Risk Aggregation and Diversification

C. Bernard S. Vanduffel

October 29, 2015

Overview

This report reviews the academic literature on risk aggregation and diversification as well as the regulatory approaches. We will point out the advantages and disadvantages of the different approaches with a focus on model risk issues.

We first discuss, in Section 1, the basic fundamentals of measuring aggregated risk. Specifically, we review the concept of a risk measure as a suitable way to measure the aggregate risk. We discuss desirable properties of risk measures and illustrate our discussion with the study of Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR).

Section 2 explores the question of diversification benefits associated with risk aggregation and the potential limitations of correlations as the only statistic to measure dependence. We go beyond correlations and explain that a full multivariate model is needed to obtain a correct description of the aggregate risk position.

We then explore the regulators approach to risk aggregation and diversification in Section 3, and provide some observations on the implicit assumption made by international regulators and different approaches that can be taken.

We end our review by highlighting that model risk becomes a key issue in measuring risk aggregation and diversification. We explore in Section 4 a framework that allows practical quantification of model risk and which has been recently developed in Bernard and Vanduffel [2015a][1](building further on ideas of Embrechts et al. [2013]). Details are provided in Appendices A and B.

Introduction

The risk assessment of high-dimensional portfolios $(X_1, X_2, ..., X_d)$ is a core issue in risk management of financial institutions. In particular, this problem appears naturally for an insurance company. An insurer is typically exposed to different risk factors (e.g., non-life risk, longevity risk, credit risk, market risk, operational risk), has different business lines or has an exposure to several portfolios of clients. In this regard, one typically attempts to measure the risk of a random sum, $S = \sum_{i=1}^{d} X_i$, in which the individual risks X_i depict losses (claims of the different customers, changes in the different market risk factors,...) using a risk measure such as the variance, the Value-at-Risk (VaR) or the

[1]This paper received the 2014 PRMIA Award for New Frontiers in Risk Management.

Tail-Value-at-Risk[2] (TVaR). It is clear that solving this problem is mainly a numerical issue once the joint distribution of $(X_1, X_2, ..., X_d)$ is completely specified. Unfortunately, estimating a multivariate distribution is a difficult task. In many cases, the actuary will be able to use mathematical and statistical techniques to describe the marginal risks X_i fruitfully but the dependence among the risks is not specified, or only partially specified. In other words, the assessment of portfolio risk is prone to model misspecification (model risk).

From a mathematical point of view, it is then often convenient to assume that the random variables X_i are mutually independent, because powerful and accurate computation methods such as Panjer's recursion and the technique of convolution can then be applied. In this case, one can also take advantage of the Central Limit Theorem, which states that the sum of risks, S, is approximately normally distributed if the number of risks is sufficiently high. In fact, the mere existence of insurance is based on the assumption of mutual independence among the insured risks, and sometimes this complies, approximately, with reality. In the majority of cases, however, the different risks will be interrelated to a certain extent. For example, a sum S of dependent risks occurs when considering the aggregate claims amount of a non-life insurance portfolio because the insured risks are subject to some common factors such as geography, climate or economic environment. The cumulative distribution function of S can no longer be easily specified.

Standard approaches to estimating a multivariate distribution among dependent risks consist in using a multivariate Gaussian distribution or a multivariate Student t distribution, but there is ample evidence that these models are not always adequate. More precisely, while the multivariate Gaussian distribution can be suitable as a fit to a dataset "on the whole", it is usually a poor choice if one wants to use it to obtain accurate estimates of the probability of simultaneous extreme ("tail") events, or, equivalently, if one wants to estimate the VaR of the aggregate portfolio $S = \sum_{i=1}^{d} X_i$ at a given high confidence interval; see McNeil et al. [2010]. The use of the multivariate Gaussian model is also based on the (wrong) intuition that correlations[3] are enough to model dependence (Embrechts et al. [1999], Embrechts et al. [2002]). This fallacy also underpins the variance-covariance standard approach that is used for capital aggregation in Basel III and Solvency II, and which also appears in many risk management frameworks in the industry. Furthermore, in practice, there are not enough observations that can be considered as tail events. In fact, there is always a level beyond which there is no observation. Therefore if one makes a choice for modeling tail dependence, it has to be somewhat arbitrary, at least not based on observed data.

There is recent literature on the development of flexible multivariate models that allow a much better fit to the data using for example pair copula constructions and vines (see e.g. Aas et al. [2009] or Czado [2010] for an overview). While these models have theoretical and intuitive appeal, their successful use in practice requires a dataset that is sufficiently rich. However, no model is perfect, and while such developments are clearly needed for an accurate assessment of portfolio risk, they are only useful to regulators and risk managers if they are able to significantly reduce the model risk that is inherent in

[2]In the literature it is also called the Expected shortfall, the Conditional Value at Risk and the Tail Value-at-Risk, among others.

[3]It should be clear that using correlations is not enough to model dependence, as a single number (i.e., the correlation) cannot be sufficient to describe the interaction between variables unless additional assumptions are made (e.g., a Gaussian dependence structure).

risk assessments.

In this review we provide a framework that allows practical quantification of model risk and which has been recently developed in Bernard and Vanduffel [2015a] (building further on ideas of Embrechts et al. [2013] and references herein). Technically, consider N observed vectors $\{(x_{1i}, ..., x_{di})\}_{i=1,...,N}$ and assume that a multivariate model has been fitted to this dataset. However, one does not want to trust the fitted multivariate model in areas of the support that do not contain enough data points (e.g., tail areas). The idea is thus to split $\mathbb{R}^d$ into two subsets, the first subset $\mathcal{F}$ is referred to as the "fixed part" and the second subset $\mathcal{U}$ is the "unfixed part", which will incorporate all the areas for the fitted model is not giving an appropriate fit. This incorporates the two directions discussed above for risk aggregation. Precisely, if one has a perfect trust in the model, then all observations are in the "fixed" part ($\mathcal{U} = \emptyset$) and there is no model risk. If one has no trust at all in the fit of the dependence, then $\mathcal{F} = \emptyset$ and we are in the setting of Embrechts et al. [2013] who derives risk bounds for portfolios when the marginal distributions of the risky components are known but no dependence information is available. The approach of Bernard and Vanduffel [2015a] makes it possible to consider dependence information in a natural way and may lead to more narrow risk bounds. This framework is also supplemented with an algorithm allowing actuaries to deal with model risk in a very practical way, as we will show in full details.

1 Measuring Aggregate Risk

Insurance companies essentially exchange premiums against (future) random claims. Consider a portfolio containing d policies and let X_i $(i = 1, 2, ..., d)$ denote the loss, defined as the random claim net of the premium, of the $i-$th policy. In order to protect policyholders and other debtholders against insolvency, the regulator will require the portfolio loss $S = X_1 + X_2 + ... + X_d$ to be "low enough" as compared to the available resources, say a capital requirement K, which means that the available capital K has to be such that $S - K$ is a "safe bet" for the debtholders. i.e., one is "reasonably sure" that the event '$S > K$' is of minor importance (Tsanakas and Desli [2005], Dhaene et al. [2012]).

It is clear that measuring the riskiness of $S = X_1 + X_2 + ... + X_d$ is of key importance for setting capital requirements. However, there are several other reasons for studying the properties of the aggregate loss S. Indeed, an important task of an Enterprise Risk Management (ERM) framework concerns capital (risk) allocation, i.e., the allocation of total capital held by the insurer across its various constituents (subgroups) such as business lines, risk types, geographical areas, among others. Indeed, doing so makes it possible to redistribute the cost of holding capital across the various constituents so that it can be transferred back to the depositors or policyholders in the form of charges (premiums). Risk allocation makes it also possible to assess the performance of the different business lines by determining the return on allocated capital for each line. Finally, the exercise of risk aggregation and allocation may help to identify areas of risk consumption within a given organization and thus to support the decision making concerning business expansions, reductions or even eliminations; see Panjer [2001], Tsanakas [2009].

When measuring the aggregate risk S, it is also important to consider the context at hand. In particular, different stakeholders may have different perceptions of riskiness. For

example, depositors and policyholders mainly care only about the probability that the company will meet its obligations. Regulators primarily share the interests of depositors and policyholders and establish rules to determine the required capital to be held by the company. However, they also care about the magnitude of the loss given that it exceeds the capital held, as this amount that needs to be funded by society when a bail out is needed. Formally, they care about the *shortfall* of the portfolio loss S with solvency capital requirement $\varrho(S)$; that is,

$$(S - \varrho(S))_+ := \max(0, S - \varrho[S]) \tag{1.1}$$

The shortfall is thus part of the total loss that cannot be covered by the insurer. It is also referred to as the *loss to society* or the *policyholders deficit*. In view of their limited liability, shareholders do not really have to care about the residual risk but rather focus on the properties of the variable $S - (S - \varrho(S))_+ . = \min(S, \varrho(S))$. In summary, various stakeholders may have different perceptions and sensitivities with respect to the meaning of the risk they run, and they may employ different paradigms to defining and measuring it.

As for measuring the risk, the two most influential risk measures are the *Value-at-Risk* (VaR) and the *Tail-Value-at-Risk (TVaR)*.[4] For a given probability level p, they are denoted by VaR_p and TVaR_p, respectively, and are defined as

$$\mathrm{VaR}_p(S) = \min\{x \mid P[S \leqslant s] \geqslant p\}, \qquad 0 < p < 1, \tag{1.2}$$

and

$$\mathrm{TVaR}_p(S) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q[S]\, dq, \qquad 0 < p < 1. \tag{1.3}$$

So, VaR_p is merely the minimum loss one observes with probability $1-p$ whereas TVaR_p is the average of all upper VaRs.

1.1 Coherent risk measures

The VaR and TVaR are merely two particular examples of risk measures. In fact, any functional ϱ mapping the random loss X (belonging to a relevant[5] set Γ of random losses) into a number $\varrho[X]$ can be used. However, it makes sense to impose certain properties (axioms) to the risk measure ϱ. Hereafter, we define a typical (and appealing) set of axioms. From a normative point of view, the "best set of axioms" is however nonexistent, as any normative axiomatic setting is based on a "belief" in its underpinning axioms. We obtain,

- *Positive homogeneity*: for any $X \in \Gamma$ and $a > 0$, $\varrho[aX] = a\varrho[X]$.
- *Translation invariance*: for any $X \in \Gamma$ and $b \in \mathbb{R}$, $\varrho[X + b] = \varrho[X] + b$.
- *Monotonicity*: for any $X, Y \in \Gamma$, $X \leqslant Y$ implies that $\varrho[X] \leqslant \varrho[Y]$.

[4]Between these two, the Value-at-Risk is currently by far the most popular risk measure in practice, among both regulators and risk managers; see, for example, Jorion [2006].

[5]In particular, the set Γ contains the random losses X_i $(i = 1, 2, ..., d)$ and we assume that $X_i, X_j \in \Gamma$ implies that $X_i + X_j \in \Gamma$, and also $aX_i \in \Gamma$ for any $a > 0$ and $X_i + b \in \Gamma$ for any real b.

- *Subadditivity*: for any $X, Y \in \Gamma$, $\varrho[X+Y] \leqslant \varrho[X] + \varrho[Y]$.

In Artzner et al. [1999], a risk measure that satisfies the aforementioned four properties of monotonicity, positive homogeneity, translation invariance and (most noticeably) subadditivity is called a *coherent* risk measure. As is well-known, the Value-at-Risk does not satisfy the subadditivity property whereas for any p the Tail-Value-at-Risk does. In fact, TVaR can be readily seen as the smallest coherent risk measure that is more conservative than VaR (which is not coherent) (for a proof, see Artzner et al. [1999] and also Dhaene et al. [2006]).

While the first three properties do not present much controversy, the desirability of the subadditivity property of a risk measure has been a major topic for research and discussion (see also Section 2.1). In the next subsection we explain that subadditivity is typically a natural constraint indeed. In this regard, we stress that the terminology "coherent" can be somewhat misleading as it may suggest that any risk measure that is not "coherent" is inadequate. Note that the well-known standard deviation principle, defined as $\varrho(X) = \mathbb{E}(X) + k\sqrt{\text{var}(X)}$ for some constant k, does not satisfy the monotonicity axiom and is thus not coherent[6]. In what follows, we assume in line with the academic literature and current practice that $\varrho(X)$ only depends on the distribution of X (i.e., $\varrho(X)$ is a functional of the distribution of X and is called a law-invariant risk measure).

1.2 Backtesting and robustness of risk measures

Backtesting: Ultimately, a model is used to assess the riskiness of S and to obtain a risk number $\varrho(S)$. In many cases, it is possible to build several competing models that are all consistent with respect to the available (incomplete) information and merely differ with respect to the ad-hoc assumptions that are made.

A natural way to compare the competing models is to use an error measure that involves the point forecasts and the realizing observations. More precisely, the performance of a particular model can be summarized by means of the average $\overline{T}$ of the scoring function T over n forecast cases, i.e.,

$$\overline{T} = \sum_{i=1}^{n} T(x_i, y_i), \tag{1.4}$$

where the $i-$th forecast case corresponds to the couple (x_i, y_i) in which x_i is the point forecast and y_i is the observation $(i = 1, 2, ..., n)$. Typical examples of scoring functions are the squared error $T(x, y) = (x - y)^2$ and the absolute error $T(x, y) = |x - y|$.

Gneiting [2011] shows that the scoring function T used should be adapted to the risk measure at hand, otherwise misguided inferences can be obtained. This author argues that one should evaluate the quality of the model (used to predict the functional $\varrho(S)$) by using a scoring function that would issue this functional as an optimal point forecast. If a scoring function is given, the optimal forecast (assuming that observations are identically

[6] A distortion risk measure, defined as $\varrho(X) = \int_0^1 F^{-1}(t)g'(1-t)\mathrm{d}t$ for an increasing function g with $g(0) = 0$, $g(1) = 1$ is coherent if g is concave on $[0, 1]$. We refer to Wang [2000], Bäuerle and Müller [2006] and Föllmer and Schied [2010] and the references therein for studies of risk measures and their properties.

and independently distributed), by applying Bayes rule, follows from

$$x^* = \arg\min_x E(T(x, S)), \tag{1.5}$$

For example, if the scoring function is the squared error, the optimal forecast is known to be the mean of S, while if the scoring function is the absolute error, the solution is given by its median. If this match between risk measure (functional) and scoring function exists, then the risk measure is "elicitable". For example, the mean and the median are elicitable. Also the VaR is elicitable, as using a (generalized) piecewise linear scoring function is consistent with VaR estimates. However, not every risk measure is elicitable: the standard deviation is not and, most notably, also the TVaR is not elicitable. See also Ziegel [2014] and Embrechts and Hofert [2014].

Risk measures that are not elicitable make it possible that there will be inconsistencies when comparing point forecasts from different models and/or from different forecasters. Suppose you have a model which is known to provide the best 99.5%-VaR estimate of the portfolio loss. However, there is also another model available that is known to give a suboptimal 99.5%-VaR estimate. Then, if you use the square error scoring function (which is not consistent with 99.5%-VaR) to evaluate the 99.5%-VaR estimates you might end up picking the suboptimal models, simply because you are using the wrong metric to assess the 99.5%-VaR estimates.

A few comments are in order: TVaR is not elicitable but it is indirectly elicitable as it can be decomposed into a conditional mean and a quantile, which are both separately elicitable. Furthermore, backtesting requires a rich data sample of predictions and observations, which is not readily available in the context of solvency assessments in which the horizon used is typically one year. Furthermore, the consistency argument used to link a risk measure to an optimal scoring function builds on the assumption that all observations are identically and independently distributed, which is not always the standard situation encountered in risk practice.

Robustness: Another important topic concerns robustness of the risk measure with respect to model misspecification and small changes in the data. From a regulator's viewpoint, the risk measure used should really be stable with respect to varying model assumptions and small changes in data sets. In the context of solvency II, two insurers holding the same portfolio should obtain the same VaR for this portfolio. However, when the correct model cannot be identified with (almost) certainty, the insurers may use two different models and obtain significantly different VaR results. For example, Chernih et al. [2010], show that it is possible to build a credit risk portfolio model that is consistent with the standard[7] MKMV credit risk model one with the exception of MKMV using a Gaussian dependence among asset returns whereas Chernih et al. [2010] employ a different copula (which, however, yields the same correlations as in MKMV). Hence, both models are perfectly consistent with the available information on exposure, loss-given default, default probabilities and correlations, but when used to estimate the 99,5% VaR of a typical loan portfolio their results can differ with a factor as high as 15; see also

[7]The MKMV model is used by many financial institutions for assessing the riskiness of credit risk portfolios. Furthermore, the Basel III standard framework relies on it to determine the required capital that banks need to hold for the credit risk they run; see Basel Committee [2010b]. Also the Solvency II framework uses this formula to decide on the amount of capital that insurers need to hold as a buffer against the adverse consequences if one or more of their reinsurance or derivative counterparts fail.

Heyde and Kou [2004], Kou et al. [2013], Bernard et al. [2013b] and Bernard et al. [2015] for more evidence and other examples. In the light of these observations Bernard et al. [2013b] warn for the use of VaR at high confidence levels (e.g., 99.5%) as a basis for capital requirements. Note also that if the external risk measure is not robust, institutions may pursue regulatory arbitrage by choosing a model that significantly reduces the capital requirements or by manipulating the input data.

2 Aggregation and Diversification

2.1 Diversification Benefits and Subadditivity

From the Canadian regulator's website (OSFI [2014]), one can read *"we define risk aggregation as the approach used to calculate the total of each and all of the risk elements. A diversification credit results when the method of aggregation of risks produces results that are less than the sum of the total of the individual risk elements."* Diversification benefits may come from pooling risks within one type of risk such as insurance risk, from pooling several types of risks (e.g., insurance risk and asset risk), across entities or across geographies. There is a careful warning that it is hard to determine the diversification benefits in period of stress. Capital requirements are determined to cover stress periods and it is especially in these stressed periods that some potential diversification benefits disappear. Reduction of capital should be granted for diversification benefits only in the case when even during stress periods, the diversification benefit stay valid. Some benefits should however be recognized. See OSFI [2014] for discussion of diversification benefits between volatility risk and respectively mortality risk, morbidity risk, longevity risk and lapse risk.

Let us consider portfolios with respective losses S_1 and S_2 and let ϱ be a risk measure used for setting capital requirements; i.e., $\varrho(S_1)$ is the capital for the first portfolio, $\varrho\left(S_2\right)$ is the capital for the second portfolio and $\varrho(S_1+S_2)$ is the capital of the combined (merged) position. Note that we assume that the losses S_1 and S_2 do not change of nature when merging the portfolios. In reality, however, merging or splitting portfolios may change management, business strategy and cost structure, among others, and may thus change the marginal distribution of the losses under consideration.

A standard definition for the diversification benefit, denoted by $DB(\varrho, S_1, S_2)$, is that

$$DB(\varrho, S_1, S_2) = \varrho\left(S_1\right) + \varrho\left(S_2\right) - \varrho\left(S_1 + S_2\right). \tag{2.6}$$

Hence, $DB(\varrho, S_1, S_2)$ provides the gain (loss) one obtains by merging two portfolios. It is clear that if ϱ is coherent (and thus subadditive) then the diversification benefit is non-negative, which corresponds to the common intuition that merging risks creates benefits. To confirm this intuition, let us observe that

$$\left(S_1 + S_2 - \varrho\left(S_1\right) - \varrho\left(S_2\right)\right)^+ \leqslant \sum_{j=1}^{2} \left(S_j - \varrho\left(S_j\right)\right)^+. \tag{2.7}$$

Inequality (2.7) states that the shortfall risk of the merged portfolio is always smaller than the sum of the shortfall risks of the stand-alone portfolios, when the solvency capital requirement is additive. It expresses that, from the viewpoint of the regulator, a merger

is beneficial in the sense that shortfall risk decreases when the capitals are summed up. The underlying reason is clear: within the merged portfolio, the shortfall of one of the entities can be compensated by the gain of the other. In summary, "*a merger decreases the shortfall*". Hence, the inequality (2.7) indicates that the solvency capital of the merged position can be smaller than the sum of the solvency capitals of the two stand-alone portfolios. These observations provide support for the common belief that a solvency capital requirement (risk measure) should be subadditive. Indeed, when merging two stand-alone portfolios, subadditivity is allowed by the regulator as long as

$$(S_1 + S_2 - \varrho(S_1 + S_2))^+ \leqslant \sum_{j=1}^{2} (S_j - \varrho(S_j))^+$$

holds. In this regard, let us notice that the requirement of subadditivity implies that

$$(S_1 + S_2 - \varrho(S_1 + S_2))^+ \geqslant (S_1 + S_2 - \varrho(S_1) - \varrho(S_2))^+, \tag{2.8}$$

and consequently, for some realizations (s_1, s_2) we may have that

$$(s_1 + s_2 - \varrho(S_1 + S_2))^+ > (s_1 - \varrho(S_1))^+ + (s_2 - \varrho[S_2])^+.$$

Hence, the use of a subadditive risk measure may give rise to a larger shortfall than the sum of the shortfalls of the stand-alone entities, i.e.,

$$(s_1 + s_2 - \varrho(S_1 + S_2))^+ > (s_1 - \varrho(S_1))^+ + (s_2 - \varrho(S_2))^+$$

may hold (Dhaene et al. [2008]). Therefore, while subadditivity is an acceptable property from the viewpoint of regulators they should restrict the degree of subadditivity in order to avoid that $(S_1 + S_2 - \varrho(S_1 + S_2))^+$ becomes too risky as compared to $(S_1 - \varrho(S_1))^+ + (S_2 - \varrho(S_2))^+$.

In this regard, it is also important to note that it is not clear cut that merging is advantageous for the shareholders. We explain this as follows. For portfolio j $(j = 1, 2)$ the end-of-the-year available funds are given by $(\varrho(S_j) - S_j)^+$. Indeed, if the loss S_j is smaller than the capital $\varrho(S_j)$, then the funds that belong to the shareholders (at the end of the reference period) will be given by $\varrho(S_j) - S_j$ whereas in the case that the loss S_j exceeds $\varrho(S_j)$, the business unit related to this portfolio gets ruined and the available funds become equal to zero. Since

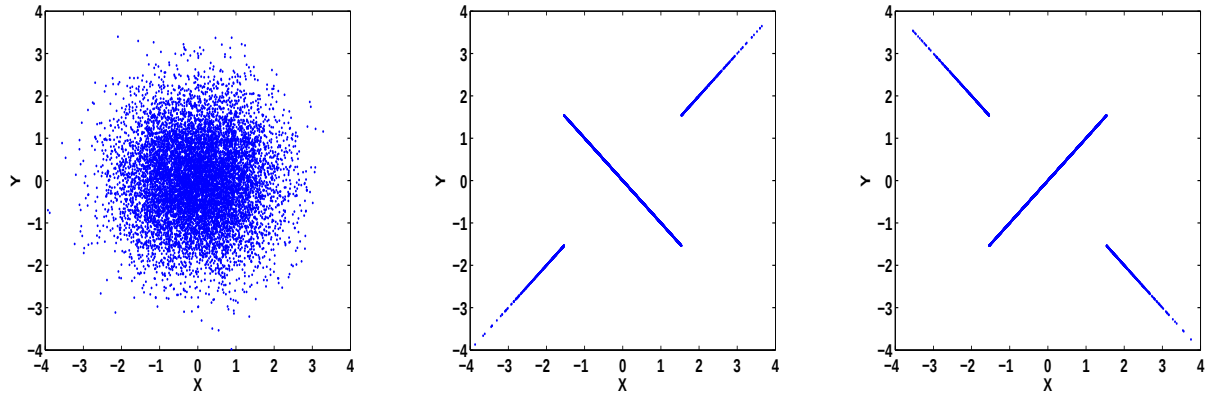
$$(\varrho(S_1) + \varrho(S_2) - S_1 - S_2)^+ \leqslant \sum_{j=1}^{2} (\varrho(S_j) - S_j)^+, \tag{2.9}$$

we observe that keeping the two portfolios separated might be preferred from the shareholders' point of view, essentially because in this case fire walls are built in, ensuring that the poor performance of one portfolio will not contaminate the other one. In fact, the shareholders and regulators have interests that are not fully aligned; see also Dhaene et al. [2008] and Dhaene et al. [2009] for more discussion.

2.2 The Fallacy of Using Correlations Only

Some practitioners appear to believe that for aggregating two risks one only needs to know their correlation coefficient. This (wrong) intuition is likely due to the widespread use and

importance[8] of the multivariate normal distribution that is fully characterized upon specification of the means, standard deviations and pairwise correlations (Embrechts et al. [1999, 2002]). However, one should be aware of the fact that the multivariate normal distribution inherits a choice of a specific (Gaussian) dependence already and that correlations are merely needed to parameterize this Gaussian dependence. Effectively, it is easy to construct two normal random variables that have a specific correlation coefficient but that are *not* jointly (bivariate) normal. To illustrate this feature, let X and Y be standard normally distributed random variables that are independent. In particular they have a Gaussian dependence with zero correlation. Next, we consider Z_c defined as $Z = -X$ if $|X| < c$ and $Z = X$ if $|X| > c$ ($c > 0$). It is easy to see that Z is also standard normally distributed: it has perfect positive correlation with X in the tails and perfect negative correlation otherwise. One can then choose c^* such that correlation between X and Z_{c^*} is zero ($c^* \approx 1.538$). Hence, when $p > \Phi(c^*)$ ($\Phi(\cdot)$ denotes the c.d.f. of the standard normal random variable), $\mathrm{VaR}_p(X + Z_{c^*}) = 2\Phi^{-1}(p)$ whereas $\mathrm{VaR}_p(X + Y) = \sqrt{2}\Phi^{-1}(p)$. A numerical illustration can be found in Figure 2.1.



Panel A Panel B Panel C

Figure 2.1: Illustration of three situations where the random variables are standard normally distributed and have zero correlation.

Another example illustrating the deficiencies of correlations concerns risk measurement of a portfolio of credit loans. To explain this idea, let us consider risks X_i ($i = 1, 2, ..., n$) indicating default events so that $S = X_i + X_2 + ... + X_n$ reflects the number of defaults of the portfolio. Specifically, Let p_i denote the probability that the i-th company defaults and denote by p_{ij} the pairwise default probability that both company i and company j default. The pairwise default correlation ρ_{ij}^D ($i, j = 1, 2, \ldots, n$) is then given as

$$\rho_{ij}^D = \frac{p_{ij} - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}. \tag{2.10}$$

In other words, correlations only reveal full information on interaction between two default event (pairwise), but not really on the manner three or more loans interact. In this regard, note that there is an intrinsic lack of sufficient default statistics (joint defaults are inherently very rare events) and one can simply not expect to be able to reliably estimate

[8]The multivariate distribution is at the center of many theories and applications such as linear regression, principal component analysis, CAPM, Markowitz mean-variance analysis, discriminant analysis, capital aggregation (e.g., Basel III and Solvency II), Credit Portfolio modeling (Moody's KMV model).

higher order joint default probabilities. In other words, assessing the risk of a credit risk portfolio is inherently subject to model uncertainty.[9] For example, the influential MKMV model links defaults of companies to the asset return behavior and assumes that asset returns are multivariate normally distributed. This assumption, however, is merely one possible choice and there are no reasons to believe this assumption is close to reality. Bernard, Rüschendorf and Vanduffel 2013b assess the impact of model uncertainty on VaR calculations. When using $p = 99.5\%$ as the basis for calculating VaR and capital requirements (as in Basel III and Solvency II), the results of industry models are typically within a wide range of possible values of VaR. By contrast, model risk appears more limited when using more moderate levels of probability to assessing the VaR. These authors conclude that it might be useful to impose additional constraints on models when used for setting capital requirements. For example, one may use the obtained VaR bounds to set a minimum value on the VaR that is obtained by the internal model, or, one may want to impose a particular model that different institutions need to use for computing capital requirements, as this provides some guarantee that capital levels can be readily compared across institutions also yielding fair competition.

2.3 The impact of micro correlations

In this section, we present another weakness of correlation. Kousky and Cooke [2012] explain how catastrophic risks are usually characterized by fat tails and dependence. With fat-tailed loss distributions, the probability of an event declines only slowly, relative to its severity, meaning that very large losses are not so exceptional (for more mathematical explanations we refer to Kousky and Cooke [2012]). Many natural catastrophes have been shown to be fat tailed. As explained in full details in Kousky and Cooke [2012], catastrophes can introduce another type of dependence, which is called tail dependence. Tail dependence refers to the probability that one variable exceeds a certain percentile, given that another has also exceeded that percentile. More simply, it means bad things are more likely to happen together. It is clear that a catastrophe will potentially hit simultaneously multiple lines of business for an insurer (houses, cars, health, businesses...). Lescourret and Robert [2006] have observed such tail dependence for lines of insurance covering over 700 storm events in France. Moreover, catastrophic risks tend to be spatially correlated because of the high dependence among the claims due to a given disaster. In practice, this correlation declines with the spatial distance between policies. When it declines to zero, it allows insurers to diversify by holding policies in different regions. Unfortunately, Kousky and Cooke [2012] find that "close to zero" does not count as zero for diversification benefits. Even small, positive, average correlations among policies, which they term "micro-correlations", can cause problems in risk aggregation.

The main issue with microcorrelation comes from the fact that the law of large numbers fails when risks are not independent even if they display a correlation coefficient that is very close to zero. This is well explained in the works of Kousky and Cooke [2009], Cooke and Kousky [2010], and Cooke et al. [2011] applied on catastrophic risks in Kousky and Cooke [2012]. The basic idea is very simple and is based on the situations in which policies have a small, average, positive correlation (say 0.04, which is the average correlation found in flood insurance claims in the U.S. at a county level in Cooke and Kousky [2010]). Cooke and Kousky show how quickly tiny, positive correla-

[9] Duffie and Singleton [2012].

tions between policies can become pernicious.

Let $X_1, ..., X_n$ and $Y_1, ..., Y_n$ be two sets of random variables with the same average variance σ^2 and average covariance C (within and between sets). The correlation of the sums of the X's and the sum of the Y's is easily found to be:

$$corr\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i\right) = \frac{n^2 C}{n\sigma^2 + n(n-1)C} \tag{2.11}$$

The main issue is that it goes to 1 as n grows, if C is non-zero (even very small) and σ^2 is finite. If all variables are independent, then $C = 0$, and the correlation in (2.11) is zero. To highlight this amplification of correlation, Kousky and Cooke [2009] use flood insurance claim data. They randomly draw pairs of US counties and compute their correlation. The green histogram in Figure 2.2 shows 500 such correlations. The average correlation is 0.04. Although a few counties have high and positive correlations, most of the correlations are very small and around zero. Instead of looking at the correlations between two randomly chosen counties, they then sum 100 randomly chosen counties and correlating this with the sum of another distinct set of 100 randomly chosen counties. After repeating this 500 times, they obtain the blue histogram where the average of 500 such correlations (of 100) is 0.23. The red histogram depicts 500 correlations (of 500) with an average value is 0.71. This dramatic increase in correlation is a result of the micro-correlations between the individual variables.

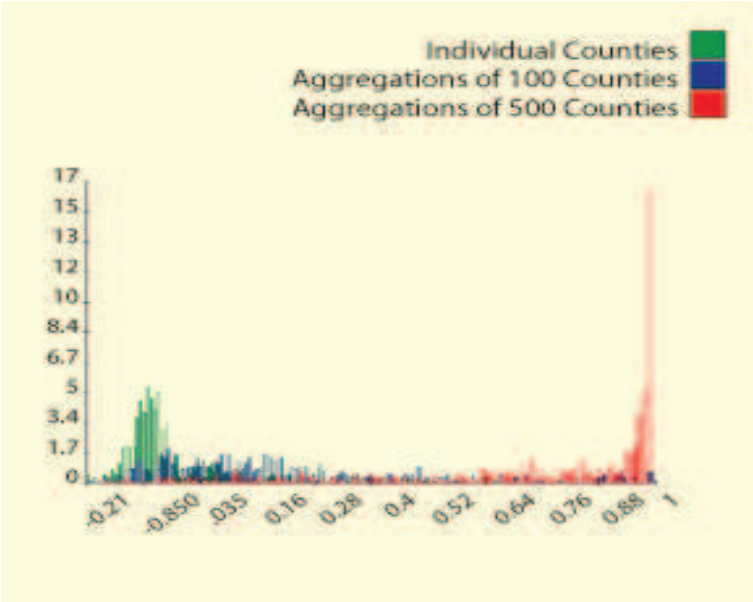


Figure 2.2: Figure 11 from Kousky and Cooke [2009] is reproduced here as an illustration.

2.4 Fitting a Multivariate Distribution

In practice, there exist efficient and accurate statistical techniques to estimate the respective marginal distributions of $\mathbf{X} = (X_1, \cdots, X_d)$. On the other hand, the joint dependence structure of $\mathbf{X}$ is often much more difficult to capture: there are computational and convergence issues with statistical inference of multi-dimensional data, and the choice of

multivariate distributions is quite limited compared to the modelling of marginal distributions. However, an inappropriate dependence assumption can have important risk management consequences. For example, (mis)using the Gaussian multivariate copula, can result in severely underestimating probability of simultaneous default in a large basket of firms (McNeil et al. [2010]).

The easiest (and therefore popular) modeling of a multivariate distribution is to use a multivariate Gaussian or multivariate Student distribution. The advantage of the multivariate Student distribution is that it displays some tail dependence. However, there are limitations of this multivariate dependence as there is a single degree of freedom parameter which drives the tail dependence of all pairs of variables.

More generally, multivariate distribution can be decomposed in the marginal distributions F_{X_i}, $i = 1, 2, ..., d$ (reflecting the stand-alone risks) and a so-called copula function C (reflecting the dependence). More precisely, Sklar [1959]'s theorem states that there exists a vector $(U_1, U_2, ..., U_d)$ of standard uniformly distributed random variables such that

$$\mathbf{X} \overset{d}{=} (F_{X_1}^{-1}(U_1), F_{X_2}^{-1}(U_2), ..., F_{X_n}^{-1}(U_d)). \tag{2.12}$$

where "$\overset{d}{=}$" reflects equality in distribution. The representation (2.12) thus shows that the distributional properties of the portfolio $\mathbf{X}$ are indeed completely specified by the marginal distributions F_{X_i} $(i = 1, 2, ..., d)$ of its risky components and the joint distribution C of $(U_1, U_2, ..., U_d)$ describing the interaction among the risks of the portfolio.

Copulas have been extensively studied by Joe [1997] and Nelsen [2007]. There are large families of two-dimensional copulas so that modeling dependence between two variables is relatively easy. The most popular two-dimensional copulas are the Archimedian ones for which an important literature exists on estimation and goodness of fit; see Joe [1997]. Bedford and Cooke [2001, 2002] have then proposed to construct a multivariate copula using pair copulas as building blocks. They also give graphical representations involving a sequence of nested trees, which they called regular vines. This multivariate model, also called pair-copula construction, allows to decompose a complex multivariate model into simpler two-dimensional building blocks. An overview is given by Czado [2010]. This approach is very flexible and allows the dependence between any subset of two variables to be different. For some estimation techniques of parameters of regular vines, one can refer to Kurowicka and Cooke [2006]. An alternative to pair copula constructions is proposed in Hofert [2012] using hierarchical model; ssee Okhrin et al. [2013] for estimation issues. The nested Archimedian copulas are studied by Hofert and Pham [2013] and used by Savu and Trede [2010]. A comprehensive overview of dependence in high dimensions can be found in Embrechts and Hofert [2013].

2.5 Summary

Taking into account the dependence among risky components is crucial to assess the aggregate risk of a portfolio. We show that subadditivity of a risk measure is justified from a regulator's viewpoint. In other words, it is justified that companies receive some diversification benefits when aggregating risks. However, some care is needed: Diversification benefits are often assessed using correlations, but correlation is a poor measure of dependence. It is merely a single number and not sufficient to describe the complex interaction among risky components. We end Section 2 by discussing how to fully describe

dependence fully.

3 Overview of current Regulation

The report of Basel Committee [2010a] describes the modeling methods used by financial firms and regulators in various countries to aggregate risk. It also aims at identifying the conditions under which these aggregation techniques perform as anticipated in the model and suggests potential improvements. The report expresses doubts about the reliability of internal risk aggregation results that incorporate diversification benefits: "Model results should be reviewed carefully and treated with caution, to determine whether claimed diversification benefits are reliable and robust." In this section, we very briefly summarize their findings as well as those of other regulators.

3.1 Regulatory frameworks

Basel III regulation for banks One calculates a bank's overall minimum capital requirement as the sum of capital requirements for the credit risk, operational risk, and market risk, without recognizing any diversification benefits between the three risk types. The idea that no diversification corresponds to the worst-case situation of the portfolio is not entirely correct. Technically, such property is verified when a coherent risk measure is used but may be violated for other risk measures such as VaR. In other words, it may be possible to aggregate risks so that the VaR of the aggregated risk is higher than the sum of the VaRs.

Within the market risk, banks have the choice between two methods. They may benefit from diversification if they use an internal model approach (IMA). With the standardized measurement method (SMM), the minimum capital requirement for market risk is the sum of the capital charges calculated for each individual risk type (interest rate risk, equity risk, foreign exchange risk, commodities risk and price risk in options).

Canadian Minimum Capital Test (MCT) and Minimum Continuing Capital and Surplus Requirements (MCCSR) Capital requirements of property and casualty insurers in Canada are based on the Minimum Capital Test (MCT). The MCT is a factor-based requirement that aggregates risks as a sum with an explicit credit for diversification between insurance risk and the sum of credit and market risk, so that the total capital required for these risks is lower than the sum of the individual requirements for these risks.

On the other hand, capital requirements of life insurance companies in Canada are computed according to OSFI's MCCSR. The MCCSR employs more sophisticated approaches in some areas. "MCCSR imposes capital requirements for the following risk components: asset default risk, mortality risk, morbidity risk, lapse risk, disintermediation risk, and segregated fund guarantee risk" (Basel Committee [2010a]). Some diversification benefits can be incorporated in the computation of mortality risk, morbidity risk and segregated funds risk but the total MCCSR is calculated as the sum of each risk without potential reduction due to diversification. Again it is (implicitly) assumed here that this is the worst possible situation. More information on the MCT and MCCR can be found on the website of OSFI (`www.osfi-bsif.gc.ca`).

Solvency II The Solvency Capital Requirement (SCR) under Solvency II is defined as the Value-at-Risk (VaR) at 99.5% and a horizon of one year. When aggregating risks, insurers may benefit from diversification: they have the option to use an internal model (without any particular method prescribed) or a standard formula. The standard formula aggregates risks using a correlation matrix (Var-Covar approach) to take into account dependencies.

Swiss Framework for Insurance Companies Since 2008, all insurers in Switzerland must use the Swiss Solvency Test (SST). Similarly as in Solvency II, there is a standard model and the possibility to use an internal model. The standard model considers the following risks separately: market risk, credit risk (counterparty default), non-life insurance risk, life insurance risk, and health insurance risk. Operational risks do not make part of the current SST. Diversification between risk categories is recognized in all cases. Life insurance companies use the Var-Covar aggregation method whereas non-life insurers aggregate risks more carefully to find the distribution of the aggregate risk and then use an Expected Shortfall (or TVaR).

US Insurance Risk Based Capital (RBC) Solvency Framework We end our brief review of regulatory frameworks used across the world in the industry by the US risk-based capital (RBC). The RBC formula is a standardized system applied to all states in the US and allowing for an easy comparison across the companies. Each type of insurer has a separate RBC formula (life, property and casualty, and health). Diversification benefits are incorporated by computing a covariance matrix among the individual risks to reduce the overall capital so that it is smaller than the sums of individual risks.

In the calculation of RBC, the formula is a square root of sum of squares. This amounts to use a very simple assumption for aggregating risks by assuming that they are fully correlated (correlation equal to one) or independent (zero correlation) (OSFI [2014]).

3.2 Comparison and Comments on International Regulatory Frameworks

Generally, regulatory rules incorporate diversification by taking into account some correlation effect to reduce the total capital (at least in some subcategories). Overall, we observe that regulators all implicitly assume that the sum of the risk numbers is the worst possible situation. "No diversification benefits" is then synonym to "adding up risk numbers (VaRs)".

The easiest method to aggregate risks is the Var-Covar approach (which is explicitly mentioned in the Solvency II and SST above and also used by the Australian regulator (OSFI [2014])). It builds on the assumption that the correlation matrix is enough to describe the dependence and that it is possible to aggregate risks based on this correlation matrix. Its strength is to be a simple approach but it is merely only a correct approach for elliptical multivariate distributions such as the Gaussian multivariate distribution. Furthermore, correlation is a linear measure of dependence and does not capture tail dependence adequately. Using such a method to aggregate risk may perhaps be fine to have some idea on the distribution "globally", but fails when it comes to assess the risk

in the tail, and note that capital requirements are typically based on tail risk measures such as Value-at-Risk at 99.5%, which essentially reflects the outcome of a 1-in-200 year scenario.

Instead of using the Var-Covar approach, one may use copulas to aggregate the individual risks. This approach is rather flexible and allows to separate the risk assessment of the marginal distribution of individual risks and their dependence. By specifying a given copula to model some dependence, it is then possible to recognize tail dependence among some risks. However, determining the "right" copula to use is a very hard task that is prone to significant model risk, as we will see later in this report. Statistical methods to fit a multivariate model involve large numbers of parameters and copula families. In addition, understanding the outputs of the model will then require a good expertise of the copula approach in order to understand the impact of each assumption made on the dependence. This is a concern and a challenge among institutions.

Another way to capture tail risks and tail dependence is to understand "where the dependence comes from", and to model the real risk drivers of the dependence among individual risks of the portfolio and understand their interactions. The report of Basel Committee [2010a] suggest to use "Scenario-based aggregation.". Aggregation through scenarios boils-down to determining the state of the firm under specific events and summing profits and losses for the various positions under the specific event. In other words, it means that one needs to incorporate information that one knows about the dependence in some specific states.

We propose in Appendix B a method to assess model risk that is somewhat in this spirit, as it allows to incorporate existing information about the dependence structure among the risks in some states of the world. The scenario based approach has a clear advantage in that the multivariate model is then based on some clearly identified risk drivers (which can then be simulated for instance) and it forces the firm to understand the chosen multivariate model: it is not anymore a complex set of copulas but dependence among factors is obtained through reasonable factors. As observed in the report of Basel Committee [2010a], the results of scenario-based aggregation are easier to interpret with more meaningful economic and financial implications but it requires again a deep expertise to identify risk drivers, derive meaningful sets of scenarios with relevant statistical properties, and then use them to obtain a full loss distribution will still be a challenging task. A lot of the inputs in these kind of models comes from experts' judgments. Overall there is no clear unique solution to solve the problem of risk aggregation. Each method has its pros and cons and may be helpful in given situations and useless in others.

4 Model Risk on Dependence

As discussed extensively in the previous sections, one of the main issues in aggregating risks arises from the difficulty in modeling the dependence among a large number of risks, i.e. risk aggregation is prone to model risk. Specifically, we showed in Section 1 that there is no unique way to measure risk, and in Section 2 that correlation is not enough to measure dependence and that the full information on dependence contains much more information. However, as it appears in Section 3, regulators over the world discuss diversification benefits and propose guidelines in estimating them. But there is no consensus. It turns out that dependence modeling carries a lot of model risk.

In appendices, we provide specific examples that can be helpful in better understanding model risk related to aggregation. Appendix A discusses how to minimize or maximize a given risk measure $\varrho(\cdot)$ of the aggregate risk when the distributions of the risky components are known but not their interdependence (consistent with the approach of Embrechts et al. [2013]). This approach is useful to assess model risk on dependence, which is one of the most important factor in assessing aggregated risk.

However, the bounds on model risk on dependence obtained by the approach described Appendix A (see also Embrechts et al. [2013]) are typically too wide to be useful in practice. They ignore all information on dependence and consider only the information about the marginal distributions. There are a few papers studying model risk with partial information on the dependence structure. See among others, Cheung and Vanduffel [2013] for convex ordering bounds with given variance; Embrechts and Puccetti [2006] for bounds on the distribution of S when the copula of $\mathbf{X}$ is bounded by a given copula; Tankov [2011] for bounds on S when $n = 2$ and when there are constraints on the copula; Bernard et al. [2013b] when an upper bound on the variance of the aggregate risk is imposed, and Bernard et al. [2014a] when high-order moments are given.

In Appendix B, we present a framework which allows practical quantification of model risk (and was developed in Bernard and Vanduffel [2015a]). Importantly, unlike Appendix A we no longer ignore the available information on dependence. We assume that risk modelers have developed an "as good as possible" multivariate model for a certain portfolio. However, no model is perfect and the extent of misspecification of the proposed model affects the risk measurement and should be assessed. Our framework includes an algorithm allowing actuaries to deal with model risk in a very practical way.

These results make it possible to identify risk measures for which additional information of a well-fitted multivariate model reduces the model risk significantly, making them meaningful candidates for use by risk managers and regulators. Our approach may lead to bounds that are significantly tighter than the (unconstrained) ones available in the literature, accounting for the available information coming from a multivariate fitted model and allowing for a more realistic assessment of model risk. However, model risk remains a significant concern and we recommend caution regarding regulation based on Value-at-Risk at a very high confidence level since such an assessment is unable to benefit from careful risk management attempts to fit a multivariate model. For instance, we observe from numerical experiments that the portfolio VaR at a very high confidence level (as used in the current Basel regulation) might be prone to such a high level of model risk that, even if one knows the multivariate distribution nearly perfectly, its range of possible values remains wide. In fact, one may then not even be able to reduce the model risk as computed in Embrechts et al. [2013] (see also Appendix A) where no information on the dependence among the risks is used at all.

We remark that it could be of interest to consider also a "global" constraint to sharpen the bounds further. A natural global statistic on the distribution of the aggregate risk is the variance and it would be relatively easy to extend our study by using techniques similar to those employed in Bernard et al. [2013b] to account for a maximum possible variance of the aggregate portfolio.

Finally, we assume that the marginal distributions are fixed and known. To capture the possible uncertainty of the marginal distributions one might consider amplifying their tails. For example, a distortion (Wang transform) could be applied when re-discretizing (instead of using $\widehat{f}_i$).

Superadditivity of VaR: We end this section on an important discussion on consequences of aggregation. Specifically, we discuss the superadditivity of VaR. Comonotonicity is the worst-case dependence according to risk averse decision makers, but that it does not yield the maximum VaR of a portfolio (more details can be found in Appendix A). The worst case VaR does not readily occur when the risks are perfectly correlated. As VaR is additive for comonotonic risks, there exists thus a dependence such that

$$\mathrm{VaR}_p(X_1 + X_2 + ... + X_n) \geqslant \mathrm{VaR}_p(X_1) + \mathrm{VaR}_p(X_2) + ... + \mathrm{VaR}_p(X_n) \tag{4.13}$$

The non-existence of diversification benefits is a situation that is hard to accept by practitioners. In addition, the use of VaR can lead to inconsistent risk rankings since the highest possible value of the risk measure does not correspond to the scenario of full dependence. An important question is when the stated inequality (4.13) is strict, i.e., when does one have (strict) superadditivity and how significant is the superadditivity. It is not difficult to show that one can always find a dependence such that the stated inequality is strict *unless* $VaR_q(X_i)$ is constant for $q \geqslant p$ (see also Bernard, Rüschendorf and Vanduffel 2013b). This observation allows us to draw the following conclusions:

- When only the marginal distributions are known and the portfolio contains *unbounded* risks then the maximum possible VaR (by finding the worst possible dependence) can be significantly larger than the VaR obtained in the comonotonic case (in which the VaR is additive). For example, Embrechts et al. [2013] show in their Figure 5 that for a portfolio of Pareto(2) distributed risks the upper bound on the VaR is about two times larger than the comonotonic VaR (i.e. when the marginal risks are assumed to be comonotonic). See also Embrechts et al. [2014]. More generally, Puccetti and Rüschendorf [2012b] show that under some mild conditions the worst Value-at-Risk behaves asymptotically as the worst Tail Value-at-Risk (TVaR). The intuition behind this result is as follows. The VaR (measured at some probability level p) of a comonotonic sum is of course just a particular point on the quantile function of this sum. Now, by changing the comonotonic dependence in the *upper* tail of the marginal supports (from level p onwards), one is able to adjust the upper quantiles of the sum (from level p onwards). As the quantile function is non-decreasing, it is then clear that the highest VaR will be obtained if one can change the dependence such that the quantile function of the sum becomes a constant on $(p, 1)$. The constant value is then the maximum VaR and is equal to the comonotonic TVaR (Bernard et al. 2013b).

  Although, the fact that for a given p, some dependence structures yield a VaR larger than the comonotonic VaR, this may not happen in real-world situations.

- Insurance companies typically have limited liability, hence the VaR cannot be (strictly) superadditive for high levels of probability (which is the standard case for solvency assessments). In fact, in this case the VaR obtained by using a particular model is likely to be subadditive. This feature is important as violation of the subadditivity property is ground for refuting a risk measure, in particular VaR.

- The situation described above stresses that information on the dependence is crucial if one wants to build models that provide risk numbers that are trustworthy in the sense that upper and lower bounds for these numbers stay in some reasonable range. For example, it might be reasonable to assume that the risks are positively

dependent, or the variance of the aggregate risk can be estimated accurately from a statistical analysis of observed losses, or some information on the copula function might be available. In this regard, the results in the literature on ranges of VaR in the presence of additional dependence information are more limited and of an ad-hoc nature. Rüschendorf [1991], Embrechts and Puccetti [2010a], Embrechts et al. [2013] consider the situation in which some of the bivariate distributions are known, and Denuit et al. [1999] study VaR bounds assuming that the joint distribution of the risks is bounded by some distribution. However, the bounds that are proposed in these papers are often hard to deal with, especially for high-dimensional and inhomogeneous portfolios, and they do not necessarily sharpen the unconstrained bounds in a significant way; see also Chernih et al. [2010] for an illustration in the context of credit risk portfolio modeling. These observations, however, contrast with the findings of Bernard et al. [2013b]. They consider the presence of a variance constraint on the portfolio sum as a source of dependence information and show that doing so can significantly tighten the (unconstrained) VaR bounds.

We recall that the risk measure that is dominantly used in regulatory frameworks is VaR. For example, the current European regulation of financial institutions (Basel III) formally relies on the concept of risk weighted assets (RWA), but is essentially a VaR based framework. Hence, an approach based on risk-weighted assets may not be appropriate if one needs to aggregate risks to computing VaR of a portfolio. The majority of the academic literature has always been arguing against the use of VaR because it does not comply with subadditivity. Recently there has then been a trend in moving away from VaR and to use TVaR instead; see Embrechts et al. [2014], Basel Committee [2012] and Basel Committee [2013].

5 Conclusions

Recent turbulent events such as the subprime crisis, have increased the pressure on regulators and financial institutions to carefully reconsider risk models and to understand the extent to which the outcomes of risk assessments based on these models are robust with respect to changes in the underlying assumptions.

Consequently, we have observed a recent and important literature on risk aggregation and diversification benefits. New approaches for dealing with risk aggregation are to be expected and the issue of model risk that is inherent in risk aggregation will be the topic of significant study as well.

Section 4 briefly summarizes the latest developments on the assessment of model risk. In Appendix B we describe a practical method to assess model risk that takes into account a typical set of available information. This information may come from statistical modeling such as a multivariate model fitted on the data at hand (and trusted wherever there is enough data) but may also arise from scenarios or experts' opinions. Assume that some information is known about extreme scenarios. For instance assume that when one large reinsurer goes bankrupt, then one knows that the insurers that are reinsured by this reinsurer will be subject to losses and thus will all incur losses simultaneously (thus showing a comonotonic situation in the tail). If such information is available, it can be incorporated and it may be well possible to redcue the bounds on Value-at-Risk at high levels.

Appendices

A Model Risk of Dependence when Aggregating Risks

The difficulty in modeling the dependence among a large number of risks is a main issue in aggregating risks, i.e. risk aggregation is prone to model risk. In this appendix, we discuss how to minimize or maximize a given risk measure $\varrho(.)$ of the aggregate risk when the distributions of the risky components are known but not their interdependence (consistent with the approach of Embrechts et al. [2013]). In the next section, we will perform the same exercise but now by assuming that additional dependence information is available (following the recent method proposed by Bernard and Vanduffel [2015a]).

In what follows, $\boldsymbol{X} = (X_1, X_2, ..., X_d)$ is the portfolio at hand with given marginal distributions F_{X_1} $(i = 1, 2, ..., d)$ and we are interested in the properties of $\varrho(S)$ where $S = \sum_{i=1}^{d} X_i$. For convenience, we assume that all means are finite.

Recall from (2.12) that the distributional properties of the portfolio $\boldsymbol{X}$ are completely specified if one also knows the copula that describes the interaction among the risks of the portfolio. In this case, the multivariate distribution of $\boldsymbol{X}$ is known and there is clearly only one possible value for $\varrho(S)$. However, when the dependence structure is unspecified, $\varrho(S)$ can take a range of possible values depending on the dependence structure chosen. We aim at finding maximum and minimum possible values for $\varrho(S)$ reflecting the degree of model risk. It is intuitive that for a strong dependence, S becomes a "more variable" risk and $\varrho(S)$ should be at the highest. Reciprocally, if there is a lot of compensation between the risks then $\varrho(S)$ should be small. A well-known device to describe the variability among risks is the so-called convex order. Mathematically, the convex[10] ordering, $\leqslant_{\text{cx}}$ between random variables X, Y is defined as follows

$$X \leqslant_{\text{cx}} Y \quad \text{ if } E(f(X)) \leqslant E(f(Y))$$

for all convex functions $f(\cdot)$ such that the expectation exists. Note that

$$E(X) \leqslant_{\text{cx}} X \tag{A.14}$$

and also that $X \leqslant_{\text{cx}} Y$ implies that X and Y have the same mean but Y has the largest variance. Convex order conforms well with the preferences of risk-averse investors and is very useful to quantify the uncertainty on $\varrho(S)$. Precisely, when the risk measure $\varrho()$ is consistent with convex order, then convex order bounds translate into bounds of the risk measure[11]. This is the case for the variance or the TVaR for instance[12]. As for the Value-at-Risk, this risk measure is not consistent with convex ordering as such, but there is still a close relationship between bounds on VaR and convex order bounds as we will also explain hereafter (see also Bernard et al. [2013b]). In any case, it is thus important to determine upper and lower convex bounds for sums of risks.

[10] For more details on this ordering in the context of actuarial science, see e.g. Müller and Stoyan [2002], Denuit et al. [2005], Denuit et al. [1999] and Dhaene et al. [2002].

[11] Convex order is a natural order in the class of *admissible risks*. Bernard, Jiang, and Wang [2014b] introduce the concept of *admissible risk* to describe all possible aggregate risk S with given marginal distributions but unknown dependence structure.

[12] All concave distortion risk measures are consistent with convex order.

A.1 Convex upper and lower bounds

The convex upper bound for a general number d of individual risks is attained when the risks are maximally dependent (i.e., co-monotonic) which is an easy to describe dependence structure. More precisely, in the comonotonic case one actually considers

$$\mathbf{X} \stackrel{d}{=} (F_{X_1}^{-1}(U_1), F_{X_2}^{-1}(U_2), ..., F_{X_n}^{-1}(U_n)), \tag{A.15}$$

in which now

$$U_1 = U_2 = ... = U_n := U, \tag{A.16}$$

It is intuitively clear that that the variables $X_i = F_i^{-1}(U)$ are fully dependent, as they are maximally increasing in each other. Hence, we obtain that for any portfolio sum $S := \sum_i X_i$ in which the risky components X_i are distributed with F_i,

$$E(S) \leqslant_{\text{cx}} S \leqslant_{\text{cx}} \sum_{i=1}^{n} F_i^{-1}(U) \tag{A.17}$$

Proofs for this result (in particular for the second inequality) can be found at many places, the earliest references being Meilijson and Nádas [1979] and and Rüschendorf [1982].

While the convex upper bound is straightforward to attain, the stated convex lower bound, i.e., $E(S)$, is not attainable (sharp) in general. In fact, getting convex lower bounds that are sharp is a very difficult problem, in particular in higher[13] dimensions. Nevertheless, in what follows we show that there exists an algorithm that makes it possible (at least for portfolios with moderate to high portfolio size, which is the case of interest) to find a dependence among the risks such that the sum S approximately behave as the constant $E(S)$. In other words, the algorithm provides approximations for a convex lower bound of S. Next, we discuss how to find maximum and minimum risk bounds for portfolios when employing the variance and the Value-at-Risk as a risk measure.

A.2 Rearrangement algorithm

The Rearrangement Algorithm of Puccetti and Rüschendorf [2012a] and further extended in Embrechts et al. [2013] can be seen as a practical method to construct dependence between the variables X_j $(j = 1, 2, \ldots, d)$, such that the portfolio sum $S = X_1 + ... + X_d$ becomes as small as possible in convex order. We recall that this algorithm is important for finding minimum bounds on the variance and Tail-Value-at-Risk (the maximum bounds are easy to find and follow from comonotonicity in this case), and turns out to be equally important for finding bounds on Value-at-Risk although VaR does not satisfy convex order.

Without loss of (practical) generality we assume that the variables X_j are discretized

[13]For $d = 2$, the convex lower bound is obtained for $X_1 = F_1^{-1}(U)$ and $X_2 = F_2^{-1}(1 - U)$ as studied by Denuit et al. [1999] and Tankov [2011] and for $d \geqslant 3$ see Bernard et al. [2014b] for some results. In fact, the existence of a sharp lower bound is closely related to the concept of complete mixability (Wang and Wang [2011]) as we explain further in the text; see also Dhaene et al. [2002], Wang and Wang [2011], Embrechts et al. [2013], Wang et al. [2013] for more background and more mathematical results.

and take n values that are put in a matrix $\mathbf{A}$ randomly[14]:

$$\mathbf{A} = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1d} \\ x_{21} & x_{22} & ... & x_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & ... & x_{nd} \end{bmatrix}. \tag{A.18}$$

The matrix $\mathbf{A}$ can be seen as a representation of a possible multivariate structure for $\boldsymbol{X} = (X_1, X_2, ..., X_d)$. Importantly, we do not change the respective marginal distributions of X_j $(j = 1, 2, ..., d)$ by rearranging the outcomes within a column but only the dependence between the X_js.

1. For i from 1 to d, Make the i^{th} column antimonotonic with the sum of the other columns.

2. Start again from column 1, and make it antimonotonic with the sums of the columns from 2 to d.

At each step of this algorithm, we make the $j-$th column antimonotonic with the sum of others, so that the columns, say X_j before rearranging, and $\tilde{X}_j$ after rearranging, verify obviously

$$var\left(\sum_{i=1}^{d} X_i\right) \geqslant var\left(\tilde{X}_j + \sum_{i\neq j} X_i\right).$$

Indeed,

$$var\left(\sum_{i=1}^{d} X_i\right) = var\left(X_j + \sum_{i\neq j} X_i\right)$$

and its minimum when X_j is antimonotonic with $\sum_{i\neq j} X_i$. At each step of the algorithm the variance decreases[15], it is bounded from below (by 0) and thus converges to a limit $\ell \geqslant 0$ (convergence of a monotone sequence of real numbers). If the variance becomes zero, we have found a perfect mixability situation, i.e., the dependence is such that the sum becomes a constant and thus is as convex small as possible (see (A.14)). Otherwise, the algorithm will converge to a local minimum. There is then no guarantee that this minimum is really the minimum of the variance of the sum optimized over all dependence structure, as this minimum may depend on the starting point. However, in practice, it turns out that the convergence is very fast and one typically approximates the situation of complete mixability in a few iterations (unless the portfolio size is very small). In particular, the algorithm works remarkably well for the case of a homogeneous portfolio (in which all X_j have the same distribution).

Remark A.1. The algorithm as described above will always stop in a situation where each column is antimonotonic with the others.[16]

[14]For example, we may put in each column of the matrix $\mathbf{A}$ the elements in increasing order, in which case we work with a comonotonic structure as the start situation (yielding a portfolio sum that is largest possible in convex order).

[15]Note that the situation in which all the columns are antimonotonic with the sum of all others is an obvious necessary condition to have a dependence structure that minimize the variance.

[16]At each step of the algorithm, if a column is not antimonotonic with the sum of the others, then

A.3 Example of Application of the RA

To illustrate the algorithm presented above, we show a very simple example based on a matrix containing 8 rows and 3 columns (i.e., we consider a portfolio containing three risks that take values under eight scenarios) that we report in a matrix similar to the general case given by (A.18)

$$\begin{bmatrix} 3 & 4 & 1 \\ 2 & 1 & 1 \\ 0 & 3 & 2 \\ 1 & 2 & 1 \\ 0 & 4 & 2 \\ 1 & 0 & 1 \\ 3 & 1 & 2 \\ 4 & 2 & 3 \end{bmatrix}. \tag{A.19}$$

Here, we start from the comonotonic structure and apply the RA sequentially as described in the above algorithm and we find (i.e., by applying sequentially the RA on the first, second and third column) that

$$\begin{bmatrix} 3 & 4 & 1 \\ 2 & 1 & 1 \\ 0 & 3 & 2 \\ 1 & 2 & 1 \\ 0 & 4 & 2 \\ 1 & 0 & 1 \\ 3 & 1 & 2 \\ 4 & 2 & 3 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 4 & 1 \\ 3 & 1 & 1 \\ 0 & 3 & 2 \\ 3 & 2 & 1 \\ 0 & 4 & 2 \\ 4 & 0 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 4 & 1 \\ 3 & 2 & 1 \\ 0 & 4 & 2 \\ 3 & 2 & 1 \\ 0 & 3 & 2 \\ 4 & 0 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 4 & 1 \\ 3 & 2 & 1 \\ 0 & 4 & 1 \\ 3 & 2 & 1 \\ 0 & 3 & 2 \\ 4 & 0 & 2 \\ 2 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix}. \tag{A.20}$$

Note that in the last matrix, we find indeed that each column is antimonotonic with the sum of the two others.

A.4 Model risk on dependence on variance

Proposition A.2 (Bounds on the variance of $\sum_{i=1}^{d} X_i$)**.** *Let $(X_1, X_2, ..., X_d)$ be a portfolio with respective marginal distribution F_i. Let $S = \sum_{i=1}^{d} X_i$. We have:*

$$var(E(S)) = 0 \leqslant var(S) \leqslant var\left(\sum_{i=1}^{d} F_i^{-1}(U)\right).$$

in which U is random variable that is uniformly distributed on (0,1).

Proposition A.2 is a straightforward consequence of the fact that variance is consistent with convex order and the convex ordering relation (A.17). Hence, the lower bound that

it is rearranged to make it antimonotonic. Doing so implies that the variance decreases strictly (as the antimonotonicity is the unique dependence structure that attains the minimum variance). The matrix has a finite size and therefore there is a finite number of possible rearrangements of this matrix and therefore the variance can only decrease strictly a finite number of times. If at some point for each column, the variance does not change, it means that each column is antimonotonic with the sum of the others and therefore the algorithm has stopped.

we propose here corresponds to the case in which the portfolio sum is constant, i.e. we have the situation of complete mixability as in Wang and Wang [2011]. In this case, we say that the stated lower bound is "sharp", as there exists no dependence structure among the risks X_i such that the sum is constant and exhibits zero variance exactly. As explained above, the RA attempts to achieve this situation but this is not always possible in which case the stated lower bound in Proposition A.2 is not sharp. In any case, the RA can be seen as a method to get an approximation for the sharp convex lower bound.

Let us illustrate these bounds with the example of 8 observations presented above. The maximum variance is obtained when the risks exhibit a comonotonic dependence (see (A.19)) and we find

$$\begin{bmatrix} 4 & 4 & 3 \\ 3 & 4 & 2 \\ 3 & 3 & 2 \\ 2 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad S := \begin{bmatrix} 11 \\ 9 \\ 8 \\ 6 \\ 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}.$$

As for the minimum variance, after applying the RA in (A.20) we find as output,

$$\begin{bmatrix} 1 & 4 & 1 \\ 3 & 2 & 1 \\ 0 & 4 & 1 \\ 3 & 2 & 1 \\ 0 & 3 & 2 \\ 4 & 0 & 2 \\ 2 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix} \qquad S := \begin{bmatrix} 6 \\ 6 \\ 5 \\ 6 \\ 5 \\ 6 \\ 5 \\ 5 \end{bmatrix}.$$

We see that the lower bound in Proposition A.2 is not attained in this particular case, i.e., there will be no dependence structure among X_1, X_2 and X_3 such that the sum is constant. However, the output of the RA can still be seen as a very good approximation for the sum that is smallest possible with respect to convex order. In other words, the algorithm makes it possible to find approximate (sharp) lower bounds for the variance.

A.5 Model risk on dependence on VaR

As comonotonicity is the worst-case dependence according to risk averse decision makers, it is intuitive that, similar to the case of the variance, this dependence yields the maximum VaR of a portfolio. We will see however that this intuition is wrong in general. Let us first observe that for any sum $S = \sum_{i=1}^{d}(X_i)$ and $0 < p < 1$,

$$\begin{aligned} \mathrm{VaR}_p(S) &\leqslant \mathrm{TVaR}_p(S) && \text{(A.21)} \\ &\leqslant B = \mathrm{TVaR}_p\left(\sum_{i=1}^{d} F_i^{-1}(U)\right) && \text{(A.22)} \end{aligned}$$

Similarly, one finds that

$$A := \text{LTVaR}_p \left(\sum_{i=1}^{d} F_i^{-1}(U) \right) \leqslant \text{VaR}_p (S) \tag{A.23}$$

where we have defined the left Tail Value-at-Risk (LTVaR) at level p $(0 < p < 1)$ as

$$\text{LTVaR}_p(X_i) = \frac{1}{p} \int_0^p \text{VaR}_u[X_i] \mathrm{d}u. \tag{A.24}$$

Note that for TVaR and LTVaR,

$$(\text{L})\text{TVaR}_p \left(\sum_{i=1}^{d} F_i^{-1}(U) \right) = \sum_i (\text{L})\text{TVaR}_p(F_i^{-1}(U)) \tag{A.25}$$

In summary, we then obtain the following result.

Theorem A.3 (Bounds on the VaR of $\sum_{i=1}^{d} X_i$.)**.** *Let $(X_1, X_2, ..., X_d)$ be a portfolio with respective marginal distribution F_i. Let $S = \sum_{i=1}^{d} X_i$ and $p \in (0, 1)$. Then,*

$$\sum_{i=1}^{n} \text{LTVaR}_q(F_i^{-1}(U)) \leqslant VaR_q \left(\sum_{i=1}^{d} X_i \right) \leqslant \sum_{i=1}^{n} \text{TVaR}_q(F_i^{-1}(U)). \tag{A.26}$$

These bounds are given and proved in Bernard et al. [2013b]. The question is then if these bounds can be sharp. To deal with this problem let us note that

$$\text{VaR}_p \left(\sum_{i=1}^{d} F_i^{-1}(U) \right) \leqslant \text{TVaR}_p \left(\sum_{i=1}^{d} F_i^{-1}(U) \right). \tag{A.27}$$

Hence, in order to attain the upper bound B, the idea is to start with the comonotonic dependence and next change it such that the inequality (A.26) turns into an equality. As TVaR_p is the average of all upper VaR_qs on the interval $[p, 1]$, it is clear that the equality is obtained if the VaR of the comonotonic sum $\sum_{i=1}^{d} F_i^{-1}(U)$ becomes constant on $[p, 1]$ (by changing this comonotonic dependence). Let G_i denote the distribution of F_i when restricted[17] to the upper p-part of F_i. In order to attain the upper bound, one thus needs to find a dependence between the risks (now with marginal distributions G_i) such that the corresponding sum becomes constant (i.e., the risks are completely mixing). In general, the mixing property does not hold and the stated bounds are thus not sharp. However, it is now clear that (approximations of) sharp VaR bounds are obtained by finding a dependence between the risks (with marginal distributions G_i) such that the corresponding sum becomes as convex *small* as possible (see also Bernard et al. [2013b]). A similar reasoning shows that in order to reach the stated lower bound as closely as possible one should change the comonotonic dependence such that the quantile function of the comonotonic portfolio sum becomes as flat as possible on the interval $[0, p]$.

We build on this idea to propose a practical algorithm to approximate sharp bounds. Hence, let us show how to find approximate sharp bounds with the discrete example

[17] Formally, G_i is the distribution of $F_i^{-1}(V)$, where V is uniformly distributed on $[q, 1]$.

discussed above when the level p used to assess the VaR is 5/8. Note that we start from the the comonotonic structure in the matrix.

$$\begin{bmatrix} 4 & 4 & 3 \\ 3 & 4 & 2 \\ 3 & 3 & 2 \\ 2 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 4 & 3 \\ 3 & 4 & 2 \\ 3 & 3 & 2 \end{bmatrix}.$$

We then apply the RA in the three corresponding rows.

$$\begin{bmatrix} 4 & 4 & 3 \\ 3 & 4 & 2 \\ 3 & 3 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 3 & 3 \\ 3 & 4 & 2 \\ 3 & 4 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 3 & 2 \\ 3 & 4 & 2 \\ 3 & 4 & 3 \end{bmatrix}.$$

So that the sums are respectively 9, 9 and 10 and thus the maximum VaR is 9. To obtain the minimum VaR, one works on the lower values of each X_i and apply the RA on these values

$$\begin{bmatrix} 4 & 4 & 3 \\ 3 & 4 & 2 \\ 3 & 3 & 2 \\ 2 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Applying the RA as described above

$$\begin{bmatrix} 2 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 0 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 2 & 1 \end{bmatrix}.$$

so that the values of the sums are 4,3,3,3 and 3. Therefore the minimum VaR is 4.

B Model Risk of Dependence when Aggregating Risks and Model Risk Quantification

In this appendix, we present a framework, which allows practical quantification of model risk (and was developed in Bernard and Vanduffel [2015a]). We assume that risk modelers have developed an "as good as possible" multivariate model for a certain portfolio $(X_1, X_2, ..., X_d)$. However, no model is perfect and we want to assess to which extent misspecification of the proposed model affects the risk measurement of $S = \sum_i X_i$. Importantly, unlike Appendix A we do no longer ignore the available information on dependence. Our framework includes an algorithm allowing actuaries to deal with model risk in a very practical way as we will show in full details.

These results make it possible to identify risk measures for which additional information of a well-fitted multivariate model reduces the model risk significantly, making them meaningful candidates for use by risk managers and regulators. For instance, we observe from numerical experiments that the portfolio VaR at a very high confidence level (as used in the current Basel regulation) might be prone to such a high level of model risk that, even if one knows the multivariate distribution nearly perfectly, its range of possible values remains wide. In fact, one may then not even be able to reduce the model risk as computed in Embrechts et al. [2013] (see also Appendix A) where no information on the dependence among the risks is used at all.

The idea pursued in our approach is intuitive and corresponds to real-world situations. Let us assume that we have observed N d-dimensional vectors of observations $\{(x_{1i}, ..., x_{di})\}_{i=1,...,N}$ and that a multivariate model has already been fitted to this dataset. In other words, there is a joint distribution of $(X_1, X_2, ..., X_d)$ available (benchmark model). However, we are aware that the model is subject to misspecification, especially due to lack of data. Hence, we split $\mathbb{R}^d$ into two subsets: $\mathcal{F}$ will be referred to as the "fixed" or "trusted" area and $\mathcal{U}$ as the "unfixed" or "untrusted"area. $\mathcal{U}$ reflects the area in which the data are not considered trustworthy (rich) enough to conclude that the fitted model is appropriate (in that area). Note that

$$\mathbb{R}^d = \mathcal{F} \bigcup \mathcal{U}.$$

If one has perfect trust in the model, then all observations reside in the "trusted" part ($\mathcal{U} = \emptyset$) and there is no model risk. On the contrary, $\mathcal{F} = \emptyset$ when there is no trust in the fit of the dependence, which corresponds to the case studied by Embrechts et al. [2013] (see also Appendix A)

A closely related problem has already been studied for two-dimensional portfolios ($d = 2$) when some information on the dependence (copula) is available; see for example Tankov [2011], Bernard et al. [2012] and Bernard et al. [2013a]. Tankov [2011] uses extreme dependence scenarios to find model-free bounds for the prices of some bivariate derivatives, whereas Bernard et al. [2014] and Bernard et al. [2014][18] use such scenarios to determine optimal investment strategies for investors with state-dependent constraints. While both applications show that finding bounds on copulas in the bivariate case can be of interest, risk management typically involves more than two risks. Unfortunately, finding bounds on copulas in the general $d-$dimensional case in the presence of constraints is not only more difficult but also less useful for risk management applications. The reason is that when $d > 2$, in most cases, the worst copula (under constraints) of a vector $(X_1, X_2, ..., X_d)$ does not give rise to the highest possible value of the risk measure at hand of $S = \sum_{i=1}^{d} X_i$, because the marginal distributions also have an impact; see e.g. Bernard et al. [2014b] for illustrations of this feature.

In addition, there are very few papers in the literature that deal explicitly with high-dimensional problems and the presence of (partial) information on the dependence structure: Rüschendorf [1991], Embrechts and Puccetti [2010b] and Embrechts et al. [2013] consider the situation in which some of the bivariate distributions are known, Denuit et al. [1999] study VaR bounds assuming that the joint distribution of the risks is bounded by some distribution and Bernard et al. [2013b] compute VaR bounds when the variance of the sum is known. They are able to show how the information on the variance of the

[18]This paper received the 2015 Redington prize from the SOA.

aggregate risk allows to reduce the distance between the maximum VaR and the minimum VaR significantly. Moreover, they provide an algorithm (Extended RA) that can be used by actuaries to assess Value-at-Risk bounds (and thus model risk on Value-at-Risk assessment) in a very practical way when only the variance of the aggregate risk is known.

The setup in all these papers is sometimes hard to reconcile with the information that is available in practice; or, it does not make use of all available dependence information. Furthermore, while the bounds that are proposed in these papers might be sharp (attainable), they are often hard to compute numerically, especially for higher dimensions with inhomogeneous risks. Note also that the bounds obtained do not always make it possible to strengthen the unconstrained bounds in a significant way, suggesting that additional dependence information is needed in order to obtain better bounds; see also Wang and Wang [2011], Embrechts et al. [2013], Wang et al. [2013] and Bernard et al. [2014b] for related results.

Hence, in this final section, we study bounds for risk measures of the aggregate risk S by using information on the *multivariate* joint distribution of its components X_i (which embeds information on the dependence) rather than using copula information. We propose two methods for deriving bounds on risk measures. The first method could be non parametric, it builds on the RA of Embrechts et al. [2013] and allows to find sharp bounds. The second method provides analytic, simple bounds but may not be sharp. In practice, we also show through examples that analytical bounds are close to be sharp.

The outline of this section is as follows. First, we present the practical method using the RA (Section B.3). This method can be performed directly using the data at hand (without fitting a model), so that in this case, model risk can be assessed in a fully non-parametric way. This method builds on the rearrangement algorithm that was recently developed by Puccetti and Rüschendorf [2012a] and further studied by Embrechts et al. [2013]. It relies on a discretized version of the problem described above and uses a matrix representation to approximate the worst case dependence structures. We then give the analytical form and illustrate them by simulations of N(0,1) risks and Pareto risks in Section B.4. We provide bounds that can be computed directly (using, for instance, Monte Carlo simulations) but that may not be sharp. Our numerical results indicate that in high dimensions the bounds computed using the direct method in Section B.4 are close to the non-parametric bounds as computed in Section B.3. In other words, while finding sharp bounds is theoretically a difficult problem, the numerical illustrations suggest that the algorithm that we propose in Section B.4 leads to nearly sharp bounds. The numerical results also show that the new bounds typically outperform the (unconstrained) ones already available in the literature and thus allow for more realistic assessment of model risk. However, model risk remains a significant concern, especially when using a risk measure that focuses on "tail type" events, such as the VaR.

B.1 Theoretical Setting and Assumptions

Let $(X_1, X_2, ..., X_d)$ be some random vector of interest. Let $\mathcal{F} \subset \mathbb{R}^d$ and $\mathcal{U} = \mathbb{R}^d \backslash \mathcal{F}$. We assume that we know

(i) the marginal distribution F_i of X_i on $\mathbb{R}$ for $i = 1, 2, ..., d$,

(ii) the distribution of $(X_1, X_2, ..., X_d) \mid \{(X_1, X_2, ..., X_d) \in \mathcal{F}\}$

(iii) and the probability $p_f := P((X_1, X_2, ..., X_d) \in \mathcal{F})$, as well as $p_u := P((X_1, X_2, ..., X_d) \in \mathcal{U}) = 1 - p_f$.

The joint distribution of $(X_1, X_2, ..., X_d)$ is thus not completely specified (unless $\mathcal{F} = \mathbb{R}^d$ and $\mathcal{U} = \emptyset$). Consequently, risk measures (e.g., the VaR) of the aggregate sum $\sum_{i=1}^d X_i$ cannot be computed precisely. In fact, there are many vectors $(Y_1, Y_2, ..., Y_d)$ that agree with $(X_1, X_2, ..., X_d)$ for the properties (i), (ii) and (iii) but have a different risk measure of their sum. In what follows, we are interested in finding the extreme possible values of the risk measure at hand, as the gap between the minimum and the maximum can be useful in measuring model risk. Formally, we use the following definition of model risk. This definition is in the same spirit as in Barrieu and Scandolo [2015].

Definition B.1 (Model risk)**.** Let $(X_1, X_2, ..., X_d)$ be a random vector satisfying (i), (ii) and (iii) and assume that one uses a (law-invariant) risk measure $\varrho(\cdot)$ to assess the risk of $\sum_{i=1}^d X_i$. Define

$$\varrho_{\mathcal{F}}^+ := \sup\left\{\varrho\left(\sum_{i=1}^d Y_i\right)\right\}, \quad \varrho_{\mathcal{F}}^- := \inf\left\{\varrho\left(\sum_{i=1}^d Y_i\right)\right\}$$

where the supremum and the infimum are taken over all other (joint distributions of) random vectors $(Y_1, Y_2, ..., Y_d)$ that agree with (i), (ii) and (iii). The model risk that one underestimates the risk by computing a direct estimate of $\varrho(\sum X_i)$ in some chosen benchmark model (i.e., when some multivariate distribution for $(X_1, ..., X_d)$ has been specified) is defined as

$$\frac{\varrho_{\mathcal{F}}^+ - \varrho(\sum_{i=1}^n X_i)}{\varrho_{\mathcal{F}}^+} \tag{B.28}$$

and, similarly, the model risk for overestimation is given as

$$\frac{\varrho(\sum_{i=1}^n X_i) - \varrho_{\mathcal{F}}^-}{\varrho_{\mathcal{F}}^-}. \tag{B.29}$$

The rest of this section aims at obtaining the maximum and minimum possible values $\varrho_{\mathcal{F}}^+$ and $\varrho_{\mathcal{F}}^-$ of $\varrho(\sum_{i=1}^d X_i)$. The recent literature on model risk estimation has dealt mainly with the case in which there is full uncertainty on the dependence among the risks X_i $(i = 1, 2, ..., d)$, i.e., when $\mathcal{F} = \emptyset$ (Appendix A where we reviewed the work of Embrechts et al. [2013] with respect to VaR. See also Bernard et al. [2014b] regarding a convex risk measure. In this section, we consider the case in which information on the dependence translates into joint distributions that are partially known.

In this respect, it will be useful to consider the indicator variable $\mathbb{I}$ corresponding to the event "$(X_1, X_2, ..., X_d) \in \mathcal{F}$"

$$\mathbb{I} := \mathbb{1}_{(X_1, X_2, ..., X_d) \in \mathcal{F}} \tag{B.30}$$

so that one can express the probabilities that a random vector takes values in $\mathcal{F}$ resp. in$\mathcal{U}$ as

$$p_f = P(\mathbb{I} = 1) \text{ and } p_u = P(\mathbb{I} = 0). \tag{B.31}$$

Let us also introduce a standard uniformly distributed random variable U independent of the event "$(X_1, X_2, ..., X_d) \in \mathcal{F}$" (and thus also independent of $\mathbb{I}$) as well as a random vector $(Z_1, Z_2, ..., Z_d)$ defined by

$$Z_i = F^{-1}_{X_i|(X_1, X_2, ..., X_d) \in \mathcal{U}}(U), \qquad i = 1, 2, ..., d, \tag{B.32}$$

where $F^{-1}_{X_i|(X_1,X_2,...,X_d)\in\mathcal{U}}$ denotes the (left) inverse of the distribution function

$$F_{X_i|(X_1,X_2,...,X_d)\in\mathcal{U}}(x) := P(X_i \leqslant x|(X_1, X_2, ..., X_d) \in \mathcal{U}).$$

Note that $F^{-1}_{X_i|(X_1,X_2,...,X_d)\in\mathcal{U}}(x)$ can be computed, as the marginal distribution of X_i is known and the joint distribution of $(X_1, X_2, ..., X_d)$ is known on $\mathcal{F}$ (see the properties (i) and (ii)). Further, all Z_i $(i = 1, 2, ..., d)$ are increasing in the (common) variable U, and thus $(Z_1, Z_2, ..., Z_d)$ is a comonotonic vector with known joint distribution. Define also

$$T := F^{-1}_{\sum_i X_i|(X_1,X_2,...,X_d)\in\mathcal{F}}(U). \tag{B.33}$$

Hence, T is a random variable with distribution $F_{\sum_i X_i|(X_1,X_2,...,X_d)\in\mathcal{F}}(x)$.

While most of our results hold generally or can be extended in a straightforward way, we will focus on bounds for the variance, the VaR and the TVaR (which have been defined in the previous section in (A.21), (A.22) and (A.24)).

B.2 Practical framework for the non-parametric approach

We follow the same setting as already introduced in Appendix A. We have N observations of the d-dimensional vector $(x_{i1}, x_{i2}, ..., x_{id})$ for $i = 1, ..., N$. Denote by $M = (x_{ij})$ the corresponding $N \times d$ matrix. These N observations may simply be N observed data vectors or N simulated vector values obtained from a fitted multivariate distribution of $(X_1, X_2, ..., X_d)$. In both cases, each observation $(x_{i1}, x_{i2}, ..., x_{id})$ occurs with probability $\frac{1}{N}$ naturally (possibly involving repetitions). We assume that the matrix M contains enough data to allow for an accurate description of the marginal distributions of X_k $(k = 1, 2, ..., d)$ so that the matrix M can effectively be seen as a representation of the random vector of interest $(X_1, X_2, ..., X_d)$. Define S_N by $S_N(i) = \sum_{k=1}^{d} x_{ik}$ for $(i = 1, 2, ..., N)$. In other words, S_N can be seen as a random variable that takes the value $S_N(i)$ in "state" i for $i = 1, 2, ..., N$. In general, it might be difficult to find sharp bounds for risk measures of $S = \sum_i X_i$. The purpose of what follows is to deal with this problem using the "sampled" counterpart S_N of S, rather than S itself.

As in the theoretical setting presented above, we suppose that the joint distribution of $(X_1, X_2, ..., X_d)$ is not completely specified. In the context of the matrix representation M for the vector $(X_1, X_2, ..., X_d)$, we assume that the matrix M is effectively split into two parts. There is a submatrix $\mathcal{F}_N$ of trusted observations $(x_{i1}, x_{i2}..., x_{id})$ and $\mathcal{U}_N$ consists of the rest of the observations. In the sequel, the set $\mathcal{F}_N$ will be referred to as the "fixed" or "trusted" part and $\mathcal{U}_N$ as the "untrusted" part. In the case in which one has perfect trust in all observations, the "untrusted" part contains no elements ($\mathcal{U}_N = \emptyset$) and S_N can be used to assess the risk of S. By contrast, if one has no trust in the observed dependence, then $\mathcal{F}_N = \emptyset$. In this case, the observations $(x_{i1}, x_{i2}, ..., x_{id})$ are useful only in modeling marginal distributions F_k $(k = 1, 2, ..., d)$ and do not allow for conclusions regarding the dependence. It is then important to observe that rearranging the values x_{ik} $(i = 1, 2, ..., N)$ within the $k-$th column does not affect the empirical marginal distribution of X_k but only changes the observed dependence ("interaction between elements of different columns").

Let us denote by ℓ_f the number of elements in $\mathcal{F}_N$ and by ℓ_u the number of elements in $\mathcal{U}_N$, such that

$$N = \ell_f + \ell_u.$$

Without any loss of generality, it is convenient to modify the matrix M by changing the order of the rows so that the "trusted area" corresponds to the ℓ_f first rows and the untrusted area corresponds to the last ones. By doing so, we have only reallocated the states $i = 1, 2, ..., N$, without impact on the adequacy of M to describe the distributional (law-invariant) properties of $(X_1, X_2, ..., X_d)$. Similarly, as per definition of the submatrix $\mathcal{U}_N$, we are allowed to rearrange the values *within the columns* of $\mathcal{U}_N$ (and thus within the corresponding parts of M), as this operation generates a new matrix that is considered as trustworthy as the initial one (note, indeed, that we do not know the dependence between the X_i, conditionally on $(X_1, X_2, ..., X_d) \in \mathcal{U}$).

Without loss of generality, we can thus always assume that the matrix $\mathcal{U}_N$ depicts a comonotonic dependence (in each column, the values are sorted in decreasing order, that is, such that $x_{m_1 k} \geqslant x_{m_2 k} \geqslant ... \geqslant x_{m_{\ell_u} k}$ for all $k = 1, 2, ..., d$). Finally, for $\mathcal{F}_N$ (and thus also for the corresponding part of $\mathcal{X}_N$) we can assume that the ℓ_f observations $(x_{i_j 1}, x_{i_j 2} ... x_{i_j d})$ appear in such a way that for the sums of the components, i.e., $s_j := x_{i_j 1} + x_{i_j 2} + ... + x_{i_j d}$ ($j = 1, 2, ..., \ell_f$), it holds that $s_1 \geqslant s_2 \geqslant ... \geqslant s_{\ell_f}$. From now on, the observed data points are reported in the following matrix M

$$M = \begin{bmatrix} x_{i_1 1} & x_{i_1 2} & ... & x_{i_1 d} \\ x_{i_2 1} & x_{i_2 2} & ... & x_{i_2 d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i_{\ell_f} 1} & x_{i_{\ell_f} 2} & ... & x_{i_{\ell_f} d} \\ x_{m_1 1} & x_{m_1 2} & ... & x_{m_1 d} \\ x_{m_2 1} & x_{m_2 2} & ... & x_{m_2 d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m_{\ell_u} 1} & x_{m_{\ell_u} 2} & ... & x_{m_{\ell_u} d} \end{bmatrix}, \tag{B.34}$$

where the grey area reflects $\mathcal{F}_N$ and the white area reflects $\mathcal{U}_N$. The corresponding vectors S_N^f and S_N^u consist of sums of the components for each observation in the trusted (respectively untrusted) part:

$$\begin{bmatrix} S_N^f \\ S_N^u \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{\ell_f} \\ \tilde{s}_1 := x_{m_1 1} + x_{m_1 2} + ... + x_{m_1 d} \\ \tilde{s}_2 := x_{m_2 1} + x_{m_2 2} + ... + x_{m_2 d} \\ \vdots \\ \tilde{s}_{\ell_u} := x_{m_{\ell_u} 1} + x_{m_{\ell_u} 2} + ... + x_{m_{\ell_u} d} \end{bmatrix}. \tag{B.35}$$

While $s_1 \geqslant s_2 \geqslant ... \geqslant s_{\ell_f}$ are trusted, the sums $\tilde{s}_i$ change when the choice of dependence in $\mathcal{U}_N$ is varied. In fact, the set $\{i_1, ..., i_{\ell_f}\}$ can be seen as the collection of states (scenarios) in which the corresponding observations are trusted, whereas the set $\{m_1, ..., m_{\ell_u}\}$ provides the states in which there is doubt with respect to the dependence structure.

For pedagogical purposes, we now provide a simple example of this setup. It will be used throughout this section to illustrate each algorithm that we propose. This toy example is not meant to represent a realistic set of observations since, in true applications, there would be a large number of observations (here $N = 8$) and possibly a large number of

variables (here $d = 3$). The eight observations are given as follows, with three observations trusted ($\ell_f = 3$), which appear in the grey area of the matrix:

$$\begin{bmatrix} 3 & 4 & 1 \\ 1 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 0 & 1 \\ 1 & 1 & 2 \\ 4 & 2 & 3 \end{bmatrix}. \tag{B.36}$$

Without loss of generality we can then consider for further analysis the following matrix M and the vectors of sums S_N^f and S_N^u, as follows:

$$M = \begin{bmatrix} 3 & 4 & 1 \\ 2 & 4 & 2 \\ 0 & 2 & 1 \\ 4 & 3 & 3 \\ 3 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad S_N^f = \begin{bmatrix} 8 \\ 8 \\ 3 \end{bmatrix}, \quad S_N^u = \begin{bmatrix} 10 \\ 7 \\ 4 \\ 3 \\ 1 \end{bmatrix}. \tag{B.37}$$

Finally, with some abuse of notation (completing by 0 so that S_N^f and S_N^u take 8 values), one also has the following representation of S_N :

$$S_N = \mathbb{I} S_N^f + (1 - \mathbb{I}) S_N^u, \tag{B.38}$$

where $\mathbb{I} = 1$ if $(x_{i1}, x_{i2}...x_{id}) \in \mathcal{F}_N$ $(i = 1, 2, ..., N)$. In fact, S_N^f can be readily seen as the sampled counterpart of the T that we used previously (see Definition B.33), whereas S_N^u is a comonotonic sum and corresponds to the sampled version of $\sum_{i=1}^d Z_i$. In this last section, we aim at finding worst case dependence structures allowing for a robust risk assessment of the portfolio sum S (S_N). This amounts to rearranging the outcomes in the columns of the untrusted part $\mathcal{U}_N$ such that the risk measure at hand for S_N becomes maximized (resp. minimized).

B.3 Bounds on a given risk measure

In this section we discuss the sharpness[19] of upper and lower bounds for the three risk measures defined early, namely the variance, TVaR and VaR.

We discussed analytical bounds and sharpness of these bounds in the previous section and in particular in Proposition A.2 that gives unconstrained bounds for variance and Proposition A.3 that gives unconstrained bounds for VaR. They can be naturally extended to the case under study here with constraints.

[19]Recall that a bound on a risk measure is "sharp" if there exists a dependence structure among the risks such that this bound is attained.

B.3.1 Theoretical Bounds on Variance and Tail Value-at-Risk:

The next proposition gives easy-to-compute upper and lower bounds for the variance of a portfolio sum $\sum_{i=1}^{d} X_i$.

Proposition B.2 (Bounds on the variance of $\sum_{i=1}^{d} X_i$)**.** *Let $(X_1, X_2, ..., X_d)$ be a random vector that satisfies properties (i), (ii) and (iii), and let $\mathbb{I}$ and $(Z_1, Z_2, ..., Z_d)$ be defined as in (B.30) and (B.32). We have:*

$$var\left(\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} E(Z_i)\right) \leqslant var\left(\sum_{i=1}^{d} X_i\right) \leqslant var\left(\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} Z_i\right).$$

The proof of Proposition B.2 can be found in Bernard and Vanduffel [2015a]. The stated upper and lower bounds in Proposition B.2 are intuitive and extend in a natural way Proposition A.2. When computing the variance of the portfolio sum $\sum_{i=1}^{d} X_i$, one needs to consider the events $(X_1, X_2, ..., X_d) \in \mathcal{F}$ and $(X_1, X_2, ..., X_d) \in \mathcal{U}$ separately. The distribution of $\sum_{i=1}^{d} X_i$ is known on the event $\{(X_1, X_2, ..., X_d) \in \mathcal{F}\}$, but unknown on the event $\{(X_1, X_2, ..., X_d) \in \mathcal{U}\}$. On $\mathcal{U}$, one then substitutes sum $\sum_i X_i$ by the constant $\sum_i E(Z_i)$ (to compute the lower bound and thus to minimize variance) and by the comonotonic sum $\sum_i Z_i$ (to compute the upper bound and thus to maximize variance). Note in particular that when $\mathcal{U} = \emptyset$, the upper bound is equal to the lower bound and there is no model risk.

Next, we discuss bounds for the TVaR as they are similar to the variance (because oth measures are consistent with convex order).

Proposition B.3 (Bounds on the TVaR of $\sum_{i=1}^{d} X_i$)**.** *Let $(X_1, X_2, ..., X_d)$ be a random vector that satisfies properties (i), (ii) and (iii), and let $\mathbb{I}$ and $(Z_1, Z_2, ..., Z_d)$ as defined in (B.30) and (B.32). We have that*

$$TVaR_p\left(\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} E(Z_i)\right) \leqslant TVaR_p\left(\sum_{i=1}^{d} X_i\right) \leqslant TVaR_p\left(\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} Z_i\right).$$

There is no model risk (the bounds reduce to the same value) when $\mathcal{U} = \emptyset$.

Some of the bounds stated in Propositions B.2 and B.3 can be sharp. In particular, the *upper* bounds for the variance and for the TVaR stated in propositions B.2 and B.3 are sharp, without further conditions. Note, indeed, that the multivariate vector

$$(\mathbb{I}X_1 + (1-\mathbb{I})Z_1, \mathbb{I}X_2 + (1-\mathbb{I})Z_2, ..., \mathbb{I}X_d + (1-\mathbb{I})Z_d) \tag{B.39}$$

satisfies conditions (i) and (ii). In contrast, the stated lower bounds may not be sharp because $\mathbb{I}X_i + (1-\mathbb{I})E(Z_i)$ is usually not distributed with F_i $(i = 1, 2, ..., d)$. In order to get close to the stated lower bounds, one should try to modify the dependence of the vector $(Z_1, Z_2, ..., Z_d)$ such that $Z_1 + Z_2 + ... + Z_d$ becomes constant (and thus equal to $E(Z_1) + E(Z_2) + ... + E(Z_d)$). We use this insight to propose an algorithm below that makes it possible to approximate the sharp bounds when the risk measure used is the standard deviation or the TVaR.

B.3.2 Practical Bounds on Standard Deviation and TVaR

From Proposition B.2 it is clear that in order to maximize the variance of S_N one needs a comonotonic scenario on $\mathcal{U}_N$. However, we have already initialized a comonotonic structure (without loss of generality), and the corresponding values of the sums are exactly the values $\tilde{s}_i$ $(i = 1, 2, ..., \ell_u)$ reported for S_N^u in (B.35). The upper bound on variance is then computed as

$$\frac{1}{N}\left(\sum_{i=1}^{\ell_f}(s_i - \bar{s})^2 + \sum_{i=1}^{\ell_u}(\tilde{s}_i - \bar{s})^2\right), \tag{B.40}$$

where the average sum $\bar{s}$ is given by

$$\bar{s} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{d} x_{ij} = \frac{1}{N}\left(\sum_{i=1}^{\ell_f} s_i + \sum_{i=1}^{\ell_u}\tilde{s}_i\right). \tag{B.41}$$

To achieve the minimum variance bound found in Proposition B.2, the values of S_N^u must be as close as possible to each other; ideally, S_N^u must be constant. In this regard, the concept of complete mixability appears as a theoretical device. "Complete mixability" refers to the dependence structure that makes the sum S_N^u constant (Wang and Wang [2011]). To make this the case, in practice, we apply the rearrangement algorithm of Embrechts et al. [2013] to the matrix U_N (untrusted part) to render it as close as possible to the complete mixability condition. For completeness, the algorithm is presented in Section A.2. Denote by $\tilde{s}_i^m$ the corresponding values of the sums of S_N^u after applying the RA. We then compute the minimum variance as follows:

$$\frac{1}{N}\left(\sum_{i=1}^{\ell_f}(s_i - \bar{s})^2 + \sum_{i=1}^{\ell_u}(\tilde{s}_i^m - \bar{s})^2\right), \tag{B.42}$$

where $\bar{s}$ is computed as in (B.41).

We illustrate the upper and lower bounds (B.40) and (B.42) for the variance derived above with the matrix M of observations provided in (B.37). We then use the comonotonic structure for the untrusted part of the matrix M and compute the vectors of sums S_N^f and S_N^u as defined above in (B.37). The average sum is $\bar{s} = 5.5$. The maximum variance is equal to

$$\frac{1}{8}\left(\sum_{i=1}^{3}(s_i - \bar{s})^2 + \sum_{i=1}^{5}(\tilde{s}_i^c - \bar{s})^2\right) \approx 8.75.$$

For the lower bound, we apply the RA to the U_N and we obtain

$$M = \begin{bmatrix} 3 & 4 & 1 \\ 2 & 4 & 2 \\ 0 & 2 & 1 \\ 1 & 1 & 3 \\ 0 & 3 & 2 \\ 1 & 2 & 2 \\ 3 & 1 & 1 \\ 4 & 0 & 1 \end{bmatrix}, \quad S_N^f = \begin{bmatrix} 8 \\ 8 \\ 3 \end{bmatrix}, \quad S_N^u = \begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \\ 5 \end{bmatrix}. \tag{B.43}$$

With an average sum $\bar{s} = 5.5$, the minimum variance can be calculated as

$$\frac{1}{8}\left(\sum_{i=1}^{3}(s_i - \bar{s})^2 + \sum_{i=1}^{5}(\tilde{s}_i^m - \bar{s})^2\right) \approx 2.5.$$

Assume that we want to fix the TVaR at probability level p, so that, for ease of exposition,

$$k := N(1-p), \tag{B.44}$$

where k is an integer. Similarly to the case of maximizing the variance, it follows from Proposition B.3 that in order to obtain the maximum TVaR one needs a comonotonic scenario on $\mathcal{U}_N$. Hence, we merely need to select the k highest values from S_N^f and S_N^u as computed in (B.35). Let us label these values by s_1^*,s_2^*,...,s_k^* (ranked in decreasing order), and we can then easily compute the maximum TVaR at probability level p. Also, the minimum TVaR is obtained similarly to the minimum variance. First, apply the RA to the untrusted part of the matrix U_N to render the variance of the (new) sum S_N^u as small as possible. Then, select the k highest values out of S_N^f and S_N^u, say: s_1^*,s_2^*,...,s_k^* (ranked in decreasing order) and compute the minimum TVaR.

Let us consider the previous example again. Let us choose $p = 5/8$, so that $k = 3$. The highest $k = 3$ values are 8, 8 and 10 and the maximum TVaR is then 26/3 (≈ 8.67). After application of the RA, we obtain (B.43) for S_N^u and thus the highest three outcomes that we observe for S_N^u and S_N^f are 8, 8 and 5. Hence, the minimum TVaR is $21/3 = 7$.

B.3.3 Theoretical Bounds on Value-at-Risk

VaR is a widely used risk measure in financial services. The following proposition provides bounds on VaR.

Proposition B.4 (VaR Bounds for $\sum_{i=1}^{d} X_i$)**.** *Let $(X_1, X_2, ..., X_d)$ be a random vector that satisfies properties (i), (ii) and (iii), and let $\mathbb{I}$, $(Z_1, Z_2, ..., Z_d)$ and U be defined as in (B.30) and (B.32). Define the variables L_i and H_i as*

$$L_i = LTVaR_U(Z_i) \text{ and } H_i = TVaR_U(Z_i)$$

and let

$$M_p := VaR_p\left(\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} H_i\right), \quad m_p := VaR_p\left(\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} L_i\right)$$

Bounds on the Value-at-Risk of the aggregate risk are given as

$$m_p \leqslant VaR_p\left(\sum_{i=1}^{d} X_i\right) \leqslant M_p. \tag{B.45}$$

The proof can be found in Bernard and Vanduffel [2015b]. Initially, the appearance of variables L_i and H_i may seem somewhat odd. However, note that the variables Z_i, which played crucial roles in Propositions B.2 and B.3, can also be expressed as $Z_i = VaR_U(Z_i)$, and here we merely use $TVaR_U(Z_i)$ and $LTVaR_U(Z_i)$ instead. Thus, Proposition B.4 has a similar form[20] to that of Proposition B.3 resp. B.2, but the bounds proposed are usually not sharp.[21] We observe that in the case of no uncertainty (i.e., $\mathcal{U} = \emptyset$) there is

[20]Note that VaR is not consistent with convex order, although there are some connections (see Bernard et al. [2013b] and Bernard et al. [2012]).

[21]Note, indeed, that the variables H_i and L_i are not distributed as $(X_i|\mathbb{I} = 0)$.

no model risk, as $\mathbb{I} = 1$. When there is full uncertainty, i.e., $\mathcal{U} = \mathbb{R}^d$, then $\mathbb{I} = 0$, and we are returned to the unconstrained lower bound on the VaR of a portfolio given in Proposition A.3 (see also Theorem 2.1 of Bernard, Rüschendorf, and Vanduffel [2013b]). Note also that the VaR bounds are not sharp in general.

For practical calculations it might be convenient to use an alternative formulation of the stated VaR bounds.

Proposition B.5 (Alternative formulation of the VaR Bounds)**.** *Let $(X_1, X_2, ..., X_d)$ be a random vector that satisfies properties (i), (ii) and (iii), and let $\mathbb{I}$, $(Z_1, Z_2, ..., Z_d)$ and T be defined as in (B.30), (B.32) and* (B.33)*. Recall that $p_f = P(\mathbb{I} = 1)$. Define*

$$\alpha_* := \inf\left\{\alpha \in (\alpha_1, \alpha_2) \mid VaR_\alpha(T) \geqslant TVaR_{\frac{p-p_f\alpha}{1-p_f}}\left(\sum_{i=1}^d Z_i\right)\right\},$$

where $\alpha_1 = \max\left\{0, \frac{p+p_f-1}{p_f}\right\}$ and $\alpha_2 = \min\left\{1, \frac{p}{p_f}\right\}$. Then, for $p \in (0,1)$,

$$M_p = \begin{cases} TVaR_{\frac{p-p_f\alpha_*}{1-p_f}}\left(\sum_{i=1}^d Z_i\right) & \text{if } \frac{p+p_f-1}{p_f} < \alpha_* < \frac{p}{p_f} \\ VaR_{\alpha_*}(T) & \text{if } \alpha_* = \frac{p}{p_f} \\ \max\left\{VaR_{\alpha_*}(T), TVaR_{\frac{p-p_f\alpha_*}{1-p_f}}\left(\sum_{i=1}^d Z_i\right)\right\} & \text{if } \alpha_* = \frac{p+p_f-1}{p_f}. \end{cases} \tag{B.46}$$

The expressions for the lower bound m_p are obtained by replacing, in the above statements, "TVaR" with "LTVaR".

The proof of Proposition B.5 is provided in the appendix of Bernard and Vanduffel [2015a]. We can illustrate Proposition B.5 in a discrete setting, in which the probability space Ω has N states. Assume that the event $\{(X_1, ..., X_d) \in \mathcal{F}\}$ corresponds to the set $\{\omega_1, ..., \omega_{\ell_f}\}$, whereas $\{(X_1, ..., X_d) \notin \mathcal{F}\}$ corresponds to the set $\{\omega_{\ell_f+1}, ..., \omega_{\ell_f+\ell_u}\}$ with $\ell_f + \ell_u = N$ (N is the number of states). Then, $p_f = \frac{\ell_f}{N}$ and $p_u = \frac{\ell_u}{N}$. Assume that $\sum_{i=1}^d Z_i$ takes ℓ_u values $\tilde{s}_1 \geqslant \tilde{s}_2 \geqslant ... \geqslant \tilde{s}_{\ell_u}$, and that T then takes ℓ_f values $s_1 \geqslant s_2 \geqslant ... \geqslant s_{\ell_f}$. Specifically, assume $N = 8$, $\ell_f = 3$, $\ell_u = 5$, $s_1 = 8 \geqslant s_2 = 8 \geqslant s_3 = 3$ and $\tilde{s}_1 = 10 \geqslant \tilde{s}_2 = 7 \geqslant \tilde{s}_3 = 4 \geqslant \tilde{s}_4 = 3 \geqslant \tilde{s}_5 = 1$, $p_f = 3/8$ and $p_u = 5/8$. For $p = 5/8$, we apply (B.46) in proposition B.5 and find that $\alpha_* = 0.75$ and that the maximum VaR_p is equal to $TVaR_{0.55}(\sum Z_i) = 8$. These values will be confirmed by the algorithm described below.

B.3.4 Practical Bounds on VaR

To compute the maximum VaR, we present an algorithm that can be applied directly to the matrix M of the observed data, and thus leads to non-parametric bounds on the VaR. Recall that the first ℓ_f rows of matrix M correspond to $\mathcal{F}_N$, whereas ℓ_u denotes the number of rows of $\mathcal{U}_\mathcal{N}$ ($N = \ell_f + \ell_u$). In the algorithm, we also make use of S_N^f and S_N^u, which we treat as random variables. To compute the VaR at probability level p, we define

$$k := N(1-p) \tag{B.47}$$

where we assume that k is an integer.

The algorithm is based on Proposition B.5 and on the following motivation. Recall from the discussion of Proposition B.4 that the stated upper and lower VaR bounds are not sharp in general. Nevertheless, we are able to propose an algorithm to approximate sharp bounds. We explain this idea further. Hence, let $p \in (0,1)$ and let us observe that, almost surely,

$$\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} Z_i \leqslant \mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} H_i.$$

In particular,[22] for all α, β in $[0,1]$ such that $p_f\alpha + (1-p_f)\beta = p$,

$$\max\left\{\mathrm{VaR}_\alpha(T), \mathrm{VaR}_\beta\left(\sum_{i=1}^{d} Z_i\right)\right\} \leqslant M_p = \max\left\{\mathrm{VaR}_{\alpha_*}(T), \mathrm{TVaR}_{\beta_*}\left(\sum_{i=1}^{d} Z_i\right)\right\} \tag{B.48}$$

where α_* is defined as in Proposition B.5 and $\beta_* = \frac{p-p_f\alpha_*}{1-p_f}$. The critical issue is to choose α and β, as well as a dependence between the components of the (comonotonic) vector $(Z_1, Z_2, ..., Z_d)$, such that the inequality (B.48) turns into an equality. Such an equality is clearly obtained when taking $\beta = \beta_*$ (thus $\alpha_* = \alpha$) and a dependence in the vector $(Z_1, Z_2, ..., Z_d)$ such that

$$\mathrm{VaR}_{\beta_*}\left(\sum_{i=1}^{d} Z_i\right) = \mathrm{TVaR}_{\beta_*}\left(\sum_{i=1}^{d} Z_i\right). \tag{B.49}$$

Hence, the best approximation for the sharp bound for $\mathrm{VaR}_p\left(\sum_{i=1}^{d} X_i\right)$ is likely to occur when the quantile (VaR) function of the $\sum_{i=1}^{d} Z_i$ can be made (nearly) flat on $[\beta_*, 1]$. In cases in which this feature cannot be (nearly) obtained, it cannot be excluded that better approximations can be found (for example, if the quantile function $\sum_{i=1}^{d} Z_i$ can be made flat on another interval $[\beta, 1]$ in which β is close to β_*). Similar reasoning shows that in order to reach the stated lower bound as closely as possible one should make the quantile function of the portfolio sum as flat as possible on the interval $[0, \beta_*]$. We build on this idea to propose a practical algorithm to approximate sharp bounds below.

Here, S_N^f plays the role of T and S_N^u plays the role of $\sum_{i=1}^{d} Z_i$ (see also (B.38)). Without loss of generality, assume that S_N^f takes values $s_1 \geqslant s_2 \geqslant ... \geqslant s_{\ell_f}$.

Algorithm for computing the maximum VaR

1. Recall that $p_f = \frac{\ell_f}{N}$. Compute $m_1 := \max\{0, \ell_f - k\}$ (so that $\alpha_1 = \frac{m_1}{\ell_f} = \max\{0, \frac{p+p_f-1}{p_f}\}$) and $m_2 := \min\{\ell_f, N-k\}$ (then $\alpha_2 = \frac{m_2}{\ell_f} = \min\{1, \frac{p}{p_f}\}$).

2. Compute α_* where

$$\alpha_* := \inf\left\{\alpha \in (\alpha_1, \alpha_2) \mid \mathrm{VaR}_\alpha(S_N^f) \geqslant \mathrm{TVaR}_{\frac{p-p_f\alpha}{1-p_f}}(S_N^u)\right\}.$$

3. Apply the RA to the first $\lfloor(1-\beta_*)\ell_u\rfloor$ rows of the untrusted part $\mathcal{U}_\mathcal{N}$ of the matrix M, where $\beta_* = \frac{p-p_f\alpha_*}{1-p_f}$ and where $\lfloor\cdot\rfloor$ denotes the floor of a number. Observe that $\lfloor(1-\beta_*)\ell_u\rfloor = k + m_* - \ell_f$ where $m_* := \lfloor\alpha_*\ell_f\rfloor$ and note that $m_1 \leqslant m_* \leqslant m_2$.

[22] See Bernard and Vanduffel [2015a] for ore details and intuition on the proof of this result.

4. By abuse of notation, denote the rearranged sums in the untrusted part as S_N^u. This is the dependence that potentially achieves the maximum VaR by making $TVaR_{\frac{\ell_u - b_*}{\ell_u}}(S_N^u)$ as close as possible to $VaR_{\frac{\ell_u - b_*}{\ell_u}}(S_N^u)$. To compute this maximum possible VaR, calculate all (row) sums for $\mathcal{U}_\mathcal{N}$ and $\mathcal{F}_\mathcal{N}$ and sort them from maximum to minimum value, $\tilde{s}_1 \geqslant \tilde{s}_2 \geqslant ... \geqslant \tilde{s}_k \geqslant ... \geqslant \tilde{s}_N$. Then, the VaR is $\tilde{s}_k$.

The above algorithm is a quick way to derive potentially attainable bounds for VaR of the aggregate risk. It requires running the rearrangement algorithm only once. However, as the RA will rarely generate a perfectly constant sum on the area where it is applied, it is possible that a better bound might be obtained by applying step 3 to the first $k+m-\ell_f$ rows of the $\mathcal{U}_\mathcal{N}$ for some other m ($m_1 \leqslant m \leqslant m_2$).

Illustration of the algorithm for obtaining the maximum VaR in the example with $d = 3$, $N = 8$, $k = 3$ with the same matrix M given in (B.37) so that $\ell_f = 3$ and $\ell_u = 5$. In this case, $\alpha_* = 0.75$, so the theoretical maximum VaR is equal to $\text{TVaR}_{\frac{p-p_f\alpha_*}{1-p_f}} = 8$. In the algorithm, $m_* = \lfloor \alpha_* \ell_f \rfloor = \lfloor 2.25 \rfloor = 2$ and the maximum VaR is obtained for $m_* = 2$ (that is, by applying the RA to the first $k + m_* - \ell_f = 2$ rows of the untrusted portion of the matrix). By going through all possible values of m, we show below that this is indeed the optimal value.

We find for the minimum and maximum value for m,

$$m_1 = \max(0, 0) = 0, m_2 = \min(3, 8 - 3) = 3,$$

so that the number of rows to which one can consider applying the RA is between 0 and 3, as

$$k + m_1 - \ell_f = 0, k + m_2 - \ell_f = 3.$$

The first VaR that we compute by taking three rows of S_N^u ($m + k - \ell_f = 3$ with $m = 3$) is equal to VaR=7:

$$M = \begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 3 & 3 \\ 3 & 2 & 2 \\ 4 & 1 & 2 \end{bmatrix}, \quad S_N^f = \begin{bmatrix} 8 \\ 8 \\ 3 \end{bmatrix}, \quad S_N^u = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}.$$

The second value is equal to VaR=8 ($m + k - \ell_f = 2$ with $m = 2$) and is already the maximum possible value:

$$M = \begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 3 & 3 \\ 4 & 2 & 2 \end{bmatrix}, \quad S_N^f = \begin{bmatrix} 8 \\ 8 \\ 3 \end{bmatrix}, \quad S_N^u = \begin{bmatrix} 9 \\ 10 \end{bmatrix}.$$

Indeed, one more iteration ($m = 1$) will not change the value of the VaR, and two more iterations ($m = 0$) will lead to a lower number.

The algorithm for computing the minimum VaR is similar to that for the maximum, where TVaR is replaced by LTVaR to compute α_*. Details and an example can be found in Bernard and Vanduffel [2015a],

B.4 Example of Bounds for risk measures of portfolios with dependence uncertainty by Monte Carlo

We discuss how to simulate the analytical bounds for variance, TVaR and VaR (obtained in the respective Propositions B.2, B.3 and B.5 with some examples.

B.4.1 Bounds on Variance

The upper and lower bounds for the variance in Proposition B.2 can be computed by numerical integration or by Monte Carlo simulation. If the number of dimensions d is high then it is clear that the best approach to computing the theoretical bounds is to use Monte Carlo techniques (using simulations from the fitted multidimensional model on $\mathcal{F}$). We illustrate Proposition B.2 with an example. In this respect, it is appropriate to use the standard deviation as the risk measure and not the variance (it is clear that in this instance the bounds are the square roots of those presented in Proposition B.2). Doing so makes it possible to compare fairly the results of this example with those of subsequent examples that use TVaR or VaR as the risk measure.

Example B.6 (multivariate normal distribution as a benchmark model)**.** Assume that $(X_1, ..., X_d)$ is a random vector with standard normally distributed marginals. Furthermore, the joint distribution of $(X_1, ..., X_d)$ is assumed to be a multivariate standard normal distribution with correlation parameter[23] ϱ on the subset $\mathcal{F} := [q_\beta, q_{1-\beta}]^d \subset \mathbb{R}^d$ (for some $\beta < 50\%$), where q_γ denotes the quantile of the standard normal random variable at probability level γ. In Table B.1, we assume that $d = 20$, and we provide the upper and lower bounds for the standard deviation of the portfolio sum for various confidence levels β and correlation levels ρ. The first column ($\beta = 0\%$) provides results for cases in which there is no uncertainty on the multivariate distribution; as such, it provides a benchmark for assessing model risk (see Definition B.1). The last column ($\beta = 50\%$) provides bounds for cases in which there is full uncertainty on the dependence; as such, it corresponds to the situation that is traditionally studied in the literature.

$\mathcal{F} = [q_\beta, q_{1-\beta}]^d$	$\mathcal{U} = \emptyset$ $\beta = 0\%$	$\beta = 0.05\%$	$\beta = 0.5\%$	$\beta = 5\%$	$\mathcal{U} = \mathbb{R}^d$ $\beta = 50\%$
$\rho = 0$	4.47	(4.4 , 5.65)	(3.89 , 10.6)	(1.23 , 19.3)	(0 , 20)
$\rho = 0.1$	7.62	(7.41 , 8.26)	(6.23 , 11.7)	(1.69 , 19.2)	(0 , 20)
$\rho = 0.5$	14.5	(13.8 , 14.6)	(11.1 , 15.4)	(3.74 , 18.6)	(0 , 20)

Table B.1: In the first column we report the standard deviation of $\sum_{i=1}^{20} X_i$ under the assumption of multivariate normality (no dependence uncertainty, i.e., $\mathcal{U} = \emptyset$). Lower and upper bounds of the standard deviation of $\sum_{i=1}^{20} X_i$ are reported as pairs $(\varrho_{\mathcal{F}}^-, \varrho_{\mathcal{F}}^+)$ for various confidence levels β. We use 3,000,000 simulations. All digits reported in the table are significant.

One observes from Table B.1 that the impact of model risk on the standard deviation can be substantial even when the joint distribution $(X_1, ..., X_d)$ is almost perfectly known, i.e., when β is close to zero (p_u is close to 0). Consider for instance $\beta = 0.05\%$

[23] A multivariate standard normal distribution with correlation coefficient ρ is such that the pairwise correlation is ρ for all pairs (X_i, X_j) with $i \neq j$.

and $\rho = 0$. In this case, $p_u = 1 - 0.999^{20} \approx 0.02$, and we find that using a multivariate normal assumption (as the benchmark) might underestimate the standard deviation by (5.65-4.47)/4.47=26.4% and overestimate it by (4.47-4.4)/4.4=1.6%. It thus seems that the assumption of multivariate normality is not particularly robust against misspecification. Here, in fact, it clearly gives rise to a situation in which one is more likely to underestimate risk than to overestimate it. Furthermore, the example shows that adding some partial information on the dependence (i.e., when $\beta < 50\%$) can change the unconstrained bounds (case in which $\beta = 50\%$) and confirms that dependence is important when assessing the risk of a portfolio. For instance, when $\beta = 0.5\%$ and $\rho = 0$, one has that $p_u = 1 - 0.99^{20} \approx 0.18$ and the unconstrained upper bound for the standard deviation shrinks by approximately 50% (from 20 to 10.6).

	$\mathcal{U} = \emptyset$				$\mathcal{U} = \mathbb{R}^d$
p_u	$\beta = 0\%$	$\beta = 0.05\%$	$\beta = 0.5\%$	$\beta = 5\%$	$\beta = 50\%$
$\rho = 0$	0	0.02	0.18	0.88	1
$\rho = 0.1$	0	0.02	0.18	0.87	1
$\rho = 0.5$	0	0.016	0.12	0.66	1

Table B.2: Probability p_u that $(X_1, ..., X_d)$ takes values outside the $d-$cube $[q_\beta, q_{1-\beta}]^d$, for a confidence level β and a correlation coefficient ρ. We use 3,000,000 simulations.

In Table B.2 we report, for the levels of correlation ρ and confidence levels β used in Table B.1, the probability p_u that $(X_1, ..., X_d)$ takes values outside the $d-$cube $\mathcal{F} = [q_\beta, q_{1-\beta}]^d$. Doing so allows us to better interpret the results of Table B.1 and will also be useful in understanding the effect of the choice of another design for the trusted area $\mathcal{F}$.

In the above example, the trusted area is based solely on the use of the marginal densities, $N(0, 1)$. More generally, assume that marginal densities have been fitted to empirical densities $\widehat{f}_i$ for $i = 1, ..., d$, respectively. Then, $\mathcal{F}$ is defined as

$$\mathcal{F} := \left\{ (x_{1i}, ..., x_{di}) \in \mathbb{R}^d \ / \ \forall j \in \{1, 2, ..., d\}, \widehat{f}_j(x_{ji}) \geqslant \varepsilon \right\} \tag{B.50}$$

In the case that the rare events correspond to either the largest or the smallest outcomes of the risks, this approach is consistent with the use of a d-cube as trusted area. Another natural criterion by which to determine the trusted part of the multivariate distribution consists in starting from a given fitted multivariate density $\widehat{f}$ (coming for instance from a multivariate Gaussian model, a multivariate Student model or a Pair-Copula Construction model (Aas et al. [2009], Czado [2010])). The trusted area is then based on the contour levels of the density. We refer to Bernard and Vanduffel [2015a] for more details on this point. These observations are also intuitive, as the standard deviation is sensitive to high outcomes and these scenarios occur frequently when considering the upper bound (as the tail events are then assumed to be fully correlated).

B.4.2 Bounds on TVaR

We now use the same illustrative example for the TVaR and compute the stated bounds using Monte Carlo simulations.

Example B.7 (multivariate normal distribution as a benchmark model)**.** Table B.3 provides for various levels of probability level p, confidence level β, and correlation ρ the

bounds on TVaR. The results are in line with those of the previous example. Model risk is already present for small levels of β, but at the same time the availability of dependence information ($\beta < 50\%$) allows for strengthening the unconstrained bounds ($\beta = 50\%$) significantly. Interestingly, the degree of model risk also depends on the interplay between the probability level p used to assess the TVaR and the degree of uncertainty on the dependence as measured by β. When p is large (e.g., $p = 99.5\%$), a small proportion of model uncertainty (e.g., $\beta = 0.05\%$) appears to have a tremendous effect on the model risk of underestimation. We can explain this observation as follows. The TVaR is essentially measuring the average of all upper VaRs and its level is thus driven mainly by scenarios in which one or more outcomes of the risks involved are high. These scenarios, however, are not considered as trustworthy for depicting the (tail) dependence with negative impact on the level of the TVaR. In fact, for a given level of p the model risk of underestimation increases sharply with an increase in the level of β and approaches its maximum already for small to moderate values of β. This effect is further emphasized when the level of p increases. In other words, the TVaR is highly vulnerable to model misspecification, especially when it is assessed at high probability levels.

		$\mathcal{U} = \emptyset$				$\mathcal{U} = \mathbb{R}^d$
$F = [q_\beta, q_{1-\beta}]^d$		$\beta = 0\%$	$\beta = 0.05\%$	$\beta = 0.5\%$	$\beta = 5\%$	$\beta = 50\%$
$p = 95\%$	$\rho = 0$	9.21	(9.12 , 11.6)	(8.49 , 27.5)	(3.36 , 41.3)	(-0.002 , 41.3)
	$\rho = 0.1$	15.7	(15.4 , 17.3)	(13.5 , 28.4)	(4.72 , 41.3)	(0.004 , 41.3)
	$\rho = 0.5$	29.9	(28.1 , 30.5)	(22.9 , 34.0)	(10.0 , 41.3)	(-0.002 , 41.3)
$p = 99.5\%$	$\rho = 0$	12.9	(12.8 , 30.4)	(12.1 , 57.9)	(7.52 , 57.9)	(-0.004 , 57.9)
	$\rho = 0.1$	22	(21.5 , 33.3)	(19.0 , 57.8)	(10.0 , 57.9)	(-0.002 , 57.9)
	$\rho = 0.5$	42	(37.4 , 47.6)	(29.6 , 57.9)	(15.2 , 57.9)	(0.019 , 57.9)

Table B.3: $\text{TVaR}_{95\%}$ and $\text{TVaR}_{99.5\%}$ of $\sum_{i=1}^{20} X_i$ are reported in the absence of uncertainty (multivariate standard normal model with $\mathcal{U} = \emptyset$). Bounds are then given for various levels of confidence β, correlation ρ and probability p. Bounds are obtained based on 3,000,000 simulations. All digits reported are significant.

Similarly to the case of the standard deviation, one can also use a trusted area that is based on the contours of the multivariate normal distribution in order to assess the upper and lower bounds. As the results are similar, we do not report them in detail.

B.4.3 Bounds on Value-at-Risk

We here assess the VaR bounds when the benchmark model is a multivariate normal distribution.

Example B.8 (Multivariate normal distribution as a benchmark model)**.** The VaR bounds reported in Table B.4 were obtained within a few minutes, using 3,000,000 Monte Carlo simulations. We make the following observations. First, model risk is clearly present even when the dependence is "mostly" known (i.e., β is small). Furthermore, the precise degree of model error depends highly on the level of the probability p that is used to assess the VaR. Let us consider the benchmark model with $\rho = 0$ (the risks are independent and standard normally distributed) and $\beta = 0\%$ (no uncertainty). We find that $\text{VaR}_{95\%}\left(\sum_{i=1}^{20} X_i\right) = \sqrt{20}\Phi^{-1}(95\%) = 7.35$ and, similarly, $\text{VaR}_{99.5\%}\left(\sum_{i=1}^{20} X_i\right) = 11.5$,

$\mathrm{VaR}_{99.95\%}\left(\sum_{i=1}^{20} X_i\right) = 14.7$. However, if $\beta = 0.05\%$, then $p_u \approx 0.02$, and the benchmark model might underestimate the 95%−VaR by (8.08-7.36)/8.08=8.9% or overestimate it by (7.36-7.27)/7.27=1.24%. However, when using the 99.5%−VaR, the degree of underestimation may rise to (30.4-11.5)/30.4=62.2%, whereas the degree of overestimation is equal only to (11.5-11.4)/11.4=0.9%. Hence, the risk of underestimation is sharply increasing in the probability level that is used to assess VaR.

$F = [q_\beta, q_{1-\beta}]^d$		$\mathcal{U} = \emptyset$ $\beta = 0\%$	$\beta = 0.05\%$	$\beta = 0.5\%$	$\beta = 5\%$	$\mathcal{U} = \mathbb{R}^d$ $\beta = 0.5$
p=95%	$\rho = 0$	7.36	(7.27 , 8.08)	(6.65 , 27.5)	(0.79 , 41.3)	(-2.17 , 41.3)
p=95%	$\rho = 0.1$	12.5	(12.2 , 13.3)	(10.7 , 27.7)	(1.51 , 41.2)	(-2.17 , 41.3)
p=95%	$\rho = 0.5$	23.8	(22.9 , 24.2)	(18.9 , 30.9)	(6.97 , 41.2)	(-2.17 , 41.3)
p=99.5%	$\rho = 0$	11.5	(11.4 , 30.4)	(10.8 , 57.8)	(6.13 , 57.8)	(-0.29 , 57.8)
p=99.5%	$\rho = 0.1$	19.6	(19.1 , 31.4)	(16.9 , 57.8)	(8.23 , 57.8)	(-0.29 , 57.8)
p=99.5%	$\rho = 0.5$	37.4	(34.3 , 45.1)	(27.4 , 57.8)	(13.5 , 57.8)	(-0.29 , 57.8)
p=99.95%	$\rho = 0$	14.7	(14.6 , 71.0)	(13.8 , 71.1)	(9.31 , 71.1)	(-0.036 , 71.1)
p=99.95%	$\rho = 0.1$	25.1	(24.2 , 71.1)	(21.5 , 71.1)	(12.1 , 71.1)	(-0.035 , 71.1)
p=99.95%	$\rho = 0.5$	47.7	(41.3 , 71.1)	(32.3 , 71.1)	(17.2 , 71.1)	(-0.036 , 71.1)

Table B.4: $\mathrm{VaR}_{95\%}$, $\mathrm{VaR}_{99.5\%}$ and $\mathrm{VaR}_{99.95\%}$ of $\sum_{i=1}^{20} X_i$ are reported in the absence of uncertainty (multivariate standard normal model with $\mathcal{U} = \emptyset$). Bounds are then given for various levels of confidence β, correlation ρ and probability p. We use $3,000,000$ simulations and all digits reported are significant.

Finally, note that when very high probability levels are used in VaR calculations ($p = 99.95\%$; see the last three rows in Table B.4), the constrained upper bounds are very close to the unconstrained upper bound, even when there is almost no uncertainty on the dependence ($\beta = 0.05\%$). The bounds computed by Embrechts et al. [2013] are thus nearly the best possible bounds, even though it seems that the multivariate model is known at a very high confidence level. This implies that any effort to accurately fit a multivariate model will not reduce the model risk on the risk measure (and the capital requirement).

Note that when no information on the dependence is available ($\beta = 50\%$) the upper and lower bounds stated in proposition B.4 reduce to $\sum_{i=1}^{d} \mathrm{TVaR}_p(X_i)$ and $\sum_{i=1}^{d} \mathrm{LTVaR}_p(X_i)$, respectively, and coincide with the lower bound A and upper bound B, given by Bernard et al. [2013b]. Using their formulas for A and B, we find that the bounds on the VaR_p of sums of 20 independent $\mathcal{N}(0,1)$ risks are

$$A = -20\frac{\phi(\Phi^{-1}(p))}{p}, \quad B = 20\frac{\phi(\Phi^{-1}(p))}{1-p}$$

and we observe that one obtains consistency with the bounds reported in Table B.4. For example, when $p = 95\%$, we find that $(A, B) = (-2.17, 41.25)$, which conforms with the numbers in Table B.4.

Example B.9 (Pareto distributed risks)**.** We provide another example with Pareto distributed risks and we find that the same results can be found. They are even amplified. We assume the individual risks are all Pareto with parameter $\theta = 3$. We assume that there are $d = 20$ risks distributed as a Pareto with parameter θ and that their dependence

is modeled by a Gaussian copula with parameter ρ (pairwise correlation). Each risk has the cdf for $x > 0$,

$$F(x) = 1 - (1 + x)^{-\theta}$$

The Value-at-Risk at level $p \in (0, 1)$ is given by

$$F_X^{-1}(p) = (1 - p)^{-1/\theta} - 1$$

Assuming that the trusted zone is based on each marginal being between the quantile of level $1 - \beta$ and β respectively

$$\mathcal{F} = \bigcap_{k=1}^{d} \{q_\beta \leqslant X_k \leqslant q_{1-\beta}\}$$

where

$$q_\beta = (1 - \beta)^{-1/\theta} - 1$$

We find from the simulation of $n = 3,000,000$ simulations of a $d = 20$ Pareto variables the following results for the bounds on Value-at-Risk.

		$\mathcal{U} = \emptyset$				$\mathcal{U} = \mathbb{R}^d$
$F = [q_\beta, q_{1-\beta}]^d$		$\beta = 0\%$	$\beta = 0.05\%$	$\beta = 0.5\%$	$\beta = 5\%$	$\beta = 0.5$
α=95%	$\rho = 0$	16.6	(16 , 18.4)	(13.8 , 37.4)	(8.62 , 61.4)	(7.29 , 61.4)
α=95%	$\rho = 0.1$	19.7	(18.3 , 20.6)	(15.9 , 37.8)	(8.82 , 61.4)	(7.29 , 61.4)
α=95%	$\rho = 0.5$	28	(26.5 , 33.5)	(20.6 , 43.2)	(10.3 , 61.4)	(7.29 , 61.4)
α=99.5%	$\rho = 0$	25.8	(21.5 , 60.7)	(17.5 , 156)	(10.7 , 156)	(9.27 , 156)
α=99.5%	$\rho = 0.1$	32.5	(27.9 , 63)	(21.8 , 156)	(11.6 , 156)	(9.27 , 156)
α=99.5%	$\rho = 0.5$	61.1	(49 , 94.7)	(31.6 , 155)	(14 , 155)	(9.26 , 155)
α=99.95%	$\rho = 0$	43.5	(26.5 , 359)	(20.5 , 360)	(12.4 , 360)	(9.83 , 359)
α=99.95%	$\rho = 0.1$	51.9	(36.3 , 357)	(26.8 , 359)	(13.9 , 358)	(9.82 , 357)
α=99.95%	$\rho = 0.5$	116	(69.6 , 361)	(40.1 , 361)	(16.8 , 359)	(9.83 , 359)

B.4.4 Further discussion on model risk

Let us consider again a random vector $(X_1, ..., X_d)$ having standard normally distributed marginals all correlated with a coefficient of 10% (benchmark). We now focus on the model risk for underestimation and overestimation; that is, we consider the quantities

$$\frac{\varrho_{\mathcal{F}}^+ - \varrho(\sum_i X_i)}{\varrho_{\mathcal{F}}^+} \text{ and } \frac{\varrho_{\mathcal{F}}^- - \varrho(\sum_i X_i)}{\varrho_{\mathcal{F}}^-}, \tag{B.51}$$

which were introduced in Section B.4 (Definition B.1 and expressions (B.28) and (B.29)). The risk measure $\varrho(\cdot)$ is the VaR and the TVaR, and, for the trusted area $\mathcal{F}$, we consider the elliptical contours such that $P((X_1, ..., X_d) \in \mathcal{F}) = p_f$.

In Figure B.3, we represent the risk of underestimating and overestimating VaR and TVaR, respectively, at various probability levels p using the risk measures (B.51) for model risk. From Figure B.3, we observe that a slight misspecification of the model already leads to a potentially significant underestimation of VaR and TVaR. By contrast, the risk of overestimating appears to be less pronounced. We can explain these observations as follows. In the benchmark model, the risks X_i $(i = 1, 2, ..., d)$ are assumed to be

multivariate normally distributed, with a correlation coefficient of only 10%. However, in the presence of uncertainty the risks are assumed to be fully dependent in the untrusted area $\mathcal{U}$ when calculating the upper bound on TVaR and to behave as a constant when calculating the lower bound (the portfolio sum is thus also constant in this instance). The latter situation is closer to the one that is present in the benchmark model, and therefore the risk of overestimating TVaR is relatively small. Conversely, the risk of underestimation is rather significant. The same pattern holds true for the bounds on VaR, as these are based on the TVaR of a comonotonic sum and thus differ from the VaR under the benchmark model (risks have low correlation).

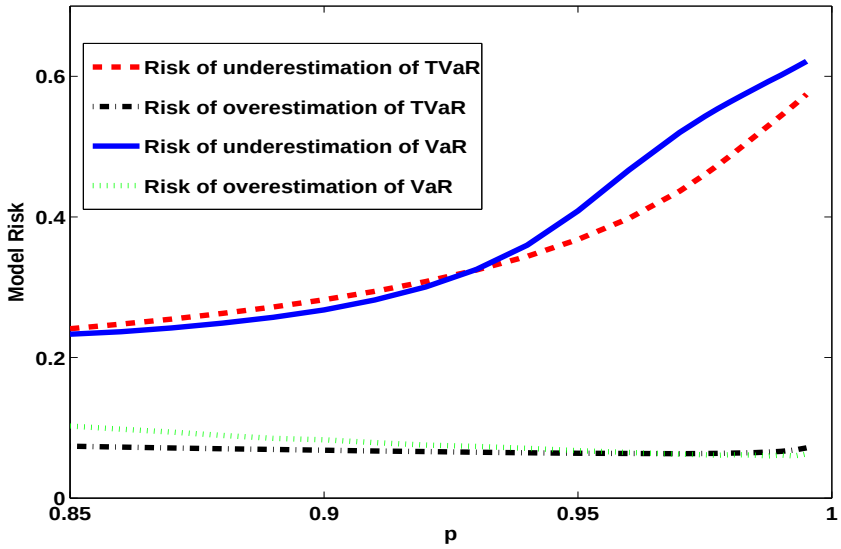


Figure B.3: We assume that $(X_1, X_2, ...X_{20})$ is a multivariate standard distribution with pair correlation $\rho = 0.1$. Let $p_f = 90\%$. We show the model risk for overestimating or underestimating VaR_p and TVaR_p as a function of p.

We also observe that when the probability level p is high, the model risk of underestimating VaR appears to be larger than the model risk of underestimating TVaR. We can explain this remarkable feature as follows. For p sufficiently large (as compared to $1 - p_u$), the worst VaR and the worst TVaR of the portfolio sum are both based on the untrusted scenarios, as the very largest outcomes for the portfolio sum usually correspond to the untrusted scenarios. Hence, in this case, the worst VaR and the worst TVaR tend to be close to each other.[24] However, the difference between the modeled TVaR and the modeled VaR, naturally, remains strictly positive. The two effects together imply that the model risk of underestimating VaR is more significant than the risk of underestimating TVaR when p is very large. The example thus suggests that VaR is more sensitive to model risk than TVaR. It also illustrates that a model may provide a good fit for the data on the whole but still not be suitable for estimating VaR at high probability levels.

[24]Puccetti and Rüschendorf (2012a) show that under mild conditions that for a given set of scenarios the worst Value-at-Risk behaves asymptotically as the worst Tail Value-at-Risk (TVaR).

References

Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

Pauline Barrieu and Giacomo Scandolo. Assessing financial model risk. *European Journal of Operational Research*, 242(2):546–556, 2015.

Basel Committee. Developments in modelling risk aggregation. *Joint Forum, Bank for International Settlements*, 2010a.

Basel Committee. Basel iii: A global regulatory framework for more resilient banks and banking systems. *preprint*, 2010b.

Basel Committee. Fundamental review of the trading book. *Consultative Document, May*, 2012.

Basel Committee. Fundamental review of the trading book: A revised market risk framework. *Consultative Document, October*, 2013.

Nicole Bäuerle and Alfred Müller. Stochastic orders and risk measures: Consistency and bounds. *Insurance: Mathematics and Economics*, 38(1):132–148, 2006.

Tim Bedford and Roger M Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268, 2001.

Tim Bedford and Roger M Cooke. Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

Carole Bernard and Steven Vanduffel. A new approach to assessing model risk in high dimensions. *Journal of Banking and Finance*, 58:166–178, 2015a.

Carole Bernard and Steven Vanduffel. Quantile of a mixture. *Dependence Modeling, forthcoming*, 2015b.

Carole Bernard, Xiao Jiang, and Steven Vanduffel. A note on 'improved fréchet bounds and model-free pricing of multi-asset options' by Tankov (2011). *Journal of Applied Probability*, 49(3):866–875, 2012.

Carole Bernard, Yuntao Liu, Niall MacGillivray, and Jinyuan Zhang. Bounds on capital requirements for bivariate risk with given marginals and partial information on the dependence. *Dependence Modeling*, 1:37–53, 2013a.

Carole Bernard, Ludger Rüschendorf, and Steven Vanduffel. VaR bounds with a variance constraint. *Journal of Risk and Insurance, forthcoming*, 2013b.

Carole Bernard, Phelim Boyle, and Steven Vanduffel. Explicit representation of cost-efficient strategies. *Finance*, 35(2):5–55, 2014.

Carole Bernard, Jit-Seng Chen, and Steven Vanduffel. Optimal portfolios under worst case scenarios. *Quantitative Finance*, 14(4):657–671, 2014.

Carole Bernard, Michel Denuit, and Steven Vanduffel. VaR bounds with high-order moments. *working paper*, 2014a.

Carole Bernard, Xiao Jiang, and Ruodu Wang. Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, 54(1):93–108, 2014b.

Carole Bernard, Ludger Rüschendorf, Steven Vanduffel, and Jing Yao. How robust is the Value-at-Risk of credit risk portfolios? *European Journal of Finance, forthcoming*, 2015.

Andrew Chernih, Luc Henrard, and Steven Vanduffel. Reconciling credit correlations. *The Journal of Risk Model Validation*, 4(2):47–64, 2010.

Ka-Chun Cheung and Steven Vanduffel. Bounds for sums of random variables when the marginal distributions and the variance of the sum are given. *Scandinavian Journal*, 13(2):103–118, 2013.

Roger Cooke and Carolyn Kousky. The limits of securitisation: Micro-correlations, fat tails, and tail dependence. *Rethinking Risk Measurement and Reporting Volume I: Uncertainty, Bayesian Analysis, and Expert Judgement, London: Risk Books*, pages 273–294, 2010.

Roger M Cooke, Carolyn Kousky, and Harry Joe. Micro correlations and tail dependence. *Dependence Modeling: Vine Copula Handbook, Chapter 5*, pages 89–112, 2011.

Claudia Czado. Pair-copula constructions of multivariate copulas. In *Copula Theory and Its Applications*, pages 93–109. Springer, 2010.

Michel Denuit, Christian Genest, and Étienne Marceau. Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25(1):85–104, 1999.

Michel Denuit, Jan Dhaene, Marc Goovaerts, and Rob Kaas. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley Online Library, 2005.

Jan Dhaene, Michel Denuit, Marc Goovaerts, Rob Kaas, and David Vyncke. The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, 31(1):3–33, 2002.

Jan Dhaene, Steven Vanduffel, Marc Goovaerts, Rob Kaas, Qihe Tang, and David Vyncke. Risk measures and comonotonicity: a review. *Stochastic Models*, 22:573–606, 2006.

Jan Dhaene, Roger JA Laeven, Steven Vanduffel, Gregory Darkiewicz, and Marc J Goovaerts. Can a coherent risk measure be too subadditive? *Journal of Risk and Insurance*, 75(2):365–386, 2008.

Jan Dhaene, Michel Denuit, and Steven Vanduffel. Correlation order, merging and diversification. *Insurance: Mathematics and Economics*, 45(3):325–332, 2009.

Jan Dhaene, Andreas Tsanakas, Emiliano A Valdez, and Steven Vanduffel. Optimal capital allocation principles. *Journal of Risk and Insurance*, 79(1):1–28, 2012.

Darrell Duffie and Kenneth J Singleton. *Credit risk: pricing, measurement, and management*. Princeton University Press, 2012.

Paul Embrechts and Marius Hofert. Risk measures and dependence modeling. In *Handbook of Insurance*, pages 135–165. Springer, 2013.

Paul Embrechts and Marius Hofert. Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application*, 1:493–514, 2014.

Paul Embrechts and Giovanni Puccetti. Bounds for functions of dependent risks. *Finance and Stochastics*, 10(3):341–352, 2006.

Paul Embrechts and Giovanni Puccetti. Risk aggregation. In *Copula Theory and Its Applications*, pages 111–126. Springer, 2010a.

Paul Embrechts and Giovanni Puccetti. Bounds for the sum of dependent risks having overlapping marginals. *Journal of Multivariate Analysis*, 101(1):177–190, 2010b.

Paul Embrechts, Alexander McNeil, and Daniel Straumann. Correlation: pitfalls and alternatives. *Risk London, Risk Magazine*, 12:69–71, 1999.

Paul Embrechts, Alexander McNeil, and Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: Value-at-Risk and beyond, M. Dempster & H. Moffatt, eds*, pages 176–223, 2002.

Paul Embrechts, Giovanni Puccetti, and Ludger Rüschendorf. Model uncertainty and VaR aggregation. *Journal of Banking & Finance*, 37(8):2750–2764, 2013.

Paul Embrechts, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. An academic response to basel 3.5. *Preprint*, 2014.

H. Föllmer and A. Schied. Coherent and convex risk measures. *Encyclopedia of Quantitative Finance, Cont, R. (ed.) New York: John Wiley & Sons*, pages 355–363, 2010.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

Chris C Heyde and Steven G Kou. On the controversy over tailweight of distributions. *Operations Research Letters*, 32(5):399–408, 2004.

Marius Hofert. A stochastic representation and sampling algorithm for nested archimedean copulas. *Journal of Statistical Computation and Simulation*, 82(9):1239–1255, 2012.

Marius Hofert and David Pham. Densities of nested archimedean copulas. *Journal of Multivariate Analysis*, 118:37–52, 2013.

Harry Joe. *Multivariate models and multivariate dependence concepts*, volume 73. CRC Press, 1997.

Philippe Jorion. *Value-at-Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill. Third edition, 2006.

Steven Kou, Xianhua Peng, and Heyde Chris C. External risk measures and basel accords. *Mathematics Operational Research*, 38(3):393–417, 2013.

Carolyn Kousky and Roger Cooke. Explaining the failure to insure catastrophic risks. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 37(2):206–227, 2012.

Carolyn Kousky and Roger M Cooke. The unholy trinity: fat tails, tail dependence, and micro-correlations. *Resources for the Future Discussion Paper*, pages 09–36, 2009.

Dorota Kurowicka and Roger M Cooke. *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons, 2006.

Laurence Lescourret and Christian Y Robert. Extreme dependence of multivariate catastrophic losses. *Scandinavian Actuarial Journal*, 2006(4):203–225, 2006.

Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques, and tools*. Princeton university press, 2010.

Isaac Meilijson and Arthur Nádas. Convex majorization with an application to the length of critical paths. *Journal of Applied Probability*, pages 671–677, 1979.

Alfred Müller and Dietrich Stoyan. Comparison methods for stochastic models and risks. *John Wiley&Sons Ltd., Chichester*, 2002.

Roger B Nelsen. *An introduction to copulas*. Springer, 2007.

Ostap Okhrin, Yarema Okhrin, and Wolfgang Schmid. On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics*, 173(2):189–204, 2013.

OSFI. Quantitative Impact Study no. 4: General - aggregation and diversification - supplementary information. *Online document*, 2014.

Harry H Panjer. *Measurement of Risk, Solvency Requirements and Allocation of Capital Within Financial Conglomerates*. University of Waterloo, Institute of Insurance and Pension Research, 2001.

Giovanni Puccetti and Ludger Rüschendorf. Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, 236(7):1833–1840, 2012a.

Giovanni Puccetti and Ludger Rüschendorf. Asymptotic equivalence of conservative VaR- and ES-based capital charges. *Working paper*, 2012b.

Ludger Rüschendorf. Random variables with maximum sums. *Advances in Applied Probability*, pages 623–632, 1982.

Ludger Rüschendorf. Bounds for distributions with multivariate marginals. *Lecture Notes-Monograph Series*, pages 285–310, 1991.

Cornelia Savu and Mark Trede. Hierarchies of archimedean copulas. *Quantitative Finance*, 10(3):295–304, 2010.

Abe Sklar. *Fonctions de répartition à n dimensions et leurs marges*, volume 8. 1959.

Peter Tankov. Improved Fréchet bounds and model-free pricing of multi-asset options. *Journal of Applied Probability*, 48(2):389–403, 2011.

Andreas Tsanakas. To split or not to split: Capital allocation with convex risk measures. *Insurance: Mathematics and Economics*, 44(2):268–277, 2009.

Andreas Tsanakas and Evangelia Desli. Measurement and pricing of risk in insurance markets. *Risk analysis*, 25(6):1653–1668, 2005.

Bin Wang and Ruodu Wang. The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360, 2011.

Ruodu Wang, Liang Peng, and Jingping Yang. Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, 17(2):395–417, 2013.

Shaun Wang. A class of distortion operations for pricing financial and insurance risks. *Journal of Risk and Insurance*, 67(10):15–36, 2000.

Johanna F Ziegel. Coherence and elicitability. *Mathematical Finance*, 2014.