ is close to one, shocks are correlated across large distances, e.g., representing country-wide shocks to real estate prices. This structure thus allows for both "local" and "global" risk dependencies in a simple setting.[7]

For simplicity, we introduce symmetry in the risk structure by requiring that

$$x_1=\rho x_N+w_1, \tag{3}$$

i.e., the relationship between x_1 and x_N is the same as that between x_{i+1} and x_i, $i=1,\ldots,N-1$. We can rewrite the risk structure in matrix notation, by defining $\mathbf{x}=[x_i]_i$, $\mathbf{w}=[w_i]_i$. The relationship (2), (3) then becomes

$$\mathbf{Ax}=\mathbf{w}.$$

Here, $\mathbf{A}$ is an invertible so-called circulant Toeplitz matrix,[8] given by

$$\mathbf{A}=\text{Toeplitz}_N[-\rho,\underline{1},\mathbf{0}_{N-2}^T,-\rho].$$

Thus, given the vector with independent noise terms, $\mathbf{w}$, the risk structure, $\mathbf{x}$, is defined by

$$\mathbf{x}\stackrel{\text{def}}{=}\mathbf{A}^{-1}\mathbf{w}.$$

The symmetry is merely for tractability and we would expect to get similar results without it (although at the expense of higher model complexity). In fact, for large N, the covariance structure in our model is very similar to that of a standard AR(1) process, defined by

$$\hat{x}_0=\hat{w}_0,$$

$$\hat{x}_i=\rho\hat{x}_i+\hat{w}_{i+1},\quad i=2,\ldots,N,$$

where the $\hat{w}_i$'s are independent, $\sigma^2(\hat{w}_1)=1$, $\sigma^2(\hat{w}_i)=1-\rho^2$, $i>1$ (although, of course, i does not denote a time subscript in our model, as it does in a standard AR(1) process). In this case, the matrix notation becomes $\hat{\mathbf{A}}\hat{\mathbf{x}}=\hat{\mathbf{w}}$, where $\hat{\mathbf{A}}=\text{Toeplitz}_N[-\rho,\underline{1}]$. The only difference between $\mathbf{A}$ and $\hat{\mathbf{A}}$ is that $\mathbf{A}_{1N}=1$, whereas $\hat{\mathbf{A}}_{1N}=0$. The covariance matrix $\hat{\Sigma}=[cov(\hat{x}_i,\hat{x}_j)]_{ij}$ has elements

$$\hat{\Sigma}_{ij}=\rho^{|i-j|}.$$

Thus, in the AR(1) setting, the covariance between risks decreases geometrically with the distance $|i-j|$. As we shall see in Theorem 1, this property carries over to the covariances in our symmetric structure when the number of risks within an asset class, N, is large.

[6] We focus on multivariate normal risks, for tractability. Similar results arise with other, thin-tailed, individual risks, e.g., Bernoulli distributions, although the analysis becomes more complex, because other distribution classes are not closed under portfolio formation so the central limit theorem needs to be incorporated into the analysis.

[7] For review and discussion of models with common shocks and modeling approaches for spatially dependent economic and financial data, see, among others, Conley (1999), Andrews (2005), Ibragimov and Walden (2007), and Ibragimov (2009b).

[8] A Toeplitz matrix $\mathbf{A}=\text{Toeplitz}_N[a_{-N+1},a_{-N+2},\ldots,a_{-1},a_0,a_1,\ldots,a_{N-2},a_{N-1}]$, is an $N\times N$ matrix with the elements given by $(\mathbf{A})_{ij}=a_{i-j}$, $1\le i\le N$, $1\le j\le N$. A Toeplitz matrix is banded if $(A)_{ij}=0$ for large $|i-j|$, corresponding to $a_i=0$ for indices i that are large by absolute value. When $a_i=0$ if $i<-k$ or $i>m$, for $k<N-1$, $m<N-1$, we use the notation $a_{-k},a_{-k+1},\ldots,\underline{a_0},\ldots,a_{m-1},a_m$ to represent the whole sequence generating the Toeplitz matrix. For example, the notation $\mathbf{A}=\text{Toeplitz}_N[a_{-1},\underline{a_0}]$ then means that $\mathbf{A}_{ii}=a_0$, $\mathbf{A}_{i,i-1}=a_{-1}$, and that all other elements of $\mathbf{A}$ are zero. For an $N\times N$ Toeplitz matrix, if $a_{N-j}=a_{-j}$, then the matrix is, in addition, circulant. See Horn and Johnson (1990) for more on the definition and properties of Toeplitz and circulant matrices.

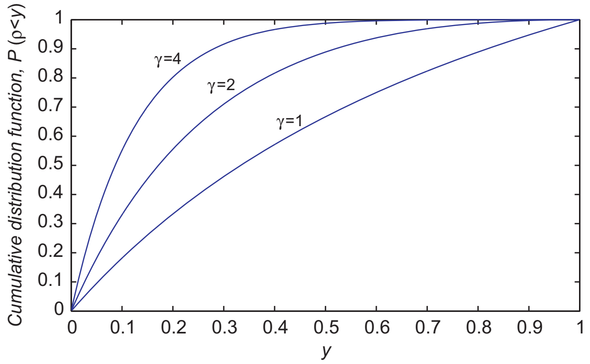

Fig. 1. Cumulative distribution function for correlation: $\mathbb{P}(\rho \le y) = 1-((1-y)/(1+y))^{\gamma}$, with tail indices, $\gamma = 1,2,4$.

We assume that the correlations between individual risks (2) are uncertain. Specifically,

$$\mathbb{P}(\rho \le y) = 1-\left(\frac{1-y}{1+y}\right)^{\gamma}, \quad y \in [0,1), \ \gamma > 0, \tag{4}$$

i.e., the probability that the correlation is close to one satisfies a power-type law with the tail index γ, where a low γ indicates that there is a substantial chance that correlations are high, whereas a high γ indicates that correlations are very likely low. The uncertainty of correlations will lead to heavier-tailed distributions for *portfolios* of risks, just as uncertain variances may cause the distribution of an *individual* risk to have a heavy tail.[9]

Relation (4) is equivalent to the condition that the ratio $(1+\rho)/(1-\rho)$ follows the Pareto distribution with the tail index γ similar to (1):

$$\mathbb{P}\left(\frac{1+\rho}{1-\rho} \ge \upsilon\right) = \upsilon^{-\gamma}, \quad \upsilon \ge 1. \tag{5}$$

In Fig. 1, we show the c.d.f. of ρ for some values of γ.

We first consider the case in which intermediaries do not trade risks with each other. At t=0, an intermediary invests in a portfolio of x_i unit risks, with the total portfolio size determined by the vector $\mathbf{c} \in \mathbb{R}^N$. The total portfolio is then $\mathbf{c}^T\mathbf{x}$.[10] Here, $\mathbf{c}_i$ represents the number of units of $\mathbf{x}_i$ risk that the intermediary invests in. The total dollar outcome of the investment in risk i, after the risks are realized, is $\mathbf{c}_i\mathbf{x}_i$ and the total dollar value of the whole portfolio, after realization, is $\mathbf{c}^T\mathbf{x}$. Since the elements of $\mathbf{x}$ can take on negative values, it is natural to think of $\mathbf{c}^T\mathbf{x}$ as containing both long and short positions. The short positions may be in risk-free capital (i.e., leverage through debt), but also in risky assets. At t=0, the intermediary also reserves capital, so that $k \le K$ is available at t=1, where K represents the maximum capital (in dollar terms) available for the intermediary to reserve and is exogenously given. For simplicity, we assume that the capital is invested in risk-free cash. We could allow the capital to be invested in the risky portfolio, but this would further complicate the formulas without any new insights. The intermediary has limited liability, so its losses are bounded by k. If the intermediary defaults, the additional losses are borne by the counter-party. The shortfall that may be imposed on the counter-party is taken into account in the asset pricing.

At t=1, the values of the t=0 risks are realized. Because of limited liability, the value of the portfolio (in dollars) is then

$$\max(k+\mathbf{c}^T\mathbf{x},0) \stackrel{\text{def}}{=} \mathbf{c}^T\mathbf{x}+k+Q, \tag{6}$$

where $Q = \max(k+\mathbf{c}^T\mathbf{x},0)-\mathbf{c}^T\mathbf{x}-k$ is the realized value of the option to default (see Ibragimov, Jaffee, and Walden, 2010). Since the intermediary reserved capital, so that k be available at $t=1$, the value of the portfolio of investments above the capital reserved at $t=1$ is $\mathbf{c}^T\mathbf{x}$ + Q.

The price the intermediary pays at $t=0$ for this portfolio of investments is its discounted expected value, $\delta E[\mathbf{c}^T\mathbf{x}+Q]$, less a premium, d, per unit risk. This premium represents the part of the surplus generated by the transaction that is captured by the intermediary. A similar feature is used in the work of Froot, Scharfstein, and Stein (1993) and Froot and Stein (1998). In a full equilibrium model, d would be endogenously derived, but for tractability we assume that it is exogenously given and that it is constant.[11]

If the intermediary defaults, it is out of business from there on, generating zero cash flows in all future time periods.[12] The sequence of events and the total cash flows at times zero and one are shown in Fig. 2. Later, we will introduce the possibility for the intermediaries to trade with each other, after t=0 (when they take on the risk), but before t=1 (when the value of the risk is realized), but for the time being, we ignore this possibility.

The ex ante value of the total cash flows to the owners of the intermediary between t=0 and 1 is then

$$\underbrace{d(\mathbf{c}^T\mathbf{1})-\delta k-\delta E[\mathbf{c}^T\mathbf{x}+Q]}_{t=0}+\underbrace{\delta E[(k+\mathbf{c}^T\mathbf{x}+Q)]}_{t=1} = d(\mathbf{c}^T\mathbf{1}). \tag{7}$$

This is thus the net present value of operating the intermediary between t=0 and 1.

[9] As we shall see, our approach leads to very tractable formulas. An alternative approach for introducing more complex correlation structures than the standard multivariate normal one is to use copulas, which may also lead to heavy-tailed portfolio distributions.

[10] We allow for short-selling. In the proof, we show that the optimal portfolio does not involve short-selling, so we could equivalently have permitted only nonnegative portfolios, $\mathbf{c} \in \mathbb{R}^N_+$.

[11] In Ibragimov, Jaffee, and Walden (2008), equilibrium premiums are derived in a model with multiple risk factors and risk-averse agents. The general model, however, is quite intractable, and the simplifying assumptions in this paper allow us to carry through a more complete study of the role of risk distributions.

[12] It may be optimal, if possible, for the owners of the intermediary to infuse more capital even if losses exceed k, to keep the option of generating future profits alive. Equivalently, they may be able to borrow against future profits. Taking such possibilities into account would increase the point of default to a value higher than k, but would qualitatively not change the results, since there would always be some realized loss level beyond which the intermediary would be shut down even with such possibilities.

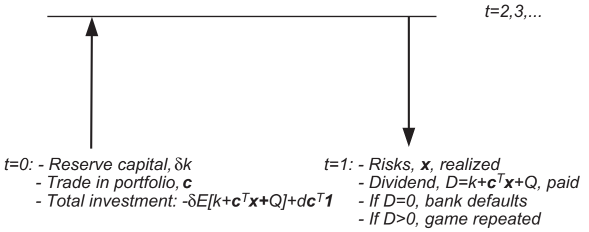

Fig. 2. Intermediary's cash flows in model. At $t=0$, the intermediary reserves capital k and trades in a risk portfolio $\mathbf{c}$. At $t=1$ risks are realized. If the intermediary defaults, it is out of business, otherwise, the game is repeated, i.e., at $t=1$, the intermediary reserves capital and trades in risk, at $t=2$ risks are realized, etc.

Now, if the intermediary survives, which it does with probability

$$q \stackrel{\text{def}}{=} \mathbb{P}(-\mathbf{c}^T\mathbf{x} \le k),$$

the situation is repeated, i.e., at $t=1$ the intermediary takes on new risk, reserves capital, and then at $t=2$ risks are realized, etc. If the intermediary defaults at any point in time, it goes out of business and its cash flows are zero from there on. In the infinite horizon case, the value of the intermediary in recursive form is therefore

$$V^B = d(\mathbf{c}^T\mathbf{1}) + \delta q V^B,$$

implying that

$$V^B = \frac{d(\mathbf{c}^T\mathbf{1})}{1-\delta q}. \tag{8}$$

We note that the counter-party of the $\mathbf{x}$ risk transaction understands that the intermediary may default, and takes this into account when the price for the risk is agreed upon. Therefore, since the price of the contract takes into account the risk level of the intermediary, the counter-party does not need to impose additional covenants.

Eq. (8) describes the trade-off the individual intermediary makes when choosing its portfolio. On the one hand, a larger portfolio increases the cash flows per unit time—increasing the numerator—but on the other hand, it also increases the risk of default—increasing the denominator. It is straightforward to show that if the intermediary could, it would take on an infinitely large portfolio. This is shown as a part of the proof of Theorem 1 in Appendix A. In terms of Eq. (8), the numerator effect dominates the denominator effect.

A regulator, representing society, therefore imposes restrictions on the probability for default, to counter-weight the risk-shifting motive. Specifically, the intermediary's one-period probability of default is not allowed to exceed β at any point in time. In other words, the intermediary faces the constraint that the value at risk at the $1-\beta$ confidence level can be at most k, $\mathrm{VaR}_{1-\beta}(\mathbf{c}^T\mathbf{x}) \le k$. Therefore, since the realized dollar losses are $\mathbf{c}^T\mathbf{x}$ and the capital reserved is k, the value at risk constraint says that the probability is at most β that the realized losses exceed the capital. In our notation, this is the same as to say that $q \ge 1-\beta$.

The VaR constraint is imposed in the model to reflect the existing management and regulatory standards through which most financial intermediaries currently operate. Most major financial intermediaries apply VaR as a management tool and regularly report their VaR values. The Basel II bank capital requirements are based on VaR. European securities firms face the same Basel II VaR requirements, while the major U.S. securities firms have become bank holding companies and thus also face these requirements. European insurance firms are being reregulated under "Solvency II," which is VaR-based, in parallel to Basel II. U.S. insurers are regulated by individual states with capital requirements that vary by state and insurance line; these are generally consistent with a VaR interpretation.

To be clear, VaR is not necessarily the optimal risk measure when financial intermediary behavior may create systemic externalities. Indeed, in this section we show that VaR requirements can lead to diversification decisions by individual intermediaries that are inconsistent with maximizing the societal welfare. In Section 4, we consider alternative proposals for intermediary regulation when systemic externalities are important.

It is natural to think of $\mathbf{c}^T\mathbf{1}/k$ as a measure of the leverage of the firm, since $\mathbf{c}^T\mathbf{1}$ represents the total liability exposure and k represents the capital that can be used to cover losses.

The assumption that there is an upper bound, K, that the intermediary faces on how much capital can be reserved is reduced form. In a full equilibrium model without frictions, the investment level would be chosen such that the marginal benefit and cost of an extra dollar of investment would be equal, and K would then be endogenously derived. In our reduced-form model, where the benefits d are constant, we assume that the marginal cost of raising capital beyond K is very high, so that K provides an effective hard constraint on the capital raising abilities of the intermediary. We note that the bound may be strictly lower than the friction-free outcome, because of other frictions on capital availability in imperfect financial markets (see, e.g., Froot, Scharfstein, and Stein, 1993 for a discussion of such frictions). Given a choice of capital, $k \le K$, the VaR constraint imposed by the regulator then imposes a bound on how aggressively the intermediary can invest, i.e., on the intermediary's size. The VaR constraint therefore also automatically imposes a capital requirement restriction on the intermediary. We next study the intermediary's behavior in this environment.

3.1. Optimal behavior of intermediary

In line with the previous arguments, the program for the intermediary is

$$\max_{\mathbf{c},k} \frac{d(\mathbf{c}^T\mathbf{1})}{1-\delta\mathbb{P}(-\mathbf{c}^T\mathbf{x} \le k)} \quad \text{s.t.,} \tag{9}$$

$$k \in [0,K], \tag{10}$$

$$\mathbf{c} \in \mathbb{R}^n, \tag{11}$$

$$k \ge \mathrm{VaR}_{1-\beta}(\mathbf{c}^T\mathbf{x}). \tag{12}$$

The following theorem characterizes the intermediary's behavior:

Theorem 1. *Given a VaR constrained intermediary solving* (9)–(12), *where the risks are on the form* (2,3), β *is close to zero, and the distribution of correlations is of the form* (4). *Then, for large N*:

(a) *For a given* ρ, *the covariance* $cov(x_i,x_j)$ *converges to* $\rho^{|i-j|}$ *for any i, j.*
(b) *The payoff of the intermediary's chosen risk portfolio,* $\mathbf{c}^T\mathbf{x}$, *is of Pareto-type with the tail index* 2γ, *and with the probability distribution function (p.d.f.)* $f(y)$, *where*

$$f(y)=\frac{\gamma 2^{\gamma}b^{-\gamma}}{\sqrt{\pi}}\times\frac{\Gamma\left(\gamma+\frac{1}{2},\frac{by^2}{2}\right)}{|y|^{1+2\gamma}}, \tag{13}$$

for some constant, $b>0$, *for all* $y\in\mathbb{R}\backslash\{0\}$. *Here,* Γ *is the lower incomplete Gamma function,* $\Gamma(a,x)\stackrel{\text{def}}{=}\int_0^x t^{a-1}e^{-t}\,dt$, $a>0$. *Also,* $f(0)=2\gamma/(b\sqrt{2\pi}(2\gamma+1))$.
(c) *Maximal capital is reserved,* $k=K$, *i.e.*, (10) *is binding.*
(d) *Maximal value at risk is chosen, i.e.*, (12) *is binding.*
(e) *If* $\gamma>1$, *then the variance of the portfolio is*

$$\sigma^2=b^2\frac{\gamma}{\gamma-1}, \tag{14}$$

else it is infinite.

We note that, since $\Gamma(a,x)=\Gamma(a)+o(1)$, as $x\to+\infty$, (where $\Gamma(a)$ is the Gamma function, $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}\,dt$, $a>0$), we obtain that $\mathbf{c}^T\mathbf{x}$ has a Pareto-type distribution in accordance with (1):

$$P(-\mathbf{c}^T\mathbf{x}>y)=\frac{2^{\gamma}b^{-\gamma}\Gamma\left(\gamma+\frac{1}{2}\right)+o(1)}{\gamma\sqrt{\pi}|y|^{2\gamma}},\quad y\to\infty.$$

Thus, the portfolio chosen by the intermediary when solving (9)–(12) is moderately heavy-tailed, or even heavy-tailed when $\gamma<\frac{1}{2}$, although the individual risks are thin-tailed. We define the random variable $\xi=\mathbf{c}^T\mathbf{x}$ and the constant $c=\mathbf{c}^T\mathbf{1}$, and show the p.d.f. of ξ for the two cases $\gamma=1$ (heavy-tailed distribution with infinite variance) and $\gamma=2.5$ (moderately heavy-tailed distribution with finite variance) in Fig. 3.

We also note that the total portfolio risk changes with the number of investments, N, in a very different way than what is implied by standard diversification results, for which the *size* of the investment is taken for given. If portfolio size were given, then the portfolio risk would vanish as N grew. However, taking into account that the intermediary can change its total exposure with the number of risks, in our framework, the risk does not mitigate, but instead converges to a portfolio risk that is much more heavy-tailed than the individual risks.

Our approach of letting the number of risks within an asset class grow has some similarities with the idea of an *asymptotically fine-grained portfolio*, (see Gordy, 2000, 2003), used for a theoretical motivation of the capital rules laid out in Basel II. Gordy shows that under the assumptions of one systematic risk factor and infinitely many small idiosyncratic risks, the portfolio invariant VaR rules of Basel II can be motivated. Similar to Gordy's setting, in our model each risk is a small part of the total asset class risk when N is large. However, in our setting all risk is not diversified away when the number of risks increases, as shown in Theorem 1. In fact, the risk that is not diversified away adds up to systematic risk in our model (which turns into systemic risk if it brings down the whole system), and since the number of risk classes, $M>1$, there will be multiple risk factors. This is contrary to the analysis in Gordy (2000), where, after diversification, all institutions hold the same one-factor risk. Therefore, the capital rules in Basel II would not be motivated in our model.

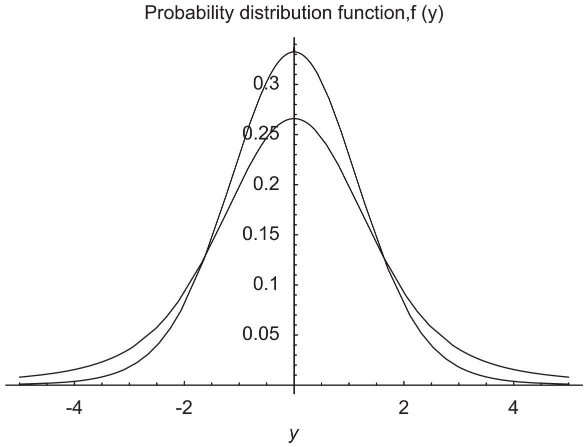

Fig. 3. Probability distribution function, f, of intermediary's chosen portfolio of risk, for $\gamma=1$, corresponding to a heavy-tailed distribution with infinite variance, and for $\gamma=2.5$, corresponding to a moderately heavy-tailed distribution with finite variance.

3.1.1. *The value to society*

We next turn to the value to society of the markets for the M separate risk classes. For the individual intermediary, the game ends if it defaults. From society's perspective, however, we would expect other players to step in and take over the business if an intermediary defaults, although it may take time to set up the business, develop client relationships, etc. Therefore, the cost to society of an intermediary's default may not be as serious as the value lost by the specific intermediary. The argument is that since there is a well-functioning market, it is quite easy to set up a copy of the intermediary that defaulted. We assume that an outside provider of capital steps in and sets up an intermediary identical to the one that defaulted. The argument per assumption, however, only works if there is a well-functioning market. In the unlikely event in which all intermediaries default, the market is not well-functioning and it may take a much longer time to set up the individual copy. Thus, individual intermediary default is less serious from society's perspective, but *massive* intermediary default (for simplicity, the case when all M intermediaries default at the same time) is much more serious. A similar argument is made in Acharya (2009).

To incorporate this reasoning into a tractable model, we use a special numerical mechanism. When only one individual intermediary fails (or a few intermediaries), the market continues to operate reasonably well. In such cases, the replacement of a failed intermediary occurs immediately if an intermediary defaults in an even period ($t=0,2,4,\ldots$), and takes just one period if it defaults in an odd period ($t=1,3,5,\ldots$). We note that in this favorable case without massive default, it takes, on average, half a period to rebuild after an intermediary's default (zero periods or one period with 50% chance each). On the other hand, when all intermediaries fail—a massive default—we assume there is a 50% chance that it will take another $2T$ periods ($2T+1-1$) to return to normal operations. Therefore, the expected extra time to recover after a massive default is $(50\%)\times(2T)+(50\%)\times(0)=T$. Henceforth, we call T the *recovery time after massive default*.

Our distinction between defaults in odd and even time periods significantly simplifies the analysis by ensuring that in even periods there are either 0 or M intermediaries in the market. Without the assumption, we would need to introduce a state variable describing how many intermediaries are alive at each point in time, which would increase the complexity of the analysis. Qualitatively similar results arise when relaxing the assumption, i.e., when instead assuming that it always takes one period to replace up to $M-1$ defaulting intermediaries and T periods if all M intermediaries default at the same time.

The assumption of longer recovery times after massive intermediary defaults can be viewed as a reduced-form description of the externalities imposed on society by intermediary defaults. The simplicity of the assumption allows us to carry out a rigorous analysis of the role of risk distributions, which is the focus of this paper. Microfoundations for such externalities have been suggested elsewhere in the literature, e.g., liquidity risk as in Allen and Gale (2000). In Allen and Gale (2000), the externality occurs when failure of intermediaries triggers failure of other intermediaries.

Society's and individual intermediaries' objectives may be unaligned. From our previous discussion, it follows that the value to society of all intermediaries at t=0 is, in recursive form:

$$V^S = Mdc+\delta qMdc+\delta^2(1-(1-q)^M)V^S+\delta^{2T+2}(1-q)^M V^S, \tag{15}$$

implying that

$$V^S=\frac{Mdc(1+\delta q)}{1-\delta^2+\delta^2(1-\delta^{2T})(1-q)^M}. \tag{16}$$

The first term on the right-hand side of (15) is the value generated between t=0 and 1, and the second term is the expected value between t=1 and 2. The third term is the discounted contribution to the value from t=2 and forward if all M intermediaries do not default (which occurs with probability $(1-(1-q)^M)$), and the fourth term is the contribution to the value if all M intermediaries default (which occurs with probability $(1-q)^M$). We shall see that this externality of massive intermediary default in the form of longer recovery time significantly decreases—or even reverses—the value of diversification from society's perspective and that the situations in which diversification is optimal versus suboptimal are easily characterized.

3.2. Risk-sharing

We next analyze what happens when the M different intermediaries, with identically distributed risk portfolios, $\xi_1,\ldots,\xi_M$, get the opportunity to share risks. We recall that the risk portfolios in the different risk classes are independently distributed. Therefore, as long as the intermediaries do not trade, a shock to one intermediary will not spread to others. If the intermediaries trade, however, systemic risk may arise when their portfolios become more similar because of trades. In what follows, we explore this idea in detail.

We assume that the intermediaries may trade risks at $t=\frac{1}{2},\frac{3}{2},\frac{5}{2},\ldots$, after they have formed their portfolio, but before the risks are realized. We assume that the VaR requirement must hold at all times. For example, even if the intermediaries trade with each other at $t=\frac{1}{2}$, the VaR requirement needs to be satisfied between $t=0$ and $\frac{1}{2}$.[13] Also, the counter-party of the risk trade correctly anticipates whether trades between intermediaries will take place at $t=\frac{1}{2}$, taking the impact of such a trade on the default option, Q, into account when the price is decided.

We focus on symmetric equilibria, in which all intermediaries choose to share risks fully. We define

$$q_M=\mathbb{P}\left(\frac{-\sum_{m=1}^{M}\xi_m}{M}\le K\right),$$

so q_M is the probability that total losses in the market are lower than total capital. Thus, $1-q_M$, is the probability for massive intermediary default if intermediaries fully share risks. We note, in passing, that systemic risk is generated through a different mechanism in our model than in Acharya (2009). In Acharya (2009), systemic risk arises when intermediaries take on correlated real investments, and intermediaries do not trade risks. In our model, the risk portfolios of different intermediaries are independent, and systemic risk arises when intermediaries become interdependent through trades. The latter type of systemic risk may have been especially important in the financial crisis.

All intermediaries hold portfolios of the same risk features, albeit in different risk classes, so when they share risks, the net "price" they pay at $t=\frac{1}{2}$ is zero. However, the states of the world in which default occurs are different when the intermediaries share risks. In fact, defaults will be perfectly correlated in the full risk-sharing situation: With probability $1-q_M$ a massive intermediary default occurs and with probability q_M no intermediary defaults. Risk-sharing could, of course, be achieved not

[13] This is not a critical assumption. Alternatively, we could have assumed that the regulator anticipates whether the individual intermediaries will trade risks, and adjusts the VaR requirements accordingly.

only by cross-ownership, but also by trading in derivatives contracts (credit default swaps, corporate debt, etc.).

For individual intermediaries, the value when sharing risks (using the same arguments as when deriving (8)) is then

$$V_M^B = \frac{dc}{1-\delta q_M}, \tag{17}$$

which should be compared with the value of not sharing, (8). Therefore, as long as

$$q_M > q, \tag{18}$$

the intermediaries prefer to trade risks, since it leads to a higher probability of survival, and thereby a higher value.

We note that there may be additional reasons for intermediaries to prefer risk-sharing. For example, if intermediaries anticipate that they will be bailed out in case of massive defaults, if managers are less punished if an intermediary performs poorly when all other intermediaries perform poorly too, or if the VaR restrictions are relaxed, this provides additional incentives for risk-sharing, which would amplify our main result.

3.2.1. The value to society

From society's perspective, a similar argument as when deriving Eqs. (15) and (16) shows that

$$V_M^S = Mdc + \delta q_M Mdc + \delta^2 q_M V_M^S + \delta^{2T+2}(1-q_M)V_M^S, \tag{19}$$

so the total value of risk-sharing between intermediaries is

$$V_M^S = \frac{Mdc(1+\delta q_M)}{1-\delta^2+\delta^2(1-\delta^{2T})(1-q_M)}. \tag{20}$$

Therefore, via (16), it follows that society prefers risk-sharing when

$$\frac{1+\delta q_M}{1+\lambda(1-q_M)} > \frac{1+\delta q}{1+\lambda(1-q)^M}, \tag{21}$$

where

$$\lambda = \frac{\delta^2(1-\delta^{2T})}{1-\delta^2}. \tag{22}$$

The value of λ determines the relative trade-off society makes between the costs in foregone investment opportunities of individual and massive intermediary default.[14] If $T=0$, there is no extra delay when massive default occurs. In this case, $\lambda=0$, and society's trade-off (21) is the same as individual intermediaries' (18). If, on the other hand, T is large and δ is close to one, representing a situation in which it takes a long while to set up a market after massive default and the discount rate is low, λ is large, and the trade-off between $(1-q_M)$ and $(1-q)^M$ in (21) becomes very important. In this case, society will mainly be interested in minimizing the risk of massive default and this risk is minimized when intermediaries do *not* share risk, since $1-q_M \geq (1-q)^M$.[15]

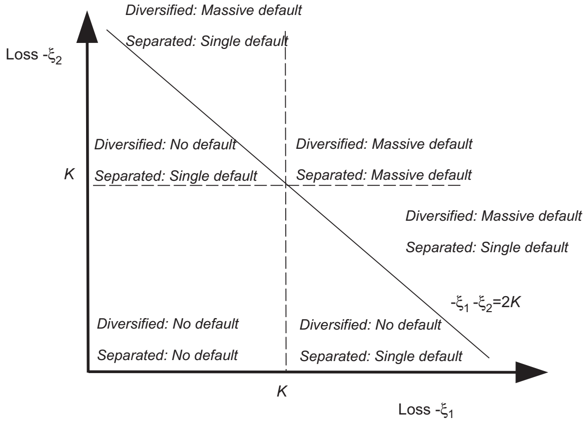

Fig. 4. Cost to society for diversified and separated cases, when there are two intermediaries. In the diversified case, massive default occurs if $-\xi_1-\xi_2 \geq 2K$. In the separated case, massive default occurs if $-\xi_1 \geq K$ and $-\xi_2 \geq K$, whereas single default occurs if $-\xi_1 \geq K$ and $-\xi_2 < K$, or if $-\xi_2 \geq K$ and $-\xi_1 < K$. Thus, massive default is rarer in the separated case than in the diversified case, but it is more common that at least one intermediary defaults in the separated case.

It is clear that society and individual intermediaries may disagree about whether it is optimal to diversify. Specifically, this occurs when (18) holds, but (21) fails. The opposite, that (18) fails but (21) holds, can never occur.[16] To understand the situation conceptually, we study the special case when there are two risk classes, $M=2$. In Fig. 4, we show the different outcomes depending on the realizations of losses in the diversified and separated cases. When both $-\xi_1 > K$ and $-\xi_2 > K$, it does not matter whether the intermediaries diversify or not, since there will be massive intermediary default either way. Similarly, if $-\xi_1 \leq K$ and $-\xi_2 \leq K$, neither intermediary defaults, regardless of whether they diversify or not. However, in the case when $-\xi_1 > K$ and $-\xi_2 \leq K$, the outcome is different, depending on the diversification strategy the intermediaries have chosen. In this case, if $-\xi_1-\xi_2 > 2K$, then a massive default occurs if the intermediaries are diversified, but only a single default if they are not. If $-\xi_1-\xi_2 \leq 2K$, on the other hand, no default occurs if the intermediaries are diversified, but a single default occurs if they are not. Exactly the same argument applies in the case when $-\xi_1 \leq K$ and $-\xi_2 > K$. Therefore, the optimal outcome depends on the trade-off between avoiding single defaults in some states of the world but introducing massive defaults in others, when diversifying. The point that risk-sharing increases the risk for joint failure is, of course, not new—it was made as early as in Shaffer (1994). Our analyses of the trade-off between diversification benefits and costs of massive failures, the importance of distributional properties, correlation uncertainty, and number of asset classes, as well as the implications for financial institutions, however, are to the best of our knowledge novel.

[14] It is easy to show that $\lambda \in [0,T]$, and that λ is increasing in δ and T.

[15] This follows trivially, since $1-q_M = \mathbb{P}(\sum_i -\xi_i > MK) \geq \mathbb{P}(\cap_i\{-\xi_i > K\}) = (1-q)^M$.

[16] This also follows trivially, since $1-q_M \geq (1-q)^M$ and therefore, if $q \geq q_M$, then $(1+\delta q)/(1+\lambda(1-q)^M) \geq (1+\delta q)/(1+\lambda(1-q_M)) \geq (1+\delta q_M)/(1+\lambda(1-q)^M)$.

It is clear that the objectives of the intermediaries and society will depend on the distributions of the $\xi-$risks and it may not be surprising that standard results from the theory of diversification apply to the individual intermediaries' problem. Society's objective function is more complex, however, since it trades off the costs of massive and individual intermediary defaults. It comes as a pleasant surprise that for low default risks (i.e., for a β close to zero in the $\text{VaR}_{1-\beta}$ constraint) given the distribution of the $\xi-$risk, we can completely characterize when the objectives of intermediaries and society are different.

Theorem 2. *Given i.i.d. asset class risks, $\xi_1,\ldots,\xi_M$, of Pareto-type (1) with the tail index $\alpha \neq 1$, and d, δ, and T as previously defined. Then, for low β,*

(a) *From an intermediary's perspective, risk-sharing is optimal if and only if $\alpha > 1$, regardless of the number of risk classes, M.*
(b) *From society's perspective, risk-sharing is optimal if and only if $\alpha > 1$ and $M > M_*$, where M_* is the diversification threshold*

$$M_* = \left(\frac{1-\delta^{2T+1}}{1-\delta}\right)^{1/(\alpha-1)}. \quad (23)$$

Thus, the break-even for individual intermediaries, from (18), is a Pareto distribution with the tail index of one, e.g., a Cauchy distribution. If the distribution is heavier, then (18) fails, and intermediaries and society therefore agree that there should be no risk-sharing. For thinner tails, intermediaries prefer risk-sharing. On the contrary, for society, what is optimal depends on M. It is easy to show that $M_* \in [1,(2T+1)^{1/(\alpha-1)})$, and that M_* is increasing in δ and T. We note that for the limit case when $\delta \to 0$, $\alpha \to \infty$, or T=0, i.e., when $M_* \to 1$, then society and intermediaries always agree.

It is useful to define

$$\eta = \frac{1-\delta^{2T+1}}{1-\delta}, \quad (24)$$

so that $M_* = \eta^{1/(\alpha-1)}$. This separates the impact on M_* of the tail behavior of the risk distributions from the other factors (δ and T). It follows immediately that $\eta \in [1,2T+1)$. In Fig. 5, we show η as a function of δ for $T=1$, 10, 20, 30.

From Theorem 2, it follows that if there is a large number of risk classes, M, society agrees with the individual intermediaries:

Corollary 1. *Given i.i.d. Pareto-type risk classes with the tail index α, if there is a large enough number of risk classes available, intermediaries and society agree on whether risk-sharing is optimal.*

In Fig. 6, we show the break-even number of risk classes, M_*, for Pareto-type risks with the tail indices $\alpha = 2,2.5,3,4$. With no externality of massive default ($\eta = 1$), society prefers diversification when $\alpha > 1$, just like individual intermediaries. However, when $\eta > 1$, diversification is suboptimal up until M_*. For example, for $\alpha = 2$, and $\eta = 10$, at least 10 risk-classes are needed for diversification to be optimal for society.

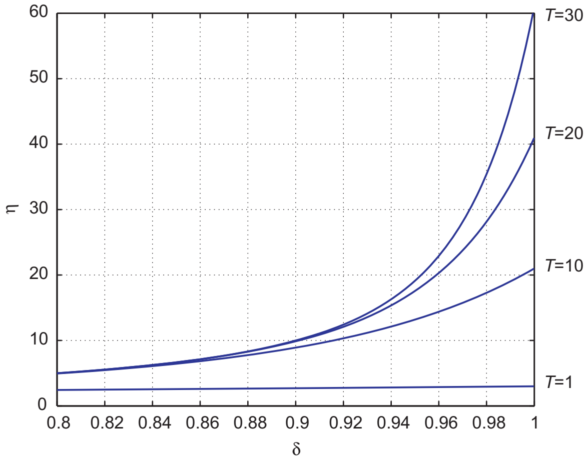

Fig. 5. The coefficient, $\eta = (1-\delta^{2T+1})/(1-\delta)$, as a function of the discount factor, δ, for different recovery times after massive default, T=1,10,20,30. All else equal, a higher η leads to a higher diversification threshold, M_*.

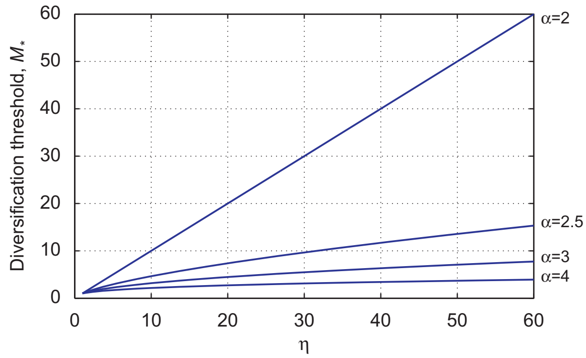

Fig. 6. Diversification threshold, M_*, for Pareto-type risk distributions with tail indices $\alpha = 2,2.5,3,4$, as a function of the coefficient $\eta = (1-\delta^{2T+1})/(1-\delta)$, where δ is the per period discount factor and T is the recovery time after massive default. The diversification threshold is the break-even number of risk classes needed for diversification to be optimal for society. Results are asymptotically valid for VaR confidence levels close to 100%, i.e., for β close to zero in (12).

Theorem 2 provides a complete characterization of the objectives of intermediaries and firms for VaR close to the 100% confidence level, by relating the number of risk classes, M, the discount rate, δ, the tail distribution of the risks, α, and the recovery time after massive default, T. It also relates to the uncertainty of correlations, γ, through the relation $\alpha = 2\gamma$. Theorem 2 is therefore the fundamental result of this paper. The theorem has several immediate empirical implications, since when (23) fails, we would expect there to exist financial regulations against risk-sharing across risk classes:

Implication 1. *Economies with heavier-tailed risk distributions should have stricter regulations for risk-sharing between risk classes.*

Equivalently, using the relationship between uncertainty of correlation structure and tail distributions,

Implication 2. *Economies with more uncertain correlation between risk classes should have stricter regulations for risk-sharing.*

We also have

Implication 3. *Economies with lower interest rates should have stricter regulations for risk-sharing.*

Implication 4. *Economies with fewer risk classes should have stricter regulations for risk-sharing.*

Implication 5. *Economies with risk classes for which it takes longer to recover after a massive default should have stricter regulations for risk-sharing.*

Implication 4 may be related to the size of the economy, in that larger economies may have more risk classes and therefore, society may be more tolerant to risk-sharing across these classes. Moreover, Implication 5 may be related to the degree to which an economy is open to foreign investments in that economies that are open may be faster in recovering after a massive default, and thereby allow more risk-sharing between risk classes. In Section 4, we apply these implications to derive policy solutions for controlling the social costs created when actions by intermediaries to diversify create investments in highly correlated market portfolios.

Theorem 2 provides an asymptotic result and therefore, does not enlighten us about what the break-even number of risk classes is when the default risks, β, is significantly different from zero. For β significantly different from zero, the analysis is less tractable. Because of the noncoherency of the value at risk measure, we suspect that diversification may be less valuable for society in such cases. Indeed, we have the following sufficient condition for diversification to be optimal from society's perspective, which holds for all β below some threshold.

Theorem 3. *Consider i.i.d. asset class risks, $\xi_1,\ldots,\xi_M$, of Pareto-type with the tail index $\alpha>2$, and η as defined in (24). Define the tail probability $F(x)\stackrel{\text{def}}{=}\mathbb{P}(-\xi_1>x)$. Assume that the variance of ξ_1 is σ^2, that $F(K_0)=\beta_0$, and that for $K>K_0$,*

$$\frac{c_1}{K^\alpha}\le F(K)\le\frac{c_2}{K^\alpha},\tag{25}$$

for constants $0<c_1\le c_2$. Then, for all $\beta\le\beta_0$, risk-sharing is optimal from society's perspective, if the number of asset classes, M, satisfies

$$c_1 2^\alpha M^\alpha - C\eta M^{\alpha/2} - Mc_2\eta\alpha^\alpha\ge 0,\tag{26}$$

where $C=(2e\alpha\sigma^2)^{\alpha/2}$, and e is the constant $e=2.71828\ldots$.

In Fig. 7 we show the break-even number of risks for the Pareto-distributed risks of Section 3.1, for which $\alpha=2\gamma$. The bound is chosen so that the results are valid for all value at risk at confidence levels above 99%, i.e., for $\beta\le 1\%$. We see that, compared with Fig. 6, M needs to be significantly higher to ensure that society prefers diversification. The bound gets even worse if the VaR is at a lower confidence level, e.g., at the 95% level.

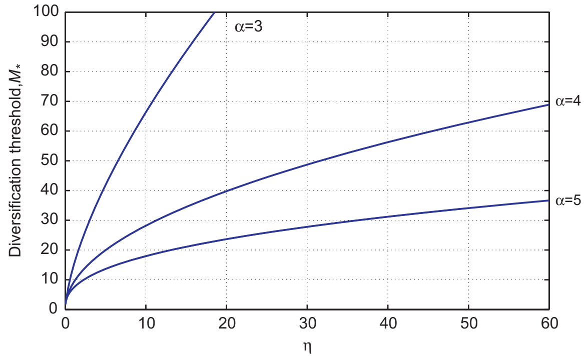

Fig. 7. Diversification threshold, M_*, i.e., sufficient condition for number of risk classes, $M>M_*$, to be large enough to guarantee that society prefers diversification, for Pareto-type risk distributions with tail indices $\alpha=3,4,5$, as a function of η. Here, $\eta=(1-\delta^{2T+1})/(1-\delta)$, δ is the per period discount factor and T is the recovery time after massive default. Results are valid for VaRs above the 99% confidence level.

As an example, assume that the average time it takes to replace a defaulted intermediary is six months if the default is individual (i.e., that one-period is six months), that the one period discount factor is $\delta=0.99$ (so that the one-year discount rate is approximately 2%), and that if there is massive intermediary default, it takes five years to restore the whole market. In this case, $\eta=(1-0.99^{10})/(1-0.99)\approx 9.6$. From Fig. 7, if the risk has a tail exponent of $\alpha=4$, there needs to be about $M=30$ risk classes to guarantee that it is optimal to diversify at a VaR at the 99% confidence level in this case. Thus, diversification may be suboptimal from society's perspective even with moderately heavy-tailed risks, as long as the number of risk classes is not large.

3.3. Extensions to general distributions

We have analyzed the simplest possible framework, with specific independent identically distributed risk classes. Here, we sketch how our results can be generalized to more general distributions.

Similar to the proof of Theorem 3, one can obtain its extension to the case of the distribution of K such that, for $K>K_0$, $c_1/K^\alpha\le F(K)\le c_2/K^\zeta$ for constants $0<c_1\le c_2$ and $\alpha\ge\zeta>1$. For such distributions, the risk-sharing outcome is optimal from society's perspective, if the number of asset classes, M is sufficiently large compared to K so that it satisfies $2^\zeta c_1 M^\zeta K^\zeta - M\eta c_2\alpha^\zeta K^\alpha - C\eta M^{\zeta-\alpha/2}K^\zeta\ge 0$, where $C=2^{\zeta-\alpha/2}(e\zeta\sigma^2)^{\alpha/2}$.

Using conditioning arguments, it is straightforward to show that the results in this paper continue to hold for (possibly dependent) scale mixtures of normals, with $w_i=V_iZ_i$, where Z_i are i.i.d. normally distributed zero-mean random variables and V_i are (possibly dependent) random variables independent of Z_i. This is a rather large class of distributions: it includes, for instance, the Student's t-distributions with arbitrary degrees of freedom (including the Cauchy distribution), the double exponential distribution, the logistic distribution, and all symmetric stable distributions.

The bounds on the number of risk classes M provided by Theorem 3 can be sharpened using extensions and refinements of probability inequality (36) provided by Lemma 1 used in the proof of the theorem. In particular, one can use probability inequalities in terms of higher moments of the independent random variables ξ (see, for instance, Theorem 1.3 and Corollaries 1.7 and 1.8 in Nagaev, 1979).

4. Potential implications for risk management and for policy makers

Capital requirements provide the most common mechanism used to control the risk of bank failure. Set at a high enough level, a simple capital-to-asset requirement can achieve any desired level of safety for an individual bank. Such capital requirements, however, impose significant costs on banks by limiting their use of debt tax shields, expanding the problem of debt overhang, and creating agency problems for the shareholders.[17] These costs have a negative impact on the overall economy because they reduce the efficiency of financial intermediation.[18] In our model, it is clear that increasing capital requirements is an imperfect tool for the regulator, since it cannot be used to specifically target negative externalities of systemic risk. In fact, there is a one-to-one relationship between VaR and capital requirements in our model—increasing the capital requirement is equivalent to decreasing the VaR.

For this reason, new proposals to control systemic risk in the banking sector recommend focused capital requirements based on each bank's specific contribution to the aggregate systemic risk. Acharya (2009), for example, advocates higher capital requirements for banks holding asset portfolios that are highly correlated with the portfolios of other banks. This follows from his model in which banks create systemic risk by choosing correlated portfolios. In a similar spirit, Adrian and Brunnermeier (2009) advocate a "CoVaR" method in which banks face higher capital requirements based on their measured contribution to the aggregate systemic risk.

Focused capital requirements will be efficient in controlling systemic risk, however, only if the source of the systemic risk is properly identified. In particular, the model in this paper creates a symmetric equilibrium in which each of the M banks is responsible for precisely $1/M$ of the systemic risk. Furthermore, systemic risk in our model arises only as a byproduct—a true negative externality—of each bank's attempt to eliminate its own idiosyncratic risk. The risks that they take on are independent, but the diversification changes the states of the world in which massive default occurs. For this reason, neither VaR constraints, nor the capital requirement plans advocated by Acharya (2009) will be effective in controlling the type of systemic risk that arises in our model. The CoVaR measure in Adrian and Brunnermeier (2009) may also be imperfect, since it does not take tail-distributions into account. Of course, systemic risk in the banking industry may reflect a variety of generating mechanisms, so we are not claiming anything like a monopoly on the proper regulatory response. But it is the case that even focused capital requirements will not be effective if the systemic risk arises because banks hedge their idiosyncratic risk by swapping into a "market" portfolio that is then held in common by all banks.

For the case developed in our model, where all banks wish to adopt the same diversified market portfolio, direct prohibitions against specific banking activities or investments in specific asset classes will be more effective than capital requirements as a mechanism to control systemic risk. For example, Volcker (2010) has recently proposed to restrict commercial banking organizations from certain proprietary and more speculative activities. While such prohibitions may seem draconian, they would apply only to activities or asset classes in which moderately heavy tails create a discrepancy between the private and public benefits of diversification. No regulatory action would be needed for asset classes with thin-tailed risks, where the banks and society both benefit from diversification, and for severely heavy-tailed risks, where the banks and society agree that diversification is not beneficial.

It is also useful to note that direct prohibitions have long existed in U.S. banking regulation. For one thing, U.S. commercial banks have long been prohibited from investing in equity shares. Even more relevant, the 1933 Glass Steagall Act forced U.S. commercial banks to divest their investment banking divisions. Subsequent legislation—specifically the 1956 Bank Holding Company Act and the Gramm-Leach-Bliley Act of 1999—provided more flexibility, by expanding the range of allowed activities for a bank holding company, although commercial banks are themselves still restricted to a "banking business." Glass-Steagall thus provides a precedent for direct prohibitions on bank activities as well as an indication that the prohibitions can be changed over time as conditions warrant.

Experience with regulating catastrophe insurance counter-party risk suggests another practical and specific regulatory approach, namely "monoline" requirements. Monoline requirements have long been successfully

[17] The debt overhang problem arises in recapitalizing a bank because the existing shareholder ownership is diluted while some of the cash inflow benefit accrues as a credit upgrade for the existing bondholders and other bank creditors. The agency problems arise because larger capital ratios provide management greater incentive to carry out risky investments that raise the expected value of compensation but may reduce expected equity returns; for further discussion, see Kashyap, Rajan, and Stein (2008). While the tax shield benefit of debt is valuable for the banking industry, it is not necessarily welfare-enhancing for society.

[18] The efficiency costs of capital requirements can be mitigated by setting the requirements in terms of contingent capital in lieu of balance sheet capital. One mechanism is based on bonds that convert to capital if bankruptcy is threatened (Flannery, 2005), but that instrument is not particularly directed to systemic risk. Kashyap, Rajan, and Stein (2008) take the contingent capital idea a step further by requiring banks to purchase "capital insurance" that provides cash to the bank if industry losses, or some comparable aggregate trigger, hits a specified threshold. This mechanism may reduce or eliminate the costs that are otherwise created by bankruptcy, but it does not eliminate the negative externality that creates the systemic risk.

imposed on insurance firms that provide coverage against default by mortgage and municipal bond borrowers; see Jaffee (2006, 2009). The monoline requirements prohibit these insurers from operating as multiline insurers that offer coverage on multiple insurance lines. The monoline restriction eliminates the possibility that large losses on the catastrophe line would bankrupt a multiline insurer, thus creating a cascade of losses for policyholders across its other insurance lines. Such monoline restrictions do create a cost in the form of the lost benefits of diversification, because a monoline insurer is unable to deploy its capital to pay claims against a portfolio of insurance risks. Nevertheless, Ibragimov, Jaffee, and Walden (2008) show that when the benefits of diversification are muted by heavy tails or other distributional features, the social benefits of controlling the systemic risk dominate the lost benefits of diversification.

As a specific example, we reference the use of credit default swaps (CDS) purchased by banks and other investors to provide protection against default on the subprime mortgage securities they held in their portfolios. The systemic problem was that the CDS protection was provided by other banks and financial service firms acting as banks, i.e., American International Group, Inc. (AIG), with the effect that a set of large banks ended up holding a very similar, albeit diversified, portfolio of subprime mortgage risks.[19] When the risks on the individual underlying mortgages proved to be highly correlated, this portfolio suffered enormous losses, creating the systemic crisis. Capital requirements actually provided the investing banks with incentive to purchase the CDS protection, so that higher capital requirements, per se, do not solve the systemic problem. Instead, there must be regulatory recognition that moderately heavy-tailed risk distributions create situations in which the social costs may exceed the private benefits of diversification.

5. Concluding remarks

The subprime financial crisis has revealed highly significant externalities through which the actions of individual intermediaries may create enormous systemic risks. The model in this paper highlights the differences between risks evaluated by individual intermediaries versus society. We develop a model in which the negative externality arises because actions to diversify that are optimal for individual intermediaries may prove to be suboptimal for society.

We show that the distributional properties of the risks are crucial: most importantly, with moderately heavy-tailed risks, the diversification actions of individual intermediaries may be suboptimal for society. We also show that when there is uncertainty about correlations between a large number of thin-tailed risks and intermediaries face value at risk constraints, this is exactly the type of risk portfolios they will choose. Also, the number of distinct asset classes in the economy, the discount rate, and the time to recover after a massive intermediary default influence the value to society of diversification. The optimal outcome from society's perspective involves less risk-sharing, but also creates a lower probability for massive intermediary collapses.

The policy implications of our model are very direct: banking regulation should be expanded in order to restrict the ability of banks to swap their loan portfolios containing idiosyncratic risk for market portfolios of loan risk. Such direct restrictions appear preferable to VaR constraints, higher capital requirements, and to CoVaR measures when the goal is to deter the massive bank defaults that may arise when banks diversify in this manner. Monoline requirements, which are already applied to insurance firms that offer insurance coverage against default on mortgages and municipal bonds, provide an example of how such direct restrictions can be implemented.

Appendix A. Proofs

Proof of Theorem 1. We first study the problem, given a fixed N and show that there exists a solution to Eqs. (9)–(12). We replace the maximum by the supremum and show that the problem has a finite solution. Of course, $\mathbf{c}=\mathbf{0}$, $k=0$, is a feasible choice that satisfies Eqs. (10)–(12), immediately,immediately leading to a lower bound of 0 for the function under the maximum sign in (9).

Now, if there is a finite upper bound on (9), then the supremum problem has a finite solution. Given the definition of Σ, it is clear that the variance of $\mathbf{c}^T\mathbf{x}$ is increasing in ρ, and since, given ρ, the distribution of the portfolio is normal, the value at risk, $\mathrm{VaR}_{1-\beta}(\mathbf{c}^T\mathbf{x})$, for a given $\mathbf{c}$ is also increasing in ρ. Therefore, the value at risk of $\mathbf{c}^T\mathbf{x}$ is greater than the value at risk for $\mathbf{c}^T\mathbf{y}$, where $\mathbf{y}$ is a vector of i.i.d. normally distributed random variables with mean 0 and variance σ^2.

Define the sets $C_{\mathbf{y}}=\{\mathbf{c}\in\mathbb{R}^N_+ : \mathrm{VaR}_{1-\beta}(\mathbf{c}^T\mathbf{y})\le K\}$, and $C_{\mathbf{x}}=\{\mathbf{c}\in\mathbb{R}^N_+ : \mathrm{VaR}_{1-\beta}(\mathbf{c}^T\mathbf{x})\le K\}$. From the previous argument, it follows that $C_{\mathbf{x}}\subset C_{\mathbf{y}}$. Now, $C_{\mathbf{y}}$ is defined by the elements $\mathbf{c}$ that satisfy $\sum_i(\mathbf{c}_i)^2\le r$, for some $0<r<\infty$, and since this is a compact set, and the objective function is continuous, the supremum of the relaxed problem (over $C_{\mathbf{y}}$) is bounded above, which obviously then is also true for the intermediary's problem. Moreover, it is clear that $C_{\mathbf{x}}$ is a closed set, and (since $C_{\mathbf{x}}\subset C_{\mathbf{y}}$), thereby compact. By continuity, it follows that the supremum to the intermediary's problem is achieved.

We will show that the p.d.f. of a uniform portfolio, $\mathbf{c}=(1/N)\mathbf{1}$, takes the form (30). Therefore, given that $K>0$, it is always possible to choose k=K, find a strictly positive ε, and choose a uniform portfolio $\mathbf{c}=(\varepsilon/N)\mathbf{1}$, such that the constraint (12) is satisfied. Clearly, for such a portfolio, (9) is strictly positive, so $\mathbf{c}=\mathbf{0}$ is not a solution to the optimization problem.

[19] It is important to note that AIG wrote its CDS contracts from its Financial Products subsidiary, which was chartered as a savings and loan association and not as an insurance firm. Indeed, AIG also owns a monoline mortgage insurer, United Guaranty, but this subsidiary was not the source of the losses that forced the government bailout.

Given that there is a strictly positive solution to Eqs. (9)–(12), we now analyze its properties. It is easy to see that (10) must be binding with $k=K$, since the program is homogeneous of degree one in k, i.e., assuming that $k<K$, $\mathbf{c}$ is a solution with $r=d\mathbf{c}^T\mathbf{1}/(1-\delta\mathbb{P}(-\mathbf{c}^T\mathbf{x}>k))>0$, then choosing capital K, and portfolio $\mathbf{c}_2=(K/k)\mathbf{c}$ will give $r'=d\mathbf{c}_2^T\mathbf{1}/(1-\delta\mathbb{P}(-\mathbf{c}_2^T\mathbf{x}>K))=(K/k)d\mathbf{c}^T\mathbf{1}/(1-\delta\mathbb{P}(-\mathbf{c}^T\mathbf{x}>k))=(K/k)r>r$, yielding a contradiction. Thus, property (c) is proved.

We next show that it is always optimal for the intermediary to choose a uniform portfolio, $\mathbf{c}=c\mathbf{1}$ for some $c>0$. Given a normally distributed risk, $y\sim N(0,\sigma^2)$ and a value at risk requirement of K_0 at the $1-\beta$ confidence level, it immediately follows that the maximum size of investment in this risk, c, and thereby the maximum premium collected (that does not break the VaR requirement) is

$$c=\frac{K_0}{\Phi^{-1}(1-\beta)}\times\frac{1}{\sigma}, \tag{27}$$

where Φ is the standard normal c.d.f.: $\Phi(x)=\int_{-\infty}^{x}e^{-t^2/2}/\sqrt{2\pi}\,dt$. From (27) it follows that any VaR-constrained portfolio manager, whose compensation is monotonically increasing in scale, $\mathbf{c}^T\mathbf{1}$—e.g., linear as assumed—who chooses among portfolios of multivariate normally distributed risks, will choose the portfolio with the lowest variance, σ.

Given ρ, standard matrix algebra yields that $\Sigma=cov(\mathbf{A}^{-1}\mathbf{x},\mathbf{A}^{-1}\mathbf{x})=(1-\rho^2)\mathbf{A}^{-1}\mathbf{A}^{-T}$. Also, for large N, standard theory of Toeplitz matrices (see, e.g., Gray, 2006) implies that $\Sigma_{ij}\to\hat{\Sigma}_{ij}=\rho^{|i-j|}$. Therefore, property (a) holds.

Conditional on ρ, to choose the portfolio with the lowest variance, the manager solves $\min_{\mathbf{q}\in S^n}\mathbf{q}^T\Sigma\mathbf{q}$, where Σ depends on ρ. We have $\Sigma^{-1}=(1/(1-\rho^2))\mathbf{A}^T\mathbf{A}$. The solution to the portfolio problem is

$$\mathbf{q}=\frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}=\frac{\mathbf{A}^T\mathbf{A}\mathbf{1}}{\mathbf{1}^T\mathbf{A}^T\mathbf{A}\mathbf{1}}=\frac{s}{sN}\mathbf{1}=\frac{1}{N}\mathbf{1}.$$

Here, the result follows, since $\mathbf{1}$ is an eigenvector to $\mathbf{A}^T\mathbf{A}$ with nonzero eigenvalue, s, since $\mathbf{A}$ is invertible. The manager thus invests uniformly, which is not surprising since the risks are symmetric. This is the optimal strategy regardless of the correlation, ρ, so even with uncertain correlations, it is always optimal to choose the uniform portfolio.

We next study the distribution of the uniform portfolio. We define $x^N=(1/\sqrt{N})\mathbf{1}_N^T\mathbf{x}_N$, where $\mathbf{x}_N$ on the right-hand side contains N elements. The c.d.f. of x^N is $F_N(y)=\mathbb{P}(x^N\le y)$. From the previous argument, it is clear that the agent's program with N risks simplifies to

$$\max_{c_N}\frac{dc_N}{1-\delta F_N\left(\frac{K}{c_N}\right)}\quad \text{s.t.}, \tag{28}$$

$$F_N\left(\frac{K}{c_N}\right)\ge 1-\beta. \tag{29}$$

Conditional on ρ, x^N is normally distributed with variance $\sigma_N^2(\rho)=((1-\rho^2)/N)\mathbf{1}^T\mathbf{A}^{-1}\mathbf{A}^{-T}\mathbf{1}$. However, since $\mathbf{A}^T\mathbf{A}$ is a symmetric positive definite matrix and $\mathbf{1}$ is an eigenvector with corresponding eigenvalue $(1-\rho)^2$, it follows that $\mathbf{1}$ is an eigenvector to $\mathbf{A}^{-1}\mathbf{A}^{-\mathrm{T}}$ with eigenvalue $1/(1-\rho)^2$. Therefore, $\sigma_N^2(\rho)=((1-\rho^2)/N)\mathbf{1}^T\mathbf{A}^{-1}\mathbf{A}^{-T}\mathbf{1}=((1-\rho^2)/N)\mathbf{1}^T(1/(1-\rho)^2)\mathbf{1}=(1+\rho)/(1-\rho)$. Using (5), we immediately obtain that the p.d.f. of x^N is

$$f_0(y)=\int_1^\infty\frac{1}{\sqrt{2\pi v}}e^{-y^2/2v}\frac{\gamma}{v^{\gamma+1}}\,dv=\frac{\gamma 2^\gamma}{\sqrt{\pi}}\times\frac{\Gamma\left(\gamma+\frac{1}{2},\frac{y^2}{2}\right)}{|y|^{1+2\gamma}}. \tag{30}$$

It is easy to show that $f_0(y)$ is decreasing in $y\in(0,\infty)$, that is, $f_0(y)$ is unimodal and symmetric (about 0): this conclusion may be obtained directly or by noting that $f_0(y)$ is a scale mixture of normal distributions and conditioning arguments (see, for instance, the proof of Theorems 5.1 and 5.2 in Ibragimov, 2009b).

We study the optimization problem (28) and (29), and rewrite it as

$$\max_z\frac{Kd}{z}\frac{1}{1-\delta F(z)}\quad\text{s.t.}, \tag{31}$$

$$F(z)\ge 1-\beta, \tag{32}$$

where $F(y)=\int_{-\infty}^{y}f_0(z)\,dz$. The constraint that β is close to zero now implies that the minimal feasible z, such that (12) is satisfied, is large.

Clearly, $(Kd/z)(1/(1-\delta F(z)))$ is positive for all z. We note that $(Kd/z)(1/(1-\delta F(z)))$ becomes arbitrarily large as $z\to 0$, so if there was no VaR constraint, it would be optimal to take on an arbitrarily large amount of risk.

The derivative with respect to z is

$$\frac{d\left[\frac{Kd}{z}\frac{1}{1-\delta F(z)}\right]}{dz}=\frac{Kd}{z^2(1-\delta F(z))^2}(\delta F(z)-1+z\delta f_0(z)).$$

Now, since $F(z)\le 1$, $\delta<1$ and $zf_0(z)\to 0$ for large z, $\delta F(z)-1+z\delta f_0(z)\le\delta-1+zf_0(z)<0$ for large z. Therefore, $d[(Kd/z)(1/(1-\delta F(z)))]/dz<0$ for large z. The feasible set for z is $[\underline{z},\infty)$, for some $\underline{z}>0$, and the VaR constraint is binding if $z=\underline{z}$ is chosen. When β is close to zero, $\underline{z}$ is large and therefore $d[(Kd/z)(1/(1-\delta F(z)))]/dz<0$ for all feasible z. Therefore, the optimum is reached at $\underline{z}$ and the VaR constraint, (32), is indeed binding. We note that this argument does not depend on the distribution, so the VaR constraint will also be binding when we study the diversified case. We have shown property (d).

This binding VaR constraint corresponds to $c=K/F^{-1}(1-\beta)$. Since there is a unique optimum, the same $c_N\equiv c$ in (28) will be chosen regardless of N, which via (30) immediately implies property (b) (with $b=1/c$).

Finally, it is easy to check that for $\gamma>1$, $Var(x^N)=\int_{-\infty}^{\infty} y^2 f_0(y)\,dy=\gamma/(\gamma-1)$, which immediately leads to property (e). We are done. □

Proof of Theorem 2. Define the tail probabilities $F(x)=\mathbb{P}(-\xi_1>x)$ and $F_M(K)=\mathbb{P}(-\sum_{m=1}^{M}\xi_m/M>K)$. From Feller (1970, pp. 275–279), it follows that since ξ_1 is of Pareto-type with $F_1(K)=((1+o(1))/K^\alpha)\ell(K)$, as $K\to+\infty$, for some slowly varying function, $\ell(K)$,

$$F_M(K)=M\left(\frac{1+o(1)}{MK}\right)^\alpha \ell(MK), \tag{33}$$

as $K\to+\infty$. For risk-sharing to be optimal from an intermediary's perspective, (18) is therefore equivalent to

$$M\left(\frac{1+o(1)}{MK}\right)^\alpha \ell(MK)<\frac{1+o(1)}{K^\alpha}\ell(K),$$

which implies

$$M^{1-\alpha}<1+o(1),$$

as $K\to+\infty$, which, for large K, holds if and only if $\alpha<1$. We have proved the first part of the theorem.

For the second part, since $q=1-F(K)$ and $q_M=1-F_M(K)$, (21) and (33) together imply that risk-sharing from society's perspective is optimal if and only if

$$(1+\delta-\delta F_M(K))(1+\lambda F(K)^M)>(1+\delta-\delta F(K))(1+\lambda F_M(K)),$$

as $K\to+\infty$, which is satisfied if and only if

$$\delta F(K)>\delta F_M(K)+\lambda(1+\delta)F_M(K)=\delta\eta F_M(K).$$

This is equivalent to

$$\frac{1+o(1)}{K^\alpha}\ell(K)>\eta M\left(\frac{1+o(1)}{MK}\right)^\alpha \ell(MK),$$

as $K\to+\infty$, which, in turn, is satisfied if and only if, for large K,

$$M^{\alpha-1}>\eta. \tag{34}$$

Since $\eta\ge 1$, (34) is never satisfied for $M>1$ and $\alpha<1$, so risk-sharing is never optimal from society's perspective when $\alpha<1$. For $\alpha>1$, on the other hand, (34) is satisfied if and only if

$$M>\eta^{1/(\alpha-1)}. \tag{35}$$

We have proved the second part of the theorem. □

Proof of Theorem 3. Define $F_M(K)=\mathbb{P}((-\sum_{m=1}^{M}\xi_m/M)>K)$. We first show that $\eta F_M(K)<F(K)$ is a sufficient condition for (21). We have

$$\begin{aligned}\eta F_M<F &\Leftrightarrow \delta F_M+(1+\delta)\lambda F_M<\delta F\\ &\Rightarrow 1+\delta-\delta F+(1+\delta)\lambda F_M-\delta F\lambda F_M<1+\delta-\delta F_M\\ &\Leftrightarrow (1+\delta-\delta F)(1+\lambda F_M)<1+\delta-\delta F_M\\ &\Leftrightarrow 1+\delta-\delta F<\frac{1+\delta-\delta F_M}{1+\lambda F_M}\\ &\Leftrightarrow 1+\delta q<\frac{1+\delta q_M}{1+\lambda(1-q_M)}\\ &\Rightarrow \frac{1+\delta q}{1+\lambda(1-q)^M}<\frac{1+\delta q_M}{1+\lambda(1-q_M)}.\end{aligned}$$

To show that $\eta F_M(K)<F(K)$, we use the following lemma.

Lemma 1 (Nagaev, 1979, *Corollary 1.11*). *For independent random variables X_i, with $E[X_i]=0$ and $\sum_{i=1}^{M}E[X_i^2]=B_M$, it is the case that*

$$\mathbb{P}\left(\sum_{i=1}^{M}X_i>x\right)\le\sum_{i=1}^{M}\mathbb{P}(X_i>y)+e^{x/y}\left(\frac{B_M}{xy}\right)^{x/y},\quad x,y>0. \tag{36}$$

Using (36) for the ξ–risks, for which $B_M=M\sigma^2$, by choosing $x=MK$ and $y=2MK/\alpha$, we get

$$F_M(K)\le MF\left(\frac{2MK}{\alpha}\right)+e^{\alpha/2}\sigma^\alpha\left(\frac{2M}{\alpha}\right)^{-\alpha/2}K^{-\alpha}.$$

From (25), it then follows that, for $K>K_0$,

$$F_M(K)\le Mc_2\alpha^\alpha(2MK)^{-\alpha}+e^{\alpha/2}\sigma^\alpha\left(\frac{2M}{\alpha}\right)^{-\alpha/2}K^{-\alpha},$$

and since $F(K)\ge c_1K^{-\alpha}$, it follows that the condition

$$\eta\left(Mc_2\alpha^\alpha(2MK)^{-\alpha}+e^{\alpha/2}\sigma^\alpha\left(\frac{2M}{\alpha}\right)^{-\alpha/2}K^{-\alpha}\right)\le c_1K^{-\alpha} \tag{37}$$

implies $\eta F_M(K)<F(K)$. Now, by multiplying (37) by $(2MK)^\alpha$, we get

$$Mc_2\eta\alpha^\alpha+\eta e^{\alpha/2}\sigma^\alpha(2M\alpha)^{\alpha/2}\le c_1 2^\alpha M^\alpha,$$

i.e.,

$$c_1 2^\alpha M^\alpha-C\eta M^{\alpha/2}-M\eta c_2\alpha^\alpha\ge 0.$$

We are done. □

List of variables

M	number of risk classes
N	number of risks within each risk class
$x_n^{t,m}$	individual risk n, within risk class m at time t
ρ	correlation between risks within risk class (for large N)
$\mathbf{x}$	vector of individual risks in first risk class
$\mathbf{c}$	vector of investments of intermediary (in first risk class)
c	total size of risk investment, $c=\mathbf{c}^T\mathbf{1}$
ξ,ξ_i	risk portfolio invested in by intermediary (in risk class i), $\xi=\mathbf{c}^T\mathbf{x}$
δ	one period discount factor
K	capital reserved by intermediary
β	value at risk at $1-\beta$ confidence level may not be higher that K, $\mathrm{VaR}_{1-\beta}(\xi)\le K$
Q	outcome of intermediary's option to default
M_*	diversification threshold, above which society prefers risk-sharing
T	recovery time after massive default
d	premium intermediary collects per unit risk of investment
γ	parameter governing uncertainty of correlations between risks in same risk class
α	tail distribution of risk class (ξ). $\alpha=2\gamma$
f	probability distribution function of portfolio risk, ξ

F cumulative loss distribution of portfolio risk, $F(x) = \mathbb{P}(-\xi > x)$
q chance for individual intermediary of not defaulting when risks are separated $q = 1 - F(K)$
F_M cumulative loss distribution of fully diversified portfolio $F_M(x) = \mathbb{P}(-\sum_i \xi_i / M > x)$
q_M chance for intermediaries of not defaulting when risks are shared $q_M = 1 - F_M(K)$
λ society's trade off between risk-sharing and separated outcomes (Eq. (22))
η $\eta = M_*^{\alpha-1}$, $\alpha > 1$
V^B value of one intermediary if separated
V^S value of one intermediary if risk-sharing
V_M^B value to society if separated
V_M^S value to society if risk-sharing

References

Acharya, V., 2009. A theory of systemic risk and design of prudent bank regulation. Journal of Financial Stability 5, 224–255.

Adrian, T., Brunnermeier, M., 2009. CoVar. Unpublished working paper. Princeton University.

Allen, F., Gale, D., 2000. Financial contagion. Journal of Political Economy 108, 1–33.

Andrews, D.W.K., 2005. Cross-section regression with common shocks. Econometrica 73, 1551–1585.

Bernanke, B.S., 1983. Nonmonetary effects of the financial crisis in the propagation of the great depression. American Economic Review 73, 257–276.

Brunnermeier, M., 2009. Deciphering the liquidity and credit crunch 2007–2008. Journal of Economic Perspectives 23, 77–100.

Brunnermeier, M., Pedersen, L.H., 2009. Market liquidity and funding liquidity. Review of Financial Studies 22, 2201–2238.

Caballero, R., Krishnamurthy, A., 2008. Collective risk management in a flight to quality episode. Journal of Finance 63, 2195–2230.

Conley, T.G., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics 92, 1–45.

Diamond, D., Dybvig, P., 1983. Bank runs, deposit insurance, and liquidity. Journal of Political Economy 91, 401–419.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, New York.

Fama, E., 1965. Portfolio analysis in a stable Paretian market. Management Science 11, 404–419.

Feller, W., 1970. An Introduction to Probability Theory and its Applications, second ed., vol. 2. Wiley, New York.

Flannery, M., 2005. No pain, no gain: effecting market discipline via 'reverse convertible debentures'. In: Scott, H. (Ed.), Capital Adequacy Beyond Basel: Banking, Securities, and Insurance. Oxford University Press, Oxford, pp. 171–197 (Chapter 5).

Froot, K.A., Scharfstein, D., Stein, J., 1993. Risk management: coordinating corporate investment and financing policies. Journal of Finance 48, 1629–1658.

Froot, K.A., Stein, J., 1998. Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach. Journal of Financial Economics 47, 55–82.

Gabaix, X., 2009. Power laws in economics and finance. Annual Review of Economics 1, 255–293.

Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, H.E., 2006. Institutional investors and stock market volatility. Quarterly Journal of Economics 121, 461–504.

Gordy, M.B., 2000. Credit VaR and risk-bucket capital rules: a reconciliation. Federal Reserve Bank of Chicago May, 406–417.

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

Gray, R.M., 2006. Toeplitz and circulant matrices: a review. Foundations and Trends in Communications and Information Theory 2, 155–239.

Guillaume, D., Dacorogna, M., Davé, M., Müller, U., Olsen, R., Pictet, O., 1997. From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets. Finance and Stochastics 1, 95–129.

Horn, R.A., Johnson, C.R., 1990. Matrix Analysis. Cambridge University Press, Cambridge (Corrected reprint of the 1985 original).

Ibragimov, R., 2009a. Heavy-tailed densities. In: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics Online. Palgrave Macmillan.

Ibragimov, R., 2009b. Portfolio diversification and value at risk under thick-tailedness. Quantitative Finance 9, 565–580.

Ibragimov, R., Jaffee, D., Walden, J., 2008. Insurance equilibrium with monoline and multiline insurers. Unpublished working paper, University of California at Berkeley.

Ibragimov, R., Jaffee, D., Walden, J., 2009. Nondiversification traps in catastophe insurance markets. The Review of Financial Studies 22, 959–993.

Ibragimov, R., Jaffee, D., Walden, J., 2010. Pricing and capital allocation for multiline insurance firms. Journal of Risk and Insurance 77 (3), 551–578.

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

Jaffee, D., 2006. Monoline restrictions, with applications to mortgage insurance and title insurance. Review of Industrial Organization 28, 88–108.

Jaffee, D., 2009. Monoline regulations to control the systemic risk created by investment banks and GSEs. The B.E. Journal of Economic Analysis & Policy 9. Available at: 〈http://www.bepress.com/bejeap/vol9/iss3/art17〉.

Jansen, D.W., de Vries, C.G., 1991. On the frequency of large stock returns: putting booms and busts into perspective. Review of Economics and Statistics 73, 18–32.

Kashyap, A., Rajan, R., Stein, J., 2008. Rethinking capital regulation. Presented at Federal Reserve Bank of Kansas City Symposium. Available at: 〈http://www.kc.frb.org/publicat/sympos/2008/KashyapRajanStein.03.12.09.pdf〉.

Kodes, L., Pritsker, M., 2002. A rational expectations model of financial contagion. Journal of Finance 57, 769–799.

Kyle, A., Xiong, W., 2001. Contagion as a wealth effect. Journal of Finance 56, 1401–1440.

Loretan, M., Phillips, P.C.B., 1994. Testing the covariance stationarity of heavy-tailed time series. Journal of Empirical Finance 1, 211–248.

Lux, T., 1998. The socio-economic dynamics of speculative markets: interacting agents, chaos and the fat tails of return distributions. Journal of Economic Behavior & Organization 33, 143–165.

Mandelbrot, B., 1997. Fractals and Scaling in Finance. In: Discontinuity Concentration Risk. Springer-Verlag, New York.

Nagaev, S.V., 1979. Large deviations of sums of independent random variables. Annals of Probability 7, 745–789.

Rachev, S.T., Menn, C., Fabozzi, F.J., 2005. Fat-tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing. Wiley, Hoboken, NJ.

Rochet, J.C., Tirole, J., 1996. Interbank lending and systemic risk. Journal of Money, Credit, and Banking 28, 762–773.

Ross, S.A., 1976. A note on a paradox in portfolio theory. Unpublished Mimeo, University of Pennsylvania.

Samuelson, P.A., 1967. General proof that diversification pays. The Journal of Financial and Quantitative Analysis 2, 1–13.

Shaffer, S., 1994. Pooling intensifies joint failure risk. Research in Financial Services 6, 249–280.

Volcker, P.A., 2010. Statement for the committee on banking, housing, and urban affairs of the United States Senate, February 2, 2010. Available at: 〈http:banking.senate.gov/public/index.cfm?FuseAction=Files.View&FileStore_id=ec787c56-dbd2-4498-bbbd-ddd23b58c1c4〉.

Wagner, W., 2010. Diversification at financial institutions and systemic crises. Journal of Financial Intermediation 19, 373–386.