

Four Theorems and a Financial Crisis

Bikramjit Das* Paul Embrechts*[†] Vicky Fasen*

ETH Zurich

In this paper we give an academic assessment of the financial crisis (crises) from our point of view and discuss where quantitative risk management went wrong. We formulate four mathematical theorems/research areas which have relevance for financial crises in general where the underlying theme is model uncertainty. Related to these theorems, key issues that will be discussed are: financial alchemy on Wall street, risk aggregation and diversification, tail dependence for a portfolio of losses, and the significance of correlation bounds.

AMS 2010 Subject Classifications: primary: 00-01, 91G99
secondary: 91B02, 91B30

Keywords: copula, correlation, extreme value theory, model uncertainty, regular variation, risk management, Value-at-Risk.

*Department of Mathematics and RiskLab, ETH Zurich, Raemistrasse 101, 8092 Zurich, Switzerland, email: bikramjit.das@math.ethz.ch, embrecht@math.ethz.ch, vicky.fasen@math.ethz.ch

[†]SFI Senior Professor

1 Introduction

At the time of writing these lines, we find ourselves going from the 2007-2008 subprime crisis to the 2009-20 xy government debt crisis, where hopefully $x = 1$ but the value of y is unclear at the moment. The former crisis no doubt had its roots in the USA, the latter more in Europe. Both, however, have considerable impact on the world at large. The extent to which they can be viewed as economically separate crises or whether they constitute one single crisis is for future historians to find out. A very readable discussion on this (in German) is Brunetti (2011), see also the excellent Kindleberger and Aliber (2011). For the purpose of this paper, we refer to *The Crisis* as the economic events which took place around the subprime crisis; the messages given however extend well beyond.

If there is one clear bubble that surely came out of The Crisis, then it is, without doubt, the number of articles, press coverings, books, conferences, etc., all devoted to the topic. On the other hand, serious warnings beforehand were rare, and this in the media, industry as well as academia. In earlier reactions on The Crisis, mathematics got some of the blame; this resulted mainly from its contribution to the financial engineering of more and more complex and opaque products on Wall Street. Products which in the end were understood by very few, if any indeed; a typical example are the so-called synthetic CDOs, leading eventually to Goldman Sachs' ABA-CUS 2700-AC1. Concerning the latter, its creator Fabrice Tourre was quoted as having said: "What if we created a *thing*, which has no purpose, which is absolutely conceptual, and highly theoretical and which nobody knows how to price?" (The Financial Times, 29/1/2007)

As academics, we have a moral obligation and a societal duty to ask ourselves: "What really went wrong?" and make sure that in our teaching and research we transmit the lessons learned. The biggest mistake for us would be "to continue as if nothing has happened, or is happening"; see Embrechts (2011). Clearly the question has to be discussed at all relevant levels of academia as well as society, including politics. In this paper, we will concentrate in a kind of "mea culpa"-way on lessons learned from and for (financial) mathematics, including more quantitative fields of economics and finance. In Donnelly and Embrechts (2010) we summarized, with numerous references, some of the earlier aspects of the debate.

As already stated above, in this paper we shall highlight some of the mathematical issues underlying aspects of The Crisis. It is to be hoped that, through the results chosen, the reader will get a better understanding of the role of mathematics in banking and finance. Of course, we cannot, and indeed will not treat all aspects of this interaction. If there is one thing that is correct uniformly throughout mathematics then it is the fact that a result holds true if and only if it is proven. A somewhat cynical view on banking would be "In banking, a result is right if and only if it is profitable"; see Rogers (2010). What constitutes a proof in mathematics is hardwired in its axiomatic foundation and based on centuries of development in mathematical logic. An integral and crucial part of any theorem are the conditions under which certain (very) precise conclusions hold. At this point the (often more applied) critic may say: "This is nice and laudable in your Platonean universe but breaks down in practice!" To a certain extent, this may be true, but at the same time hides a dangerous argument. From a more technical, methodological point of view, The Crisis saw numerous examples where "practice" fully misunderstood the conditions under which some mathematical concepts or results could be applied. Or indeed where models

were applied to totally insufficient or badly understood data; the typical “garbage in, garbage out” syndrome. Also, mathematicians have failed in their effort to communicate such conditions more forcefully, and more broadly. One major theme that mathematicians should have stressed more is *Model Uncertainty*.

As the title promises, in this paper we will exemplify the above through four theorems, each of which has some bearing on The Crisis. In some cases, this will be more philosophical, in others very concrete. For each of the theorems discussed, we give a precise mathematical formulation, together with a reformulation or translation to circumstances linked to The Crisis. In Section 2, we start with the *Banach-Tarski* Theorem (also referred to as Paradox). Investment Banking, over the recent years, seems to have tried very hard to “apply” this result in practice and as a consequence did strive for financial product constructions which, for a while and for those directly involved, led to ridiculous multiplications of gains and wealth. We then move on, in Section 3, to some results related to (non-)aggregation properties of risk measures used in capital adequacy calculations within the financial industry. The theorem we focus on goes under the name of *Delbaen* and finds its origin in earlier work on the Loss Distribution Approach for Operational Risk. In Section 4 we highlight a result of *Sibuya* related to dependence modelling of multivariate extremal events. This copula-related result lies at the heart of the early accusations of the mathematics used for pricing and hedging of senior tranches of Collateralized Debt Obligations (CDOs). The catch-word here is the Gaussian copula. The final theorem we want to stress is due to *Fréchet* and *Höfding*. This result warns for the model uncertainty underlying the pricing of many products in finance and insurance where the underlying assumptions include statistical information on the marginal risk factors together with some idea(s) on interdependence. We conclude in Section 6 by giving an outlook on quantitative risk management research in a post-crisis period; where it is hoped that “post” means “fairly soon”.

2 The first theorem: Banach-Tarski

In an interview in 1999, the second author made the following statement: “Die Finanzwelt ist die einzige Welt, wo die Leute immernoch glauben, dass sich Eisen in Gold verwandeln lässt”; see Embrechts (1999). Translated into English: “The world of finance is the only one in which people still believe in the possibility of turning iron into gold”. The above statement was made in the wake of the 1998 LTCM hedge fund crisis; little did we know at the time how true this statement would become about ten years later!

For The Crisis, asset-backed securities, like CDOs, were in the popular press often likened to magical financial engineering tools allowing to cut a pizza into several pieces, reassemble them and end up with two pizzas each in size equal to the one we started from. In the language of modern finance, the second pizza would be referred to as “a free lunch”. This is the point where Banach-Tarski enters; the lesson to be learned is to always understand in detail the conclusions of a mathematical theorem.

Theorem 1 (Banach and Tarski (1924))

Given any two bounded sets A and B in the three-dimensional space \mathbb{R}^3 , each having non-empty interior, one can partition A into finitely many (at least five) disjoint parts and rearrange them

by rigid motions (translation, rotation) to form B .

Proof. See Wagon (1993). □

Remark 1

- (a) *An excellent and very readable account of this result, including a detailed historical discussion is found in Wapner (2005). The latter text also contains some interesting speculations on how Theorem 1 may become useful in high-energy particle physics, and especially the Big Bang Theory in cosmology.*
- (b) *Theorem 1 remains true if one replaces \mathbb{R}^3 by a general d -dimensional Euclidean space \mathbb{R}^d , $d \geq 3$. Interestingly, Theorem 1 does not hold in \mathbb{R}^2 .*
- (c) *Though the statement of Theorem 1 sounds very paradoxical, and hence is often referred to as the Banach-Tarski Paradox, it is in fact a theorem that follows fully logically from the standard axioms of mathematics. Before we discuss this, let us first reformulate the theorem in some of its more paradoxical, folkloristic, but mathematically vague versions and then return to The Crisis.*

Version 1

Given a three-dimensional solid ball (of gold, say), it is possible to cut this ball in finitely many pieces and reassemble these to form two solid balls, each identical in size to the first one.

Version 2

Any solid, a pea, say, can be partitioned into a finite number of pieces, then reassembled to form another solid of any specified shape, say the Sun. For this reason, Theorem 1 is often referred to as “The Pea and the Sun Paradox”.

As already stated above, Theorem 1 can be proved within the standard axiomatic, so-called Zermelo-Fraenkel logical framework of mathematics, often abbreviated as ZFC. Within ZFC there is one axiom, the *Axiom of Choice* (hence the C), which logicians have discussed a lot; for mainstream mathematicians, its acceptance is not an issue, despite consequential results like Theorem 1. The paradox disappears from Theorem 1 if one looks more carefully at its proof (and also what the theorem does not say); in particular, though one does have existence of the partition of A , its constituents are non-measurable (here the Axiom of Choice enters in an essential way), so the standard notion of volume does not apply to these pieces. As a result, one cannot *construct* the solution. And yet, over and over again, Wall Street tries to convince the public that such magical tricks of creating something out of nothing are possible. As a consequence, one should not be surprised that “quants” (financial engineers) are often referred to as “The Alchemists or Wizards of Wall Street”; see Kazanjian (2000), or also Braithwaite (2011) who wrote: “ So by financial alchemy, assets can be transmuted from garbage to gold - and therefore, requires less capital.”

As already stated in the Introduction, this first section (Theorem 1) may seem a bit far-fetched. Its main contribution is that with any concept, result, new methodology, one has to look very carefully at the precise definitions and possible conclusions. And concerning “creating something out of nothing” in finance, this not only reflects on financial engineering on Wall Street.

We strongly advise the interested reader to have a critical look at the history and the development of fractional-reserve banking (FRB), *the* corner stone of the modern banking industry, and draw his/her own conclusions on whether Banach-Tarski is so far off. Of course, FRB is the initial liquidity provided to the banking system facilitating the transformation of short term deposits into long term loans; the *raison d'être* of a banking system. The title of Rothbard (2008), *The Mystery of Banking*, shows that our discussion above is perhaps not all that distant from economic reality. This so-called Austrian point of view on economic theory is presented in full detail in the monumental Huerta de Soto (2009). See also The Economist (2011) for a discussion on the alternative schools on macro-economics. It is no coincidence that The Crisis has led to an increased interest into these alternative theories. And finally, Dewdney (1989) gives an amusing account of his friend Arlo Lipof who claimed to have been able to physically realize the Banach-Tarski construction. However, be careful, “Arlo Lipof” is an anagram of “April Fool”.

3 The second theorem: Delbaen

The second theorem concerns the concepts of risk aggregation and diversification, especially in the presence of (very) heavy-tailed or catastrophic risks. For a loss random variable X with distribution function F , a heavy-tailed model is typically characterized through power-tail (or Pareto-type) behavior, i.e., for some $\delta \geq 0$,

$$\bar{F}(x) := 1 - F(x) = x^{-\delta}L(x), \quad (1)$$

where the measurable function $L : (0, \infty) \rightarrow (0, \infty)$ is *slowly varying* at ∞ in Karamata's sense:

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1 \quad \text{for all } x > 0. \quad (2)$$

Standard notation is $L \in SV(\infty)$ and $\bar{F} \in RV_{-\delta}(\infty)$, i.e., \bar{F} is *regularly varying at ∞ with index of regular variation $-\delta \leq 0$* . Typical examples for distributions which have tails in $RV_{-\delta}(\infty)$ are stable distributions with index of stability $0 < \delta < 2$ and Pareto distributions with index $\delta > 0$. The theory and applications of functions satisfying (1) (and (2)) are encyclopedically summarized in Bingham et al. (1987). The key result proved by Jovan Karamata, Karamata (1930), is that the convergence in (2) holds uniformly on all compact subsets in $(0, \infty)$. From this, an extremely powerful theory can be worked out with numerous deep applications to (mainly limit theorems in) probability theory and statistics. The importance of this was early on realized by William Feller in his classic Feller (1971). Interestingly, both Feller and Karamata came from the same city of Zagreb.

For our purposes, we consider L^p -spaces of random variables X (also called *risks*) defined on some atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $X \in L^p$ iff $\mathbb{E}(|X|^p) < \infty$. It is well-known that for $\bar{F} \in RV_{-\delta}(\infty)$, $X \in L^p$ for $p < \delta$ and for $p > \delta$, $\mathbb{E}(|X|^p) = \infty$. In particular, for $\delta < 1$ the first moment of X is infinite. From a risk management point of view, such extremely heavy-tailed models do occur in the modeling of catastrophic events. Examples include the modeling

of Operational Risk under Basel (II/ III), earthquake damage, nuclear plant disasters, pyroclastic flows, internet traffic data, etc. McNeil et al. (2005) contains a discussion and further references on the former, Operational Risk. A typical question, from a risk management perspective, is how to measure the risk associated with such extremely heavy-tailed models. Indeed most popular methodologies of risk measurement typically assume the existence of the second moment, or at least the first moment. The second theorem in this paper clarifies this issue to some extent. It says that if we restrict to risks on the space L^p , $0 < p < 1$, there does not exist any risk measure that admits diversification. The mathematical background of this result is to be found in functional analysis; see Rudin (1973), p.35-36. A formulation, immediately applicable to quantitative risk management, is given in Delbaen (2009). For this let E be a vector space of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ which satisfies the following two conditions:

- (i) *Rearrangement invariant*: If X and Y have the same distribution and $X \in E$, then also $Y \in E$.
- (ii) *Solid*: If $|Y| \leq |X|$ and $X \in E$, then also $Y \in E$.

In particular, any L^p -space is both rearrangement invariant and solid. How can we now quantify the risk of the financial loss X in E ? This is done through the notion of a risk measure. Starting from the pioneering work of Paul Samuelson in the 1960's *risk* entered the portfolio manager's equation next to *return*. Over the last decades, there has been a real explosion of papers on the measurement of *risk* in the financial industry, and this partly due to international regulatory pressure. Perhaps startling is the fact that, despite this academic research output, at the height of The Crisis we went half a century backwards and took *risk* again out of the equation by solely focussing on *return*.

For the purpose of this section, we concentrate on a very basic, somewhat simplified interpretation of risk. Throughout, a one-period risk will be denoted as a random variable X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$; it denotes the value (or sometimes the change in value) of a financial or insurance position at time 1 (future), viewed from time 0 (now). We denote the distribution function of X by F . The most popular risk measures currently in use throughout the financial industry are the VaR (Value-at-Risk) entering so-called Pillar 1 capital adequacy calculations under Basel (II/III) and Solvency II, and the $\mathbb{E}\mathbb{S}$ (Expected Shortfall) used in the Swiss Solvency test. The VaR at confidence level $\alpha \in (0, 1)$ is then defined as the α -quantile, $\text{VaR}_\alpha(X) = F^{\leftarrow}(\alpha)$, where F^{\leftarrow} is the generalized inverse $F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$. The Expected Shortfall at confidence $\alpha \in (0, 1)$ is defined as $\mathbb{E}\mathbb{S}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(X) du$ if $\mathbb{E}|X| < \infty$. Note that, for F continuous, $\mathbb{E}\mathbb{S}_\alpha(X) = \mathbb{E}(X|X > \text{VaR}_\alpha(X))$, hence its name. In general a risk measure ρ is only defined as a map from E to \mathbb{R} . Within the banking regulatory framework, the value of $\rho(X)$ can be interpreted as the amount of capital that should be kept aside or invested in a risk-free asset in order to avoid a shortfall on the risk X .

Thus, we understand $\rho(X)$ as the *risk capital* needed for holding the position X . For a risk measure to be reasonable, it should satisfy certain basic properties (axioms). Convex and coherent risk measures are among the very widely used axiomatic systems in the literature. The study of coherent risk measures started with Artzner et al. (1997, 1999) and Delbaen (2000); for more details on convex risk measures see Frittelli and Gianin (2002) and Föllmer and Schied

(2011). The latter contains numerous references, also to the earlier economics literature. Kaina and Rüschendorf (2009) investigate in particular the properties of convex risk measures on L^p -spaces.

Definition 1 (Convex risk measure)

A convex risk measure $\rho : E \rightarrow \mathbb{R}$ satisfies for all $X, Y \in E$:

- (a) $\rho(0) = 0$.
- (b) *Monotonicity*: If $X \leq Y$ a.s., then $\rho(X) \leq \rho(Y)$.
- (c) *Translation-Invariance*: If $\eta \in \mathbb{R}$, then $\rho(X + \eta) = \rho(X) + \eta$.
- (d) *Convexity*: $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$ for $0 \leq \lambda \leq 1$.

If $\rho(X) \leq 0$ then the risk X is acceptable and no additional capital is necessary. The first property says that without risk, there is no reserve capital requirement. Moreover, if X is less risky than Y , then capital requirement for X is also less than that of Y (see (b)). The meaning of translation-invariance (see (c)) is that if we add the risk free amount η to X then the risk capital has to be increased by η ; this sounds odd but only reflects the fact that we consider losses as well as risk capital as positive. Finally, by convexity the portfolio of X and Y requires less risk capital than the proportional individual risk capitals. This property encourages diversification.

A subset of the set of convex risk measures are the *coherent risk measures*, where (d) is replaced by the axioms of positive homogeneity and subadditivity as follows. For all $X, Y \in E$:

- (d1) *Positive Homogeneity*: $\rho(\lambda X) = \lambda\rho(X)$ for any $\lambda \geq 0$.
- (d2) *Subadditivity*: $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

The choice of axioms (d1) and (d2) are perhaps debatable from a practical point of view and there exists a whole body of research on this topic. Positive homogeneity is often criticized as it does not take concentration of risk and liquidity risk sufficiently into account, whereas subadditivity, which reflects diversification, is criticized as it does not necessarily hold for several important situations relevant in practice; for the latter, see Embrechts and Puccetti (2010), Section 3.

Now we are ready to present the second theorem along with some interpretations. In contrast to the original paper Delbaen (2009) we formulate this theorem for risk measures instead of utility functions (i.e., $u(X) = -\rho(-X)$ satisfying the conditions in Delbaen (2009)).

Theorem 2 (Delbaen (2009), Theorem 4)

Let E be a vector space which is rearrangement invariant and solid, and $\rho : E \rightarrow \mathbb{R}$ be a convex risk measure, then $E \subset L^1$.

Thus, any non-trivial risk measure on L^p , $0 < p < 1$, is not convex and hence in particular, not coherent. The $\mathbb{E}\mathbb{S}$, the classical example of a coherent risk measure, can by construction only be defined on L^1 and not on L^p , $0 < p < 1$, and hence is in line with Theorem 2. However, VaR defined on L^p , $0 < p < 1$, can by Theorem 2 not be convex and hence cannot be coherent. Why

is the above result relevant for finance in general and risk management in particular? Theorem 2 raises the question to what extent can catastrophic events (in particular the ones statistically pointing at infinite mean models) be insured. We will not enter into the more applied economic discussion of this issue, but refer the reader to some relevant literature precisely on this topic; see below. For our purposes, and indeed historically the reason why we include Theorem 2 in this paper, is the key example of the quantitative (so-called Pillar 1) modeling of Operational Risk under Basel II/III; see Chapter 10 in McNeil et al. (2005). In this context, a pivotal publication was the regulatory document Moscadelli (2009). In the latter paper, based on over 40,000 operational risk losses, numerous business lines conformed with an infinite mean model, i.e., $\delta < 1$ in (1). Of course one can discuss the wisdom of allowing such models in our range of possibilities. The same can of course be said for any model with some divergent moment (i.e., all power-tail models (1), $0 < \delta < \infty$) or indeed any distributional model with unbounded support. The point is that, if careful statistical analysis (in the case of Moscadelli (2009) based on *extreme value theory*) statistically points at $0 < \delta < 1$ in (1), then extreme care has to be taken on the risk management conclusions for such portfolios. And this is exactly where Theorem 2 enters. Within the context of Operational Risk, this issue was first raised in Nešlehová et al. (2006). The latter paper gives further references on the topic from the realm of economics and finance.

These early papers have given rise to a whole industry of results on risk aggregation and risk diversification in the presence of catastrophic risks. A key observation underlying these results is the fact that for X and Y independent and identically distributed with distribution function F so that $\bar{F} \in RV_{-\delta}(\infty)$, $0 < \delta < 1$, we have that for α sufficiently close to 1,

$$\text{VaR}_\alpha(X + Y) > \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y). \quad (3)$$

As a consequence, if one interprets $\text{VaR}_\alpha(X)$ as the risk capital needed for holding the position X , diversification benefit (reflected in subadditivity) breaks down. In this example diversification leads to an increase in risk. Specific examples include $\bar{F}(x) = x^{-1/2}$ for $x \geq 1$, i.e., $\delta = 1/2$ (here (3) can be calculated explicitly), and δ -stable distributions with $\delta \in (0, 1)$. In these cases (3) holds for all $0.5 < \alpha < 1$ (see Ibragimov and Walden (2007), Proposition 1). Several of these results were already discussed in Embrechts et al. (2002). For a discussion on the economic consequences of situations where (3) occurs, see for instance Ibragimov et al. (2009, 2011) and the references therein.

Inequality (3) is not only valid under the independence assumption. An example relevant for practice is given in Embrechts et al. (2009); for this we need the following definition.

Definition 2 (Elliptical distribution)

A random vector \mathbf{X} has an elliptical distribution with mean $\mu \in \mathbb{R}^d$ and dispersion matrix Σ in $\mathbb{R}^{d \times d}$, if there exist R, \mathbf{A} and \mathbf{U} satisfying

$$\mathbf{X} \stackrel{d}{=} \mu + R\mathbf{A}\mathbf{U}$$

where

- (a) $R \geq 0$ is a non-negative random variable with distribution function F_R ;

(b) \mathbf{U} is uniformly distributed on the unit sphere $\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = 1\}$ and independent of R ;

(c) $\mathbf{A} \in \mathbb{R}^{d \times d}$ with $\mathbf{A}\mathbf{A}' = \Sigma$.

It is well known (see McNeil et al. (2005), Theorem 6.8) that for (X, Y) elliptically distributed and $\alpha \in (0.5, 1)$,

$$\text{VaR}_\alpha(X + Y) < \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y).$$

This means that VaR_α , $\alpha \in (0.5, 1)$, restricted to the space of elliptical distributions is a coherent risk measure for both finite and infinite mean models. We investigate now the special case where

(a) $\bar{F}_R \in RV_{-\delta}(\infty)$, $\delta \geq 0$,

(b) $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$ for some $\rho \in (0, 1)$,

and we define

$$(\tilde{X}, \tilde{Y}) = (X, Y) | (X \geq 0 \text{ and } Y \geq 0).$$

If $\delta \in (0, 1)$ and α is close to 1, then

$$\text{VaR}_\alpha(\tilde{X} + \tilde{Y}) > \text{VaR}_\alpha(\tilde{X}) + \text{VaR}_\alpha(\tilde{Y}).$$

For the risks \tilde{X} and \tilde{Y} , diversification (measured through VaR) increases the risk although for X and Y it is contrary. This shows that not only the dependence structure and the tail behavior but also the support of the distribution functions have an influence on diversification. In the symmetric, hence two-sided case, large losses of X can be compensated by large gains of Y and vice versa. This results in the subadditivity of VaR_α for α close to 1. The reasoning clearly breaks down in the (\tilde{X}, \tilde{Y}) -case, where diversification, in fact, increases the risk. Similar constructions can be made if one cuts out other parts of the support of (X, Y) in order to obtain (\tilde{X}, \tilde{Y}) ; a construction which in practice often can be made through derivatives or reinsurance, say. For a more mathematical view on aggregation of dependent risks see, e.g., Embrechts et al. (2009) and Degen et al. (2010). A thorough analysis based on the theory of multivariate regular variation is to be found in Embrechts and Mainik (2012). For a discussion in the realm of econometrics, see Daniélsson (2011), Section 4.4. The reader should however be warned that the summary statement at the beginning of Section 4.4.3 of the aforementioned text contains some imprecise statements, the corrections of which are to be found in the more mathematical papers listed above. However, we want to point out that the non-diversification results of VaR are not only an issue for catastrophic events. Superadditivity of VaR can also occur for very skewed risks (typically to be found in credit risk; see McNeil et al. (2005), Example 6.7) or risks with given nice margins, $N(0, 1)$, say, but special dependence structures. The latter can be achieved through a special copula construction; see McNeil et al. (2005), Example 6.22 and Ibragimov and Walden (2007), Theorem 1, for a related result.

4 The third theorem: Sibuya

The single most quoted mathematical concept related to The Crisis is without any doubt the Gaussian- or normal-copula model. In his web-publication, Salmon (2009), the author puts the title “The formula that killed Wall Street” above the formula

$$\Pr [T_A < 1, T_B < 1] = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)), \gamma). \quad (4)$$

The paper was recently reprinted as Salmon (2012) in the American Statistical Society and Royal Statistical Society’s Significance, and this as the winner of the ASA’s Excellence in Statistical Reporting Award for 2010. Formula (4), properly explained, lies at the heart of one of the standard pricing formulas for tranches of Collateralized Debt Obligations, CDOs; see Li (2000). In (4), T_A (respectively T_B) denotes the time to default of company A (respectively B). The left hand side hence denotes the joint probability of default (before the end of period 1) of *both* companies A and B . The right hand side contains as components a bivariate Gaussian distribution Φ_2 with correlation parameter $\gamma \in [-1, 1]$, the quantile function $\Phi^{-1} = \Phi^{\leftarrow}$ of a one-dimensional standard normal distribution function, as well as the marginal survival probability distribution functions $F_A(t) = \mathbb{P}(T_A \leq t)$ and $F_B = \mathbb{P}(T_B \leq t)$ for $t \geq 0$.

The proper interpretation of (4) uses the notion of a 2-dimensional copula C , which is a distribution function on $[0, 1]^2$ with uniformly distributed margins. Given marginal distribution functions F_X and F_Y , one can always define a joint distribution function

$$F(x, y) = C(F_X(x), F_Y(y)) \quad \text{for } (x, y) \in \mathbb{R}^2. \quad (5)$$

Conversely, for any joint distribution function F , there exists a copula C such that (5) holds. For F_X and F_Y continuous, C is unique:

$$C(u, v) = F(F_X^{\leftarrow}(u), F_Y^{\leftarrow}(v)) \quad \text{for } (u, v) \in [0, 1]^2.$$

This forms the content of Sklar’s Theorem, as for instance discussed in McNeil et al. (2005), Theorem 5.3. A version of (5) applies also to the joint survival function $\bar{F}(x, y) = \mathbb{P}(X > x, Y > y)$ of the bivariate random vector (X, Y) with distribution F , margins F_X and F_Y , and tails $\bar{F}_X(x) = 1 - F_X(x)$ and $\bar{F}_Y(y) = 1 - F_Y(y)$. Then there exists again a copula \bar{C} , the *survival copula*, such that

$$\bar{F}(x, y) = \bar{C}(\bar{F}_X(x), \bar{F}_Y(y)) \quad \text{for } (x, y) \in \mathbb{R}^2. \quad (6)$$

In the bivariate case C and \bar{C} are related as follows:

$$\bar{C}(u, v) = u + v - 1 + C(1 - u, 1 - v) \quad \text{for } (u, v) \in [0, 1]^2.$$

In finance, no doubt the most (in-)famous copula model is the Gaussian copula:

$$C_{\Phi, \gamma}(u, v) = \Phi_2(\Phi^{\leftarrow}(u), \Phi^{\leftarrow}(v)) \quad \text{for } (u, v) \in [0, 1]^2,$$

with Φ_2 and Φ^{\leftarrow} as defined earlier. For the Gaussian copula, we can check that

$$C_{\Phi,\gamma}(u,v) = \bar{C}_{\Phi,\gamma}(u,v). \quad (7)$$

Copula theory has taken finance and econometrics by storm, and this from around 1997-8. As explained in Genest et al. (2009), the main paper that triggered this interest in copulas in finance is Embrechts et al. (2002), available as a RiskLab preprint from 1998 onwards. Starting from examples like in Section 5 below, the latter paper highlighted several pitfalls and fallacies in the use of linear correlation in finance in general and quantitative risk management in particular. For a paper summarizing a more historical perspective on the topic of copulas, see Embrechts (2009). With these definitions, (4) reduces to a Gaussian copula model $C_{\Phi,\gamma}$ with correlation coefficient γ applied to the marginal default (or marginal) distribution functions F_A and F_B , i.e.,

$$\mathbb{P}(T_A < 1, T_B < 1) = C_{\Phi,\gamma}(F_A(1), F_B(1)).$$

One easily extends this formula to d companies A_1, \dots, A_d yielding a formula for the joint default probability $\mathbb{P}(T_{A_1} < 1, \dots, T_{A_d} < 1)$. A critical aspect of such joint default modeling in credit risk management, say, is the possibility of high joint default probabilities. It is however a fact that the Gaussian copula model (4) does not allow for such events and consequently may severely underestimate the probabilities of joint default in periods of stress. The reason for this can be found in Theorem 3 below. Before we state the theorem, let us recall a few concepts of joint extremes of random variables; see for instance Beirlant et al. (2004) and McNeil et al. (2005) for the relevant background from extreme value theory. We restrict ourselves to two dimensions for conceptual ease; although most of the results can be extended to general finite dimensions $d \geq 2$.

Suppose $(X, Y) \in \mathbb{R}^2$ is a random vector denoting losses incurred from two separate, but perhaps related investments. The investor would be quite concerned about the possibility of having high losses in both investments together, in other words, he/she would want to know the probability $\mathbb{P}(X > t | Y > t)$ for large thresholds t . In such a context, for a random vector $(X, Y) \in \mathbb{R}^2$ with identically distributed, possibly dependent components and right end-point $x^* = \sup\{t \in \mathbb{R} : \mathbb{P}(X \leq t) < 1\}$, *asymptotic independence* is the property that

$$\lim_{t \uparrow x^*} \mathbb{P}(X > t | Y > t) = \lim_{t \uparrow x^*} \frac{\mathbb{P}(X > t, Y > t)}{\mathbb{P}(Y > t)} = 0. \quad (8)$$

Independent random vectors are trivially asymptotically independent. The property of asymptotic independence can be quite nicely described via copula functions as follows. The *coefficients of upper and lower tail dependence* of a bivariate random vector (X, Y) with distribution function F , margins F_X and F_Y , copula C as in (5) and survival copula \bar{C} as in (6) are defined as

$$\lambda_U = \lim_{u \downarrow 0} \frac{\bar{C}(u, u)}{u} \quad \text{and} \quad \lambda_L = \lim_{u \downarrow 0} \frac{C(u, u)}{u} \quad (9)$$

given that these limits exist. If we assume that $F_X = F_Y$ with common right end-point x^* , then

asymptotic independence of F is equivalent to having $\lambda_U = 0$ since,

$$\lambda_U = \lim_{u \downarrow 0} \frac{\bar{C}(u, u)}{u} = \lim_{t \uparrow x^*} \frac{\bar{C}(\bar{F}_X(t), \bar{F}_X(t))}{\bar{F}_X(t)} = \lim_{t \uparrow x^*} \frac{\mathbb{P}(X > t, Y > t)}{\mathbb{P}(Y > t)} = \lim_{t \uparrow x^*} \mathbb{P}(X > t | Y > t) = 0;$$

more details can be found in Reiss (1989), Chapter 7, and Resnick (2008), Chapter 5. Hence, for asymptotically independent random variables X and Y it is very unlikely that X and Y are large at the same time. Moreover from (7) and (9) we can already conclude that for a Gaussian copula $\lambda_L = \lambda_U$. This leads to the third theorem.

Theorem 3 (Sibuya (1960))

Suppose (X, Y) is a random vector following a bivariate normal distribution with correlation coefficient $\gamma \in [-1, 1)$. Then X and Y are asymptotically independent.

Proof. See Sibuya (1960), Theorem 3 and also McNeil et al. (2005), Example 5.32. □

In other words, what Theorem 3 states is that it is highly unlikely that jointly normal variables are both large together if they have correlation coefficient strictly less than 1; a property one can easily illustrate through simulation. A further consequence of (9) is that asymptotic tail independence as defined above is strictly a *copula property* and hence does not depend on, nor can be influenced by, the marginal distributions. Hence it is relevant to look more in detail at the asymptotic property of (9). In terms of “The formula that killed Wall street” given in (4), if we assume that the one-period survival probabilities for both companies A and B are equal, i.e., $F_A(1) = F_B(1) = u$, then Sibuya’s Theorem, (7) and (9) imply that

$$\begin{aligned} \mathbb{P}(T_A < 1, T_B < 1) &= \Phi_2(\Phi^{\leftarrow}(u), \Phi^{\leftarrow}(u); \gamma) \\ &= C_{\Phi, \gamma}(u, u) = \bar{C}_{\Phi, \gamma}(u, u) = o(u) \quad \text{as } u \rightarrow 0, \end{aligned} \tag{10}$$

for any correlation coefficient $\gamma \in [-1, 1)$. Thus joint default probabilities are of lower order than the individual default probabilities under a Gaussian dependence structure and this is true whatever the marginal distribution F_A and F_B . This phenomenon ignores the empirically observed clustering of defaults in extreme circumstances. Donnelly and Embrechts (2010) lists the drawbacks of a (Gaussian) copula-based model in modeling credit risk which we do not dwell upon further in this paper. Our point of view is that while blind acceptance is not always wise, as has happened with the Gaussian copula model in credit risk modeling, blind rejection is not prudent either. It is necessary to understand where a model may fail or succeed before applying it in practice. Thus, here we dig a little bit deeper into the world of joint tail dependence. The implication of Theorem 3, while quite trivial and obvious for the correlation coefficient $\gamma = 0$, is not quite as intuitive when $\gamma \neq 0$. Let us consider the following example to illustrate how much of the details we miss if we do not consider other concepts of tail dependence.

Example 1

Suppose (X, Y) corresponds to a loss vector having a Gaussian copula dependence with correlation γ and margins $F_X = F_Y = N(2, 1)$. Then $\mathbb{P}(X > 4) = \mathbb{P}(Y > 4) = 0.023$. Now if we

want to calculate the probability of at least one loss being greater than 4 then by the asymptotic independence property we are tempted to estimate

$$\mathbb{P}(\max(X, Y) > 4) = \mathbb{P}(X > 4) + \mathbb{P}(Y > 4) - \mathbb{P}(X > 4, Y > 4) \approx \mathbb{P}(X > 4) + \mathbb{P}(Y > 4) = 0.046,$$

for any $\gamma < 1$. But actually calculating the probability, we see that

$$\mathbb{P}(\max(X, Y) > 4) = \begin{cases} 0.045 & \text{if } \gamma = 0, \\ 0.041 & \text{if } \gamma = 0.5, \\ 0.032 & \text{if } \gamma = 0.9. \end{cases} \quad (11)$$

Hence, for $\gamma = 0$ the true probability of 0.045 and the approximate probability 0.046 are very close. However, for γ close to one the approximate probability differs significantly from the true probability.

After all asymptotic independence does not mean that there is no dependence between X and Y if both take high values. Therefore we will introduce some well known dependence concepts. Analogous to regularly varying functions at ∞ as defined in (1), we define regularly varying functions at 0. A measurable function $f : (0, \infty) \rightarrow (0, \infty)$ is regularly varying at 0 with index δ if $\lim_{t \rightarrow 0} f(tx)/f(t) \rightarrow x^\delta$ for any $x > 0$ and we write $f \in RV_\delta(0) = SV(0)$. Now, following Hua and Joe (2011), we introduce the definition of *tail order*.

Definition 3 (Tail order)

Let F be a bivariate distribution with copula C and survival copula \bar{C} as given in (5) and (6), respectively. If there exist $\kappa_L > 0$ and $L \in SV(0)$ such that

$$C(u, u) \sim u^{\kappa_L} L(u) \quad \text{as } u \rightarrow 0, \quad (12)$$

then we call κ_L the lower tail order of F . In a similar manner, if there exist $\kappa_U > 0$ and $L \in SV(0)$ such that

$$\bar{C}(u, u) \sim u^{\kappa_U} L(u) \quad \text{as } u \rightarrow 0, \quad (13)$$

then we call κ_U the upper tail order of F .

If the upper tail order $\kappa_U > 1$, then X and Y with distribution F are asymptotically independent. Under the assumption that X and Y are unit Fréchet distributed, i.e., $F_X(x) = F_Y(x) = \exp(-1/x)$ for $x \geq 0$, the definition of upper tail order is equivalent to the definition of *coefficient of tail dependence* as defined in Ledford and Tawn (1996). In essence, the latter paper introduced the ideas which have been further investigated as *tail order* and the more general concept of *hidden regular variation* which is discussed later.

Further investigation into the Gaussian dependence structure reveals that for $\gamma \in (-1, 1)$, the exact rate of the joint copula tail is

$$C_{\Phi, \gamma}(u, u) = \bar{C}_{\Phi, \gamma}(u, u) \sim u^{\kappa} L(u) \quad \text{as } u \rightarrow 0, \quad (14)$$

where $\kappa = \kappa_U = \kappa_L = \frac{2}{\gamma+1}$ are the tail orders and $L \in SV(0)$ with

$$L(u) \sim O((- \log u)^{-\gamma/(\gamma+1)}) \quad \text{as } u \rightarrow 0;$$

see Reiss (1989), Chapter 7 and Ledford and Tawn (1996). Clearly for independent X and Y ($\gamma = 0$) we have $\kappa = 2$ and $L(u) = 1$. Equation (14) gives us the exact order of the joint tail. Thus, although (X, Y) exhibits asymptotic independence for $\gamma \neq 0$, there is still some kind of intermediate order dependence in the extremes, which is not neglectable. The notion of tail order is very closely related to the notion of *hidden regular variation* given in Resnick (2002) which describes intermediate order dependence in a more general framework; in a certain sense one could speak of hidden dependence in models exhibiting tail independence.

In order to talk about hidden regular variation we resort to multivariate regular variation (still restricting to two dimensions). The connection between extreme value theory and regular variation has been emphasized time and again. The space where our random variables (risks) lie is the non-negative orthant $[0, \infty)^2$. For topological convenience this space is compactified and then $\{(0, 0)\}$ is removed from it to get $\mathbb{E} = [0, \infty]^2 \setminus \{(0, 0)\}$. Moreover, we work on the space $\mathbb{E}_0 = (0, \infty]^2 = \mathbb{E} \setminus \{\text{the axes}\}$. In the space \mathbb{E} , sets bounded away from $\{(0, 0)\}$ are relatively compact, hence we have the already well-developed theory of (vague) convergence, denoted by $\xrightarrow{\nu}$, of (Radon) measures in order to find asymptotic limits of probability measures for such tail sets; for more details see Resnick (2007).

Definition 4 (Multivariate and hidden regular variation)

Suppose (X, Y) is a non-negative random vector with joint distribution function F .

- (a) We say that (X, Y) is multivariate regularly varying on \mathbb{E} , if there exist a function $b(t) \uparrow \infty$ as $t \rightarrow \infty$ and a Radon measure $\nu \neq 0$ on the Borel sets of \mathbb{E} such that as $t \rightarrow \infty$,

$$tF(b(t) \cdot) = t\mathbb{P}\left(\frac{(X, Y)}{b(t)} \in \cdot\right) \xrightarrow{\nu} \nu(\cdot)$$

in the space of Radon measures on \mathbb{E} . We write $F \in MRV(b, \nu)$ and $(X, Y) \in MRV(b, \nu)$.

- (b) Let $(X, Y) \in MRV(b, \nu)$. We say that (X, Y) exhibits hidden regular variation on \mathbb{E}_0 , if there exist a function $b_0(t) \uparrow \infty$ as $t \rightarrow \infty$ with $\lim_{t \rightarrow \infty} b(t)/b_0(t) = \infty$ and a Radon measure $\nu_0 \neq 0$ on the Borel sets of \mathbb{E}_0 such that as $t \rightarrow \infty$,

$$tF(b_0(t) \cdot) = t\mathbb{P}\left(\frac{(X, Y)}{b_0(t)} \in \cdot\right) \xrightarrow{\nu_0} \nu_0(\cdot)$$

in the space of Radon measures on \mathbb{E}_0 . We write $F \in HRV(b, b_0, \nu, \nu_0)$ and $(X, Y) \in HRV(b, b_0, \nu, \nu_0)$.

If $F \in MRV(b, \nu)$ then $b \in RV_{1/\delta}(\infty)$ for some $\delta > 0$ and δ is called the *index of regular variation*. If F exhibits additionally hidden regular variation then $b_0 \in RV_{1/\delta_0}(\infty)$ for some $\delta_0 \geq \delta$. The index δ_0 is called the *index of hidden regular variation*. To illustrate the notion of hidden regular variation we want to present an example.

Example 2

Suppose X and Y are independent and identically distributed random variables (alternatively, think of a risk or loss vector) with distribution function $F_X(x) = 1 - \frac{1}{x}$ for $x \geq 1$; for $b(t) = t, t \geq 0$, and $x, y > 0$, we have that

$$t\mathbb{P}\left(\frac{(X,Y)}{b(t)} \in [(0,0), (x,y)]^c\right) = \frac{1}{x} + \frac{1}{y} - \frac{1}{txy} \xrightarrow{t \rightarrow \infty} \frac{1}{x} + \frac{1}{y} =: \mathbf{v}([(0,0), (x,y)]^c)$$

and

$$t\mathbb{P}\left(\frac{(X,Y)}{b(t)} \in (x, \infty] \times (y, \infty]\right) \xrightarrow{t \rightarrow \infty} 0 = \mathbf{v}((x, \infty] \times (y, \infty]). \quad (15)$$

Thus, $(X, Y) \in MRV(b, \mathbf{v})$ with index 1 and the limit measure \mathbf{v} concentrates on the axes of \mathbb{E} , i.e., \mathbf{v} is non-zero only on $\{0\} \times (0, \infty]$ and $(0, \infty] \times \{0\}$. But clearly there is some mass on $\mathbb{E}_0 = \mathbb{E} \setminus \{\text{the axes}\}$ which vanished with the normalization $b(t) = t$. This is what hidden regular variation is designed to capture. For $b_0(t) = t^{1/2}$,

$$t\mathbb{P}\left(\frac{(X,Y)}{b_0(t)} \in (x, \infty] \times (y, \infty]\right) = \frac{1}{xy} =: \mathbf{v}_0((x, \infty] \times (y, \infty]). \quad (16)$$

Hence, the index of hidden regular variation is $\delta_0 = 2$. In this example (X, Y) has independence copula $C_{\perp}(u, u) = u^2$ with upper tail order $\kappa_U = 2$. This means $\kappa_U = \delta_0$. Although (15) suggests that X and Y can not be large at the same time, the different normalization in (16) shows that there is still (positive but a low) probability that this can happen. Hidden regular variation reflects the extremal behavior on $\mathbb{E} \setminus \{\text{the axes}\}$, yielding information which can be lost by classical multivariate regular variation. \square

Hence, what hidden regular variation captures is a part of the measure that has escaped due to the stronger normalization under multivariate regular variation. This phenomenon is observed in Example 2, but it holds more generally. We can also observe such phenomenon if we use an equally distributed bivariate distribution with Gaussian copula with correlation parameter γ and Pareto(1) margins. If we look at tail dependence on the non-negative orthant, with a normalization $b(t) = t$, we get the same limit measure \mathbf{v} on \mathbb{E} as in Example 2, i.e., the limit measure lies on the axes. But if $\gamma < 1$ with a softer normalization $b_0(t) \in RV_{(\gamma+1)/2}$ we get a different limit measure \mathbf{v}_0 on \mathbb{E}_0 as in Example 2 (for $\gamma = 0$). This is visualized in Figure 1, where we look at empirical estimates of the limiting densities of \mathbf{v} and \mathbf{v}_0 . Since \mathbf{v}_0 has no closed form if $\gamma \neq 0$, we simulated the density of \mathbf{v}_0 . The graphs portray some bias resulting from approximation in the simulations. The top two figures show the densities of \mathbf{v} and \mathbf{v}_0 , respectively restricted to $[0.1, 0.3]^2$ for $\gamma = 0.1$. The bottom figures do the same for the case $\gamma = 0.9$. Whereas, there is hardly any mass in the left figures (in reality $\mathbf{v}|_{[0.1, 0.3]^2} = 0$ so that there is no mass on $[0.1, 0.3]^2$), the right figures show different concentrations of mass for $\gamma = 0.1$ and $\gamma = 0.9$. Both $\gamma = 0.1$ and $\gamma = 0.9$ exhibit hidden regular variation but with different limit measures \mathbf{v}_0 and with different normalizations: $b_0(t) \in RV_{0.55}$ for $\gamma = 0.1$ and $b_0(t) \in RV_{0.95}$ for $\gamma = 0.9$. Thus the model with $\gamma = 0.1$ requires a softer normalization than the model with

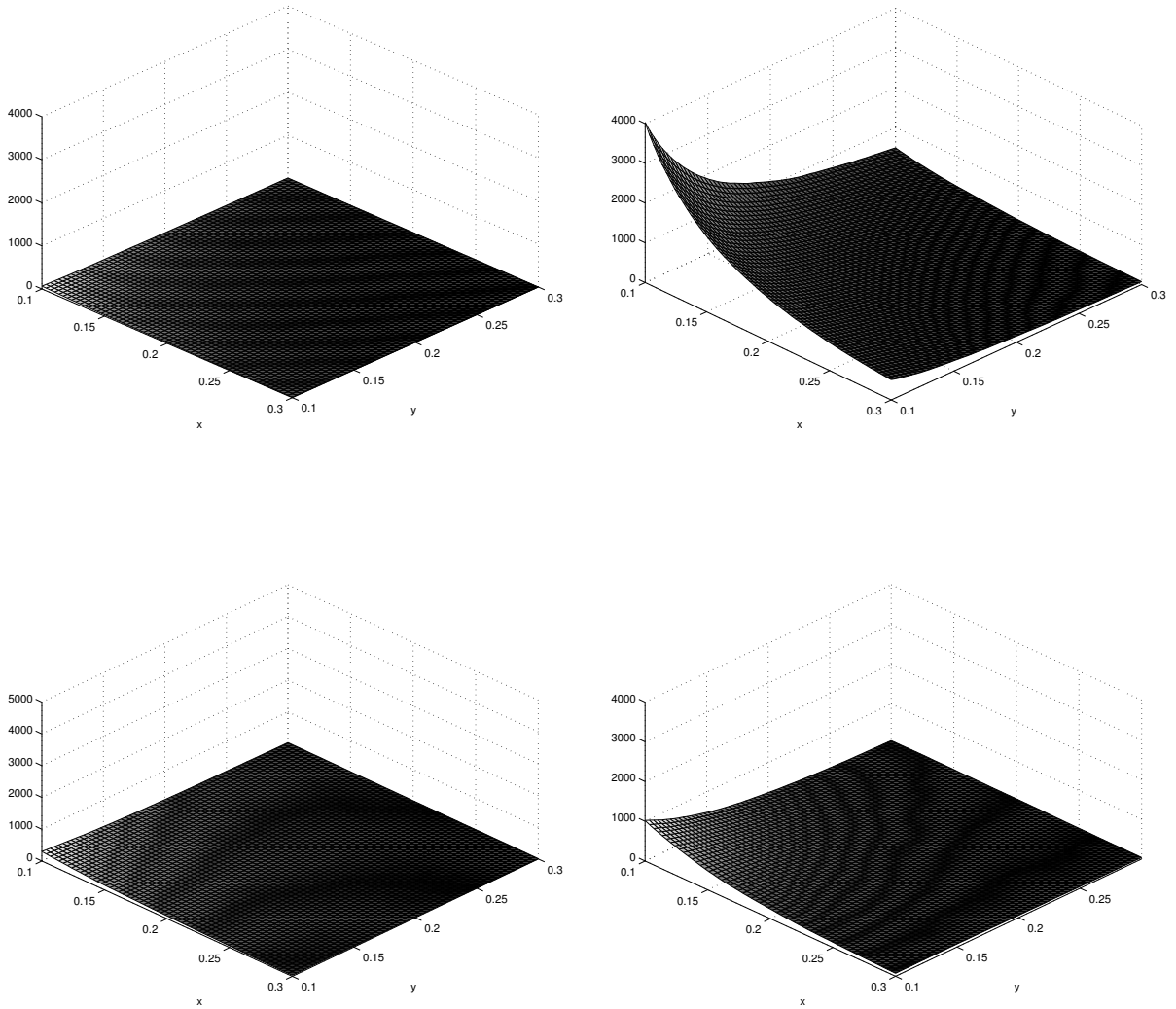


Figure 1: Empirical density v for Pareto(1) margins with Gaussian copula restricted to $[0.1, 0.3]^2$ (left) and v_0 restricted to $[0.1, 0.3]^2$ (right) with $\gamma = 0.1$ (top) and $\gamma = 0.9$ (bottom) from 2000,000 simulations for both distributions.

$\gamma = 0.9$, so that the hidden limit measures do not blow up in compact intervals, as also observed empirically in the right two figures.

The following proposition relates now hidden regular variation to tail orders and asymptotic independence.

Proposition 1

Let $F \in HRV(b, b_0, v, v_0)$ with index of regular variation $\delta > 0$ and hidden regular variation

$\delta_0 \geq \delta$. Moreover, the margins F_X and F_Y have the same distribution and are continuous. Then F has upper tail order $\kappa_U = \delta_0/\delta$. If $\delta_0 > \delta$ then X and Y are asymptotically independent.

Now coming back to the Gaussian dependence structure, we know that the Gaussian copula $C_{\Phi,\gamma}$ admits asymptotic independence as exhibited in Theorem 3 and furthermore we had the upper (and lower) tail order given by $\kappa = \frac{2}{\gamma+1}$. But something more is true in general. If F is a bivariate distribution function with margins which are identical Pareto distributions with parameter $\delta > 0$ and Gaussian copula $C_{\Phi,\gamma}$ with $-1 < \gamma < 1$, then in fact F is multivariate regularly varying with index δ and exhibits hidden regular variation with index $\delta_0 = \frac{2\delta}{\gamma+1}$; see Reiss (1989), Example 7.2.7 for the exact calculations. Therefore, the tail order $\kappa = \frac{2}{\gamma+1}$ is in fact a consequence of Proposition 1 here.

We conclude that even within the framework of asymptotic independence where joint extremes are quite unlikely, much more can be said about the structure of the dependence in the tails. We want to stress this idea since, while considering models where joint extremes may occur (e.g., pricing CDO tranches), there exist many models which exhibit asymptotic independence like the Gaussian copula, but with a dependence structure which exhibits hidden regular variation. In such cases, the underlying hidden regular variation index may be more conducive to the study of joint extremes. It is therefore important, for models used in practice, to investigate the tail behavior in the extremes also using hidden regular variation or tail order (a consequence of hidden regular variation via Proposition 1).

5 The fourth theorem: Fréchet-Höfding

Since the mid-nineties, RiskLab has on several occasions been contacted by practitioners with questions of the following type (as example): “In a risk management context, simulate from a bivariate model with marginal lognormal distributions $\text{LN}(0,1)$, $\text{LN}(0,4)$ and linear correlation 70%”. At first sight, this seems a totally harmless question, examples of which abound in actuarial science, economics and finance. Somewhat more generally, the question reformulates as: “Find a model for a multivariate vector of random variables (X_1, \dots, X_d) with given marginal distribution functions (F_1, \dots, F_d) and some dependence structure”. Formulated in this generality, the first question coming to mind should be: “Does such a multivariate model exist?” And indeed, the answer to the lognormal question above is “No!”. Given $X \sim F_X = \text{LN}(0,1)$ and $Y \sim F_Y = \text{LN}(0,4)$, the maximum linear correlation across *all* bivariate models with these marginal distributions is 0.6658. And since $0.7 > 0.6658$, the answer is: “Such a model does not exist!”. Also, if the condition of 70% were to be lowered to 60%, then there are infinitely many solutions, hence we enter the realm of *Model Uncertainty*.

The solution to the above problem is based on results of Wassily Höfding and Maurice Fréchet in the 1940s. Before we can formulate these results, we need some notation. Suppose X and Y are non-degenerate random variables with finite second moment and joint distribution function

F , then the linear (also called Pearson) correlation γ between X and Y is defined as

$$\gamma = \gamma(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \in [-1, 1].$$

An alternative representation of the correlation coefficient is

$$\gamma(X, Y) = \frac{\int_0^1 \int_0^1 [C(F_X(x), F_Y(y)) - F_X(x)F_Y(y)] dx dy}{\sqrt{\text{Var}(X) \text{Var}(Y)}}, \quad (17)$$

here C is the copula in (5); this result is referred to as Höfddings Lemma, see McNeil et al. (2005), Lemma 5.24. The representation (17) nicely exhibits the influence on $\gamma(X, Y)$ of both the marginal distributions and the copula. Also recall that if there exist an increasing function Ψ_1 and an increasing (decreasing) Ψ_2 and a random variable Z so that $X = \Psi_1(Z)$ and $Y = \Psi_2(Z)$ almost surely then (X, Y) is called *co-(counter-)monotonic*. The copula for a comonotonic pair of random variables, the *comonotonic copula*, is given by

$$C_{\max}(u, v) = \min(u, v) \quad \text{for } (u, v) \in [0, 1]^2$$

and analogously the *countermonotonic copula* is given by

$$C_{\min}(u, v) = \max(u + v - 1, 0) \quad \text{for } (u, v) \in [0, 1]^2.$$

Both co- and countermonotonicity reflect a strong (even functional) dependence. The former can easily be generalized to arbitrary dimensions; countermonotonicity though is a two-dimensional concept.

Theorem 4 (Höfdding (1940, 1941), Fréchet (1957))

Let (X, Y) be a bivariate random vector with finite variances, non-degenerate marginal distribution functions F_X and F_Y and an unspecified joint distribution function F . The following statements holds.

1. The attainable correlations from any joint model F with the above specifications form a closed interval

$$[\gamma_{\min}, \gamma_{\max}] \subset [-1, 1]$$

with $-1 \leq \gamma_{\min} < 0 < \gamma_{\max} \leq 1$.

2. The minimum correlation γ_{\min} is attained if and only if (X, Y) is countermonotonic; the maximum correlation γ_{\max} if and only if (X, Y) is comonotonic.
3. $\gamma_{\min} = -1$ if and only if X and $-Y$ are of the same type; $\gamma_{\max} = 1$ if and only if X and Y are of the same type.

Proof. See McNeil et al. (2005), p. 205. □

Though the result is well known to probabilists and statisticians, it was mainly its explicit appearance in Embrechts et al. (2002) that brought it to the attention of the more applied risk management community. In our experience, Theorem 4, though being truly fundamental to linear correlation based quantitative risk management, still needs broader knowledge. For a full discussion, see McNeil et al. (2005), Section 5.2. Often, especially in credit risk management, stress testing of credit risk portfolios is performed by “moving the correlations up to 1”. Theorem 4 warns us that, if correlation is to be interpreted as linear correlation, one has to be careful by doing so, not to move the stressed model out of the range of existing models! Also note that, from Theorem 4 it follows that $\gamma = \gamma_{\max}$ corresponds to comonotonicity of the risk vector (X, Y) , hence strong (monotone) functional dependence, and

$$\gamma_{\max} = \gamma_{\max}(F_X, F_Y) = \frac{\int_0^1 \int_0^1 [\min(F_X(x), F_Y(y)) - F_X(x)F_Y(y)] dx dy}{\sqrt{\text{Var}(X)\text{Var}(Y)}}. \quad (18)$$

Similarly, $\gamma = \gamma_{\min}$ corresponds to countermonotonicity of the risk vector (X, Y) and

$$\gamma_{\min} = \gamma_{\min}(F_X, F_Y) = \frac{\int_0^1 \int_0^1 [\max(F_X(x) + F_Y(y) - 1, 0) - F_X(x)F_Y(y)] dx dy}{\sqrt{\text{Var}(X)\text{Var}(Y)}}. \quad (19)$$

One easily shows that for $X \sim \text{LN}(0, \sigma_1^2)$, $Y \sim \text{LN}(0, \sigma_2^2)$, $\sigma_1, \sigma_2 > 0$,

$$\begin{aligned} \gamma_{\max}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) &= \frac{e^{\sigma_1 \sigma_2} - 1}{\sqrt{(e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}}, \\ \gamma_{\min}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) &= \frac{e^{-\sigma_1 \sigma_2} - 1}{\sqrt{(e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}}; \end{aligned}$$

see McNeil et al. (2005), Example 5.26. Thus, for the lognormal example, maximum (as well as minimum) correlation can be made arbitrarily small by increasing one of σ_1^2 and σ_2^2 , while keeping the other one fixed, i.e.

$$\lim_{\sigma_1 \rightarrow \infty} \gamma_{\max}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) = \lim_{\sigma_1 \rightarrow \infty} \gamma_{\min}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) = 0,$$

while $\lim_{|\sigma_1 - \sigma_2| \rightarrow 0} \gamma_{\max}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) = 1$. Near perfect negative correlation $\gamma_{\min}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) \rightarrow -1$ is achieved if both σ_1^2 and σ_2^2 are close to zero. Note that if σ_1^2 and σ_2^2 tend to zero then both X and Y become more and more symmetric around 1, so that $X - 1$ and $-(Y - 1)$ are approximately equally distributed, which relates to Theorem 4, making X and $-Y$ being of the same type; thus $\gamma_{\min}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) \sim -1$. Similarly we obtain $\gamma_{\max}(\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2)) \sim 1$ if σ_1^2, σ_2^2 tend to zero. This example also clearly shows that one has to be careful with statements like “small (linear) correlation implies close to independence” and this especially in very skewed portfolios; indeed for σ_1 increasing and σ_2 small,

$\text{LN}(0, \sigma_1^2)$ and $\text{LN}(0, \sigma_2^2)$ have very different skewness.

In Table 1 we provide a list of maximum and minimum attainable correlations for some bivariate distributions which have marginal distributions F_X and F_Y belonging to the same parametric family. Since univariate normal distributions are symmetric and any two of them are always of the same type, the first result in the table is immediate.

F_X, F_Y	γ_{\max}	γ_{\min}
$\text{N}(0, \sigma_1^2), \text{N}(0, \sigma_2^2), \sigma_1, \sigma_2 > 0$	1	-1
$\text{LN}(0, \sigma_1^2), \text{LN}(0, \sigma_2^2), \sigma_1, \sigma_2 > 0$	$\frac{e^{\sigma_1 \sigma_2} - 1}{\sqrt{(e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}}$	$\frac{e^{-\sigma_1 \sigma_2} - 1}{\sqrt{(e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}}$
$\text{Pareto}(\alpha), \text{Pareto}(\beta), \alpha, \beta > 2$	$\frac{\sqrt{\alpha\beta(\alpha-2)(\beta-2)}}{\alpha\beta - \alpha - \beta}$	$\frac{\sqrt{(\alpha-2)(\beta-2)} \left((\alpha-1)(\beta-1) \text{Beta}(1 - \frac{1}{\alpha}, 1 - \frac{1}{\beta}) - \alpha\beta \right)}{\sqrt{\alpha\beta}}$
$\text{Beta}(1, 1), \text{Beta}(\alpha, 1), \alpha > 0$	$\frac{\sqrt{3\alpha(\alpha+2)}}{(2\alpha+1)}$	$-\frac{\sqrt{3\alpha(\alpha+2)}}{(2\alpha+1)}$

Table 1: Table of $\gamma_{\max}(F_X, F_Y)$ and $\gamma_{\min}(F_X, F_Y)$ for different pairs of marginal distributions F_X and F_Y .

In most cases however, analytic expressions for γ_{\min} or γ_{\max} are not available and one has to resort to numerical evaluations of (18) and (19). Figure 2 contains as an example the attainable values of the correlation γ for the Gamma distribution, $\Gamma(\alpha, \beta)$ for $\alpha, \beta > 0$ which is quite often encountered in insurance mathematics and credit risk management. The scale parameter β is fixed at $\beta = 1$ for both marginals; indeed we have that

$$\gamma_{\min/\max}(\Gamma(1, 1), \Gamma(\alpha, 1)) = \gamma_{\min/\max}(\Gamma(1, \beta_1), \Gamma(\alpha, \beta_2)),$$

for any $\alpha > 0$ and $\beta_1, \beta_2 > 0$, thus β_1, β_2 have no influence on the correlation bounds. Note further that, specifically for the case where both X and Y are identically distributed as $\Gamma(1, 1) \equiv \text{Exp}(1)$, we have that

$$\gamma_{\min}(\Gamma(1, 1), \Gamma(1, 1)) = 1 - \frac{\pi^2}{6} \quad \text{and} \quad \gamma_{\max}(\Gamma(1, 1), \Gamma(1, 1)) = 1.$$

Interestingly, this result is closely related to the Pareto case presented in Nešlehová et al. (2006) (see Table 1 and Figure 3) where

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \gamma_{\min}(\text{Pareto}(\alpha), \text{Pareto}(\alpha)) &= 1 - \frac{\pi^2}{6} = \gamma_{\min}(\Gamma(1, 1), \Gamma(1, 1)), \\ \lim_{\alpha \rightarrow \infty} \gamma_{\max}(\text{Pareto}(\alpha), \text{Pareto}(\alpha)) &= 1 = \gamma_{\max}(\Gamma(1, 1), \Gamma(1, 1)). \end{aligned} \quad (20)$$

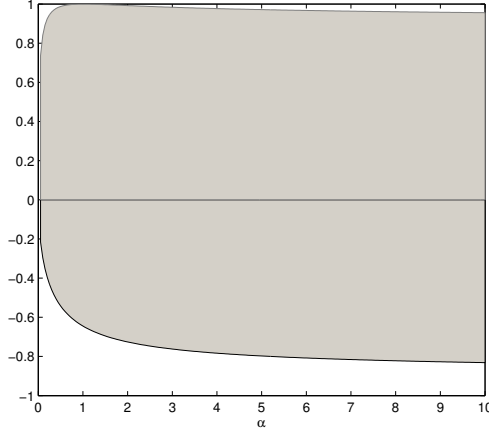


Figure 2: The shaded portion gives the attainable correlations for $F_X \sim \Gamma(1, 1)$ and $F_Y \sim \Gamma(\alpha, 1)$ depending on α .

This can be explained by the following argument. Let X_α denote a Pareto distributed random variable with distribution function $F_\alpha(x) = 1 - x^{-\alpha}$ for $x \geq 1$ and X_{Exp} be an exponentially distributed random variable with distribution function $F_{\text{Exp}}(x) = 1 - \exp(-x)$ for $x \geq 0$. Then for any $0 < u < 1$,

$$\lim_{\alpha \rightarrow \infty} \frac{F_\alpha^{\leftarrow \prime}(u)}{\sqrt{\text{Var}(X_\alpha)}} = \frac{F_{\text{Exp}}^{\leftarrow \prime}(u)}{\sqrt{\text{Var}(X_{\text{Exp}})}},$$

where the prime means that we take the derivative. Therefore the integrands in (18) (and (19)) are approximately the same if both the margins are either F_{Exp} or F_α with very high value of α and hence, (20) follows. This example further highlights that the linear (Pearson) correlation coefficient, while good in detecting linear dependence, might not perform so well to assess other forms of dependence; see also the monograph of Drouot and Kotz (2001) on correlation and dependence where many more examples are to be found. If, however, correlation stands for rank correlation (Kendall's tau, Spearman's rho), then the results are quite different. *Kendall's tau* γ_τ and *Spearman's rho* γ_S have the copula representations

$$\gamma_\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad \text{and} \quad \gamma_S(X, Y) = 12 \int_0^1 \int_0^1 [C(u, v) - uv] dudv.$$

They do not depend on the marginal distributions in contrast to the linear correlation in (17). Let $\lambda \in [-1, 1]$, define the joint distribution of (X, Y) as

$$F(x, y) = \frac{1 + \lambda}{2} C_{\max}(F_X(x), F_Y(y)) + \frac{1 - \lambda}{2} C_{\min}(F_X(x), F_Y(y)).$$

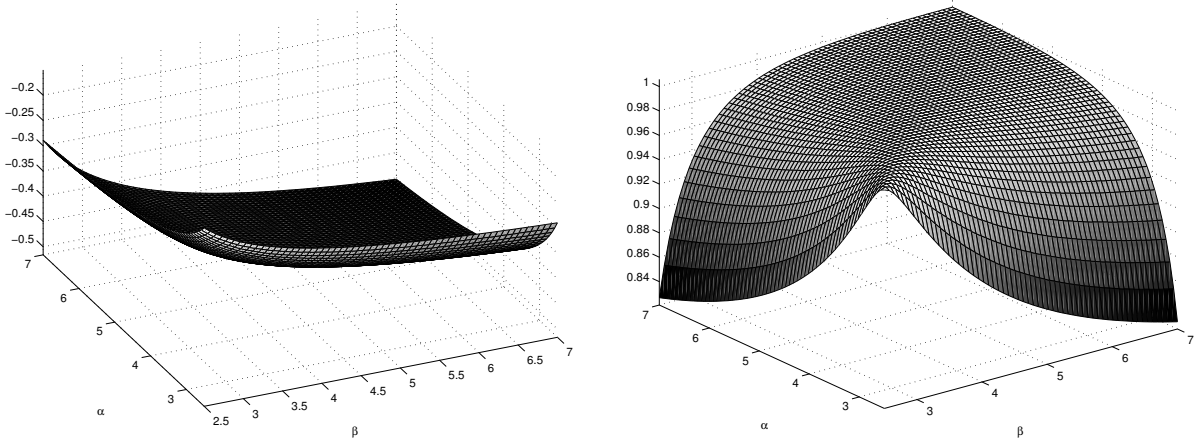


Figure 3: The left figure shows $\gamma_{\min}(\text{Pareto}(\alpha), \text{Pareto}(\beta))$ for $(\alpha, \beta) \in [2.5, 7]^2$ and the right figure shows $\gamma_{\max}(\text{Pareto}(\alpha), \text{Pareto}(\beta))$ for $(\alpha, \beta) \in [2.5, 7]^2$.

One easily obtains that

$$\gamma_{\tau}(X, Y) = \gamma_S(X, Y) = \lambda.$$

Thus we can attain any correlation (in the sense of Spearman's rho or Kendall's tau) given arbitrary marginal distributions. So also in the introductory LN-example from this section, if we want bivariate distributions with margins $\text{LN}(0,1)$, $\text{LN}(0,4)$ and given rank correlation $\gamma_{\tau}(X, Y) = 0.7$ or $\gamma_S(X, Y) = 0.7$ we always have a solution. This is the good news and also a main reason that in quantitative risk management, one should consider these alternative correlation coefficients more seriously. The bad news however is that typically, infinitely many solutions to the above question exist, hence leading to *Model Uncertainty*. There exists a huge literature on calculating bounds for risk measures on financial and insurance positions where only partial information on the marginal distributions and/or the dependence structure (the copula) is given. The interested reader is advised to search for (numerous) publications by Ludger Rüschendorf and Giovanni Puccetti; see Embrechts and Puccetti (2010) for a start.

In Theorem 4 we see that for two random variables with fixed marginal distributions, the maximum (respectively minimum) correlation is attained when the joint distribution is *comonotonic* (respectively *countermonotonic*). A question that naturally follows is whether there is an extension of Theorem 4 for vectors \mathbf{X}, \mathbf{Y} in \mathbb{R}^d which represent vectors of losses for different lines or policies. Unfortunately, there is no general notion of *comonotonicity* in this case, and for instance, $\min(C_{\mathbf{X}}(u_1, \dots, u_d), C_{\mathbf{Y}}(v_1, \dots, v_d))$ is not necessarily a $2d$ -dimensional copula where $C_{\mathbf{X}}$ and $C_{\mathbf{Y}}$ are the copulas of \mathbf{X} and \mathbf{Y} . Hence, the case $d > 1$ is clearly not as straightforward as $d = 1$. There have been several attempts at defining multivariate comonotonicity; see Puccetti

and Scarsini (2010) for an overview on different definitions and their pros and cons. We present below a few of the most common examples. For this, let \mathbf{X}, \mathbf{Y} be random vectors in \mathbb{R}^d with distributions $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$.

- (a) (\mathbf{X}, \mathbf{Y}) is *s*(tronly)-comonotonic if there exist a random variable Z and increasing functions $f_1, \dots, f_d, g_1, \dots, g_d$ such that $(\mathbf{X}, \mathbf{Y}) \stackrel{d}{=} ((f_1(Z), \dots, f_d(Z)), (g_1(Z), \dots, g_d(Z)))$.
- (b) (\mathbf{X}, \mathbf{Y}) is π -comonotonic if there exist a random vector $\mathbf{Z} = (Z_1, \dots, Z_d)$ and increasing functions $f_1, \dots, f_d, g_1, \dots, g_d$ such that $(\mathbf{X}, \mathbf{Y}) \stackrel{d}{=} ((f_1(Z_1), \dots, f_d(Z_d)), (g_1(Z_1), \dots, g_d(Z_d)))$.
- (c) (\mathbf{X}, \mathbf{Y}) is *c*-comonotonic if $\sup\{\mathbb{E}(\langle \tilde{\mathbf{X}}, \tilde{\mathbf{Y}} \rangle) : \tilde{\mathbf{X}} \sim F_{\mathbf{X}}, \tilde{\mathbf{Y}} \sim F_{\mathbf{Y}}\} < \infty$ and

$$\mathbb{E}(\langle \mathbf{X}, \mathbf{Y} \rangle) = \sup\{\mathbb{E}(\langle \tilde{\mathbf{X}}, \tilde{\mathbf{Y}} \rangle) : \tilde{\mathbf{X}} \sim F_{\mathbf{X}}, \tilde{\mathbf{Y}} \sim F_{\mathbf{Y}}\},$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product.

- (d) (\mathbf{X}, \mathbf{Y}) is μ -comonotonic if μ is a probability measure on \mathbb{R}^d that vanishes on Borel subsets of Hausdorff dimension $d - 1$ and there exists a random vector $\mathbf{V} \sim \mu$ such that

$$\mathbf{V} \in \arg \max_{\tilde{\mathbf{V}}} \{\mathbb{E}(\langle \mathbf{X}, \tilde{\mathbf{V}} \rangle) : \tilde{\mathbf{V}} \sim \mu\} \quad \text{and} \quad \mathbf{V} \in \arg \max_{\tilde{\mathbf{V}}} \{\mathbb{E}(\langle \mathbf{Y}, \tilde{\mathbf{V}} \rangle) : \tilde{\mathbf{V}} \sim \mu\}.$$

In the one-dimensional case $d = 1$ the above definitions (a)-(d) of comonotonicity are equivalent to the usual definition of comonotonicity given earlier. In particular, the definition in (c) is a conclusion of Theorem 4. However, in the multivariate case $d > 1$, (a) represents the strongest version of comonotonicity. The notions of comonotonicity get weaker, in fact strictly weaker from (a) to (d) (only from (c) to (d) we require additional assumptions). Are there further differences between the cases $d = 1$ to $d > 1$? For example when $d = 1$ and given any pair of marginal distributions F_X and F_Y in \mathbb{R} there exists a bivariate comonotonic random vector with margins F_X, F_Y . Unfortunately when $d > 1$, given any pair of marginal distributions $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$ in \mathbb{R}^d there does not necessarily exist a *s*- (π -)comonotonic pair of random vectors with margins $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$. On the other hand, given margins $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$ there always exists a *c*- (μ -)comonotonic pair of random vectors with these margins. One common element of all four definitions of comonotonicity mentioned here is that if \mathbf{X} and \mathbf{Y} have the same distribution and (\mathbf{X}, \mathbf{Y}) is *s*- (π -, *c*- or μ -)comonotone then $\mathbf{X} = \mathbf{Y}$ almost surely. Besides, if $\mathbb{E}(\mathbf{X}) = \mathbb{E}(\mathbf{Y}) = \mathbf{0}$ and $\mathbb{E}(\mathbf{X}\mathbf{X}') = \mathbb{E}(\mathbf{Y}\mathbf{Y}') = I_d$ (the d -dimensional identity matrix), then the set $\{\mathbb{E}(\langle \tilde{\mathbf{X}}, \tilde{\mathbf{Y}}' \rangle) : \tilde{\mathbf{X}} \sim F_{\mathbf{X}}, \tilde{\mathbf{Y}} \sim F_{\mathbf{Y}}\} \subset \mathbb{R}^{d \times d}$ is a compact and convex set and for $d = 1$ it is equal to $[\gamma_{\min}(F_X, F_Y), \gamma_{\max}(F_X, F_Y)]$ (see Villani (2003)).

The most commonly used definition of multivariate comonotonicity is that of μ -comonotonicity, which plays quite an important role in the study of risk measures, in particular, in the context of multivariate *law-invariant convex* and *law-invariant coherent* risk measures. A multivariate convex (or coherent) risk measure $\rho : E^d \rightarrow \mathbb{R}$ is defined as the one-dimensional convex (or coherent) risk measure of Definition 1, where $\mathbf{X} \leq \mathbf{Y}$ a.s. is to be understood componentwise and in the definition of translation-invariance η is replaced by ηe_i where e_i is the i^{th} -unit vector.

Law-invariance means that $\rho(\mathbf{X}) = \rho(\mathbf{Y})$ if $F_{\mathbf{X}} = F_{\mathbf{Y}}$. As a consequence of Meilijson and Nadas (1979) for $d = 1$, $X, Y \in L^1$ and for all law invariant convex risk measures ρ on L^1

$$\rho(F_X^{\leftarrow}(U) + F_Y^{\leftarrow}(U)) = \sup\{\rho(\tilde{X} + \tilde{Y}) : \tilde{X} \sim F_X, \tilde{Y} \sim F_Y\},$$

where U is uniformly distributed on $(0, 1)$. Thus, the comonotonic pair $(F_X^{\leftarrow}(U), F_Y^{\leftarrow}(U))$ is the *worst case couple* for any law invariant convex risk measure. In the multivariate setting unfortunately there does not exist a similar result; in particular there does not exist one worst case couple for any law invariant convex risk measure. The concept of multivariate *max-correlation risk measures* was defined in Rüschemdorf (2006) as

$$\rho_{\mathbf{Z}}(\mathbf{X}) = \sup\{\mathbb{E}(\langle \tilde{\mathbf{X}}, \mathbf{Z} \rangle) : \tilde{\mathbf{X}} \sim F_{\mathbf{X}}\}$$

which is the correlation coefficient up to normalization and where $\mathbf{Z} = (Z_1, \dots, Z_d)$ satisfies $Z_i \geq 0$ and $\mathbb{E}(Z_i) = 1$ for $i = 1, \dots, d$. Then the random vector $\mathbf{X}^* \sim F_{\mathbf{X}}$ for which $\mathbb{E}(\langle \mathbf{X}^*, \mathbf{Z} \rangle) = \sup\{\mathbb{E}(\langle \tilde{\mathbf{X}}, \mathbf{Z} \rangle) : \tilde{\mathbf{X}} \sim F_{\mathbf{X}}\}$ is called the *worst case scenario* for the risk $\mathbf{X} \sim F_{\mathbf{X}}$ and the dependence structure of $(\mathbf{X}^*, \mathbf{Z})$ is the *worst case dependence structure*. For $d = 1$, a conclusion of Kusuoka (2001) is that the max-correlation risk measures are the only law-invariant, comonotone additive (i.e., $\rho(X + Y) = \rho(X) + \rho(Y)$, if (X, Y) is comonotone) coherent risk measures. Hence, the comonotone dependence structure represents the worst case dependence structure which exhibits no diversification. This fundamental result was extended in Ekeland et al. (2010) to the multivariate setup by replacing comonotonic additivity by μ -comonotonic additivity. They concluded that $\mathbf{X}^* + \mathbf{Y}^*$ is the worst case scenario for $\mathbf{X} + \mathbf{Y}$ with $\mathbf{X} \sim F_{\mathbf{X}}$, $\mathbf{Y} \sim F_{\mathbf{Y}}$ regarding $\rho_{\mathbf{Z}}$ if and only if $(\mathbf{X}^*, \mathbf{Y}^*)$ are $F_{\mathbf{Z}}$ -comonotone, i.e., if $(\mathbf{X}^*, \mathbf{Y}^*)$ are $F_{\mathbf{Z}}$ -comonotone then $\rho_{\mathbf{Z}}(\mathbf{X}^* + \mathbf{Y}^*) = \rho_{\mathbf{Z}}(\mathbf{X}^*) + \rho_{\mathbf{Z}}(\mathbf{Y}^*)$ reflecting *no diversification*. Rüschemdorf (2006, 2012b) summarizes that any multivariate law-invariant, coherent risk measure can be characterized by a class of max-correlation risk measures and their worst case scenarios by μ -comonotonic random vectors. Typically, multivariate law-invariant convex risk measures do admit a diversification effect. However for the subclass of *translated max-correlation measures* there exist examples of pairs (\mathbf{X}, \mathbf{Y}) without diversification effect. The area of risk management in a multivariate setting is a very active research area today, and there still remain many unanswered questions; see the forthcoming monograph Rüschemdorf (2012a).

6 And a financial crisis: Conclusion

It should be clear by now that a common thread through the above examples (theorems) is *Model Uncertainty*. An important lesson to be learned is that technical questions, like pricing and hedging, asked in the financial industry often are (or have to be) based on highly incomplete model assumptions. Economics definitely, and to a high degree finance and insurance, involve an important social/human factor which cannot be fully captured by rationality. This leads in various ways to a relatively high degree of uncertainty; it is no coincidence that as a consequence of The Crisis the classic Knight (1921) on Risk, Uncertainty, and Profit has been rediscovered. Also the various combinations of the Knowns and Unknowns, as for instance discussed in Diebold et al.

(2010), have become popular. Finally, extreme views as for instance exposed in Taleb (2007) doubt whether any value can be given to a scientific/rational approach to finance. As always, the truth will lie somewhere in the middle between the totally rational and the fully behavioral, say, points of view. The history of life insurance for instance has shown us how useful actuarial (mathematical) calculations can be. Similarly, the over-rational hypes during The Crisis brought many of us with their feet back on solid ground. Concentrating on quantitative (mathematical) finance, there is no doubt that mathematical finance has been very successful in relating today's prices, but is much less confident in explaining (predicting) tomorrow's ones (Hans Bühlmann, private communication). Mathematicians are well (or at least ought to be) aware of this.

Going forward, mathematicians (*a*) have to get much more involved with more applied issues in insurance and finance, and (*b*) have to keep on stressing the conditions needed to be fulfilled in order for certain results to be applied. Our paper contains numerous examples of this. Whereas (appointed) actuaries in insurance have a long-standing tradition in this respect, in finance there is much more work to be done. It is therefore not possible to single out one concrete theme where we would encourage young researchers to start working on. One key observation is that Quantitative Risk Management (QRM) will gain in importance; and this not just by providing tools and techniques, like for instance done in McNeil et al. (2005), but more importantly by drawing a clear line for practice with a sign "careful, those questions asked are with current knowledge impossible to answer!". Examples are the pricing and hedging of synthetic CDOs or the 99.9% VaR estimation of yearly Operational Risk losses.

We see two main areas where QRM research will have to dig deeper:

- (*a*) in going from more frequency oriented "if" questions to a more severity oriented "what if" approach, and this at several levels, and
- (*b*) thinking more about the (Q)RM landscape in a changing world: "Within the financial industry, are we covering with our RM radar all relevant corners?"

Let us end with some thoughts on (*b*) by recalling three main dimensions of risk management:

- Dimension 1: Scope.
 - Micro: the individual firm, the trading floor, the client, ...
 - Macro: the country or even worldwide system, networks.
- Dimension 2: Time.
 - Short: high frequency trading, \ll 1 year (or quarter).
 - Medium: Solvency 2/ Basel II/III, \sim 1 year.
 - Long: social/ life insurance, \gg 1 year.
- Dimension 3: Level.
 - Quantitative versus qualitative risk assessment.

The above carving up of the RM-cake is of course rather simplistic, though it is fair to say that most of the (more mathematical) work has centered on Micro/ Medium/ Quantitative. Mathematicians have to become more aware of the need to spend more time and effort on the other (to some extent more important) combinations! We very much hope that our paper will arouse some interest on these issues.

Acknowledgement

This paper formed the basis for the Keynote Lecture by the second author at the Fifth Conference of the Thailand Econometric Society; he takes pleasure in thanking the organizers of that conference for the support given. Bikramjit Das and Vicky Fasen thank RiskLab, ETH Zurich, for the financial support.

References

- Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1997). Thinking coherently. *RISK*, **10**, 68–71.
- Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.
- Banach, S. and Tarski, A. (1924). Sur la décomposition des ensembles de points en parties respectivement congruentes. *Fundamenta Mathematicae*, **6**, 244–277.
- Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes*. John Wiley & Sons, Chichester.
- Bingham, N., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. Cambridge University Press, Cambridge.
- Braithwaite, T. (2011). Banks turn to financial alchemy in search for capital. October 25, Financial Times, p. 14.
- Brunetti, A. (2011). *Wirtschaftskrise ohne Ende?* 2nd edn. h.e.p. Verlag, Bern.
- Daniélsson, J. (2011). *Financial Risk Forecasting*. John Wiley & Sons, Chichester.
- Degen, M., Lambrigger, D.D. and Segers, J. (2010). Risk concentration and diversification: second-order properties. *Insurance: Mathematics & Economics*, **46**(3), 541–546.
- Delbaen, F. (2000). Coherent risk measures on general probability spaces. Preprint ETH Zurich.
- Delbaen, F. (2009). Risk measures for non-integrable random variables. *Mathematical Finance*, **19**(2), 329–333.

- Dewdney, A.K. (1989). A matter fabricator provides matter for thought. *Computer Recreations. Scientific American*, April 1989, 86–89.
- Diebold, F.X., Doherty, N.A. and Herring, R.J. (2010). *The Known, the Unknown and the Unknowable in Financial Risk Management*. Princeton University Press, Princeton.
- Donnelly, C. and Embrechts, P. (2010). The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *ASTIN Bulletin*, **40**, 1–33.
- Drouot, D. and Kotz, S. (2001). *Correlation and Dependence*. Imperial College Press, London.
- Ekeland, I., Galichon, A. and Henry, M. (2010). Comonotonic measures of multivariate risk. *Mathematical Finance*, **20**, 1–24.
- Embrechts, P. (1999). Die Zauberformel gibt es nicht. In: G. Wüstholtz (Ed.), *Mathematik – das geistige Auge*, pages 28–29. Departement Mathematik, ETH-Zurich.
- Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, **76**(3), 639–650.
- Embrechts, P. (2011). Ceterum censeo carthaginem esse delendam. In: A. Herzberg (Ed.), *Proceedings of the Conference on Statistics, Science and Public Policy*.
- Embrechts, P., Lambrigger, D.D. and Wüthrich, M.V. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, **12**(2), 107–127.
- Embrechts, P. and Mainik, G. (2012). Asymptotic diversification effects in multivariate regularly varying models. Preprint.
- Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In: M.A.H. Dempster (Ed.), *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press, Cambridge.
- Embrechts, P. and Puccetti, G. (2010). Risk aggregation. In: P. Jaworski, F. Durante, W. Haerdle and T. Rychlik (Eds.), *Copula Theory and its Applications*, pages 111–126. Springer, Berlin.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol. 2. 2nd edn. Wiley, New York.
- Föllmer, H. and Schied, A. (2011). *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, Berlin. 3rd, extended edition.
- Fréchet, M. (1957). Sur les tableaux de corrélations dont les marges sont données. *Annales de l'Université de Lyon, Sciences Mathématiques et Astronomie, Série A*, **9**, 53–57.
- Frittelli, M. and Gianin, E.R. (2002). Putting order in risk measures. *Journal of Banking & Finance*, **26**, 1473–1486.

- Genest, C., Gendron, M. and Bourdeau-Brien, M. (2009). The advent of copulas in finance. *European Journal of Finance*, **15**(7-8), 609–618.
- Höfding, W. (1940). Maßstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, **5**, 181–233.
- Höfding, W. (1941). Maßstabinvariante Korrelationstheorie für diskontinuierliche Verteilungen. *Archiv für mathematische Wirtschafts- und Sozialforschung*, **7**, 49–70.
- Hua, L. and Joe, H. (2011). Tail order and intermediate tail dependence of multivariate copulas. *Journal of Multivariate Analysis*, **102**(10), 1454–1471.
- Huerta de Soto, J. (2009). *Money, Bank Credit, and Economic Cycles*. Ludwig von Mises Institute, Atlanta.
- Ibragimov, R., Jaffee, D. and Walden, J. (2009). Nondiversification traps in catastrophe insurance markets. *Review of Financial Studies*, **22**, 959–993.
- Ibragimov, R., Jaffee, D. and Walden, J. (2011). Diversification disasters. *Journal of Financial Economics*, **99**(2), 333–348.
- Ibragimov, R. and Walden, J. (2007). The limits of diversification when losses may be large. *Journal of Banking & Finance*, **31**, 2551–2569.
- Kaina, M. and Rüschendorf, L. (2009). On convex risk measures on L^p -spaces. *Mathematical Methods of Operations Research*, **69**(3), 475–495.
- Karamata, J. (1930). Sur un mode de croissance régulière des fonctions. *Mathematica (Cluj.)*, **4**, 38–53.
- Kazanjian, K. (2000). *Wizards of Wall Street*. Aspen publishers.
- Kindleberger, C. and Aliber, R. (2011). *Manias, Panics, and Crashes: A History of Financial Crises*. Wiley, New York.
- Knight, F.H. (1921). *Risk, Uncertainty, and Profit*. Houghton Mifflin Company, Boston.
- Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.
- Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, **83**(1), 169–187.
- Li, D. (2000). On default correlation: a copula function approach. *Journal of Fixed Income*, **9**(4), 43–54.
- McNeil, A., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton.

- Meilijson, I. and Nadas, A. (1979). Convex majorization with an application to the length of critical paths. *Journal of Applied Probability*, **6**, 671–677.
- Moscadelli, M. (2009). The modeling of operational risk: experience with the analysis of the data collected by the Basel committee. Banca d’Italia, Temi di discussione del Servizio Studi, No. 517.
- Nešlehová, J., Embrechts, P. and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, **1**(1), 3–25.
- Puccetti, G. and Scarsini, M. (2010). Multivariate comonotonicity. *Journal of Multivariate Analysis*, **101**(1), 291–304.
- Reiss, R.D. (1989). *Approximate Distributions of Order Statistics*. Springer-Verlag, New York.
- Resnick, S.I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, **5**, 303–336.
- Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- Resnick, S.I. (2008). *Extreme Values, Regular Variation and Point Processes*. 2nd edn. Springer, New York.
- Rogers, L.C.G. (2010). *Mathematical Finance: the P&L*. Plenary Lecture at the 34th Conference on Stochastic Processes and Their Applications, Osaka, Japan, September 6, 2010.
- Rothbard, M.N. (2008). *The Mystery of Banking*. Ludwig von Mises Institute, Atlanta.
- Rudin, W. (1973). *Functional Analysis*. McGraw-Hill Book Co., New York.
- Rüschendorf, L. (2006). Law invariant convex risk measures for portfolio vectors. *Statistics & Decisions*, **24**(1), 97–108.
- Rüschendorf, L. (2012a). *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Book manuscript, to appear.
- Rüschendorf, L. (2012b). Worst case portfolio vectors and diversification effects. *Finance & Stochastics*, **16**, 155–175.
- Salmon, F. (2009). Recipe for disaster: the formula that killed Wall Street. February 23, Wired Magazine.
- Salmon, F. (2012). The formula that killed Wall Street. *Significance*, **9**(1), 16–20.
- Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, **11**, 195–210.

Taleb, N.N. (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House Publishing Group.

The Economist (2011). Briefing: Heterodox economies. Marginal revolutionaries. The Economist. December 31, 49–52.

Villani, C. (2003). *Topics in Optimal Transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence.

Wagon, S. (1993). *The Banach-Tarski Paradox*. Cambridge University Press, Cambridge.

Wapner, L.M. (2005). *The Pea and The Sun: A Mathematical Paradox*. A K Peters/CRC Press.