SOCIETY OF ACTUARIES

2014 Enterprise Risk Management Symposium
Sept. 29 - Oct. 1, 2014, Chicago, IL

A New Approach to Assessing Model Risk in High Dimensions

By Carole Bernard* and Steven Vanduffel†‡

Actuaries
Risk is Opportunity.℠

A New Approach to Assessing Model Risk in High Dimensions

Carole Bernard* and Steven Vanduffel†‡

March 3, 2015

Abstract

A central problem for regulators and risk managers concerns the risk assessment of an aggregate portfolio defined as the sum of d individual dependent risks X_i. This problem is mainly a numerical issue once the joint distribution of $(X_1, X_2, \ldots, X_d)$ is fully specified. Unfortunately, while the marginal distributions of the risks X_i are often known, their interaction (dependence) is usually either unknown or only partially known, implying that any computed risk measure of the portfolio is subject to model uncertainty.

Previous academic research has focused on the maximum and minimum possible values of a given risk measure of the portfolio, in the case in which only the marginal distributions are known. This approach leads to wide bounds, as all information on the dependence is ignored.

In this paper, we integrate in a natural way available information on the multivariate dependence and provide easy-to-compute bounds for the risk measure at hand. We observe that incorporating the information of a well-fitted multivariate model may, or may not, lead to much tighter bounds, a feature that also depends on the risk measure used. We illustrate this point by showing that the Value-at-Risk at a very high confidence level (as used in Basel II) is typically prone to very high model risk, even if one knows the multivariate distribution almost completely.

Our results make it possible to determine which risk measures can benefit from adding dependence information (i.e., leading to narrower bounds when used to assess portfolio risk), and, hence, to identify those models for which it would be meaningful to develop accurate multivariate models.

Key-words: Model risk, VaR, TVaR, variance, tail dependence, tail correlation.

*Carole Bernard, Department of Statistics and Actuarial Science at the University of Waterloo (email: c3bernar@uwaterloo.ca).

†Corresponding author : Steven Vanduffel, Department of Economics and Political Sciences at Vrije Universiteit Brussel (VUB) (e-mail: steven.vanduffel@vub.ac.be).

‡C. Bernard gratefully acknowledges support from the Natural Sciences and Engineering Research Council of Canada, the Humboldt Research Foundation and the hospitality of the chair of mathematical statistics of Technische Universität München where the paper was completed. S. Vanduffel acknowledges the financial support of the BNP Paribas Fortis Chair in Banking.

1 Introduction

The risk assessment of high dimensional portfolios $(X_1, X_2, \ldots, X_d)$ is a core issue in the regulation of financial institutions and in quantitative risk management. In this regard, one usually attempts to measure the risk of the aggregate portfolio (defined as the sum of individual risks X_i) using a risk measure (such as the variance or the Value-at-Risk (VaR)). It is clear that solving this problem is mainly a numerical issue once the joint distribution of $(X_1, X_2, \ldots, X_d)$ is completely specified. Unfortunately, estimating a multivariate distribution is a difficult task, and thus the assessment of portfolio risk is prone to model misspecification (model risk). At present, there is no generally accepted framework for quantifying model risk. A natural way to do so consists in finding the minimum and maximum possible values of a chosen risk measure evaluated in a family of candidate models. For example, Cont (2006) found bounds on prices of contingent claims, incorporating model risk on the choice of the risk neutral measure used for pricing. In the same spirit, Kerkhof, Melenberg, and Schumacher (2010) assess model risk in the context of management of market risk by computing the worst-case VaR across a range of models chosen based on econometric estimates involving past and present data. A related but different approach, moreover, can be found in Alexander and Sarabia (2012). These authors compare VaR estimates of the model actually used with those of a benchmark model (i.e., the regulatory model) and use the observed deviations to estimate a capital charge supplement to cover for VaR model risk.

In a recent paper, Embrechts, Puccetti, and Rüschendorf (2013) find bounds on the VaR of high dimensional portfolios, assuming that marginal distributions of the individual risks are known (or prone to negligible model risk) and that the dependence structure (also called the copula) among the risks is not specified or only partially specified. This assumption is natural, as fitting the marginal distribution of a single risk X_i $(i = 1, 2, ..., d)$ can often be performed in a relatively accurate manner, whereas fitting a multivariate model for $(X_1, X_2, \ldots, X_d)$ is challenging, even when the number of observations is large. The bounds derived by Embrechts, Puccetti, and Rüschendorf (2013) are wide, as they neglect all information on the interaction among the individual risks. In this paper, we propose to integrate in a natural way the information from a fitted multivariate model.

Standard approaches to estimating a multivariate distribution use a multivariate Gaussian distribution or a multivariate Student's 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, Frey, and Embrechts (2010). There is recent literature dealing with the development of flexible multivariate models that allow a much better fit to the data. 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 inherent in risk assessments.

In this paper, we develop a framework that allows practical quantification of model risk. Our 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. In particular, 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, Puccetti, and Rüschendorf (2013), 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, Puccetti, and Rüschendorf (2013).

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, Jiang, and Vanduffel (2012) and Bernard, Liu, MacGillivray, and Zhang (2013). Tankov (2011) uses extreme dependence scenarios to find model-free bounds for the prices of some bivariate derivatives, whereas Bernard, Boyle, and Vanduffel (2014) and Bernard, Chen, and Vanduffel (2014) 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, portfolio risk management 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, Jiang, and Wang (2014) for illustrations of this feature.

Hence, in this paper 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. There are few papers in the literature that deal explicitly with the presence of (partial) information on the dependence structure: Rüschendorf (1991), Embrechts and Puccetti (2010) and Embrechts, Puccetti, and Rüschendorf (2013) consider the situation in which some of the bivariate distributions are known, Denuit, Genest, and Marceau (1999) study VaR bounds assuming that the joint distribution of the risks is bounded by some distribution and Bernard, Rüschendorf, and Vanduffel (2013) compute VaR bounds when the variance of the sum is known. However, the setup in these papers is often harder 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, Puccetti, and Rüschendorf (2013); Wang, Peng, and Yang (2013) and Bernard, Jiang, and Wang (2014) for related results.

We propose two methods for deriving bounds on risk measures. The first method is presented in Section 2 and provides bounds that can be computed directly (using, for instance, Monte Carlo simulations) but that may not be sharp. The second method (Section 3) is more practical as we provide an algorithm to approximate the sharp bounds, which 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 (2012) and further studied by Embrechts, Puccetti, and Rüschendorf (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 illustrate the theoretical results by various examples and apply both approaches to assessing the model risk of a portfolio of financial indexes. Our numerical results indicate that in high dimensions the bounds computed using the direct method in Section 2 are close to the non-parametric bounds as computed in Section 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 2 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.

2 Bounds for Risk Measures of Portfolios with Dependence Uncertainty

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 this paper, 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 in this paper the following definition of model risk. This definition is in the same spirit as in Barrieu and Scandolo (2013).

Definition 2.1 (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 $\rho(\cdot)$ to assess the risk of $\sum_{i=1}^{d} X_i$. Define

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

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 $\rho(\sum X_i)$ in some chosen benchmark model (i.e., when some multivariate distribution for $(X_1, \ldots, X_d)$ has been specified) is defined as

$$\frac{\rho_{\mathcal{F}}^{+} - \rho(\sum_{i=1}^{n} X_i)}{\rho_{\mathcal{F}}^{+}} \tag{1}$$

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

$$\frac{\rho(\sum_{i=1}^{n} X_i) - \rho_{\mathcal{F}}^{-}}{\rho_{\mathcal{F}}^{-}}. \tag{2}$$

The rest of the paper aims at obtaining the maximum and minimum possible values $\rho_{\mathcal{F}}^{+}$ and $\rho_{\mathcal{F}}^{-}$ of $\rho(\sum_{i=1}^{d} X_i)$. 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 $\mathcal{F} = \emptyset$. See for example Embrechts, Puccetti, and Rüschendorf (2013) with respect to VaR and Bernard, Jiang, and Wang (2014) regarding a convex risk measure. In this paper 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 "$(X_1, X_2, ..., X_d) \in \mathcal{F}$"

$$\mathbb{I} := \mathbb{1}_{(X_1, X_2, ..., X_d) \in \mathcal{F}} \tag{3}$$

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

$$p_f = P(\mathbb{I} = 1) \text{ and } p_u = P(\mathbb{I} = 0). \tag{4}$$

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

$$Z_i = F^{-1}_{X_i|(X_1, X_2, ..., X_d) \in \mathcal{U}}(U), \qquad i = 1, 2, ..., d, \tag{5}$$

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

$(i), (ii)$ and (iii)). 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

$$T := F^{-1}_{\sum_i X_i | (X_1, X_2, ..., X_d) \in \mathcal{F}}(U). \tag{6}$$

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 Tail Value-at-Risk (TVaR). Let us recall their definitions. For $p \in (0, 1)$, we denote by $\mathrm{VaR}_p(X)$ the VaR of X at level p, or a quantile function at level p,

$$\mathrm{VaR}_p(X) = F^{-1}_X(p) = \inf\{x \in \mathbb{R} \mid F_X(x) \geqslant p\}. \tag{7}$$

By convention, $\inf\{\emptyset\} = \infty$ and $\inf\{\mathbb{R}\} = -\infty$, so that $\mathrm{VaR}_p(X)$ is properly defined by (7) for all $p \in [0, 1]$. Furthermore, $\mathrm{TVaR}_p(X)$ denotes the Tail Value-at-Risk (TVaR) at level p

$$\mathrm{TVaR}_p(X) = \frac{1}{1-p} \int_p^1 \mathrm{VaR}_u(X) \mathrm{d}u, \qquad p \in (0, 1).$$

Observe that $p \to \mathrm{TVaR}_p$ is continuous. We define $\mathrm{TVaR}_1(X) = \lim_{p \nearrow 1} \mathrm{TVaR}_p(X)$. TVaR_p is a weighted average of all upper VaRs from probability level p onwards. Similarly, we define the left Tail Value-at-Risk (LTVaR) at level p as $\mathrm{LTVaR}_p(X) = \frac{1}{p} \int_0^p \mathrm{VaR}_u(X) \mathrm{d}u$ and $\mathrm{LTVaR}_0(X) = \mathrm{LTVaR}_{0^+}(X)$.

2.1 Bounds on Variance

In this section, we derive easy-to-compute upper and lower bounds for commonly used risk measures of a portfolio sum $\sum_{i=1}^d X_i$. We start with the variance.

Proposition 2.2 (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 (3) and (5). 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 2.2 is provided in Appendix A.

The stated upper and lower bounds in Proposition 2.2 are intuitive. 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.

The upper and lower bounds for the variance in Proposition 2.2 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 2.2 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 2.2). 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 2.3 (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[1] 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 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 2.1). 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.

One observes from Table 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 $\beta = 0.05\%$ 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).

$\mathcal{F} = [q_\beta, q_{1-\beta}]^d$	$\mathcal{U} = \emptyset$ $\beta = 0\%$	$\beta = 0.05\%$	$\beta = 0.5\%$	$\beta = 5\%$	$\mathcal{U} = \mathbb{R}^d$ $\beta = 50\%$
$\rho = 0$	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 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 $(\rho_{\mathcal{F}}^-, \rho_{\mathcal{F}}^+)$ for various confidence levels β. We use 3,000,000 simulations. All digits reported in the table are significant.

[1]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 $i \neq j$.

In Table 2 we report, for the levels of correlation ρ and confidence levels β used in Table 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 1 and will also be useful in understanding the effect of the choice of another design for the trusted area $\mathcal{F}$.

	$\mathcal{U} = \emptyset$				$\mathcal{U} = \mathbb{R}^d$
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 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 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

$$\mathcal{F} := \left\{ (x_{1i}, ..., x_{di}) \in \mathbb{R}^d \ / \ \forall j \in \{1, 2, ..., d\}, \widehat{f_j}(x_{ji}) \geqslant \varepsilon \right\}. \tag{8}$$

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, Czado, Frigessi, and Bakken (2009); Czado (2010))). The trusted area is then based on the contour levels of the density,

$$\mathcal{F} := \left\{ (x_{1i}, ..., x_{di}) \in \mathbb{R}^d \ / \ \widehat{f}(x_{1i}, ..., x_{di}) \geqslant \varepsilon \right\}. \tag{9}$$

The above example amounts to considering a trusted area $\mathcal{F}$ defined using the contours of the multivariate normal distribution (which are ellipsoids). Let us denote by $\chi_2(p_f, d)$ the quantile at level p_f of a χ^2 distribution with d degrees of freedom. It is then well-known that $P((X_1, X_2, ..., X_d) \in \mathcal{F}) = p_f$ where $\mathcal{F}$ is the $d-$dimensional elliptical disk given as

$$\mathcal{F} = \mathcal{C}_{p_f} := \left\{ \mathbf{X} := (x_1, x_2, ..., x_d) \ / \ (\mathbf{X} - \mu)^t \Sigma^{-1} (\mathbf{X} - \mu) < \chi_2(p_f, d) \right\}. \tag{10}$$

We computed bounds as in Table 1, but we do now for $\mathcal{F}$ determined by the contour of level p_u (varying p_u from 0 to 1 instead of varying β from 0 to 1). We do not report these results in a table, but instead draw the bounds in Figure 1.On the x-axis, we plot the probability p_u that $P((X_1, X_2, ..., X_d) \in \mathcal{U})$. Full trust is obtained with $p_u = 0$, in which case there is no model risk. The graph shows that the model risk for an underestimation of the standard deviation is more pronounced than for an overestimation. Note, indeed, that the distance between the estimate (respectively 4.47 for $\rho = 0$ and 7.62 for $\rho = 0.1$) and the upper bound $\rho^+_{\mathcal{F}}$ is larger than the distance between the estimate and the lower bound $\rho^-_{\mathcal{F}}$. 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). All in all, the numerical results are very similar to those obtained in Table 1, and there is little difference whether one uses a $d-$cube or a $d-$dimensional elliptical area as trusted area.

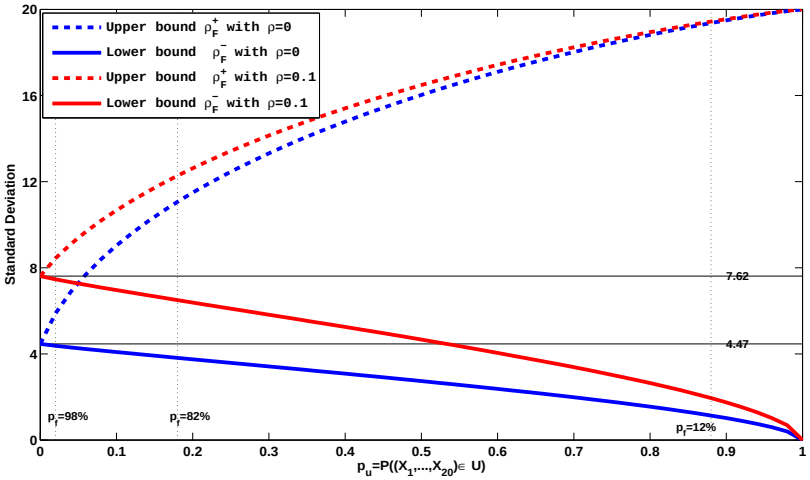


Figure 1: Bounds on standard deviation when the trusted area is $\mathcal{C}_{p_f}$ as a function of p_u and for the correlation coefficients $\rho = 0$ (blue curves) and $\rho = 0.1$ (red curves).

2.2 Bounds on TVaR

Next, we discuss bounds for the TVaR.

Proposition 2.4 (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 (3) and (5). 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 $\mathcal{U} = \emptyset$. We now use the same illustrative example for the variance and compute the stated bounds using Monte Carlo simulations.

Example 2.5 (multivariate normal distribution as a benchmark model)**.** Table 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.

		$\mathcal{U}=\emptyset$				$\mathcal{U}=\mathbb{R}^d$
$F=[q_\beta, q_{1-\beta}]^d$		$\beta=0\%$	$\beta=0.05\%$	$\beta=0.5\%$	$\beta=5\%$	$\beta=50\%$
	$\rho=0$	9.21	(9.12 , 11.6)	(8.49 , 27.5)	(3.36 , 41.3)	(-0.002 , 41.3)
$p=95\%$	$\rho=0.1$	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)
	$\rho=0$	12.9	(12.8 , 30.4)	(12.1 , 57.9)	(7.52 , 57.9)	(-0.004 , 57.9)
$p=99.5\%$	$\rho=0.1$	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 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.

Remark 2.6 (Generalization). Note that the upper and lower bounds for the variance (standard deviation) and the TVaR display similar forms (Propositions 2.2 and 2.4, remark 2.7 when $d=2$). In fact, both risk measures are consistent with the so-called convex order, and one can show that for all risk measures ρ that are consistent with this order the bounds have the same structure. In particular, this feature holds for all concave distortion risk measures. Further details are provided in the Appendix, in the proofs of Propositions 2.2 and 2.4.

Remark 2.7 (Bounds on copulas). Propositions 2.2 and 2.4 are stated in the general case of a d-dimensional problem. We are able, in Proposition 2.2, to exhibit the dependence structure such that the upper bound is sharp. For the lower bound, given that the lower Fréchet bound exists for $d=2$, it is straightforward to construct a sharp bound by using the antimonotonic dependence. In fact, for all risk measures ρ that are consistent with convex order,

$$\rho\left(\mathbb{I}(X_1+X_2)+(1-\mathbb{I})(A_1+A_2)\right) \leqslant \rho\left(X_1+X_2\right) \leqslant \rho\left(\mathbb{I}(X_1+X_2)+(1-\mathbb{I})(Z_1+Z_2)\right),$$

where $A_1=Z_1$ and $A_2=F^{-1}_{X_2|(X_1,X_2)\in\mathcal{U}}(1-U)$ for the same U as defined earlier. Hence, for $d=2$ the lower bound in Proposition 2.2 is sharp if and only if A_1+A_2 is constant, which is not usually the case. However, we will show further in this paper that in high dimensions the lower bounds can be expected to be nearly sharp. Our approach is connected to the work of Tankov (2011). This author provides (for the case in which

$d = 2$) the pointwise minimum and maximum over all dependence structures (copulas) such that the joint distribution is fixed on some compact $\mathcal{F}$ of $[0,1]^2$. Generally, the bounds obtained from Tankov (2011) are not copulas but quasi-copulas, and thus they lead to non-sharp bounds for $\rho(X_1 + X_2)$. Our setting does not require assumptions regarding the compactness of $\mathcal{F}$ and generates sharp upper bounds and nearly sharp lower ones. Importantly, the results in Tankov (2011) and extensions of Bernard, Jiang, and Vanduffel (2012) and Bernard, Liu, MacGillivray, and Zhang (2013) are restricted to $d = 2$, whereas our approach holds for general d, making it suitable for application in portfolio risk management.

2.3 Bounds on VaR

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

Proposition 2.8 (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 (3) and (5). 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 VaR of the aggregate risk are given as

$$m_p \leqslant VaR_p\left(\sum_{i=1}^{d} X_i\right) \leqslant M_p. \tag{11}$$

The proof is provided in Appendix A.3. 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 2.2 and 2.4, 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 2.8 has a similar form[2] to that of Proposition 2.4 resp. 2.2, but the bounds proposed are usually not sharp.[3] We observe that in the case of no uncertainty (i.e., $\mathcal{U} = \emptyset$) there is 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 Theorem 2.1 of Bernard, Rüschendorf, and Vanduffel (2013). 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. To this end, we make use of a lemma that is of interest on its own. In particular, it provides the inspiration for designing, in Section 3, an algorithm that makes it possible to approximate *sharp* upper and lower bounds for the portfolio VaR (see Section 3.1).

[2]Note that VaR is not consistent with convex order, although there are some connections (see Bernard, Rüschendorf, and Vanduffel (2013) and Bernard, Jiang, and Vanduffel (2012)). Hence Remark 2.6 does not apply.

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

Lemma 2.9 (Computing VaR)**.** *Consider a sum* $S = \mathbb{I}X + (1-\mathbb{I})Y$*, where* $\mathbb{I}$ *is a Bernoulli distributed random variable with parameter* p_f *and where the components* X *and* Y *are independent of* $\mathbb{I}$*. Define* $\alpha_* \in [0,1]$ *by*

$$\alpha_* := \inf\left\{\alpha \in (0,1) \mid \exists \beta \in (0,1) \left\{ \begin{array}{c} p_f\alpha + (1-p_f)\beta = p \\ VaR_\alpha(X) \geqslant VaR_\beta(Y) \end{array} \right\}\right\}$$

and let $\beta_* = \frac{p-p_f\alpha_*}{1-p_f} \in [0,1]$*. Then, for* $p \in (0,1)$*,*

$$VaR_p(S) = \max\left\{VaR_{\alpha_*}(X),\, VaR_{\beta_*}(Y)\right\}. \tag{12}$$

The proof is provided in Appendix A.4. It is clear that in many cases[4] $\mathrm{VaR}_{\alpha_*}(X) = \mathrm{VaR}_{\beta_*}(Y)$. Using Lemma 2.9, we are now ready to generate other (more explicit) expressions for the upper bound M_p and the lower bound m_p.

Proposition 2.10 (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 (3), (5) and* (6)*. 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* $p \in (0,1)$*,*

$$M_p = \begin{cases} TVaR_{\frac{p-p_f\alpha_*}{1-p_f}}\left(\sum_{i=1}^d Z_i\right) & \text{if } \frac{p+p_f-1}{p_f} < \alpha_* < \frac{p}{p_f} \\ VaR_{\alpha_*}(T) & \text{if } \alpha_* = \frac{p}{p_f} \\ \max\left\{VaR_{\alpha_*}(T), TVaR_{\frac{p-p_f\alpha_*}{1-p_f}}\left(\sum_{i=1}^d Z_i\right)\right\} & \text{if } \alpha_* = \frac{p+p_f-1}{p_f}. \end{cases} \tag{13}$$

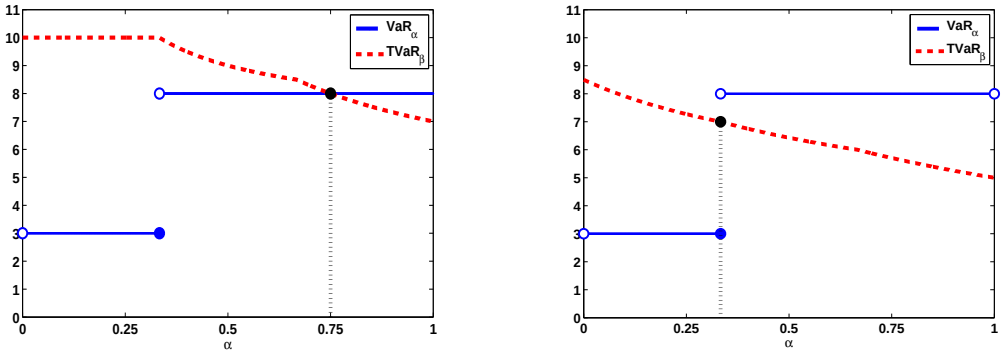
The expressions for the lower bound m_p *are obtained by replacing, in the above statements, "TVaR" with "LTVaR."*

The proof of Proposition 2.10 is provided in Appendix A.4.

Illustration: We illustrate the proposition 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 (13) in Proposition 2.10 and find that $\alpha_* = 0.75$ and that the maximum VaR_p is equal to $TVaR_{0.55}(\sum Z_i) = 8$. This value is illustrated in Figure 2.

Note that TVaR is continuous. When $\max\left\{0, \frac{p-p_f}{1-p_f}\right\} < \beta_* < \min\left\{1, \frac{p}{1-p_f}\right\}$, then there are two possible cases illustrated by Panel A and Panel B of Figure 2.

[4]For example, it is sufficient that the distribution functions of X and Y are strictly increasing with unbounded support. See also the proof for more cases.



Panel A Panel B

Figure 2: VaR_α and $TVaR_{\beta(\alpha)}$ for $\alpha \in (0,1)$

Note that the setting of this illustration will also be used in the following section to discuss a non-parametric approach to computing bounds. Next, we assess the VaR bounds when the benchmark model is a multivariate normal distribution.

Example 2.11 (Multivariate normal distribution as a benchmark model)**.** The VaR bounds reported in Table 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 $\mathrm{VaR}_{95\%}\left(\sum_{i=1}^{20} X_i\right) = \sqrt{20}\Phi^{-1}(95\%) = 7.35$ and, similarly, $\mathrm{VaR}_{99.5\%}\left(\sum_{i=1}^{20} X_i\right) = 11.5$, $\mathrm{VaR}_{99.95\%}\left(\sum_{i=1}^{20} X_i\right) = 14.7$. 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.

Finally, note that when very high probability levels are used in VaR calculations ($p = 99.95\%$; see the last three rows in Table 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, Puccetti, and Rüschendorf (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 2.8 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, Rüschendorf, and Vanduffel (2013). 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 4. For example, when $p = 95\%$, we find that $(A, B) = (-2.17, 41.25)$, which conforms with the numbers in Table 4.

$F = [q_\beta, q_{1-\beta}]^d$		$\mathcal{U} = \emptyset$ $\beta = 0\%$	$\beta = 0.05\%$	$\beta = 0.5\%$	$\beta = 5\%$	$\mathcal{U} = \mathbb{R}^d$ $\beta = 0.5$
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 4: $\text{VaR}_{95\%}$, $\text{VaR}_{99.5\%}$ and $\text{VaR}_{99.95\%}$ 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.

2.4 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

$$\frac{\rho^+_{\mathcal{F}} - \rho(\sum_i X_i)}{\rho^+_{\mathcal{F}}} \text{ and } \frac{\rho^-_{\mathcal{F}} - \rho(\sum_i X_i)}{\rho^-_{\mathcal{F}}}, \tag{14}$$

which were introduced in Section 2 (Definition 2.1 and expressions (1) and (2)). The risk measure $\rho(\cdot)$ is the VaR and the TVaR, and, for the trusted area $\mathcal{F}$, we consider the elliptical contours as in (10) such that $P((X_1, ..., X_d) \in \mathcal{F}) = p_f$.

In Figure 3, we represent the risk of underestimating and overestimating VaR and TVaR, respectively, at various probability levels p using the risk measures (14) for model risk. From Figure 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

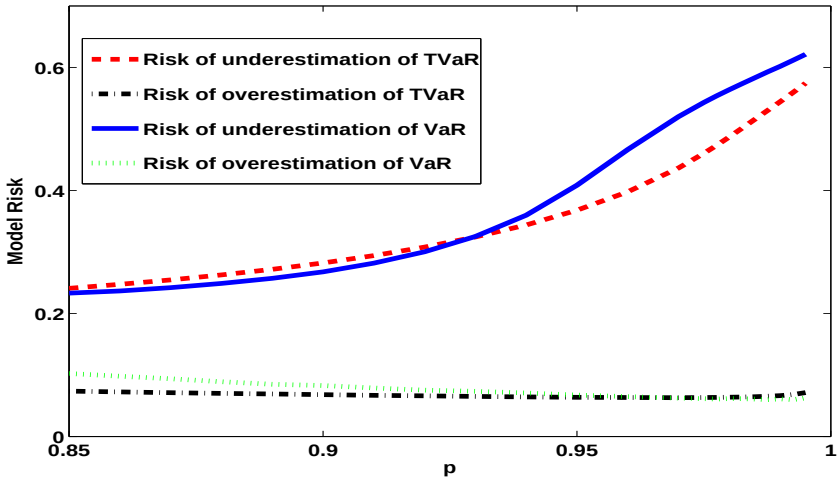


Figure 3: We assume that $(X_1, X_2, ...X_{20})$ is a multivariate standard distribution with pair correlation $\rho = 0.1$. Let $p_f = 90\%$. We show the model risk for overestimating or underestimating VaR_p and TVaR_p as a function of p.

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).

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.[5] 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.

3 Approximating Sharp Bounds

A bound on a risk measure is "sharp" if there exists a dependence structure among the risks such that this bound is attained. In this section we discuss the sharpness of upper and lower bounds for the three risk measures considered in the previous section.

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

3.1 Sharpness of the Bounds

Variance and TVaR: Some of the bounds stated in Propositions 2.2 and 2.4 can be sharp. In particular, the *upper* bounds for the variance and for the TVaR stated in Propositions 2.2 and 2.4 are sharp, without further conditions. Note, indeed, that the multivariate vector

$$(\mathbb{I}X_1 + (1-\mathbb{I})Z_1, \mathbb{I}X_2 + (1-\mathbb{I})Z_2, ..., \mathbb{I}X_d + (1-\mathbb{I})Z_d) \tag{15}$$

satisfies conditions (i), (ii) and (iii). 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.

VaR: Recall from the discussion of Proposition 2.8 that the stated upper and lower VaR bounds are not sharp in general. Nevertheless, thanks to Lemma 2.9, 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, from Lemma 2.9, for all α, β in $[0,1]$ such that $p_f\alpha + (1-p_f)\beta = p$,

$$\max\left\{\mathrm{VaR}_\alpha(T), \mathrm{VaR}_\beta\left(\sum_{i=1}^{d} Z_i\right)\right\} \leqslant M_p = \max\left\{\mathrm{VaR}_{\alpha_*}(T), \mathrm{TVaR}_{\beta_*}\left(\sum_{i=1}^{d} Z_i\right)\right\} \tag{16}$$

where α_* is defined as in Proposition 2.10 (consistent with Lemma 2.9) 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 (16) 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

$$\mathrm{VaR}_{\beta_*}\left(\sum_{i=1}^{d} Z_i\right) = \mathrm{TVaR}_{\beta_*}\left(\sum_{i=1}^{d} Z_i\right). \tag{17}$$

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 in the following section.

3.2 Approximations

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 before, 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 modelling 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 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, without any loss of generality, the observed data points are reported in

the following matrix M

$$M = \begin{bmatrix} x_{i_1 1} & x_{i_1 2} & ... & x_{i_1 d} \\ x_{i_2 1} & x_{i_2 2} & ... & x_{i_2 d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i_{\ell_f} 1} & x_{i_{\ell_f} 2} & ... & x_{i_{\ell_f} d} \\ x_{m_1 1} & x_{m_1 2} & ... & x_{m_1 d} \\ x_{m_2 1} & x_{m_2 2} & ... & x_{m_2 d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m_{\ell_u} 1} & x_{m_{\ell_u} 2} & ... & x_{m_{\ell_u} d} \end{bmatrix}, \tag{18}$$

where the gray area reflects $\mathcal{F}_N$ and the white area reflects $\mathcal{U}_{\mathcal{N}}$. 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:

$$\begin{bmatrix} S_N^f \\ S_N^u \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{\ell_f} \\ \tilde{s}_1 := x_{m_1 1} + x_{m_1 2} + ... + x_{m_1 d} \\ \tilde{s}_2 := x_{m_2 1} + x_{m_2 2} + ... + x_{m_2 d} \\ \vdots \\ \tilde{s}_{\ell_u} := x_{m_{\ell_u} 1} + x_{m_{\ell_u} 2} + ... + x_{m_{\ell_u} d} \end{bmatrix}. \tag{19}$$

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 the paper 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 gray area of the matrix:

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

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:

$$M = \begin{bmatrix} 3 & 4 & 1 \\ 2 & 4 & 2 \\ 0 & 2 & 1 \\ 4 & 3 & 3 \\ 3 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad S_N^f = \begin{bmatrix} 8 \\ 8 \\ 3 \end{bmatrix}, \quad S_N^u = \begin{bmatrix} 10 \\ 7 \\ 4 \\ 3 \\ 1 \end{bmatrix}. \tag{21}$$

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 :

$$S_N = \mathbb{I} S_N^f + (1 - \mathbb{I}) S_N^u, \tag{22}$$

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 6 and Proposition 2.10), whereas S_N^u is a comonotonic sum and corresponds to the sampled version of $\sum_{i=1}^d Z_i$. In this paper, we aim at finding worst-case dependences 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).

3.3 Bounds on Standard Deviation

From Proposition 2.2 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 (19). The upper bound on variance is then computed as

$$\frac{1}{N} \left(\sum_{i=1}^{\ell_f} (s_i - \bar{s})^2 + \sum_{i=1}^{\ell_u} (\tilde{s}_i - \bar{s})^2 \right), \tag{23}$$

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

$$\bar{s} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{d} x_{ij} = \frac{1}{N} \left(\sum_{i=1}^{\ell_f} s_i + \sum_{i=1}^{\ell_u} \tilde{s}_i \right). \tag{24}$$

To achieve the minimum variance bound found in Proposition 2.2, 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 (RA) of Embrechts, Puccetti, and Rüschendorf (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 Appendix B of this paper. 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:

$$\frac{1}{N}\left(\sum_{i=1}^{\ell_f}(s_i-\bar{s})^2+\sum_{i=1}^{\ell_u}(\tilde{s}_i^m-\bar{s})^2\right), \tag{25}$$

where $\bar{s}$ is computed as in (24).

We illustrate the upper and lower bounds (23) and (25) for the variance derived above with the matrix M of observations provided in (21). 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 (21). 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

$$M=\begin{bmatrix}3&4&1\\2&4&2\\0&2&1\\1&1&3\\0&3&2\\1&2&2\\3&1&1\\4&0&1\end{bmatrix},\quad S_N^f=\begin{bmatrix}8\\8\\3\end{bmatrix},\quad S_N^u=\begin{bmatrix}5\\5\\5\\5\\5\end{bmatrix}. \tag{26}$$

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

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

3.4 Bounds on TVaR

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

$$k:=N(1-p), \tag{27}$$

where k is an integer. Similarly to the case of maximizing the variance, it follows from Proposition 2.4 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 (19). 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 (26) 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 $21/3=7$.

3.5 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. We compute the VaR at probability level p, so that, for ease of exposition, $k := N(1-p)$, where we assume that k is an integer.

The algorithm is based on Proposition 2.10 and is further inspired by the motivation provided in subsection 3.1. 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 (22)). 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 $m_1 \leqslant m_* \leqslant m_2$.

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 ($m_1 \leqslant m \leqslant m_2$).

We now illustrate the algorithm for obtaining the maximum VaR in the example with $d = 3$, $N = 8$, $k = 3$ with the same matrix M given in (21) so that $\ell_f = 3$ and $\ell_u = 5$. In this case, $\alpha_* = 0.75$ (as illustrated in Panel A of Figure 2 and as discussed in the preceding section), so the theoretical maximum VaR is equal to $\mathrm{TVaR}_{\frac{p-p_f\alpha_*}{1-p_f}} = 8$. 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 that $m_1 = \max(0, 0) = 0$ and $m_2 = \min(3, 8-3) = 3$, so that the number of rows to which one can consider applying

the RA is between 0 and 3, as $k + m_1 - \ell_f = 0$ and $k + m_2 - \ell_f = 3$. 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_*. It can be found in Appendix C.

4 Application to a Portfolio

We illustrate the results in the paper in order to discuss the model risk of an investment portfolio. The portfolio we consider consists of five blue chips, namely 3M (MMM), Bank of America (BAC), Alcoa (AA), American Express (AXP) and General Electric (GE). Let X_1, X_2, ... X_5 denote their respective daily returns. Using Yahoo Finance, we built a history of daily returns from Oct. 1, 1986 to Sept. 20, 2013, i.e., 6,807 observations in total. We summarize the characteristics of the observed returns for the stocks involved. In this respect, we report the annualized standard deviation (obtained by multiplying the daily standard deviation by $\sqrt{252}$).

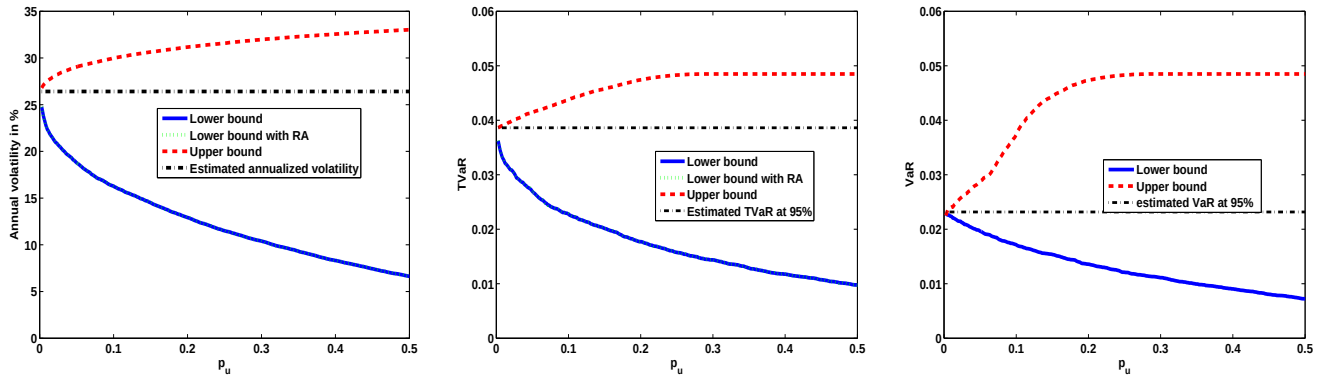
	MMM X_1	AA X_2	BAC X_3	AXP X_4	GE X_5
daily mean	$0.56\ 10^{-3}$	$0.45\ 10^{-3}$	$0.63\ 10^{-3}$	$0.70\ 10^{-3}$	$0.58\ 10^{-3}$
daily stdev	0.0152	0.0238	0.0273	0.0234	0.0184
annualized stdev	24.1%	37.7%	43.3%	37.1%	29.2%

Table 5: Summary statistics

We assume that the portfolio is always equally invested in the five stocks. Hence, the daily return of the portfolio can be computed as

$$S = \frac{X_1 + X_2 + ... + X_5}{5}.$$

We are interested in assessing the risk (standard deviation, TVaR and VaR) of the portfolio return S. To this end, it is more convenient to work with the variable $L := -S$. To estimate the (benchmark) standard deviation, VaR and TVaR we make use of the collected data in a straightforward way. Next, we use the results and the algorithms described above to assess the upper and lower bounds on these estimates. For example, the observed portfolio has an estimated annualized volatility of 26.42%. In the presence of full uncertainty ($p_u = 100\%$) on the dependence (the stock returns are all comonotonic), we observe that the annualized volatility becomes 34.14%, which provides an absolute upper bound on the volatility. Note that the portfolio presented here thus contains risks that are already highly dependent, as the distance between the observed volatility and the maximum volatility is small.



Panel A: standard deviation Panel B: TVaR Panel C: VaR

Figure 4: Panel A: Bounds on standard deviation (volatility): The lower bound is computed either by Proposition 2.2 or by the algorithm described in Section 3.3. Panel B: Bounds on TVaR of $L = -S$ at 95%: The lower bound is computed either by Proposition 2.4 or by the algorithm described in Section 3.4. Panel C: Bounds on VaR of $L = -S$ at 95%: The lower bound is computed either by Proposition 2.10 or by the algorithm described in Section 3.5.

Let us next compute upper and lower bounds for the different risk measures by varying the level of the p_u. It is clear that, as the initial situation represents risks that are already highly dependent, the maximum bounds on standard deviation, VaR and TVaR are closer to the observed ones (assuming that there is no uncertainty on the dependence) than are the minimum bounds. In line with the observations made in Section 2.4 we see that when p becomes large (as compared to $1 - p_u$), the worst VaR and TVaR tend to match each other. Consequently, the risk of underestimating the VaR is higher than the risk of underestimating the TVaR. The results suggest that a regulation based on VaR at lower probability levels (say 90%) would be more robust to model misspecification than a regulation based on VaR at the high confidence level $p = 99.5\%$ (i.e., the current situation under Solvency II and Basel III).

5 Final Remarks

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.

The measurement of model risk may be considered reasonably under control when only the marginal distributions are assumed to be known (unconstrained bounds); see Embrechts, Puccetti, and Rüschendorf (2013). However, these bounds are wide, as they ignore the (partial) information on dependence that might be available. In this paper, we integrate in a natural way information on the multivariate structure. We assume that the marginal distributions are known while the joint model is only partially known (on some subset $\mathcal{F}$). We design a data-driven algorithm that approximates sharp bounds, and numerical illustrations show that these bounds correspond closely to the easy-to-compute bounds that we derived in the first section, and thus that they are nearly sharp in practical situations.

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 VaR at a very high confidence level since such an assessment is unable to benefit from careful risk management attempts to fit a multivariate model. We remark that it could be of interest to consider also a "global" constraint, in addition to the constraints (i), (ii) and (iii) (see Section 2), 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, Rüschendorf, and Vanduffel (2013) to account for a maximum possible variance of the aggregate portfolio.

Finally, the following comments and considerations may be interesting for further research. In the paper, we do not study how to optimally choose the trusted area $\mathcal{F}$; rather, we provide two ways to do so. The first way is based solely on the use of the marginal densities, as in (8), and typically reduces the trusted area to a d dimensional cube. The other criterion, in (9), is based on the contours of a given fitted multivariate density $\widehat{f}$ (deriving, for instance, from a multivariate Gaussian model, a multivariate Student model or a Pair-Copula Construction model (Aas, Czado, Frigessi, and Bakken (2009), Czado (2010))). We leave it for future research to determine which criterion is best to define the "trusted area". The difficulty here resides in finding an appropriate measure for the observed "density" of data points that is able to accurately reflect the level of trust that one can have in a fitted model.

If there is a lack of data and yet one believes in the marginal distributions that have been fitted, then it is possible to improve the efficiency of the algorithm by re-discretizing using the fitted marginal $\hat{f}_i$. The idea is simple: in the "trusted" part, one repeats the same observations k times and, in the "untrusted part,"one discretizes so that there are k times more observations. Then, one applies to the new matrix the algorithms that are described in this paper. This is carried out in the same spirit as bootstrapping.

Finally, in the paper 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 $\widehat{f}_i$).

A Proofs

A.1 Proof of Proposition 2.2

It is well-known that for any vector $(Y_1, Y_2, ..., Y_d)$ and any convex function $v(x)$, it holds that

$$\mathrm{E}\left(v\left(\sum_{i=1}^{d} Y_i\right)\right) \leqslant \mathrm{E}\left(v\left(\sum_{i=1}^{d} F_{Y_i}^{-1}(U)\right)\right), \tag{28}$$

where U is a uniformly distributed random variable on $(0,1)$; see Meilijson and Nádas (1979). A simple conditioning argument and taking into account Jensen's inequality then shows that for all convex functions $v(x)$,

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

Since $v(x) = x^2$ is convex and all sums involved in the above inequality have the same first moment, the stated result follows. □

A.2 Proof of Proposition 2.4

The proof is immediate by the fact that $\mathrm{TVaR}_p(X) \leqslant \mathrm{TVaR}_p(Y)$ for all $0 < p < 1$ if and only if $E(v(X)) \leqslant E(v(Y))$ for all convex functions $v(\cdot)$.

In fact, the ordering result in the proposition holds for any risk measure that respects convex order, and the variance and TVaR are merely two particular examples. An important family of risk measures for which the ordering result holds is the family of concave distortion risk measures (which contains the TVaR). □

A.3 Proof of Proposition 2.8

For a given random vector $(X_1, X_2, ..., X_d)$ satisfying properties (i), (ii) and (iii), there exists a certain vector $(Y_1, Y_2, ...Y_d)$ with marginals Y_i that have the same distribution as Z_i such that

$$\sum_{i=1}^{d} X_i =_d \left(\mathbb{I}\sum_{i=1}^{d} X_i + (1-\mathbb{I})\sum_{i=1}^{d} Y_i\right).$$

As per definition of the VaR, it follows for all $p \in (0,1)$,

$$\mathrm{VaR}_p\left(\sum_{i=1}^{d} X_i\right) = \inf\left\{x \in \mathbb{R} \mid p_f F_{\left(\sum_{i=1}^{d} X_i|\mathbb{I}=1\right)}(x) + (1-p_f)F_{\sum_{i=1}^{d} Y_i}(x) \geqslant p\right\}, \tag{29}$$

where $p_f := P(\mathbb{I} = 1)$. Note that for all $p \in (0,1)$,

$$F^{-1}_{\sum_{i=1}^{d} Y_i}(p) \leqslant \mathrm{TVaR}_p\left(\sum_{i=1}^{d} Y_i\right) \leqslant \mathrm{TVaR}_p\left(\sum_{i=1}^{d} Z_i\right),$$

where the second inequality follows from the fact that the Z_i are comonotonic (while having the same distribution as the Y_i). Thus,

$$F^{-1}_{\sum_{i=1}^d Y_i}(U) \leqslant R := \mathrm{TVaR}_U(\sum_{i=1}^d Z_i) \text{ a.s.}$$

which can be also written in terms of their cdf. Therefore, for all $x \in \mathbb{R}$,

$$F_{\sum_{i=1}^d Y_i}(x) \geqslant F_R(x). \tag{30}$$

From (30) it follows that

$$\mathrm{VaR}_p\left(\sum_{i=1}^d X_i\right) \leqslant \inf\left\{x \in \mathbb{R} \mid p_f F_{\left(\sum_{i=1}^d X_i|\mathbb{I}=1\right)}(x) + (1-p_f)F_R(x) \geqslant p\right\}.$$

We observe that the right-hand side of the above equation is by definition the VaR of a sum of mutually exclusive variables, it follows that

$$\begin{aligned}\mathrm{VaR}_p\left(\sum_{i=1}^d X_i\right) &\leqslant \mathrm{VaR}_p\left(\mathbb{I}\sum_{i=1}^d X_i + (1-\mathbb{I})TVaR_U\left(\sum_{i=1}^d Z_i\right)\right)\\ &= \mathrm{VaR}_p\left(\mathbb{I}\sum_{i=1}^d X_i + (1-\mathbb{I})\sum_{i=1}^d H_i\right),\end{aligned}$$

where the last equality follows from the fact that TVaR is additive for the comonotonic sum $\sum_{i=1}^d Z_i$. The proof for the lower bound is similar and omitted. □

A.4 Proof of Lemma 2.9 and Proposition 2.10

Proof of Lemma 2.9 Denote by $F(x)$ and $G(x)$ the distributions of X resp. Y. Since X and Y are independent of $\mathbb{I}$ we find for the distribution of $S = \mathbb{I}X + (1-\mathbb{I})Y$,

$$F_S(x) = p_f F(x) + p_u G(x) \qquad x \in \mathbb{R}.$$

Let $p \in (0,1)$ and denote $\mathrm{VaR}_p(S)$ by s_p,

$$s_p = \inf\{x \in \mathbb{R} \mid p_f F(x) + p_u G(x) \geqslant p\}.$$

In what follows, when considering $\alpha, \beta \in (0,1)$ we always assume that they satisfy $p_f\alpha + (1-p_f)\beta = p$. Note that we define α_* as

$$\alpha_* := \inf\{\alpha \in (0,1) \mid \exists \beta \in (0,1) \ / \ p_f\alpha + (1-p_f)\beta = p \text{ and } \mathrm{VaR}_\alpha(X) \geqslant \mathrm{VaR}_\beta(Y)\} \tag{31}$$

and $\beta_* = \frac{p - p_f\alpha_*}{1-p_f}$. The proof consists in verifying that s_p can always be expressed as

$$s_p = \max\{\mathrm{VaR}_{\alpha_*}(X), \mathrm{VaR}_{\beta_*}(Y)\}. \tag{32}$$

To this end, we compute s_p by distinguishing the 4 following cases for $F(\cdot)$ and for $G(\cdot)$:

Case 1: F is continuous in s_p and for all $z < s_p$, $F(z) < F(s_p)$
Case 2: F is continuous in s_p and there exists $z < s_p$, $F(z) = F(s_p)$
Case 3: F is discontinuous in s_p and for all $z < s_p$, $F(z) < F(s_p^-)$
Case 4: F is discontinuous in s_p and there exists $z < s_p$, $F(z) = F(s_p^-)$
Case a: G is continuous in s_p and for all $z < s_p$, $G(z) < G(s_p)$
Case b: G is continuous in s_p and there exists $z < s_p$, $G(z) = G(s_p)$
Case c: G is discontinuous in s_p and for all $z < s_p$, $G(z) < G(s_p^-)$
Case d: G is discontinuous in s_p and there exists $z < s_p$, $G(z) = G(s_p^-)$

We have summarized the computations of s_p in Table 6. From Table 6, it is clear that (32) is proved. Let us now make the calculations case by case.

—	(a)	(b)	(c)	(d)
(1)	$\alpha_* = F(s_p)$ $\beta_* = G(s_p)$ $s_p = VaR_{\alpha_*}(X)$ $= VaR_{\beta_*}(Y)$	$\alpha_* = F(s_p)$ $\beta_* = G(s_p)$ $s_p = VaR_{\alpha_*}(X)$ $> VaR_{\beta_*}(Y)$	$\alpha_* = F(s_p)$ $s_p = VaR_{\alpha_*}(X)$ $= VaR_{\beta_*}(Y)$	$\alpha_* = F(s_p)$ if $F_S(s_p^-) < p$, $s_p = VaR_{\alpha_*}(X)$ $= VaR_{\beta_*}(Y)$ if $F_S(s_p^-) = p$, $s_p = VaR_{\alpha_*}(X)$ $> VaR_{\beta_*}(Y)$
(2)	$\alpha_* = F(s_p)$ $\beta_* = G(s_p)$ $s_p = VaR_{\beta_*}(Y)$ $> VaR_{\alpha_*}(X)$	Impossible	$\alpha_* = F(s_p)$ $s_p = VaR_{\beta_*}(Y)$ $> VaR_{\alpha_*}(X)$	$\alpha_* = F(s_p)$ $s_p = VaR_{\beta_*}(Y)$ $> VaR_{\alpha_*}(X)$
(3)	$\beta_* = G(s_p)$ $s_p = VaR_{\alpha_*}(X)$ $= VaR_{\beta_*}(Y)$	$\beta_* = G(s_p)$ $s_p = VaR_{\alpha_*}(X)$ $> VaR_{\beta_*}(Y)$	$s_p = VaR_{\alpha_*}(X)$ $= VaR_{\beta_*}(Y)$	if $F_S(s_p^-) < p$, $s_p = VaR_{\alpha_*}(X)$ $= VaR_{\beta_*}(Y)$ if $F_S(s_p^-) = p$, $s_p = VaR_{\alpha_*}(X)$ $> VaR_{\beta_*}(Y)$
(4)	$\beta_* = G(s_p)$ if $F_S(s_p^-) < p$, $s_p = VaR_{\beta_*}(Y)$ $= VaR_{\alpha_*}(X)$ if $F_S(s_p^-) = p$, $s_p = VaR_{\beta_*}(Y)$ $> VaR_{\alpha_*}(X)$	$\beta_* = G(s_p)$ $s_p = VaR_{\alpha_*}(X)$ $> VaR_{\beta_*}(Y)$	if $F_S(s_p^-) < p$, $s_p = VaR_{\alpha_*}(X)$ $= VaR_{\beta_*}(Y)$ if $F_S(s_p^-) = p$, $s_p = VaR_{\beta_*}(Y)$ $> VaR_{\alpha_*}(X)$	Impossible

Table 6: Summary of all cases for the VaR of a mixture where $s_p = VaR_p(S)$. In all cases, α_* is defined as (31) and $\beta_* = \frac{p - p_f \alpha_*}{1 - p_f} \leqslant G(s_p)$, $\alpha* = \frac{p - (1-p_f)\beta_*}{p_f} \geqslant F(s_p)$.

Case 1: F is continuous in s_p and for all $z < s_p$, $F(z) < F(s_p)$

In this case we always have that $s_p = \mathrm{VaR}_{F(s_p)}(X)$. Hence, we only need to show that $\alpha_* = F(s_p)$ (i.e. $\beta_* = \frac{p - p_f F(s_p)}{1 - p_f}$) and that $s_p = \mathrm{VaR}_{\alpha_*}(X) \geqslant \mathrm{VaR}_{\beta_*}(Y)$ as in this case (32) will obviously hold.

Since $VaR_p(S) = s_p$ then $F_S(s_p^-) = p_f F(s_p^-) + p_u G(s_p^-) \leqslant p \leqslant F_S(s_p) = p_f F(s_p) +$

$p_u G(s_p)$. Thus, by continuity of F, $p_f F(s_p) + p_u G(s_p^-) \leqslant p \leqslant p_f F(s_p) + p_u G(s_p)$. Thus,

$$G(s_p^-) \leqslant \frac{p - p_f F(s_p)}{1 - p_f} \leqslant G(s_p) \tag{33}$$

(1a): G is continuous in s_p and for all $z < s_p$, $G(z) < G(s_p)$. Then, $s_p = \mathrm{VaR}_{G(s_p)}(Y)$. It is also clear that for $\alpha < F(s_p)$ and thus $\beta > G(s_p)$, one has that $\mathrm{VaR}_\alpha(X) < \mathrm{VaR}_\beta(Y)$. Hence, as per definition of α_*, one has $\alpha_* = F(s_p)$, $\beta_* = G(s_p)$ and $s_p = \mathrm{VaR}_{\alpha_*}(X) = \mathrm{VaR}_{\beta_*}(Y)$.

(1b): G is continuous in s_p and there exists $z < s_p$, $G(z) = G(s_p)$ (thus, G is constant on the interval (z, s_p)). Then, $\mathrm{VaR}_{G(s_p)}(Y) < s_p = \mathrm{VaR}_{F(s_p)}(X)$. However, for $\alpha < F(s_p)$ and thus $\beta > G(s_p)$, one has that $\mathrm{VaR}_\alpha(X) < \mathrm{VaR}_\beta(Y)$. Hence, as per definition of α_*, $\alpha_* = F(s_p)$, $\beta_* = G(s_p)$ and $s_p = \mathrm{VaR}_{\alpha_*}(X) > \mathrm{VaR}_{\beta_*}(Y)$. Thus, $s_p = VaR_{\alpha_*}(X) > \mathrm{VaR}_{\beta_*}(Y)$.

(1c): G has a discontinuity in s_p and for all $z < s_p$, $G(z) < G(s_p^-)$. From (33), in this case, $\mathrm{VaR}_{\frac{p-p_f F(s_p)}{1-p_f}}(Y) = s_p$. For $\alpha < F(s_p)$ and thus $\beta > \frac{p-p_f F(s_p)}{1-p_f}$, $\mathrm{VaR}_\alpha(X) < \mathrm{VaR}_\beta(Y)$. Hence, as per definition of α_*, $\alpha_* = F(s_p)$, $\beta_* = \frac{p-p_f F(s_p)}{1-p_f}$ and $s_p = \mathrm{VaR}_{\alpha_*}(X) = \mathrm{VaR}_{\beta_*}(Y)$.

(1d): G has a discontinuity in s_p and there exists $z < s_p$, $G(z) = G(s_p^-)$ so that G is constant on some interval (r, s_p) with $r < s_p$. From (33),

$$\mathrm{VaR}_{\frac{p-p_f F(s_p)}{1-p_f}}(Y) \leqslant s_p.$$

If $\frac{p-p_f F(s_p)}{1-p_f} > G(s_p^-)$ (or equivalently, $F_S(s_p^-) < p$), then $\mathrm{VaR}_{\frac{p-p_f F(s_p)}{1-p_f}}(Y) = \mathrm{VaR}_{F(s_p)}(X) = s_p$. Clearly, for $\alpha < F(s_p)$ and thus $\beta > \frac{p-p_f F(s_p)}{1-p_f}$, one has that $\mathrm{VaR}_\alpha(X) < \mathrm{VaR}_\beta(Y)$. Hence, as per definition of α_*, one has $\alpha_* = F(s_p)$, $\beta_* = \frac{p-p_f F(s_p)}{1-p_f}$ and $s_p = \mathrm{VaR}_{\alpha_*}(X) = \mathrm{VaR}_{\beta_*}(Y)$. If $\frac{p-p_f F(s_p)}{1-p_f} = G(s_p^-)$ (or equivalently, $F_S(s_p^-) = p$), then this implies that $\mathrm{VaR}_{\frac{p-p_f F(s_p)}{1-p_f}}(Y) < s_p$. When $\alpha < F(s_p)$ thus $\beta > \frac{p-p_f F(s_p)}{1-p_f}$ one has that $\mathrm{VaR}_\alpha(X) < \mathrm{VaR}_\beta(Y) \leqslant s_p$. Hence, as per definition of α_*, one has $\alpha_* = F(s_p)$, $\beta_* = \frac{p-p_f F(s_p)}{1-p_f}$ and $s_p = \mathrm{VaR}_{\alpha_*}(X) > \mathrm{VaR}_{\beta_*}(Y)$.

Case 2: F is continuous in s_p and there is a $z < s_p$, $F(z) = F(s_p)$ ($F(\cdot)$ is constant on (z, s_p))

(2a): this case can be obtained from (1b) by changing the role of X and Y.

(2b): G is continuous in s_p and there exists $z < s_p$, $G(z) = G(s_p)$. Thus G is constant on some interval (r, s_p) with $r < s_p$. Hence, $VaR_p(S) \leqslant \min(z, z) < s_p$ which contradicts the definition of $s_p = VaR_p(S)$. The case (2b) is impossible.

(2c): G is discontinuous in s_p and for all $z < s_p$, $G(z) < G(s_p^-)$. From (33), in this case, $\mathrm{VaR}_{\frac{p-p_f F(s_p)}{1-p_f}}(Y) = s_p > \mathrm{VaR}_{F(s_p)}(X)$. However, for all $\alpha > F(s_p)$ and thus $\beta < \frac{p-p_f F(s_p)}{1-p_f}$ it holds that $\mathrm{VaR}_\alpha(X) > \mathrm{VaR}_\beta(Y)$. Hence, as per definition of α_*, $\alpha_* = F(s_p)$, $\beta_* = \frac{p-p_f F(s_p)}{1-p_f}$ and $s_p = \mathrm{VaR}_{\beta_*}(Y) > \mathrm{VaR}_{\alpha_*}(X)$.

(2d): G is discontinuous in s_p and there exists $z < s_p$, $G(z) = G(s_p^-)$. From (33), $\mathrm{VaR}_{\frac{p-p_f F(s_p)}{1-p_f}}(Y) \leqslant s_p$. If $\frac{p-p_f F(s_p)}{1-p_f} > G(s_p^-)$ (or equivalently, $F_S(s_p^-) < p$), then $\mathrm{VaR}_{\frac{p-p_f F(s_p)}{1-p_f}}(Y) =$

$s_p > \mathrm{VaR}_{F(s_p)}(X)$. For $\alpha > F(s_p)$ and thus $\beta < \frac{p-p_f F(s_p)}{1-p_f}$ one has that $\mathrm{VaR}_\alpha(X) > \mathrm{VaR}_\beta(Y)$. Hence, as per definition of α_*, $\alpha_* = F(s_p)$, $\beta_* = \frac{p-p_f F(s_p)}{1-p_f}$ and $\mathrm{VaR}_{\alpha_*}(X) < \mathrm{VaR}_{\beta_*}(Y) = s_p$. The case that $\frac{p-p_f F(s_p)}{1-p_f} = G(s_p^-)$ is excluded as it implies that $\mathrm{VaR}_p(S) < s_p$ should hold (similar to the case (2b)) which is a contradiction with the definition of s_p.

Case 3: F has a discontinuity in s_p and for all $z < s_p$, $F(z) < F(s_p^-)$

In this case, $s_p = \mathrm{VaR}_{F(s_p)}(X)$. This situation is merely identical to previous cases.

(3a): it is the same as (1c) by changing the role of X and Y.

(3b): it is the same as (2d) by changing the role of X and Y.

(3c): Observe that $VaR_\alpha(X) = s_p$ for all $F(s_p^-) \leqslant \alpha \leqslant F(s_p)$ and also that $VaR_\beta(Y) = s_p$ for all $G(s_p^-) \leqslant \beta \leqslant G(s_p)$.

We also know that $F_S(s_p^-) \leqslant p \leqslant F_S(s_p)$ hence there exists $F(s_p^-) \leqslant \alpha_1 \leqslant F(s_p)$ and $G(s_p^-) \leqslant \beta_1 \leqslant G(s_p)$ so that $p_f\alpha_1 + p_u\beta_1 = p$ and $VaR_{\alpha_1}(X) = VaR_{\beta_1}(Y) = s_p$. Therefore, $VaR_{\alpha_*}(X) = VaR_{\beta_*}(Y) = s_p$.

(3d): Observe that $VaR_\alpha(X) = s_p$ for all $F(s_p^-) \leqslant \alpha \leqslant F(s_p)$ and also that $VaR_\beta(Y) = s_p$ for all $G(s_p^-) < \beta \leqslant G(s_p)$.

We also know that $F_S(s_p^-) \leqslant p \leqslant F_S(s_p)$ and there are two possibilities:

In the case when $F_S(s_p^-) < p$, then there exists $\alpha_1 \in (F(s_p^-), F(s_p))$ and $\beta_1 \in (G(s_p^-), G(s_p))$ so that $p_f\alpha_1 + p_u\beta_1 = p$ and $VaR_{\alpha_1}(X) = VaR_{\beta_1}(Y) = s_p$. Therefore, $VaR_{\alpha_*}(X) = VaR_{\beta_*}(Y) = s_p$.

In the case when $F_S(s_p^-) = p$, then $p_f F(s_p^-) + p_u G(s_p^-) = p$ and one has that $VaR_{F(s_p^-)}(X) > VaR_{G(s_p^-)_*}(Y)$, while for $\alpha < F(s_p^-)$ and $\beta > G(s_p^-)$ one has that $VaR_\alpha(X) < VaR_\beta(Y)$. Hence, $\alpha_* = F(s_p^-)$, $G(s_p^-) = \beta_*$ and $s_p = VaR_{\alpha_*}(X) > VaR_{\beta_*}(Y)$.

Case 4: F has a discontinuity in s_p and there exists $z < s_p$, $F(z) = F(s_p^-)$

By changing the role of X and Y we have that the case (4a) corresponds to (1d), the case (4b) corresponds to (2d) and the case (4c) corresponds to (3d). Finally the case of (4d) is treated as follows. In the case (4d), both F and G are discontinuous at s_p, and there exists z_1 and z_2 such that $F(z_1) = F(s_p)$ and $G(z_2) = G(s_p)$ so that F is constant on (z_1, s_p) and G is constant on (z_2, s_p). Then $VaR_p(S) \leqslant \min(z_1, z_2) < s_p$ which contradicts the definition of $s_p = VaR_p(S)$. This case is thus impossible. □

Proof of Proposition 2.10 The proof follows as a direct application of Lemma 2.9. Consider $X = T$ with distribution F, and $Y = \sum_{i=1}^d TVaR_U(Z_i)$ with distribution G. From Lemma 2.9,

$$M_p = \max\left\{\mathrm{VaR}_{\alpha_*}(T), \mathrm{TVaR}_{\beta_*}(Y)\right\}. \tag{34}$$

It is clear that the df G of Y is continuous and strictly increasing on its support. Let $0 < G(M_p) < 1$. By inspection of the Table in proof of Lemma 2.9 we are in the situation of the cases (1a), (1b), (1c) and (1d). We observe that $\beta_* = G(M_p)$ and

$$M_p \;=\; \mathrm{VaR}_{\beta_*}(Y) = \mathrm{VaR}_{\beta_*}\left(TVaR_U\left(\sum_{i=1}^d Z_i\right)\right) = TVaR_{\beta_*}\left(\sum_i Z_i\right). \tag{35}$$

Thus, for all $0 < \beta_* < 1$, or, equivalently, $\frac{p+p_f-1}{p_f} < \alpha_* < \frac{p}{p_f}$ one has that $M_p =$

$TVaR_{\beta_*}\left(\sum_i Z_i\right)$. When $G(M_p) = 0$, we are always in the cases (1b), (3b) and (4b) so that $M_p = \mathrm{VaR}_{\alpha_*}(T)$ with $\beta_* = 0$ and $\alpha_* = \frac{p}{p_f}$. (this is also clear from the fact that in this case $VaR_0\left(Y\right) = -\infty$). When $G(M_p) = 1$, we are either in the cases (1b), (3b) and (4b) or in the cases (1a), (2a), (3a) and (4a). In the first situation, it follows from inspection of the table again that $M_p = \mathrm{VaR}_{\alpha_*}(T)$ with $\alpha_* = \frac{p}{p_f}$ and note that $\mathrm{VaR}_{\alpha_*}(T) > \mathrm{VaR}_1(Y) (= TVaR_1\left(\sum_i Z_i\right))$. In the second situation, one always has $\mathrm{VaR}_{\beta_*}(Y) > \mathrm{VaR}_{\alpha_*}(T)$ with $\beta_* = 1$ and $\alpha_* = \frac{p-p_f}{1-p_f}$. Hence, one concludes that if $G(M_p) = 1$, then $M_p = \mathrm{VaR}_{\alpha_*}(T)$ with $\alpha_* = \frac{p-p_f}{1-p_f}$ unless $\mathrm{VaR}_1(Y) > \mathrm{VaR}_{\frac{p}{p_f}}(T)$ in which case, $M_p = TVaR_1\left(\sum_i Z_i\right)$.

The proof of the expression for m_p follows by applying Lemma 2.9 again, where we now take Y as $Y = \sum_{i=1}^{d} LTVaR_U(Z_i)$. □

B Rearrangement Algorithm of Embrechts, Puccetti, and Rüschendorf (2013)

The rearrangement algorithm (RA) can be seen as a method to construct dependence between the variables X_j $(j = 1, 2, \ldots, d)$, such that the distribution of the variance of the sum $S = X_1 + ... + X_d$ is as small as possible (more generally, the sum becomes as small as possible in convex order).

Assume that the variables X_j are discretized and take n values put in a matrix $\mathbf{A}$ randomly:

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

Recall that 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.

Rearrangement algorithm

1. For i from 2 to d, Make the i^{th} column anti-monotonic with the sum of the other columns.

2. Start again from column 1, and make it anti-monotonic with the sum of the columns from 2 to d.

At each step of this algorithm, we make column j anti-monotonic, so that the columns, say X_j before rearranging and $\tilde{X}_j$ after rearranging verify obviously:

$$var(\sum_{i=1}^{d} X_i) \geqslant var(\sum_{i=1}^{d} \tilde{X}_j).$$

Indeed

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

and its minimum when X_j is anti-monotonic with $\sum_{i\neq j} X_i$. At each step of the algorithm the variance decreases,[6] it is bounded from below (by 0) and thus converges. If it gets to 0, we have found a perfect mixability situation in which the dependence makes the sum constant. Otherwise the algorithm will converge to a local minimum. There is no guarantee that this minimum is the minimum of the variance of the sum optimized over all dependence structure, this minimum may depend on the starting point. However, in practice it turns out to converge very fast and to approximate the situation of complete mixability in a few iterations. It works remarkably well for the case of homogeneous portfolio or close distributions among the X_j.

This algorithm was successfully used to compute (approximate) VaR bounds on the sum of n dependent risks with given marginal distributions by Embrechts, Puccetti, and Rüschendorf (2013) by applying the RA to the largest rows of the matrix that depicts the comonotonic structure between the risks.

In the paper we make use of it on submatrices throughout algorithms to obtain the minimum variance, minimum TVaR and bounds on VaR.

C Algorithm for Computing the Minimum VaR

1. Compute m_1 and m_2, as in Step 1 in the algorithm for the maximum VaR (see 3.5).

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

3. Apply the RA to the last $\lceil \beta_* \ell_u \rceil$ 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 $\lceil\cdot\rceil$ denotes that we take the ceiling. Observe that $\lceil \beta_* \ell_u \rceil = \ell_u - (k + m_* - \ell_f) = \ N - k - m_*$, where $m_* := \lceil \alpha_* \ell_f \rceil$, and note that $m_1 \leqslant m_* \leqslant m_2$.

4. If $m_* > m_1$. Apply the RA the last $b_* := N - k - m_*$ rows of the untrusted part $\mathcal{U}_\mathcal{N}$ of the matrix M. By abuse of notation, denote the rearranged sums in the untrusted part as S_N^u. This is the dependence that potentially achieves the minimum VaR by making $LTVaR_{\frac{\ell_u - b_*}{\ell_u}}(S_N^u)$ as close as possible to $VaR_{\frac{\ell_u - b_*}{\ell_u}}(S_N^u)$. To compute this minimum 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 $\tilde{s}_k$.

Illustration of the algorithm for obtaining minimum VaR in the example with $d = 3$, $N = 8$, $k = 3$ with the same matrix M given in (21). In this case, $\alpha_* = 1.3 > \alpha_1$, so that when applying Proposition 2.10, the theoretical minimum VaR is equal to $\mathrm{LTVaR}_{\frac{p-p_f\alpha_*}{1-p_f}} = 3.75$. In the algorithm, $m_* = \lceil \alpha_* \ell_f \rceil = \lceil 0.12 \rceil = 1$ and the minimum VaR is likely (but not certain) to be obtained for $m_* = 1$ (which corresponds to applying the RA to the last $N - k - m_* = 4$ rows of the untrusted matrix). One can show that $m_* = 1$ is the optimal value, indeed, by applying Step 3 to the last $N - k - m$ rows of the $\mathcal{U}_\mathcal{N}$ for $m_1 \leqslant m \leqslant m_2$.

[6] 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 minimizes the variance.

Indeed, we first apply the RA on the last two rows of the $\mathcal{U}_N$ (case $m = 3 = m_2$, so that $N-k-m = 2$ rows on the $\mathcal{U}_N$ will be rearranged), which yields a rearranged matrix and the corresponding sums $\tilde{S}$:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \tilde{S} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

The sums are now $\{8, 8, 3, 2, 2, 10, 7, 4\}$, so that the first VaR is equal to VaR $= 7$ (as the 4^{th} largest value).

Next, we apply the RA to the last three rows of the $\mathcal{U}_N$ (case $m = 2$, so that $N-k-m = 8-3-2 = 3$ rows on the $\mathcal{U}_N$ are rearranged), which gives:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \quad \tilde{S} = \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}.$$

The sums are now $\{8, 8, 3, 3, 3, 2, 10, 7\}$, so that the second VaR is equal to VaR $= 7$ (as the 4^{th} largest value).

Applying the RA to the last four rows of $\mathcal{U}_N$ (corresponding to the case $m = 1$) gives rise to:

$$\begin{bmatrix} 3 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \\ 0 & 2 & 2 \end{bmatrix}, \quad \tilde{S} = \begin{bmatrix} 4 \\ 3 \\ 4 \\ 4 \end{bmatrix}.$$

The sums are now $\{8, 8, 3, 4, 4, 4, 3, 10\}$ so that the third VaR is equal to VaR $= 4$ (as the 4^{th} largest value); it is smaller than the others so we keep it.

The last step consists of applying the RA to the five rows of the $\mathcal{U}_N$ (case $m = 0$), which we have already done when computing the minimum variance and the minimum TVaR. We find that the sums are now $\{8, 8, 3, 5, 5, 5, 5, 5\}$, so that the VaR computed is VaR $= 5$ (as the 4^{th} largest value). This value is larger than in the previous case, and we cannot improve the VaR obtained previously.

References

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009): "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, 44(2), 182–198.

Alexander, C., and J. M. Sarabia (2012): "Quantile uncertainty and Value-at-Risk model risk," *Risk Analysis*, 32(8), 1293–1308.

Barrieu, P., and G. Scandolo (2013): "Assessing financial model risk," *arXiv preprint: 1307.0684*.

Bernard, C., P. Boyle, and S. Vanduffel (2014): "Explicit representation of cost-efficient strategies," *Finance*, 25, 5–66.

Bernard, C., J. Chen, and S. Vanduffel (2014): "Optimal portfolios under worst case scenarios," *Quantitative Finance, Special issue on behavioral finance*, 14, 657–671.

Bernard, C., X. Jiang, and S. Vanduffel (2012): "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.

Bernard, C., X. Jiang, and R. Wang (2014): "Risk aggregation with dependence uncertainty," *Insurance: Mathematics and Economics*, 54, 93–108.

Bernard, C., Y. Liu, N. MacGillivray, and J. Zhang (2013): "Bounds on capital requirements For bivariate risk with given marginals and partial information on the dependence," *Dependence Modelling*, 1, 37–53.

Bernard, C., L. Rüschendorf, and S. Vanduffel (2013): "VaR bounds with a variance constraint," *Working Paper*.

Cont, R. (2006): "Model uncertainty and its impact on the pricing of derivative instruments," *Mathematical finance*, 16(3), 519–547.

Czado, C. (2010): "Pair-copula constructions of multivariate copulas," in *Copula Theory and Its Applications*, pp. 93–109. Springer.

Denuit, M., C. Genest, and É. Marceau (1999): "Stochastic bounds on sums of dependent risks," *Insurance: Mathematics and Economics*, 25(1), 85–104.

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

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

Kerkhof, J., B. Melenberg, and H. Schumacher (2010): "Model risk and capital reserves," *Journal of Banking & Finance*, 34(1), 267–279.

McNeil, A. J., R. Frey, and P. Embrechts (2010): *Quantitative risk management: concepts, techniques, and tools*. Princeton university press.

Meilijson, I., and A. Nádas (1979): "Convex majorization with an application to the length of critical paths," *Journal of Applied Probability*, pp. 671–677.

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

Rüschendorf, L. (1991): "Bounds for distributions with multivariate marginals," *Lecture Notes-Monograph Series*, pp. 285–310.

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

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

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