A clustering approach and a rule of thumb for risk aggregation

F. Marta L. Di Lascio[a], Davide Giammusso, Giovanni Puccetti[b,*]

[a] Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy
[b] Department of Economics, Management and Quantitative Methods, University of Milano, Italy

Abstract

The problem of establishing reliable estimates or bounds for the (T)VaR of a joint risk portfolio is a relevant subject in connection with the computation of total economic capital in the Basel regulatory framework for the finance sector as well as with the Solvency regulations for the insurance sector. In the computation of total economic capital, a financial institution faces a considerable amount of model uncertainty related to the estimation of the interdependence amongst the marginal risks. In this paper, we propose to apply a clustering procedure in order to partition a risk portfolio into independent subgroups of positively dependent risks. Based on available data, the portfolio partition so obtained can be statistically validated and allows for a reduction of capital and the corresponding model uncertainty. We illustrate the proposed methodology in a simulation study and two case studies considering an Operational and a Market Risk portfolio. A rule of thumb stems from the various examples proposed: in a mathematical model where the risk portfolio is split into independent subsets with comonotonic dependence within, the smallest VaR-based capital estimate (at the high regulatory probability levels typically used) is produced by assuming that the infinite-mean risks are comonotonic and the finite-mean risks are independent. The largest VaR estimate is instead generated by obtaining the maximum number of independent infinite-mean sums.

Keywords: (T)VaR bounds, Model uncertainty, Hierarchical clustering, Risk aggregation, Operational risk, Market Risk
AMS 2010 Subject Classification: 62P05 (primary), 91G60, 91B30 (secondary)

1. Motivation

The problem of establishing reliable estimates or bounds for the (T)VaR of a joint risk portfolio is a relevant subject in connection with the computation of total economic capital in the Basel regulatory framework for the finance sector as well as with the Solvency regulations for the insurance sector. For instance, the new regime for banks using internal models (Basel Committee on Banking Supervision, 2016) prescribes the use of TVaR in order to compute capital requirements for Market Risk, and also requires a rigorous backtesting based on VaR.

The computation of the (T)VaR of an aggregate position calls for a statistical model for the individual risk factors as well for their interdependence. The statistical estimation of the marginal distributions of the standalone risks is mature and relatively robust even in the case of a small data sample, and usually occurs in the form of historical data, or models chosen in a stress-testing environment or resulting from a simulating mechanism. On the contrary, the estimation of a multivariate dependence structure (often performed via the notion of a copula) is a much more challenging task where the scarcity of joint observations often leads to a large statistical uncertainty.

If one assumes full knowledge of the distributions of the individual risk factors held, such dependence uncertainty can be quantified using the Rearrangement Algorithm (RA) described in Embrechts et al. (2013). However, the upper VaR bound produced by the RA, despite being mathematically sharp, is too large for practical use and is attained by a seemingly unrealistic multivariate scenario; see Aas and Puccetti (2014).

Recent research has focused on methods to reduce dependence uncertainty based on extra (statistical) information to be added on the top of the knowledge of the marginal distributions. For instance, in Sect. 4 of Embrechts et al. (2013) it is shown that higher order (typically two-dimensional) marginals information on the joint portfolio, when available, may lead to improved VaR bounds. The worst VaR value can be similarly reduced by estimating the values of the copula on some subset of its domain (Bernard et al., 2013) or putting a variance constraint on the total position (Bernard et al., 2015). An approach using factor models is provided in Bernard et al. (2017).

Intuitively, one might think that assuming that the risks are *positively dependent* or *positively correlated* would imply a considerably smaller value for the worst VaR estimate. It is however shown in Bignozzi et al. (2015) that additional positive dependence information added on top of the marginal distributions (for instance via the notions of PLOD risks) does not substantially lower the estimate of the worst-possible VaR. In particular, the assumption of a lower bound on the copula of the risk portfolio does not allow one to lower the worst VaR estimate below the sum of marginal VaRs; see Puccetti et al. (2016).

A different approach combining prior information, scarce

*Corresponding author. Address: Department of Economics, Management and Quantitative Methods, Via Conservatorio 7, 20122 Milano

Email addresses: Marta.DiLascio@unibz.it (F. Marta L. Di Lascio), davide.giammusso@gmail.com (Davide Giammusso), giovanni.puccetti@unimi.it (Giovanni Puccetti)

observations and expert opinion in order to reduce estimation uncertainty is described in Arbenz and Canestraro (2012). The problem of portfolio selection under dependence uncertainty is treated in Pflug and Pohl (2017).

In this paper, we start from the observation (Puccetti et al., 2017) that in practice some subgroups of the risk portfolio can be assumed to be independent. Hence, we propose to apply a clustering procedure to partition a risk portfolio into independent clusters with (maximal) positive dependence within the risks of each subgroup. The portfolio partition so obtained can be statistically validated and allows for a reduction of the estimate of economic capital, and the corresponding model uncertainty. We will illustrate the methodology in two case studies related to Operational and Market Risk.

A further motivation for other application areas of the clustering approach discussed here comes from the paper Prettenthaler et al. (2017), where the authors quantify the VaR-based diversification potential for flood risk in Europe by identifying pools of countries that should group together if an overall agreement on risk management cannot be achieved.

VaR diversification

If only the marginal distributions of a risk portfolio are fixed, one can always find a dependence structure (copula) under which the VaR is superadditive, i.e. the VaR of the sum of the marginals is greater than the sum of marginal numbers; see the various examples given in Embrechts et al. (2014).

If the tails of the marginal distributions are extremely heavy (i.e. they have infinite mean), then also diversification increases portfolio riskiness. In other words, assuming that all the marginal risks of a given portfolio are independent does *not* always yield a safe lower bound on a VaR-based economic capital. This fallacy of VaR has been identified in the identical marginals setting since the examples provided in Embrechts et al. (2002). Then, it has been formalized and extended to different dependence scenarios in the paper Ibragimov (2009); see also Mainik et al. (2013), Ibragimov and Prokhorov (2016) and the references given therein. A comprehensive review of these results is given in Ibragimov et al. (2015).

Whereas the case of homogeneous risks has been studied under a variety of dependence assumptions, the case of *inhomogeneous* summands seems to be more intricate to fully characterize. Stemming from the various examples proposed in this paper, a novel conclusion can be formulated in the form of a simple rule of thumb: if a risk portfolio is divided into independent subsets with comonotonic dependence within, *the smallest VaR-based capital estimate (at the high regulatory probability levels typically used) is produced by assuming that the infinite-mean risks are comonotonic and the finite-mean risks are independent. The largest VaR estimate is instead generated by obtaining the maximum number of independent infinite-mean sums.*

Since a clustering procedure produces several possible outcomes that can be equally assessed as statistically relevant based on the available observations, this rule of thumb can help to choose a specific model amongst all the feasible partitions of the same risk portfolio.

A discussion on infinite-mean models

The overall analysis carried out in this paper relies on the assumption that the marginal distributions of the risk portfolios under study are known. In particular, the proposed rule of thumb assumes the existence and inclusion of marginal models with an infinite mean.

A large number of studies in economics, finance and insurance have documented financial and economic variables with infinite mean; see for instance the papers Nešlehová et al. (2006) and Chavez-Demoulin et al. (2006). Without a doubt, infinite-mean models arise quite often for certain kinds of data, particularly insurance claim data and operational loss data; see Chavez-Demoulin et al. (2016) and Section 5. However, the use of infinite-mean models in practice remains controversial: it is not the risks themselves that have an infinite-mean, but rather certain models we might choose to describe them.

Even if financial institutions will generally have historical data or a simulating mechanism to estimate the margins, there is always an element of statistical uncertainty and discretion in the choice of the marginal models. The inclusion of even a single infinite-mean model has a significant effect on the quantification of economic capital.

Infinite-mean models generally produce unreliable or huge aggregate economic capitals, they do not allow for the use of any coherent risk measure like the TVaR, they make the lower and upper bounds produced by the RA too wide to make sense in application.

A way out of the infinite-mean world is to take into account notional upper limits to loss distribution: no loss can destroy more than a bank's or a company's entire value. It may also be possible to choose alternative models where the tail mimics a power tail within the sample and then decays more quickly at very extreme loss levels. A very good paper from which to enlarge this discussion is Cirillo and Taleb (2016), where a *no-loss-is-infinite* principle inspires a smooth procedure to provide TVaR-based capital allocations in the presence of apparently infinite-mean models. Also, if an infinite mean is statistically identified, for a financial institution it might be preferable to transfer part of the risk rather than to manage it.

From a mathematical viewpoint, the message of this paper is that infinite-mean models in an inhomogeneous risk portfolio deliver higher economic capitals if assumed to be independent rather than comonotonic. As the use of infinite-mean models is potentially disruptive, a financial institution should carefully consider their inclusion and treatment at the governance level.

Summary

The paper is organized as follows. In Section 2 we formalize our mathematical model and set our notation, while in Section 3 we describe the proposed clustering methodology. In Section 4 we illustrate the above stated rule of thumb with some toy Pareto models. Then, we provide two case studies of clustering selection with application to an Operational Risk (Section 5) and a Market Risk (Section 6) portfolio. In Section 7 we draw our conclusions, relegating the more technical mathematical results and computational details to the Appendix.

2. Mathematical framework and notation

We assume that a financial institution holds a number of d risk positions over a predetermined time horizon, represented by the random variables $X_1, \dots, X_d$ over some probability space $(\Omega, \mathcal{F}, P)$. The institution's aggregate exposure is given by the sum of the individual risk positions

$$X_d^+ = X_1 + \cdots + X_d.$$

The total position X_d^+ is then mapped into the economic capital $\mathrm{VaR}_\alpha(X_d^+)$ or $\mathrm{TVaR}_\alpha(X_d^+)$.

The *Value-at-Risk* (VaR) of a loss random variable X, computed at the probability level $\alpha \in (0, 1)$, is the α-quantile of its distribution, defined as

$$\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) > \alpha\},$$

where $F_X(x) = P(X \leq x)$ is the distribution function of X.

For a random variable X with finite mean, the *Tail-Value-at-Risk* (TVaR), computed at the probability level $\alpha \in (0, 1)$, is defined as

$$\mathrm{TVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_q(X)\, dq.$$

The TVaR, often referred to as the Expected Shortfall (ES), is simply the average of all VaR values between the probability levels α and 1.

The computation of the (T)VaR of the aggregate position X_d^+ calls for a statistical model for the joint distribution of the risk vector $(X_1, \dots, X_d)$, in the form of marginals information regarding the risk factors and of a dependence structure coupling them. It is realistic to assume that the marginal distributions $F_1, \dots, F_d$ of the individual random variables $X_1, \dots, X_d$ are known by the financial institution. On the contrary, the estimation of a multivariate dependence structure, typically performed via the notion of a copula (see for instance Ch. 7 in McNeil et al., 2015), requires a dataset of joint occurrences that is seldom (if ever) available. As a practical example, consider the six-dimensional portfolio of DNB, the largest Norwegian bank, which has been extensively studied in Aas and Puccetti (2014). A proper data-based statistical estimation of a six-dimensional copula for a portfolio including Credit (X_1), Market (X_2), Ownership (X_3), Operational (X_4), Business (X_5), and Insurance (X_6) risk is simply out of reach. As a result, the bank faces a considerable *model risk* related to the choice of the multivariate dependence structure (copula) on top of its marginals information.

A successful numerical technique to assess the robustness of a benchmark model with respect to dependence uncertainty is the Rearrangement Algorithm (RA)[1] described in Embrechts et al. (2013). The RA evaluates

$$\underline{\mathrm{VaR}}_\alpha(X_d^+) = \inf\{\mathrm{VaR}_\alpha(X_d^+) : X_j \sim F_j, 1 \leq j \leq d\}, \tag{2.1a}$$

$$\overline{\mathrm{VaR}}_\alpha(X_d^+) = \sup\{\mathrm{VaR}_\alpha(X_d^+) : X_j \sim F_j, 1 \leq j \leq d\}, \tag{2.1b}$$

that represent the best and worst possible VaR for the total position X_d^+, given solely the knowledge of the marginal distributions $F_1, \dots, F_d$. Similarly, the RA is able to compute corresponding best and worst possible values for the TVaR as

$$\underline{\mathrm{TVaR}}_\alpha(X_d^+) = \inf\{\mathrm{TVaR}_\alpha(X_d^+) : X_j \sim F_j, 1 \leq j \leq d\}, \tag{2.2a}$$

$$\overline{\mathrm{TVaR}}_\alpha(X_d^+) = \sup\{\mathrm{TVaR}_\alpha(X_d^+) : X_j \sim F_j, 1 \leq j \leq d\}. \tag{2.2b}$$

A benchmark probabilistic model for the risk portfolio is represented by *comonotonic* scenario, under which the portfolio individual risks X_j are all assumed to be perfectly positively dependent, i.e. almost surely increasing functions of a common random factor. Standard references on comonotonicity are Dhaene et al. (2002) and McNeil et al. (2015, Sec. 7.2.1). The comonotonic model implies additivity of the joint (T)VaR positions, that is

$$\mathrm{VaR}_\alpha(X_d^+) = \mathrm{VaR}_\alpha(X_1) + \cdots + \mathrm{VaR}_\alpha(X_d), \tag{2.3a}$$

$$\mathrm{TVaR}_\alpha(X_d^+) = \mathrm{TVaR}_\alpha(X_1) + \cdots + \mathrm{TVaR}_\alpha(X_d). \tag{2.3b}$$

Since TVaR is a *coherent* risk measure, the worst-case bound $\overline{\mathrm{TVaR}}_\alpha(X_d^+)$ coincides with the sum of marginal numbers, i.e.

$$\overline{\mathrm{TVaR}}_\alpha(X_d^+) = \mathrm{TVaR}_\alpha(X_1) + \cdots + \mathrm{TVaR}_\alpha(X_d).$$

On the contrary, for general marginal distributions the worst-case VaR bound $\overline{\mathrm{VaR}}_\alpha(X_d^+)$ is (considerably, for risks with infinite mean) larger than the sum of marginal VaR estimates. In the paper Embrechts et al. (2014), many examples are reported where

$$\overline{\mathrm{VaR}}_\alpha(X_d^+) >> \mathrm{VaR}_\alpha(X_1) + \cdots + \mathrm{VaR}_\alpha(X_d).$$

To reduce the worst possible (T)VaR estimate, in this paper we propose to apply a clustering procedure in order to partition a risk portfolio into independent clusters with (maximal) positive dependence within the risks of each subgroup.

Formally, we assume that the marginal components of a d-dimensional risk vector $(X_1, \dots, X_d)$ have fixed distributions $F_1, \dots, F_d$ and the risk vector is split into k subgroups I_i. Let $\mathcal{I} = \{I_1, \dots, I_k\}$ be a partition of $\{1, \dots, d\}$, that is $\bigcup_{i=1}^k I_i = \{1, \dots, d\}$ with $I_i \cap I_j = \emptyset$, $i \neq j$. Let $X_{I_i} = (X_j, j \in I_i)$ denote the risk subvector of the i-th subgroup. For a given partition $\mathcal{I}$, denote by

$$\mathrm{VaR}_\alpha^{\mathcal{I}}(X_d^+) \quad \text{and} \quad \mathrm{TVaR}_\alpha^{\mathcal{I}}(X_d^+) \tag{2.4}$$

the VaR and the TVaR of the aggregate position X_d^+ under the assumption that:

> the risk subvectors $X_{I_1}, \dots, X_{I_k}$ are *independent* while the risks $(X_j, j \in I_i)$ within each subgroup are *comonotonic*; see Figure 1.

In the applications that follow, we will apply a clustering data-driven procedure in order to select feasible choices for the partition $\mathcal{I}$ and we will see that the values $\mathrm{VaR}_\alpha^{\mathcal{I}}(X_d^+)$ and $\mathrm{TVaR}_\alpha^{\mathcal{I}}(X_d^+)$ will deliver a large reduction of (T)VaR-based economic capital and of the corresponding model uncertainty.

[1] See `https://sites.google.com/site/rearrangementalgorithm/`.

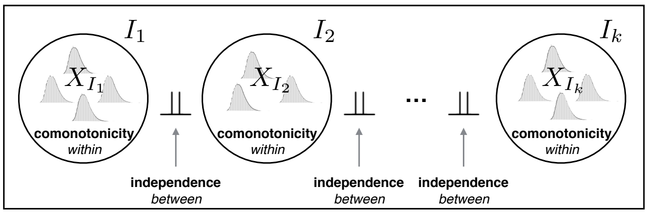

Figure 1: Partitioning a risk portfolio into independent subgroups $I_1, \dots, I_k$ with comonotonic dependence within the risks of each subgroup.

3. The hierarchical clustering methodology

In this section, we illustrate a clustering, data-driven statistical procedure that allows one to split a risk portfolio into independent subsets with (close to) maximal dependence within the subsets so as to fit the mathematical framework described so far.

Clustering is a data-driven method that attempts to discover structures in data, grouping together observations or variables into sets, called clusters. In our work we focus the attention on the clustering of (random) variables and on the hierarchical clustering in the agglomerative approach, as described in Everitt et al. (2011). This method creates a hierarchy of nested clusters and, in the agglomerative version, the procedure starts with each single variable forming a cluster and, at every step, it proceeds by merging one pair of clusters at a time. The procedure stops when the last two clusters are merged to form the whole group of variables.

In the hierarchical clustering context, the decision on which clusters should be merged is based on a combination of two measures: a measure of *dissimilarity* between two *variables*, and a measure, called *linkage rule*, that specifies the dissimilarity between two *clusters* as a function of the pairwise dissimilarities of variables in the clusters. The choice of these two measures influence the resulting clustering outcome: the variables belonging to the same cluster are similar in some sense, depending on the dissimilarity measure/linkage rule used.

In our set-up we employ a correlation-based distance as measure of pairwise dissimilarity, that considers two random variables X_r and X_j to be similar (dissimilar) if their observations are highly (slightly) correlated. As a dissimilarity measure, we use

$$d(X_r, X_j) = \sqrt{2(1 - \rho_S(X_r, X_j))}, \qquad (3.1)$$

where $\rho_S(X_r, X_j)$ is the sample Spearman's rank correlation coefficient computed as the standard sample Pearson correlation coefficient between the *ranked* observations from X_r and X_j; see p. 404 in Everitt and Skrondal (2010). The dissimilarity measure (3.1) captures the comonotonic dependence relationship between the two compared variables. In fact, if X_r and X_j are comonotonic, then $\rho_S(X_r, X_j) = 1$ and $d(X_r, X_j) = 0$. As a result, the final clustering will have maximum within-group dependence.

As for the way to evaluate the dissimilarity between two clusters of variables (or between a cluster of variables and a single variable), we employ the well-known and most used linkages: the complete, the single, and the average linkage. All these provide a way to compute the dissimilarity between two clusters of variables, say A and B, as a function of all pairwise dissimilarities between the variables in A and the variables in B. To be specific, the single, the complete and the average linkage rule compute the dissimilarity between A and B, respectively as the smallest, the largest and the average dissimilarity amongst all pairwise dissimilarities between the variables in A and those in B. For the mathematical details of the three linkages we refer to Everitt et al. (2011) and James et al. (2013). Here, we employ all three methods to evaluate the robustness of the analysis and the effect of the linkage on the hierarchy of nested clustering. The interested reader can retrieve the R code and a full description of the performed analyses via the RunMyCode online repository[2].

The sequence of nested partitions obtained through the hierarchical clustering procedure is visualized by means of a tree-based representation of the variables called a *dendrogram*. The dendrogram illustrates the process and the partition produced at each step of the clustering procedure. As we move higher up the tree, branches themselves merge, either with leaves or other branches. The height of this fusion, as measured on the vertical axis through the chosen linkage rule and dissimilarity measure, indicates how similar the two (clusters of) variables are: variables that merge at the very bottom of the tree are quite similar to each other, whereas variables that merge close to the top will tend to be quite different. A specific partition in the nested sequence produced is selected by cutting the dendrogram at a particular height.

Thus, the hierarchical clustering approach does not require to choose *a priori* the number of clusters, k, which is an added advantage over other unsupervised classification methods, such as k-means clustering, self-organized maps, etc. The ideal number of clusters can be selected by looking at the dendrogram. Moreover, the hierarchical clustering methodology does not require a starting classification which is instead needed, for example, in other clustering methods (e.g. k-means and mixture-based clustering). In general, a drawback of the hierarchical method is its computational cost, which is relevant only for high-dimensional data and does not affect our analysis.

4. A simulation study

Before illustrating the proposed clustering procedure in some real case studies, it is illustrative to compute the (T)VaR values (2.4) in a simulation study considering all the possible partitions of a risk portfolio. We consider six five-dimensional ($d = 5$) risk portfolios where each risk X_j follows a Pareto(θ_j) marginal distribution

$$P(X_j \le x) = 1 - (1 + x)^{-\theta_j}, \; x \ge 0, \qquad (4.1)$$

for some tail parameters $\theta_j > 0$. Recall that for the Pareto model (4.1) the mean of X_j is infinite when $\theta \le 1$. We consider the following six different portfolios:

[2]See `http://www.runmycode.org/companion/view/2955`.

P1. A portfolio of homogeneous finite-mean risks with

$$\theta_j = 3, 1 \leq j \leq 5;$$

P2. A portfolio of homogeneous infinite-mean risks with

$$\theta_j = 0.98, 1 \leq j \leq 5;$$

P3. An inhomogeneous portfolio of mixed risks with

$$\theta_1 = 0.95, \theta_2 = 0.98, \theta_3 = 1.6, \theta_4 = 4, \theta_5 = 5.$$

P4. An inhomogeneous portfolio of mixed risks with

$$\theta_1 = 0.6, \theta_2 = 0.98, \theta_3 = 1.6, \theta_4 = 4, \theta_5 = 5.$$

P5. An inhomogeneous portfolio of mixed risks with

$$\theta_1 = 0.98, \theta_2 = 1.5, \theta_3 = 1.6, \theta_4 = 4, \theta_5 = 5.$$

P6. An inhomogeneous portfolio of mixed risks with

$$\theta_1 = 1.5, \theta_2 = 1.5, \theta_3 = 1.6, \theta_4 = 4, \theta_5 = 5.$$

In Table 1 we report estimates for:

- the minimum and maximum possible estimates for the VaR of the total loss exposure, computed over the set $\mathcal{P}_d$ of all possible partitions of $\{1, \ldots, d\}$, and defined as

$$\underline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+) := \min_{\mathcal{I} \in \mathcal{P}_d} \mathrm{VaR}^{\mathcal{I}}_\alpha(X_d^+), \quad (4.2a)$$

$$\overline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+) := \max_{\mathcal{I} \in \mathcal{P}_d} \mathrm{VaR}^{\mathcal{I}}_\alpha(X_d^+); \quad (4.2b)$$

  In the cases considered ($d = 5$), the total number of possible partitions is equal to 52.

- for portfolios P1 and P6, the minimum and maximum possible estimates for the TVaR of the total loss exposure, computed over all the possible partitions, and similarly defined as

$$\underline{\mathrm{TVaR}}^{\mathcal{P}}_\alpha(X_d^+) := \min_{\mathcal{I} \in \mathcal{P}_d} \mathrm{TVaR}^{\mathcal{I}}_\alpha(X_d^+), \quad (4.3a)$$

$$\overline{\mathrm{TVaR}}^{\mathcal{P}}_\alpha(X_d^+) := \max_{\mathcal{I} \in \mathcal{P}_d} \mathrm{TVaR}^{\mathcal{I}}_\alpha(X_d^+); \quad (4.3b)$$

- the best and worst possible (T)VaR values for the total position X_d^+, given solely the knowledge of the marginal distributions and computed via the RA; see (2.1) and (2.2);

- the comonotonic values (2.3) obtained as the simple sum of marginal numbers, and corresponding to the trivial partition $\{1, 2, 3, 4, 5\}$.

For the homogeneous risk portfolio with finite-mean risks (P1), the lowest capital estimate is obtained when all the risks are assumed to be independent ($\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ is the best partition), while the greatest capital stems from a full comonotonic model ($\{1, 2, 3, 4, 5\}$ is the worst partition). For the homogeneous portfolio with infinite-mean risks (P2), these results are reversed. The figures for portfolios P1-P2 are similar to those obtained in Ibragimov (2009, Th. 4.1-4.2) for convolutions of stable distributions. Within our mathematical set-up, it is easy to prove the complete ordering of partitions for Pareto-like homogeneous portfolios; see Appendix A for the mathematical details.

For an inhomogeneous portfolio mixing finite and infinite-mean risks, the situation is more intricate and has not been formalized in the literature. From the results in Table 1 and the examples illustrated in the remainder of the paper, a simple rule of thumb for risk aggregation emerges. The least VaR-based capital estimate (at the high probability levels used) is produced by assuming that the infinite-mean risks are comonotonic and the finite-mean risks are independent. For example, the best partition for portfolios P3 and P4 is $\{1, 2\}, \{3\}, \{4\}, \{5\}$.

The largest VaR estimate over all the partitions is generated by obtaining the maximum number of independent infinite-mean sums, that is by assigning each infinite-mean risk to a different subgroup. Thus, for portfolios P3 and P4, the highest VaR-based capital estimate is obtained when the infinite-mean risks X_1 and X_2 belong to different subgroups and are then assumed to be independent. The behaviors of the inhomogeneous portfolios P5 (only one infinite-mean risk) and P6 (no infinite-mean risks) are analogous to that of P1 and are consistent with our rule of thumb. However, the above stated rule of thumb heavily depends on the probability levels used, which must be sufficiently large; see Section 5 on this.

Since the VaR estimate for an aggregate risk is asymptotically driven by the heaviest tail (see Proposition A.1.), the inclusion of even a single infinite-mean model has a significant impact on the quantification of economic capital. First, an infinite marginal mean implies that the TVaR of the aggregate sum is not defined. Second, in the presence of infinite-mean models the VaR will tend to deliver extremely large values, especially for the very high confidence level prescribed by regulation.

One can immediately appreciate the magnifying effect when comparing portfolio P5 to portfolio P6 in Table 1, where the inclusion of a single infinite-mean model more than doubles the corresponding best and worst VaR estimates at the 99% level. The inclusion of a more extremely heavy-tailed case ($\theta = 0.6$) in portfolio P4 drastically increases the VaR-based capital estimates deteriorating their accuracy, measured by larger confidence intervals.

In conclusion, the behavior of (T)VaR estimates amongst the possible partitions is inextricably intertwined with the selection of the marginal distributions and, in particular, their tail behaviors.

5. An Operational Risk case study

We now propose two real case studies in which to apply the clustering procedure described in Section 3: one belongs to the heavy-tailed, data-sparse realm of Operational Risk, the other

P1									
	$\underline{\mathrm{VaR}}_\alpha(X_d^+)$	$\underline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{VaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	1.8233	5.6554	[5.6397, 5.6707]	{1}, {2}, {3}, {4}, {5}	8.5720	[8.5338, 8.6109]	{1, 2, 3, 4, 5}	8.5721	14.6756
$\alpha = 0.975$	2.4199	7.0264	[7.0008, 7.0526]	{1}, {2}, {3}, {4}, {5}	12.1003	[12.0295, 12.1720]	{1, 2, 3, 4, 5}	12.0998	19.7897
$\alpha = 0.990$	3.6415	9.2439	[9.1912, 9.2965]	{1}, {2}, {3}, {4}, {5}	18.2092	[18.0567, 18.3628]	{1, 2, 3, 4, 5}	18.2079	28.6447
	$\underline{\mathrm{TVaR}}_\alpha(X_d^+)$	$\underline{\mathrm{TVaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{TVaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{TVaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	5.9894	8.1459	[8.1049, 8.1867]	{1}, {2}, {3}, {4}, {5}	15.3576	[15.2393, 15.4780]	{1, 2, 3, 4, 5}	15.3581	15.3581
$\alpha = 0.975$	7.7853	10.0471	[9.9753, 10.1196]	{1}, {2}, {3}, {4}, {5}	20.6481	[20.4390, 20.8634]	{1, 2, 3, 4, 5}	20.6496	20.6496
$\alpha = 0.990$	10.9104	13.2317	[13.0821, 13.3885]	{1}, {2}, {3}, {4}, {5}	29.8069	[29.3648, 30.2706]	{1, 2, 3, 4, 5}	29.8119	29.8119

P2									
	$\underline{\mathrm{VaR}}_\alpha(X_d^+)$	$\underline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{VaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	20.2607	101.3018	[100.3890, 102.2353]	{1, 2, 3, 4, 5}	123.0954	[122.1062, 124.0797]	{1}, {2}, {3}, {4}, {5}	101.3045	385.2407
$\alpha = 0.975$	41.1273	210.6203	[207.9323, 213.3341]	{1, 2, 3, 4, 5}	239.4450	[236.5916, 242.2609]	{1}, {2}, {3}, {4}, {5}	210.6379	786.6000
$\alpha = 0.990$	108.8486	544.3595	[533.3733, 555.5652]	{1, 2, 3, 4, 5}	588.4729	[577.4243, 599.6371]	{1}, {2}, {3}, {4}, {5}	544.2706	2011.3560

P3									
	$\underline{\mathrm{VaR}}_\alpha(X_d^+)$	$\underline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{VaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	22.4154	47.0963	[46.6884, 47.5069]	{1, 2}, {3}, {4}, {5}	54.2111	[53.7668, 54.6667]	{1}, {2, 3, 4, 5}	50.1153	121.8281
$\alpha = 0.975$	47.5702	94.8298	[93.6694, 96.0371]	{1, 2}, {3}, {4}, {5}	106.5601	[105.2982, 107.8572]	{1, 3, 4, 5}, {2}	101.3346	238.5554
$\alpha = 0.990$	126.4209	241.5839	[236.8072, 246.4360]	{1, 2}, {3}, {4}, {5}	263.8447	[258.7939, 268.9411]	{1, 3, 4, 5}, {2}	255.7386	584.8387

P4									
	$\underline{\mathrm{VaR}}_\alpha(X_d^+)$	$\underline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{VaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	146.3589	170.3672	[169.3407, 171.3924]	{1, 2}, {3}, {4}, {5}	193.1536	[192.0576, 194.2430]	{1 4 5}, {2 3}	174.0609	373.1082
$\alpha = 0.975$	466.8276	512.7423	[508.3446, 517.2904]	{1 2}, {3}, {4}, {5}	556.9200	[552.2811, 561.6270]	{1 3 4 5}, {2}	520.6063	992.6268
$\alpha = 0.990$	2153.2570	2265.9795	[2233.3220, 2297.6110]	{1 2}, {3}, {4}, {5}	2371.9938	[2338.6924, 2404.9576]	{1 3 4 5}, {2}	2282.7458	3796.9055

P5									
	$\underline{\mathrm{VaR}}_\alpha(X_d^+)$	$\underline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{VaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	20.2607	30.6411	[30.4228, 30.8659]	{1}, {2}, {3}, {4}, {5}	34, 0695	[33.7975, 34.3451]	{1, 2, 3, 4, 5}	34.0677	71.4667
$\alpha = 0.975$	42.1267	55.6864	[55.0941, 56.2909]	{1}, {2}, {3}, {4}, {5}	64.4557	[63.7459, 65.2037]	{1, 2, 3, 4, 5}	64.4595	127.9082
$\alpha = 0.990$	108.8486	128.0283	[125.7431, 130.4692]	{1}, {2}, {3}, {4}, {5}	149.8718	[147.1879, 152.5900]	{1, 2, 3, 4, 5}	149.8554	278.7862

P6									
	$\underline{\mathrm{VaR}}_\alpha(X_d^+)$	$\underline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{VaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{VaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	6.3681	16.6669	[16.5810, 16.7554]	{1} {2}, {3}, {4}, {5}	20.1757	[20.0482, 20.3046]	{1, 2, 3, 4, 5}	20.1749	41.5147
$\alpha = 0.975$	10.6959	25.3693	[25.1798, 25.5576]	{1}, {2}, {3}, {4}, {5}	33.0265	[32.7511, 33.3150]	{1, 2, 3, 4, 5}	33.0280	65.8104
$\alpha = 0.990$	20.5436	44.5180	[43.9817, 45.0640]	{1}, {2}, {3}, {4}, {5}	61.5505	[60.7545, 62.3554]	{1, 2, 3, 4, 5}	61.5457	119.8284
	$\underline{\mathrm{TVaR}}_\alpha(X_d^+)$	$\underline{\mathrm{TVaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	best partition	$\overline{\mathrm{TVaR}}^{\mathcal{P}}_\alpha(X_d^+)$	CI	worst partition	comonotonic value	$\overline{\mathrm{TVaR}}_\alpha(X_d^+)$
$\alpha = 0.950$	41.2347	44.3256	[42.8527, 48.6568]	{1}, {2}, {3}, {4}, {5}	61.0083	[58.8910, 67.2682]	{1, 2, 3, 4, 5}	61.6463	61.6463
$\alpha = 0.975$	65.5188	68.4279	[65.5100, 77.0906]	{1}, {2}, {3}, {4}, {5}	96.6062	[92.4150, 109.2116]	{1, 2, 3, 4, 5}	97.8895	97.8895
$\alpha = 0.990$	120.3731	122.1367	[114.9876, 143.7750]	{1}, {2}, {3}, {4}, {5}	175.7929	[165.4839, 207.4321]	{1, 2, 3, 4, 5}	179.0431	179.0431

Table 1: Estimates for the (T)VaR bounds defined in (2.1), (2.2), (2.3), (4.2), and (4.3) for the six Pareto risk portfolios described in Section 4. More detailed results (at the partitions level) for the homogeneous portfolios P1 and P2 are to be found in Appendix A, and full computational details in Appendix B.

to the more statistically manageable domain of Market Risk. Of course, the methodologies here described might well apply to the broader variety of risks to which financial and insurance regulations apply.

The current regulatory framework for OpRisk allows banks to opt for increasingly sophisticated approaches, the most mathematically involved of which being the Advanced Measurement Approach (AMA); see Basel Committee on Banking Supervision (2011). However, the Committee has decided to substitute all the current existing approaches (including the AMA) by one single standardised approach starting from Jan 1, 2022; we discuss this regulatory change below.

The AMA is commonly utilized via the Loss Distribution Approach (LDA) based on a loss-frequency plus loss-severity stochastic model, where financial institutions are given full freedom concerning the modeling assumptions used. The risk measure prescribed is the VaR of the 1-year aggregate exposure, at the $\alpha = 99.9\%$ probability level. For the LDA, the Basel framework supports the structuring of a financial institution into a number of business lines (BLs) and loss types (LTs); based on data size, aggregation of OpRisk losses to the business line level is widely chosen – similarly to what was done in Moscadelli (2004).

In this paper, we consider the database of OpRisk losses collected by Willis Professional Risks from public media and made available to us by Willis Tower Watson. The database consists

Business line (BL)	data size
Not Available (NA)	61
Corporate Finance	42
Trading & Sales	201
Retail Banking	233
Commercial Banking	210
Payment & Settlement	12
Agency Services	18
Asset Management	79
Retail Brokerage	49
Insurance (life)	17
Insurance (non-life)	35
Total	957

Table 2: Data size for each Business Line in the Willis dataset.

BL (initials)	i	ξ_i	β_i
CF + P&S + AS + INS	1	1.41	22.56
T&S	2	0.88	128.47
RBa	3	1.66	28.76
CB	4	1.20	83.49
AM + RBr	5	1.06	20.18

Table 3: Parameter values for the GPD-distributed risks. For our data analysis, we adjust the gross loss amounts for inflation to 2013 prices exactly as done in Chavez-Demoulin et al. (2016).

of 1,413 OpRisk events reported in the public media since 1970. For our analysis, we consider only the 957 available, inflation-adjusted, gross losses in Million GBP from 1974–2013 and the corresponding BLs as reported. Given the scarce availability to the (academic) public of real OpRisk datasets, we use this data even if it may not reflect all the features of an individual company or consortium dataset.

For the estimation of the marginal distributions of each business line, we follow a classical EVT-POT approach, for which we refer for instance to the textbook reference McNeil et al. (2015, Ch. 5). In order to achieve a significant number of observations leading to a more robust statistical analysis we have merged some BLs, discarded the observations with NA and hence worked with a total number of $d = 5$ BLs $X_1, \ldots, X_5$; see Table 3. A different aggregation scheme leading to different estimates is possible, see the paper Embrechts and Puccetti (2008) for the underlying methodological issues. Considering that the above cited Moscadelli study considered $d = 8$ BLs with a total amount of 47,000(!) observations, it is natural to consider what follows just as an illustrative example.

Data size left alone, the reliability of the POT method highly depends on the choice of the (sufficiently large) threshold beyond which a Generalized Pareto distribution (GPD) distribution is fitted. In our study, we select a different threshold for each BL so that the fit to a GPD distribution is good and the number of threshold exceedances is the same across all BLs. This choice has two immediate advantages: - it allows a proper application of the clustering procedure to follow within the framework of an EVT analysis of exceedances; - it makes our study not dependent on a subjective and possibly controversial choice of a threshold (typically the Achille's heel of the POT methodology).

Finally, for each BL we estimate the parameters of a GPD and we assume that

$$P(X_j \le x) = F_j(x) = 1 - \left(1 + \xi_j \frac{x}{\beta_j}\right)^{-1/\xi_j}, x \ge 0,\ 1 \le j \le d. \tag{5.1}$$

The estimated parameters for the $d = 5$ marginal distributions are collected in Table 3: recall that for the GPD model (5.1) whenever $\xi_j \ge 1$, the mean of X_j is infinite. The GPDs distribution are here assumed to hold over the entire loss domain $X_j \ge 0$.

We do not consider a full LDA approach (we do not model frequencies) for one practical and one mathematical reason. First, a stochastic model for the frequency of claims in each BL very much depends on one's individual dataset at hand and its full inclusion would distract the reader from the clustering procedure which is the real aim of this paper. Second, it is known that a total loss compound distribution inherits the tail properties of the single claim distribution (assumed to be homogeneous within each BL); this has similarly been done in the study carried out in Chavez-Demoulin et al. (2006).

Hence, we focus on the risk portfolio described in Table 3 and apply our clustering technique. All three linkage rules used provide the same final clusterings; see Figure 2. The partitions so-obtained, by varying the number of clusters from 2 to 4, are $(\{1,2,4,5\},\{3\})$, $(\{1,4,5\},\{2\},\{3\})$ and $(\{1,4\},\{2\},\{3\},\{5\})$, respectively.

The three nested partitions are reported in the middle part of Table 4 with the corresponding estimates of the aggregate VaR at different probability levels. All these partitions are valid models for economic capital calculation and all yield a huge reduction of capital estimate with respect to the worst possible VaR value $\overline{\mathrm{VaR}}_\alpha(X_d^+)$. The final model selection depends on the firm-wide governance and risk management, and can be delegated for instance to expert opinion.

Along with nested partitions as produced by the clustering procedure, we also report VaR estimates for the partitions that should yield the best possible VaR estimate $(\{1,3,4,5\},\{2\})$ and one of the worst VaR estimates $(\{1,2\},\{3\},\{4\},\{5\})$ according to our rule of thumb. One can observe that, while the partition $\{1,2\},\{3\},\{4\},\{5\}$ always provides a larger VaR value, at the levels $\alpha = 0.95, 0.99$ the comonotonic value (corresponding to the partition $\{1,2,3,4,5\}$, i.e. $k = 1$) provides a smaller VaR estimate if compared to $\{1,3,4,5\},\{2\}$. Then, at $\alpha = 99.9\%$, the partition $(\{1,3,4,5\},\{2\})$ delivers a smaller estimate than the comonotonic one.

We take this example to remark that the illustrated rule of thumb has not to be taken as a mathematical theorem and must be used responsibly. First of all, it is an asymptotic statement, where *asymptotic* means that it holds for sufficiently large probability levels α, but it might *not* hold at a prescribed *finite* level.

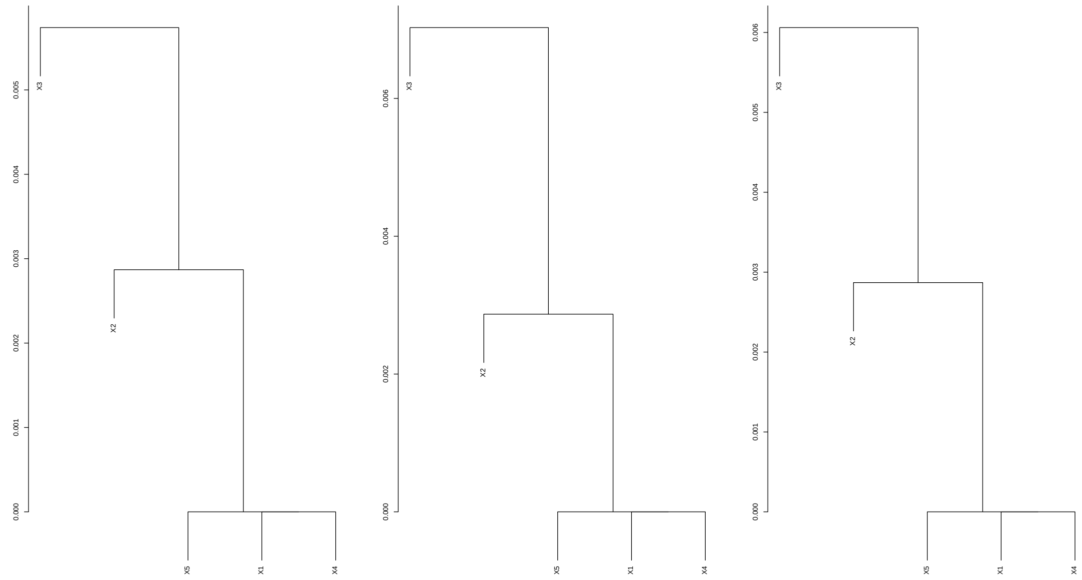

Figure 2: Hierarchical clustering results for the OpRisk dataset described in Section 5 based on dissimilarity measure (3.1) and by varying linkage rule. From left to right: single, complete and average method (the resulting partitions are identical).

	OpRisk, $d = 5$						
k		$\alpha = 0.95$	CI	$\alpha = 0.99$	CI	$\alpha = 0.999$	CI
	$\overline{\text{VaR}}_\alpha(X_d^+)$	$4.066 \cdot 10^4$		$3.440 \cdot 10^5$		$8.677 \cdot 10^6$	
4	$\{1,2\},\{3\},\{4\},\{5\}$	$1.272 \cdot 10^4$	$[1.269, 1.277] \cdot 10^4$	$1.104 \cdot 10^5$	$[1.094\,,\ 1.135] \cdot 10^5$	$3.201 \cdot 10^6$	$[3.109\,,\ 3.299] \cdot 10^6$
4	$\{1,4\},\{2\},\{3\},\{5\}$	$1.199 \cdot 10^4$	$[1.195\,,\ 1.203] \cdot 10^4$	$1.020 \cdot 10^5$	$[1.011\,,\ 1.028] \cdot 10^5$	$2.991 \cdot 10^6$	$[2.904\,,\ 3.084] \cdot 10^6$
3	$\{1,4,5\},\{2\},\{3\}$	$1.165 \cdot 10^4$	$[1.161\,,\ 1.169] \cdot 10^4$	$1.004 \cdot 10^5$	$[0.995\,,\ 1.012] \cdot 10^5$	$2.970 \cdot 10^6$	$[2.883\,,\ 3.061] \cdot 10^6$
2	$\{1,2,4,5\},\{3\}$	$1.119 \cdot 10^4$	$[1.115\,,\ 1.123] \cdot 10^4$	$1.001 \cdot 10^5$	$[0.992\,,\ 1.009] \cdot 10^5$	$2.976 \cdot 10^6$	$[2.889\,,\ 3.069] \cdot 10^6$
2	$\{1,3,4,5\},\{2\}$	$9.082 \cdot 10^3$	$[9.051\,,\ 9.113] \cdot 10^3$	$7.627 \cdot 10^4$	$[7.563\,,\ 7.691] \cdot 10^4$	$2.294 \cdot 10^6$	$[2.226\,,\ 2.362] \cdot 10^6$
1	comonotonic value	$8.355 \cdot 10^3$		$7.489 \cdot 10^4$		$2.296 \cdot 10^6$	
	$\underline{\text{VaR}}_\alpha(X_d^+)$	$2.485 \cdot 10^3$		$3.618 \cdot 10^4$		$1.654 \cdot 10^6$	

Table 4: Estimates of $\text{VaR}_\alpha^{\mathcal{I}}(X_d^+)$ at different probability levels for the sum of $d = 5$ GPD random variables following the distributions in Table 3.

The new regulatory framework for Operational Risk

In December 2017, the Basel Committee published a revision (Basel Committee on Banking Supervision, 2017a) to the current OpRisk regulatory framework. Starting from January 1, 2022, the so-called standardised approach for measuring minimum operational risk capital requirements will replace all existing approaches in the Basel II framework. As a consequence, the operational risk capital charge will be computed on the basis of a *standard formula* valid for all internationally active banks on a consolidated basis. At the moment, no change is planned for the Solvency OpRisk regulatory framework.

It is undeniable that the correct application of the AMA approach for OpRisk comes with many troublesome statistical issues (data scarcity, limited collection periods, reporting bias, cut-off values, access to reliable datasets, data contamination; see Embrechts and Hofert, 2011) that we have only touched or not considered in our study. All of the above problems are amplified by the choice of the OpRisk regulatory level. Looking at the wide empirical confidence intervals for the 99.9% capital estimates in Table 4, it is immediately clear why the choice of $\alpha = 0.999$ for the VaR capital charge has always been highly contested; see, for instance Daníelsson et al. (2001) and the examples in Embrechts et al. (2014).

However, it is also true that the possibility of opting for an advanced modeling methodology has inspired mathematically sound and very sophisticated techniques to deal with such issues, as for instance the one described in Chavez-Demoulin et al. (2016). In this sense, a brutal stop on the AMA would probably imply an end to the mathematical development of further enhanced methodologies.

It must be said that also the standard formula (SF) comes with a number of deficiencies that cannot be deemed less relevant if compared to the AMA approach. We agree with Scherer and Stahl (2018) (an English summary of this paper can be found in Stahl (2017)) that: being a *world formula*, the SF is prone to significant model uncertainty; the SF does not satisfy the axioms defining coherent risk measures; the SF is not law-invariant; the SF is not back-testable and not necessarily con-

j	Ticker	ν_j	μ_j	σ_j
1	GS	5.09	-4.90e-04	5.81e-03
2	MMM	3.05	-2.93e-04	2.63e-03
3	BA	3.12	-1.27e-04	4.21e-03
4	UNH	3.70	-5.30e-04	3.63e-03
5	HD	3.72	-2.37e-05	3.69e-03
6	IBM	3.70	-3.23e-04	3.75e-03
7	MCD	4.55	-5.14e-05	3.14e-03
8	AAPL	2.97	-1.65e-04	4.06e-03
9	JNJ	3.17	-1.97e-04	2.40e-03
10	TRV	3.14	-1.40e-04	2.92e-03
11	UTX	2.84	-2.27e-04	2.98e-03
12	CAT	4.60	-5.36e-04	5.66e-03
13	CVX	3.73	-4.63e-04	4.67e-03
14	DIS	3.27	1.41e-05	3.12e-03
15	V	4.27	-1.04e-05	4.33e-03
16	PG	6.43	-9.85e-05	3.29e-03
17	JPM	2.15	-4.61e-04	3.54e-03
18	DD	4.36	-1.68e-04	4.76e-03
19	AXP	2.80	-1.09e-04	3.70e-03
20	XOM	3.17	-2.53e-04	3.51e-03
21	WMT	3.05	-2.07e-04	3.22e-03
22	MSFT	3.09	-1.95e-04	4.04e-03
23	MRK	4.09	-1.87e-04	3.85e-03
24	NKE	7.49	3.56e-04	5.27e-03
25	VZ	6.89	-2.48e-04	3.76e-03
26	KO	3.13	6.12e-05	2.57e-03
27	INTC	3.42	-8.87e-05	4.11e-03
28	PFE	3.87	-1.06e-05	3.65e-03
29	CSCO	3.08	-1.84e-04	3.72e-03
30	GE	2.98	-2.47e-05	3.05e-03

Table 5: Composition of the Dow Jones Industrial Average, with stocks ordered by their weights in the index as of June 20, 2016. For the daily log-loss L^j we assume $L^j \sim \mu_j + \sigma_j \; t_{\nu_j}$, where t_ν denotes a t-distribution with ν degrees of freedom. MLE estimates for the marginal parameters ν_j, μ_j, and σ_j are also reported.

servative. Some of these topics are also discussed in more detail in Stahl (2016).

If a solution of the conundrum AMA versus SF is clearly beyond the scope of this paper, the insight on our methodology given by the Willis dataset remains valid to illustrate the diversification interplay between finite and infinite-mean distributions.

6. A Market Risk case study

Financial markets constitute a framework where one can conduct a higher dimensional, more robust statistical analysis. Compared to OpRisk, Market Risk data are abundant, freely available and less heavy-tailed. These factors translate into more reliable (T)VAR estimates.

The revised regulatory framework (taking effect on January 1, 2022; see Basel Committee on Banking Supervision, 2017b) for Market Risk – Internal Models Approach – prescribes a four-step adjustment procedure based on the TVaR computed (on a daily basis) at the quantile level $\alpha = 97.5\%$ over a holding period of 10 days; see Basel Committee on Banking Supervision (2016, pp.52-69). Here, we focus on the rigorous backtesting of internal models prescribed by the new regime, which includes the computation of two daily VaRs for each desk calibrated to a 99.0% and 97.5% level and a daily TVaR calibrated to 97.5%. As an illustrative example within this new regulatory set-up, we analyse the 252 daily log-losses reported in 2016 for the $d = 30$ stocks composing the Dow Jones Industrial Average (DJIA) index. Denoted by P_t^j the close price at the end of day t of the j-th stock, we study the time series

$$L_t^j = -\log\left(\frac{P_t^j}{P_{t-1}^j}\right),$$

for $1 \leq j \leq 30$ and $2016/1/4 \leq t \leq 2016/12/30$. One can of course easily reproduce the analysis which follows for a different holding period (for instance 10 days); and this possibly using the same data as overlapping observations are admitted by the Basel guidelines. Close prices P_t^j have been freely downloaded via Google Finance.

Using standard MLE techniques, for all the $d = 30$ marginal stock log-losses we fit (with impressive accuracy) the parameters of a t-distribution. The composition of the DJIA index and the corresponding estimated marginal stock distributions are collected in Table 5. Similarly to Section 5, we focus on a statistically viable and illustrative example of the methodology treated in this paper, leaving the overall procedure to compute and backtest a Market Risk economic capital to the specific trading desks of one's institution.

As for the cluster analysis of the marginal stock log-losses, we apply the hierarchical methodology as introduced and described in Section 5. Figure 3 presents the dendrogram of the 30 marginal stock log-losses for each linkage rule used. Contrary to what happens in the OpRisk case study, here the kind of linkage impacts the final clustering. Our further analysis focuses on the nested partitions of the clustering based on the average linkage, which are reported in Table 6. This example shows the applicability of a clustering technique also in a higher dimensional environment. When treating high dimensional portfolios, however, the possible choices of the final clustering also grow with the dimension and will call for a further (possibly, non statistical) selection (prior source of information or beliefs, expert opinion, etc.).

Having all the risks with a finite mean, the estimates of the joint (T)VaR for the (linearized) loss operator X_d^+ are perfectly ordered in decreasing order on k starting from $k = 1$ (comonotonic scenario). This behavior is known since Proschan (1965); see also Ju and Pan (2016) and references therein. Along with the estimates produced by the cluster analysis, we provide the portfolio (T)VaR when a multivariate t-distribution is fitted to the data, and when a t-copula is estimated from data pseudo-samples and then applied to the marginal distributions in Table 5. As one can see from Table 6, the two multidimensional t-models fall within the range of less diversified partitions. In general, we support the use of many lower dimensional sub-models rather than (very) high dimensional ones; see the warnings about using copulas in high dimensions given in Mikosch (2006) and Mai and Scherer (2013). Following this viewpoint, valid alternatives are offered by *vines* (Brechmann et al., 2012) and *hierarchical models* (Arbenz et al., 2012).

Concluding, if all marginal distributions in a risk portfolio have a finite mean, for a large enough value of α, the smallest VaR estimate is produced by the assumption of complete independence, whereas the largest by complete comonotonicity.

The full independence assumption is more realistic at the bank-level rather than at the portfolio-level, since it is common to assume that some risks such as catastrophic risk and operational risk are independent from market-driven risks (e.g. market risk, credit risk).

To this latter respect, a typical example is given by the already mentioned DNB portfolio. This portfolio has been studied in Aas and Puccetti (2014) where the ICAAP economic capital has been computed under a variety of assumptions. The DNB portfolio consists of six ($d = 6$) risks with a finite mean (Table 1 in Aas and Puccetti, 2014) and therefore, at the 99.97% confidence level at which (T)VaR estimates are computed, the independence and comonotonicity cases are to be considered the best and worst possible partitions, respectively. One can see from Tables 5-6 in Aas and Puccetti (2014) how both these scenarios deliver a reduction of the (T)VaR dependence range as computed by the RA.

7. Conclusions

An appropriate risk aggregation framework is fundamental for adequate firm-wide risk management, but poses a number of significant statistical and mathematical challenges. In particular, when computing the total economic capital associated to a joint risk portfolio, a financial institution faces a considerable model uncertainty related to the modeling of interdependence amongst the risks held.

This paper contributes to the literature of easily computable and practical bounds on (T)VaR which has been initiated in Bignozzi et al. (2015) and further extended in Puccetti et al. (2016) and Puccetti et al. (2017). In particular, we propose to apply a clustering procedure in order to partition a risk portfolio into independent subgroups of positively dependent risks. Based on available data, the portfolio partition so obtained allows for a great reduction of total economic capital and model uncertainty, under a statistically viable and mathematically sound environment. The proposed rule of thumb might serve as a guideline to avoid time consuming simulations and facilitate the choice of the aggregation model used for economic capital computation.

We conclude the paper with a mathematical appendix formalizing some of the results described in the text and a brief description of how the estimates throughout the paper have been obtained.

Appendix A. A mathematical appendix

Recall the mathematical framework described at the end of Section 2. For each different partition $\mathcal{I}$ of $\{1, \ldots, d\}$, Tables A.7-A.8 give (T)VaR estimates for the homogeneous portfolios P1 and P2 introduced in Section 4, at the quantile level $\alpha = 0.975$.

$\alpha = 0.975$				
partition	$\mathrm{VaR}^{\mathcal{I}}_\alpha(X^+_d)$	CI	$\mathrm{TVaR}^{\mathcal{I}}_\alpha(X^+_d)$	CI
$\{1,2,3,4,5\}$	12.1003	[12.0295, 12.1720]	20.6481	[20.4390, 20.8634]
$\{1,2,3,4\},\{5\}$	10.4506	[10.3944, 10.5060]	17.3080	[17.1385, 17.4831]
$\{1,2,3\},\{4,5\}$	9.6377	[9.5908, 9.6863]	15.3021	[15.1624, 15.4423]
$\{1,2,3\},\{4\},\{5\}$	8.9734	[8.9298, 9.0176]	14.2156	[14.0870, 14.3485]
$\{1,2\},\{3,4\},\{5\}$	8.4992	[8.4611, 8.5380]	12.9841	[12.8763, 13.0957]
$\{1,2\},\{3\},\{4\},\{5\}$	7.7798	[7.7472, 7.8122]	11.6279	[11.5354, 11.7255]
$\{1\},\{2\},\{3\},\{4\},\{5\}$	7.0264	[7.0008, 7.0526]	10.0471	[9.9753, 10.1196]

Table A.7: Estimates of $\mathrm{VaR}^{\mathcal{I}}_\alpha(X^+_d)$ and $\mathrm{TVaR}^{\mathcal{I}}_\alpha(X^+_d)$ at the quantile level $\alpha = 0.975$ for the sum of $d = 5$ Pareto(3) random variables.

$\alpha = 0.975$		
partition	$\mathrm{VaR}^{\mathcal{I}}_\alpha(X^+_d)$	CI
$\{1,2,3,4,5\}$	210.6203	[207.9323, 213.3341]
$\{1,2,3,4\},\{5\}$	220.3223	[217.6049, 223.0579]
$\{1,2,3\},\{4,5\}$	224.1656	[221.3665, 226.9672]
$\{1,2,3\},\{4\},\{5\}$	228.5470	[225.7553, 231.3990]
$\{1,2\},\{3,4\},\{5\}$	230.7306	[227.9037, 233.5825]
$\{1,2\},\{3\},\{4\},\{5\}$	235.0663	[232.2665, 237.9174]
$\{1\},\{2\},\{3\},\{4\},\{5\}$	239.4450	[236.5916, 242.2609]

Table A.8: Estimates of $\mathrm{VaR}^{\mathcal{I}}_\alpha(X^+_d)$ at the quantile level $\alpha = 0.975$ for the sum of $d = 5$ Pareto(0.98) random variables.

The ordering of partitions illustrated in the above tables has been shown in Theorems 4.1-4.2 of Ibragimov (2009) for convolutions of stable distributions. Using a classical result on subexponential distributions, we prove the analogous asymptotic result for Pareto-like marginals.

To a given partition $\mathcal{I} = \{I_1, \ldots, I_k\}$ of $\{1, \ldots, d\}$, associate a vector $w = (w_1, \ldots, w_d) \in \mathbb{R}^d$ defined by $w_j = \#I_j$, for $1 \leq j \leq k$, and $w_j = 0$ for $k + 1 \leq j \leq d$. Notice that $\sum_{j=1}^d w_i = d$ holds for any choice of the partition $\mathcal{I}$. In our mathematical framework, the vector w can be seen as a measure of the portfolio diversification implied by $\mathcal{I}$. For instance, $w = (1, 1, \ldots, 1)$ corresponds to the assumption of independent risks, while $v = (d, 0, \ldots, 0)$ corresponds to a full comonotonic model. A formal way to compare the portfolio diversification implied by two different partitions is offered by the notion of *majorization*.

Denote by $w_{[i]}$ the i-largest component of w ($w_{[1]}$ is the maximal, $w_{[d]}$ is the minimal). A vector $w \in \mathbb{R}^d$ is majorized by a vector $v \in \mathbb{R}^d$, $w \prec v$, if $\sum_{r=1}^j w_{[r]} \leq \sum_{r=1}^j v_{[r]}$, $1 \leq j \leq d-1$, and $\sum_{r=1}^d w_{[r]} = \sum_{r=1}^d v_{[r]}$. In our context, for example we have that

$$(1, 1, \ldots, 1) \prec (d, 0, \ldots, 0).$$

Hence $w \prec v$ means that v describes a portfolio which is less diversified than that of w. If one considers the two partitions $\mathcal{I}_1 = \{\{1, 2\}, \{3, 4\}, \{5\}\}$ and $\mathcal{I}_2 = \{\{1, 2, 3\}, \{4\}, \{5\}\}$, for the corresponding weights w^1 and w^2 one has

$$w^1 = (2, 2, 1, 0, 0) \prec (3, 1, 1, 0, 0) = w^2.$$

A function $h : \mathbb{R}^d \to \mathbb{R}$ which preserves the ordering $\prec$ of majorization is called Schur-convex. Formally, h is said to be *Schur-convex* if $w \prec v$ implies $h(w) \leq h(v)$. If the last inequality is strict whenever $w \prec v$ but v is not a permutation of w, h is said to be *strictly* Schur-convex. The function h is said to be (strictly) *Schur-concave* if $-h$ is (strictly) Schur-convex. For

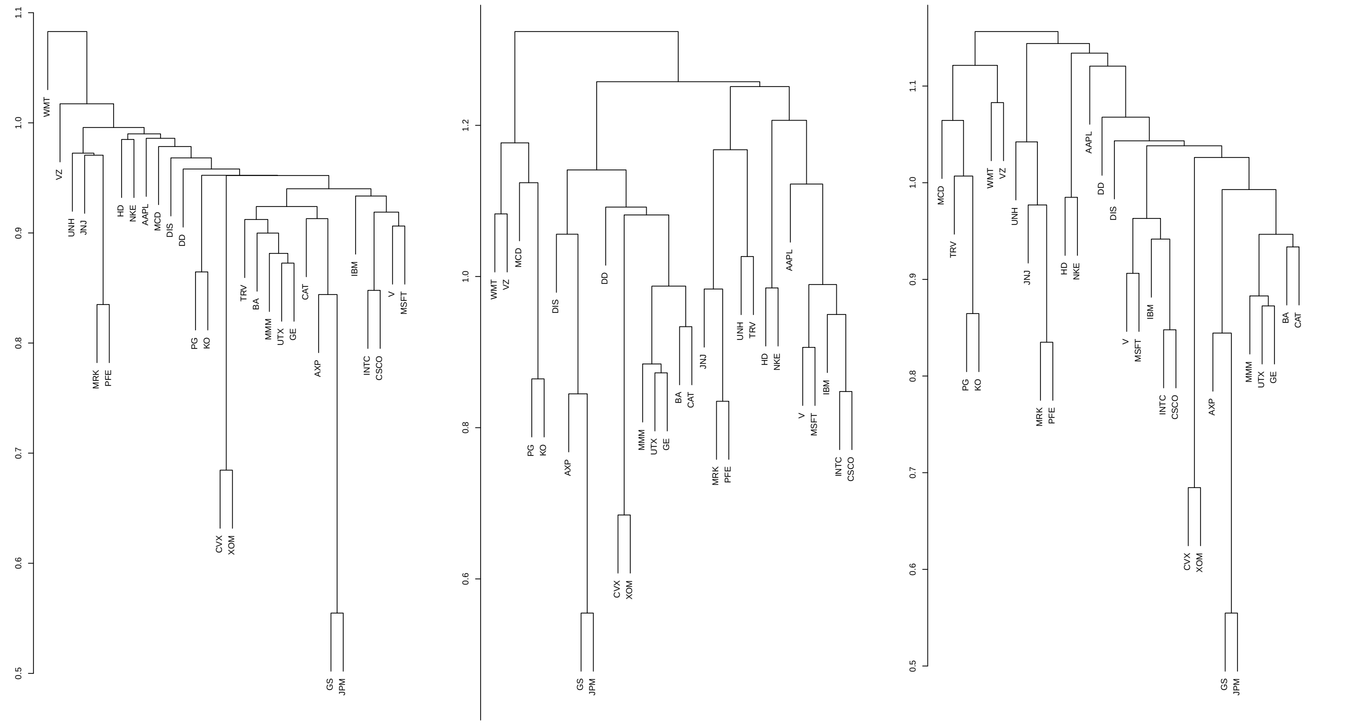

Figure 3: Hierarchical clustering results for the Market Risk dataset described in Section 6 based on the dissimilarity measure (3.1) and by varying linkage rule. From left to right: single, complete and average method.

example, the function $h(w) = \sum_{j=1}^{d} w_j^{\theta}$ is strictly Schur-convex for $\theta > 1$ and strictly Schur-concave for $\theta < 1$. For this and many more results on Schur-convexity, we refer to Marshall et al. (2011, Ch. 3).

For a given $\mathcal{I}$, denote by $X_{d,\mathcal{I}}^{+}$ the sum of the risks $X_1, \dots, X_d$ under the assumption that the risk subvectors $X_{I_1}, \dots, X_{I_k}$ are independent while the risks $(X_j, j \in I_i)$ inside each subgroup are comonotonic. If the risks X_j are also assumed to be identically distributed, one immediately has that

$$P\left(X_{d,\mathcal{I}}^{+} > x\right) = P\left(\sum_{j=1}^{k} w_j Y_j > x\right), \qquad \text{(A.1)}$$

where $Y_1, \dots, Y_k$ are iid random variables with $Y_1 \sim X_1$. Equation (A.1) holds because identically distributed risks belonging the same subgroup will be (under the assumption of comonotonicity inside the subgroup) equal with probability one.

Assume now that the distribution of each X_j (hence of each Y_j) is *regularly varying* with a common tail index $\theta > 0$, written $\mathcal{R}(-\theta)$, that is

$$\lim_{x\to\infty} \frac{P(X_j > tx)}{P(X_j > x)} = t^{-\theta}, \text{ for all } t > 0, \text{ and } 1 \le j \le d.$$

Notice that a Pareto(θ) distribution is $\mathcal{R}(-\theta)$, with the tail index θ separating distributions with finite mean ($\theta > 1$) from those with infinite mean ($\theta \le 1$).

As regularly varying distributions are subexponential (see for instance Appendix A3 in Embrechts et al. (1997) or Cooke et al. (2014, Cor. 4.1)), one can use the very well-known approximation result on subexponential distribution (Tang and Tsitsiashvili, 2003) that

$$P\left(\sum_{j=1}^{k} w_j Y_j > x\right) \overset{x\to\infty}{\sim} \sum_{j=1}^{k} P\left(w_j Y_j > x\right) \overset{x\to\infty}{\sim} \sum_{j=1}^{k} w_j^{\theta} P(X_1 > x). \qquad \text{(A.2)}$$

In the comparison of two partitions $\mathcal{I}$ and $\mathcal{I}'$ with corresponding weights w and v one obtains from (A.1) and (A.2) that

$$\lim_{x\to\infty} \frac{P(X_{d,\mathcal{I}}^{+} > x)}{P(X_{d,\mathcal{I}'}^{+} > x)} = \frac{\sum_{j=1}^{d} w_j^{\theta}}{\sum_{j=1}^{d} v_j^{\theta}}. \qquad \text{(A.3)}$$

As the function $h : \mathbb{R}^d \to \mathbb{R}$, $h(w) = \sum_{j=1}^{d} w_j^{\theta}$ is strictly Schur-convex when $\theta > 1$ and strictly Schur-concave when $\theta < 1$, the following proposition immediately follows from (A.3).

Proposition A.1. Let $\mathcal{I}$ and $\mathcal{I}'$ two partitions of $\{1, \dots, d\}$ with corresponding weights w and v such that $w \prec v$ and w is not a permutation of v. If the common distribution of the random variables $X_1, \dots, X_d$ is $\mathcal{R}(-\theta)$, and under the set-up introduced above, we have that

$$\lim_{x\to\infty} \frac{P(X_{d,\mathcal{I}}^{+} > x)}{P(X_{d,\mathcal{I}'}^{+} > x)} = \begin{cases} < 1, & \text{if } \theta > 1; \\ > 1, & \text{if } \theta < 1. \end{cases}$$

Translated into the language of VaR, this means that

$$\lim_{\alpha\to 1^-} \frac{\mathrm{VaR}_{\alpha}^{\mathcal{I}}(X_d^{+})}{\mathrm{VaR}_{\alpha}^{\mathcal{I}'}(X_d^{+})} = \begin{cases} < 1, & \text{if } \theta > 1; \\ > 1, & \text{if } \theta < 1. \end{cases}$$

Market Risk $d = 30$

k	VaR_α	$\alpha = 0.975$	CI	$\alpha = 0.99$	CI
	$\overline{\mathrm{VaR}}_\alpha(X_d^+)$	0.49122		0.66399	
	comonotonic value	0.32431		0.45049	
2	{1, 2, 3, 4, 5, 6, 8, 9, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 23, 24, 27, 28, 29, 30}, {7, 10, 16, 21, 25, 26}	0.27652	[0.27518, 0.27781]	0.38355	[0.38094, 0.38605]
3	{1, 2, 3, 5, 6, 8, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 24, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 21, 25, 26}	0.24148	[0.24033, 0.24254]	0.33395	[0.33165, 0.33613]
3	*t-copula + marginals*	0.23364	[0.23233, 0.23488]	0.34063	[0.33789, 0.34320]
4	{1, 2, 3, 6, 8, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 21, 25, 26}, {5, 24}	0.22049	[0.21945, 0.22152]	0.30561	[0.30361, 0.30760]
-	*multivariate t-distribution*	0.21417	−−	0.29073	−−
5	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 21, 25, 26}, {5, 24}, {8}	0.20862	[0.20763, 0.20954]	0.28807	[0.28619, 0.28995]
6	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 26}, {5, 24}, {21, 25}, {8}	0.20555	[0.20459, 0.20647]	0.28521	[0.28327, 0.28703]
7	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 26}, {5, 24}, {3 singletons}	0.20517	[0.20426, 0.20608]	0.28485	[0.28290, 0.28673]
8	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 26}, {5, 24}, {4 singletons}	0.19327	[0.19244, 0.19418]	0.26855	[0.26671, 0.27028]
9	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {10, 16, 26}, {5, 24}, {5 singletons}	0.19226	[0.19138, 0.19320]	0.26768	[0.26581, 0.26945]
10	{1, 2, 3, 6, 11, 12, 13, 15, 17, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {10, 16, 26}, {5, 24}, {6 singletons}	0.18343	[0.18258, 0.18429]	0.25499	[0.25327, 0.25672]
11	{1, 2, 3, 6, 11, 12, 13, 15, 17, 19, 20, 22, 27, 29, 30}, {9, 23, 28}, {10, 16, 26}, {5, 24}, {7 singletons}	0.18185	[0.18096, 0.18272]	0.25352	[0.25172, 0.25526]
12	{1, 2, 3, 11, 12, 13, 17, 19, 20, 30}, {9, 23, 28}, {6, 15, 22, 27, 29}, {10, 16, 26}, {5, 24}, {7 singletons}	0.13974	[0.13915, 0.14039]	0.18934	[0.18818, 0.19056]
13	{1, 2, 3, 11, 12, 17, 19, 30}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {10, 16, 26}, {5, 24}, {7 singletons}	0.12291	[0.12241, 0.12345]	0.16421	[0.16335, 0.16519]
14	{1, 2, 3, 11, 12, 17, 19, 30}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {5, 24}, {8 singletons}	0.12172	[0.12121, 0.12222]	0.16297	[0.16203, 0.16396]
15	{1, 17, 19}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {5, 24}, {8 singletons}	0.10297	[0.10259, 0.10338]	0.13393	[0.13319, 0.13462]
16	{1, 17, 19}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {4}, {10 singletons}	0.10179	[0.10138, 0.10217]	0.13270	[0.13197, 0.13342]
17	{1, 17, 19}, {13, 20}, {23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {11 singletons}	0.10023	[0.09984, 0.10062]	0.13091	[0.13016, 0.13158]
18	{1, 17, 19}, {13, 20}, {23, 28}, {6, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {15, 22}, {11 singletons}	0.09163	[0.09130, 0.09196]	0.11895	[0.11832, 0.11956]
19	{1, 17, 19}, {13, 20}, {23, 28}, {6, 27, 29}, {16, 26}, {2, 11, 30}, {15, 22}, {3, 12}, {11 singletons}	0.08317	[0.08287, 0.08346]	0.10668	[0.10618, 0.10720]
20	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {2, 11, 30}, {15, 22}, {3, 12}, {12 singletons}	0.08008	[0.07980, 0.08038]	0.10262	[0.10216, 0.10309]
21	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {2, 11, 30}, {15, 22}, {14 singletons}	0.07763	[0.07737, 0.07790]	0.09959	[0.09911, 0.10007]
22	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {2, 11, 30}, {16 singletons}	0.07571	[0.07546, 0.07600]	0.09724	[0.09679, 0.09773]
23	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {11, 30}, {17 singletons}	0.07355	[0.07330, 0.07383]	0.09426	[0.09384, 0.09474]
24	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {19 singletons}	0.07220	[0.07196, 0.07248]	0.09247	[0.09205, 0.09293]
25	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {21 singletons}	0.07136	[0.07111, 0.07161]	0.09154	[0.09110, 0.09203]
26	{1, 17, 19}, {13, 20}, {23, 28}, {23 singletons}	0.06933	[0.06909, 0.06959]	0.08889	[0.08848, 0.08936]
27	{1, 17}, {13, 20}, {23, 28}, {24 singletons}	0.06462	[0.06441, 0.06486]	0.08171	[0.08134, 0.08208]
28	{1, 17}, {13, 20}, {26 singletons}	0.06299	[0.06278, 0.06323]	0.07972	[0.07934, 0.08011]
29	{1, 17}, {28 singletons}	0.06074	[0.06053, 0.06096]	0.07672	[0.07638, 0.07709]
	$\underline{\mathrm{VaR}}_\alpha(X_d^+)$	−0.01813		−0.01216	

k	TVaR_α	$\alpha = 0.975$	CI	$\alpha = 0.99$	CI
	$\overline{\mathrm{TVaR}}_\alpha(X_d^+)$	0.49108		0.66417	
2	{1, 2, 3, 4, 5, 6, 8, 9, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 23, 24, 27, 28, 29, 30}, {7, 10, 16, 21, 25, 26}	0.41891	[0.41566, 0.42213]	0.56711	[0.56054, 0.57422]
-	*t-copula + marginals*	0.38311	[0.37906, 0.38674]	0.54315	[0.53521, 0.55146]
3	{1, 2, 3, 5, 6, 8, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 24, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 21, 25, 26}	0.36522	[0.36242, 0.36807]	0.49438	[0.48853, 0.50058]
4	{1, 2, 3, 6, 8, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 21, 25, 26}, {5, 24}	0.33536	[0.33276, 0.33807]	0.45574	[0.45020, 0.46177]
5	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 21, 25, 26}, {5, 24}, {8}	0.31573	[0.31332, 0.31830]	0.42795	[0.42270, 0.43353]
6	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 26}, {5, 24}, {21, 25}, {8}	0.31289	[0.31044, 0.31540]	0.42534	[0.42017, 0.43084]
7	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 26}, {5, 24}, {3 singletons}	0.31258	[0.31013, 0.31509]	0.42509	[0.41989, 0.43062]
-	*multivariate t-distribution*	0.30972	−−	0.40577	−−
8	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {7, 10, 16, 26}, {5, 24}, {4 singletons}	0.29530	[0.29294, 0.29769]	0.40249	[0.39744, 0.40773]
9	{1, 2, 3, 6, 11, 12, 13, 14, 15, 17, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {10, 16, 26}, {5, 24}, {5 singletons}	0.29445	[0.29206, 0.29686]	0.40176	[0.39670, 0.40709]
10	{1, 2, 3, 6, 11, 12, 13, 15, 17, 19, 20, 22, 27, 29, 30}, {4, 9, 23, 28}, {10, 16, 26}, {5, 24}, {6 singletons}	0.28049	[0.27824, 0.28281]	0.38248	[0.37765, 0.38757]
11	{1, 2, 3, 6, 11, 12, 13, 15, 17, 19, 20, 22, 27, 29, 30}, {9, 23, 28}, {10, 16, 26}, {5, 24}, {7 singletons}	0.27905	[0.27683, 0.28138]	0.38124	[0.37646, 0.38637]
12	{1, 2, 3, 11, 12, 13, 17, 19, 20, 30}, {9, 23, 28}, {6, 15, 22, 27, 29}, {10, 16, 26}, {5, 24}, {7 singletons}	0.20744	[0.20589, 0.20922]	0.27867	[0.27538, 0.28249]
13	{1, 2, 3, 11, 12, 17, 19, 30}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {10, 16, 26}, {5, 24}, {8 singletons}	0.17930	[0.17792, 0.18085]	0.23852	[0.23559, 0.24191]
14	{1, 2, 3, 11, 12, 17, 19, 30}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {5, 24}, {8 singletons}	0.17808	[0.17673, 0.17966]	0.23735	[0.23437, 0.24078]
15	{1, 17, 19}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {5, 24}, {8 singletons}	0.14357	[0.14262, 0.14467]	0.18526	[0.18318, 0.18762]
16	{1, 17, 19}, {13, 20}, {9, 23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {10 singletons}	0.14244	[0.14147, 0.14354]	0.18423	[0.18218, 0.18659]
17	{1, 17, 19}, {13, 20}, {23, 28}, {6, 15, 22, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {11 singletons}	0.14070	[0.13972, 0.14176]	0.18234	[0.18031, 0.18476]
18	{1, 17, 19}, {13, 20}, {23, 28}, {6, 27, 29}, {16, 26}, {2, 3, 11, 12, 30}, {15, 22}, {11 singletons}	0.12791	[0.12706, 0.12890]	0.16536	[0.16355, 0.16748]
19	{1, 17, 19}, {13, 20}, {23, 28}, {6, 27, 29}, {16, 26}, {2, 11, 30}, {15, 22}, {3, 12}, {11 singletons}	0.11424	[0.11349, 0.11512]	0.14615	[0.14448, 0.14807]
20	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {2, 11, 30}, {15, 22}, {3, 12}, {12 singletons}	0.11002	[0.10926, 0.11084]	0.14081	[0.13916, 0.14264]
21	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {2, 11, 30}, {15, 22}, {14 singletons}	0.10702	[0.10624, 0.10784]	0.13736	[0.13566, 0.13920]
22	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {2, 11, 30}, {16 singletons}	0.10468	[0.10393, 0.10554]	0.13470	[0.13304, 0.13653]
23	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {11, 30}, {17 singletons}	0.10141	[0.10065, 0.10222]	0.13026	[0.12862, 0.13208]
24	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {16, 26}, {19 singletons}	0.09950	[0.09877, 0.10030]	0.12777	[0.12618, 0.12956]
25	{1, 17, 19}, {13, 20}, {23, 28}, {27, 29}, {21 singletons}	0.09861	[0.09789, 0.09940]	0.12689	[0.12532, 0.12879]
26	{1, 17, 19}, {13, 20}, {23, 28}, {23 singletons}	0.09586	[0.09517, 0.09668]	0.12343	[0.12194, 0.12521]
27	{1, 17}, {13, 20}, {23, 28}, {24 singletons}	0.08669	[0.08615, 0.08733]	0.10902	[0.10781, 0.11050]
28	{1, 17}, {13, 20}, {26 singletons}	0.08475	[0.08420, 0.08544]	0.10682	[0.10563, 0.10838]
29	{1, 17}, {28 singletons}	0.08160	[0.08107, 0.08221]	0.10278	[0.10160, 0.10422]

Table 6: Estimates of $\mathrm{VaR}_\alpha^{\mathcal{I}}(X_d^+)$ and $\mathrm{TVaR}_\alpha^{\mathcal{I}}(X_d^+)$ at different probability levels for the sum of $d = 30$ random variables following the distributions in Table 5. We also provide the portfolio (T)VaR when a multivariate t-distribution is fitted to the data (*multivariate t-distribution*, in this case one can provide analytical estimates), and when a t-copula is estimated from data pseudo-samples and then applied to the marginal distributions (*t-copula+marginals*).

Appendix B. Computational details

The figures reported in the tables throughout the paper are median estimates evaluated over 10^4 (10^3 for Table 6) repetitions of 10^6 (10^7 for Table 4 and $5 \cdot 10^6$ for Table 1(P4)) Monte Carlo simulations under each prescribed model. Simulations for Table 1 (P1, P2) and Tables A.7–A.8 have been obtained using MATLAB (2017a) on a Asus K501UX (Intel i7-6500U, 2 cores, 2.5GHz, 12 GB RAM); simulations for all remaining tables have been obtained on INDACO Cluster (16 nodes each with CPU Intel Xeon E5-22683V4 2.1GHz, 16 cores, 256GB RAM) using MATLAB (2017b). For each figure, empirical Confidence Intervals (CI) at the 95% confidence level are also reported. Estimates of $\underline{\mathrm{VaR}}_\alpha(X_d^+)$ and $\overline{\mathrm{VaR}}_\alpha(X_d^+)$ are obtained via the RA function of the R package `qrmtools` with $2 \cdot 10^6$ ($4 \cdot 10^6$ for Table 4) discretization points; estimates of $\underline{\mathrm{TVaR}}_\alpha(X_d^+)$ are obtained via the RA using the numerical methodology described in Jakobsons and Vanduffel (2015) with $2 \cdot 10^6$ discretization points.

Acknowledgments

The authors are grateful to two referees and an Associate Editor for many useful suggestions that improved an earlier version of this paper. This research was inspired by fruitful conversations between Hansjoerg Albrecher and Giovanni Puccetti. F. Marta L. Di Lascio acknowledges the support of the Free University of Bozen-Bolzano, Faculty of Economics and Management, via the project *Aggregation functions for Innovation and Data Analysis* (AIDA). Giovanni Puccetti acknowledges the financial support of the University of Milan via the project *Piano di sostegno alla ricerca 2015-2017* and the use of the platform UNITECH INDACO. The three authors are listed in alphabetical order and their individual contribution to this paper is equal.

References

Aas, K. and G. Puccetti (2014). Bounds for total economic capital: the DNB case study. *Extremes 17*(4), 693–715.

Arbenz, P. and D. Canestraro (2012). Estimating copulas for insurance from scarce observations, expert opinion and prior information: a Bayesian approach. *Astin Bull. 42*(1), 271–290.

Arbenz, P., C. Hummel, and G. Mainik (2012). Copula based hierarchical risk aggregation through sample reordering. *Insurance Math. Econom. 51*(1), 122–133.

Basel Committee on Banking Supervision (2011). *Operational Risk – Supervisory Guidelines for the Advanced Measurement Approaches*. Bank for International Settlements, Basel.

Basel Committee on Banking Supervision (2016). *Minimum Capital Requirements for Market Risk*. Bank for International Settlements, Basel.

Basel Committee on Banking Supervision (2017a). *Basel III: Finalising post-crisis reforms*. Bank for International Settlements, Basel.

Basel Committee on Banking Supervision (2017b). *High-level Summary of Basel III Reforms*. Bank for International Settlements, Basel.

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. *Depend. Model. 1*, 37–53.

Bernard, C., L. Rüschendorf, and S. Vanduffel (2015). Value-at-risk bounds with variance constraints. *J. Risk Insur. 84*(3), 923–959.

Bernard, C., L. Rüschendorf, S. Vanduffel, and R. Wang (2017). Risk bounds for factor models. *Finance Stoch. 21*(3), 631–659.

Bignozzi, V., G. Puccetti, and L. Rüschendorf (2015). Reducing model risk via positive and negative dependence assumptions. *Insurance Math. Econom. 61*, 17–26.

Brechmann, E. C., C. Czado, and K. Aas (2012). Truncated regular vines in high dimensions with application to financial data. *Canad. J. Statist. 40*(1), 68–85.

Chavez-Demoulin, V., P. Embrechts, and M. Hofert (2016). An extreme value approach for modeling operational risk losses depending on covariates. *J. Risk Insur. 83*(3), 735–776.

Chavez-Demoulin, V., P. Embrechts, and J. Nešlehová (2006). Quantitative models for operational risk: Extremes, dependence and aggregation. *J. Bank. Financ. 30*(10), 2635–2658.

Cirillo, P. and N. N. Taleb (2016). Expected shortfall estimation for apparently infinite-mean models of operational risk. *Quant. Finance 16*(10), 1485–1494.

Cooke, R. M., D. Nieboer, and J. Misiewicz (2014). *Fat-Tailed Distributions: Volume 1*. John Wiley & Sons, Hoboken NJ.

Daníelsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin (2001). An academic response to Basel II. London School of Economics Financial Markets Group, Special Paper 130.

Dhaene, J., M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom. 31*(1), 3–33.

Embrechts, P. and M. Hofert (2011). Practices and issues in operational risk modeling under Basel II. *Lith. Math. J. 51*(2), 180–193.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

Embrechts, P., A. J. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk management: Value at Risk and Beyond*, pp. 176–223. Cambridge Univ. Press.

Embrechts, P. and G. Puccetti (2008). Aggregating operational risk across matrix structured loss data. *J. Operational Risk 3*(2), 29–44.

Embrechts, P., G. Puccetti, and L. Rüschendorf (2013). Model uncertainty and VaR aggregation. *J. Bank. Financ. 58*(8), 2750–2764.

Embrechts, P., G. Puccetti, L. Rüschendorf, R. Wang, and A. Beleraj (2014). An academic response to Basel 3.5. *Risks 2*(1), 25–48.

Everitt, B. S., S. Landau, M. Leese, and D. Stahl (2011). *Cluster Analysis* (Fifth ed.). John Wiley & Sons, Chichester.

Everitt, B. S. and A. Skrondal (2010). *The Cambridge Dictionary of Statistics* (Fourth ed.). Cambridge Univ. Press.

Ibragimov, M., R. Ibragimov, and J. Walden (2015). *Heavy-tailed Distributions and Robustness in Economics and Finance*. Springer, Cham.

Ibragimov, R. (2009). Portfolio diversification and value at risk under thick-tailedness. *Quant. Finance 9*(5), 565–580.

Ibragimov, R. and A. Prokhorov (2016). Heavy tails and copulas: Limits of diversification revisited. *Econ. Lett. 149*, 102 – 107.

Jakobsons, E. and S. Vanduffel (2015). Dependence uncertainty bounds for the expectile of a portfolio. *Risks 3*(4), 599–623.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An Introduction to Statistical Learning*. Springer, New York.

Ju, S. and X. Pan (2016). A new proof for the peakedness of linear combinations of random variables. *Statist. Probab. Lett. 114*, 93–98.

Mai, J.-F. and M. Scherer (2013). What makes dependence modeling challenging? Pitfalls and ways to circumvent them. *Stat. Risk Model. 30*(4), 287–306.

Mainik, G., , and P. Embrechts (2013). Diversification in heavy-tailed portfolios: properties and pitfalls. *Ann. Actuar. Science 7*(1), 26–45.

Marshall, A. W., I. Olkin, and B. C. Arnold (2011). *Inequalities: Theory of Majorization and its Applications* (Second ed.). Springer, New York.

McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative Risk Management. Concepts, Techniques and Tools. Revised Edition.* Princeton University Press.

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

Moscadelli, M. (2004). The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee. Temi di discussione 517, Banca d'Italia.

Nešlehová, J., P. Embrechts, and V. Chavez-Demoulin (2006). Infinite-mean models and the LDA for operational risk. *J. Operational Risk 1*(1), 3–25.

Pflug, G. C. and M. Pohl (2017). A review on ambiguity in stochastic portfolio optimization. *Set-Valued Var. Anal.*, to appear.

Prettenthaler, F., H. Albrecher, P. Asadi, and J. Köberl (2017). On flood risk

pooling in europe. *Nat. Hazards 88*(1), 1–20.
Proschan, F. (1965). Peakedness of distributions of convex combinations. *Ann. Math. Statist. 36*, 1703–1706.
Puccetti, G., L. Rüschendorf, and D. Manko (2016). VaR bounds for joint portfolios with dependence constraints. *Depend. Model. 4*(1), 368–381.
Puccetti, G., L. Rüschendorf, D. Small, and S. Vanduffel (2017). Reduction of Value-at-Risk bounds via independence and variance information. *Scand. Actuar. J. 2017*(3), 245–266.
Scherer, M. and G. Stahl (2018). Die Solvency II-Standardformel: Mythos und Wirklichkeit. Working paper.
Stahl, G. (2016). Model uncertainty in a holistic perspective. In J. Kallsen and A. Papapantoleon (Eds.), *Advanced Modelling in Mathematical Finance*, pp. 189–215. Springer, Cham.
Stahl, G. (2017). Model uncertainty of the standard formula and implications for internal models. Presented at the *Workshop: Risk Measurement and Regulatory Issues in Business* at the CRM Montreal, Canada.
Tang, Q. and G. Tsitsiashvili (2003). Randomly weighted sums of subexponential random variables with application to ruin theory. *Extremes 6*(3), 171–188 (2004).