Report 99-054

Second Order Tail Effects

Casper G. de Vries

ISSN 1389-2355

Second Order Tail Effects

Casper G. de Vries*

Tinbergen Institute Rotterdam, Erasmus Universiteit Rotterdam and NIAS

Semi-parametric extremal analysis can be a useful tool to calculate the Value-at-Risk (VaR) for loss probabilities which are at and below the inverse of the sample size. We first review the standard estimation procedures and VaR implications on the basis of the first order expansion to the tail probabilities of heavy tail distributed random variables. Subsequently we present some new results that are based on using a second order expansion of the tail risk. In particular, we discuss the issue of efficiency in estimation using high or low frequency data; and we investigate the relation between the VaR over a short and a long investment horizon.

*Summary of presentation for the 'Workshop on Statistics in Finance', Hong Kong, July 1999. Some of this material was first presented at the conference on 'Extremes, Risk and Safety' in Gothenburg, August 1998. The paper is partially based on joint work with M. Dacorogna, J. Danielsson, J. Geluk, L. de Haan, U. Muller, L. Peng and O. Pictet. I am grateful to R.Brinkman for helpful discussion and to a referee for careful reading of the manuscript.

1 Introduction

Financial asset data sets nowadays cover millions of high frequency price quotes. These data sets are well suited for studying the market risk on very large losses. Regulators of the financial industry currently require that commercial banks be able to report, on a daily basis, a loss estimate over a ten-day trading horizon for their entire trading portfolio given a certain preassigned low risk level. The loss estimate is called the Value-at-Risk (VaR). For internal risk management purposes the larger investment banks also back out a VaR estimate for a one-day trading horizon. Non-financial corporations nowadays do include long horizon VaR forecasts in their yearly statements. Out of convenience the continuously compounded asset returns are often presumed to be normally distributed, see J.P.Morgan (1995), Jorion (1997), and Dowd (1998). As it happens, however, asset returns are heavy tailed distributed. If we work from this assumption, the VaR can be well estimated by employing extreme value techniques, see e.g. Dacorogna et al. (1995), Longin (1997), Danielsson and De Vries (1997, 1998) and Dowd (1998). The approach is a go between the traditional finance based normal approach and the historical simulation based non-parametric approach.

In the paper we first briefly review the motivation behind the by now standard estimation procedures by means of a first order expansion to the tail probabilities of heavy tail distributed random variables. We discuss how the first order approach implies a particular relationship between the VaR over short and longer investment horizons. Subsequently we present some new results that are based on using a second order expansion of the tail risk. In particular we discuss the issue of efficiency in estimation using high and low frequency data; and we investigate the relation between the VaR over a short and a long investment horizon.

2 The First Order Approach to Heavy Tails and VaR

Suppose that the returns are i.i.d. and have tails which vary regularly at infinity. In that case

$$F(-x) = ax^{-\alpha}[1 + o(1)] \quad \text{as } x \to \infty, \quad \text{and } a > 0. \tag{1}$$

These distributions are said to exhibit heavy tails since the m-th moment $E[X^m]$ is unbounded when $\alpha < m$, whereas in case of e.g. the normal d.f. for any finite m the $E[X^m]$ is bounded. Given parameter estimates for the scale coefficient a and tail index α, the VaR x can be calculated upon inverting $ax^{-\alpha}$ for a given small risk level p: $x_p \approx (a/p)^{1/\alpha}$. We first discuss how the parameters can be estimated, and subsequently discuss the VaR application in more detail.

2.1 estimation

The standard estimation procedures can be motivated as follows. Suppose the Pareto law $G(-x) = ax^{-\alpha}$ holds exact below a certain threshold $-s$, where $s > 0$. The conditional distribution reads $G_{X|X\leq -s}(-x) = (x/s)^{-\alpha}$. One can go from this to the associated conditional density with tail index $\alpha + 1$: $g_{X|X\leq -s}(-x) = \alpha(x/s)^{-\alpha-1}(1/s)$. Take logarithms to get

$$\log g_{X|X\leq -s}(-x) = \log \alpha - (\alpha + 1)\log \frac{x}{s} - \log s.$$

Substitute in this expression the random variable $-X_i$ for the x, whenever $X_i < -s$. Differentiate with respect to α, sum the result over the observations X_i which fall below $-s$, and equate to 0 in order to obtain the Maximum Likelihood estimator of the tail index:

$$\widehat{\frac{1}{\alpha}} = \frac{1}{M}\sum_{i=1}^{M} \log \frac{-X_i}{s}, \quad X_i < -s, \tag{2}$$

and where M is the random number of extreme observations X_i that fall below the threshold $-s$. For a large enough s, the conditional Pareto density $g_{X|X\leq -s}(-x)$ may also be a good approximation to the true conditional density $f_{X|X\leq -s}(-x)$, when the conditional distribution is not exactly Pareto but rather satisfies (1). The estimator (2) applied to the extreme observations from a heavy tailed distribution that adheres to (1) is known as the Hill (1975) estimator. We note that the estimator (2) is conditional on the appropriate choice of the threshold s; but how this choice has to be made cannot be discussed without going into the second order expansion. The assumption of independence is also crucial; although the estimator can be shown to be consistent for important classes of stochastic processes.

Likewise we can motivate the estimator for the extreme quantiles or VaR. Let x_p and x_t be two extreme quantiles with associated probabilities p and t respectively, that adhere to the law $G_{X|X\leq -s}(-x) = (x/s)^{-\alpha}$. Then $t/p = (x_t/x_p)^{-\alpha}$, and hence $x_p = x_t\,(t/p)^{1/\alpha}$. Suppose $p < 1/n < t$, where n is the sample size; moreover let t be such that M, $M < n$, is the closest integer equal to nt. Then we can estimate the VaR x_p by

$$\widehat{x_p} = x_t \left(\frac{M/n}{p} \right)^{\widehat{1/\alpha}}. \tag{3}$$

Since the statistical properties of $\widehat{x_p}$ are dominated by the properties of the exponent $\widehat{1/\alpha}$, we can limit the discussion towards discussing the properties of the tail index estimator.

2.2 value at risk at different horizons

Suppose a bank has estimated its one-day VaR from past daily return observations. It also has to calculate the VaR for a ten-day investment horizon to fulfill its regulatory requirements. The industry often works from the assumption of normality and calculates the ten-day VaR by sizing up the one-day estimate with a factor $\sqrt{10}$, since this is the well known convolution rule for summing i.i.d. normal random variables. The square-root procedure reduces the burden of estimation on risk managers. If the observations are heavy tailed distributed, this simple convolution rule no longer applies. Nevertheless for the tail risk, aggregation is still simple under the i.i.d. assumption.

Let the returns X_i have a distribution as in (1). For the sum $\Sigma_i^k X_i$ (holding k fixed), we have by Feller's theorem (1971, VIII.8)

$$P\left\{\Sigma_i^k X_i < -x\right\} = kax^{-\alpha}[1 + o(1)], \quad as\ x \to \infty, \tag{4}$$

and where the scale factor 'a' is as in (1). We pointed out that banks for internal purposes often calculate the VaR over a one day investment horizon, but that regulators require a longer horizon. Corporations for their yearly reports need an even longer horizon, see the recently launched CorporateMetrics (1999) product by the RiskMetrics group. The question therefore is how to go from the high frequency estimate to the low frequency estimate without having to reestimate the parameters on a reduced sample size, and thus possibly losing efficiency. In Dacorogna et al. (1995, 1998) the following rule was presented:

Proposition 1 (The α-root rule) *Suppose X has finite variance, so that $\alpha > 2$. At a constant risk level p, increasing the time horizon k increases the*

VaR for the normal model percentagewise by more, i.e. by $\sqrt{k}$, than for the fat tailed model, where the increase is a factor $k^{1/\alpha}$.
Proof. *Rescale x on the left hand side in (4) by $k^{1/\alpha}$, this gives $ax^{-\alpha}$ on the right hand side and hence equals the first order term in (1).* ■

The opposite holds if the distribution is so heavy tailed that the second moment is unbounded, i.e. if $\alpha < 2$ then $k^{1/\alpha} > \sqrt{k}$. In the related economics literature on diversification it has been noted that the effect of diversification is less pronounced in comparison with the normal distribution, if the returns are sum-stable distributed with $\alpha < 2$, see Fama and Miller (1972, p. 270). They note that for $\alpha < 1$ diversification actually increases the dispersion. We are not aware of a discussion in the finance literature of the case $\alpha > 2$ but finite, for either the issue of diversification nor for the issue of tail risk (VaR) aggregation over time

3 The Second Order Approach to Heavy Tails

Throughout this section we assume that the following second order expansion applies:

$$F(-x) = ax^{-\alpha}[1 + bx^{-\beta} + o(1)], \quad \text{as } x \to \infty, \quad \text{and } a > 0. \tag{5}$$

Freely floating foreign exchange rate returns are often more or less symmetrically distributed about a zero mean. Therefore, in what follows we will often assume that the lower and upper side tails are similar up to and including the second order term

$$\begin{aligned} P\{X &\leq -x\} = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})), \\ P\{X &> x\} = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})). \end{aligned} \tag{6}$$

The differences may come from the o-terms. Note that the second order term is assumed to be of the same type as the first order term. Some motivation for this choice can be found in the following observations. If the second order term were of the form $\log x$, some of the results below would not apply due to the slower rate of convergence; for other functional forms like $\exp(-x)$ convergence is so rapid that the second order term plays no role of importance. The expansion (5) applies for symmetric heavy tailed distributions like the Student-t, which is often used to model the unconditional distribution of asset returns, and it applies to the stationary distribution of the ARCH(1) process, which is used for modelling the conditional asset returns.

3.1 statistical properties

On basis of the expansion (5) one can derive the first two moments of the Hill estimator (2) by elementary calculus. The conditional $k-$th order log empirical moment from a sample $X_1, ..., X_n$ of n i.i.d. draws from $F(x)$ is defined as follows:

$$u_k(s_n) \equiv \frac{1}{M} \sum_{i=1}^{M} \chi_{(X_i < -s_n)} (\log \frac{-X_i}{s_n})^k, \tag{7}$$

where s_n is a threshold that depends on n, M is the random number of left tail excesses, and where $\chi_{(.)}$ is the indicator function. Note that $u_k(s_n)$ is a function of the highest realizations only. We will sometimes suppress the reference to n in s_n when this does not create confusion. The theoretical properties of the Hill estimator $u_1(s_n)$ are well documented by e.g. Hall (1982) and Goldie and Smith (1987).

The properties of the Hill estimator derive from the following Lemma

Lemma 2 *Given the model (5), for $k \geq 1$, and as n, $s_n \to \infty$, while $s_n/n \to 0$,*

$$\mathrm{E}[u_k(s)] = \Gamma(k+1)\left(\frac{1}{\alpha^k} + \frac{bs^{-\beta}}{(\alpha+\beta)^k}\right) + o\left(s^{-\beta}\right). \tag{8}$$

Proof. From calculus after two transformations of variables we have the following result:

$$\begin{aligned}
\alpha \int_s^\infty (\log \frac{x}{s})^k x^{-\alpha-1} dx &= \alpha s^{-\alpha} \int_1^\infty (\log y)^k y^{-\alpha-1} dy \\
&= \alpha s^{-\alpha} \int_0^\infty t^k \left(e^t\right)^{-\alpha-1} e^t dt \\
&= \alpha^{-k} s^{-\alpha} \int_0^\infty x^k e^{-x} dx \\
&= \frac{\Gamma(k+1)}{\alpha^k} s^{-\alpha}.
\end{aligned}$$

Hence, the conditional expectation in (8) follows from the assumption (5) and the calculus result

$$\begin{aligned}
\mathrm{E}[u_k(s)] &= \frac{1}{1-F(s)} \int_s^\infty \left(\log \frac{x}{s}\right)^k f(x) dx \\
&= \frac{\Gamma(k+1)}{1+bs^{-\beta}} \left[\frac{1}{\alpha^k} + \frac{bs^{-\beta}}{(\alpha+\beta)^k}\right] + o\left(s^{-\beta}\right).
\end{aligned}$$

■

It immediately follows that for $k=1$:

Corollary 3 *The asymptotic bias of the Hill estimator $u_1(s_n)$ from (2) is*

$$\mathrm{E}\left[u_1(s_n)-\frac{1}{\alpha}\right]=-\frac{b\beta}{\alpha(\alpha+\beta)}s_n^{-\beta}+o\left(s_n^{-\beta}\right). \tag{9}$$

After some manipulation and application of the Lemma (2) for $k=1,2$, one obtains the asymptotic variance of the Hill estimator.

Corollary 4 *For the threshold $s_n\to\infty$, but $s_n^{\alpha}/n\to 0$,*

$$\mathrm{Var}\left[u_1(s_n)-\frac{1}{\alpha}\right]=\frac{s_n^{\alpha}}{an}\frac{1}{\alpha^2}+o\left(\frac{s_n^{\alpha}}{n}\right). \tag{10}$$

These two results can be readily combined to obtain the asymptotic mean squared error (AMSE) of $u_1(s_n)$

$$\mathrm{AMSE}\left(u_1(s_n)\right)\approx\frac{1}{a\alpha^2}\frac{s_n^{\alpha}}{n}+\frac{b^2\beta^2}{\alpha^2(\alpha+\beta)^2}s_n^{-2\beta}. \tag{11}$$

From this expression it is easy to see that for $n\to\infty$, the rate by which $s_n\to\infty$ determines which of the two terms in (11) asymptotically dominates the other, or that they just balance.

Rewrite (11) in shorthand notation as AMSE $=An^{-1}s^{\alpha}+Ds^{-2\beta}$. From the first order condition $\alpha An^{-1}s^{\alpha-1}-2\beta Ds^{-2\beta-1}=0$, the unique AMSE minimizing threshold level $\overline{s}$ is found as

$$\overline{s}_n=\left(\frac{2\beta D}{\alpha A}\right)^{\frac{1}{2\beta+\alpha}}n^{\frac{1}{2\beta+\alpha}}.$$

To summarize, we have the following result:

Proposition 5 *As $n\to\infty$ the* AMSE *minimizing asymptotic threshold level $\overline{s}_n$ is*

$$\overline{s}_n(u_1)=\left[\frac{2ab^2\beta^3}{\alpha(\alpha+\beta)^2}\right]^{\frac{1}{(\alpha+2\beta)}}n^{\frac{1}{(\alpha+2\beta)}}. \tag{12}$$

And the associated asymptotically minimal MSE *of $u_1(s_n)$ is*

$$\begin{aligned}\overline{AMSE}\left[u_1(\overline{s}_n)\right] &= \frac{1}{a\alpha}\left[\frac{1}{\alpha}+\frac{1}{2\beta}\right]\left[\frac{2ab^2\beta^3}{\alpha(\alpha+\beta)^2}\right]^{\frac{\alpha}{2\beta+\alpha}}n^{-\frac{2\beta}{2\beta+\alpha}} \\ &\quad +o\left(n^{-\frac{2\beta}{2\beta+\alpha}}\right).\end{aligned} \tag{13}$$

From (11-13) it is straightforward to show that if s_n tends to infinity at a rate below $n^{1/(2\beta+\alpha)}$, the bias part in the MSE dominates, while conversely the variance part dominates if s_n tends to infinity more rapidly than $n^{1/(2\beta+\alpha)}$. It is also easy to see that the number of excedances M is such that

$$n^{\frac{-2\beta}{\alpha+2\beta}} M\left(u_1\left(\overline{s}_n\right)\right) \rightarrow a\left[\frac{2ab^2\beta^3}{\alpha\left(\alpha+\beta\right)^2}\right]^{-\frac{\alpha}{\alpha+2\beta}} \text{ in } p \text{ as } n \rightarrow \infty. \tag{14}$$

Further asymptotic properties of the Hill estimator, like asymptotic normality given that $\overline{s}_n$ is used in (2), are shown in e.g. Goldie and Smith (1987). Danielsson et al. (1997) discuss how a bootstrap of the AMSE can be used to back out the optimal threshold $\overline{s}_n$ in practice, such that the Hill estimator retains its asymptotic normality property. In this bootstrap procedure the empirical minimum of the bootstrapped MSE is used to estimate $\overline{s}_n$ consistently, and the procedure guarantees that the rate conditions assumed in the above results are automatically satisfied. By doing so one balances the two vices of bias squared and variance such that these disappear at the same rate. For dependent data it is sometimes known how the variance is affected, see e.g. the recent work by Drees (1999) and Starica (1999) for the ARCH(1) process, but other aspects, like the choice of the threshold s_n, are still open issues.

3.2 time aggregation and efficiency

The log-returns are time additive, i.e. the two week return is the sum of the one week returns. Nowadays financial data sets can be obtained at even the finest time grid around, which is the trading time scale. The question is which data should be used for estimation purposes. In particular we ask ourselves the following question, if one needs results for a long investment horizon, should one nevertheless use the high frequency data for estimation, and then use a rule like the α-root rule to extrapolate to the low frequency level? We give an answer in terms of the asymptotic mean squared error efficiency.

Assume that $\alpha > 2$, because this is the relevant case for most financial data. In that case both the mean and the variance are bounded. We first obtain a general lemma on second order convolution behavior. This result is needed because, as was shown above, the $AMSE$ of the tail index estimator is a function of the first and second order parameters. The existing literature only gives a result on second order convolution behavior for positive random variables, see Geluk, De Haan, Resnick and Starica (1997). But since the log-asset returns can be positive and negative, we need to analyze this case afresh. To restrict the number of different combinations that will arise, we assume that the tails are similar. We find that because the distribution of asset returns is two-sided, a new factor depending on $E[X^2]$ enters.

Lemma 6 (Second order convolution) *Suppose that the tails are second order similar, i.e. as* $x \to \infty$

$$\begin{aligned} P\{X &\leq -x\} = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})), \\ P\{X &> x\} = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})), \end{aligned} \tag{15}$$

and $a > 0, b \neq 0$. *Moreover, assume that* $\alpha > 2$ *and* $\beta > 0$ *so that* $E[X]$ *and* $E[X^2]$ *are bounded. Suppose* X_1 *and* X_2 *are i.i.d. and satisfy (15). Then for the 2-convolution*

$$\begin{aligned} P\{X_1 + X_2 &> s\} = P\{X_1 + X_2 \leq -s\} \\ &= 2as^{-\alpha}(1 + bs^{-\beta} + \alpha E[X]s^{-1} + \frac{\alpha(\alpha+1)}{2}E[X^2]s^{-2}) \\ &\quad + o(s^{-\alpha-2}) + o(s^{-\alpha-\beta}) \end{aligned} \tag{16}$$

as $s \to \infty$.

The Lemma (6) was obtained in Dacorogna et al. (1998) by elaborate calculus arguments. We develop some intuition for the result by a novel argument. The probability $P\{X_1 + X_2 > s\}$ can be split into just two parts:

$$\begin{aligned} P\{X_1 + X_2 &> s\} \approx P\{X_1 + X_2 > s, X_2 \leq \frac{s}{2}\} + \\ P\{X_1 + X_2 &> s, X_1 \leq \frac{s}{2}\} \end{aligned} \tag{17}$$

The remaining other part $P\{X_1 > \frac{s}{2}, X_2 > \frac{s}{2}\} = P\{X_1 > \frac{s}{2}\}^2 = O(s^{-2\alpha})$ is of smaller order and can be ignored since it is assumed that $\alpha > 2$.

To determine $P\{X_1 + X_2 > s, X_2 \leq \frac{s}{2}\}$, we first compute the conditional probability $P\{X_1 + X_2 > s \mid X_2 = c\} = P\{X_1 + c > s\}$, say. This conditional probability is obtained from the marginal by translation. Consider the law $P\{X > x\} = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta}))$ as $x \to \infty$, and suppose we shift X by adding the constant c. This changes the probability into $P\{X + c >$

$x\} = a(x-c)^{-\alpha}(1+b(x-c)^{-\beta}+o(x^{-\beta}))$. Use the Taylor expansion to write, assuming that $x > c$,[1]

$$\begin{aligned}(x-c)^{-\gamma} &= x^{-\gamma}(1-\frac{c}{x})^{-\gamma} \\ &= x^{-\gamma}\{1+\gamma\frac{c}{x}+\frac{\gamma(\gamma+1)}{2}(\frac{c}{x})^2+O((\frac{c}{x})^3)\}.\end{aligned}$$

Use this twice to rewrite $P\{X+c > x\}$ as:

$$\begin{aligned}P\{X+c &> x\} = ax^{-\alpha}[1+\alpha cx^{-1}+\frac{\alpha(\alpha+1)}{2}c^2x^{-2}+bx^{-\beta} \\ &+o(x^{-\beta})+o(x^{-2})].\end{aligned} \tag{18}$$

The following conditional probability can be split into three parts

$$\begin{aligned}P\{X_1+X_2 &> s, -\frac{s}{2} \le X_2 \le \frac{s}{2}\} = \int_{-\infty}^{\infty} P\{X+c > s\}dF(c) \\ -\int_{-\infty}^{-s/2} P\{X+c &> s\}dF(c) - \int_{s/2}^{\infty} P\{X+c > s\}dF(c).\end{aligned}$$

In all three integrals substitute the right hand side form of (18) for $P\{X+c > s\}$. The second and third integral are of small order $O(s^{-2\alpha})$. For example, since for $s \to \infty$

$$\begin{aligned}\int_{s/2}^{\infty} P\{X+c &> s\}dF(c) = \\ &\int_{s/2}^{\infty} as^{-\alpha}(1+o(1))\{\alpha x^{-\alpha-1}(1+o(1))\}dx = \\ &O(s^{-2\alpha}).\end{aligned}$$

The first probability can be found by using the translation result

$$\int_{-\infty}^{\infty} P\{X_1+c > s\}dF(c) = E_c[P\{X_1+c > s\}]$$

$$= E_c[as^{-\alpha}\{1+\alpha cs^{-1}+\frac{\alpha(\alpha+1)}{2}c^2s^{-2}+bs^{-\beta}+o(s^{-\beta})+o(s^{-2})\}]$$

$$= as^{-\alpha}\{1+bs^{-\beta}+\alpha E[X_2]s^{-1}+\frac{\alpha(\alpha+1)}{2}E[X_2^2]s^{-2}+o(s^{-\beta})+o(s^{-2})\}.$$

[1]See also Dacorogna et al.(1995) where this expansion is used to show that the Hill estimator is not location invariant.

The last expression gives $P\{X_1 + X_2 > s, -\frac{s}{2} \leq X_2 \leq \frac{s}{2}\}$, but we need $P\{X_1 + X_2 > s, X_2 \leq \frac{s}{2}\}$, see (17). However, as before, the probability $P\{X_1+X_2 > s, X_2 \leq -\frac{s}{2}\}$ is of small order and can be ignored. By symmetry the same result is obtained for $P\{X_1 + X_2 > s, X_1 \leq \frac{s}{2}\}$. Putting these two probabilities together yields the claim.

From this second order convolution result we can infer how the $AMSE$ will be affected by the choice of the return frequency in the estimation, see Dacorogna et al. (1995,1998):

Proposition 7 *Suppose the X_i are i.i.d. with a distribution $F(x)$ that is symmetric around zero, $E[X] = 0$, and varies regularly at infinity as in (5) with $\alpha > 2$. Then a w-convolution affects the leading term in the $\overline{AMSE}\left[u_1\left(\overline{s}_n\right)\right]$ from (13) as follows:*

(i) $\beta < 2$. There is no effect;

(ii) $\beta = 2$. The $AMSE$ changes by a factor

$$\left[\left(1 + \frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]/b\right)^2\right]^{\alpha/(2\beta+\alpha)};$$

(iii) $\beta > 2$. The $AMSE$ changes by a factor

$$c\left[\frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]\right]^{\frac{2\alpha}{4+\alpha}}\left(\frac{1}{b^2}\right)^{\alpha/(2\beta+\alpha)},$$

and where

$$c = \frac{4+\alpha}{2\beta+\alpha}\left(\frac{2}{\alpha+2}\right)^{\frac{2\alpha}{4+\alpha}}\left(\frac{\alpha+\beta}{\beta}\right)^{\frac{2\alpha}{2\beta+\alpha}}\left(\frac{\alpha}{4an}\right)^{\frac{4}{4+\alpha}}\left(\frac{2\beta an}{\alpha}\right)^{\frac{2\beta}{2\beta+\alpha}}.$$

The upshot of Proposition 7 is that either time aggregation has no effect, i.e. when $\beta < 2$, or that the $AMSE$ deteriorates, possibly only after the first few convolutions when $b < 0$ and $\beta = 2$. If $\beta > 2$ the $AMSE$ always deteriorates after the first convolution. While it can thus not be ruled out that higher frequencies deteriorate the $AMSE$ properties of $\widehat{\alpha}$ for the first few convolutions, the majority of the cases goes into the other direction. For this reason it may be advisable to use the highest frequency data available for estimation, and subsequently to extrapolate to obtain the lower frequency result by means of a rule like the α-root rule from Proposition 1.

3.3 second order VaR

Suppose one follows the advice from the previous subsection and estimates the low frequency VaR from the high frequency VaR. By doing this one exploits the efficiency that the high frequency data deliver. On the negative side however, one may loose from the fact that the α-root rule from Proposition 1 is based on a first order approximation We investigate the possible loss in precision that may arise from neglecting the second order terms. Assume the mean is $E[X] = 0$. Consider the convolution result (16), but inflate the VaR s by a factor $2^{1/\alpha}$. This gives

$$\begin{aligned} P\{X_1 + X_2 \ \leq \ & -2^{1/\alpha}s\} = \\ & as^{-\alpha}\{1 + b2^{-\beta/\alpha}s^{-\beta} + \frac{\alpha(\alpha+1)}{2}E[X^2]2^{-2/\alpha}s^{-2}\} \\ & +o(s^{-\alpha-2}) + o(s^{-\alpha-\beta}). \end{aligned}$$

Let $P\{X \leq -s\} = as^{-\alpha}(1 + bs^{-\beta} + o(s^{-\beta})) = p$, say, and use this to rewrite the above

$$\begin{aligned} P\{X_1 + X_2 \ \leq \ & -2^{1/\alpha}s\} = \\ & p + as^{-\alpha}\{-b(1 - 2^{-\beta/\alpha})s^{-\beta} + \frac{\alpha(\alpha+1)}{2}E[X^2]2^{-2/\alpha}s^{-2}\} + \\ & o(s^{-\alpha-2}) + o(s^{-\alpha-\beta}). \end{aligned}$$

If $b > 0$ and $\beta < 2$, then for sufficiently large s the α-root rule is overly conservative, since the second order term $-b(1 - 2^{-\beta/\alpha})s^{-\beta}$ is negative. If, however, $b < 0$, or if $\beta > 2$, then the second order term is positive, and the α-root rule is not prudent enough. To circumvent the bias in the low frequency VaR estimates that stems from the α-root rule, one could redo the quantile estimation on the low frequency data by means of (3), while retaining the tail index estimate from the high frequency data. Which procedure is better is an issue for further research.

4 Conclusion

The paper first reviews the standard estimation procedures and VaR implications on the basis of a first order expansion for the tail probabilities of heavy tail distributed random variables. Subsequently, it was argued why second order results are needed for determining the properties of the estimators.

We developed a new intuitive derivation of the second order convolution result. This second order convolution result is useful for the discussion of the efficiency in estimation. While for most cases using the high frequency data is mean-square efficient, we showed that there are some exceptions. The second order convolution result also enables one to determine the precision of the rule by which the VaR over a short investment horizon is related to the VaR over a long investment horizon.

References

Dacorogna, M.M., U.A. Muller, O.V. Pictet and C.G. de Vries, 1995, *Extremal returns in extremely large data sets,* Tinbergen Institute discussion paper, TI95-70.

Dacorogna, M.M., U.A. Muller, O.V. Pictet and C.G. de Vries, 1998, *Extremal forex returns in extremely large data sets,* mimeo, submitted.

Dowd, K., 1998, Beyond value at risk, the new science of risk management, Wiley, Chichester.

Danielsson, Jansen, D., J. and De Vries C.G., 1996. *The methods of moments ratio estimator for the tail shape parameter,* Communications in Statistics, Theory and Methods, 25, 711-720.

Danielsson J., L. de Haan, L. Peng and C.G. de Vries, 1997, *Using a bootstrap method to choose the sample fraction in tail index estimation,* Tinbergen Institute discussion paper, TI97-016/4, forthcoming in Journal of Multivariate Analysis.

Danielsson J., and C.G. de Vries, 1997, *Tail index and quantile estimation with very high frequency data,* Journal of Empirical Finance, 4, 241-257.

Danielsson J., and C.G. de Vries, 1998, *Value-at-Risk and extreme returns,* Tinbergen Institute discussion paper TI98-017/2

Dekkers A.L.M., Einmahl, J.H.J. and de Haan, L., 1989. *On the estimation of the extreme-value index and large quantile estimation,* Annals of Statistics, 17, 1795-1832.

Drees, H., 1999, *Weighted approximations of tail processes under mixing conditions,* University of Cologne, mimeo.

Fama E.F. and M.H. Miller, 1972, *The theory of finance,* Dryden Press, Hinsdale.

Feller W., 1971, *An introduction to probability theory and its applications, volume II,* John Wiley, New York, 2nd edition.

Geluk J., L. de Haan, S. Resnick and C. Starica, 1997, *Second order regular variation, convolution, and the central limit theorem,* Stochastic Processes and their Applications, 69, 139-159.

Goldie, C.M. and Smith R.L., 1987, *Slow variation with remainder: Theory and applications*, Quarterly Journal of Mathematics, Oxford 2nd series, 38, 45-71.

Hall, P., 1982, *On some simple estimates of an exponent of regular variation*, Journal of the Royal Statistical Society, Series B, 44, 37-42.

Hill, B.M., 1975, *A simple general approach to inference about the tail of a distribution*, Annals of Statistics, 3, 1163-1173.

Jorion, P.,1997, *Value-at-Risk*, Irvin: McGraw Hill.

Morgan Guarantee Trust Company,1995, *RiskMetrics Technical Document*, New York: J.P.Morgan Bank.

Longin, F.M., 1997, *From value-at-risk to stress testing:the extreme value approach,* CERSSEC working paper 97-004.

Peng, L. 1997. *Second order condition and Extreme value estimation,* Ph.D. dissertation #178, Tinbergen Institute, Erasmus University Rotterdam.

RiskMetrics Group, 1999, *CorporateMetrics Technical Document,* www.riskmetrics.com.

Starica, C., 1999, *On the tail empirical process of solutions of stochastic difference equations,* Chalmers University, mimeo.