ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Statistics and Quantitative Risk Management for Banking and Insurance

Paul Embrechts

RiskLab, Department of Mathematics and Swiss Finance Institute, ETH Zurich, 8092 Zurich, Switzerland, embrechts@math.ethz.ch

Marius Hofert

RiskLab, Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland, marius.hofert@math.ethz.ch.

The author (Willis Research Fellow) thanks Willis Re for financial support while this work was being completed.

Keywords

Statistical methods, risk measures, dependence modeling, risk aggregation, regulatory practice

Abstract

As an emerging field of applied research, Quantitative Risk Management (QRM) poses a lot of challenges for probabilistic and statistical modeling. This review provides a discussion on selected past, current, and possible future areas of research in the intersection of statistics and quantitative risk management. Topics treated include the use of risk measures in regulation, including their statistical estimation and aggregation properties. An extensive literature provides the statistically interested reader with an entrance to this exciting field.

1. Introduction

In 2005, the first author (Paul Embrechts), together with Alexander J. McNeil and Rüdiger Frey, wrote the by now well-established text McNeil et al. (2005). Whereas the title read as "Quantitative Risk Management: Concepts, Techniques, Tools", a somewhat more precise description of the content would be reflected in "**Statistical Methods for** Quantitative Risk Management: Concepts, Techniques, Tools". Now, almost a decade and several financial crises later, more than ever, the topic of Quantitative Risk Management (QRM) is high on the agenda of academics, practitioners, regulators, politicians, the media, as well as the public at large. The current paper aims at giving a brief historical overview of how QRM as a field of statistics and applied probability came to be. We will discuss some of its current themes of research and attempt to summarize (some of) the main challenges going forward. This is not an easy task as the word "risk" is omnipresent in modern society and consequently its (statistical) quantification. One therefore has to be careful not to lose sight of the forest for the trees! Under the heading "QRM: The Nature of the Challenge", in McNeil et al. (2005, Section 1.5.1) it is written: "We set ourselves the task of defining a new discipline of QRM and our approach to this task has two main strands. On the one hand, we have attempted to put current practice onto a firmer **mathematical** footing. On the other hand, the second strand of our endeavor has been to put together material on techniques and tools which go **beyond** current practice and address some of the **deficiencies** that have been raised repeatedly by critics". Especially in the wake of financial crises, it is not always straightforward to defend a more mathematical approach on "risk" (more on this later). For the purpose of the paper, we interpret **risk** at the more mathematical level of (random) uncertainty in both frequency as well as severity of loss events. We will not enter into the, very important, discussion around "Risk and Uncertainty" (Knight (1921)) on the various "Knowns, Unknowns and Unknowables" (Diebold et al. (2010)). Concerning "mathematical level", let the following quotes speak for themselves. First, in their (Nobel-)path breaking paper Gale and Shapley (1962) are addressing the question "What is mathematics?" the authors wrote "…any argument which is carried out with sufficient precision is mathematical,…". Lloyd Shapley (together with Alvin Roth) got the 2012 Nobel Prize for economics; see Roth and Shapley (2012). As a second quote on the topic we like Norbert Wiener's; see Wiener (1923):

> "Mathematics is an experimental science … It matters little … that the mathematician experiments with pencil and paper while the chemist uses test-tube and retort, or the biologist stains and the microscope … The only great point of divergence between mathematics and other sciences lies in the far greater permanence of mathematical knowledge, in the circumstance that experience only whispers 'yes' or 'no' in reply to our questions, while logic shouts."

It is precisely statistics that has to straddle the shouting world of mathematical logic with the whispering one of practical reality.

2. A whisper from regulatory practice

In the financial (banking and insurance) industry, solvency regulation has been around for a long time. The catchwords are Basel for banking and Solvency for insurance. The former derives from the Basel Committee on Banking Supervision, a supra-national institution, based in Basel, Switzerland, working out solvency guidelines for banks. Basic frameworks go under the names of Basel I, II, III together with some intermediate stages. The website `www.bis.org/bcbs` of the Bank for International Settlements (BIS) warrants a regular visit from anyone interested in banking regulation. An excellent historic overview of the Committee's working is to be found in Tarullo (2008). Similar developments for the insurance industry go under the name of Solvency, in particular the current Solvency II framework. For a broader overview, see Sandström (2006). Whereas Solvency II is not yet in force, since January 1, 2011, the Swiss Solvency Test (SST) is. The key principles on which the SST, originating in 2003, is based are: (a) it is risk based, quantifying market, insurance and credit risk, (b) it stresses market consistent valuation of assets and liabilities, and (c) advocates a total balance sheet approach. Before, actuaries were mainly involved with liability pricing and reserve calculations. Within the SST, stress scenarios are to be quantified and aggregated for capital requirement. Finally, as under the Basel regime for banking, internal models are encouraged and need approval by the relevant regulators (the FINMA in the case of Switzerland). An interesting text exposing the main issues both from a more methodological as well as practical point of view is SCOR (2008). As in any field of applications, if as a statistician one really wants to have an impact one has to get more deeply involved with practice. From this point of view, some of the above publications are must reads. McNeil et al. (2005, Chapter 1) gives a somewhat smooth introduction for those lacking the necessary time.

Let us concentrate on one particular example highlighting the potentially interesting interactions between methodology and practice, and hence the title of this section. As already stated, the Basel Committee, on a regular basis, produces documents aimed at improving the resilience of the international financial system. The various stakeholders, including academia, are asked to comment on these new guidelines before they are submitted to the various national agencies to be cast into local law. One such document is BIS (2012). In the latter consultative document, p. 41, the following question (Nr. 8) is asked: "What are the likely operational constraints with moving from VaR to ES, including any challenges in delivering robust backtesting, and how might these be best overcome?"; for VaR and ES, see Definition 2.1. This is an eminently statistical question the meaning of which, together with possible answers, we shall discuss below. This example has to be viewed as a blueprint for similar discussions of statistical nature within the current QRM landscape. Some further legal clarification may be useful at this point: by issuing consultative documents, the Basel Committee wants to intensify its consultation and discussions with the various stakeholders of the financial system. In particular, academia and industry are invited to comment on the new proposals and as such may have an influence on the precise formulations going forward. The official replies are published on the website of the BIS. We will not be able to enter into all relevant details but merely highlight the issues relevant from a more statistical point of view; in doing so, we will make several simplifying assumptions.

The first ingredient is a portfolio $\mathcal{P}$ of financial positions each with a well-defined value at any time t. For t (today, say) the value of the portfolio is v_t. Though crucial in practice, we will not enter into a more detailed economic and/or actuarial discussion of the precise meaning of **value**: it suffices to mention possibilities like "fair value", "market value" (mark-to-market), "model value" (mark-to-model), "accounting value" (often depending on the jurisdiction under which industry is regulated) and indeed various combinations of the above! Viewed from today, the value of the portfolio one time period in the future is a random variable V_{t+1}. In (market) risk management one is interested in the (positive) tail of the so-called **Profit-and-Loss (P&L)** distribution function (df) of the random variable (rv)

$$L_{t+1} = -(V_{t+1} - v_t). \tag{1}$$

The one-period ahead value V_{t+1} is determined by the **structure** $S(t+1, \boldsymbol{Z}_{t+1})$ of $\mathcal{P}$, where $\boldsymbol{Z}_{t+1}$ in $\mathbb{R}^d$ is a (typically) high-dimensional random vector of **risk factors** with corresponding **risk-factor changes** $\boldsymbol{X}_{t+1} = \boldsymbol{Z}_{t+1} - \boldsymbol{z}_t$. S is also referred to as the **mapping**. As an example consider for $\mathcal{P}$ a linear portfolio consisting of β_j stocks $S_{t,j}$, $j \in \{1, \dots, d\}$ ($S_{t,j}$ denotes the value of stock j at time t and β_j the number of shares of stock j in $\mathcal{P}$). Furthermore, assume the risk factors to be

$$\boldsymbol{Z}_t = (Z_{t,1}, \dots, Z_{t,d}), \quad Z_{t,j} = \log S_{t,j}$$

and the corresponding risk-factor changes to be the log-returns (here again, t denotes today):

$$X_{t+1} = \log(S_{t+1,j}) - \log(s_{t,j}) = \log\Bigl(\frac{S_{t+1,j}}{s_{t,j}}\Bigr).$$

In this case, the portfolio structure is

$$S(t+1, \boldsymbol{Z}_{t+1}) = \sum_{j=1}^{d} \beta_j S_{t+1,j} = \sum_{j=1}^{d} \beta_j \exp(Z_{t+1,j}). \tag{2}$$

(1) together with $V_{t+1} = S(t+1, \boldsymbol{Z}_{t+1})$ then implies that

$$\begin{aligned} L_{t+1} &= -\sum_{j=1}^{d} \beta_j (\exp(Z_{t+1,j}) - \exp(z_{t,j})) \\ &= -\sum_{j=1}^{d} \beta_j (\exp(z_{t,j} + X_{t+1,j}) - \exp(z_{t,j})) = -\sum_{j=1}^{d} \tilde{w}_{t,j} (\exp(X_{t+1,j}) - 1), \end{aligned}$$

where $\tilde{w}_{t,j} = \beta_j s_{t,j}$; $w_{t,j} = \tilde{w}_{t,j}/v_t$ is then the relative value of stock j in $\mathcal{P}$. For this, and further examples, see McNeil et al. (2005, Section 2.1.3).

Based on models and statistical estimates (computed from data) for $(\boldsymbol{X}_{t+1})$, the problem is now clear: find the df (or some characteristic, a risk measure) of L_{t+1}. Here, "the df" can be interpreted unconditionally or conditionally on some family of σ-algebras (a filtration of historical information). As a consequence, the mathematical set-up can become arbitrarily involved!

Analytically, (1) becomes

$$\begin{aligned} L_{t+1} &= -(S(t+1, \boldsymbol{Z}_{t+1}) - S(t, \boldsymbol{z}_t)) \\ &= -(S(t+1, \boldsymbol{z}_t + \boldsymbol{X}_{t+1}) - S(t, \boldsymbol{z}_t)). \end{aligned} \tag{3}$$

If, for ease of notation, we suppress the time dependence, (3) becomes

$$L = -(S(\boldsymbol{z} + \boldsymbol{X}) - S(\boldsymbol{z})) \tag{4}$$

with $\boldsymbol{z} \in \mathbb{R}^d$ and $\boldsymbol{X}$ a d-dimensional random vector denoting the one-period changes in factor values; d is typically high, $d \geq 1000$! In general, the function S can be highly non-linear. This especially holds true in cases where derivative financial instruments (like options) are involved. If time is modeled explicitly (which in practice is necessary) then (3) is a time series $(L_t)_t$ (stationary or not) driven by a d-dimensional stochastic process $(\boldsymbol{X}_t)_t$ and a deterministic, though mostly highly non-linear portfolio structure S. Going back (for ease of notation) to the static case (4), one has to model the df $F_L(x) = \mathbb{P}(L \leq x)$ under various assumptions on the input. Only for the most trivial case, like $\boldsymbol{X}$ is multivariate normal and S linear, can this be handled analytically. In practice, various approximation methods are used:

(A1) Replace S by a linearized version S^{Δ} using Taylor expansion and approximate F_L by the df $F_{L^{\Delta}}$ of the linearized loss L^{Δ}; here one uses smoothness of S (usually fine) combined with a condition that $\boldsymbol{X}$ in (4) is stochastically small (a more problematic assumption). The latter implies that one-period changes in the risk factors are small; this is acceptable in "normal" periods but not in "extreme" periods. Note that it is especially for the latter that QRM is needed!

(A2) Find stochastic models for $(\boldsymbol{X}_t)_t$ closer to reality, in particular beyond Gaussianity but for which F_L or $F_{L^{\Delta}}$ can be calculated/estimated readily. This is a non-trivial task in general.

(A3) Rather than aiming for a full model for F_L, estimate some characteristics of F_L relevant for solvency capital calculations, that is, estimate a so-called **risk measure**. Here the VaR and ES abbreviations in the above BIS (2012) quote enter, both are risk measures; they are defined as follows.

Definition 2.1

Suppose L as given above and let $0 < \alpha < 1$, then

i) The **Value-at-Risk** of L at confidence level α is given by

$$\mathrm{VaR}_\alpha(L) = \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}$$

(i.e. $\mathrm{VaR}_\alpha(L)$ is the $100\alpha\%$ quantile of F_L).

ii) The **Expected Shortfall** of L at confidence level α is given by

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\,du.$$

Remark 2.2

1) $\mathrm{VaR}_\alpha(L)$ defined above coincides with the so-called **generalized inverse** $F_L^{\leftarrow}(\alpha)$ at the confidence level α; see Embrechts and Hofert (2013a). Its introduction around 1994, linked to (4) as a basis for solvency capital calculation, truly led to a new risk management benchmark on Wall Street; see Jorion (2007). A simple Google search will quickly confirm this. An interesting text aiming at improving VaR-usage and warning about the various misunderstandings and misuses is Wong (2013).

2) For F_L continuous, the definition of Expected Shortfall is equivalent to

$$\mathrm{ES}_\alpha(L) = \mathbb{E}[L \mid L > \mathrm{VaR}_\alpha(L)],$$

hence its name. Alternative names in use throughout the financial industry are "conditional VaR" and "conditional tail expectation".

3) Note that ES does not exist for infinite mean models, i.e. when $\mathrm{ES}_\alpha |L| = \infty$. This reveals a potential problem with ES as a risk measure in case of very heavy-tailed loss data.

4) Both risk measures are used in industry and regulation; VaR more in banking, ES more in insurance, the latter especially under the SST. Depending on the risk class under consideration, different values of α (typically close to 1) and the holding period ("the one period" above, see (1)) are in use; see McNeil et al. (2005) for details.

5) A whole industry of papers on estimating VaR and/or ES has emerged, and this under a wide variety of model assumptions on the underlying P&L process $(L_t)_t$. McNeil et al. (2005) contains a summary of results stressing the more statistical issues. An excellent companion text with a more econometric flavour is Daníelsson (2011). More recent statistical techniques in use are for instance extreme quantile estimation based on regime-switching models and lasso-technology, see Chavez-Demoulin et al. (2013a). Estimation based on self-exciting (or Hawkes) processes is exemplified in Chavez-Demoulin and McGill (2012). For an early use on Hawkes processes in finance, see Chavez-Demoulin et al. (2005). McNeil et al. (2005, Section 7.4.3) contains an early text-book reference to QRM. Extreme Value Theory (EVT) methodology is for instance presented in McNeil and Frey (2000); Chavez-Demoulin et al. (2013b) use it in combination with generalized additive models in order to model VaR for Operational Risk based on covariates.

6) From the onset, especially VaR has been heavily criticized as it only looks at the frequency of extremal losses (α close to 1) but not at the severity. The important "what if" question is more addressed by ES. Furthermore, in general VaR is not a subadditive risk measure, i.e. it is possible that

for two positions L_1 and L_2, $\mathrm{VaR}_\alpha(L_1 + L_2) > \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2)$ for some $0 < \alpha < 1$. This makes risk aggregation and diversification arguments difficult. ES as defined in Definition 2.1 ii) above is always subadditive. These issues played a role in the subprime crisis of 2007–2009; see Donnelly and Embrechts (2010) and Das et al. (2013) for some discussions on this. McNeil et al. (2005) contains an in-depth analysis together with numerous references for further reading. Whereas industry early on (mid nineties) did not really take notice of the warnings concerning the problem with VaR in markets away from "normality", by now the negative issues underlying VaR-based regulation have become abundantly clear and the literature is full of examples on this. Below we have included a particular example from the realm of credit risk stressing, hopefully exposing in a pedagogically clear way, some of the problems. It is a slightly expanded version of McNeil et al. (2005, Example 6.7) and basically goes back to Albanese (1997). By now, there exist numerous similar examples.

Example 2.3 (Non-subadditivity of VaR)
Assume we have given 100 bonds with maturity T equal to one year, nominal value 100, yearly coupon 2%, and default probability 1% (no recovery). The corresponding losses (assumed to be independent) are

$$L_i = \begin{cases} -2, & \text{with probability } 0.99, \\ 100, & \text{with probability } 0.01, \end{cases} \qquad i \in \{1, \ldots, 100\}.$$

(Recall the choice of negative values to denote gains!) Consider two portfolios,

$$\mathcal{P}_1 : \sum_{i=1}^{100} L_i \qquad \text{and} \qquad \mathcal{P}_2 : \ 100L_1. \tag{5}$$

$\mathcal{P}_1$ is a so-called diversified portfolio and $\mathcal{P}_2$ clearly a highly concentrated one. As this point the reader may judge for himself/herself which of the two portfolios he/she believes to be less risky (and thus would assign a smaller risk measure to).

Figure 1 shows the boundaries of this decision according to VaR (the shift by 201 in y scale is for plotting the y axis in log-scale). Rather surprisingly, for $\alpha = 0.95$, for example, VaR_α is superadditive, i.e.

$$\mathrm{VaR}_{0.95}(\mathcal{P}_1) > \mathrm{VaR}_{0.95}(\mathcal{P}_2).$$

By the interpretation of a risk measure as the amount of capital required as a buffer against future losses, this implies that the diversified portfolio $\mathcal{P}_1$ can be riskier (thus requiring a larger risk capital) than the concentrated $\mathcal{P}_2$ according to VaR. Indeed, one can show that for two portfolios as in (5) based on n bonds with default probability p, VaR_α is superadditive if and only if $(1-p)^n < \alpha \leq 1-p$. This formula holds independently of the coupon and nominal value. For $n = 100$ and $p = 1\%$ we obtain that VaR_α is superadditive if and only if $0.3660 \approx 0.99^{100} < \alpha \leq 0.99$.

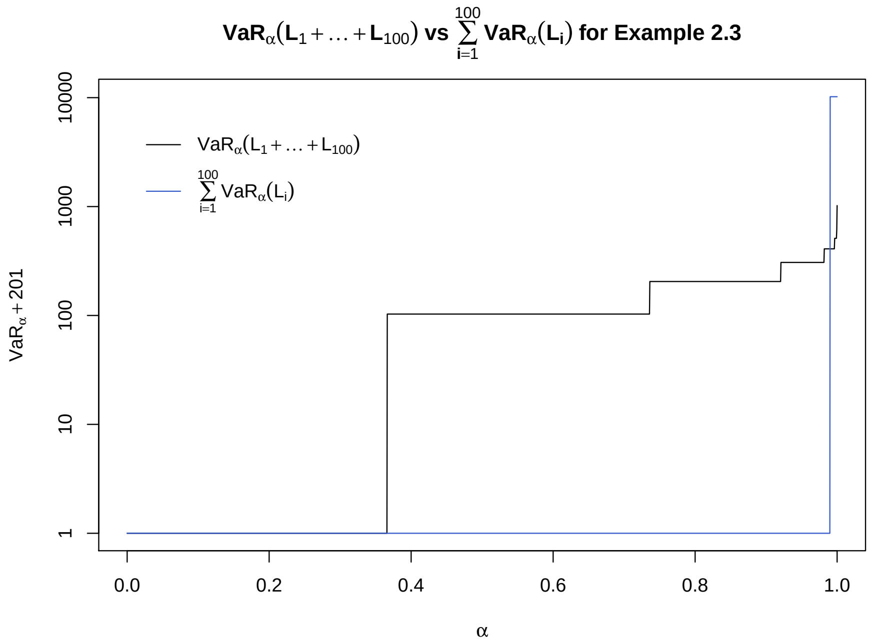

Figure 1

VaR_α as a function in α for the two portfolios $\mathcal{P}_1$ and $\mathcal{P}_2$.

It is clear that for $\alpha > 0.95$, say, corresponding to values typical for financial and insurance regulation, both $\mathrm{VaR}_\alpha(L)$ and $\mathrm{ES}_\alpha(L)$ are difficult to estimate. One needs an excellent fit for $1 - F_L(x)$, x large; at this point, EVT as summarized in McNeil et al. (2005, Chapter 7) comes in as a useful tool. Not that EVT is a panacea for solving such high-quantile estimation problems, more importantly, it offers guidance on the kind of questions from industry and regulation which are beyond the boundaries of sufficiently precise estimation. In the spirit of McNeil et al. (2005, Example 7.25) we present Figure 2 based on the Danish fire insurance data. For these typical loss data from non-life insurance, the linear log-log plot for the data above a threshold of $u = 5.54$ (the 90%-quantile of the empirical distribution; in million Danish Krone) clearly indicates power-tail behavior. The solid line through the data stems from a Peaks-Over-Threshold fit (EVT); see McNeil et al. (2005, Chapter 7) for details. The dotted curves are the so-called profile likelihoods, the width of these at a specific confidence level (second y axis on the right-hand side) indicates EVT-based confidence intervals for the risk measure under consideration (VaR on the left-hand side, ES on the right-hand side of Figure 2). Do note the considerable uncertainty around these high-quantile estimates!

We now return to Question 8 in BIS (2012) concerning a possible regulatory regime switch for market risk management from VaR to ES and the issue of “robust backtesting”. A summary of our current knowledge on Question 8 we summarize below. We compare and contrast both risk measures on the basis of four broadly defined criteria: (C1) Existence; (C2) Ease of accurate statistical estimation; (C3) Subadditivity, and (C4) Robust forecasting and backtesting. The “loud” answer to the above “whispering” question, coming

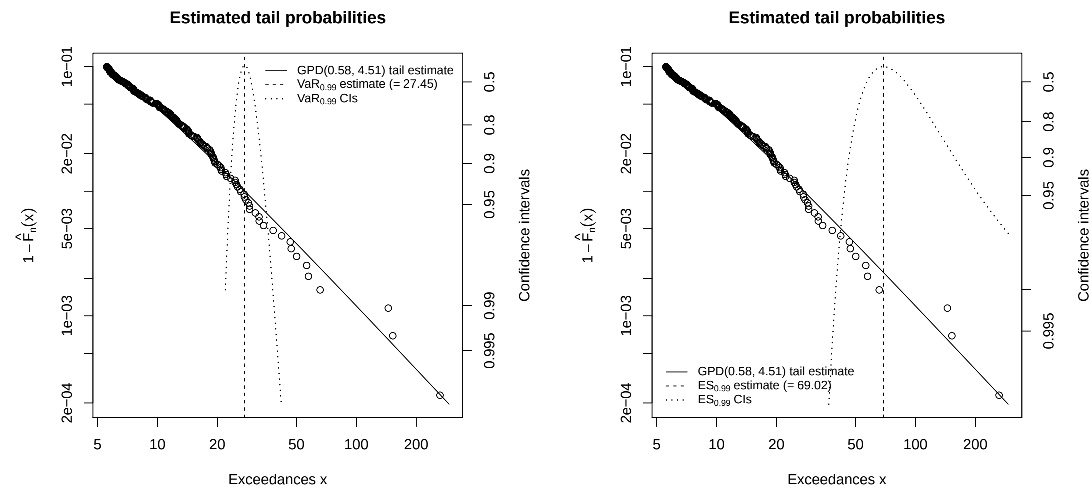

Figure 2

Estimated tail probabilities including risk measures $\mathrm{VaR}_{0.99}$ (left) and $\mathrm{ES}_{0.99}$ (right) with likelihood-based confidence intervals for the Danish fire insurance data.

from mathematical statistics, is expertly summarized in Gneiting (2011). In the latter paper, the author offers a theory to assess the quality of statistical forecasts introducing the notion of elicitability. For the ease of this discussion, we will refer to a forecast as **elicitable** if it is "properly" backtestable. For the precise mathematical meaning of "properly" we refer to the latter paper and the references therein. It is shown that in general, VaR is elicitable whereas ES is not (note that in Gneiting (2011), ES is referred to as Conditional Value-at-Risk, CVaR). In Section 3.4 the author states that: "This negative result may challenge the use of the CVaR functional as a predictive measure of risk, and may provide a partial explanation for the lack of literature on the evaluation of CVaR forecasts, as opposed to quantile or VaR forecasts, for which we refer to Berkowitz and O'Brien (2002), Giacomini and Komunjer (2005), Balo et al. (2006), among others". See also Jorion (2007, Chapter 6). The Basel II guidelines also contain a multiplier penalty matrix for solvency capital, the severity of the multiplier being a function of the historical statistical accuracy of the backtesting results; see Jorion (2007, Table 6-3, p. 148). A further comment on the difficulty of ES-forecasting is to be found in Daníelsson (2011, Remark 2, p. 89 and Section 8.4) mainly concentrating on the need for much more data for ES as compared to VaR, data which often is not available. This leads to the interesting question to what extent elicitability is really fundamental towards risk-measure backtesting in insurance and finance. Whereas we are in no doubt that the notion is important, more research on this topic is needed. A first relevant paper beyond Gneiting (2011) is Ziegel (2013). Both VaR and ES are law invariant risk measures, in risk-measure terminology, ES is coherent (in particular, subadditive), whereas VaR in general is not (as we saw in Example 2.3). This raises the question whether there exist non-trivial, law-invariant coherent risk measures. The answer, from Ziegel (2013), is that so-called expectiles are, but the important subclass of law-invariant spectral risk measures are not elicitable. For some more information on expectiles, see Rémillard (2013) and the references therein. Ziegel (2013) contains an interesting discussion on how backtesting procedures based on exceedance residuals as in McNeil and Frey (2000) or McNeil et al. (2005, Section 4.4.3) may possibly be put into a decision-theoretic framework akin to elicitability. As already stated above, this is no doubt an important area of future research. Clearly, concerning (C1), $\mathrm{VaR}_\alpha(L)$, when properly defined as in Definition 2.1, always exists. For $\mathrm{ES}_\alpha[L]$ to exist, one needs $\mathbb{E}|L| < \infty$. The latter condition may seem without any problem, see however the discussion on Operational Risk in McNeil et al. (2005, Section 10.1.4). Concerning (C2), and this especially for α close to 1, both measures are difficult to estimate with great statistical accuracy. Of course, confidence intervals around ES are often much larger than corresponding ones around VaR; see McNeil et al. (2005, Figure 7.6) or Figure 2 above. Concerning (C3), we already discussed the fact that in general VaR is not subadditive whereas ES is. This very important property tips the balance clearly in favor of ES; of course ES also contains by definition important loss information of the "what if" type.

It is clear that for criteria (C1)–(C4) above one needs much more detailed information specifying precise underlying model assumptions, data availability, specific applications etc. The references given yield this background. The backtesting part of (C4) we already discussed above. When robustness in Basel's Question 8 is to be interpreted in a precise statistical sense, like in Huber and Ronchetti (2009) or Hampel et al. (2005), then relevant information is to be found in Dell'Aquila and Embrechts (2006) and Mancini and Trojani (2011); these papers concentrate on robust estimation of risk measures like VaR. In Cont et al. (2010) the authors conclude that "…historical Value-at-Risk, while failing to be subadditive, leads to a more robust procedure than alternatives such as Expected Shortfall". Moreover, "the conflict we have noted between robustness of a risk measurement procedure and the subadditivity of the risk measure show that one cannot achieve robust estimation in this framework while preserving subadditivitiy." This conclusion seems in line with the results in Gneiting (2011) replacing elicitability by robustness. However also here the discussion is not yet closed: robustness can be interpreted in several ways and it is not clear what kind of robustness the Basel Committee has in mind. For instance, Filipović and Cambou (private communication), within the context of the SST, formulate a "robustness criterium" which is satisfied by ES but not by VaR. Hence also here, more work is clearly needed.

For the above discussion we did not include an overview of the explosive literature on the theory of risk measures. Some early references with a more economic background are Gilboa and Schmeidler (1989) and Fishburn (1984). A paper linking up more with recent developments in financial econometrics is Andersen et al. (2013). An excellent textbook treatment at the more mathematical level is Föllmer and Schied (2011). From a more mathematical point of view, recent research on risk measures borrows a lot from Functional Analysis; see Delbaen (2012). Concerning the question on robustness properties of risk measures: it is historically interesting that the mathematical content of the main result in Artzner et al. (1999, Proposition 4.1) is contained in the classic on statistical robustness Huber (1981, Chapter 10, Proposition 2.1).

SUMMARY POINTS

The above discussion clearly shows the importance of (mathematical) statistics and probability theory already at the level of very basic questions in QRM of a highly relevant nature. In Section 4 we will add a further example of this important discourse, and this under the name of model uncertainty and risk aggregation. To this end, statisticians with a substantial background in practically relevant QRM should also be able and willing to give their guiding advice, in particular concerning questions like the above Question 8. Given all the practical and scientific information we know of at the moment, our advice to the regulators would be "stick with the devil", i.e. stay with VaR-based capital solvency laws if a risk-measure based regime is wanted. Be very aware of all of VaR's weaknesses (and these are many), especially in terms of crises. We would prefer statistical estimation for lower confidence level α in VaR_α together with regulatory defined stress factors (already partly in place, but for α's too close to 1, i.e. too far out in the tails of the underlying P&L's where statistical, as well as model uncertainty is considerable). Banks should be encouraged to ask, and report on the "what if" question, as such, reporting on ES_α is definitely useful (if $\mathbb{E}|L| < \infty$). Finally, though we did not report on the holding period (the plus 1 in (1)), time scaling of VaR_α, like for market risk from 1-day to 10-days (= two trading weeks), our research indicates that (overall) it is difficult to beat the square-root-of-time rule; see Embrechts et al. (2005) and Kaufmann (2004), and, for an introduction to this rule, Jorion (2007, p. 98). This issue also figures prominently in BIS (2012); see for instance Section 3.3 on "Factoring in market liquidity".

3. Modeling interdependence

Returning to (3), a main statistical modeling task remains the (high-dimensional) risk factor process $(\boldsymbol{Z}_t)_t$ or the risk factor changes $(\boldsymbol{X}_t)_t$. This either in a stationary or a non-stationary environment. Also here, a huge, mainly econometric literature exists; interesting texts are Engle (2009) and Christian Gourieroux (2001). From a more static modeling point of view, Genton (2004) offers several industry models. A review of models in use within QRM is to be found in McNeil et al. (2005, Chapter 1, 3, and Sections 4.5, 4.6). These textbooks also contain numerous references for further reading.

Keeping in line with the overall tenor of the paper, below we single out some areas of current research where QRM and statistics have an interesting scientific intersection as well as offer research of practical relevance. The modeling of interdependence between risk factors or risk-factor changes is a typical such example, it also lies at the very basis of most relevant QRM questions. Examples are to be found in market risk management already discussed in the previous section. The field of credit risk offers a clear playground for interdependence modelers; indeed most credit-risk portfolios exhibit interdependence, especially under extreme market conditions. The various financial crises (subprime, sovereign,...) give ample of proof. For instance the famous AAA-guarantee given by the rating agencies for the (super-)senior tranches of Collateralized Debt Obligations (CDOs) turned out to be well below this low-risk level when the American housing marked collapsed around 2006. An early warning for the consequences of even a slight increase in interdependence between default-events of credit positions for the evaluation of securitized products like CDOs is to be found in McNeil et al. (2005, Figure 8.1); see also Hofert and Scherer (2011). The bigger methodological umbrella to be put over such examples is Model Uncertainty. The latter is no doubt one of the most active and practically relevant research fields in QRM at the moment. We shall only highlight some aspects below, starting first with the notion of copula.

In the late nineties, the concept of copula took QRM by storm. A **copula** is simply a multivariate df with standard uniform univariate margins. Copulas provide a convenient and comparably simple (mainly static) way of describing the dependence between the components $X_1, \dots, X_d$ of a random vector $\boldsymbol{X}$ (the risk-factor changes in (4) for example). A publication which sparked this hype no doubt is Embrechts et al. (2002); see Figure 3 based on the data collected by Genest et al. (2009) (unfortunately, the latter database was not maintained beyond 2005). This paper was available as a RiskLab preprint since early 1998

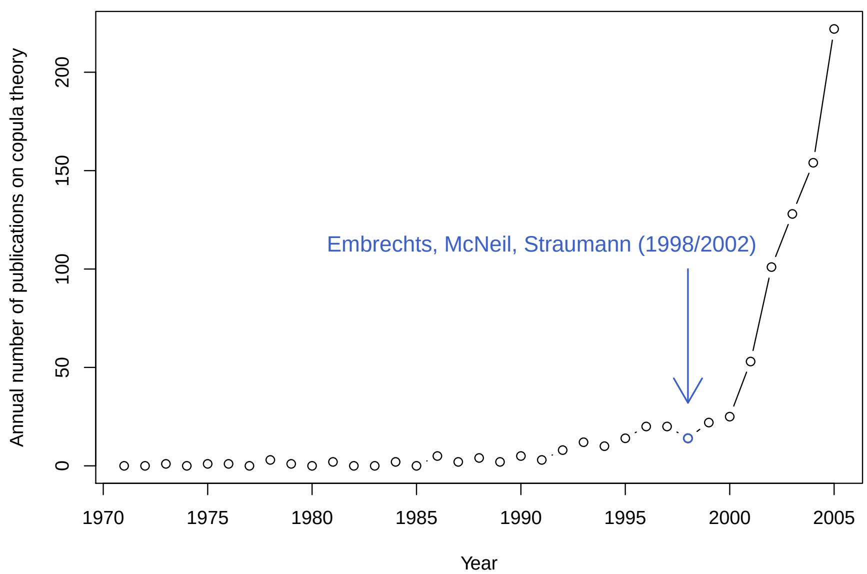

Figure 3

Number of publications on copula theory.

and was also published in a much abbreviated form in Embrechts et al. (1999). For the record, Daniel Straumann summarized the main findings of Embrechts et al. (1999) at the first ETH Risk-Day on September 25, 1998 in a talk "Tempting fallacies in the use of correlation". The first author, Paul Embrechts, gave a similar talk at the 3rd Columbia-JAFEE conference at Columbia University on March 27, 1999, this time with a title "Insurance analytics: actuarial tools in financial risk management". For some of this QRM-relevant background, see

Donnelly and Embrechts (2010). We would like to stress, however, that the copula concept has been around almost for as long as modern probability and statistics emerged in the early 20th century. Related work can be traced back to Fréchet (1935) and Hoeffding (1940), for example.

The central theorem underlying the theory and applications of copulas is **Sklar's Theorem**; see Sklar (1959). It consists of two parts. The first one states that for any multivariate df H with margins $F_1, \dots, F_d$, there exists a copula C such that

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \ \boldsymbol{x} \in \mathbb{R}^d. \tag{6}$$

This decomposition is not necessarily unique, but it is in the case where all margins $F_1, \dots, F_d$ are continuous. The second part provides a converse statement: Given any copula C and univariate dfs $F_1, \dots, F_d$, H defined by (6) is a df with margins $F_1, \dots, F_d$. An analytic proof of Sklar's Theorem can be found in Sklar (1996), a probabilistic one in Rüschendorf (2009).

In what follows we assume the margins $F_1, \dots, F_d$ to be continuous. In the context of finance and insurance applications, this assumption is typically not considered restrictive, but it is, for example, when counting data plays an important role, such as for medical or pharmacological applications. In general, it certainly becomes a more restrictive assumption if d is large. For the case of non-continuous margins (with all its pitfalls in thinking about dependence) we refer the interested reader to Genest and Nešlehová (2007). For now, we focus on the case of continuous margins and thus have a unique representation (6) for H in terms of the copula C and the margins of H. From this representation we see that the copula C is precisely the function which combines (or "couples", hence the name) the margins $F_1, \dots, F_d$ of H to the joint df H. To understand the importance of Sklar's Theorem for statistics and QRM, in particular the aforementioned hype, we now interpret (6) in terms of two data sets shown in Figure 4. The data sets show $n = 500$ realizations of $(X_1, X_2) \sim H$ with F_1 and F_2 being standard normal dfs on the left-hand side, and F_1 and F_2 being standard exponential dfs on the right-hand side. The generated data sets could be historical risk-factor changes, for example. Questions which immediately come up, are:

- How can the dependence between X_1 and X_2 be modeled based on these two data sets?
- Which of the data sets shows "stronger" dependence between X_1 and X_2?

Although graphics such as Figure 4 are used in various fields of application to answer such questions, it is highly critical to do so. In the two plots in Figure 4, the margins live on quite different scales and thus disguise the underlying dependence structure, the C in (6). To make a comparison possible, we transform the margins to $\mathrm{U}[0,1]$, i.e. the standard uniform distribution. More precisely, we consider $\boldsymbol{U} = (F_1(X_1), F_2(X_2))$; note that if $X_j \sim F_j$ and F_j is continuous then $F_j(X_j) \sim \mathrm{U}[0,1]$. The results are shown in Figure 5. Although the data sets in Figure 4 do not seem to have much in common, after removing the influence of the margins, we suddenly see that they do! We obtained the exact same samples

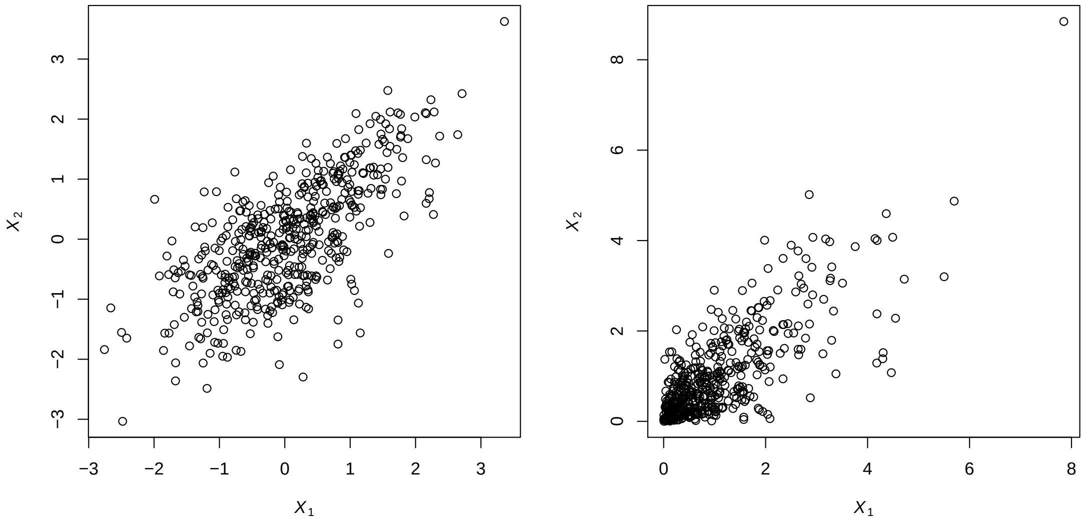

Figure 4

500 samples from $(X_1, X_2) \sim H$ where $F_1 = F_2 = F$ with $F(x) = \Phi(x)$, i.e., the df of the standard normal distribution $\mathrm{N}(0,1)$ (left), and $F(x) = 1 - \exp(-x)$, i.e., the df of the standard exponential distribution Exp(1) (right).

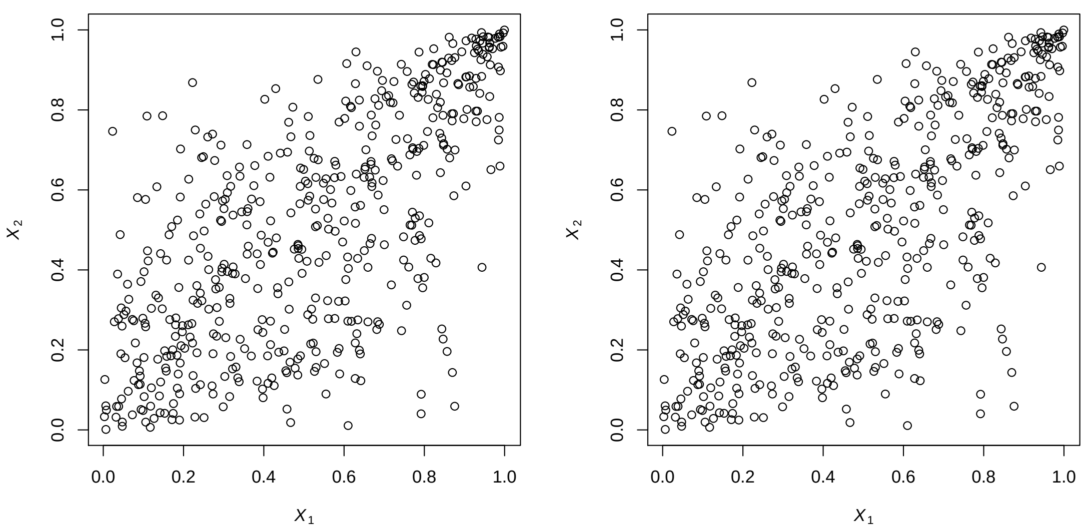

Figure 5

The 500 samples from $(U_1, U_2) = (F_1(X_1), F_2(X_2)) \sim C$ underlying the corresponding plots in Figure 4. They are (even exactly; not only in distribution) identical.

(not only samples equal in distribution). Since they have U$[0, 1]$ margins, we indeed see a sample from the copula C in (6). The first part of Sklar's Theorem thus allows us to make C visible in this way, to decompose H (or (X_1, X_2)) into its margins F_1, F_2 (or X_1, X_2; note the missing vector notation here) and the underlying dependence structure C (or (U_1, U_2)).

In a practical situation, one does not know the margins F_1, F_2. In this case one typically replaces them by their (slightly scaled) empirical dfs in order to make (an approximation to) C visible. One obtains the so-called **pseudo-observations** $\hat{\boldsymbol{U}}_i$, $i \in \{1, \dots, n\}$, given by

$$\hat{U}_{ij} = \frac{n}{n+1}\hat{F}_{n,j}(X_{ij}) = \frac{R_{ij}}{n+1}, \quad j \in \{1, \dots, d\}, \tag{7}$$

where $\hat{F}_{n,j}$ denotes the empirical df of the jth margin and R_{ij} denotes the rank of X_{ij} among $X_{1j}, \dots, X_{nj}$. A plot of the pseudo-observations similar to Figure 5 is therefore referred to as **rank plot**. It may easily be computed with the free statistical software R via `plot(apply(x, 2, rank, na.last = "keep") / (nrow(x) + 1))` where `x` denotes the (n, d)-matrix whose rows contain realizations of $\boldsymbol{X}$.

The first part (decomposition) of Sklar's Theorem proves especially useful if d is large. Assume $d = 100$ and each of the margins to have two parameters. Furthermore, assume the copula to have one parameter. Computing a maximum likelihood estimator for all parameters in such a set-up would amount to solving an optimization problem in 201 dimensions. By estimating the parameters of the margins individually and estimating the copula parameter based on the pseudo-observations (the scaling $n/(n + 1)$ in (7) is used so that the copula density can be evaluated at the pseudo-observations), one only has to solve 100 optimization problems in two dimensions and one univariate optimization problem. Note that all of these can be carried out in parallel as they do not depend on each other. This approach is not a panacea for solving all estimation problems of multivariate models since the pseudo-observations are not independent anymore (due to the ranks). This influences the estimator, especially in high dimensions; see Embrechts and Hofert (2013b). However, it makes parameter estimation feasible and comparably fast even in high dimensions (depending on the problem at hand) and is therefore widely used in practice.

The second part of Sklar's Theorem basically reads (6) backwards. Given any marginal dfs $F_1, \dots, F_d$ (continuous or not) and any copula C, H in (6) is a multivariate df with margins $F_1, \dots, F_d$. This provides us with a concrete construction principle for new multivariate dfs. Important for finance and insurance practice is that it gives us additional modeling options besides the multivariate normal or t distribution, it thus allows us to construct more flexible, realistic multivariate models for $\boldsymbol{X}$. This is also used for stress testing to answer questions like

- How does a VaR estimate react if one keeps the margins fixed but change the dependence?
- How does it change if the margins vary but the dependence remains fixed?

Another important aspect of the second part of Sklar's Theorem is that it gives us a sampling algorithm for dfs H constructed from copula models. This is heavily used in practice as realistic models are rarely analytically tractable and thus are often evaluated based on simulations. The algorithm implied by the second part of Sklar's Theorem for sampling $\boldsymbol{X} \sim H$ where H is constructed from a copula C and univariate dfs $F_1, \dots, F_d$ is as follows:

Algorithm 3.1

1) Sample $\boldsymbol{U} \sim C$;
2) Return $\boldsymbol{X} = (X_1, \dots, X_d)$ where $X_j = F_j^{\leftarrow}(U_j)$, $j \in \{1, \dots, d\}$.

It should by now be clear to the reader how we constructed Figures 4 and 5. In line with Algorithm 3.1, we first constructed a sample of size 500 from $\boldsymbol{U} \sim C$ for C being a Gumbel copula (with parameter such that Kendall's tau equals 0.5). This is shown in Figure 5. We then used the quantile function of the standard normal distribution to map $\boldsymbol{U}$ to $\boldsymbol{X}$ having $\mathrm{N}(0,1)$ margins (shown on the left-hand side of Figure 4) and similarly for the standard exponential margins on the right-hand side of Figure 4.

After this brief introduction to the modeling of stochastic dependence with copulas, a few remarks are in order. Summaries of the main results partly including historical overviews can be found in the textbooks of Schweizer and Sklar (1983, Chapter 6), Nelsen (1999), Cherubini et al. (2004), or Jaworski et al. (2010). The relevance for finance is summarized in Embrechts (2009). Adding Schweizer (1991), both papers together give an idea of the key developments. Embrechts (2009) also recalls some of the early discussions between protagonists and antagonists of the usefulness of copulas in insurance and finance in general and QRM in particular.

History has moved on and by now copula technology forms a well-established tool set for QRM. Like VaR, also the copula concept is marred with warnings and restrictions when it comes to its applications. An early telling example is the use of the Gaussian copula (the copula C corresponding to H being multivariate normal via (6)) in the realm of CDO-pricing, an application which for some lay at the core of the subprime crisis; see for instance the online publication (blog) Salmon (2009). Somewhat surprisingly, the statistical community bought the arguments; see Salmon (2012). This despite the fact that the weakness (even impossibility) of the Gaussian copula for modeling joint extremes was known for almost 50 years: Sibuya (1959) showed that the Gaussian copula does not exhibit a notion referred to as tail dependence; see McNeil et al. (2005, Section 5.2.3). Informally, the latter notion describes the probability that one (of two) rvs takes on a large (small) value, given that the other one takes on a large (small) value; in contrast to the Gaussian copula, a Gumbel copula (being part of the Archimedean class) is able to capture (upper) tail dependence and we may already suspect that by looking at Figure 5. With the notion of tail dependence at hand, it is not difficult to see its importance for CDO pricing models (or more generally, intensity-based default models) such as the one based on Gaussian copulas (see Li (2000)), Archimedean copulas (see Schönbucher and Schubert (2001)), or nested Archimedean copulas (see Hofert and Scherer

(2011)); the latter two references introduce models which are able to capture tail dependence. Due to its importance for QRM, Sibuya (1959) and, more broadly, tail dependence was referred to in the original Embrechts et al. (2002, Section 4.4). We repeat the main message from the latter paper: copulas are a very useful concept for better understanding the concept of stochastic dependence, not a panacea for its modeling. The title of Lipton and Rennie (2007) is telling in this context; see also the discussions initiated by Mikosch (2006).

As mentioned before, within QRM, copulas have numerous applications towards risk aggregation and stress testing, and as such have entered the regulatory guidelines. Below we briefly address some of the current challenges; they are mainly related to high-dimensionality, statistical estimation, and numerical calculations.

A large part of research on copulas is conducted and presented in the bivariate case ($d = 2$). Bivariate copulas are typically convenient to work with, the case $d = 2$ offers more possibilities to construct copulas (via geometric approaches), required quantities can often be derived by hand, graphics (of the df or, if it exists, its density) can still be drawn without losing information, and numerical issues are often not severe or negligibly small. The situation is different for $d > 2$ and fundamentally different for $d \gg 2$ (even if we stay far below $d = 1000$). In general, the larger the dimension d, the more complicated it is to come up with a numerically tractable model which is flexible enough to account for real data behavior. Joe (1997, pp. 84) presents several desirable properties of multivariate distributions. However, to date, no model is known which exhibits all such properties.

The t copula belongs to the most widely used copula models; see Demarta and McNeil (2005) for an introduction and possible extensions. Besides an explicit density, it admits tail dependence and can be sampled comparably fast; it also incorporates the Gaussian copula as limiting case, which, despite all its limitations, is still a popular model. This makes the t copula especially attractive for practical applications even in large dimensions. From a more technical point of view, a d-dimensional t copula with $\nu > 0$ degrees of freedom and correlation matrix P can be written as

$$C(\boldsymbol{u}) = t_{\nu,P}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)), \quad \boldsymbol{u} \in [0,1]^d,$$

where $t_{\nu,P}$ denotes the df of a multivariate t distribution with ν degrees of freedom and correlation matrix P (corresponding to the dispersion matrix Σ in terms of which the multivariate t distribution is defined in general; see Demarta and McNeil (2005)) and t_ν^{-1} denotes the quantile function of a univariate t distribution with ν degrees of freedom. From this formula we already see that the components involved in evaluating a t copula are non-trivial. Neither $t_{\nu,P}$ nor t_ν^{-1} are known explicitly. Especially the former is a problem and for $d > 3$, randomized quasi Monte Carlo methods are used (in the well-known `R` package `mvtnorm`) to evaluate the d-dimensional integral; see Genz and Bretz (2009, Section 5.5.1). Furthermore, the algorithm requires ν to be an integer, which is a limitation for which no easy solution is available yet.

The Gaussian and t copula families belong to the class of elliptical copulas and are given **implicitly** by solving (6) for C in the case where H is the df of an elliptical distribution. A different ansatz stems from Archimedean copulas, which are given **explicitly** in terms of a so-called **generator** $\psi : [0,\infty) \to [0,1]$ via

$$C(\boldsymbol{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)), \quad \boldsymbol{u} \in [0,1]^d; \tag{8}$$

see McNeil and Nešlehová (2009). For the Gumbel copula family, $\psi(t) = \exp(-t^{1/\theta})$, $\theta \geq 1$. Archimedean copulas allow for a fast sampling algorithm, see Marshall and Olkin (1988), and, in contrast to elliptical copulas, are not limited to radial symmetry. Elliptical copulas put the same probability mass in the upper-right corner than in the lower-left corner of the distribution and thus value joint large gains in the same way as joint large losses. On the other hand, it directly follows from (8) that Archimedean copulas are symmetric, that is, invariant under permutation of the arguments. This is also visible from the symmetry with respect to the diagonal in Figure 5. Still, the flexibility in the (upper or lower) tail of copulas such as the Gumbel are interesting for practical applications and Archimedean copula families such as the Clayton or Gumbel belong to the most widely used copula models (the Gumbel family additionally belongs to the class of extreme value copulas and is one of the rare comparably tractable examples in this class).

The rather innocent functional form (8) bears interesting numerical problems, already in small to moderate dimensions which have even led to wrong statements in the more recent literature. The copula density is easily seen to be

$$c(\boldsymbol{u}) = (-1)^d \psi^{(d)}(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)) \prod_{j=1}^{d} (-\psi^{-1})'(u_j), \quad \boldsymbol{u} \in (0,1)^d.$$

In particular, it involves the dth generator derivative. For d large, these are difficult to access. For the Gumbel copula, an explicit form for $(-1)^d\psi^{(d)}$ has only recently been found; see Hofert et al. (2012). The formula is

$$(-1)^d \psi^{(d)}(t) = \frac{\psi(t)}{t^d} \sum_{k=1}^{d} a_{dk}(\alpha) t^{\alpha k},$$
$$a_{dk}(\alpha) = (-1)^{d-k} \sum_{j=k}^{d} \alpha^j s(d,j) S(j,k) = \frac{d!}{k!} \sum_{j=1}^{k} \binom{k}{j} \binom{\alpha j}{d} (-1)^{d-j},$$

where $\alpha = 1/\theta \in (0,1]$ and s and S denote the **Stirling numbers of the first** and **the second kind**, respectively. The mathematical beauty disguises the numerical problems in evaluating $(-1)^d\psi^{(d)}(t)$; see Hofert et al. (2013) and the source code of the `R` package `copula` for solutions. Even if we can evaluate the generator derivatives, Figure 6 shows what we try to achieve by going towards higher dimensions; note the scale of the y axis and the steep curve near zero even in log-scale.

Derivatives of Archimedean generators not only appear in densities of Archimedean copulas, they appear in other, more flexible dependence models based on Archimedean copulas, such as Khoudraji-transforms, Archimedean Sibuya

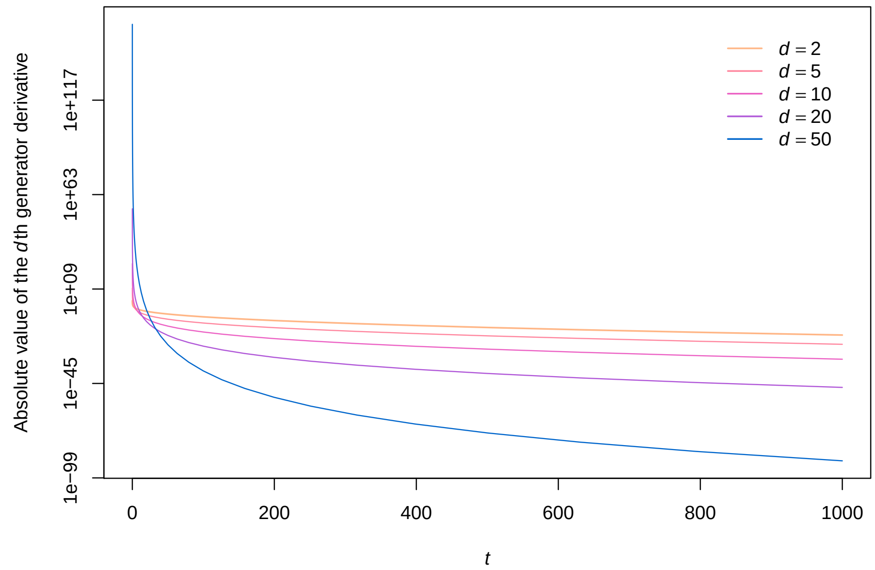

Figure 6

Generator derivatives for Gumbel copulas with parameter such that Kendall's tau equals 0.5.

copulas, nested Archimedean copulas, but also in statistically highly relevant quantities such as conditional distributions or Kendall dfs. Problems as the above also appear for example when computing the densities of copulas describing the dependence in multivariate extreme value distributions. Besides the theoretical challenge to derive explicit mathematical formulas, it is generally an (even larger) challenge to evaluate the formulas in a fast and numerically reliable way in computer arithmetic. In high dimensions, one soon enters the world of multi-precision arithmetic (mainly to test, but also to be able to compute certain quantities at all) and parallel computing (nowadays of special interest due to multi-core processors even in standard notebooks).

Let us now turn to another aspect of high-dimensional models in general. Assume we fitted a d-dimensional copula C to the pseudo-observations obtained from realizations of the vector $\boldsymbol{X}$ of risk-factor changes. Suppose we would like to compute the probability that $\boldsymbol{X}$ falls in a non-empty hypercube $(\boldsymbol{y}, \boldsymbol{z}]$ (the most basic shapes we typically compute probabilities of). This would require us to evaluate C at 2^d many points, the corners of $(\boldsymbol{a}, \boldsymbol{b}] \subseteq [0,1]^d$, where $\boldsymbol{a} = (F_1(y_1), \dots, F_d(y_d))$ and $\boldsymbol{b} = (F_1(z_1), \dots, F_d(z_d))$. This **$C$-volume** can be computed via

$$\Delta_{(\boldsymbol{a},\boldsymbol{b}]}C = \sum_{\boldsymbol{j}\in\{0,1\}^d} (-1)^{\sum_{k=1}^d j_k} C(a_1^{j_1} b_1^{1-j_1}, \dots, a_d^{j_d} b_d^{1-j_d}).$$

Although theoretically "just" a sum, we indeed have 2^d summands of alternating sign. If we can not exactly evaluate C (and this is even a problem if C is explicit such as in the Gumbel case), we soon, for increasing d, end up with a volume outside $[0,1]$. Furthermore, 2^d grows very fast, it is thus not only numerically critical but also time consuming to compute probabilities of falling in hypercubes. Indeed, for $d \approx 270$ the number of corners of a d-dimensional hypercube is roughly equal to the number of atoms in the universe. Clearly, for $d = 1000$ approximation schemes and possibly new approaches must be developed.

New high-dimensional dependence models for $\boldsymbol{X}$ that recently became of interest are hierarchical models. Through a matrix of parameters, elliptical copulas such as the Gaussian or t copulas allow for $\binom{d}{2}$-many (an additional one for the t copula) parameters. Symmetric models such as (8) typically allow for just a handful of parameters, depending on the parametric generator under consideration. Hierarchical models try to go somewhere in-between, which is often a good strategy for d large. While estimating $\binom{d}{2}$-many parameters may be time-consuming and, for example in the case of a small sample size, not justifiable by the data at hand or task in mind, using the assumption of symmetry for blocks or groups of variables among $\boldsymbol{X}$ allows to reduce the number of parameters to fall in between a fully symmetric model and a model considering parameters for all $\binom{d}{2}$ pairs of variables. Extending (8) to allow for hierarchies can be achieved in various ways, see for example McNeil (2008) or Hofert (2012). Reducing the parameters in elliptical models can be done by considering block matrices; see Torrent-Gironella and Fortiana (2011) which is one of the rare publications carefully considering numerical issues behind the scenes as well. For a different idea to incorporate groups in t copulas, see Demarta and McNeil (2005).

SUMMARY POINTS

Overall, to our experience, there are many challenges in the area of high dimensional dependence modeling. It does not seem to be a pleasing task to go there as numerics will heavily enter soon after one leaves the "comfort zone" of d being as small as two or three. However, from a practical point of view it is necessary. Working in high dimensions may require starting from scratch, model building already has to take numerics into account, not only mathematical beauty or convenience. The same applies to the (even more complicated) case of time dependent models not discussed here. To give an example, sampling dependent Lévy processes with the notion of Lévy copulas to model dependence between jumps is typically also numerically a non-trivial task and limited to rather small dimensions. Although mathematically d is just a natural number and can be increased arbitrarily in the above discussion and formulas, solutions to the problems mentioned above heavily depend on d. In the end of the (bank's) day, models, estimators, stress tests etc. have to be **computed** and thus, additionally, require reliable software behind the scenes. It is to be encouraged that academia contributes here as well (and gets scientifically rewarded for entering this time-consuming process), as new statistical models and procedures (including their limitations) are typically best understood by those who invented them. Being pressed for time in business practice often requires to implement a model in a small amount of time, basically until it "works". However, proper testing of statistical software has to go far beyond this point and extensive test suites have to be put in place.

Open source statistical software (slowly finding its way into business practice but still not used to a sufficient degree) typically guarantees more stable and reliable procedures as anyone can report bugs, for example. Furthermore, open source increases the understanding of new models and the recognition of their limitations. Software development also has to be communicated on an academic level as it is already part of (but also limited to) statistics courses. Preparing students for (or at least to be aware of) the computational challenges they will meet in practice should rather become the rule than the exception.

4. Risk aggregation and model uncertainty

The subadditivity property of a risk measure is often defended in order to achieve a coherent, consistent way to aggregate risks across different entities (business lines, say) and also in order to find methodologically consistent ways to disaggregate, also referred to risk allocation. Further, in the background lies the concept of risk diversification. In this section we will give a very brief overview of some of the underlying mathematical issues. Given a general risk measure R (examples treated so far include $R = \mathrm{VaR}$, $R = \mathrm{ES}$), for d risks $X_1, \dots, X_d$, we define the coefficient of diversification as

$$D_R(\boldsymbol{X}) = \frac{R(\sum_{j=1}^{d} X_j)}{\sum_{j=1}^{d} R(X_j)}.$$

In the superadditive case, $D_R(\boldsymbol{X})$ may be larger than 1; an important question is "by how much?", and this in particular as a function of the various probabilistic assumptions on the joint df H of $\boldsymbol{X}$. Model Uncertainty in particular enters if we only have partial information on the marginal dfs $F_1, \dots, F_d$ and a/the copula C; see our discussion in the previous section. The mathematical problem underlying the calculation of $D_R(\boldsymbol{X})$ has a long history. Below we will content ourselves with a very brief summary pointing the reader to some of the basic literature; we will also present some examples aimed at understanding the range of values for $D_R(\boldsymbol{X})$ if only information on $F_1, \dots, F_d$ is available (the unconstrained problem).

The reference on the topic, from a more mathematical point of view is Rüschendorf (2013). As can be seen from the title, it contains all relevant topics to be addressed in this section. It also contains an extensive list of historically relevant references. We know that in the case of $R = \mathrm{VaR}$, $D_R(\boldsymbol{X}) > 1$ typically happens for the F_j's being very heavy tailed (e.g. independent and infinite mean), very skew (see Example 2.3) or for the risks $X_1, \dots, X_d$ exhibiting a special dependence structure (even with $F_1 = \cdots = F_d$ being $\mathrm{N}(0, 1)$); McNeil et al. (2005, Example 6.22). Interesting questions for practice now are: (Q1) Given a specific model leading to superadditivity of $D_R(\boldsymbol{X})$, by how much can $D_R(\boldsymbol{X}) > 1$, and (Q2) which dependence structures between $X_1, \dots, X_d$ lead to such extreme cases. Are these realistic from a practical point of view? And finally, (Q3) under what kind of dependence information between $X_1, \ldots, X_d$ can such bounds be computed. Besides the above textbook reference, questions (Q1)–(Q3) are discussed in detail in Embrechts et al. (2013). From the latter paper we borrow the example below.

Example 4.1
Suppose $X_j \sim \text{Pareto}(2)$, $j \in \{1, \ldots, d\}$, i.e. $1 - F_j(x) = \mathbb{P}(X_j > x) = (1+x)^{-2}$, $x \geq 0$. For $d = 8$ and $\alpha = 0.99$, $\text{VaR}_\alpha(X_j) = 9$ resulting in $\sum_{j=1}^{8} \text{VaR}_\alpha(X_j) = 72$, the so-called comonotonic case for C; see McNeil et al. (2005, Proposition 6.15). An exact upper bound on $D_R(\boldsymbol{X})$ can be given in this case, it equals about 2; see Table 4 in Embrechts et al. (2013). The latter paper also contains lighter tailed examples like log-normal and gamma; see Figure 4 in the latter paper. Here the upper bound for $D_R(\boldsymbol{X})$ for $\alpha = 0.99$ is between 1.15 and 1.5. The main lesson learned is that $D_R(\boldsymbol{X})$ can be considerably larger than 1 in the unconstrained case. Constraining the rvs to e.g. positive quadrant dependence does not change the bounds by much. For $\boldsymbol{X}$ being elliptically distributed, or for $R = \text{ES}$, the upper bound is always 1! Note that for $D_R(\boldsymbol{X}) > 1$, we enter the world of non-coherence, i.e., the risk of the overall portfolio position $\sum_{j=1}^{d} X_j$ is not covered, in a possibly conservative way, by the sum of the marginal risk contributions. In finance parlance, people would say that in such a case, "diversification does not work".

For the above kind of problems, techniques from the realm of Operations Research will no doubt turn out to be relevant. Key tools to look for are Robust and Convex Optimization; see for instance Ben-Tal et al. (2009) and Boyd and Vandenberghe (2004). Early work, stressing the link between Operations Research and QRM, especially in the context of the calculation of ES is by Stan Uryasev and R. Tyrell Rockafellar; see for instance Uryasev and Rockafellar (2013) and Chun et al. (2012).

SUMMARY POINTS

In this final section, we only briefly entered into the field of Model Uncertainty. Going forward, and in the wake of the recent financial crisis, this area of research will gain importance. Other areas of promising research, which for reasons of space limitations we were not able to discuss, are High Frequency (Inter Day) Data in finance, and Systemic Risk (Network Theory). For the former, the classic "that started it all" is Dacorogna et al. (2001). By now, the field is huge! A more recent edited volume is Viens et al. (2012). For a start on some of the more mathematical work on Systemic Risk and Network Theory, we suggest interested readers to start with papers on the topic by Rama Cont (Imperial College, London) and Tom Hund (McMaster University, Hamilton, Canada).

Acknowledgements

We apologize in advance to all the investigators whose research could not be appropriately cited owing to space limitations.

References

Albanese, C (1997), Credit exposure, diversification risk and coherent VaR, preprint, Department of Mathematics, University of Toronto.

Andersen, TG, Bollerslev, T, Christoffersen, PF, and Diebold, Fx (2013), Financial risk measurement for financial risk management, **Handbook of the Economics of Finance**, ed. by G Constantinides, M Harris, and R Stulz, vol. 2, Amsterdam: Elsevier Science B. V., 1127–1220.

Artzner, P, Delbaen, F, Eber, JM, and Heath, D (1999), Coherent measures of risk, **Mathematical Finance**, 9, 203–228.

Balo, Y, Lee, TH, and Saltoğlu, B (2006), Evaluating predictive performance of Value-at-Risk models in emerging markets: a reality check, **Journal of Forecasting**, 25, 101–128.

Ben-Tal, A, El Ghaoui, L, and Nemirovski, A (2009), Robust Optimization, Princeton University Press.

Berkowitz, J and O'Brien, J (2002), How accurate are Value-at-Risk models at commercial banks? **Journal of Finance**, 57, 1093–1111.

BIS (2012), Fundamental review of the trading book, `http://www.bis.org/publ/bcbs219.pdf` (2012-02-03).

Boyd, S and Vandenberghe, L (2004), Convex Optimization, Princeton University Press.

Chavez-Demoulin, V and McGill, JA (2012), High-frequency financial data modeling using Hawkes processes, **Journal of Banking and Finance**, 36(12), 3415–3426.

Chavez-Demoulin, V, Davison, AC, and McNeil, AJ (2005), Estimating value-at-risk: a point process approach, **Quantitative Finance**, 5(2), 227–234.

Chavez-Demoulin, V, Embrechts, P, and Sardy, S (2013a), Extreme-quantile tracking for financial time series, **Journal of Econometrics**, to appear.

Chavez-Demoulin, V, Embrechts, P, and Hofert, M (2013b), An extreme value approach for modeling operational risk losses depending on covariates.

Cherubini, U, Luciano, E, and Vecchiato, W (2004), Copula Methods in Finance, Wiley.

Christian Gourieroux, JJ (2001), Financial Econometrics: Problems, Models and Methods, Princeton: Princeton University Press.

Chun, SY, Shipiro, A, and Uryasev, S (2012), Conditional Value-at-Risk and Average Value-at-Risk: Estimation and Asymptotics, **Operations Research**, 60(4), 739–756.

Cont, R, Deguest, R, and Scandolo, G (2010), Robustness and sensitivity analysis of risk measurement procedures, **Quantitative Finance**, 10(6), 593–606.

Dacorogna, MM, Gençay, R, Müller, U, Olsen, RB, and Pictet, OV (2001), An Introduction to High-Frequency Finance, Academic Press.

Daníelsson, J (2011), Financial Risk Forecasting, Chichester: Wiley.

Das, B, Embrechts, P, and Fasen, V (2013), Four Theorems and a Financial Crisis, **International Journal of Approximate Reasoning**, to appear, `http://www.math.ethz.ch/~embrechts/ftp/Four_Theorems_2012.pdf` (2013-02-03).

Delbaen, F (2012), Monetary Utility Functions, Osaka University Press.

Dell'Aquila, R and Embrechts, P (2006), Extremes and robustness: a contradiction? **Financial Markets and Portfolio Management**, 20, 103–118.

Demarta, S and McNeil, AJ (2005), The t Copula and Related Copulas, **International Statistical Review**, 73(1), 111–129.

Diebold, FX, Doherty, NA, and J., HR (2010), The Known, the Unknown, and the Unknowable in Financial Risk Management: Measurement and Theory Advancing Practice, Princeton University Press.

Donnelly, C and Embrechts, P (2010), The devil is in the tails: actuarial mathematics and the subprime mortgage crisis, **ASTIN Bulletin**, 40(1), 1–33.

Embrechts, P (2009), Copulas: A personal view, **Journal of Risk and Insurance**, 76, 639–650.

Embrechts, P, McNeil, AJ, and Straumann, D (1999), Correlation: Pitfalls and Alternatives, **RISK Magazine**, 69–71, `http://www.math.ethz.ch/~mcneil/ftp/risk.pdf` (2013-02-16).

Embrechts, P, McNeil, AJ, and Straumann, D (2002), Correlation and Dependency in Risk Management: Properties and Pitfalls, **Risk Management: Value at Risk and Beyond**, ed. by M Dempster, Cambridge University Press, 176–223.

Embrechts, P, Puccetti, G, and Rüschendorf, L (2013), Model uncertainty and VaR aggregation, **Journal of Banking and Finance**, 37(8), 2750–2764.

Embrechts, P and Hofert, M (2013a), A note on generalized inverses, **Mathematical Methods of Operations Research**, `doi:http://dx.doi.org/10.1007/s00186-013-0436-7`.

Embrechts, P and Hofert, M (2013b), Statistical inference for copulas in high dimensions: A simulation study, **ASTIN Bulletin**, 43(2), 81–95, `doi:http://dx.doi.org/10.1017/asb.2013.6`.

Embrechts, P, Kaufmann, R, and Patie, P (2005), Strategic long-term financial risks: single risk factors, **Computational Optimization and Applications**, 32(1–2), 61–90.

Engle, R (2009), Anticipating Correlations: A New Paradigm for Risk Management, Princeton University Press.

Fishburn, PC (1984), Foundations of Risk Measurement. I. **Management Science**, 30, 396–406.

Föllmer, H and Schied, A (2011), Stochastic Finance: An Introduction in Discrete Time, 3rd ed., de Gruyter.

Fréchet, M (1935), Généralisations du théorème des probabilités totales, **Fundamenta Mathematica**, 25, 379–387.

Gale, D and Shapley, LS (1962), College admissions and stability of marriage, **The American Mathematical Monthly**, 69(1), 9–15.

Genest, C and Nešlehová, J (2007), A primer on copulas for count data, **The Astin Bulletin**, 37, 475–515.

Genest, C, Gendron, M, and Bourdeau-Brien, M (2009), The Advent of Copulas in Finance, **The European Journal of Finance**, 15, 609–618.

Genton, MG (2004), Skew-Elliptic Distributions and Their Applications: A Journey Beyond Normality, Boca Raton: Chapman & Hall/CRC.

Genz, A and Bretz, F (2009), Computation of Multivariate Normal and t Probabilities, Springer.

Giacomini, R and Komunjer, I (2005), Evaluation and combination of conditional quantile forecasts, **Journal of Business & Economic Statistics**, 23, 416–431.

Gilboa, I and Schmeidler, D (1989), Maximum expected utility with a non-unique prior, **Journal of Mathematical Economics**, 18, 141–153.

Gneiting, T (2011), Making and evaluating point forecasts, **Journal of the Americal Statistical Association**, 106, 746–762.

Hampel, FR, Ronchetti, EM, Rousseeuw, PJ, and Stahel, WA (2005), Robust Statistics: The Approach Based on Influence Functions, Wiley-Interscience.

Hoeffding, W (1940), Massstabinvariante Korrelationstheorie, **Schriften des mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin**, 5, 181–233.

Hofert, M (2012), A stochastic representation and sampling algorithm for nested Archimedean copulas, **Journal of Statistical Computation and Simulation**, 82(9), 1239–1255, `doi:http://dx.doi.org/10.1080/00949655.2011.574632`.

Hofert, M and Scherer, M (2011), CDO pricing with nested Archimedean copulas, **Quantitative Finance**, 11(5), 775–787, `doi:http://dx.doi.org/10.1080/14697680903508479`.

Hofert, M, Mächler, M, and McNeil, AJ (2012), Likelihood inference for Archimedean copulas in high dimensions under known margins, **Journal of Multivariate Analysis**, 110, 133–150, `doi:http://dx.doi.org/10.1016/j.jmva.2012.02.019`.

Hofert, M, Mächler, M, and McNeil, AJ (2013), Archimedean Copulas in High Dimensions: Estimators and Numerical Challenges Motivated by Financial Applications, **Journal de la Société Française de Statistique**.

Huber, PJ (1981), Robust Statistics, Wiley.

Huber, PJ and Ronchetti, EM (2009), Robust Statistics, 2nd ed., Wiley.

Jaworski, P, Durante, F, Härdle, WK, and Rychlik, T, eds. (2010), Copula Theory and Its Applications, vol. 198, Lecture Notes in Statistics – Proceedings, Springer.

Joe, H (1997), Multivariate Models and Dependence Concepts, Chapman & Hall/CRC.

Jorion, P (2007), Value at Risk: The New Benchmark for Managing Financial Risk, 3rd ed., New York: McGraw-Hill.

Kaufmann, R (2004), Long-Term Risk Management, no. 15595, PhD thesis, Department of Mathematics, ETH Zurich.

Knight, FH (1921), Risk, Uncertainty and Profit, Houghton Mifflin Co.

Li, DX (2000), On Default Correlation: A Copula Function Approach, **The Journal of Fixed Income**, 9(4), 43–54.

Lipton, A and Rennie, A (2007), Credit Correlation: Life After Copulas, World Scientific Publishing Company.

Mancini, L and Trojani, F (2011), Robust Value-at-Risk prediction, **Journal of Financial Econometrics**, 9, 281–313.

Marshall, AW and Olkin, I (1988), Families of Multivariate Distributions, **Journal of the American Statistical Association**, 83(403), 834–841.

McNeil, AJ (2008), Sampling nested Archimedean copulas, **Journal of Statistical Computation and Simulation**, 78(6), 567–581.

McNeil, AJ and Frey, R (2000), Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, **Journal of Empirical Finance**, 7, 271–300.

McNeil, AJ and Nešlehová, J (2009), Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions, **The Annals of Statistics**, 37(5b), 3059–3097.

McNeil, AJ, Frey, R, and Embrechts, P (2005), Quantitative Risk Management: Concepts, Techniques, Tools, Princeton University Press.

Mikosch, T (2006), Copulas: Tales and facts, **Extremes**, 9(1), 3–20.

Nelsen, RB (1999), An Introduction to Copulas, Springer.

Rémillard, B (2013), Statistical Methods For Financial Engineering, Chapman & Hall/CRC.

Roth, A and Shapley, L (2012), Game, set and match, `http://www.economist.com/news/finance-and-economics/21564836-alvin-roth-and-lloyd-shapley-have-won-year%E2%80%99s-nobel-economics` (2013-01-16).

Rüschendorf, L (2009), On the distributional transform, Sklar's Theorem, and the empirical copula process, **Journal of Statistical Planning and Inference**, 139(11), 3921–3927.

Rüschendorf, L (2013), Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios, Springer.

Salmon, F (2009), Recipe for Disaster: The Formula That Killed Wall Street, `http://www.wired.com/techbiz/it/magazine/17-03/wp_quant?currentPage=all` (2013-02-16).

Salmon, F (2012), The formula that killed Wall Street, **Significance**, 9(1), 16–20.

Sandström, A (2006), Solvency: Models, Assessment and Regulation, Boca Raton: Chapman & Hall/CRC.

Schönbucher, PJ and Schubert, D (2001), Copula-Dependent Default Risk in Intensity Models, `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=301968` (2009-12-30).

Schweizer, B (1991), Thirty years of copulas, **Advances in Probability Distributions with Given Marginals**, ed. by G Dall'Aglio, S Kotz, and G Salinetti, Kluwer Academic Publishers, 13–50.

Schweizer, B and Sklar, A (1983), Probabilistic Metric Spaces, New York: North-Holland.

SCOR (2008), From Principle-Based Risk Management to Solvency Requirements, `http://www.scor.com/images/stories/pdf/scorpapers/sstbook_second_edition_final.pdf` (2013-02-16).

Sibuya, M (1959), Bivariate extreme statistics, I, **Annals of the Institute of Statistical Mathematics**, 11(2), 195–210.

Sklar, A (1959), Fonctions de répartition à n dimensions et leurs marges, **Publications de L'Institut de Statistique de L'Université de Paris**, 8, 229–231.

Sklar, A (1996), Random variables, distribution functions, and copulas – a personal look backward and forward, **Distributions with Fixed Marginals and Related Topics**, 28, 1–14.

Tarullo, DK (2008), Banking on Basel: The Future of International Financial Regulation, Washington, D. C.: Peterson Institute for International Economics.

Torrent-Gironella, G and Fortiana, J (2011), Simulation of High-Dimensional t-Student Copulas with a Given Block Correlation Matrix, ASTIN Colloquium Madrid 2011.

Uryasev, S and Rockafellar, RT (2013), The fundamental risk quadrangle in risk management, optimization and statistical estimation, *Surveys in Operations Research and Management Science*. 18.

Viens, FG, Mariani, MC, and Florescu, I (2012), Handbook of Modeling High-Frequency Data in Finance, Wiley.

Wiener, N (1923), On the nature of mathematical thinking, **Australasian Journal of Psychology and Philosophy**, 1(4), 268–272.

Wong, MCY (2013), Buble Value at Risk: A Countercyclical Risk Management Approach, revised edition, Wiley.

Ziegel, J (2013), Coherence and elicitability, preprint, University of Bern.