Fat Tails, VaR and Subadditivity*

Jón Daníelsson
London School of Economics

Bjørn N. Jorgensen
University of Colorado at Boulder

Gennady Samorodnitsky
Cornell University

Mandira Sarma
Jawaharlal Nehru University, Delhi

Casper G. de Vries
Erasmus University Rotterdam, Tinbergen Institute Rotterdam

This version: March 2012. First version 2005.
Forthcoming Journal of Econometrics

Abstract

Financial institutions rely heavily on Value–at–Risk (VaR) as a risk measure, even though it is not globally subadditive. First, we theoretically show that the VaR portfolio measure is subadditive in the relevant tail region if asset returns are multivariate regularly varying, thus allowing for dependent returns. Second, we note that VaR estimated from historical simulations may lead to violations of subadditivity. This upset of the theoretical VaR subadditivity in the tail arises because the coarseness of the empirical distribution can affect the apparent fatness of the tails. Finally, we document a dramatic reduction in the frequency of subadditivity violations, by using semi–parametric extreme value techniques for VaR estimation instead of historical simulations.

KEY WORDS: Value–at–Risk, subadditivity, fat tailed distribution, extreme value estimation

JEL Classification: G00, G18

---

*Corresponding author Jón Daníelsson, `j.danielsson@lse.ac.uk`. Danielsson acknowledges the financial support of the EPSRC grant no. GR/S83975/01. Samorodnitsky's research was partially supported by NSF grant DMS-0303493 and NSA grant MSPF-02G-183 at Cornell University. Sarma and De Vries acknowledge that major part of their work on the paper was done when they were associated with EURANDOM, Eindhoven University of Technology. We thank the referees for valuable comments. We benefitted from discussions with the participants of the Ecares conference on latest developments in heavy–tailed distributions. The first version of the paper in 2005 was called "Sub–additivity re–examined: the case for Value–at–Risk"

1 Introduction

Risk measurements have become an integral part of the operation of financial institutions and financial regulations, and most proposals for regulatory reform due to the crisis emphasize better understanding of risk. While a large number of risk measures exist, value–at–risk (VaR) remains the most widely used risk measure. The reason is that its practical advantages are perceived to outweigh its theoretical deficiencies. We argue that such a preference is often theoretically and empirically justified.

VaR has been an integral part in banks' risk management operations ever since being mandated by the 1996 *amendment to incorporate market risk* to the Basel I Accord, and continuing with Basel II. Over time, its importance has increased, with financial institutions and non–financials alike, routinely using VaR in areas such as internal risk management, economic capital and compensation.

VaR has remained preeminent even though it suffers from the theoretical deficiency of not being subadditive as demonstrated by Artzner et al. (1999). In spite of this deficiency, both industry and regulators in the banking sector have a clear preference for VaR over subadditive risk measures such as expected shortfall (ES) because of its practical advantages, primarily smaller data requirements, ease of backtesting and, in some cases ease of calculation. By contrast, the use of ES is becoming more prevalent in insurance. From an industry and regulatory perspective it is important to identify whether such a practically motivated preference is justified.

VaR is known to be subadditive in some special cases such as when asset returns are normally distributed in the area below the mean, or more generally for all log–concave distributions, see Ibragimov (2005). This is, however, not all that relevant since asset return distributions exhibit fat tails, see e.g. Mandelbrot (1963), Fama (1965) and Jansen and de Vries (1991). The implications of this for VaR are discussed in Daníelsson et al. (2005) and Ibragimov (2005). Using majorization theory, Ibragimov and Walden (2007), Ibragimov (2009) and Garcia et al. (2007) demonstrate that the VaR measure is subadditive for the infinite variance stable distributions provided the mean return is finite, the latter for general Pareto tails; see also the review in Marshall et al. (2011, ch. 12). Their results extend earlier work of Fama and Miller (1972, page 270) who discuss the effects of portfolio diversification when returns follow stable distributions. Daníelsson et al. (2005), Ibragimov (2005) and Garcia et al. (2007) also discuss cases of VaR subadditivity for distributions with Pareto type tails when the variance is finite.

Most asset returns belong to neither category, normal or the infinite variance stable, and it is of considerable practical importance to know whether the industry preference for VaR is reasonable in such cases. Our main motivation is to investigate the subadditivity of VaR for fat–tailed distributions in general, and we arrive at three key results.

First, we identify sufficient conditions for VaR to be subadditive in the relevant tail region for fat–tailed and dependent distributions. In this context, fat tails means that the tails vary regularly, so that they approximately follow a multivariate power law such as the Pareto distribution. Note that the infinite variance stable distributions are a subset of this class. Specifically, we prove that VaR is subadditive in the relevant tail region when asset returns exhibit multivariate regular variation, for both independent and cross sectionally dependent returns provided the mean is finite. Interestingly, Ibragimov (2005, 2009) shows that this holds for distributions that are in the intersection of the alpha–symmetric class and the regularly varying class; the multivariate Student–t distributions are part of this intersection. But the class of distributions with regularly varying tails is much broader than this intersection, as is the class of alpha–symmetric distributions. We construct an explicit example of interdependent returns based on the portfolio view of interbank connectedness as discussed in e.g. Shin (2009). The only exception is asset returns that are so extremely fat tailed that the first moment — the mean — becomes infinite, what we label *super fat tails*, the case discussed by Ibragimov and Walden (2007), Garcia et al. (2007) and Ibragimov (2009). But in that case, any risk subadditive measure dependent on the existence of the first moment, such as ES, is not defined.

Second, we investigate these asymptotic results by means of Monte Carlo simulations, and find that this asymptotic result may not hold in practice because of small sample sizes and choice of estimation methods. In particular, estimation of VaR by historical simulation (HS) is prone to deliver violations of subadditivity in some cases, especially for increasingly extreme losses and small sample sizes. The reason is what we call the *tail coarseness* problem. When only using a handful of observations in the estimation of HS, where the estimate is equal to one of the most extreme quantiles, the uncertainty about the location of a specific quantile is considerable, and one could easily get draws whereby a particular loss quantile of a relatively fat distribution is lower than the same quantile from a thinner distribution. This could also induce failures of subadditivity in empirical applications, even though theoretically subadditivity holds.

Finally, we demonstrate how this estimation problem can be remedied by employing the extreme value theory (EVT) semi–parametric estimation method for VaR, proposed by Daníelsson and de Vries (2000). Their EVT–based estimator corrects for most empirical subadditivity failures by exploiting a result from EVT which shows that regardless of the underlying distribution, so long as the data is fat tailed, the asymptotic tail follows a power law, just like the Pareto distribution. In effect, this method is based on fitting a power law through the tail, thus smoothing out the tail estimates and rendering the estimated VaR much less sensitive to the uncertainty surrounding any particular quantile. Ultimately this implies that subadditivity violations are mostly avoided.

The rest of the paper is organized as follows. Section 2 discusses the concept of sub-additivity. In section 3 we formally define fat tails. Our main theoretical results are obtained in section 4 with extensive proofs relegated to an Appendix. The Monte Carlo experiments are discussed in section 5 along with the estimator comparisons. Section 6 concludes the paper.

2 Subadditivity

Artzner et al. (1999) propose a classification scheme for risk measures whereby a risk measure $\rho(\cdot)$ is said to be "coherent" if it satisfies the four requirements of homogeneity, monotonicity, translation invariance and subadditivity. VaR[1] satisfies the first three requirements, but fails subadditivity. Let X_1 and X_2 denote the random returns to two financial assets. A risk measure $\rho(\cdot)$ is subadditive if

$$\rho\left(X_1 + X_2\right) \leq \rho\left(X_1\right) + \rho\left(X_2\right).$$

Subadditivity is a desirable property for a risk measure because, consistent with the diversification principle of modern portfolio theory, a subadditive measure should generate lower measured risk for a diversified portfolio than for a non–diversified portfolio.

In response to the lack of subadditivity for the VaR risk measure, several alternatives have been proposed. The most common of these alternative risk measures are expected shortfall, ES, proposed by Acerbi et al. (2001) and worst conditional expectation proposed by Artzner et al. (1999). While these risk measures are theoretically considered superior to VaR, because they are subadditive, they have not gained much traction in practice.[2] Subadditiv-

[1] Let X_1 be the return, then for the probability p, VaR is the loss level such that $\mathrm{VaR} = -\sup\left\{x \middle| \Pr(X_1 \leq x) \leq p\right\}$.

[2] See e.g. Yamai and Yoshiba (2002) for more on the practical problems with alternative risk measures.

ity of positive homogeneous risk measures guarantees their convexity, which facilitates the identification of optimal portfolios, see e.g. Pflug (2005); Stoyanov et al. (2007). For example, Daníelsson et al. (2007) show that ES remains very useful in portfolio optimization problems, since it imposes a linear constraint, while VaR is a non–linear constraint resulting in the optimization problem being NP complete.

2.1 Statistical Violations of Subadditivity

That VaR can violate subadditivity is easily demonstrated.[3] A simple example with continuous distributions is:

Example 1 *Consider two assets X_1 and X_2 that are usually normally distributed, but subject to the occasional independent shocks:*

$$X_i = \epsilon_i + \eta_i, \ \epsilon_i \sim \text{IID}\,\mathcal{N}(0,1), \ \eta_i = \begin{cases} 0 & \text{with probability} \quad 0.991 \\ \\ -10 & \text{with probability} \quad 0.009 \end{cases} \qquad i = 1, 2.$$

The 1% VaR for X_1 is 3.1, which is only slightly higher than the VaR if the shocks η would not happen, in which case they would be 2.3. Asset X_2 follows the same distribution as asset X_1, whilst being independent from X_1. Compare a portfolio composed of one X_1 and one X_2 to a portfolio of 2 X_1. In the former case, the 1% portfolio VaR is 9.8, because for $(X_1 + X_2)$ the probability of getting the -10 draw for either X_1 or X_2 is higher than 1%.

$$\text{VaR}(X_1 + X_2) = 9.8 > \text{VaR}(X_1) + \text{VaR}(X_2) = 3.1 + 3.1 = 6.2.$$

This example is especially relevant in the area of credit risk where credit events are represented by the -10 outcome.

Alternatively, we can illustrate subadditivity violations with the following discrete example. The discrete case is of interest when we turn to the Monte Carlo study, as data samples are necessarily discrete.

Example 2 *Suppose we throw two dice five times and obtain the following results*

[3] See e.g. Artzner et al. (1999); Acerbi and Tasche (2001); Acerbi et al. (2001).

	Dice 1	*Dice 2*	*Dice 1 + Dice 2*
Throw 1	*2*	*4*	*6*
Throw 2	*3*	*1*	*4*
Throw 3	*4*	*5*	*9*
Throw 4	*5*	*6*	*11*
Throw 5	*6*	*6*	*12*

The VaR estimates at probability 1/3 are:

	Dice 1	*Dice 2*	*Sum*	*Dice 1 + Dice 2*
VaR	*-3*	*-4*	*-7*	*-6*

Note that the VaR at $p = 1/3$ are the realizations at the second lowest throw since $1/3 \leq 2/5$, see the definition in Footnote 1. One shows that theoretically the VaR of rolling two dice is subadditive below the mean. But in this experiment, the VaR happens *not* to be subadditive below the mean as $-6 > -7$. Recall that definition of VaR in Footnote 1 and the fact that all outcomes are positive, imply that the VaR is a negative number.

3 Fat–tailed Asset Returns

Empirical studies have long established that the distribution of speculative asset returns tend to have fatter tails than the normal distribution, see e.g. Mandelbrot (1963), Fama (1965) and Jansen and de Vries (1991). Fat tailed distributions are often defined in terms of higher than normal kurtosis. However, kurtosis captures the mass of the distribution in the center relative to the tails, which may be thin. Distributions exhibiting high kurtosis but having truncated tails, and hence thin tails, are easy to construct.[4]

An alternative, formal, definition of a fat tailed distribution is that the tails are regularly varying at infinity, i.e., the tails have a Pareto distribution–like power expansion at infinity.

Definition 1 *A cumulative distribution function $F(x)$ varies regularly at*

[4]The issue of kurtosis is discussed e.g. in the example on page 480 of Campbell et al. (1997).

minus infinity with tail index $\alpha > 0$ if[5]

$$\lim_{t\to\infty} \frac{F(-tx)}{F(-t)} = x^{-\alpha} \quad \forall x > 0 \tag{1}$$

and at plus infinity if

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha} \quad \forall x > 0.$$

This implies that a regularly varying distribution has a tail of the form

$$F(-x) = x^{-\alpha}L(x),\ x > 0,$$

where the constant $\alpha > 0$ is called the *tail index* and L is a *slowly varying function*, e.g. a logarithm.[6] An often used particular class of these distributions has a tail comparable to the Pareto distribution:

$$F(-x) = Ax^{-\alpha}[1 + o(1)],\ x > 0, \quad \text{for } \alpha > 0, \tag{2}$$

where the parameter $A > 0$ is known as the *scale coefficient*. A regularly varying density implies regularly varying tails for the distribution as defined in (1). Under a weak extra condition regarding monotonicity, the converse also holds, i.e. for large x condition (1) implies

$$f(-x) \approx \alpha L(x)x^{-\alpha-1} \quad x > 0, \quad \text{for } \alpha > 0 \text{ and } A > 0\,. \tag{3}$$

This means that the density declines at a power rate $x^{-\alpha-1}$ far to the left of the center of the distribution, which contrasts with the much faster than exponentially declining tails of the Gaussian. The power is outweighed by the explosion of x^m in the computation of moments of order $m > \alpha$. Thus, moments of order $m > \alpha$ are infinite and α therefore determines the number of finite moments and hence the thickness of the tails. Finiteness of the moments is determined by α, apart from the boundary case of moment of order α, in which case the slowly varying function plays a role.[7]

For example, the Student–t distributions vary regularly at infinity, have degrees of freedom equal to the tail index and satisfy the above approximation.

[5]For an encyclopedic treatment of regular variation, see Bingham et al. (1987) or Resnick (1987).

[6]A function $L(x)$ is slowly varying if $L(tx)/L(t) \to 1$ as $t \to \infty$ for any $x > 0$.

[7]In the case of a two–sided power law, the sum of the two tails determines finiteness of the moments (since α could be the same in both cases).

Likewise, the stationary distribution of the GARCH(1,1) process has regularly varying tails, see de Haan et al. (1989) and Basrak et al. (2002). Moreover, the non–normal stable distributions investigated by Fama and Miller (1972, page 270), Ibragimov and Walden (2007) and Ibragimov (2009) also exhibit regularly varying tails at infinity. See also Davis and Mikosch (1998) and Campbell et al. (1997).

The first moment of most financial assets appears to be finite, indicating a tail index higher than one, see e.g. Jansen and de Vries (1991), Embrechts et al. (1996) and Daníelsson and de Vries (1997, 2000). We demonstrate below that for all assets with (jointly) regularly varying non–degenerate tails, subadditivity holds in the tail region provided the tail index is exceeds one.

An example of assets with such a distribution is the return distributions of non–life insurance portfolios which are characterized by tails with α values that hover around 1 (which is one explanation for why most insurance treaties are capped). For example, weather insurance is plagued by occasional bad weather leading heavy damage claims, after many years without any noticeable storms. But for other applications in finance, a finite mean (when $\alpha > 1$) or a finite variance ($\alpha > 2$), is more common.

4 Subadditivity of VaR in the Tail

While the normal distribution with linear dependence delivers subadditive VaR below the mean, our interest is in the empirically more relevant fat tailed distributions. We only need to focus on the lower tail, since the theoretical results apply equally to the upper tail since one can turn it into the other tail by multiplying returns with minus one, accomplished e.g. in a short sale.

As before, let X_1 and X_2 be two asset returns, each having a regularly varying tail with the same tail index $\alpha > 0$. We consider the effect of combining the assets into one portfolio, which requires studying the tail of the convolution that is determined by the joint tail behavior of the two assets. The corresponding formal mathematical definition of jointly regularly varying tails allows X_1 and X_2 to be dependent:

Definition 2 *A random vector (X_1, X_2) has regularly varying right tails with tail index α if there is a function $a(t) > 0$ that is regularly varying at infinity with exponent $1/\alpha$ and a nonzero measure μ on $(0, \infty)^2 \setminus \{0\}$ such that*

$$t\,P\left((X_1, X_2) \in a(t)\cdot\right) \to \mu \tag{4}$$

as $t \to \infty$ vaguely in $(0, \infty]^2 \setminus \{0\}$ (see e.g. Resnick, 1987).

The measure μ has a scaling property

$$\mu(cA) = c^{-\alpha}\mu(A) \tag{5}$$

for any constant $c > 0$ and any Borel set A. The non-degeneracy assumption in Proposition 1 below means that the measure μ is not concentrated on e.g. a straight line $\{ax = by\}$ for some $a, b \geq 0$.

The following general proposition, which is our main theoretical result, allows for arbitrary dependence between the returns. If the tail indices of the two assets are different, a slightly weaker form of subadditivity holds; see the Appendix.

Proposition 1 *Suppose that X_1 and X_2 are two asset returns with jointly regularly varying non–degenerate tails with tail index $\alpha > 1$. Then VaR is subadditive sufficiently deep in the tail region.*

Proof. See the Appendix

Proposition 1 guarantees that at sufficiently low probability levels, the VaR of a portfolio position is lower than the sum of the VaRs of the individual positions, if the return distribution exhibits fat tails. For example, this applies to a multivariate Student–t distribution with degrees of freedom larger than 1. Ibragimov (2009) shows that for models with common shocks and convolutions of finite mean stable distributions subadditivity holds, regardless of the value of the loss probability. Ibragimov (2005, 2009) also shows that subadditivity holds for the class of finite variance alpha–symmetric distributions with regularly varying tails, such as the multivariate Student–t distribution.

Remark 1 *From the proof to Proposition 1 in the Appendix, we see that even without the non-degeneracy assumptions and, in particular, if the two assets have different tail indices, we still have*

$$\limsup_{p \to 0} \frac{\mathrm{VaR}_p(X_1 + X_2)}{\mathrm{VaR}_p(X_1) + \mathrm{VaR}_p(X_2)} \leq 1\,,$$

which is a weaker form of subadditivity in the tails.

The following example is well known.

Example 3 *Suppose X_1 and X_2 have independent unit scale Pareto loss distributions,* $\Pr\{X_1 < -x\} = \Pr\{X_2 < -x\} = x^{-\alpha}, x \geq 1$. *By inversion,* $\mathrm{VaR}_p(X_1) = \mathrm{VaR}_p(X_2) = p^{-1/\alpha}$. *Using Feller's convolution theorem (Feller, 1971, VIII.8), we have for sufficiently low p:*

$$p = \Pr\{X_1 + X_2 \leq -\mathrm{VaR}_p(X_1 + X_2)\} \approx 2[\mathrm{VaR}_p(X_1 + X_2)]^{-\alpha}.$$

Hence, if $\alpha > 1$ then for low p:

$$\mathrm{VaR}_p(X_1 + X_2) - [\mathrm{VaR}_p(X_1) + \mathrm{VaR}_p(X_2)] \approx p^{-1/\alpha}\left(2^{\frac{1}{\alpha}} - 2\right) < 0.$$

A caveat is that diversification may not work for super fat tails, i.e. if $\alpha < 1$. Data falling into this category are characterized by a large number of very small outcomes inter–dispersed with very large outcomes. This result was noted by Fama and Miller (1972). Ibragimov and Walden (2007) and Ibragimov (2009) extend these results to the VaR risk measure for the class of sum stable distributions and possibly dependent processes. These issues are further discussed by Embrechts et al. (2008).

4.1 Affine Dependent Returns

Proposition 1 establishes that subadditivity is not violated for fat tailed data, deep in the tail area, regardless of its dependency structure. We can illustrate this result by an example of assets with linear dependence, via a factor structure. Other relevant cases are discussed in de Vries (2005) for the financial assets and Geluk and de Vries (2006) for insurance. Garcia et al. (2007) considered the case of two independent returns.

Consider the standard single factor model, where X_1 and X_2 are two assets, which are dependent via a common market factor:

$$X_i = \beta_i\ R + \varepsilon_i,\ i = 1, 2 \tag{6}$$

where R denotes the risky return of the market portfolio, β_i the constant market factor loading and ε_i the idiosyncratic risk of asset X_i. The random variable ε_i and R are independent of each other; and individual ε_is are independent of each other. Thus, the only source of cross–sectional dependence between X_1 and X_2 is the common market risk.

Since R and ε_i are independent, we can use Feller (1971)'s convolution theorem to approximate the tails of X_1 and X_2, depending upon the tail behavior of R, ε_1 and ε_2. We can further use it to approximate the tail of $X_1 + X_2$.

To illustrate this, we present below one particular case, viz., the case where R, ε_1 and ε_2 have regularly varying Pareto-like tails with the same tail index α, but with different scale coefficients A, see (2).[8]

Corollary 1 *Suppose that asset returns X_1 and X_2 can be modelled by the single index market model, where R, ε_1 and ε_2 all have Pareto–like tails with tail index $\alpha > 1$, and scale coefficients $A_r > 0$, $A_1 > 0$ and $A_2 > 0$ respectively, as in (2). Then the VaR measure is subadditive in the tail region.*

Proof. See the Appendix.

In general, the single index market model (6) may not describe the true nature of the dependence between X_1 and X_2 since ε_i's may not be cross sectionally independent, even when each of them is independent from the common market factor R. For example, apart from the market risk, the assets X_1 and X_2 may be dependent on industry specific risk, depicted by the movement of an industry index I. Moreover, typically the number of factors is larger. Such industry specific factors may lead to dependence between ε_1 and ε_2. We may model cross sectional dependence by generalizing model (6) by incorporating a sector specific factor I.

$$X_i = \beta_i\ R + \tau_i\ I + \varepsilon_i,\ i = 1, 2 \tag{7}$$

where I is the risky industry specific factor and the constant τ_i represents the effect of the industry specific risk on the asset X_i. If X_i has Pareto–like tails with scale coefficient A and tail index α, then again under the assumption of Proposition 1 for sufficiently small p:

$$\mathrm{VaR}_p(X_1 + X_2) \le \mathrm{VaR}_p(X_1) + \mathrm{VaR}_p(X_2).$$

To show the full scope of Proposition 1, we now consider a case where there is zero correlation, but where portfolios may nevertheless be dependent.

Example 4 *Consider two independent random returns X_1 and X_2, and the following two portfolios $X_1 + X_2$ and $X_1 - X_2$. Assume alternatively that the returns are standard normally distributed, or Student-t with $\alpha > 2$ degrees of freedom. It is immediate that $E[(X_1 + X_2)(X_1 - X_2)] = 0$, and hence the correlation is zero. So under normality the two portfolios are independent.*

[8]The result in the following Corollary is e.g. shown in the first version of this paper, Daníelsson et al. (2005). Ibragimov (2005, 2009) and Ibragimov and Walden (2007) considered the case $\beta_i = 1$ and when R and ϵ_i are part of the alpha–symmetric distributions.

In the case of the Student–t, however, the two portfolios are dependent in the tail area since the extremes line up along the two diagonals.

The implication for the VaR of the portfolio is as follows. For the normal case, below the mean the VaR is known to be subadditive. For the non-linear dependent case with the Student-t risk drivers, one can calculate the VaR sufficiently deep into the tail area by using Feller's convolution theorem. Since for large s

$$\begin{aligned} p &= \Pr(X_1 + X_2 > s) = \Pr(X_1 - X_2 > s) \\ &\simeq 2s^{-\alpha}, \end{aligned}$$

upon inversion, the univariate VaR's are $s \simeq (2/p)^{1/\alpha}$.

The VaR of the combination of the portfolios is obtained from

$$p = \Pr(X_1 + X_2 + X_1 - X_2 > s) = \Pr(2X_1 > s) \simeq 2^{\alpha} s^{-\alpha}$$

upon inversion $s \simeq 2\,(1/p)^{1/\alpha}$. *It follows immediately that this VaR is smaller than the sum of the individual VaRs* $2\,(2/p)^{1/\alpha}$.

In a stylized way, the first portfolio could be interpreted as belonging to a bank that is lending long in two sectors, while the other portfolio might be from a hedge fund, short one sector, long the other.

5 Monte Carlo Study and Empirical Results

The theoretical subadditivity property established in Proposition 1 only holds in the tail region, and conceivably might only hold for more extreme probabilities than those encountered in practical applications, or in very large data sets.

To investigate this issue, we conducted Monte Carlo experiments with two asset returns X_1 and X_2, assumed to follow a Student–t distribution with ν degrees of freedom; recall that the tail index $\alpha = \nu$ for the t distribution. We consider several different values for ν, i.e. $1, 2, 3, 4$.

We constructed linearly dependent random variables X_1 and X_2 by taking linear combinations of two independent Student–t variates using the Choleski decomposition of the correlation matrix, i.e. $X_2 = \rho X_1 + \sqrt{(1-\rho^2)}\tilde{X}_2$ where $\tilde{X}_2$ is independent from X_1.[9] The data are bivariate regularly varying by

[9] In a strict sense, the terminology of covariance matrix is not appropriate for the case of $\alpha \leq 2$, since then the second moment does not exist. However, one can still create linear combinations and dependency as we do here.

construction. But the linear combination of Student–t variates behind the dependent X_1 and X_2 implies that the data are not bivariate Student–t, as the convolution of two Student variates preserves the fat tail property, but does not conform to the multivariate Student–t distribution.

We chose the sample sizes, N, to represent both very large samples, expected to give asymptotic results, as well as a smaller samples representing typical applications. The largest sample size is set to 100,000 while the smaller sample sizes are 1000 and 300. In each case, we simulate two asset returns and form an equally–weighted portfolio of the returns to estimate the VaR.

The probability levels, p, are chosen to capture those typically used in practice, i.e., 5% and 1%, as well as some much smaller probability levels for the larger samples to explore the asymptotic properties. These lower probability levels are representative for levels that are used in stress tests and worst case analysis. In the tables the probability levels are indicated by p.

5.1 VaR Subadditivity Violations

Comparing Tables 1 through 3 for probability levels of 1% and 5%, we observe that the frequency of VaR subadditivity violations decreases in the sample size when $\nu > 1$.

Subadditivity fails most frequently when $\nu = 1$, and less so when the degrees of freedom increase. Our simulation results for $\nu = 1$ are in line with Fama and Miller (1972), Ibragimov and Walden (2007) and Ibragimov (2009). When $\nu = 1$, we are at the border between the situation where diversification is counterproductive and productive, since when $\nu < 1$, diversification increases risk.

Reading across the rows in the Tables, VaR subadditivity violations decrease as the probability levels are increased if $\nu > 1$. In some cases the magnitudes of the VaR subadditivity violations is nevertheless substantial. Figure 1 shows the histogram of the magnitudes, for $\nu = 2$, $N = 300$, 100,000 simulations and $p = 1\%$.

At first glance, these results may run counter to Proposition 1. The explanation for this is the finite sample properties of the data, as explained by the following experiment. Let $\nu = 3$ with independent variables and $N = 300$. We record the number of violations at all probability levels $2/N$, $3/N$, $4/N$, until $1/2$. The results are shown in Figure 2. Note the J–shaped pattern.[10]

[10] For the normal distribution one observes the same J–shape at a lower scale; the explanation of this phenomenon is analogous to the case of the Student–t.

The Student–t distribution is subadditive below the mean in the case of independent returns. Thus, at $p = 0.5$, we expect a violation in 50% of the cases, since at that point we expect the VaR for either X_1 or X_2 to have switched sign. Moving to the left, away from the mean into the tail region, thus lowers the number of violations, as seen in Figure 2.

Deep into the tail area, however, at $p = 0.1$, the number of violations starts to increase again, since estimation of VaR by historical simulation (HS) is prone to deliver violations of subadditivity in some cases, especially for increasingly extreme losses and small sample sizes. The reason is that as p decreases, the VaR is estimated by a quantile increasingly close to the minimum, where the empirical distribution becomes very coarse in comparison to the true distribution so far out in the tail.

In other words, the tail is sampled imprecisely in this area because of what we call the *tail coarseness* problem. When only using a handful of observations in the estimation of HS, i.e. where the estimate is equal to one of the most extreme quantiles, the uncertainty about the location of a specific quantile is considerable. This implies that one could easily obtain draws whereby a particular quantile of a relativity fat distribution is less extreme than the same quantile from a thinner distribution. This can imply an upset of subadditivity. See Example 2 above for a demonstration of this result.

Figures 3 and 4 further illustrate this, with the latter showing the 99% quantiles of HS estimation of VaR as we vary the threshold from 2 to 20 in a sample of size 1000.

5.2 VaR from Estimating the Tail

In this section we offer a remedy for the tail coarseness problem identified above, suggesting an alternative estimator for the VaR. We propose to use the quantile estimator of Daníelsson and de Vries (2000) which is based on extreme value theory (EVT).

For fat–failed distributions, the tail asymptotically follows a power law, i.e. the Pareto distribution,

$$F(x) = 1 - Ax^{-\alpha}$$

Given a sample of size n and $m < n$ sufficiently small, one can estimate α by the Hill estimator

$$\frac{1}{\hat{\alpha}(m)} = \frac{1}{m} \sum_{i=1}^{m} \log \frac{X_{(i)}}{X_{(m+1)}} \tag{8}$$

where $X_{(i)}$ indicates order statistics. The quasi maximum likelihood VaR estimator is

$$\mathrm{VaR}(p) = X_{(m+1)} \left(\frac{m/n}{p} \right)^{\widehat{1/\alpha(m)}}. \tag{9}$$

The two estimators, 8 and 9, are asymptotically normally distributed, see de Haan and Ferreira (ch. 2, 2006).

Effectively, EVT estimation is based on fitting a smooth function (Pareto) to the tails. Because this function is estimated by using all observations in the tail, the estimates are less sensitive to the tail coarseness problem. This power law behavior can be reliably estimated by using more data than just the most extreme observations.[11] Subsequently, one joins the parametrically estimated tail to the empirical distribution in the region where there are sufficient observations. Refer to Figure 3 for how EVT provides an estimate of a smooth tail.

We compare the number of VaR subadditivity violations obtained by using HS with the EVT method for $N = 1,000$ and $N = 10,000$ in Tables 4 and 5. We vary the probability levels as in the previous tables, but focus on the $\nu = 2$ case. Finally, we use several thresholds for the EVT estimation, i.e. m from (8). Note that the number of subadditivity violations for the HS from Table 4 are comparable to the results from Table 2 (for example there are 12,897 violations with $\rho = 0.5$ and $p = 0.003$ in Table 4, while with a hundred times as many simulations in Table 2 there are 1,294,552 violations.

EVT reduces the number of violations considerably. For example, for $N = 1,000$, $p = 0.01$, and $m = 100$, HS has 926 subadditivity violations, out of 100,000 simulations, whilst EVT has only 4. Similar results obtain in other cases. In the worst case we get about 30% reduction in violations, and in the best cases 100%. This is further supported in Figure 3 which presents the empirical upper tail and the EVT estimated tail, and Figure 4 which shows the empirical 99% confidence bounds for the VaR estimates from HS and EVT. The EVT bounds are much tighter than the HS bounds.

5.3 Empirical Study

We finally investigate the frequency of subadditivity violations for the stock returns making up the S&P–500 index. If we could use all 500 stocks we would get 124,750 pairs of stock returns for the analysis. The sample size

[11] In fact, for any sequence $m/n \to 0, m \to \infty$, this approach is better than relying on the empirical distribution. The latter approach only guarantees asymptotic normality if $m/n \to c \geq 0$, see de Haan and Ferreira (ch. 3 and 4, 2006).

is 5,000, and since not all stocks in the S&P–500 have 5,000 observations (about 20 years), some had to be removed from the sample. Furthermore, we eliminated from the sample all stock pairs where not all the dates did match. This results in 49,141 stock pairs.

The results are reported in Table 6. We apply both the HS and EVT methods, use two EVT thresholds, 50 and 125 (1% and 2.5%), and employ a range of probabilities for VaR. The average correlation across all the stock pairs is 20.6%, and the average tail index for the smaller threshold is 3.27, whilst it is 2.77 for the larger threshold.

We do not find any subadditivity violations for non-extreme VaR probabilities (1%) but as we move into the tail, the frequency of violations increases. HS is much more likely to violate subadditivity than EVT, consistent with the Monte Carlo simulations, but still a few violations are found for EVT.

We also report the average tail index and correlations in cases where we observe subadditivity violations. The correlations are much higher than for the entire sample (the lowest, 30.4%, for HS where $p = 0.1\%$ and highest for EVT where $p = 0.1\%$ compared to 20.6% for all stock pairs) and the average tail indices are always lower than for the full sample. The number of violations, given the number of observations, is in line with the numbers found in the simulations, e.g. those reported in Table 5.

6 Conclusion

We first show that VaR is subadditive in the relevant tail region when asset returns are multivariate regularly varying, and possibly dependent. Second, Monte Carlo simulations show that coarseness of the empirical distribution can upset the subadditivity of VaR in practice. The final contribution of the paper is that the use of semi–parametric extreme value techniques, dramatically reduces the frequency of subadditivity failures in practice. This approach exploits the fact that the tail of the distribution eventually becomes smooth and can only take on a specific parametric form.

Appendix

Proposition 1 deals with left tails, but for notational simplicity the argument below treats right tails.

Proof of Proposition 1. For $p > 0$ small,

$$\mathrm{VaR}_p(X_1) \sim \Big(\mu\Big\{(1,\infty)\times(0,\infty)\Big\}\Big)^{1/\alpha}\, a\left(\frac{1}{p}\right),$$

$$\mathrm{VaR}_p(X_2) \sim \Big(\mu\Big\{(0,\infty)\times(1,\infty)\Big\}\Big)^{1/\alpha}\, a\left(\frac{1}{p}\right)$$

and

$$\mathrm{VaR}_p(X_1+X_2) \sim \Big(\mu\Big\{x\geq 0,\, y\geq 0 :\, x+y>1\Big\}\Big)^{1/\alpha}\, a\left(\frac{1}{p}\right)$$

as $p \to 0$.

The scaling property (5) means that there is a finite measure η on $B_1 = \{x \geq 0,\, y \geq 0 :\, x+y=1\}$ such that

$$\mu(A) = \int_{B_1}\int_0^\infty \mathbf{1}((u,v)r \in A)\,\alpha r^{-(1+\alpha)}\, dr\, \eta(du,dv)\,. \tag{10}$$

Then

$$\mu\Big\{(1,\infty)\times(0,\infty)\Big\} = \int_{B_1} u^\alpha\, \eta(du,dv)\,,$$

$$\mu\Big\{(0,\infty)\times(1,\infty)\Big\} = \int_{B_1} v^\alpha\, \eta(du,dv)\,,$$

and

$$\mu\Big\{x\geq 0,\, y\geq 0 :\, x+y>1\Big\} = \int_{B_1} (u+v)^\alpha\, \eta(du,dv)\,.$$

Since by the triangular inequality in $L^\alpha(\eta)$

$$\left(\int_{B_1} (u+v)^\alpha\, \eta(du,dv)\right)^{1/\alpha}$$

$$< \left(\int_{B_1} (u)^\alpha\, \eta(du,dv)\right)^{1/\alpha} + \left(\int_{B_1} (v)^\alpha\, \eta(du,dv)\right)^{1/\alpha},$$

with the strict inequality under the non-degeneracy assumption, we conclude that

$$\mathrm{VaR}_p(X_1+X_2) - \mathrm{VaR}_p(X_1) - \mathrm{VaR}_p(X_2) < 0$$

holds for all $p > 0$ small enough. ■

Proof of Corollary 1. Results follow from the previous proof in the general case. But we also provide a constructive proof here.

Suppose that R has a regularly varying tail with index α and $\varepsilon_i,\ i = 1, 2$ has a regularly varying tail with index α. Further, suppose that R has a symmetric distribution. Thus, to a first order approximation,

$$\Pr\{R \leq -x\} \approx A_r\ x^{-\alpha}, \Pr\{R \geq x\} \approx A_r\ x^{-\alpha}.$$

If $\beta_i > 0$ then

$$\Pr\{\beta_i R \leq -x\} = \Pr\left\{R \leq -\frac{x}{\beta_i}\right\} \approx A_r \beta_i^{\alpha} x^{-\alpha}$$

If $\beta_i < 0$ then

$$\Pr\{\beta_i R \leq -x\} = \Pr\{-|\beta_i| R \leq -x\} = \Pr\{|\beta_i| R \geq x\} \approx A_r |\beta_i|^{\alpha}\ x^{-\alpha}.$$

Thus

$$\Pr\{\beta_i R \leq -x\} \approx A_r |\beta_i|^{\alpha}\ x^{-\alpha}, \quad \beta_i \in R$$

For the individual assets ε_1 and ε_2

$$\Pr\{\varepsilon_i \leq -x\} \approx A_i\ x^{-\alpha},\ i = 1, 2$$

By Feller's convolution theorem

$$\Pr\{X_i \leq -x\} \approx |\beta_i|^{\alpha} A_r\ x^{-\alpha} + A_i x^{-\alpha}.$$

Thus

$$p \approx x^{-\alpha} \left(A_i + |\beta_i|^{\alpha} A_r\right),$$

and upon inversion

$$x \approx p^{-\frac{1}{\alpha}}\ \left(A_i + |\beta_i|^{\alpha} A_r\right)^{\frac{1}{\alpha}}.$$

Similarly

$$\Pr\{X_1 + X_2 \leq -x\} \approx |\beta_1 + \beta_2|^{\alpha} A_r\ x^{-\alpha} + A_1 x^{-\alpha} + A_2 x^{-\alpha}$$

Thus,

$$VaR_p(X_1) \approx p^{-\frac{1}{\alpha}}\ \left(A_1 + |\beta_1|^{\alpha} A_r\right)^{\frac{1}{\alpha}},$$
$$VaR_p(X_2) \approx p^{-\frac{1}{\alpha}}\ \left(A_2 + |\beta_2|^{\alpha} A_r\right)^{\frac{1}{\alpha}},$$

and

$$VaR_p(X_1 + X_2) \approx p^{-\frac{1}{\alpha}} \left[(A_1 + A_2 + |\beta_1 + \beta_2|^\alpha A_r)^{\frac{1}{\alpha}} \right].$$

To establish the sub-additivity we proceed as follows:

$$\begin{aligned}
VaR_p(X_1 + X_2) &\approx p^{-\frac{1}{\alpha}} \left[(A_1 + A_2 + |\beta_1 + \beta_2|^\alpha A_r)^{\frac{1}{\alpha}} \right] \\
&\leq p^{-\frac{1}{\alpha}} \left[A_r \left(|\beta_1| + |\beta_2| \right)^\alpha + (A_1 + A_2) \right]^{\frac{1}{\alpha}} \\
&= p^{-\frac{1}{\alpha}} \left[A_r \left(|\beta_1| + |\beta_2| \right)^\alpha + \left((A_1 + A_2)^{\frac{1}{\alpha}} \right)^\alpha \right]^{\frac{1}{\alpha}} \\
&\leq p^{-\frac{1}{\alpha}} \left[A_r \left(|\beta_1| + |\beta_2| \right)^\alpha + \left(A_1^{\frac{1}{\alpha}} + A_2^{\frac{1}{\alpha}} \right)^\alpha \right]^{\frac{1}{\alpha}} \\
&= p^{-\frac{1}{\alpha}} \left[\left(A_r^{\frac{1}{\alpha}} |\beta_1| + A_r^{\frac{1}{\alpha}} |\beta_2| \right)^\alpha + \left(A_1^{\frac{1}{\alpha}} + A_2^{\frac{1}{\alpha}} \right)^\alpha \right]^{\frac{1}{\alpha}} \\
&\leq p^{-\frac{1}{\alpha}} \left[\left(\left(A_r^{\frac{1}{\alpha}} |\beta_1| \right)^\alpha + \left(A_1^{\frac{1}{\alpha}} \right)^\alpha \right)^{\frac{1}{\alpha}} + \left(\left(A_r^{\frac{1}{\alpha}} |\beta_2| \right)^\alpha + \left(A_2^{\frac{1}{\alpha}} \right)^\alpha \right)^{\frac{1}{\alpha}} \right], \\
&= p^{-\frac{1}{\alpha}} \left(A_r |\beta_1|^\alpha + A_1 \right)^{\frac{1}{\alpha}} + p^{\frac{1}{\alpha}} \left(A_r |\beta_2|^\alpha + A_2 \right)^{\frac{1}{\alpha}} \\
&= VaR_p(X_1) + VaR_p(X_2).
\end{aligned}$$

Where in the second step we use the triangular inequality and in the fourth step the C_α inequality for $\alpha > 1$. The sixth step relies on Minkowski's inequality for $\alpha > 1$. Thus, for $\alpha > 1$, VaR is sub-additive in the tail region.
■

7 Tables and Graphs

Table 1: Number of subadditivity violations from a Student–t with HS estimation of VaR. N=300. Number of simulations is 10,000,000

The columns are degrees of freedom of the Student–t, ν, the correlation coefficient ρ and the number of VaR subadditivity violations corresponding to various probability levels p (1% to 5%).

		VaR probabilities p		
ν	ρ	0.01	0.02	0.05
1	0.0	$2,873,140$	$3,724,601$	$4,265,379$
1	0.5	$3,067,383$	$3,978,156$	$4,442,592$
2	0.0	$594,762$	$346,238$	$104,406$
2	0.5	$1,426,493$	$1,366,805$	$974,471$
3	0.0	$147,372$	$40,576$	$4,131$
3	0.5	$783,880$	$598,916$	$323,415$
4	0.0	$50,053$	$8,499$	413
4	0.5	$533,671$	$354,449$	$162,767$

Table 2: Number of subadditivity violations from a Student–t with HS estimation of VaR. N=1,000 Number of simulations is 10,000,000

The columns are degrees of freedom of the Student–t, ν, the correlation coefficient ρ and the number of VaR subadditivity violations corresponding to various probability levels p.

		VaR probabilities p			
ν	ρ	0.003	0.005	0.01	0.05
1	0.0	$2,860,556$	$3,541,288$	$4,048,271$	$4,610,815$
1	0.5	$3,044,504$	$3,798,688$	$4,278,082$	4707254
2	0.0	$530,151$	$325,367$	$91,850$	246
2	0.5	$1,294,552$	$1,241,907$	$842,265$	$131,120$
3	0.0	$100,874$	$27,688$	$1,599$	0
3	0.5	$594,967$	$439,985$	$181,330$	$5,926$
4	0.0	$24,332$	$3,427$	60	0
4	0.5	$337,345$	$206,357$	$59,990$	753

Table 3: Number of subadditivity violations from a Student–t with HS estimation of VaR. N=10,000, Number of simulations is 10,000,000

The columns are degrees of freedom of the Student–t, ν, the correlation coefficient ρ and the number of VaR subadditivity violations corresponding to various probability levels p .

		VaR probabilities p				
ν	ρ	0.0003	0.0005	0.001	0.01	0.05
1	0.0	$2,857,166$	$3,538,949$	$4,049,058$	$4,717,315$	$4,877,893$
1	0.5	$3,036,434$	$3,793,203$	$4,275,794$	$4,793,464$	$4,909,396$
2	0.0	$499,603$	$284,187$	$60,018$	0	0
2	0.5	$1,214,032$	$1,144,662$	$698,389$	453	0
3	0.0	$76,748$	$15,161$	284	0	0
3	0.5	$457,543$	$302,667$	$80,975$	0	0
4	0.0	$12,970$	908	3	0	0
4	0.5	$190,711$	$91,285$	$11,205$	0	0

Table 4: Number of subadditivity violations for a simulation from a Student–t(2) with EVT and HS estimation. N=1000. Number of simulations is 100,000

The first two columns are the EVT threshold m (i.e. number of observations in the tail to estimate tail index) and the VaR probability p. The last four columns record the number of the violations for both EVT and HS where the data was generated with two correlation coefficients, $\rho = 0$ and $\rho = 0.5$.

			Violations		
			$\rho = 0$		$\rho = 0.5$
m	p	EVT	HS	EVT	HS
200	0.01	32	926	3,692	8,398
100	0.01	4	926	1,125	8,398
50	0.01	0	926	583	8,398
200	0.003	414	5,316	8,559	12,897
100	0.003	147	5,316	5,722	12,897
50	0.003	262	5,316	6,007	12,897
10	0.003	1,071	5,316	8,059	12,897

Table 5: Number of subadditivity violations for a simulation from a Student–t(2) with EVT and HS estimation. N=10,000. Number of simulations is 100,000

The first two columns are the EVT threshold m (i.e. number of observations in the tail to estimate tail index) and the VaR probability p. The last four columns record the number of the violations for both EVT and HS where the data was generated with two correlation coefficients, $\rho = 0$ and $\rho = 0.5$.

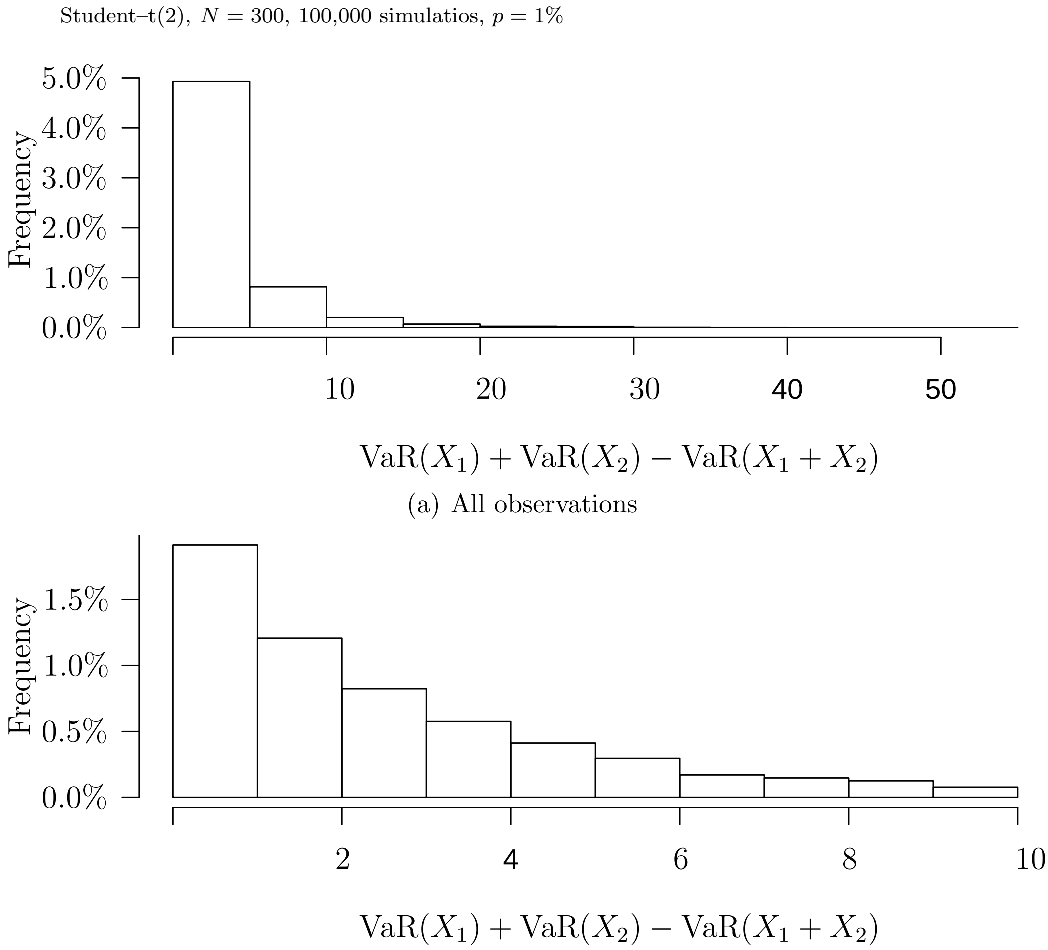
			Violations		
			$\rho = 0$		$\rho = 0.5$
m	p	EVT	HS	EVT	HS
1000	0.01	0	0	0	6
500	0.01	0	0	0	6
200	0.01	0	0	0	6
1000	0.001	0	541	6	7,002
500	0.001	0	541	15	7,002
200	0.001	0	541	127	7,002
100	0.001	0	541	213	7,002
50	0.001	0	541	242	7,002
1000	0.0003	0	4,978	46	12,087
500	0.0003	0	4,978	179	12,087
200	0.0003	2	4,978	1,338	12,087
100	0.0003	27	4,978	3,246	12,087
50	0.0003	151	4,978	5,068	12,087

Table 6: Number of subadditivity violations for S&P–500 stocks

Daily returns from April 31, 1991 to March 31, 2011, $n = 5,000$. VaR only estimated for stock pairs where 5,000 observations were available and all dates for both stocks correspond. This results in 49,141 pairs of stocks. The average tail index for the the smaller and higher threshold are $\bar{\alpha}(50) = 3.27$ and $\bar{\alpha}(125) = 2.77$, respectively, while the average correlation is $\bar{\rho} = 20.6\%$. The first two columns are the EVT threshold indicated by m (i.e. number of observations in the tail to estimate tail index) and the VaR probability at p. This is followed by two pairs of three columns, first pair for EVT, the second for HS. Within each pair the first column is the number of subadditivity violations, the second ($\bar{\alpha}$) is the average tail index and the the last ($\bar{\rho}$) the average correlations, for the subset in which subadditivity is violated. Note that the HS results are necessarily the same for the two cases of m.

		EVT			HS	
m	p	violations	$\bar{\alpha}$	$\bar{\rho}$	violations	$\bar{\rho}$
50	1.0%	0			0	
50	0.5%	0			12	57.3%
50	0.1%	89	2.41	46.8%	410	30.4%
125	2.0%	0			0	
125	1.0%	0			0	
125	0.5%	0			12	57.3%
125	0.1%	59	2.14	50.0%	410	30.4%

Figure 1: Magnitude of VaR violations



(b) Observations less than 10

Figure 2: Number of VaR subadditivity violations for a Student–t(3). N=300

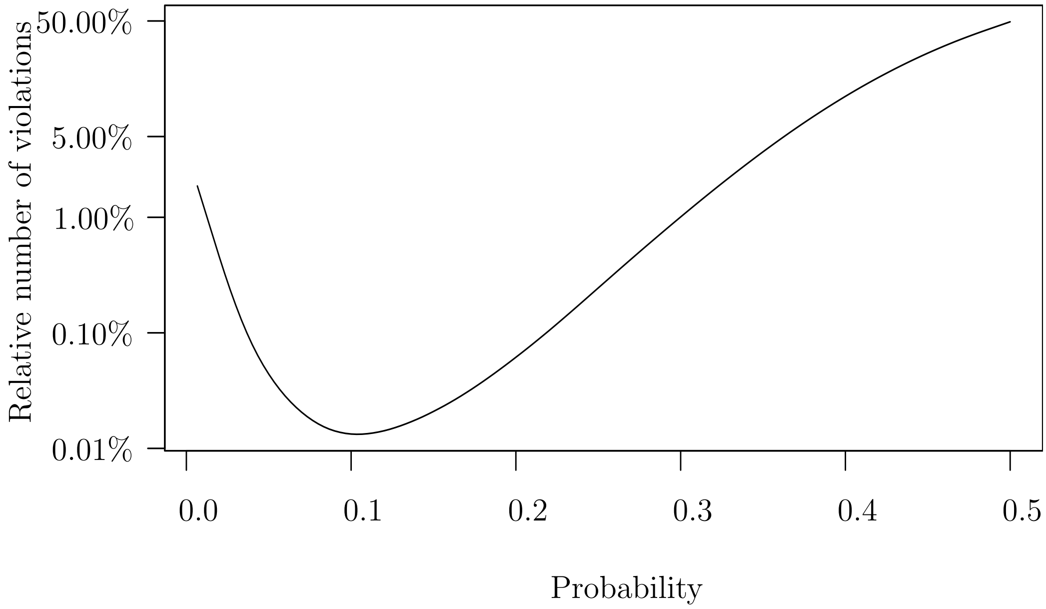


Figure 3: Empirical tail of Student–t(3). N=1000, and EVT fit.

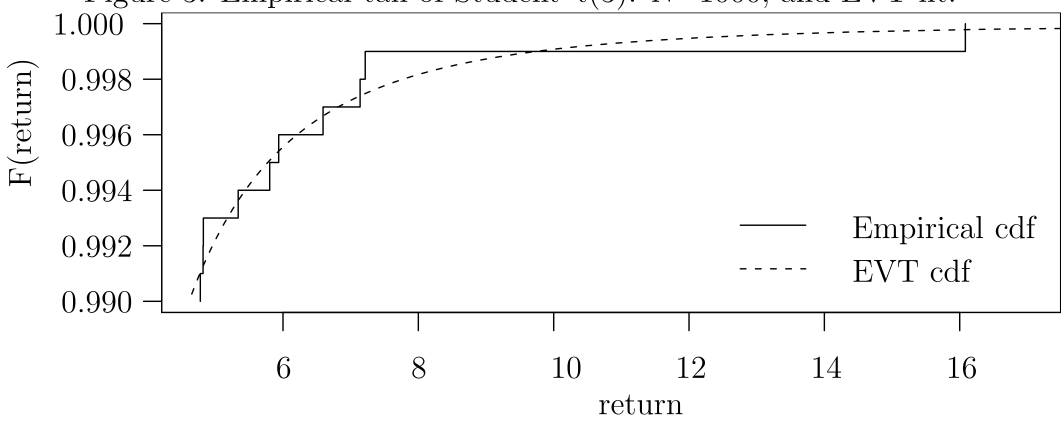
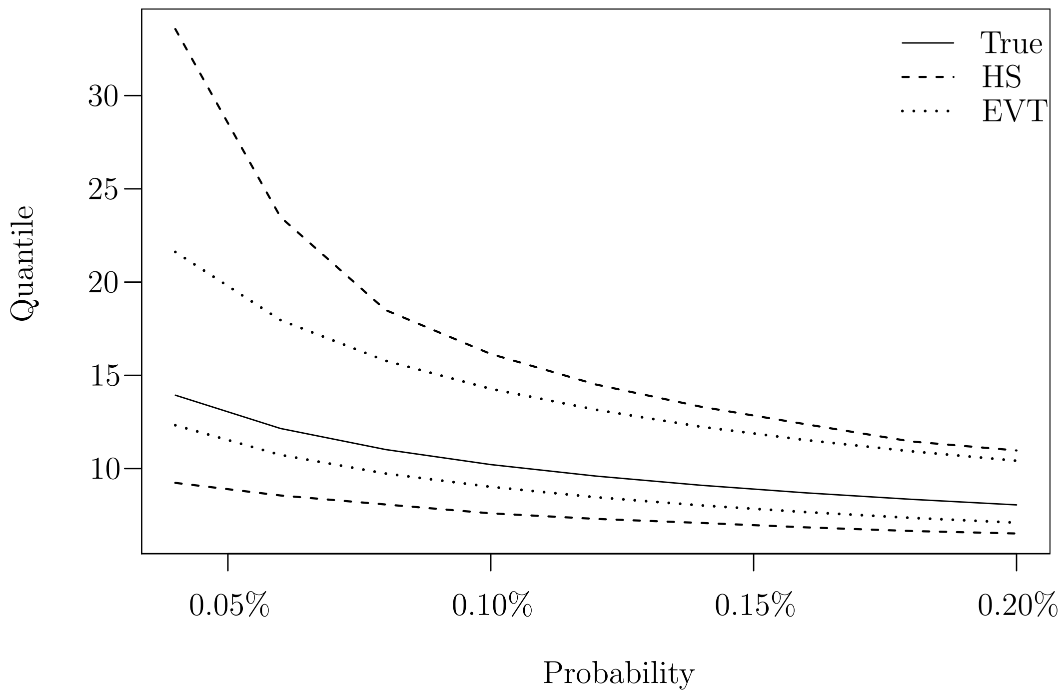


Figure 4: 1% and 99% empirical confidence bounds for VaR

VaR for a Student–t(3), estimated with HS and EVT for sample size $N = 5000$, and probabilities, $m = 2/N, \ldots, 10/N$. The EVT threshold, m, is 200. The solid line is the true quantile, and the dotted/dashed lines are 1% and 99% empirical quantile from repeating the estimation with 5,000 random samples.



References

Acerbi, C., Nordio, C., and Sirtori, C. (2001). Expected Shortfall as a Tool for Financial Risk Management. Abaxbank Working Paper arXiv:cond-mat/0102304.

Acerbi, C. and Tasche, D. (2001). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, 31:379–388.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent Measures of Riak. *Mathematical Finance*, 9:203–228.

Basel Committee on Banking Supervision (1996). Ammendment to the capital accord to incorporate market risks. Committee Report 24, Basel Committee on Banking Supervision, Basel, Switzerland.

Basrak, B., Davis, R. A., and Mikosch, T. (2002). A characterization of multivariate regular variation. *Annals of Applied Probability*, 12:908–920.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press, Cambridge.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Daníelsson, J. and de Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4:241–257.

Daníelsson, J. and de Vries, C. G. (2000). Value at risk and extreme returns. *Annales D'Economie et de Statistique*. `www.RiskResearch.org`. Reprinted in Extremes and Integrated Risk Management, Risk Books. 2000.

Daníelsson, J., Jorgensen, B. N., Sarma, M., Samorodnitsky, G., and de Vries, C. G. (2005). Subadditivity re–examined: The case for value–at–risk. *www.RiskResearch.org*.

Daníelsson, J., Jorgensen, B. N., Yang, X., and de Vries, C. G. (2007). Optimal portfolio allocation under a probabilistic risk constraint and the incentives for financial innovation. *Annals of Finance*, 4(3):345–367.

Davis, R. A. and Mikosch, T. (1998). The sample autocorrelations of heavy–tailed processes with applications to ARCH. *Annals of Statistics*, 26:2049–2080.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory, an introduction.* Springer.

de Haan, L., Resnick, S. I., Rootzen, H., and de Vries, C. G. (1989). Extremal Behaviour of Solutions to a Stochastic Difference Equation with Applications to Arch-Processes. *Stochastic Processes and their Applications*, 32:213–224.

de Vries, C. (2005). The simple economics of bank fragility. *Journal of Banking and Finance*, 29:803–825.

Embrechts, P., Kuppelberg, C., and Mikosch, T. (1996). *Modelling Extremal Events for Insurance and Finance (Applications of Mathematics).* Berlin: Springer Verlag.

Embrechts, P., Lambrigger, D. D., and Wuthrich, M. V. (2008). Multivariate extremes and the aggregation of dependent risks: Examples and counter–examples. *Extremes*, 12:107–127.

Fama, E. (1965). The behavior of stock-market prices. *Journal of Business*, 38(1):34–105.

Fama, E. T. and Miller, M. H. (1972). *The theory of finance.* Dryden Press.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley & Sons.

Garcia, R., Renault, E., and Tsafack, G. (2007). Proper conditioning for coherent var in portfolio management. *Management Science*, 53(3):483–494.

Geluk, J. and de Vries, C. (2006). Weighted sums of subexponential random variables and asymptotic dependence between returns on reinsurance equities. *Insurance Mathematics and Economics*, 38:39–56.

Ibragimov, R. (2005). *New Majorization Theory In Economics And Martingale Convergence Results In Econometrics.* PhD thesis, Yale University.

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

Ibragimov, R. and Walden, J. (2007). The limits of diversification when losses may be large. *Journal of Banking and Finance*, 31:2551–2569.

Jansen, D. W. and de Vries, C. G. (1991). On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *The Review of Economics and Statistics*, 73(1):18–24.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, 36:392 – 417.

Marshall, A. W., Olkin, I., and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer, 2nd edition.

Pflug, A. G. G. (2005). Value at risk in portfolio optimization: properties and computational approach. *Journal of Risk*, 7:1–31.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Process.* Springer-Verlag.

Shin, H. S. (2009). Securitization and financial stability. *Economic Journal*, 119:309 – 332.

Stoyanov, S., Rachev, S., and Fabozzi, F. (2007). Optimal financial portfolios. *Applied Mathematical Finance*, 14:401–436.

Yamai, Y. and Yoshiba, T. (2002). Comparative analyses of expected shortfall and VaR: their estimation error, decomposition, and optimization. *Monetary and Economic Studies*, 20:87–121. IMES Discussion Paper Series 2001-E-12, http://www.imes.boj.or.jp/english/publication/edps/2001/01-E-12.pdf.