Proper Conditioning for Coherent VaR in Portfolio Management*

René Garcia
Université de Montréal, CIRANO and CIREQ

Éric Renault
University of North Carolina at Chapel Hill, CIRANO and CIREQ

Georges Tsafack
Université de Montréal, CIRANO and CIREQ

First Draft: November 2003, This Version: April 2006

Abstract

Value at Risk (VaR) is a central concept in risk management. As stressed by Artzner et al. (1999), VaR may not possess the subadditivity property required to be a coherent measure of risk. The key idea of this paper is that, when tail thickness is responsible for violation of subadditivity, eliciting proper conditioning information may restore VaR rationale for decentralized risk management. The argument is threefold. First, since individual traders are hired because they possess a richer information on their specific market segment than senior management, they just have to follow consistently the prudential targets set by senior management to ensure that decentralized VaR control will work in a coherent way. The intuition is that if one could build a fictitious conditioning information set merging all individual pieces of information, it would be rich enough to restore VaR subadditivity. Second, in this decentralization context, we show that if senior management has access ex-post to the portfolio shares of the individual traders, it amounts to recovering some of their private information. These shares can be used to improve backtesting in order to check that the prudential targets have been enforced by the traders. Finally, we stress that tail thickness required to violate subadditivity, even for small probabilities, remains an extreme situation since it corresponds to such poor conditioning information that expected loss appears to be infinite. We then conclude that lack of coherence of decentralized VaR management, that is VaR non-subadditivity at the richest level of information, should be an exception rather than a rule.

Keywords: Value at risk, Decentralized risk management, Coherent measures of risk, Subadditivity of VaR, Heavy-tail distributions, Stable distributions

JEL classification: G1, C1, C4

*The authors gratefully acknowledge financial support from the Institut de Finance Mathématique de Montréal (IFM2), the Fonds québécois de la recherche sur la société et la culture (FQRSC), the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Network of Centres of Excellence MITACS. The first author also thanks Hydro-Québec and the Bank of Canada for financial support. Address for correspondence: Département de Sciences Économiques, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal (Québec) H3C 3J7, Canada.

1 Introduction

Value at risk (VaR) - the amount of money such that there is typically a 95% or 99% probability of a portfolio losing less than that amount over a certain horizon, has become a central concept in risk management[1]. Financial institutions, regulators as well as non-financial corporations use this method to measure financial risk. Although the VaR risk measure seems to agree with a concept of maximum loss popular with practitioners, it is not a coherent measure of risk, as stressed by Artzner et al.(1999), since it is not subadditive.

In practice, VaR is also used as a tool to manage and control risk. Several authors have recently pointed out that VaR as a risk magement tool may have perverse effects. Basak and Shapiro (2001) show that VaR risk managers choose a larger exposure to risky assets than non-risk managers and as a result incur larger losses in the worst states of the world. In general equilibrium, risk regulation may have the perverse effect of amplifying price fluctuations, as demonstrated by Basak and Shapiro (2001) and Danielsson, Shin and Zigrand (2004)).

In this paper, we focus on a decentralized portfolio management system, widespread among financial institutions, that relies on VaR as a risk measure and a risk control tool. In such a system, subadditivity appears as a natural requirement. Typically, the supervisory unit in charge of portfolio management wants to decentralize the management of certain parts of the portfolio to market segment specialists. If the unit wants to impose a global VaR amount on the whole portfolio, subadditivity will allow it to decentralize its VaR constraint into several VaR constraints, one per specialist. The supervisory unit will then be assured that VaR of the global risk will not surpass the sum of individual VaRs. The problem is that, even if one is ready to assume that individual risks are mutually independent, Artzner et al.(1999) have given examples of VaR nonsubadditivity in the case of aggregation of independent risks. A notable exception is the case where all risks are jointly normally (or more generally elliptically) distributed, since the quantiles satisfy subadditivity as long as probabilities of exceedence are smaller than a half.

We provide an analysis of the feasibility of decentralized risk management through VaR objectives even when subaddivity is violated. The main idea is that, when tail thickness is responsible for violation of subadditivity, eliciting proper conditioning information may restore a rationale for a decentralized risk management system based on VaR. In such a Rent-a-Trader system, as it is often called, the specialists are hired because they have access to specific information on which they condition their portfolio decisions. Therefore, the central management unit possesses only a subset of the conditioning information which belongs to each specialist. Naturally, in such a context, a distribution always appears more

[1] See, for example, Jorion (2001).

leptokurtic to the central unit than to the specialist. VaR may appear non-subadditive to the central management unit due to a lack of information, but without bad consequences for the actual risk incurred. We are then able to show that decentralized portfolio management with a VaR allocation to each specialist will work despite evidence to the contrary. VaR can thus be decentralized if specialists obey their VaR requirements.

Of course, central management will still want to assess whether or not specialists meet their VaR requirements. We distinguish the case where central management has access to information from a situation where they must rely on unconditional information only. We provide an illustration where traders have access to private information, which is unobservable to both central management and other traders, but where they communicate their individual portfolio shares to central management. We discuss ways for central management to improve backtesting of VaR requirements in this context.

To provide a theoretical underpinning to the validity of such a Rent-a-Trader system for risk control, we proceed in two steps. First, we show that violations of VaR subadditivity in case of aggregation of independent risks are basically due to a perverse effect of fat tails. More precisely, we show that when tails are sufficiently thin to ensure the existence of finite absolute expected returns, a rather realistic assumption, VaR subadditivity is at least guaranteed for sufficiently large confidence levels, or equivalently for sufficiently small levels of the fixed probability of VaR exceedence. The example of stable distributions is well suited to assess how small is small. In this setting, it is shown that with a reasonable level of skewness in asset returns, VaR at common confidence levels will be subadditive when absolute expected returns have a finite expectation. A contrario, we show that arbitrarily thick tails may produce arbitrarily large violations of the subaddivity property.

Second, we provide the key argument in favor of the rent-a-trader approach by linking fat tailness to a lack of conditioning information. Through the consideration of higher order moments, we extend the argument of Clark (1973) to note that, in general, the more information we condition the returns of a portfolio upon, the thinner are the tails of the resulting distribution. This argument works in particular for scale mixing variables like stochastic volatility. Moreover, by appealing to a scale mixture representation property of stable distributions, we show that our framework nests the family of stable distributions. In other words, we have a way of recovering VaR subadditivity through proper conditioning, which is valid in particular for stable distributions. This is a satisfactory result in light of the early work by Mandelbrot (1963) and Fama (1965). They show that stable distributions accommodate heavy-tailed financial series and therefore produce measures of risk based on distribution tails, such as VaR, which are more reliable (see in particular Mittnik and Rachev, 1993, and Mittnik, Paolella and Rachev, 2000).

The rest of the paper is organized as follows. Section 2 describes the Rent-a-Trader system whereby portfolio management is decentralized to specialists. Section 3 appeals to some elements of probability theory to put forward the logical relationships between fat tails, violation of VaR subadditivity and conditioning information. In section 4 we show that deconditioning by central management always increases tail fatness and spuriously makes VaR look non-subadditive. In section 5 we provide a simple illustration with private signals to traders and show how the transmission of information in the form of portfolio shares helps to assess risk more accurately. Section 6 concludes. Proofs of the various propositions are collected in an Appendix.

2 VaR Decentralization with Differential Information

In this section we describe a decentralized portfolio management system that uses VaR as a tool for risk management. Senior management is interested in the value at risk $VaR_p(X)$ associated with the random value X of its portfolio:

$$P[X \leq -VaR_p(X)] = p. \tag{2.1}$$

This value X will be conformable to a VaR requirement VaR_p^* if and only if: $VaR_p(X) \leq VaR_p^*$, that is if:

$$P[X \leq -VaR_p^*] \leq p. \tag{2.2}$$

Suppose that senior management hires n traders to manage parts of its portfolio. Then $X = \Sigma_{j=1}^n X_j$ is the aggregation of the net results X_j of n traders $j = 1, \ldots, n$. Failure of subadditivity for senior management in this context means that $VaR_p(\sum_{j=1}^n X_j)$ may exceed $\sum_{j=1}^n VaR_p(X_j)$. We will show in the next section that violation of subadditivity is a perverse effect of fat tails in the distribution of X. We will then see that fat tails can be reduced by relying on some conditioning information.

Trader j has obtained result X_j by building a portfolio $\theta_j(I_j)$, which is a function of his private information I_j. A decentralized management system is used precisely to exploit this private information I_j since it is inaccessible to central management. Let us consider that trader j has received from senior management a target S_j in terms of VaR, that is:

$$VaR_p(X_j \,|I_j\,) \leq S_j \tag{2.3}$$

where:

$$P[X_j \leq -VaR_p(X_j \,|I_j\,) \,|I_j\,] = p. \tag{2.4}$$

Note that S_j is a given number chosen by senior management. Typically, the quantity $VaR_p(X_j\,|I_j\,)$, which depends on private information, cannot be observed at the central level. Therefore, senior management cannot check directly that the requested target (2.3) has been met or equivalently that:

$$P[X_j \leq -S_j\,|I_j] \leq \; p. \tag{2.5}$$

Therefore it must rely on backtesting, with two objectives. First, as usual, senior management must check on a time series of portfolios returns that $X = \Sigma_{j=1}^{n} X_j$ fulfills the VaR requirement (2.2). It is quite natural for central management to imagine that this requirement will be ensured by the enforcement of targets S_j if and only if these targets are chosen ex-ante in order to fulfill:

$$\Sigma_{j=1}^{n} S_j \leq \; VaR_p^*. \tag{2.6}$$

Second, even though (2.5) cannot be observed, senior management should be interested to seek valuable information about individual trader j behavior. Of course, historical observation allows him to test for a weak consequence of (2.5), that is:

$$P[X_j \leq -S_j] \leq \; p. \tag{2.7}$$

But, for the targets to appear credible, a tighter control should be performed. Often senior management will request that traders communicate their portfolio shares. We will see how this information can help exercise a better control, but even if this information on individual portfolio shares is transmitted, senior management will never recover fully in practice the information I_j of the individual traders.

In section 4, we will show that, under a set of natural assumptions, both goals of backtesting may be met. In other words, it will be true that the enforcement of targets conformable to (2.6) will ensure (2.2). Moreover, senior management will have at its disposal some relevant information for a more powerful test of (2.5) than only through its weak consequence (2.7).

3 Conditioning Information, Tails and VaR Subadditivity

We show in a first subsection that violation of subadditivity in the case of aggregation of independent risks is basically a perverse effect of fat tails. In a second subsection, we study the logical relationships between conditioning information and fat tailness, in particular in the context of scale mixtures of distributions.

3.1 Tails and Subaddivity

Let us consider two stochastically independent real variables X and Y with cumulative distribution functions:

$$
\begin{aligned}
F_X(x) &= P[X \leq x] \\
F_Y(y) &= P[Y \leq y]
\end{aligned}
$$

A straightforward adaptation of the proof of Feller convolution theorem (Feller, 1971, p 278) allows us to claim that, when the variable x tends to $(-\infty)$, the distribution function $F_{X+Y}(x) = P[X + Y \leq x]$ of the sum $(X + Y)$ is equivalent to the sum of the distribution functions:

$$
F_{X+Y}(x) \sim F_X(x) + F_Y(x) \tag{3.8}
$$

Let $VaR_p(X)$ and $VaR_p(Y)$, as defined in (2.1), be the values at risk respectively associated with the random values X and Y of some portfolio at a given horizon. A first implication of the convolution property (3.8) is that VaR subadditivity is not really an issue if one of the two risks has much thinner tails than the other. Suppose for instance that, when x tends to $(-\infty)$, $F_Y(x)$ is infinitely small with respect to $F_X(x)$. Then, large losses in the aggregate portfolio $(X + Y)$ will likely come from X and thus, for sufficiently small levels p of probability exceedence, $VaR_p(X)$ is the right measure of risk to control in order to control the aggregate $VaR_p(X + Y)$, irrespective of possible violations of subadditivity. In other words, violations of subadditivity may occur but they are negligible for sufficiently small levels p .

A more interesting case occurs when X and Y both have a distribution function with a left-tail behaviour of the same type. Let us consider that for some given function g increasing and unbounded on the positive real line, there are two positive real numbers a_X and a_Y such that:

$$
\begin{aligned}
a_X &= Lim_{x \longmapsto -\infty} g(-x) F_X(x) \\
a_Y &= Lim_{x \longmapsto -\infty} g(-x) F_Y(x)
\end{aligned}
$$

Hence, by application of (3.8), we have:

$$
a_X + a_Y = Lim_{x \longmapsto -\infty} g(-x)[F_{X+Y}(x)]
$$

Then, if we assume for notational simplicity that the cumulative distribution functions F_X and F_Y are continuous and strictly increasing in the neighborhood of $(-\infty)$ to allow an unambiguous definition of inverse functions, we conclude that for sufficiently small levels of probability p of VaR exceedence:

$$\begin{aligned} VaR_p(X) &\sim g^{-1}(\frac{a_X}{p}), \\ VaR_p(Y) &\sim g^{-1}(\frac{a_Y}{p}), \\ VaR_p(X+Y) &\sim g^{-1}(\frac{a_X+a_Y}{p}). \end{aligned}$$

The following proposition is easily deduced from these asymptotic equivalences.

Proposition 3.1

If X and Y are two independent random variables such that for some positive numbers a_X and a_Y *and a given function g strictly increasing and continuous on the positive real line:*

$a_X = Lim_{x\longmapsto -\infty} g(-x)F_X(x)$ *and* $a_Y = Lim_{x\longmapsto -\infty} g(-x)F_Y(x)$
we have:
(i) If for all positive u and *v: $g(u+v) > g(u) + g(v)$,*
There exists $p_0 \in]0,1]$ such that, for any $p \in]0, p_0[$:

$$VaR_p[X+Y] < VaR_p(X) + VaR_p(Y)$$

(ii) If for all positive u and *v: $g(u+v) < g(u) + g(v)$,*
There exists $p_0 \in]0,1[$ such that, for any $p \in]0, p_0[$

$$VaR_p[X+Y] > VaR_p(X) + VaR_p(Y)$$

Proof: See the Appendix.

It is worth noticing that the subadditivity property, maintained above on all the real line either for the function g or for the function $(-g)$, is actually binding only for arbitrarily large u and v. Since the function g has been defined to characterize the tail behaviour of the distribution function, only its behaviour in the neighborhood of $(+\infty)$ really matters.

Distributions with such left-tail behavior are all distributions with Pareto-like left tails given in Feller (1971), $F_X(x) \sim_{x\longrightarrow -\infty} a_X[-x]^{-\alpha}[Log(-x)]^{\gamma}$, with $\alpha > 0$ and γ any real number. Conditions *(i)* and *(ii)* about the concavity or convexity of g translate into conditions on α, i.e. $\alpha > 1$ for *(i)* and $\alpha < 1$ for *(ii)*[2]. Therefore, up to the limit case $\alpha = 1$,

[2] For this distribution, $g(x) \sim_{x\longrightarrow +\infty} x^{\alpha}[Log(x)]^{-\gamma}$
and $g'(x) \sim_{x\longrightarrow +\infty} x^{\alpha-1}[Log(x)]^{-\gamma}\left[\alpha - \frac{\gamma}{Log(x)}\right] > 0$ for $x > Max\left(1, e^{\gamma/\alpha}\right)$.

subadditivity of VaR_p for sufficiently small levels p of probability exceedence is tantamount to finite expectation for absolute returns. In this respect, non-subadditive VaR remains quite an extreme situation[3]. However, in the case of very heavy-tailed distributions (α close to zero), we show in the Appendix that violation of subadditivity may be arbitrarily extreme.

What is most important for our purpose is to be able to ensure subadditivity for sufficiently small levels p of probability of exceedence. We want to ensure that the commonly used small values of p like 1%, 5% or 10% are within the range of maintained subadditivity. To shed more light on the relevant order of magnitude, we propose to consider the case of stable distributions as a benchmark example of variables with Pareto-like tails.

A random variable X is said to follow a stable distribution[4] $S_\alpha(\sigma, \beta, \mu)$ for $2 > \alpha > 0, \alpha \neq 1, \sigma > 0, \mid \beta \mid < 1$ and μ any real number if its characteristic function is given by:

$$E \exp(i\theta X) = \exp\left\{-\sigma^\alpha \left|\theta\right|^\alpha \left[1 - i\beta \left(sign\ \theta\right) \tan \frac{\Pi\alpha}{2}\right] + i\mu\theta\right\}. \tag{3.9}$$

The characteristic function is real if and only if μ=β=0. The parameters σ, β and μ are uniquely defined. μ is a location parameter , σ is a scale parameter and β characterizes the skewness of the distribution: a positive (resp. negative) β implies a distribution skewed to the right (resp. to the left) while a zero β gives a symmetric distribution around μ. In particular:

$$X \rightsquigarrow S_\alpha\left[\sigma, \beta, \mu\right] \Leftrightarrow \frac{X - \mu}{\sigma} \rightsquigarrow S_\alpha\left[1, \beta, 0\right], \tag{3.10}$$

and:

$$X \rightsquigarrow S_\alpha\left[\sigma, \beta, 0\right] \Leftrightarrow (-X) \rightsquigarrow S_\alpha\left[\sigma, -\beta, 0\right].$$

In all cases, the support of the distribution is the whole real line. Note that for the sake of expositional simplicity, we have excluded the limit cases $\alpha = 2$ (normal distribution) $\alpha = 1$, and $|\beta| = 1$ (distribution concentrated on one half of the real line).

This parametric family of distributions has Pareto-like tails. If $X \rightsquigarrow S_\alpha\left[\sigma_X, \beta, \mu_X\right]$, we have:

$$Lim_{x \longmapsto -\infty}(-x)^\alpha F_X(x) = a_X,$$

with:

$$a_X = \frac{(1-\alpha)(1-\beta)\sigma_X^\alpha}{2\Gamma(2-\alpha)\cos(\frac{\Pi\alpha}{2})}.$$

[3]Recently, Ibragimov (2004) obtained a similar result with a different approach based on the analysis of majorization properties of linear combinations of random variables.

[4]See Samorodnitsky and Taqqu (1994) for a thorough treatment of stable distributions.

The advantage is that in this case we can assess the level p_0 below which the subadditivity property is characterized. We can therefore state the following corollary to proposition 3.1.

Corollary 3.2

If X and Y are two independent stable variables with similar tails and the same degree of skewness:

$$X \rightsquigarrow S_\alpha\left[\sigma_X, \beta, \mu_X\right] \text{ and } Y \rightsquigarrow S_\alpha\left[\sigma_Y, \beta, \mu_Y\right],$$

and we consider a probability p of VaR exceedence such that:

$$p < P[S_\alpha(1, \beta, 0) < 0].$$

Then:

(i) If $\alpha > 1$:

$$VaR_p\left[X+Y\right] < VaR_p(X) + VaR_p(Y).$$

(ii) If $\alpha < 1$:

$$VaR_p\left[X+Y\right] > VaR_p(X) + VaR_p(Y).$$

It is important to stress that this corollary can be applied for values of the probability p which are not excessively small. For instance, if $\beta = 0$, it applies for $p < 1/2$. Irrespective of the value of β, it applies exactly when $VaR_p(X) > \mu_X$ and $VaR_p(Y) > \mu_Y$. Note that when $\alpha > 1$, the shift parameter μ is equal to the mean. In other words, it is sufficient to have absolute returns with finite means and to consider possible amounts of losses $VaR_p(X)$ and $VaR_p(Y)$ beyond the opposite of the respective means μ_X and μ_Y to ensure subaddivity of the VaR[5].

3.2 Tails and Conditioning Information

Given the importance of tail thickness for VaR subadditivity, we want to argue in this section that, in general, the larger the conditioning information set is, the thinner the tails should be. Of course, this claim rests upon some measurement of tail thickness. Extending the common idea of kurtosis measurement, we characterize tail thickness through higher-order moments.

[5]Not that we have assumed that X and Y have the same skewness parameter. This assumption, which was not needed to apply the convolution property, may appear overly restrictive to the point where only $\beta = 0$ has some practical content. Hopefully, the subadditivity should not be lost for not too different skewness parameters.

Let us consider some random variable Y such that $|Y|^n$ has a finite expectation for some positive real number n. Let m be another real number such that $0 < m < n$. We argue that the larger the ratio $\frac{E\{|Y|^n\}}{\{E\{|Y|^m\}\}^{\frac{n}{m}}}$ is, the fatter the tails of the distribution of Y should be. According to Malevergne and Sornette (2005), the major contribution to the magnitude of the moment $E\{|Y|^n\}$ comes from the values of Y in the vicinity of the maximum of $|y|^n f_Y(y)$, where $f_Y(y)$ is the probability density function (pdf) of Y. The magnitude of this quantity increases fast with the order of the moment we consider. The fatter are the tails of the pdf of Y, the faster it increases. The above ratio is the standard kurtosis measurement when $m = 2$ and $n = 4$, with Y measured in deviations from the mean. To accomodate the case of variables with possibly infinite variance and even infinite mean, we generalize the standard argument to moments of any order for variables not expressed in mean-deviation form. To characterize the effect of conditioning information, we extend the result previously derived by Clark (1973)[6] to the case of kurtosis.

Let Y and Z two random variables, where, for notational simplicity, we assume that Y is a positive real variable and : $E[(Y)^n] < +\infty$, $0 < m < n$. The tight relationship between conditioning and tail thickness, as measured by the comparison between higher and lower order moments, amounts to say that, more often than not:

$$E\left[\frac{E[(Y)^n \mid Z]}{\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}}\right] < \frac{E[(Y)^n]}{\{E[(Y)^m]\}^{\frac{n}{m}}}. \tag{3.11}$$

In other words, conditioning on the variable Z reduces the distance between higher and lower order moments. We specialize the result to scale mixtures, with Z as a mixture variable, in the following proposition.

Proposition 3.3

If the distribution of the random variable Y is a scale mixture with Z as a mixture variable, that is if $Y = \sigma(Z)u$, with Z and u stochastically independent, and if, in addition: $E\{|u|^n\} < +\infty, E\{|\sigma(Z)|^n\} < +\infty, 0 < m < n$,

Then:

$$E\left[\frac{E[|Y|^n \mid Z]}{\{E[|Y|^m \mid Z]\}^{\frac{n}{m}}}\right] < \frac{E[|Y|^n]}{\{E[|Y|^m]\}^{\frac{n}{m}}}.$$

Proof: See the Appendix.

The inequality of proposition 3.3 is very likely to hold in general[7]. It does hold for a number of common models that are actually scale mixture models. A popular example is

[6] See in particular corollary 4.1.

[7] For instance, in the classical case of a zero-mean variable Y with $m = 2$ and $n = 4$, inequality (6.22) indicates that it would take a perversely high positive correlation betwen conditional kurtosis and conditional variance to reverse the inequality.

the stochastic volatility model without leverage effect, as first introduced by Taylor (1982) as a dynamic extension of Clark (1973). A less-known example is the case of symmetric stable distributions, which can always be seen as scale mixtures of stable distributions with less fat tails. Indeed, according to Samorodnitsky and Taqqu (1994)[8], if X is a random variable with a symmetric α-stable distribution, $X \rightsquigarrow S_{\alpha}(\sigma, 0, 0)$, $0 < \alpha < 2$, then X can be seen as a scale mixture of stable distributions: $X|A \rightsquigarrow S_{\alpha'}(\sigma A^{1/\alpha'}, 0, 0), 0 < \alpha < \alpha'$, where the probability distribution of the mixing variable A is defined by its Laplace transform: $E(\exp(-\gamma A)) = \exp(-\gamma^{\alpha/\alpha'})$. Therefore, a random variable with a symmetric stable distribution[9] can be viewed as a mixture of stable distributions with less fat tails (higher α). This illustrates the general proposition above in terms of higher moments.

In this section we have shown that subadditivity of VaR is intimately related to fat tails and that fat tails are in turn very closely linked to conditioning information. In the next section, we want to use these two main principles to spell out conditions under which a decentralized portfolio management system will work in terms of risk control. These conditions will ensure that the VaR requirement is respected, that is $P[\Sigma_{j=1}^{n} X_j \leq -VaR_p^*] \leq p$.

4 Proper Conditioning for Subadditive VaR

We put forward in this section two crucial assumptions that will ensure VaR subadditivity in the decentralized management system described in section 2. We will assume that these assumptions are valid at a given probability level p, which will be seen as a relevant confidence level for VaR calculations such as 1% or 5%.

The first assumption amounts to consider that, even though subadditivity of VaR is not guaranteed at the senior management level, there exists a latent information I, nesting all individual information sets, such that the conditioning by this information would restore subadditivity of VaR. Of course this conditioning will be unfeasible in practice, but it suffices that traders obey their VaR target for the risk control system to work. Moreover, it shows that eliciting some information from traders such as portfolio shares will be useful in terms of ex-post risk control or backtesting.

Assumption 1: *There exists* $I \supset \cup_{j=1}^{n} I_j$ *such that* $VaR_p(\Sigma_{j=1}^{n} X_j \,|I) \leq \Sigma_{j=1}^{n} VaR_p(X_j \,|I)$.

As seen in section 2, the larger the conditioning information set is, the thinner the tails

[8] See Proposition 1.3.1 p. 20.

[9] This symmetry assumption is rather realistic for distributions produced by portfolio traders. Indeed, the central unit does not need to give a benchmark to the traders in the context of a decentralization portfolio management system. Therefore, the distribution of interest is not the deviations of the trader's portfolio returns from the benchmark, which ought to be skewed to the right, but simply the raw returns of the trader's strategy. The right skewness of the latter returns is less of a necessity.

are. In this case, VaR subadditivity is more likely to hold. In particular, Assumption 1 will be automatically fulfilled if the joint distribution of the vector $(X_j)_{1\leq j\leq n}$ of returns is a multivariate scale mixture, that is for some n-dimensional variable $(u_j)_{1\leq j\leq n}$ conformable to subadditivity (for instance a Gaussian vector with $p \leq \frac{1}{2}$) and independent from conditioning information, $X_j = \sigma_j(I)u_j$, for $j = 1, ..., n$.

The second assumption stated below will be fulfilled if in addition $\sigma_j(I)$ depends on information I only through trader's j information I_j. This appears as a rather natural assumption in such a delegated system where each trader is hired because he holds a specific information.

Assumption 2: *For all* $j = 1, \ldots, n$:$VaR_p(X_j\,|I\,) \leq \; VaR_p(X_j\,|I_j\,)$.

In other words, latent information other than I_j is irrelevant for forecasting the result X_j of trader's j investment. This latter condition, a kind of cross-sectional equivalent to a non-causality assumption (external information does not cause X_j given I_j), is fairly natural if one imagines trader j as an expert of his market segment. Trader j has at his disposal all the relevant information for his market segment[10].

However, assumption 2 is more general than this special case of cross-sectional non-causality. It only means that the part of latent information that trader j does not observe does not affect his perceived potential loss with probability p. In particular, we have:

Proposition 4.1: *Assumption 2 implies that,* I_j *almost certainly:*

$$X_j \leq -VaR_p(X_j\,|I_j\,) \Leftrightarrow X_j \leq -VaR_p(X_j\,|I\,)\,.$$

Under assumption 2, conditioning on the larger latent information set does not change the occurrence of VaR exceedence for j, I_j almost surely[11]. The most important result of this section is stated in the following proposition.

Proposition 4.2: *Under assumptions 1 and 2,* $\Sigma_{j=1}^{n} S_j \leq \; VaR_p^*$ *and* $VaR_p(X_j\,|I_j\,) \leq S_j$ *for all* j *implies that:*

$$P[\Sigma_{j=1}^{n} X_j \leq \; -VaR_p^*] \leq \; p.$$

[10]Note that given I_j, $VaR_p(X_j\,|I\,)$ is a random variable which can be constantly below $VaR_p(X_j\,|I_j\,)$ (with a common level of probability p) only if these two variables actually coincide almost surely. In other words, Assumption 2 is a non-causality property in terms of VaR_p . It is fulfilled in particular in case of global non-causality, that is if for all j, the conditional distributions of X_j given I or I_j coincide.

[11]Note that the converse is not true in general even though we have, by the law of iterated expectations: $I_j \subset I \Longrightarrow \; P[X_j \leq -VaR_p(X_j\,|I\,)|I_j] = p = P[X_j \leq -VaR_p(X_j\,|I_j\,)|I_j]$. But the equality of probabilities does not imply the equality of events.

Proofs for these two propositions are provided in the Appendix.

In other words, the VaR target $VaR_p(X_j\,|I_j\,) \le S_j$ imposed to each specialist j ensures that the VaR of the portfolio $\sum_{j=1}^{n} X_j$ will not exceed $\sum_{j=1}^{n} S_j$. It is worth emphasizing that this result has been obtained while VaR may not be subadditive for senior management, that is $VaR_p(\sum_{j=1}^{n} X_j)$ may exceed $\sum_{j=1}^{n} VaR_p(X_j)$. This convenient result has basically been obtained through an additional conditioning that has restored subadditivity without introducing additional perceived risk thanks to assumptions 1 and 2.

As already mentioned, assumption 2 may also be useful for the purpose of backtesting, or more precisely for ex-post control of the risk-taking behavior of the specialists. Senior management would like to check that the announced target S_j has been respected by specialist j, that is:

$$VaR_p(X_j\,|I_j\,) \le S_j. \tag{4.12}$$

Although senior management cannot observe the information set I_j, it has access to some partial information such as the specialists' actions. Let us assume, at it is often the case in practice, that the portfolio composition $\theta_j(I_j) \in I_j$, selected by each trader j is available to central management. Then, by the law of iterated expectations, (2.3)-(2.4) implies that:

$$P[X_j \le -S_j\,|\theta_j(I_j)] \le p \text{ for all } j = 1, \ldots, n. \tag{4.13}$$

If, as it is often the case, each specialist's information completely defines the return distribution of the fund in which the corresponding trader invests and the private information signals $I_j, j = 1, \cdots, n$, are mutually independent, then (4.13) actually means:

$$P[X_j \le -S_j\,|\theta_k(I_k),\ k = 1, \ldots, n] \le p \text{ for all } j = 1, \ldots, n. \tag{4.14}$$

This condition can actually be tested by senior management from an econometric model of conditional probability distributions, including for instance ARCH effects (Engle, 1982). We discuss this issue in the next section. Note that, without maintaining a joint independence assumption, the non-causality assumption 2 actually implies even more since it ensures that:

$$P[X_j \le -S_j\,|I] \le p \text{ for all } j = 1, \ldots, n. \tag{4.15}$$

Then, the control over trader j behavior appears a priori much more powerful than the solely unconditional control $P[X_j \le -S_j] \le\ p$ that could have been performed without taking advantage of the observation of specialists' actions and possibly resorting to assumption 2.

5 A Simple Illustration of a Rent-a-Trader System

We provide an illustration of the general propositions of the previous sections in a simple setting. The goal of the illustration is to provide a concrete yet basic example where conditioning on the private information of traders restore VaR subadditivity (assumption 1) and where only the information of trader j is relevant in forecasting the result X_j of trader's j investment (assumption 2). We also discuss in the framework of this example how eliciting information for traders can be helpful for backtesting VaR exceedences.

We assume that two traders can each choose a portfolio made up of one risk-free asset and two risky funds. The returns of the two risky funds depend on two state variables s_1 and s_2. State variable s_1 is observable to trader 1, but unobservable to trader 2 and to central management. Similarly, trader 2 is the only one to observe s_2. The two state variables are assumed to be *i.i.d. Bernoulli* (λ).

We can write the returns as $\widetilde{R}_1 = s_1 R_1^1 + (1 - s_1) R^0$, and $\widetilde{R}_2 = s_2 R_2^1 + (1 - s_2) R^0$, where R_1^1, R_2^1 and R^0 are mutually independent with R_1^1 and R_2^1 following the same probability distribution $N\left(\mu_1, \overline{\sigma}^2\right)$ and R^0 following $N\left(\mu_0, \overline{\sigma}^2\right)$. Moreover, the unconditional mean $[\lambda\mu_1 + (1-\lambda)\mu_0]$ of the two distributions is assumed to be equal to the risk-free rate[12]. The unconditional variance of $\widetilde{R}_1$ and $\widetilde{R}_2$ is denoted σ^2. These assumptions imply that, without any information on the state variables, a risk averse agent will only invest in the risk-free asset. Therefore, central management will have an incentive to hire traders 1 and 2, who have private information on state variables s_1 and s_2 respectively. In this context, if each trader forms his portfolio such that the VaR requirement imposed by central management is satisfied conditionally to any possible value of his private information, then the VaR requirement of the global portfolio will be satisfied and the apparent violation of subadditivity will not matter.

We further assume that each trader communicates his portfolio shares to central management. We show how, based on this information, central management can recover statistically the parameters of the conditional distributions of the traders' portfolio returns and assess whether traders have respected the VaR requirement or not. It is important to realize that this is just an example while in the general setting considered above, it has never been assumed that the knowledge of these individual portfolio shares was a sufficient statistic to recover fully the conditioning information of traders and, by the same token, to restore subadditivity.

[12]It is important to realize that funds 1 and 2 have the same conditional means μ_1 in state 1 and μ_0 in state 0 and the same conditional variance $\overline{\sigma}^2$ in any state. They differ only by the realization of the states, which do not necessarily coincide. For example, fund 1 could be in state 0 when fund 2 is in state 1.

5.1 Traders' Behavior

We start by computing the optimal mean-variance portfolio of traders 1 and 2 given their private information on s_1 and s_2 respectively. We assume for simplicity that the VaR of their optimal portfolio is always below S_j[13], the target set by central management.

We can write the portfolio return of trader 1 as $\widetilde{R}^1 = R_f + \theta_{11}\left(\widetilde{R}_1 - R_f\right) + \theta_{12}\left(\widetilde{R}_2 - R_f\right)$. The expectation and the variance conditional on the state are:

$$\begin{aligned} E\left(\widetilde{R}^1 \,|s_1 = i\right) &= R_f + \theta_{11}\left(\mu_i - R_f\right), \\ Var\left(\widetilde{R}^1 \,|s_1 = i\right) &= \theta_{11}^2\overline{\sigma}^2 + \theta_{12}^2\sigma^2. \end{aligned} \tag{5.16}$$

Normalizing initial wealth to one, and denoting the risk aversion coefficient of trader 1 by γ_1, the optimal portfolio is solution of:

$$\underset{\theta_1}{Max}\left\{\left(R_f + \theta_{11}\left(\mu_i - R_f\right)\right) - \tfrac{\gamma_1}{2}\left(\theta_{11}^2\overline{\sigma}^2 + \theta_{12}^2\sigma^2\right)\right\}$$

with $\theta_1 = (\theta_{11}, \theta_{12})$. The solution $\widehat{\theta}_1 = \left(\widehat{\theta}_{11}, \widehat{\theta}_{12}\right)$ is given by:

$$\begin{aligned} \widehat{\theta}_{11} &= \left(\mu_i - R_f\right)/\gamma_1\overline{\sigma}^2, \\ \widehat{\theta}_{12} &= 0. \end{aligned} \tag{5.17}$$

The proportion invested in the risk-free asset is $1 - \widehat{\theta}_{11}$. Trader 1 never invests in fund 2 for which he has no information[14]. Moreover, traders will always include a non-zero share of their respective risky fund in their optimal portfolio along with the risk-free asset. In the good state, they will hold a long position, in the bad state they will sell the risky fund short[15].

Overall, we are typically in a situation where each specialist's information completely defines the return distribution of the fund in which the corresponding trader invests and the private information signals $I_j, j = 1, \ldots n$, are mutually independent. Therefore, assumption 2 is fulfilled and the condition to test for backtesting is just (4.13). Note moreover that assumption 1 is trivially fulfilled for any p smaller than 1/2 since, given the private signals, the joint conditional probability distribution of traders' portfolio returns is normal.

[13] When the VaR constraint of trader 1 (resp. 2) binds, it can be shown that he may have to invest a nonzero part in asset 2 (resp. 1). Therefore, the distribution given the portfolio shares will be a mixture of normals and not a normal. Conditioning will still make the tails less fat as discussed in the earlier sections, but we prefer to keep things simple and recover normality and hence restore subadditivity.

[14] In a general framework, Merton (1987) assumes this result and justifies his assumption by the fact that the portfolios held by actual investors contain only a small fraction of the thousand of traded securities available. In our setting, the result follows directly from the private information held by the traders.

[15] In a framework with only two assets (a risk-free asset and a risky asset with the same unconditional expected return), Sentana (2005) assumes and rationalizes the fact that the wealth invested in the risky asset is a linear function of the information that the investor has on this asset.

5.2 Subadditivity Issues

Since assumptions 1 and 2 are fulfilled, we know from our general analysis above that the Rent-a-Trader system ensures a coherent risk management. However, in the type of setting described in the previous subsection, VaR may typically appear to violate subadditivity from the central management point of view, even for very small levels of confidence probability p. To see this, let us assume for simplicity that both traders have the same risk aversion γ and that their initial wealth is normalized to one. We further assume that both information variables s_1 and s_2 are in the good state ($s_1 = s_2 = 0$). Therefore, it follows that $\widehat{\theta}_{11} = \widehat{\theta}_{22} = \theta > 0$, traders' portfolios are $\widetilde{R}^1 = (1-\theta)\,R_f + \theta\widetilde{R}_1$ and $\widetilde{R}^2 = (1-\theta)\,R_f + \theta\widetilde{R}_2$ and the aggregate portfolio at central management level is $\widetilde{R}^1 + \widetilde{R}^2 = 2\,(1-\theta)\,R_f + \theta\left(\widetilde{R}_1 + \widetilde{R}_2\right)$.

In such a context, since the central management does not observe the private signals, it is confronted with a mixture of normals for which subadditivity of VaR may be violated even at small probability levels p :

Proposition 5.1: *For any mixture probability values* $\lambda < 1/2$, *at the level* $p = p\,(\lambda, \mu_0, \mu_1) = P\left(\widetilde{R}_1 + \widetilde{R}_2 \leq \mu_1 + \mu_0\right)$ *we have*

$$VaR_p\left(\widetilde{R}^1 + \widetilde{R}^2\right) > VaR_p\left(\widetilde{R}^1\right) + VaR_p\left(\widetilde{R}^2\right).$$

Moreover,

$$\begin{cases} p\,(\lambda, \mu_0, \mu_1) = \lambda^2 \Phi\left(-\overline{\mu}/\sqrt{2}\overline{\sigma}\right) + \lambda\,(1-\lambda) + (1-\lambda)^2\,\Phi\,(\overline{\mu}/2\overline{\sigma}) \underset{\overline{\mu}\to-\infty}{\longrightarrow} \lambda \\ p\,(\lambda, \mu_0, \mu_1) \underset{\overline{\mu}\to 0}{\longrightarrow} 1 \end{cases}$$

where Φ *denotes the standard normal distribution function and* $\overline{\mu} \equiv \mu_1 - \mu_0 < 0$.

Proof : see the Appendix.

In other words, the level p where violation of subadditivity occurs can be arbitrary close to λ when the expected return spread between the good and the bad states is sufficiently large. Then p can take any value between 0 and 1 when λ is sufficiently small. Therefore, a small probability λ of occurrence of the bad state will produce a violation of subadditivity of VaR, even for small levels of the confidence probability level p. We thus have an example of the surprising situation where subadditvity is violated even though decentralized risk management works, insofar as traders remain true to the VaR requirements sent to them by central management.

5.3 Backtesting VaR requirements

In this simple model, a central unit can test condition (4.13), which can be written $P[X_j \leq -S_j\,|\theta_k(I_k), k = 1, 2] \leq p$ for all j, when the shares of the portfolios held by traders 1 and 2 are known. It is important to realize that in this setting, knowing the portfolio composition of traders 1 and 2 is equivalent to knowing their private information s_1 and s_2. Indeed, if trader 1 takes a long position in risky fund 1 it means that s_1 is in the good state and inversely if he short-sells fund 1. Similarly, the position of trader 2 will be fully revealing. We can write $s_j = 1_{\{\theta_{jj}<0\}}$. Since each private information completely defines the return distribution of the fund in which the corresponding trader invests and given that the two informations s_1 and s_2 are independent, the condition to test is exactly $P[X_j \leq -S_j \, |\theta_j(s_j)] \leq p$ for $j = 1, 2$.

By the conditional normality of the return distributions, central management needs only to infer the means and variances in the good and bad states for each trader in order to test if each trader obeys his VaR limit. In section 4, we assumed that central management knew the underlying model. In practice, central management must estimate the model based on time series of returns and compute VaR conditionally on past returns. In this dynamic setting, assumption A2 can be rewritten as:

$$VaR_p \left[X_j|I_t, I_\tau, \tau \leq t\right] \leq VaR_p \left[X_j|I_{jt}, I_\tau, \tau \leq t\right]. \tag{5.18}$$

In other words, we will assume that all past information becomes common knowledge. Propositions equivalent to Propositions 4.1 and 4.2 can be derived in a dynamic context. Engle and Manganelli (2004) provide a useful framework to estimate value at risk in a dynamic context. They remark that VaR is simply a particular quantile of future portfolio values, conditional on current information. They provide a specification (CAViaR, Conditional Autoregressive Value at Risk) for calculating VaR_t as a function of variables known at time $t-1$ (which in our case could be the portfolio shares of the traders) and a set of parameters that are estimated using Koenker and Bassett's (1978) regression quantile framework.

6 Conclusion

In this paper, we have argued that, in the context of decentralized portfolio management, central management possesses only a fraction of information which belongs to each specialist. In such a context, a distribution appears always thicker to the central unit than to the specialist. Therefore, because of a lack of information, VaR may appear fallaciously non subadditive to the central management unit. We were then able to show that decentralized portfolio management with a VaR allocation to each specialist will work despite evidence to the contrary. Furthermore, we have shown that value at risk remains subadditive in many situations of practical interest. In the case of heavy-tail distributions, we have shown, at least for sufficiently small probabilities, that VaR remains subadditive when the possible loss has finite expectation. In this respect, non-subadditive VaR remains quite an extreme situation.

Even though we show that for all practical purposes VaR is not as incoherent a measure of risk as it is often argued, it remains that portfolio optimization practices using VaR as a simple substitute to variance (i.e. maximization of expected return under a VaR constraint) may generate perverse effects. In particular, there is a risk that a manager who is controlled only through a maximal loss level with probability $(1 - p)$ will be enticed to expose himself to huge possible losses with probability p, as demonstrated by Basak and Shapiro (2001). To control for such a risk, one can add to VaR the expected loss beyond the VaR or consider a parameterized family of more limited possible risks. Alexander and Baptista (2004) compare VaR and conditional VaR constraints on portfolio selection with a mean-variance model. Ortobelli, Rachev and Schwartz (2000) provide a thorough analysis of optimal portfolio allocation with stable distributed returns, including a safety-first optimal allocation problem, whereby investors maximize their expected wealth while minimizing at the same time the risk of loss. Efficient frontiers in the latter case are a function of the threshold VaR.

References

[1] Alexander, G. J. and A. M. Baptista (2004), "A Comparison of VaR and CVaR Constraints on Portfolio Selection with the Mean-Variance Model," *Management Science,* 50, 9, 1261–1273.

[2] Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 3, 203-228.

[3] Basak, S. and A. Shapiro (2001), "Value-at-risk-based risk management: Optimal policies and asset prices," *Review of Financial Studies,* 14, 371–405.

[4] Clark, P. K. (1973), "A subordinated stochastic process model with finite variance for speculative prices," Econometrica, 41,135-156.

[5] Danıelsson, J., H. S. Shin and J.-P. Zigrand (2004), "The impact of risk regulation on price dynamics," *Journal of Banking and Finance, 28, 1069–1087.*

[6] Engle, R. (1982) "Autoregressive Conditional Heteroskedasticity Models with estimation of Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

[7] Engle, R. and S. Manganelli (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics,* Volume 22, Number 4, 367-381

[8] Fama, E. (1965), "Behavior of Stock Market Prices," Journal of Business, 38, 34-105.

[9] Feller, W. (1971), An introduction to Probability Theory and Its Applications, Vol. II, 2nd edition. Wiley, New York.

[10] Ibragimov, R. (2004), "On the Robustness of Economic Models to Heavy-Tailedness Assumptions", Working Paper, Yale University.

[11] Jorion, P. (2001) *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, New York.

[12] Koenker, R. and G. Bassett (1978), "Regression Quantiles," Econometrica, 46, 33-50.

[13] Malevergne, Y. and D. Sornette (2005), *Extreme Financial Risk: From dependence to risk management*, Springer Verlag.

[14] Mandelbrot, B. (1963), "The Variation of Speculative Prices", *Journal of Business*, 36, 394-419.

[15] Merton, R.C., (1987), "A Simple Model of Capital Market Equilibrium with Incomplete Information", *Journal of Finance*, 42, 483-510.

[16] Mittnik, S. and S. T. Rachev (1993), "Modeling Asset Returns with Alternative Stable Distributions", *Econometric Reviews*, 12(3), 261-330.

[17] Mittnik, S., M. S. Paolella,and S. T. Rachev (2000), "Diagnosis and treating the fat tails in financial returns data", *Journal of Empirical Finance*, 7, 389-416.

[18] Ortobelli, S., S. T. Rachev and E. Schwartz (2000), "The Problem of Optimal Asset Allocation with Stable Distributed Returns", Working Paper, Anderson Graduate School of Management.

[19] Samorodnitsky and M. S. Taqqu (1994), "Stable Non-Gaussian Random Processes - Stochastic Models with Infinite Variance," Chapman &Hall, New York.

[20] Sentana, E. (2005), "Least Squares Predictions and Mean-Variance Analysis", *Journal of Financial Econometrics,* 3(1), 56-78.

[21] Taylor, S. J. (1982), "Financial returns modelled by the product of two stochastic processes: a study of daily sugar prices 1961-79," in O. D. Anderson (Ed.), Time Series Analysis: Theory and Practice, 1, 203-226. Amsterdam: North-Holland.

Appendix

Proof of Proposition 3.1

To prove this proposition we need the following Lemma.

Lemma: *For an increasing function g, statements (a) and (b) below are equivalent*

(a) $g(u+v) > g(u) + g(v)$ *for all positive u and v,*

(b) $g^{-1}(u+v) < g^{-1}(u) + g^{-1}(v)$, *for all positive u and v.*

Proof of Lemma:

Suppose that (a) is true but (b) is not. So there exists positive u_0 and v_0 such that $g^{-1}(u_0+v_0) \geq g^{-1}(u_0) + g^{-1}(v_0)$. Since g is increasing, we have:

$g\left(g^{-1}(u_0+v_0)\right) \geq g\left(g^{-1}(u_0) + g^{-1}(v_0)\right) \overset{\text{by (a)}}{>} g\left(g^{-1}(u_0)\right) + g\left(g^{-1}(v_0)\right)$, that is $u_0 + v_0 > u_0 + v_0$, an impossibility. Therefor, by contradiction (a) $\Rightarrow$ (b) and similarly (b) $\Rightarrow$ (a), and the lemma follows.

For the proposition, we have:

$$\begin{aligned} g^{-1}(\frac{a_X}{p}) &= Lim_{p\longmapsto 0} VaR_p(X), \\ g^{-1}(\frac{a_Y}{p}) &= Lim_{p\longmapsto 0} VaR_p(Y), \\ g^{-1}(\frac{a_X+a_Y}{p}) &= Lim_{p\longmapsto 0} VaR_p(X+Y) \end{aligned}$$

so, $Lim_{p\longmapsto 0}\left[VaR_p(X+Y) - VaR_p(X) - VaR_p(Y)\right] = g^{-1}(\frac{a_X+a_Y}{p}) - g^{-1}(\frac{a_X}{p}) - g^{-1}(\frac{a_Y}{p})$.

(i) If for all positive u and v: $g(u+v) > g(u)+g(v)$, then $g(\frac{a_X+a_Y}{p}) > g(\frac{a_X}{p})+g(\frac{a_Y}{p})$. This implies

$g^{-1}(\frac{a_X+a_Y}{p}) < g^{-1}(\frac{a_X}{p}) + g^{-1}(\frac{a_Y}{p})$. Therefore:

$Lim_{p\longmapsto 0}\left[VaR_p(X+Y) - VaR_p(X) - VaR_p(Y)\right] < 0$. So, there exists $p_0 \in \left]0,1\right]$ such that, for any $p \in \left]0, p_0\right[$:

$$VaR_p\left[X+Y\right] < VaR_p(X) + VaR_p(Y).$$

(ii) If for all positive u and v: $g(u+v) < g(u)+g(v)$, then $g(\frac{a_X+a_Y}{p}) < g(\frac{a_X}{p})+g(\frac{a_Y}{p})$. This implies

$g^{-1}(\frac{a_X+a_Y}{p}) > g^{-1}(\frac{a_X}{p}) + g^{-1}(\frac{a_Y}{p})$. Therefore:

$Lim_{p\longmapsto 0}\left[VaR_p(X+Y) - VaR_p(X) - VaR_p(Y)\right] > 0$.So, there exists $p_0 \in \left]0,1\right]$ such that, for any $p \in \left]0, p_0\right[$:

$$VaR_p\left[X+Y\right] < VaR_p(X) + VaR_p(Y), \mathbf{Q.E.D.}$$

Measuring the Degree of Violation of Subadditivity:

We propose below a way to measure the degree of violation of subadditivity. While nonsubadditivity means:

$$VaR_p\,(X+Y) > VaR_p(X) + VaR_p(Y), \tag{6.19}$$

that is the loss of the portfolio $(X+Y)$ can exceed $VaR_p(X)+VaR_p(Y)$ with probability p, the question is with what probability $kp, k \geq 1$, the loss of the portfolio $(X+Y)$ can exceed $VaR_p(X)+VaR_p(Y)$ with probability p, that is:

$$VaR_{kp}(X+Y) > (VaR_p(X) + VaR_p(Y). \tag{6.20}$$

While violation (6.19) of subadditivity means that (6.20) is fulfilled with $k=1$, it cannot be fulfilled with $k=2$ since, for any random variables X and Y:

$$VaR_{2p}(X+Y) \leq VaR_p(X) + VaR_p(Y) \tag{6.21}$$

Indeed:

$$P\left[X+Y \leq -VaR_p(X) - VaR_p(Y)\right] \leq 2p$$

$$\begin{aligned} & P\left[X+Y \leq -VaR_p(X) - VaR_p(Y)\right] \\ \leq\ & P\left[X \leq -VaR_p(X)\right] + P\left[Y \leq -VaR_p(Y)\right] = 2p. \end{aligned}$$

However, we can show for any k in $]0,2[$, there exists α_0 in $]0,1[$ such that for any α in $]0,\alpha_0[$, for $F_X(x) \sim_{x\longrightarrow -\infty} a_X[-x]^{-\alpha}$, there exists p_0 in $]0,1[$ such that for any p in $]0,p_0[$:

$$VaR_{kp}(X+Y) > VaR_p(X) + VaR_p(Y)$$

As a function of $\alpha \in\ ]0,1[$, the function $\dfrac{[\sigma_X+\sigma_Y]^\alpha}{\sigma_X^\alpha+\sigma_Y^\alpha}$ is continuous and increasing, from the value $1/2$ for $\alpha=0$ to the value 1 for $\alpha=1$. Then, for k given in $]0,2[$, there exists α_0 such that, for any α in $]0,\alpha_0[$:

$$\frac{[\sigma_X+\sigma_Y]^\alpha}{\sigma_X^\alpha+\sigma_Y^\alpha} < \frac{1}{k},$$

that is:

$$\sigma_X + \sigma_Y < k^{-1/\alpha}\left[\sigma_X^\alpha + \sigma_Y^\alpha\right]^{1/\alpha} = k^{-1/\alpha}\sigma_Z.$$

Since $\sigma_X = \lim\limits_{p\to 0} p^{1/\alpha} VaR_p(X)$, $\sigma_Y = \lim\limits_{p\to 0} p^{1/\alpha} VaR_p(Y)$
and $\sigma_Z = \lim\limits_{p\to 0}(kp)^{1/\alpha} VaR_{kp}(z)$, we have the existence of p_0 such that for any p in $]0,p_0[$:

$$p^{1/\alpha} VaR_p(X) + p^{1/\alpha} VaR_p(Y) < k^{-1/\alpha}(kp)^{1/\alpha} VaR_{kp}(Z)$$

that is $VaR_p(X) + VaR_p(Y) < VaR_{kp}(Z)$.

Proof of Proposition 3.3

Proposition 3.3. is a direct consequence of equation (3.11). Indeed, one can write:

$$\begin{aligned} E[(Y)^n] &= E\{E[(Y)^n \mid Z]\} = E\{k_{m,n}(Z)\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} \\ &= E\{k_{m,n}(Z)\}E\{\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} + Cov\{k_{m,n}(Z), \{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} \end{aligned}$$

with $k_{m,n}(Z) = \frac{E[(Y)^n|Z]}{\{E[(Y)^m|Z]\}^{\frac{n}{m}}}$. Note that, since $0 < m < n$, Jensen's inequality gives:

$$E\{\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} \geq \{E\{E[(Y)^m \mid Z]\}\}^{\frac{n}{m}} = \{E[(Y)^m]\}^{\frac{n}{m}}.$$

The inequality becomes strict as soon as the conditioning information Z is not independent from Y. The spread widens when the information content of Z about Y increases. Therefore, we conclude that, as soon as Z and Y are not independent:

$$\frac{E[(Y)^n]}{\{E[(Y)^m]\}^{\frac{n}{m}}} > E\{k_{m,n}(Z)\} + \frac{Cov\{k_{m,n}(Z), \{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\}}{E\{k_{m,n}(Z)\}} \tag{6.22}$$

To prove Proposition 3.3, we simply note that, in the case of a scale mixture, $k_{m,n}(Z)$ is a constant equal to $\frac{E[|u|^n]}{\{E[|u|^m]\}^{\frac{n}{m}}}$.

Proof of Proposition 4.1:

For A random event, we define the random variable:

$$\mathbf{1}_A = \begin{cases} 1 \text{ if } A \text{ occurs} \\ 0 \text{ otherwise.} \end{cases}$$

Assumption A2 implies that:

$$\mathbf{1}_{[X_j \leq -VaR_p(X_j|I_j)]} \leq \mathbf{1}_{[X_j \leq -VaR_p(X_j|I)]}.$$

But these two random variables have the same conditional expectation given I_j since by definition:

$$\begin{aligned} E\left[\mathbf{1}_{[X_j \leq -VaR_p(X_j|I_j)]} \,|I_j\right] &= P\{X_j \leq -VaR_p(X_j\,|I_j)\,|I_j\} \\ &= p \end{aligned}$$

and by the law of iterated expectations:

$$\begin{aligned} E\left[\mathbf{1}_{[X_j \leq -VaR_p(X_j|I)]} \,|I_j\right] &= E\left[E\left[\mathbf{1}_{[X_j \leq -VaR_p(X_j|I)]} \,|I\right] |I_j\right] \\ &= E\left[P\{X_j \leq -VaR_p(X_j\,|I)\,|I\}\,|I_j\right] \\ &= E\left[p\,|I_j\right] \\ &= p \end{aligned}$$

Therefore, these two random variables must coincide I_j almost surely. This achieves the proof of proposition 4.1.

Proof of Proposition 4.2:

Since for all $j, VaR_p(X_j | I_j) \leq S_j$ and, by assumption A2:

$$VaR_p(X_j | I) \leq VaR_p(X_j | I_j)$$

we have:

$$p(I) = P\left[\sum_{j=1}^{n} X_j \leq -\sum_{j=1}^{n} S_j | I\right] \leq P\left[\sum_{j=1}^{n} X_j \leq -\sum_{j=1}^{n} VaR_p(X_j | I) | I\right]$$

But, by assumption A1:

$$VaR_p\left(\sum_{j=1}^{n} X_j | I\right) \leq \sum_{j=1}^{n} VaR_p(X_j | I).$$

Thus:

$$p(I) \leq P\left[\sum_{j=1}^{n} X_j \leq -VaR_p(\sum_{j=1}^{n} X_j | I) | I\right] = p.$$

Since this inequality is identically true, for all possible values of the random information set I, we conclude by considering unconditional expectations that:

$$P\left[\sum_{j=1}^{n} X_j \leq -\sum_{j=1}^{n} S_j\right] \leq p.$$

A fortiori:

$$P\left[\sum_{j=1}^{n} X_j \leq -VaR_p^*\right] \leq p$$

Proof of Proposition 5.1:

Let us define the following variables

$$X_1 \equiv R_1^1 - \mu_0 \sim N\left(\overline{\mu}, \overline{\sigma}^2\right)$$
$$X_2 \equiv R_2^1 - \mu_0 \sim N\left(\overline{\mu}, \overline{\sigma}^2\right)$$
$$Z \equiv R^0 - \mu_0 \sim N\left(0, \overline{\sigma}^2\right)$$

we have

$$\widetilde{R}_1 = s_1 X_1 + (1 - s_1) Z + \mu_0$$
$$\widetilde{R}_2 = s_2 X_2 + (1 - s_2) Z + \mu_0$$

and

$$\begin{aligned} p &= P\left[\widetilde{R}_1 + \widetilde{R}_2 \leq \mu_1 + \mu_0\right] \\ &= P\left[s_1 X_1 + (1 - s_1) Z + s_2 X_2 + (1 - s_2) Z \leq \mu_1 - \mu_0\right] \\ &= P\left[s_1 s_2 (X_1 + X_2) + s_1 (1 - s_2)(X_1 + Z) + \right. \\ &\qquad \left. + s_2 (1 - s_1)(X_2 + Z) + 2(1 - s_1)(1 - s_2) Z \leq \overline{\mu}\right] \\ &= \lambda^2 F_{X_1 + X_2}(\overline{\mu}) + \lambda (1 - \lambda) F_{X_1 + Z}(\overline{\mu}) + \lambda (1 - \lambda) F_{X_2 + Z}(\overline{\mu}) + (1 - \lambda)^2 F_{2Z}(\overline{\mu}) \\ &= \lambda^2 F_{X_1 + X_2}(\overline{\mu}) + \lambda (1 - \lambda) + (1 - \lambda)^2 F_{2Z}(\overline{\mu}) \end{aligned}$$

i.e. $p = \lambda^2 \Phi\left(-\overline{\mu}/\sqrt{2}\overline{\sigma}\right) + \lambda (1 - \lambda) + (1 - \lambda)^2 \Phi\left(\overline{\mu}/2\overline{\sigma}\right)$

Let us define

$$U_1 \equiv s_1 X_1 + (1 - s_1) Z$$
$$U_2 \equiv s_2 X_2 + (1 - s_2) Z$$

We can see that $U_j = \widetilde{R}_j - \mu_0$, $j = 1, 2$. Therefore the proposition is equivalent to $VaR_p(U_1 + U_2) > VaR_p(U_1) + VaR_p(U_2)$. By definition of p, we have $VaR_p(U_1 + U_2) = -\overline{\mu}$, and since $U_1 \overset{d}{=} U_2$ then $VaR_p(U_1) = VaR_p(U_2)$ and the problem becomes $VaR_p(U_1 + U_2) > 2VaR_p(U_1)$ i.e. $VaR_p(U_1) < -\overline{\mu}/2$ or equivalently $\Pr(U_1 \leq \overline{\mu}/2) < p$.

Therefore, we have:

$$\begin{aligned} P(U_1 \leq \overline{\mu}/2) &= P\left[s_1 X_1 + (1 - s_1) Z \leq \overline{\mu}/2\right] \\ &= \lambda F_{X_1}(\overline{\mu}/2) + (1 - \lambda) F_Z(\overline{\mu}/2) \end{aligned}$$

i.e. $P(U_1 \leq \overline{\mu}/2) = \lambda \Phi(-\overline{\mu}/2\overline{\sigma}) + (1 - \lambda) \Phi(\overline{\mu}/2\overline{\sigma})$

so

$$\begin{aligned} p - P(U_1 \leq \overline{\mu}/2) &= \lambda^2 \left[\Phi\left(-\overline{\mu}/\sqrt{2}\overline{\sigma}\right) + \Phi(\overline{\mu}/2\overline{\sigma}) - 1\right] \\ &> 0 \end{aligned}$$

since $\Phi\left(-\overline{\mu}/\sqrt{2}\overline{\sigma}\right) > \Phi(-\overline{\mu}/2\overline{\sigma})$ and $\Phi(-\overline{\mu}/2\overline{\sigma}) + \Phi(\overline{\mu}/2\overline{\sigma}) = 1$

i.e. $P(U_1 \leq \overline{\mu}/2) < p$, and the proposition follows. **Q.E.D**