The Quantification of Operational Risk *

Markus Leippold†

University of Zürich

Paolo Vanini‡

University of Southern Switzerland and Zürcher Kantonalbank

This Version: November 3, 2003

*The authors thank the seminar participants at FINRISK Zurich, RISK Italia, IBM Research Lab (Switzerland), and the International Workshop on Risk and Regulation at the Collegium Budapest. In particular, we want to thank Claudio Albanese, Ulrich Anders, Giovanni Barone-Adesi, Giorgos Cheliotis, Michel Dacorogna, Hans Geiger, Chris Kenyon, and René Stulz for their comments. The authors thank Silvan Ebnöther for research assistance and many fruitful discussions. Markus Leippold and Paolo Vanini acknowledge the financial support of the Swiss National Science Foundation (NCCR FINRISK).

†**Correspondence Information:** Markus Leippold, Plattenstrasse 14, University of Zürich, Zürich, Switzerland, tel: +41 (01) 634 3962, mailto:leippold@isb.unizh.ch

‡**Correspondence Information:** Paolo Vanini, Via Buffi 15, University of Southern Switzerland, Lugano, Switzerland, tel: +41 (01) 292 70 39, mailto:paolo.vanini@lu.unisi.ch

The Quantification of Operational Risk

Abstract

We examine the quantification of operational risk for banks. We adopt a financial-economics approach and interpret operational risk management as a means of optimizing the profitability of an institution along its value chain. We start by defining operational risk and then propose a framework to model risk mitigation through the bank's value chain over time. Using analytical and numerical methods, we obtain answers concerning capital allocation, network stability, risk figures, and diversification issues. Interpreting the results shows that the usual intuition gained from market and credit risk does not apply to the quantification of operational risk.

JEL Classification Codes: C19, C69, G18, G21.

Key Words: Operational Risk Management, Stochastic Systems, Diversification, Profitability

1 Introduction

In this paper we are concerned with the characterization and quantification of operational risk for banks. Basel II gave rise to vigorous and recurring discussions on the regulation of operational risk that have spurred great interest by both academics and practitioners. In its consultative document on the New Basel Capital Accord, the Basel Committee for Banking Supervision continues its drive to increase stability of financial markets in the realms of market and credit risk, and, most recently, operational risk. However, in this paper we do not discuss how to regulate operational risk. Instead, we consider operational risk from a purely economic and business point of view.

We define operational risk as the risk a bank faces in producing goods and services for its clients. In that respect, operational risk plays a key role for the bank's profitability. When measuring and managing operational risk, some major questions emerge for the bank's management. These management concerns fall into three main categories.

1. Definition of operational risk:

   - Can we find an appropriate definition for operational risk and, in the same spirit as for market and credit risk, define assets and portfolios of operational risk?
   - Can we make the definition of operational risk "operational"? In particular, does the definition allow for a calibration to institutional characteristics?

2. Measuring operational risk:

   - Given the definition of operational risk, can we obtain quantitative risk figures?
   - Can these risk figures be interpreted in such a way that they can build the basis for management decision? Is it possible to show the trade-off between operational risk and costs, such as prevention, investment, and maintenance costs?

3. Managing operational risk:

   - Can we find a reasonable notion of diversification for operational-risk assets?
   - Can we evaluate and quantify the benefits from altering value chains and work flow structures?

This paper provides answers to these questions. To do so, we adopt a work flow view of operational risk.

Our model setup belongs to the class of the so-called "functional dependence" models. This model class allows for two types of dependence structure, the standard stochastic dependence among the risk factors and a topological dependence. The latter dependence structure is due to the system architecture and organization of the different work steps that produce the firm's outputs. Therefore, compared to market or credit risk, by buying and selling operational-risk assets the functional dependence structure complicates the risk transfer considerably. Operational-risk assets are topologically linked. They do not exist as independent objects.

The functional dependency has far-reaching consequences for possible outsourcing strategies. From a cost and risk point of view, it might be more efficient to outsource the whole support infrastructure than to focus on partial outsourcing activities only. Some recent examples include large institutions, e.g., Deutsche Bank and J.P. Morgan Chase in the financial sector, and ABB Ltd. in the nonfinancial sector.

Our approach is somewhat like the literature on IT networks. However, our approach is more general. Our work flows need not be related only to IT networks. Furthermore, many of the contributions on IT networks develop models of very low granularity levels. Our approach is primarily intended to generate information for the firm's management. Therefore, our discussion is complementary in the following respects. First, because the pressure to maintain competitiveness and the new regulatory requirements force banks to implement a firm-wide operational risk framework. Second, in making their decisions, managers are not only interested in risk and performance, but also in the costs related to operational risk events and the efficiency of counter measures. To include these aspects in our model, we use an approach that builds on arguments from financial economics.

The topological structure of our approach is similar to some credit risk models that examine contagion between creditors (see, e.g., Egloff, Leippold and Vanini (2003), and Giesecke and Weber (2002)). In the context of operational risk, such a model was first proposed by Kühn and Neu (2003). Their model derives an analogy to models of disordered phenomena of statistical mechanics. We extend their approach in several directions: First, we take into account explicit dependence structures as modeled by graphs. Second, we include costs that arise in case of operational risk

events. Third, we consider path-dependent counter measures.

We organize the paper as follows. In Section 2 we develop and present appropriate definitions of operational risk and related quantities. Section 3 presents the dynamic setup of our model and Section 4 analyzes its properties. We address the calibration of our model in Section 5. Section 6 presents a numerical exploration of our model. Section 7 concludes.

2 An Operational Definition of Operational Risk

Our model focuses on operational activities that are essential to the bank's business model. We leave aside the nonessential work flows that do not contribute to growth and profitability. To select the key operational activities, we group such activities in value chains or work flows, which we see as integrated client-to-client networks.

To develop and apply a formal model for operational risk, we consider a generic "risk management" value chain. Value in risk management is generated by information flows between trading, sales, and risk management units. Figure 1 illustrates a possible structure from a function or banking unit point of view.

In Figure 1, the top level defines six different functions, the sales functions in work flows 1 and 2, the trading unit (core work flow 3), and the risk management functions in the work flows 4 to 6. The risk management function splits up into the limit management work flows 4 and 5 and a risk measurement work flow 6. The figure shows the information flow, indicated by arrows, between the different functions or units.

A convenient mathematical concept that captures the structure illustrated in Figure 1 is graph theory. A graph $\mathcal{G}$ is a collection of nodes and of edges connecting the nodes.[1] At the nodes in Figure 1, employees make decisions or carry out a work step. The graph's edges connect the nodes, and therefore serve as information interfaces. Since information is directed, we prefer to model work flows as a collection of directed graphs.

The whole information flow in Figure 1 has several starting and end points. The end points represent product and services. For example, the different reports or the trades shown in Figure

[1] Sometimes, we also use the term "states" for nodes.

1. Thus, when modeling business activities, a graph $\mathcal{G}_j$ consists of paths linking input nodes with output nodes, where the output nodes are linked to the bank's internal and external clients and investors.

The set of the graphs $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n\}$ constitutes a network $\mathcal{N}$. However, the business work flow view in Figure 1 is not yet comprehensive, because we have not yet looked at the associated IT systems and the information flow between them. In Figure 2 we illustrate a possible IT system network for trading, sales, and risk management. The information flow occurs on different layers, on a business, an IT system, a database, and a controlling layer. The union of all nodes and edges over all layers defines the network.

An operational risk model must be flexible enough to account for the complexity due to the different layers, and it should also allow us to consider subsets of a network across different layers. For example, if we are interested in the IT systems defined by the nodes $n_{10}, n_{11}, n_{12}, n_{14}, n_{15}, n_{18}, n_{19}$ in Figure 2, we call the corresponding graph the IT risk management graph. This graph is only a subset of the whole trading, sales, and risk management IT systems network, which itself is a subset of the network defined over all layers.

In addition, flexibility is not only required in the aggregation and disaggregation on and between the different layers, but also in the granularity of modeling information flow. According to the needs of the institution, information modeling can vary from a simple, abstract two-state "running/down" modeling to a more refined modeling.

We complete the geometric skeleton structure of a work flow illustrated in Figure 2 by attaching operational risk to the elements of the graphs. Randomness occurs at two places, at the nodes and at the edges of the graph. The randomness at the nodes is due to risk factors such as, e.g., system failure, theft, fraud, model risk, and human error. Although a list of risk factors can comprise several dozen entries, not all factors matter at different nodes or edges in the network. For example, in the core work flow risk measurement on the business layer view (see Figure 1), we expect human error and model risk to be significant risk factors, but we can neglect fraud. Risk factors on the edges typically impact the information flow, e.g., in form of capacity problems or performance instabilities.

Apart from the risk factors that measure the losses from an operational risk event, other at-

tributes must be associated to the nodes. For management purposes, cost factors play a predominant role. These factors can be split up into two parts, a nonrandom component $d_{i,j}$ for the fixed costs, and a stochastic cost component $a_{i,j}(t)$ depending on the frequency and severity of operational risk events. We can adapt this latter component, which we call "counter measure," to the information generated by the risk factors. We collect the two cost components of a single graph $\mathcal{G}_j$ in the vectors $\mathbf{a}_j$ and $\mathbf{d}_j$.

Since the nodes represent systems or an individual's activities, they can be either in a running, a down, or an intermediate state. Therefore, we model the status of the nodes by stochastic processes $n_{i,j}(t)$, which take values in $[0,1]$. The value zero corresponds to an optimally operating system. If $n_{i,j}(t) = 1$, a complete breakdown of the system occurs at time t. Hence, the state $n_{i,j}(t)$ depends on an m-dimensional vector $X_{i,j}(t) = \{X^1_{i,j}(t), X^2_{i,j}(t), \ldots, X^m_{i,j}(t)\}$ and models the functioning or malfunctioning of node i in graph j. We collect by $\mathbf{X}$ the set of all risk factors on the network.

So far, risk affects only the nodes, but the edges can also suffer from risky events. Just as we did for the nodes, we define $Y_{i,j}(t) = \{Y^1_{i,j}(t), Y^2_{i,j}(t), \ldots, Y^p_{i,j}(t)\}$ as the risk factors that affect a specific outgoing edge of the node $n_{i,j}$. The associated performance functions $q^r_{i,j}, r = 1, 2, \ldots, p$ attribute to the risk factors $Y_{i,j}(t)$ a value in $[0,1]$, conditional on the state $n_{i,j}(t)$. We write $\mathbf{Y}$ for the set of all risk factors and $\mathbf{q}, \mathbf{n}$ for the set of all performance and node work flows, respectively.

Using the formal construct above, we can now rigorously define the terms *a)* operational-risk asset, *b)* operational-risk portfolio and *c)* operational risk.

Definition 1. *Given a probability space* $(\Omega, \mathbf{F}, \mathbb{P})$*,* $\mathcal{G}_j$ *a directed graph and* $\mathbf{X}, \mathbf{Y}, \mathbf{a}, \mathbf{d}, \mathbf{q}, \mathbf{n}$*, and the filtration* $\mathbf{F}$ *generated by the work flows* $\mathbf{X} \otimes \mathbf{Y}$*, we define:*

1. *The tuplet*

$$\mathcal{A}_j = (\mathcal{G}_j, \mathbf{X}_j, \mathbf{Y}_j, \mathbf{a}_j, \mathbf{d}_j, \mathbf{q}_j, \mathbf{n}_j),$$

   defines an operational-risk asset for business activity j*.*

2. *The operational-risk portfolio for* k *operational-risk assets distributed over* m *graphs is given by*

$$\mathcal{P} = (\mathcal{N}, \bigotimes_{j=1}^{k} \mathbf{X}_j, \bigotimes_{j=1}^{k} \mathbf{Y}_j, \mathbf{a}, \mathbf{d}, \bigcup_{j=1}^{k} \mathbf{q}_j, \bigcup_{j=1}^{k} \mathbf{n}_j),$$

where $\mathcal{N} = \bigcup_{i=1}^{m} \mathcal{G}_i$.

3. *Operational risk for an operational-risk asset* j *is the probability distribution of* $\mathbf{X}_j \otimes \mathbf{Y}_j$ *on* $\mathcal{G}_j$. *Accordingly, the operational risk of an operational-risk portfolio is the probability distribution of* $\left(\bigotimes_{j=1}^{k} \mathbf{X}_j\right) \otimes \left(\bigotimes_{j=1}^{k} \mathbf{Y}_j\right)$ *on* $\mathcal{N}$.

In other words, operational risk is the internal probability distribution on a bank's activities to collect, generate, transfer, and transmit information to provide goods and services on a client-to-client basis. This distribution is adjusted by the producing costs and the counter-measure costs that originate from operational-risk events.

Definition 1 suggests that operational-risk assets and portfolios are more complicated objects than are their market-risk and credit-risk counterparts. However, the generality of Definition 1 allows fitting in the definitions for market-risk and credit-risk assets.

In the traditional approach to market risk, the network would degenerate to a point work flow that could be identified as the price of financial instruments. The states $\mathbf{n}_i = n_i$ are the price functionals or positions, and X_i is the factor causing random price changes. Thus, a market-risk asset is given by $\mathcal{A}_i^M = (X_i, n_i)$. Since market-risk assets possess only a stochastic dependence and, moreover, market-risk portfolios are additive at any given time, a portfolio is given as $\mathcal{P}^M = \left(\bigotimes_{j=1}^{k} \mathbf{X}_j, \sum_{j=1}^{k} n_j\right)$.

The same structure applies to characterize credit-risk models. Recently, there has been an effort to take into account possible contagion and infection effects. Such credit models implicitly encompass a graph structure representing dependencies and define a credit-risk asset as $\mathcal{A}_i^C = (\mathcal{G}_i, \mathbf{X}_i, \mathbf{Y}_i, \mathbf{q}_i, \mathbf{n}_i)$.

However, neither market-risk nor credit-risk assets depend on the cost components $\mathbf{a}_i$ and $\mathbf{d}_i$, which highlights the importance of taking costs into account when modeling operational risk. As a result, some of the techniques that worked so well for market and credit risks become inadequate and, in some cases, unreliable for applying to operational risk.

3 The Model

Definition 1 allows us to establish a static snapshot of operational risk. However, we still lack a model framework for the network dynamics. Without loss of generality, we consider those IT systems layer and operational-risk events that occur only at the nodes. We assume that the edges possess no riskiness.

Given a state $n_{i,j}(t)$, we look for a description of its transition to the state $n_{i,j}(t+\Delta)$. This transition depends on the risk factors that affect the node under consideration, and on the states of the preceding nodes. Therefore, we must consider both a standard stochastic dependence structure and topological dependence. Thus, we introduce an auxiliary function $h_{i,j}(t)$ for each node, which we call the support function. This function turns out to be useful for formulating the dynamics of the operational-risk assets.

3.1 The Support Function

The support function $h_{i,j}(t)$ can comprise, e.g., human resources and the labor force, hardware and software, inputs from other work flows, and information inputs. A decrease in support can either be caused by an exogenous risk event or it can be caused by some failure of the neighboring nodes. The former is a pure stochastic effect. The latter is induced by the topological dependency of operational-risk assets.

For a given monotone and decreasing function f, the support functions can be uniquely linked to the states by setting

$$n_{i,j}(t) = f(h_{i,j}(t)). \tag{1}$$

We say that an operational-risk event occurs whenever the support $h_{ij}(t)$ falls below a critical threshold level $\bar{h}_{ij}$. Thus, an increase in the support for a specific node decreases the probability of a breakdown.

Possible specifications of f encompass, e.g., the step functions such as those in Kühn and Neu (2003), where a binary model "system running" and "system down" is analyzed. Here, we prefer a

specification of the form

$$f(x) = \frac{\text{erfc(x)}}{2}, \tag{2}$$

with erfc$(\cdot)$ as the complementary error function. The specification in equation (2) allows us to consider intermediate states. This more refined function f measures the quality of the node or system. In the sequel, we will work with a support dynamic[2]

$$dh_i = s_i(h_i)\left(b_i(h_i) - (w \circ f(h))_i\right) dt - \sigma_i dW_i - \varsigma_i dZ \ , \ \forall i \in \mathcal{G}_j, \tag{3}$$

where:

- $s_i(h_i)$: This work flow defines how fast a deviation of $h_i(t)$ from the expected long-run equilibrium value is set off. Hence, we call it the speed function. In the simplest case, s_i is a constant. However, such an assumption is hardly in line with real-world situations, since deviations of the network status from a prespecified target value are usually counteracted to ensure an acceptable risk level. Such counter measures are costly, and the work flow s_i can be split into the two cost components, the fix costs d_i and the variable costs a_i, according to

$$s_i(h_i) = d_i + a_i(h_i). \tag{4}$$

  Since a system triggers stronger counter measures when it is down for more than one period, we introduce a path-dependent structure for the counter measures $a_i(h_i)$. A possible functional form for $a_i(h_i)$ taking into account such a path-dependency is

$$a_i(h_i)(t) = \psi \ \mathbb{1}_{\{h_i(s) < \bar{h}_i\}} \left(\int_0^t n_i(s)\, ds \right)^{\gamma}, \tag{5}$$

  where $\mathbb{1}$ is the indicator function. Therefore, counter measures are only activated when the support function falls below a critical value $\bar{h}_i$. If this does not happen, a constant maintenance support for the node i is guaranteed. The parameter $\gamma \geq 0$ in equation (5) determines the curvature of $a_i(h_i)$. It measures how the counter measures are intensified in case of an prolonged operational-risk event. The parameter ψ measures the effectiveness of the counter

[2]For notational convenience, we drop the graph index. This should not be confusing.

measures.

In equation (5), when an operational-risk event occurs, the counter measures induce an increase in the drift of the support function. This increase is intensified when effectiveness is high and n_i is becoming large, i.e., when a total breakdown of the node i becomes more likely.

- $b_i(h_i)$: This term is part of the long-term mean of state n_i and captures the threshold in service level agreements (SLA). SLAs are used to restrict the employees responsibility in a network to a local neighborhood. That is, employees need only be concerned with the SLAs to which they contracted. Therefore, the function $b_i(h_i)$ is a simple but efficient method to implement the system engineers' incentives and to mimic the real-life reduction of network complexity.

- $(w \circ f(h))_i$: This expression quantifies the topological dependency within the network and is defined by

$$(w \circ f(h))_i(t) \quad = \quad \sum_{m=1}^{N_1(i)} w_{i,m} f(h_m) + \sum_{m=1,k=1}^{N_2(i)} w_{i,m} w_{i,k} f(h_m) f(h_k) + \dots \ . \tag{6}$$

  The constants w describe the coupling strength among the different interacting systems. The first summation in equation (6) describes the support levels of all input nodes, where information can reach system i in one and only one way. For system i, there are $N_1(i)$ of such systems. The second summation term in equation (6) models the case where systems are mirrored. Only if both input systems in the double sum are in a down state will the system i be affected.

- $\sigma_i dW_i$: The noise W_i captures the different operational risk factors. Its impact on the support function of system i is scaled by σ_i.

- $\varsigma_i dZ$: This term reflects an external risk factor, which affects all systems of the network. Not all systems are equally affected by external events. Therefore, we scale the factor by a node specific constant ς_i. We assume that $\mathbb{E}(dW_i dZ) = \rho_i$.

3.2 Cost and Loss Functions

There is a subtle difference between costs and losses. As argued in Section 2, costs are related to prevention and counter measures. Therefore, we subsume to these costs all outlays that are needed to recover a network which suffered from the operational-risk event. In contrast, losses are linked to the operational-risk event through the value-generating work flow of system i. The system no longer fulfills its assignments of business activities and impairs the functionality of the value chain and the value-generating work flow.

To parameterize the cost and the loss function, we propose the following approach. From Figure 3, we see that an operational-risk event is triggered as soon as the support function falls below its critical value $\bar{h}_i$ (point A). The operational-risk event can be characterized by its duration and its severity. For a node i, the duration D_{ik} of the operational-risk event k is given by the time interval defined by two stopping times $[\underline{\tau}_{ik}, \overline{\tau}_{ik}]$. We define the severity S_{ik} as $S_{ik} = \inf_{t \in D_{ik}} h_{ik}(t)$.

We use the duration and the severity to parameterize the cost and loss function for an operational-risk event. A possible specification for the cost function is

$$C_{ik} = C_{ik}(S_{ik}, D_{ik}) = \int_{D_{ik}} g(|h_{ik}(s)|)\ ds, \tag{7}$$

for some positive and increasing function g. In Figure 3, the shadowed area C_{ik} represents the cost function.

As argued above, the loss function ℓ_{ik} requires the specification of a value-generating work flow for node i. These work flows are determined by the value chains. A value chain can be represented as a path p through the network under consideration. We assume that the contribution of node i to the value-generating work flow of path p is given by a positive stochastic variable v_{ip}. With P paths running through system i, the loss function reads

$$\ell_{ik} = \ell_{ik}\left(S_{ik}, D_{ik}, \sum_{P} v_{ip}\right). \tag{8}$$

We can define the loss function as an expected loss difference

$$\ell_{ik}\left(S_{ik}, D_{ik}, \sum_{P} v_{ip}\right) = \mu_{ip}(v_{ip}, 0, 0) - \mu_{ip}(v_{ip}, D_{ik}, S_{ik}), \tag{9}$$

where $\mu_{ip}(\cdot)$ is the mean of the pth value work flow acting on node i. The specification in equation (9) allows the possibility of modeling reputational losses through the duration D_{ik}. Such losses can arise when a system i is down and an important service for key clients cannot be provided over an extended period.

4 Analytical Results

4.1 Prioritization

The dynamic specification of the support function in equation (3) results in a coupled system of nonlinear stochastic differential equations. Nevertheless, using analytical methods, we can explore the effect of prioritizing the network through the allocation of resources by the bank's management, and the effect of prioritization on the network's stability.

To emphasize the role of the topological dependency, we leave the stochastic dependency aside. Suppose that the bank's management aims to maintain a uniform service level on the whole network or portfolio of operational-risk assets by allocating resources in such a way that the infection rates between the different nodes is the same for all nodes at all times. In other words, we require the matrix of coupling constants $W := (w_{ij})$ in equation (3) to be symmetric.

Proposition 2. *Suppose that $a_i(h_i)/f'(h_i) > 0$ is constant for all i and h, the statistical risk is zero, and f is at least twice continuously differentiable. If the matrix W is symmetric, the network is asymptotically stable.*

Proposition 2 holds for arbitrary differentiable counter measure functions $a_i(h_i)$, threshold level functions $b_i(h_i)$, and functions f. Therefore, Proposition 2 indicates some strong policy recommendations. Once management has committed itself to a uniform service level on the portfolio, the nodes will evolve to their stationary equilibrium state. Unstable or even a chaotic portfolio

behavior is not possible in this uniform service level approach. Instead, if management prioritizes the allocation of resources, limit cycles can occur. Put differently, since the matrix W becomes nonsymmetric, the portfolio can move on isolated and closed trajectories where stationary points are no longer feasible.[3]

As an example, consider two systems A and B. If management decides that A should infect B much less than B should affect A,(because B is a much more sensitive system), then portfolios based on such a prioritizing structure can exhibit limit-cycle behavior. Therefore, given a specific topological dependency structure, management faces a trade-off between prioritizing the portfolio's service levels and the possibility of network instability. To maintain stability in the above sense, huge resources are necessary. But since resources are scarce, there is always a tendency to prioritize, and indeed, such practice is widely observed in the industry.

To the topological dependencies of the above arguments, we now add statistical dependencies. With the nonlinear dynamics in equation (3), a general analytical discussion is not feasible. Therefore, we linearize the system in (3) by choosing

$$a_i(h_i) = a_i \ , \ b_i(h_i) = b_i h_i \ , \ f(h_m) = f_m h_m \ , \tag{10}$$

with a_i, b_i, f_m positive constants.

Proposition 3. *Consider the support dynamics in equation (3) with the function specifications given in (10).*

1. *Suppose that the matrix M with entries $m_{ij} = a_i \left(b_i h_i - w_{i,j} f_j h_j\right)$ has only eigenvalues with positive real part. Then the network is asymptotically stable and the stationary covariance ϕ satisfies the algebraic equation*

$$M\phi + \phi M^\top = \sigma\sigma^\top. \tag{11}$$

2. *Assume that the system (3) is two-dimensional. Then, the stationary covariance is explicitly*

[3]The proof of this fact is due to a theorem of Vilela Mendes and Duarte (1992). Basically, the dynamics of the nodes are given by a superposition of gradient and Hamiltonian vector fields. The latter is responsible for positive cyclical and chaotic behavior.

given by:

$$\phi = \frac{\det(M)\sigma\sigma^\top + (M - \mathrm{tr}(M)\mathbf{I})\,\sigma\sigma^\top\,(M - \mathrm{tr}(M)\mathbf{I})^\top}{2\mathrm{tr}(M)\det(M)} \tag{12}$$

The asymptotic stability property given in Proposition 3 can be extended to the case of fully nonlinear portfolio dynamics if the linear dynamics are asymptotically stable, the linearized system is close to the nonlinear system, and the nonlinear work flows satisfy some regularity conditions. If these conditions hold, the stochastic stability of the nonlinear network follows (see Freidlin and Wentzell (1999)). Therefore, in a stochastic setup, a network maintains its stability only if the operational risk factors act as small perturbations on the fixed network topology. This result rationalizes, at least in part, the practice of counteracting, mitigation, or elimination of operational risk.

4.2 Topological Diversification

The above considerations lead us to the problem of how to define diversification in the context of operational risk. An ongoing debate in the banking industry is the question of centralizing/decentralizing IT networks. Since the notion of costs and risks conflicts, an answer to the above question is not straightforward. To put the discussion on firm ground, we start with the static notion of topological diversification. We define the degree $\sharp(\mathcal{M})$ of a subset $\mathcal{M} \subset \mathcal{N}$ by

$$\sharp(\mathcal{M}) = \sum_{j \in \mathcal{M}} \sharp(n_j) = \sum_{j \in \mathcal{M}} \frac{|k_j|}{|\mathcal{N}|}$$

with $|k_j|$ the number of incoming and outgoing edges in the node n_j. Therefore, centralizing a network increases the degree $\sharp(\mathcal{M})$.

Definition 4. *We fix n output nodes and k input nodes. Let n_m be the minimum number of systems (nodes) necessary to produce the outputs and let k_m be the associated minimal edges.*

- *The portfolio (n_m, k_m) is the least topologically diversified portfolio.*
- *A portfolio (n', k') is more topologically diversified than a portfolio (n'', k'') if and only if*

$n' \geq n''$ and for any graph $\mathcal{G}$:

$$\max_{n_i \in \mathcal{G}'} \sharp(\mathcal{G}'(n_i)) \leq \max_{n_i \in \mathcal{G}''} \sharp(\mathcal{G}''(n_i)) \ .$$

The definition states that a portfolio becomes more topologically diversified, if the number of systems increases and at the same time the number of interfaces does not. To put it differently, the most topologically diversified portfolio is given by a portfolio in which each operational-risk asset has no interface with any other asset.

Figure 4 shows two graphs with two outputs with different topologies. In Panel A both graphs have $\max_{n_i \in \mathcal{G}_j^A}(G^A) = 1$ for $j = 1, 2$, but in Panel B one has $\max_{n_i \in \mathcal{G}_j^B}(G^B) = 1/2$ for $j = 1, 2$. Since the number of nodes in Panel B also exceeds those of Panel A, the portfolio in Panel B is more topologically diversified than the one in Panel A. Although topological diversification increases with the parallelization of the portfolio, such solutions are much more costly. Hence, managers face a trade-off between a high degree of topological diversification and high investment and maintenance costs.

If we compare the idea of operational-risk diversification with the one for standard financial assets, some striking features show up. In the latter case, diversification means to exploit the correlation among the assets, since the risk of a portfolio should be smaller than the sum of the individual risks. In contrast, an operational-risk portfolio is topologically more diversified the less dependence there is between the assets. Hence, from a traditional finance point of view, diversification for operational-risk assets is understood as an antidiversification.

Beside these differences, the costs to diversify are also significantly different. For market assets, these costs are basically transaction costs, which are small for liquid assets and so are neglected most of the time. In contrast, for operational-risk assets, costs comprise, e.g., investment costs, which are a much more significant factor and are likely to dominate the benefits from diversification.

These hurdles for active operational risk management and the tendency to prioritize networks rationalize the following market practices. First, management often increases diversification only for the key systems, e.g., by mirroring systems. Second, if management commits itself to an outsourcing strategy, such a strategy targets large entities of operational-risk assets.

4.3 Dynamic Diversification

The notion of diversification provided by Definition 4 is based on the architecture of the portfolio, which reflects much of common decision-making in network planning. However, Definition 4 takes a purely static point of view, and hence does not incorporate the evolution and the migration of statistical risk over time in the portfolio. For a given operational-risk portfolio, it is not a priori clear how the topological dependence distributes the stochastic risks over time. Put differently, what is initially topologically a well-diversified portfolio might become the opposite in a dynamic context.

To illustrate this issue, we proceed in two steps. First, we use a simple example to show that the static notion of diversification is of little use in a dynamic context. Second, we give a dynamic definition of diversification.

To start, we consider the case with two operational-risk assets and the stationary variance given in equation (12) with two specifications A and B. In specification A, the matrix W is an arbitrary symmetric 2×2 matrix. In specification B, the matrix W is of a lower triangular type, i.e., $w_{12} = 0$ and all other parameters are the same as in A. Hence, the specification A (B) corresponds to the less-diversified portfolio in Panel A (B) of Figure 4. We then calculate the stationary variance of a portfolio with two nodes for both configurations (see Figure 5).

To conclusions emerge: First, if we consider both the dynamic and the statistical risk, the topology-based ordering is not maintained in equilibrium. There are regions, i.e., values of the supports, where the risk of configuration A dominates the risk of B and vice versa. In other words, using only architectural planning of networks to manage operational risk does not allow us to determine the true risk figure. The portfolio that is more diversified at the beginning might end up strictly dominated by other, less-diversified portfolios.

Second, risk is less for the less-diversified portfolio in the region where the portfolio fractions are similar, i.e., close to the diagonal in Figure 5. This observation reflects the results in Proposition 2, but now extends to the case in which the dynamics of the risk factors matter. In an increasingly complex network, to obtain a low risk level a uniform allocation of resources becomes more desirable. Compared to the upper panel in Figure 5, we no longer rely on the symmetry of the coupling matrix

W for the two plots in the lower panel. This specification is equivalent to prioritizing the portfolio. We observe in our numerical example that prioritization leads to a decrease in the variance of the operational-risk portfolio.

To overcome the drawbacks of a purely static notion of diversification based on network topologies, we propose a dynamic definition of diversification for operational risk.

Definition 5. *Consider a fixed number of outputs and inputs and two portfolios generating the outputs. Let* $\mathcal{P}_i = (\mathcal{A}_{i_1}, \mathcal{A}_{i_2}, \ldots, \mathcal{A}_{i_n}), i = 1, 2,$ *be two portfolios of operational-risk assets, and let* $\tau = T - t$ *be a fixed time horizon. The operational-risk portfolio* $\mathcal{P}_1$ *is more diversified than portfolio* $\mathcal{P}_2$ *during time* τ *if and only if*

$$\mathbb{E}_{1,t}\left(\int_t^T \mathcal{P}_1(s)ds\right) \geq \mathbb{E}_{2,t}\left(\int_t^T \mathcal{P}_2(s)ds\right) \ , \tag{13}$$

where

$$\mathbb{E}_{i,t}\left(\int_t^T \mathcal{P}_i(s)ds\right) = \int_t^T \mathbb{E}_{i,t}\left(n_{i_1}(s), n_{i_2}(s), \ldots, n_{i_k}(s) \in \mathbb{A}_i\right) ds \ , \ \forall n_{i_j} \in \mathcal{P}_i \ . \tag{14}$$

The set $\mathbb{A}_i \subset [0,1]^{|\mathcal{S}_i|}$ *defines the operational risk acceptance set with* $|\mathcal{S}_i|$ *the number of states in the portfolio.*

Intuitively, if the expected value of moving to states outside a predefined risk acceptance level is lower than for the reference portfolio, an operational-risk portfolio is more diversified than a reference portfolio. Definition 5 is based on an internal criterion, the choice of the risk acceptance levels, which are usually imposed by the bank's management. As operational risk is institution-based and a bank's risk figures are not comparable to those of other banks, such an internal criterion is particularly apposite.

5 Calibration

Once we have formulated our model, our next step is to calibrate the operational-risk data. Financial institutions typically acquire operational risk data in two ways. First, they rely on self-

assessments and simulations to generate enough actual data. Second, they set up databases to collect historical operational-risk data. Both activities are encouraged by the regulatory authorities.

However, the acquisition of adequate data poses serious problems. Although self-assessment based on expert knowledge seems easy to implement, the pitfalls in human decision-making under risk and uncertainty make this a challenging task.[4] Furthermore, operational-risk data in the low frequency/high impact domain are very sparse. Therefore, some institutions pool their data sets. However, the applicability of these data sets suffers from the fact that operational-risk data are, by Definition 1, both time- and institution-dependent.

In our model setup, we describe calibration for the self-assessment approach. We consider counter measure dynamics that are state-dependent and given as in equation (3). Hence, we linearly approximate the support function f and the speed function s around their expected value and the fixed costs, respectively. Carrying out the analysis, we get

$$dh_i(t) = (M_{i1} - M_{i2}h_i(t))dt + \sigma_i dW_i \ , \tag{15}$$

where M_{i1} and M_{i2} are the ith diagonal element of the matrix M_1 and M_2, respectively. The matrices M_1 and M_2 are given in Appendix A.

The model in equation (15) captures the fixed costs in the node dynamics, the sensitivity of the counter measures to a change in the support function, the mean node support, and the sensitivity of the node's state on a change in the underlying support. The solution of (15) is

$$h_i(t) = e^{-M_{i2}t}\left(h_i(0) + \int_0^t e^{M_{i2}s}M_{i1}ds + \int_0^t e^{M_{i2}s}\sigma_i dW_i\right) \tag{16}$$

If we assume that the node sensitivity matrix is symmetric, we can decouple the system using standard linear algebra. Hence, we obtain explicit formulæ for the probability laws of either the nodes or the support dynamics. We link this information to the expert knowledge as follows. We

[4]See, e.g., Rabin (1998), Ebnoether, Vanini, McNeil and Antolinez (2003), and Doebeli, Leippold and Vanini (2003).

write

$$p^{\mathrm{Op}}_{i,\Delta t} = \mathbb{P}\left[h_i(t+\Delta t) \leq \overline{h}_i |\ h_j(t) \geq \overline{\overline{h}}_j\ ,\ \forall j \in \mathcal{N}\right] \tag{17}$$

for the probability that the support for system i will be below the required level $\overline{h}_i$ at time $t+\Delta t$ conditional on an operating environment. That is, the support of all systems at time t is larger than the support $\overline{\overline{h}}_j$ needed to fulfill all duties which are fixed, e.g., in a service level agreement (SLA). We chose time Δt such that all systems can fully recover in this time interval. We obtain the probability $p^{\mathrm{Op}}_{i,\Delta t}$ explicitly in terms of several constants that we still need to calibrate. The constants are the coupling constants w_{ij}, the volatility σ and the performance index. Management chooses the threshold values $\overline{h}$ and $\overline{\overline{h}}$ and the fixed costs.

A second expression that we can use for calibration is

$$p^{\mathrm{Fail}}_{i,k,\Delta t} = \mathbb{P}\left[h_i(t+\Delta t) \leq \overline{h}_i |\ h_k(t) \leq \overline{h}_k, h_j(t) \geq \overline{\overline{h}}_j\ ,\ \forall j \in \mathcal{N}\backslash\{k\}\right]\ . \tag{18}$$

Equation (18) gives the probability that the system i will be down given that all other systems are running except system k.

Given our model setup, there are many other expressions that we can calculate. Such expression lend themselves for calibration. However, from the many possibilities that we can use to calculate probabilities, conditional probabilities, and moments, we should choose these quantities that can be related in the simplest way to the expert knowledge. We obtain such knowledge by using questionnaires. Below, we give an example of a question to extract expert knowledge. The expressions in brackets relate to the model and are not part of a questionnaire:

Choose a time period (T), such that you can give an answer to the following question: How often in the chosen period (T) do you expect your system (i) to be down ($\mathbb{E}[F_i]$) if all systems (j) which provide inputs to your system (i) will be fully operational during the chosen period?

If T_Δ equals the largest integer smaller than T/Δ, where we assume that the recovery time and

the period Δt are measured in units of hours, we get

$$\mathbb{E}\left[F_i\right] = \sum_{j=1}^{T_\Delta} j\Delta t(1-p_{i,\Delta t}^{\mathrm{Op}})^{j-1} p_{i,\Delta t}^{\mathrm{Op}} = \frac{1-\left(1-p_{i,\Delta t}^{\mathrm{Op}}\right)^{T_\Delta}\left(1+T_\Delta p_{i,\Delta t}^{\mathrm{Op}}\right)}{p_{i,\Delta t}^{\mathrm{Op}}} \ . \tag{19}$$

Hence, we obtain an equation for the unspecified probability $p_{i,\Delta t}^{\mathrm{Op}}$ that can be solved numerically. Its solution gives us an expert-based estimate $\hat{p}_{i,\Delta t}^{\mathrm{Op}}$ of the model probability $p_{i,\Delta t}^{\mathrm{Op}}$. To calibrate the model, we can ask additional questions of the same type as above. The evaluation of the questionnaires gives us a vector of estimated expressions symbolized as $\hat{x}$. We then equalize these expressions to the analytical counterparts $x(\Phi)$, which depend on a vector of model parameters Φ. Solving for estimated parameters $\hat{\Phi}$, we have determined the dynamics of the support function and the nodes.

6 Numerical Example

In this section, we apply the model to the IT systems network for a bank's risk management, as illustrated in Figure 2. Our discussion splits into two parts. In the first part, we explore the two quantities of severity S and duration D for the risk management network. We denote this network as $\mathcal{N}_1$. In the second part, we discuss the effect of an alternative network architecture.

For both $\mathcal{N}_1$ and $\mathcal{N}_2$, the network dynamic is captured by the support function h_i acting on node n_i. We assume that uncertainty is generated by the node-specific standard Brownian motions W_i and an external factor Z that acts on the whole network. We assume that the support function follows

$$dh_i = a_i(h_i)\left(b_i(h_i) - \sum_{m=1}^{N_1(i)} w_{i,m} f(h_m)\right) - \sigma_i dW_i - \varsigma_i dZ. \tag{20}$$

For the function f in equation (1), which defines the trigger for the operational-risk event, we use the complementary error function as in equation (2). Whether a node n_i is running or down depends on the level of support function relative to the threshold reflected in the function $b_i(h_i)$. For simplicity, we assume $b_i(h_i) = b =$ constant. We further simplify the model by setting $\sigma_i = \sigma$ and $\varsigma_i = \varsigma$. The dW_i's and dZ are correlated by a constant ρ. The counter measure function $a_i(h_i)$

is given by equation (5).

Having defined the static and dynamic characteristics of the network, we simulate the severity S and the duration D for each node and aggregate them for paths, graphs, and for the whole network. These two key figures then build the basis for the calculation of operational risk losses and costs (see Section 3.2). For the simulation of $\mathcal{N}_1$, we discretize model dynamics for the states and set the time increment Δt equal to one day. We use $10,000$ simulation runs and a time horizon of 30 days. This time horizon can be appropriate for some value chains such as risk management. An operational-risk portfolio supporting the trading floor requires a much shorter time horizon. For our analysis, we concentrate on the total of the risk-management nodes and on node n_1, which lies outside the risk-management network.

Given the value-generating work flows, we can calculate the costs and the losses of an operational-risk event and express them in terms of interpretable risk figures such as value-at-risk, expected shortfall, or maximum loss. As this is a straightforward exercise, we focus on the difference between two different operational risk management practices and their impact on the two key quantities, duration and severity. The first risk-management practice provides both a constant-level support and state-dependent counter measures whenever an operational-risk event occurs. The second practice provides only a constant support. In the latter case, we expect that the length of time a system malfunctions will be considerably prolonged and the severity will be much larger.

Panels (A1) and (B1) of Figure 6 plot the conditional distributions of duration and severity for $\mathcal{N}_1$ when counter measures are taken into account. In this case, both constant and state-dependent measures counteract an operational-risk event. Panel (A1) plots the distribution for node n_1, and Panel (B1) plots the distribution for the aggregated nodes of the risk-management network. The panels (A2) and (B2) plot the corresponding conditional distributions for the situation in which we impose no counter measures and the bank's management acts, given an operational-risk event, only through the constant support function.

The above risk distributions can be used for profitability management as follows. For the process owner of the risk-management value chain, the distributions serve as a valuation tool of the SLA between the IT system layer (see Figure 2) and the business layer (see Figure 1). The distribution allows the owner to calculate risk figures, such as expected values and value-at-risk, and to evaluate

the performance of the IT systems. In addition, the counter measures, which express stochastic costs, and the fixed maintenance costs characterize the IT systems from a second point of view.

Using these two data sets of risks and costs, the process owner can decide according to her preferences whether the risks and costs of the supporting systems is acceptable given the contracted SLAs. If this is not the case, the process owner can increase/reduce counter measures or fixed costs, thereby achieving acceptable risk and cost levels. This procedure defines a consistent and comprehensive management tool for the value chain's profitability.

Our model setup also allows us to explore the impact of a change in the effectiveness of counter measures by changing the parameter ψ in equation (5). Figure 7 plots the realized difference in duration and severity together with different confidence contours, when the effectiveness of the state-dependent counter measures drops to a lower level. As expected, the drop in the counter measures' effectiveness leads to an increase in both the severity and duration of operational-risk events. Such an analysis of the effectiveness eventually allows the management to quantify the gains from measures that target at an increase in the effectiveness of operational-risk management.

To discuss the impact of altering the network structure, we consider the system structure illustrated in Figure 8 as our alternative. We label the resulting network as $\mathcal{N}_2$. The motivation behind the specification $\mathcal{N}_2$ lies in the current efforts of many banks to reduce the number of systems used. These efforts are based on two causes: first, a large network complexity is costly, and second, most IT systems providers now offer highly integrated systems with multiple functionalities. Compared to $\mathcal{N}_1$, $\mathcal{N}_2$ in Figure 8 reduces the systems involved in the risk-management work flow from eight to only three.

From Figures 2 and 8, we see that the risk-management work flow in $\mathcal{N}_2$ is topologically less diversified. However, this observation does not necessarily mean that $\mathcal{N}_2$ is also dynamically less diversified. Indeed, it turns out that the diversification index, as proposed in Definition 4, only differs slightly between $\mathcal{N}_1$ and $\mathcal{N}_2$. The numbers we obtain are 0.9936 and 0.9876. Whether the difference between these two numbers is significant from a business point of view depends on the value chain under consideration. If the current risk level in network $\mathcal{N}_1$ is acceptable for the management, the switch to the less (topologically) complex network $\mathcal{N}_2$ might be justified by investment and maintenance-cost considerations.

However, if management is considering a processing value chain that involves a huge amount of different daily work steps, the difference could turn out to be significant. In that case, the management would be well-advised to further increase the efficiency of counter measures so as to further decrease the risk figures of the network.

7 Conclusion

The quantification of operational risk is a challenging task. Often, it is not clear how the management of operational risk differs from the management of market or credit risk, or how it can add value to an institution. We use a financial-economics approach to interpret operational risk management as a means to optimize the profitability of an institution along its value chain. To develop the formal model, we introduce a realistic risk-management value chain that guides us through the theoretical construction of our model.

Often, practitioners adopt only a static perception of operational risk. We propose a dynamic approach that builds on functional dependencies among different systems within a network. Based on our definition of operational risk, we use an analytical perspective to address the problem of network prioritization in terms of costs and diversification. We find that prioritizing the network can lead to instable behavior. Furthermore, when considering operational risk, we suggest splitting diversification into topological and dynamic diversification. A topologically more diversified portfolio need not be dynamically more diversified, and vice versa.

The problem of decentralizing or centralizing under operational risk threats relates to the topological dependence structure embedded in the stochastic model. Aside from stability issues, risk is likely to lessen over time. A purely static architectural planning based on the diversification can lead to a hazardous misperception about the true risk exposure. Therefore, to account for the mixing between stochastic and topological dependence over time, a dynamic modeling of operational risk is indispensable.

We use a numerical application on the risk-management value chain to highlight the usefulness of our approach for management decisions. We first quantify the interdependent operational risks and costs. These two data sets generate risk and costs figures for the management. We analyze the

influence of counter measures and compare the resulting key figures to the case in which counter measures are neglected. Finally, we quantify the impact of altering the network architecture. We conclude that our model can serve as a valuable decision tool for a bank's operational-risk management.

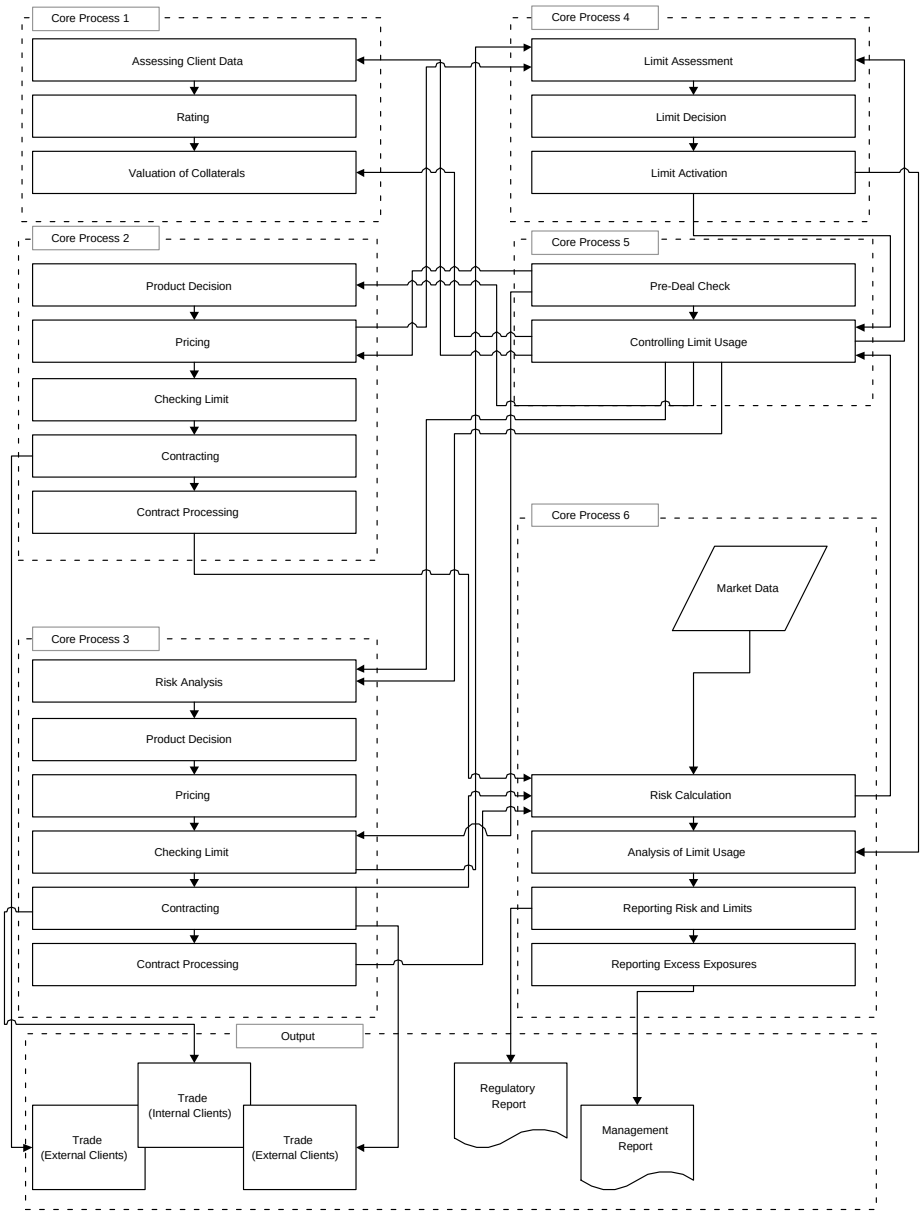

Figure 1: Work flow view for trading, sales, and risk management. Work flows 1 and 2 define the sales, work flow 3 the trading unit, and work flows 4 to 6 the risk management.

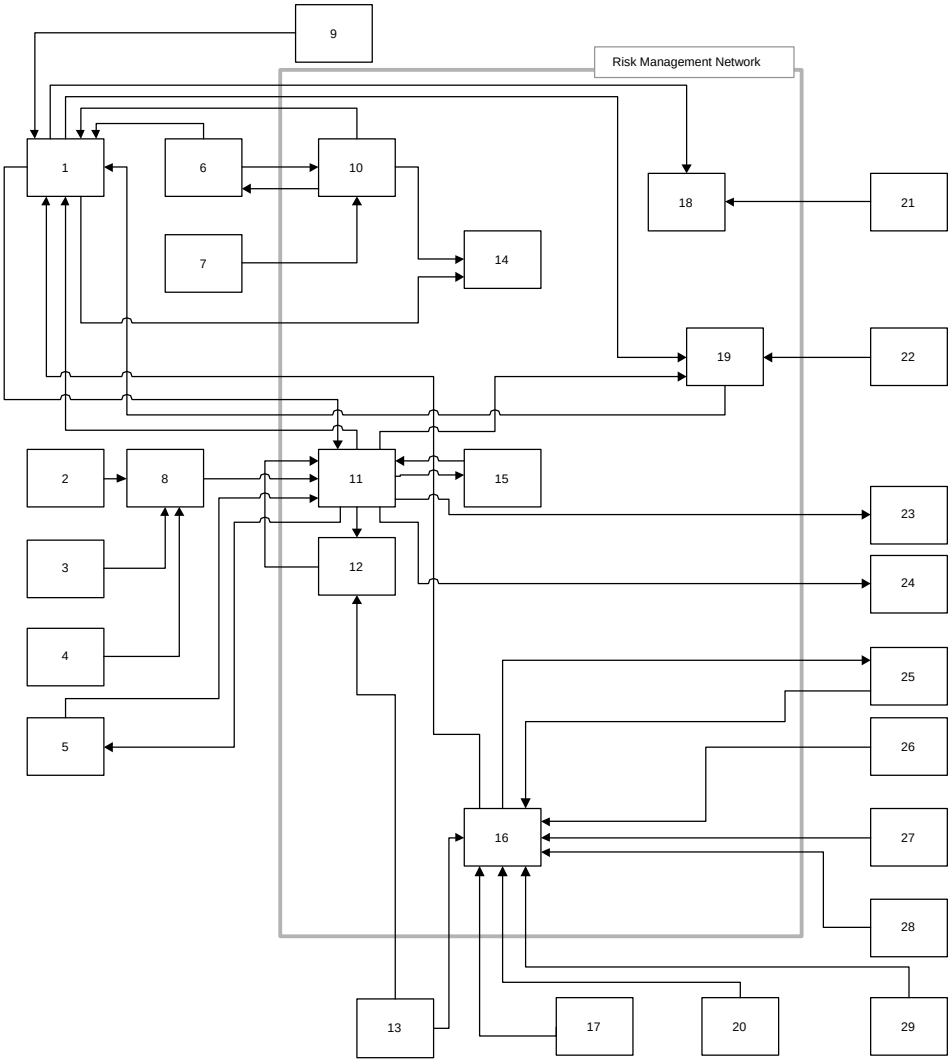

Figure 2: System view of trading, sales, and risk management work flows and the subnetwork of limit and risk management work flows.

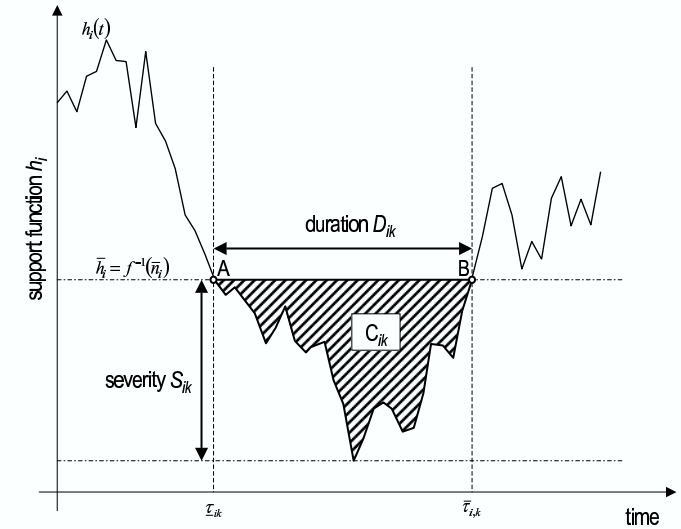

Figure 3: Parameterizing cost and loss functions. When the value of the support function $h_i(t)$ falls below the threshold $\bar{h}_i$, an operational-risk event k occurs. The operational-risk event can be characterized by the time interval, during which the support function stays below the threshold, and by the minimum value the support function obtains during this time interval. We call these quantities duration D_{ik} and severity S_{ik}, respectively.

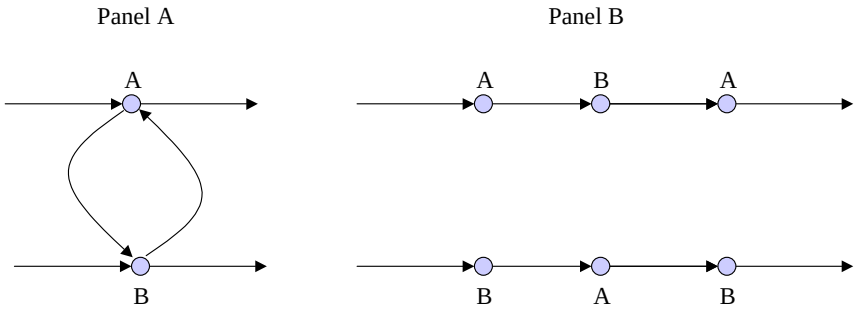

Figure 4: Two different topologies for two graphs.

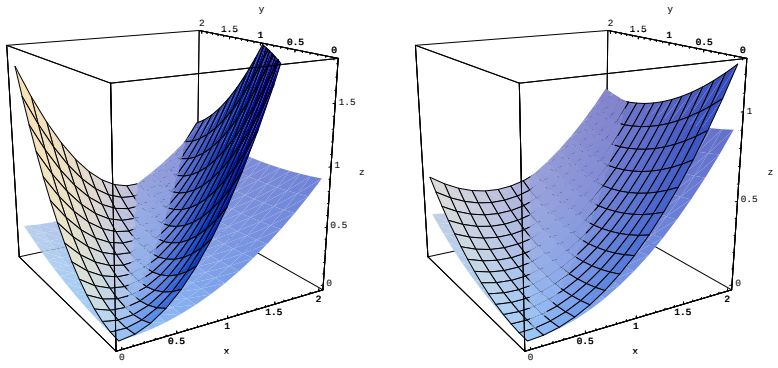

Figure 5: The stationary covariance structure for specification A (meshed surface) and B. The left panel plots the covariance structure for the symmetric case and the second panel plots the asymmetric case. We set $\sigma_{11} = 0.22$, $\sigma_{12} = 0.1$, $\sigma_{22} = 0.15$ and, for the matrix M, $w_{11} = 0.25$, $w_{12} = w_{21} = 0.3$, $w_{22} = 0.4$ and $w_{12} = 0.2$ for the asymmetric case. The z-axis denotes the variance of the stationary distribution.

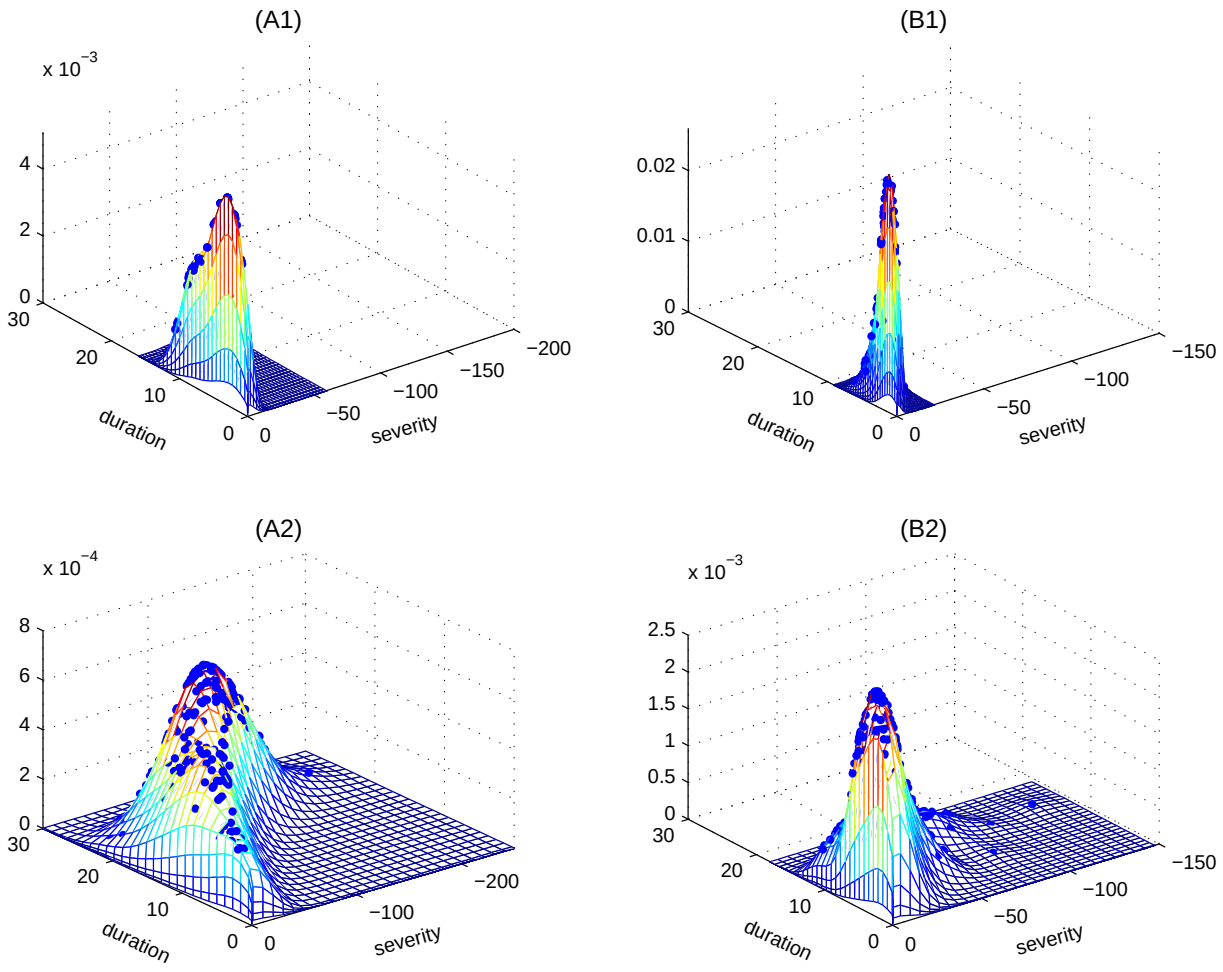

Figure 6: Conditional distribution of duration and severity with and without counter measures. Panels (A1) and (B1) plot the conditional distributions of duration and severity for node n_1 and for the aggregated risk-management nodes. Panels (A2) and (B2) plot the conditional distributions when no counter measures are imposed. The distributions are generated by a normal kernel density estimation based on a simulation with 10,000 runs.

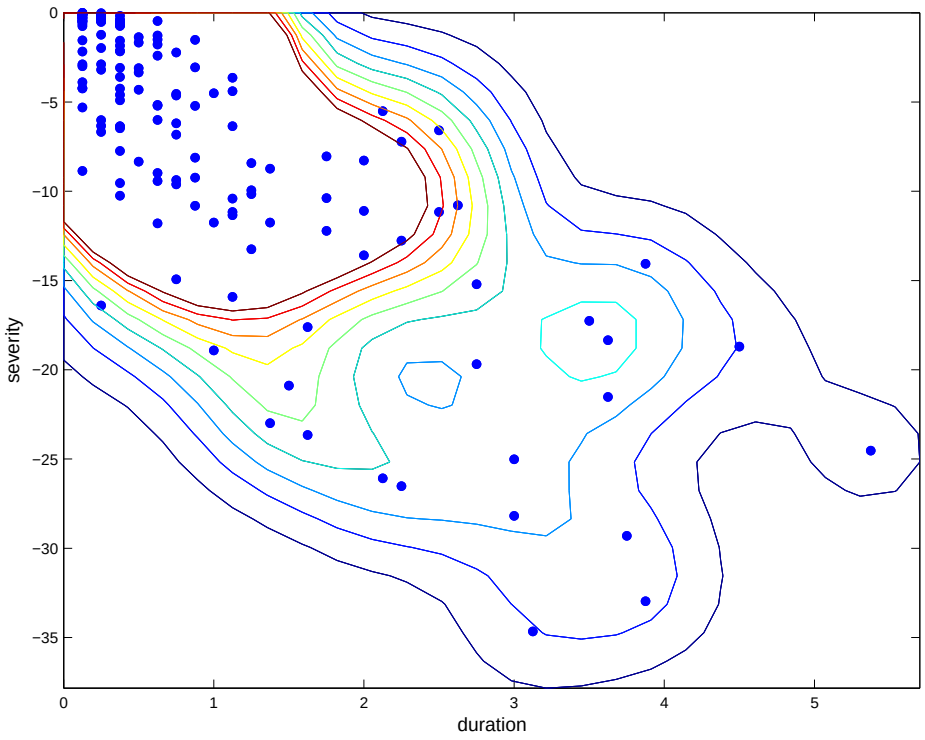

Figure 7: Impact of Effectiveness. The figure shows the difference in duration and severity, when the effectiveness of counter measures is reduced. The plot is based on a simulation with 10'000 runs with normal kernel density estimation.

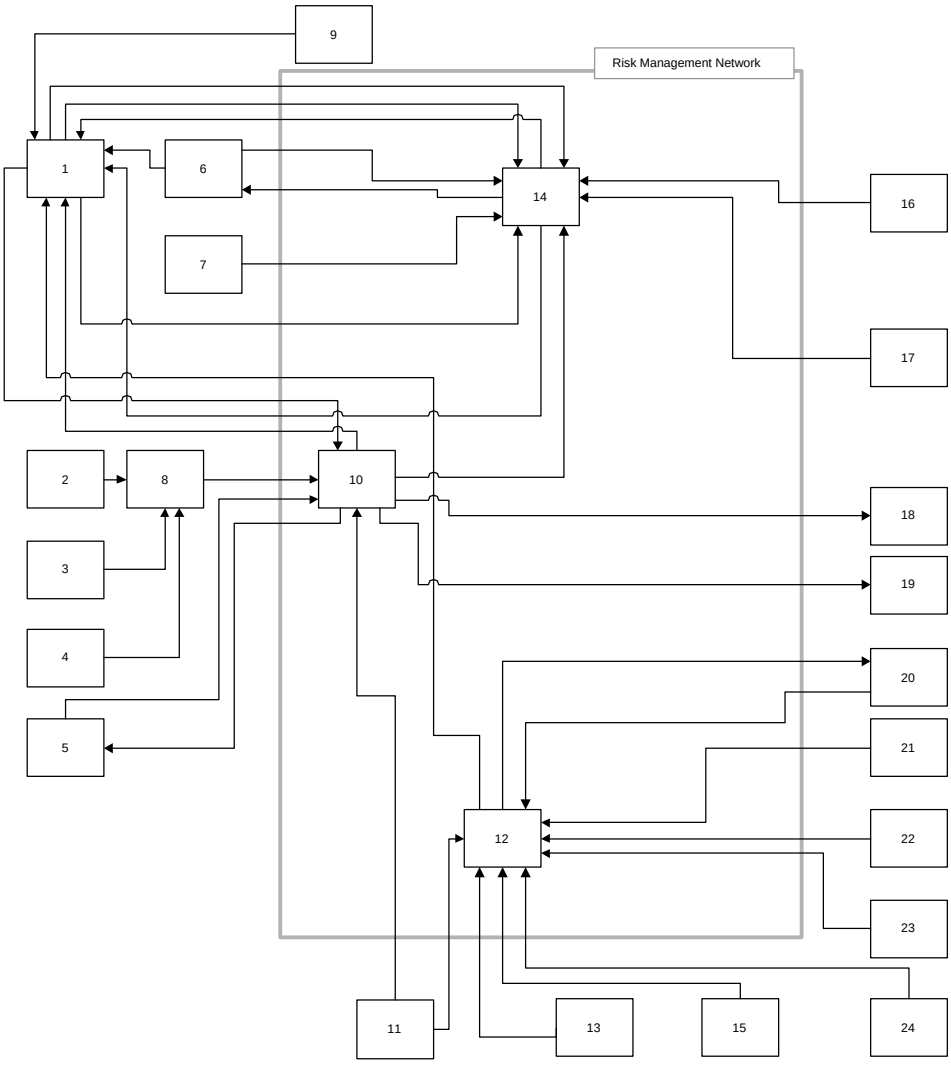

Figure 8: Alternative graph structure $\mathcal{G}_2$.

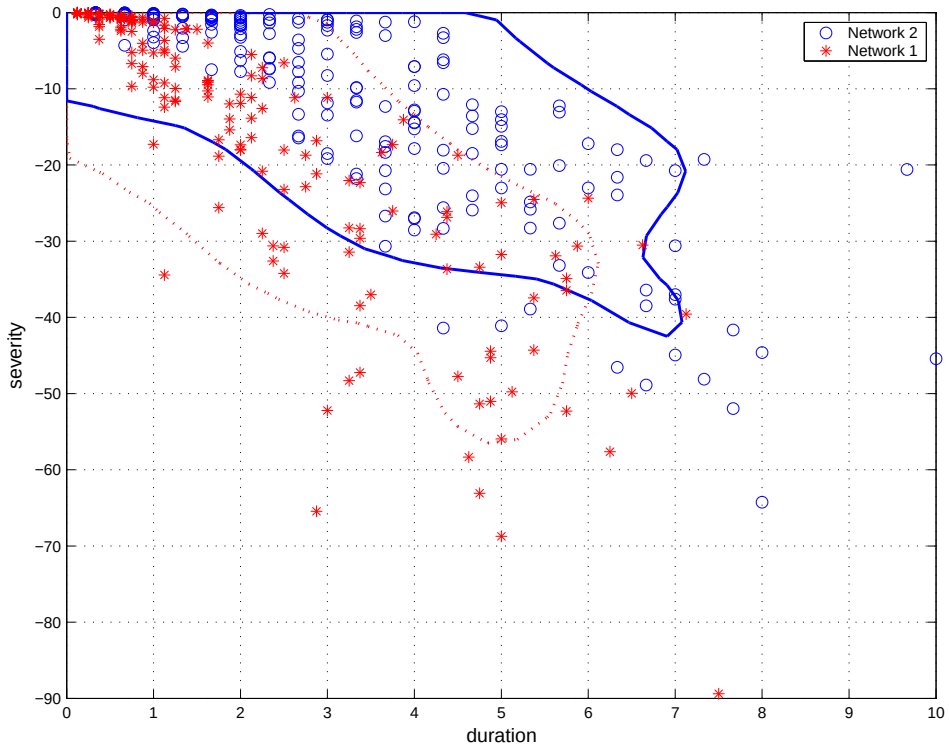

Figure 9: Altering the network architecture. The sample size of the simulation is 10,000 runs. For network $\mathcal{N}_1$, we observe 208 operational-risk events and 236 events for network $\mathcal{N}_2$.

Appendix A

We use the first-order approximations

$$\begin{aligned} s(h) &= s(\bar{h}) + \frac{\partial s(\bar{h})}{\partial h}(h - \bar{h}) + O(h^2), \\ f(h) &= f(\bar{h}) + \frac{\partial f(\bar{h})}{\partial h}(f - \bar{f}) + O(f^2), \end{aligned}$$

for the speed function s and for the function f. We then get the linearized support dynamics, where we write h for the increment $h - \bar{h}$,

$$\begin{aligned} dh_i &= \left(s_i(\bar{h}) + \partial s_i h_i\right)\left(b_i(\bar{h}) + \partial b_i h_i - \sum w_{ij}\left(f_j(\bar{h}) + \partial f_j h_j\right)\right) dt, && \text{(A.1)} \\ &= \underbrace{s_i(\bar{h})\left(b_i(\bar{h}) - \sum w_{ij} f_j(\bar{h})\right)}_{=y_i} dt + \underbrace{\partial s_i\left(b_i(\bar{h}) - \sum w_{ij} f_j(\bar{h})\right) + s_i(\bar{h})\partial b_i)}_{=x_i} h_i dt \\ &\quad -s_i(\bar{h}) \sum w_{ij} \partial f_j h_j dt, && \text{(A.2)} \end{aligned}$$

with the shorthand notation $\partial s_i = \frac{\partial s_i(h)}{\partial h}$. In vector notation, for $h = (h_1, \ldots, h_k)^\top$ and $y = (y_1, \ldots, y_k)^\top$, we obtain

$$dh = ydt + \operatorname{diag}(x_1, \ldots, x_k) h dt - \begin{pmatrix} s_1 w_{11} \partial f_1 & \ldots & s_1 w_{1k} \partial f_k \\ \vdots & \ddots & \vdots \\ s_k w_{k1} \partial f_1 & \ldots & s_k w_{kk} \partial f_k \end{pmatrix} h dt. \qquad \text{(A.3)}$$

Thus, the matrices M_1 and M_2 are given by

$$M_1 = \operatorname{diag}(y_1, \ldots, y_k), \qquad \text{(A.4)}$$

$$(M_2)_{ij} = \begin{cases} x_i - s_i w_{ii} \partial f_i, & \text{if } i = j; \\ -s_i w_{ij} \partial f_j, & \text{if } i \neq j. \end{cases} \qquad \text{(A.5)}$$

Appendix B

Proof of Proposition 2

The absence of mirrored systems implies that

$$dh_i = a_i(h_i)\left(b_i(h_i) - \sum_{m=1}^{N_1(i)} w_{i,m} f(h_m)\right), \forall i \in \mathcal{N} \ .$$

We next apply the following fundamental theorem of dynamical systems, using the following definitions. First, we decompose the matrix W in its symmetric and anti-symmetric parts:

$$W^{(S)} = \frac{1}{2}(W + W^\top) \ , \ W^{(A)} = \frac{1}{2}(W - W^\top).$$

Second, we introduce the gradient U and Hamiltonian H functions:

$$\begin{aligned} U &= -\sum_{i=1}^{\mathcal{N}} \int^{h_i} b_i(\xi_i) f_i'(\xi_i) d\xi_i + \frac{1}{2}\sum_{j,k} w_{j,k}^{(S)} f_j(h_j) f_k(h_k), \\ H &= \sum_{i=1}^{\mathcal{N}} \int^{h_i} \frac{f_i(\xi_i)}{a_i(\xi_i)} d\xi_i \ . \end{aligned}$$

Finally, we define the Riemannian metric g_{ij} and the symplectic form Λ_{ij}:

$$g_{ij} = \frac{a_i(h_i)}{f_i'(h_i)}\delta^{ij} \ , \ \Lambda_{ij}(h) = -\left(a_i(h_i)\right)^{-1}\left(\left(W^{(A)}\right)^{-1}\right)_{ij}\left(a_j(h_j)\right)^{-1}.$$

with δ the Kronecker symbol.

Theorem 6. *$a_i(h_i)/f'(h_i) > 0$ is a constant for all i and h. If the matrix $w^{(A)}$ is invertible, the vector field decomposition*

$$\dot{h}_i = \dot{h}_i^{(G)} + \dot{h}_i^{(H)}.$$

follows, with

$$\begin{aligned} \dot{h}_i^{(G)} &= -\sum_j g_{ij}\frac{\partial U}{\partial h_j}, \\ \dot{h}_i^{(H)} &= \sum_j \Lambda_{ij}\frac{\partial H}{\partial h_j} \ . \end{aligned}$$

For the proof of the theorem see Vilela Mendes and Duarte (1992). Their proof extends the fundamental

Cohen-Grossberg Theorem (1983) to the case in which W no longer needs to be symmetric. Since we assume that the matrix W is symmetric, the Hamiltonian part vanishes and only the gradient field enters the dynamics. We then get:

$$\begin{aligned}
\dot{h}_i^{(G)} &= -\sum_j g_{ij} \frac{\partial U}{\partial h_j} \\
&= -\sum_j g_{ij} \left(-b_j(h_j) f_j'(h_j) + \sum_k w_{k,j}^{(S)} f_k(h_k) f_j'(h_j) \right) \\
&= -a_i(h_i) \left(-b_i(h_i) + \sum_k w_{k,i}^{(S)} f_k(h_k) \right) \ .
\end{aligned}$$

From these facts follows that the operational risk network with symmetric couplings possesses a Lyapunov function which, in turn, guarantees global asymptotic stability of the dynamics. Since the support function and the node states are related by a monotone transformation, we proved the claim. □

Proof of Proposition 3

The linearized dynamics reads

$$dh_i = a_i \left(b_i h_i - \sum_{m=1}^{N_1(i)} w_{i,m} f_m h_m \right) + \sigma dW_t, \forall i \in \mathcal{N} \ .$$

In matrix notation, we write it as

$$dh_t = M h_t dt + \sigma dW_t,$$

with the matrix $M_{ij} = a_i \left(b_i h_i - w_{i,j} f_j h_j \right)$. If M has only eigenvalues with positive real part and if we set the initial time at $-\infty$ instead at zero, the solution of the system reads

$$h_t = \int_{-\infty}^{t} e^{-M(t-s)} \sigma dW_s \ . \tag{A.6}$$

Equation (A.6) implies a mean of zero and a covariance matrix

$$E(h_t h_s^\top) = \int_{-\infty}^{\min(t,s)} e^{-M(t-v)} \sigma \sigma^\top e^{-M^\top (t-v)} dv \ . \tag{A.7}$$

with $A^\top$ the transpose of the matrix A. If we set $\phi(t) := E(h_t h_t^\top)$ for the stationary covariance matrix, we have

$$\begin{aligned} M\phi + \phi M^\top &= \int_{-\infty}^{t} M e^{-M(t-v)} \sigma\sigma^\top e^{-M^\top (t-v)} dv \\ &\quad + \int_{-\infty}^{t} e^{-M(t-v)} \sigma\sigma^\top e^{-M^\top (t-v)} M^\top dv \\ &= \int_{-\infty}^{t} \frac{d}{dv} \left(e^{-M(t-v)} \sigma\sigma^\top e^{-M^\top (t-v)} \right) dv \\ &= \sigma\sigma^\top - \lim_{v \to -\infty} e^{-M(t-v)} \sigma\sigma^\top e^{-M^\top (t-v)} \\ &= \sigma\sigma^\top \ . \end{aligned}$$

by the assumption on the eigenvalues of M. To prove the second claim, we note that each 2×2 matrix satisfies the characteristic equation

$$M^2 - \mathrm{tr}(M) M + \det M = 0 \ .$$

This implies that e^{-Mt} is a polynomial of first order in M, since in the power series expansion higher orders can be re-expressed as a first-order polynomial by using the characteristic equation. Furthermore, from (11) and (A.7) follows that ϕ can be expressed as

$$\phi = \alpha\sigma\sigma^\top + \beta(M\sigma\sigma^\top + \sigma\sigma^\top M^\top) + \vartheta M\sigma\sigma^\top M^\top \ . \tag{A.8}$$

Using the characteristic equation, we find that (11) is satisfied if and only if the unknown α, β and ϑ satisfy a linear equation system. Solving this system for the unknown and inserting them into (A.8) proves the second claim. □

References

Cohen, M. and Grossberg, S.: 1983, Absolue stability of global pattern formulation and parallel memory storage by competitive neural networks, *IEEE Transactions on Systems, Management and Cybernetics* **13**, 815–826.

Doebeli, B., Leippold, M. and Vanini, P.: 2003, From operational risk to operational excellence, *Advances in Operational Risk Management*, RISK Books, 2nd Edition,Risk Waters Group, London, chapter 15.

Ebnoether, S., Vanini, P., McNeil, A. and Antolinez, P.: 2003, Operational risk: A practitioner's view, *Journal of Risk* **3**, 1–18.

Egloff, D., Leippold, M. and Vanini, P.: 2003, Credit risky portfolios with micro- and macro-economic dependencies, *Working paper*, University of Southern Switzerland.

Freidlin, M. and Wentzell, A.: 1999, *Random Perturbations of Dynamical Systems*, Springer, New York.

Giesecke, K. and Weber, S.: 2002, Cyclical correlations, credit contagion, and portfolio losses, *Working paper*, Cornell University and Technische Universität Berlin.

Kühn, R. and Neu, P.: 2003, Functional correlation approach to operational risk organizations, *Physica A* **322**, 650–666.

Rabin, M.: 1998, Psychology and economics, *Journal of Economic Literature* **Vol. XXXVI**, 11–46.

Vilela Mendes, R. and Duarte, J.: 1992, Vector fields and neural networks, *Complex Systems* **6**, 21–30.