

Tail index and quantile estimation with very high frequency data

Jón Daníelsson^a, Casper G. de Vries^{b,*},¹

^a *Tinbergen Institute and University of Iceland, Reykjavik, Iceland*

^b *Tinbergen Institute, Erasmus University, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands*

Abstract

A precise estimation of the tail shape of forex returns is of critical importance for proper risk assessment. We improve upon the efficiency of conventional estimators that rely on a first order expansion of the tail shape, by using the second order expansion. Here we advocate a moments estimator for the second term. The paper uses both Monte Carlo simulations and the high frequency foreign exchange recordings collected by the Olsen corporation to illustrate the technique. © 1997 Elsevier Science B.V.

1. Outliers, risk and high frequency benefits

To introduce the subject we consider two problems in the analysis of risk. The board of a pension fund has to decide over the maximum percentage of the managed funds that can be allocated to the most risky investment categories (equities, foreign currency and derivatives). The quotas are a function of the solvency and liquidity requirements of the pension fund; the fund must be able to pay out every period. Specifically, the board is concerned about a single extremely negative return that wipes out so much capital that the fund becomes illiquid or insolvent. A very similar problem concerns the allocation of upper limits on open

* Corresponding author. E-mail: cdevries@few.eur.nl or jond@hi.is

¹ We are grateful for the comments by our discussant, Paul Embrechts and the questions from the audience at the HFDF conference, as well as the excellent and stimulating comments of two anonymous referees. Daníelsson benefitted from an HCM fellowship of the EU; De Vries likes to thank the CES of the University of Munich for its hospitality.

positions to foreign currency dealers by the treasurers of the forex dealing room of an international bank. The trading limits are a function of the probability p on a single negative return so large that the solvency of the bank is endangered. These problems can be formulated in a static or dynamic fashion. In the static version the unconditional distribution is used and the wealth levels are updated. Accurate assessment of the value of p requires as much data as possible, because the extremes are rare by their very nature. The Olsen data set is thus conducive to estimating the values of p accurately.

Let X_i denote the single period return on a risky investment² and define the minimum return, X_{\min} , in a single sample of n periods as

$$X_{\min} = \min_i \{X_1, \dots, X_n\}. \quad (1)$$

Suppose the management specifies a critical loss level, $s < 0$, which represents the maximum loss that can be incurred without running into solvency problems. A simple way to determine the maximum allowable investment I , is to take $I = s/(X_{\min})$ assuming that $X_{\min} < 0$. In fact this worst case analysis is implemented by some well known banks. The procedure is, however, limited and leaves the associated uncertainty rather implicit. It is limited because the focus is solely on the worst in sample realization, thereby neglecting the probability that $X_{n+1} < X_{\min}$. It is like building a dike to the level of the highest realized flood level, whereas one might want to safeguard against worse outcomes. And the procedure leaves implicit the associated probability on outcomes below X_{\min} . Whereas the management might want to choose from probability-loss combinations (p, x_p) by considering the trade-off that exists between p and x_p . Hence a table of (p, x_p) combinations could enhance the decision process. Moreover, using the lowest order statistics X_{\min} is typically not a very reliable estimate if the true quantile, x_p , associated with $p = 1/n$, since this estimator has a very high variance.

To remedy the two deficiencies of the simple procedure, the management is asked to explicitly impute a crash probability p on outcomes worse than s . The level p signifies the insolvency risk the management deems 'acceptable'. Presumably, this probability is set very low; most likely it is taken to be $p \ll 1/n$, i.e. far outside the sample. In this case one has to find ways to extrapolate the empirical distribution of the minimum order statistic, in combination with some extra conditions, to construct a list of probability-quantile combinations (p, x_p) such that

$$\Pr\{X_{\min} \leq x_p\} = p \quad \text{when } p < 1/n. \quad (2)$$

Once x_p is determined, the trading limit I directly follows from $I = s/(x_p)$.

² Throughout the paper we develop the theory by assuming that X_i are i.i.d. Nevertheless, most results survive under the weaker assumptions of stationarity. For the particular relevant case of ARCH process, see De Haan et al. (1989).

Before we continue with the econometric and data aspects of the problem, it is useful to point out the relation with the literature. In De Haan et al. (1994) the concern about extreme downside risk as formulated above is related to the Roy (1952) safety first criterion. A full blown portfolio selection problem for this case is worked out in Jansen et al. (1995) by using diversification arguments of Arzac and Bawa (1977).

One might also wonder why the management is just concerned about a single bad draw, while a series of consecutive smaller negative returns may add up to the same highly negative result. The reason why this is of lesser concern, is that during a more gradual decline of the market the exposure can be reduced in the meantime. If the exposure is reduced, the loss level is reduced as well, given a certain return level. This is very similar to the philosophy behind the marking-to-market daily settlement procedure on futures markets. In this way the clearing house only runs the risk on a party not being able to meet its margin call for a single day, rather than letting the losses accumulate until the contract expires. Thus it is straightforward to implement this static risk of loss approach in a dynamic fashion.

How do we benefit from the magnificent data set made available by Olsen and Associates? An immediate benefit of the highest frequency data is the gain in efficiency in constructing the negative outliers. Because

$$\min\{X_1, X_2, \dots, X_{n-1}, X_n\} \leq \min\left\{\frac{X_1 + X_2}{2}, \dots, \frac{X_{n-1} + X_n}{2}\right\} \quad (3)$$

the higher frequency data are more informative about the tail shape of the return distribution. For example, if X_i are stable distributed, it follows from their invariance under convolution that the higher frequency recordings increase the efficiency of the tail shape estimators. A more detailed analysis of this claim is available in Dacorogna et al. (1995).

A less obvious but rather important gain from this data set is that its sheer size permits estimation methods which are just not feasible on the typical size of the more common financial data sets. In order to construct the probabilities or quantiles from Eq. (2) one relies on the limit law for the minimum. The reason is obvious: one does not exactly know which distribution generates the data. The limit law, however, is only an approximation. In general the estimates will therefore be biased. In fact, it can be shown that it is optimal, according to the MSE criterion, to have the bias and variance vanish at the same rate as $n \rightarrow \infty$. This rate is a non-linear function of n and m , where m is the number of lowest order statistics that are used to estimate Eq. (2). To determine the optimal number m in the MSE sense, one needs an estimate of the bias. Estimating the bias is a nontrivial exercise, however, because the linearity of the estimator defies direct use of a conventional procedure like the bootstrap (see e.g. Shao and Ti, 1995, p. 14).

To salvage this problem Hall (1990) suggested to use subsample bootstraps. This idea exploits the fact that the bias is nonlinear in the sample size by comparing bootstrap estimates from subsamples to a full sample estimate and then extracting the bias if the subsamples have size $n_1 = n^\gamma$, $\gamma < 1$. In the comparison the true values cancel, and the bias term from the full sample estimate can be ignored because it has an order of magnitude smaller than the bias term from the subsample estimate. From this, the optimal m can be estimated. Details of the subsample bootstrap procedure are given in the next section, or see Hall (1990).

To be able to implement such a procedure one needs to construct bootstrap subsamples which are magnitudes smaller than the full sample. On the other hand, because the outliers are rare by their very nature, one needs subsamples which are still quite sizable. The pay-off of the Olsen and Associates data set is, that it allows one to sensibly exploit the subsample bootstrap method. Conventional financial data sets are just too small for this procedure, even though they may be large compared to the usual economic data sets.

As of today, the Hall (1990) subsample bootstrap is the only known procedure to determine the optimal number of order statistics m that have to be taken into account in Eq. (2). It was first applied in economics by Dacorogna et al. (1995). The method as advocated by Hall, nevertheless, still has one deficiency. It presumes a very particular form of the bias term. The theoretical innovation of this paper consists in relaxing this assumption by using a more general form. This does not come without a cost, because it necessitates estimating an extra parameter β which determines the asymptotic behavior of the second order term of the law in Eq. (2) as $x \rightarrow -\infty$. We employ a novel estimator of β that is consistent. The estimator is based on recent results from Daniélsso et al. (1996). Again, the large size of the data set is crucial for obtaining decent estimates of β because the estimation relies on being able to separate the first and second order terms.

2. Extremal theory

The presentation of the theory will be concise. An excellent reference on extremal theory is the book by Leadbetter et al. (1983). Estimation issues are dealt with in Dacorogna et al. (1995) and Daniélsso et al. (1996). An introductory account is given in De Vries (1994). Without loss of generality, this section considers the problem of the distribution of the maximum order statistic. We start from the well known fact that forex return distributions are fat tailed. Focusing on the upper tail by taking $-X_{\min}$, the class of fat tailed distributions is characterized by the regular variation at infinity property

$$\lim_{t \rightarrow \infty} \left[\frac{1 - F(tx)}{1 - F(t)} \right] = x^{-\alpha}, \quad (4)$$

where $\alpha > 0$ and $F(x)$ is any fat tailed distribution. Under some mild extra conditions this class admits the following second order expansion for large x :

$$F(x) \approx 1 - ax^{-\alpha} [1 + bx^{-\beta}], \tag{5}$$

where $a > 0$, $b \in \mathcal{R}$, $\beta > 0$. The case $\beta = 0$ constitutes the expansion $F(x) \approx 1 - ax^{-\alpha} [1 + b \log x]$. As is shown in De Haan and Stadtmuller (1996), these are the only two second order expansions with non-trivial properties; e.g. a second order term like $\exp(-x)$ converges uninterestingly rapidly. Note that Hall (1990) imposes the more stringent condition of $\alpha = \beta$. This covers certain distributions, like part of the stable laws and the type II extreme value distribution. But it does not apply to e.g. the Student- t class which is often used in financial time series analysis. For the Student- t class the expansion in Eq. (5) is valid, though, with α equal to the degrees of freedom ν and $\beta = 2$.

Let $p > 1/n$, but close to $1/n$, and $t \ll 1/n$, where n is the sample size and p, t denote exceedance probabilities. The problem is to estimate the quantile x_t by using the empirical counterpart of the in sample quantile x_p . By definition and using the expansion Eq. (5):

$$p \approx ax_p^{-\alpha} [1 + bx_p^{-\beta}] \text{ and } t \approx ax_t^{-\alpha} [1 + bx_t^{-\beta}] \tag{6}$$

Division of p by t and rearranging gives

$$x_t \approx x_p \left(\frac{p}{t} \right)^{1/\alpha} \left(\frac{1 + bx_t^{-\beta}}{1 + bx_p^{-\beta}} \right)^{1/\alpha} \tag{7}$$

Ignore the second order term in Eq. (7). Estimate x_p from Eq. (2) by the descending order statistic $X_{(m)}$ i.e. $X_{(1)} \geq \dots \geq X_{(m)} \geq \dots \geq X_{(n)}$ for which m/n is closest to p . Replace p by m/n . (The choice of m is discussed below.) The extreme quantile estimator now reads:

$$\hat{x}_t = X_{(m)} \left(\frac{m}{nt} \right)^{1/\alpha} \tag{8}$$

This estimator is still conditional on the tail shape parameter α .

The α can be estimated by the Hill procedure as follows. For the largest in sample quantiles $x_q \geq x_p$, say, the second order term in Eq. (5) contributes little. Hence the Pareto distribution

$$G(x_q) = 1 - ax_q^{-\alpha}, \quad x_q \geq x_p \tag{9}$$

can be used as a first order approximation to the tail of the true distribution; recall Eq. (4). The maximum likelihood estimator of α conditional on $x_q \geq x_p$ is easily found from the log-density

$$\frac{1}{\hat{\alpha}} = \frac{1}{m-1} \sum_{i=1}^{m-1} \log \frac{X_{(i)}}{X_{(m)}}, \tag{10}$$

where again $X_{(m)}$ is used to approximate x_p and the $X_{(i)}$ are the empirical counterparts of x_q . The estimator $1/\hat{\alpha}$ is named after Hill.

A generic and long standing problem in the area is the determination of the number m of extreme values that should be used in the estimation procedures. Recently Hall (1990) has proposed the solution by means of a subsample bootstrap. We outline the idea. It is known that the estimators \hat{x}_t and $1/\hat{\alpha}$ are asymptotically normal (see e.g. Goldie and Smith, 1987). From this the asymptotic mean squared error (AMSE) can be calculated. One can show that it is optimal, in the AMSE sense, to have the bias and variance vanish at the same rate. This optimal rate determines the optimal number of order statistics as

$$m^* = cn^{2\beta/(2\beta + \alpha)}, \quad c > 0 \tag{11}$$

where c depends on the parameters a, b, α, β . For example in case of the student- t class when $\alpha = \nu$

$$c = \left[v(v+2)^2 \left(\frac{\Gamma((v+1)/2)}{\Gamma(v/2)} \frac{v^{(v-1)/2}}{\sqrt{v\pi}} \right)^{4/\nu} \right. \\ \left. / \left(16 \left(-\frac{v^2}{2} \frac{v+1}{v+2} \right)^2 \right)^{\nu/(4+\nu)} \right].$$

To determine m^* empirically, one might try to minimize bootstrap estimates of the MSE. This method fails, however, because the estimator $1/\hat{\alpha}$ is log-linear in the observations. The bootstrap therefore calculates the bias as essentially zero. This is unfortunate because the bias and variance have to disappear at the same rate for the AMSE to be minimal. But the bias and MSE are nonlinear in the sample size. Hall (1990) therefore suggested to employ subsamples of order $n_1 = n^\gamma, \gamma < 1$ to calculate bootstrap estimates of $1/\alpha$ on these subsamples and to calculate the mean squared deviation of the subsample estimates from an initial full sample estimate $1/\alpha_0$. The initial estimate of α_0 is obtained by specifying an arbitrary low threshold level, say 1% of n and using Eq. (10) to estimate α_0 . This can be further improved by iterating on the bootstrap several times. This yields a proper estimate of the MSE as it comprises both a variance and a bias part. The bias part is retained in the subsamples bootstrap estimate because the biases of the initial full sample estimate and the subsample estimate do not cancel. The former is of small order in comparison to the latter due to the fact that $n_1 < n$. Hence by varying m_1 the subsample optimal m_i^* can be estimated by minimizing the subsample MSE. Using Eq. (11), the full sample m^* is then found from

$$m^* = m_1^* \left(\frac{n}{n_1} \right)^{2\beta/(2\beta + \alpha)} \tag{12}$$

But this requires knowledge of the exponent $2\beta/(2\beta + \alpha)$. Hall (1990) then

argues that $\beta = \alpha$ is a plausible restriction because several known distributions satisfy this condition. But, examples like the Student-*t*, where α is the degrees of freedom and $\beta = 2$ show that this restriction needs not hold in models that are relevant for modelling of asset returns. Thus the restriction $\beta = \alpha$ makes the bootstrap procedure non robust. This leaves the whole subsample bootstrap endeavor in an awkward position, because it appears that the problem of not knowing m^* has been replaced by the problem of not knowing β . We have to face the problem of estimating β ³.

In Danielsson et al. (1996) it is argued that the estimation of the tail parameter β of the second order term in the expansion Eq. (5) is not straightforward. Nevertheless, they develop a nonlinear estimator which is consistent. The procedure is as follows. The first step is to obtain an estimate of m_1^* through the subsample bootstrap procedure of Hall (1990). It is easily seen from the arguments in Goldie and Smith (1987) that this yields a consistent estimate of m_1^* (as $n \rightarrow \infty$) regardless of whether $\beta = \alpha$ or not. In the second and third steps consistent estimates of α and β are obtained from the subsamples. The estimate for $1/\alpha$ is given in Eq. (10), where in theory $m = m_1^* - 1$ so that $m \rightarrow \infty$ and $(m/n)^{2\beta/(2\beta + \alpha)} \rightarrow 0$ as $n \rightarrow \infty$. By using less order statistics than m_1^* , $1/\hat{\alpha}$ and is consistent, see Hall (1982, th. 2). To introduce the estimator for β , step 3, we need the empirical log-moments.

The *j*th empirical log-moment is defined as follows:

$$\tilde{m}^{(j)} = \frac{1}{m-1} \sum_{i=1}^{m-1} \left[\log \frac{X_{(i)}}{X_{(m)}} \right]^j \tag{13}$$

Note that Eq. (10) is the special case of Eq. (13) with $j = 1$. Define Δ to be the statistic:

$$\Delta = \frac{\tilde{m}^{(1)} - \tilde{m}^{(2)}/2\tilde{m}^{(1)}}{\tilde{m}^{(3)}/3\tilde{m}^{(2)} - \tilde{m}^{(4)}/4\tilde{m}^{(3)}} \tag{14}$$

Danielsson et al. (1996) propose the following estimator for β :

$$\hat{\beta} = \hat{\alpha}(\sqrt{\Delta} - 1). \tag{15}$$

Here $\hat{\alpha}$ is the inverse of $1/\hat{\alpha}$ from the second step. In computing the $\tilde{m}^{(j)}$, $j = 1, 2, 3, 4$ where in theory $m = m_1^* + 1$ is taken. Note that there is a difference between the $\tilde{m}^{(1)}$ used here and $1/\hat{\alpha}$ and from the previous step⁴. It signifies that

³ A referee gave us the interesting suggestion that the exponent $2/(2 + \alpha/\beta)$ might be estimated from calculating an m_2^* from a second bootstrap on an even further reduced subsample size $n_2 < n_1 < n$, and using m_1^*/m_2^* to estimate the exponent. We leave for future research to investigate how the further sample size reduction affects the efficiency, and how this procedure compares with the method used in this paper.

⁴ In the results presented in this paper we set $m = m^*$.

$m/n \rightarrow 0$, while $(m/n)^{2\beta/(2\beta+\alpha)} \rightarrow \infty$. With these rates, elementary integration shows, after linearizing the ratio, $\tilde{m}^{(j)}/j\tilde{m}^{(j-1)}$, that as $n \rightarrow \infty$

$$p \lim \left[\frac{\tilde{m}^{(j)}}{j\tilde{m}^{(j-1)}} - \frac{1}{\alpha} \right] \left(\frac{n}{m} \right)^{\beta/\alpha} = -b\beta\alpha^{j-2}(\alpha + \beta)^{-j}.$$

It is then easy to show that as $m/n \rightarrow 0$ while $(m/n)^{2\beta/(2\beta+\alpha)} \rightarrow \infty$

$$p \lim \Delta = \left(\frac{\alpha + \beta}{\alpha} \right)^2 \quad (16)$$

and hence the estimator $\hat{\beta}$ in Eq. (15). Lastly, these subsample estimates for α and β are used in Eq. (12) to convert m_1^* to m^* and to estimate α by Eq. (10) taking $m = m^*$. By being able to estimate β and to break away from the restriction $\beta = \alpha$ in Eqs. (5) and (12), the estimation procedure gains in robustness. Section 3 evaluates this estimator by means of a MonteCarlo experiment and it is applied to the Olsen and Associates data set.

3. The value of large data sets

In this section we first provide a simulation study that evaluates the Hill estimator $1/\hat{\alpha}$ and from Eq. (10), the second order index estimator $\hat{\beta}$ in Eq. (15), and the quantile estimator in Eq. (8) with the iterated subsample bootstrap procedure. Finally, we apply these methods to the Olsen data set.

3.1. Simulation study

The distributions chosen for the Monte Carlo study were Student- t , extreme value type II, i.e. $F(x) = \exp(-x^{-\alpha})$ and ARCH(1). The Student- t was simulated with respectively degrees of freedom 1, 5 and 11 and the extreme value distribution was simulated with, respectively, degrees of freedom 1 and 11. For both of these distributions the degrees of freedom equals the tail index. We also simulated from an ARCH(1) process, by using normal innovations, an intercept of 0.1 in the variance equation and autoregressive coefficients with values 0.7 and 0.3, respectively. These values correspond to a theoretical tail index of an ARCH process with values 3.17 and 8.36, respectively. The second order limit expansion Eq. (5) of the stationary ARCH distribution is not known, but they are known for the Student- t and extreme value distributions. The theoretical β value for the Student- t is $\beta = 2$ and for the extreme value it is $\beta = \alpha$. On the basis of this second order expansion, the optimal m^* can be calculated by Eq. (11). The parameters of the Monte Carlo experiment are given in Table 1. The number of observations was set at 20,000 and each experiment was repeated 250 times. The results from the Monte Carlo experiments are reported in Tables 2–4 and box

Table 1
Parameters of Monte Carlo experiment

No. of observations (<i>n</i>)	20,000
No. of simulations (<i>S</i>)	250
Subsample (<i>n</i> ₁)	2000
No. subsamples (<i>NS</i>)	100
Initial <i>m</i> value (<i>m</i> ₀)	200
Bootstrap iterations (<i>BI</i>)	4

Table 2
Iterative bootstrap, Student-*t*₍₅₎

Iteration	α_0	m_i	m^*	α_1	β_1	# $\Delta > 0$	Exponent	Δ
1	4.05 (0.28)	39.07 (13.56)	81.34 (35.60)	3.90 (0.32)	0.99 (0.73)	80.74 (6.70)	0.31 (0.11)	2.09 (17.46)
2	4.41 (0.54)	31.00 (17.29)	66.55 (42.71)	4.15 (0.51)	1.13 (0.69)	79.14 (6.46)	0.33 (0.12)	-3.66 (64.11)
3	4.58 (0.76)	29.43 (16.30)	62.63 (39.48)	4.23 (0.61)	1.18 (0.72)	78.37 (6.73)	0.33 (0.12)	1.27 (1.49)
4	4.66 (0.94)	29.73 (17.15)	64.67 (46.00)	4.21 (0.59)	1.20 (0.71)	79.08 (6.53)	0.34 (0.11)	1.54 (3.72)
True value	5	26	72	5	2	100	44	1.96

Table 3
Results, Student-*t*₍₅₎

	α	m^*	X_{\max}	\hat{q}_n	\hat{q}_{3n}	Δ	$\Delta > 0$	# $\Delta > 0$ (%)	β
Mean	4.66	64.67	12.97	11.4	14.68	0.42	2.23	72	0.85
S.E.	(0.94)	(46.00)	(3.65)	(1.53)	(2.51)	(14.78)	(9.53)		4.55
RMSE	[1.00]	[46.49]	[4.05]	[1.55]	[3.42]	[14.84]	[9.50]		4.68
TRUE	5	72	11.18	11.18	14.01	1.96	1.96		2

Table 4
Tail index and quantile forecasting for simulated data

Distribution function	Tail index			True q_{3n}	Quantile estimation		
	$\hat{\alpha}$	S.E.	RMSE		\hat{q}_{3n}	S.E.	RMSE
<i>t</i> (1)	1.01	(0.05)	[0.05]	19099	19146	(5738)	[5726]
<i>t</i> (5)	4.66	(0.94)	[1.00]	17.01	14.68	(2.51)	[3.42]
<i>t</i> (11)	7.24	(1.52)	[4.06]	6.71	7.20	(0.81)	[0.95]
Extreme(1)	1.00	(0.06)	[0.06]	59999	62831	(19882)	[20043]
Extreme(11)	11.01	(0.65)	[0.65]	2.72	2.72	(0.09)	[0.09]
ARCH (3.17)	3.13	(0.55)	[0.55]		3.97	(0.99)	
ARCH (8.36)	6.74	(1.27)	[2.06]		0.80	(0.10)	

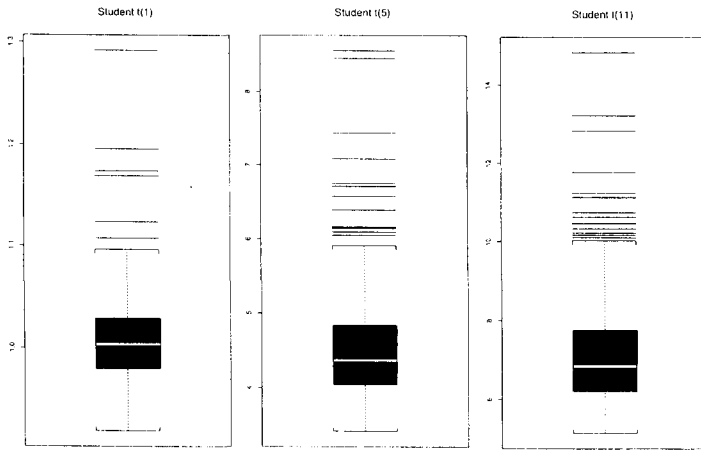


Fig. 1. Box plots for alpha simulation estimates (Student- t).

plots of the α estimates are reported in Figs. 1–3. The box plot shows the median of the data set, surrounded by a box that spans the center half of the data set, i.e. from the lower to the upper quartile. The whiskers extend 1.5 times the inter-quartile range, with the values outside the whiskers denoted as outliers.

In some samples the Δ statistic from Eq. (14) turned out to be negative. This is inherent to the usage of a moments based procedure. When $\Delta < 0$ Eq. (15) makes no sense. Given the stochastic nature of β , negative β 's when $0 \leq \Delta < 1$ are plausible and we use all cases where the β 's are negative as long as Δ was positive.

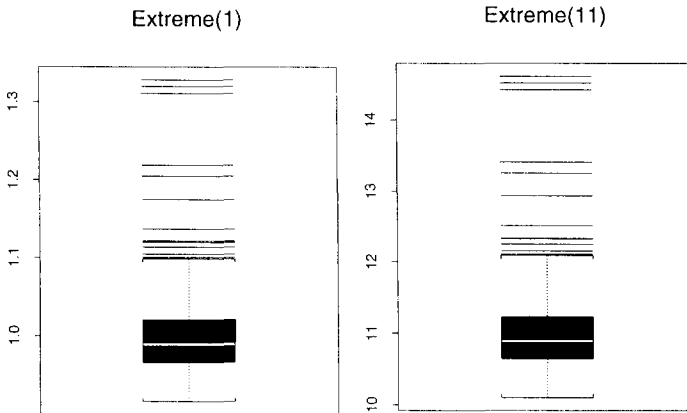


Fig. 2. Box plots for alpha simulation estimates (extreme).

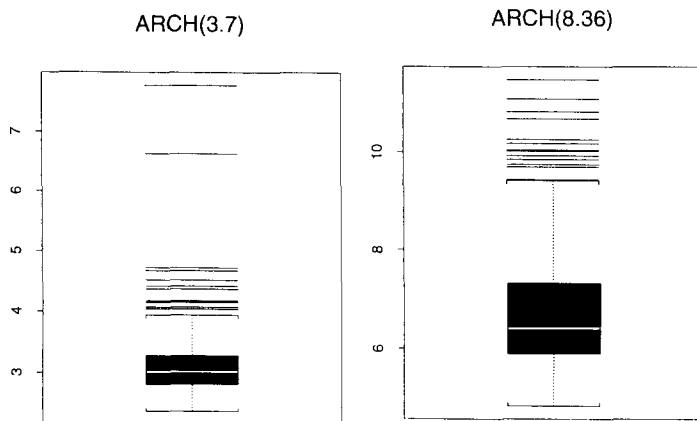


Fig. 3. Box plots for alpha simulation estimates (ARCH).

In Table 2 we report details of the bootstrap procedure for the Student- $t_{(5)}$ simulation. Standard errors are in parenthesis. The first iteration is obtained by specifying the initial threshold level arbitrarily at $m_0 = 1\%$. The estimate of α_0 clearly improves with the iterations, as can be seen in column 1. At the first iteration the value is 4.05 but it increases to 4.66 in the fourth iteration, however the standard error increases as well. These results are fairly typical, in that the estimates of the tail index improve with iterations, but the standard error increases as well. The reason must be that at m^* the bias and standard error are inversely related. In our experience, no further improvement can be expected beyond the fourth iterations in most cases. We report the bootstrap tail index, α_1 , in column four and the second order term, β_1 , in the fifth column. We see that both of those estimates improve with iterations and are getting closer to their theoretical values. Only about 80 Δ 's out of 100 are positive.

The final result of the simulation is reported in Table 3. The tail index estimate is 4.66 which has to be compared to the true value of 5. Our procedure predicts a value for the maximum order statistic in a sample of 20,000 as 11.4, which is close to the theoretical value. We regard this result as being close for the following reason. From the simulations the average value of the maximum order statistic, X_{\max} , was 12.97 with standard error 3.65. The maximum order statistic is a simple estimate of $x_{1/n}$. But due to its large variance, it is clear that the extreme value method provides a considerable improvement. The out-of-sample prediction for the maximum order statistic on an imaginary sample of size 60,000 is 14.68, compared with a theoretical value of 14.01.

Table 4 reports summary results from simulations of several distributions. We generated 250 samples of size 20,000 for each distribution, and applied our procedure to the estimation of the tail index, $\hat{\alpha}$ and the out-of-sample quantiles,

\hat{q}_{3n} , for an imaginary sample of size 60,000. The α estimates for the Student- t improve with lower α , the thicker the tail the better the estimates. In the case of the extreme value distribution, our method does quite well in estimating the tail index. For the ARCH distribution, we also find that the thicker the tail the better our method performs. For the ARCH $_{(\alpha=3.17)}$ we get an estimate of 3.13, but for ARCH $_{(\alpha=8.36)}$ the estimate is only 6.74. The quantile forecasts for an imaginary sample of 60,000, \hat{q}_{3n} , or three times the original sample size, are in all cases close to the theoretical values, \hat{q}_{3n} . However, the standard error and root mean squared error are high, especially for the thick tailed data sets, the Student- $t_{(1)}$ and extreme $_{(1)}$. The theoretical quantiles are not known for the ARCH distribution.

3.2. *Olsen data set*

The Olsen data set comprises one year of data on three forex contracts: yen–DM, yen/USD and DM/USD. The Olsen company continuously tapes bid and ask quotes; here we use logarithmic middle price quotes. These data are first automatically filtered by the Olsen company in order to remove data processing and transmission errors; a description of the type of data set is given in Muller et al. (1990). The filtered data were turned into standard return data as follows. Single second return data were created by taking the log difference between two consecutive quotes and dividing by the time difference in seconds. The few quotes that occurred in the same second, were artificially assumed to have happened in two consecutive seconds. Because several authors argue that the data on a higher frequency than 10 min behave differently from data with a frequency of 10 min or less, we also constructed 10 min return data. This was done by dividing the time axis into equal slices of 10 min intervals. The data were then searched to see whether a new record was entered in an interval one minute surrounding the endpoints of the 10 min intervals. If some were found, the quote closest to the exact endpoint was used to construct the 10 min return. If no new record was entered within the one minute interval, linear interpolation was used to construct an artificial 10 min return. Recalling the inequality in Eq. (3), linear interpolation seems an innocuous procedure for our purposes. The construction of the 10 min return data corresponds to the procedure followed by the Olsen company. Descriptive statistics are given in Table 5.

The large size of the Olsen data set makes it especially conducive to the bootstrap technique. The subsample size is 10% of the data series, or 5256 for the 10 min data. The α_0 was initially estimated by using 1% of the highest order statistics, and through iteration this threshold level was improved. Because we are interested in large changes in the exchange rate in either direction, we analyzed the properties of both lower and upper tails of the returns. The estimation results can be found in Table 6. The tail index estimates for either exchange rate hover around 4 at the 10 min aggregation level but are closer to 3 at the 1 s frequency. These values are in line with previous work that used much lower frequency data and

Table 5
Sample statistics for the Olson dataset

Data set	Aggregation	<i>T</i>	Mean	Variance	Min	Max
Yen–DM	10 minute	52558	−5.13e−6	2.66e−7	−0.00678	0.006630
Yen–DM	1 second	158482	−9.88e−9	7.55e−11	−0.00062	0.000470
Yen–USD	10 minute	52558	−2.36e−6	3.09e−7	−0.00842	0.010718
Yen–USD	1 second	567758	−3.03e−8	1.39e−9	−0.00185	0.001432
DM–USD	10 minute	52558	2.78e−6	3.13e−7	−0.00693	0.008626
DM–USD	1 second	1466945	−1.36e−8	1.42e−9	−0.00099	0.001373

concluded that the fourth unconditional moment is probably just unbounded, i.e. $\alpha < 4$. Most of the empirical parametric work on the conditional distribution which proceeds by estimating an ARCH process also comes to the same conclusion. Comparing the left tail index with the right tail index, the tail shapes appear to be very similar, suggesting tail symmetry. We see that for the most frequently traded contract, DM–USD, the tail index appears almost constant under aggregation, while this not the case for the yen–USD or the yen–DM contracts. This may be due to the infrequent trading phenomenon. Linear aggregation may have

Table 6
Results from the Olsen dataset

Data set	10 minute tails		1 second tails	
	upper	lower	upper	lower
Yen–DM				
$\hat{\alpha}_0$	3.66	3.56	2.59	2.41
$\hat{\alpha}_1$	3.72	3.60	2.53	2.39
$\hat{\beta}_1$	0.55	0.36	0.56	0.27
\hat{m}^*	91	78	679	686
$\hat{\alpha}$	4.94	4.25	2.33	2.24
Yen–USD				
$\hat{\alpha}_0$	3.03	3.08	2.85	2.81
$\hat{\alpha}_1$	2.89	3.14	2.83	2.79
$\hat{\beta}_1$	0.20	0.11	2.83	2.79
\hat{m}^*	112	64	2906	2780
$\hat{\alpha}$	3.51	4.40	2.89	2.90
DM–USD				
$\hat{\alpha}_0$	3.22	3.24	3.22	3.22
$\hat{\alpha}_1$	3.28	3.14	3.23	3.24
$\hat{\beta}_1$	3.28	0.23	0.37	0.23
\hat{m}^*	251	114	3055	2075
$\hat{\alpha}$	3.37	3.93	3.44	3.54

removed too many extremes, thereby artificially thinning the tail of the 10 min data.

We notice that in every case there is a change from the initial tail index, α_0 , to the final estimate, α . In most cases this difference is not large, except for the 10 min yen–DM data, where the difference is 1.28. This indicates that the threshold level may not matter much in many cases, but in others it can lead us astray. The threshold level is in every case considerably lower than the initial level of 1%. The iterative subsample bootstrap estimates of the second order tail shape parameter β vary considerably and the distribution of β appears to be bi-modal. It indicates the difficulty of estimating second order term parameters of an Taylor expansion of a distribution at infinity.

The quantile forecasts are presented in Table 7. We only present results for the 10 min aggregated data, the results for the 1 second data are similar. For each final estimate of α we estimate quantiles, x_t , associated with different probabilities, t . The probabilities are reported as a function of the sample size, i.e. $t = (kn)^{-1}$. We have observations on the empirical distribution function for $k \leq 1$. However our procedure is developed for out of sample forecasts and since the second order term

Table 7
Quantile forecasting. 10 minute aggregation

k	Lower tail		Upper tail	
	observed	forecast	observed	forecast
DM–USD				
0.95	0.0009	0.0012	0.0009	0.0010
1.0	0.0069	0.0086	0.0086	0.0108
1.5		0.0095		0.0122
2.0		0.0103		0.0133
2.5		0.0109		0.0142
3.0		0.0114		0.0150
Yen–USD				
0.95	0.0005	0.0012	0.0008	0.0011
1.0	0.0084	0.0087	0.0107	0.0105
1.5		0.0096		0.0118
2.0		0.0103		0.0128
2.5		0.0108		0.0136
3.0		0.0112		0.0144
Yen–DM				
0.95	0.0008	0.0012	0.0008	0.0013
1.0	0.0068	0.0074	0.0066	0.0062
1.5		0.0082		0.0067
2.0		0.0087		0.0071
2.5		0.0092		0.0075
3.0		0.0096		0.0077

increases the bias as we move into the sample the quantile forecasts perform worse for $k = 0.95$ than for $k = 1.0$, or at the largest order statistic.

Concentrating on the DM–USD contract, we see that the probability of a 1.1% increase in the exchange rate during a single episode of 10 min is about one per year and we would expect 1.5% upward increase in the exchange during a single 10 min interval once every three years. From the vantage point of risk management, it is of interest to compare these numbers with those of the Yen contracts. It appears that the yen contracts and especially the Yen–DM contract, is less risky than the DM–dollar contract displaying less extravagant down and up movements.

4. Conclusion

In this paper we study the improvement of conventional tail index and quantile estimator by incorporating an estimate of the second order term of the tail expansion. The moment estimator advocated here appears to be effective when it is most needed, i.e. when the tail index estimate is heavily biased. There is logic behind this result. If the tail index estimate is very precise, then the first order expansion of the tail shape suffices. In other words, the second order term contributes little and is therefore difficult to estimate.

Turning to the Olsen data set, the second order estimates indicated that the number of order statistics to be used in estimating the tail index should be lower than the initial 1%. Nevertheless, the tail index values were hardly affected and are in line with previous work. On the one hand this may seem surprising given that the 10 min and especially the 1 s data contain so much extra noise over and above the noise that is absent in the more conventional and lower frequency data sets. This noise is e.g. caused by the idiosyncrasies of the individual trader and makes these highest frequency data unfit for addressing several economic issues. But for the analysis of extremes this extra noise does not appear to be a devastating vice. On the other hand, we have the result that the tail index is invariant under convolution, Feller (1971, VIII.8). This result is complementary to the central limit theorem. An appropriately rescaled sum of n i.i.d. Student- t variates with 3 d.f. is over a wide range approximately normal distributed. But this does not hold for the tail shape of this sum; the tail shape remains as it was. Formally, the sum of two regularly varying random numbers with tail index α , is again regularly varying with index α .

The quantile estimator seemed to perform well, in the sense that the $\hat{x}_{1/n}$, estimates are close to the empirical quantile at $t = 1/n$. In contrast, the empirical quantile is not a particularly reliable estimator of $x_{1/n}$ and not to much should be made out of this proximity. The usefulness of the quantile estimator lies in the fact that it can generate out of sample (P, Q) combinations. For risk management this means that the trade-off which exists between crash probabilities and a threshold return is quantified.

Appendix A. Computational procedures

A.1. Monte Carlo experiments

The Monte Carlo experiments were implemented in C++, using the NAG library for random number generation. Several thick tailed distributions were chosen with a range of α and β values. The same general parameters were used in the simulation and estimation, see Table 1. Of these parameters the most important are the number of observations, n and the number of simulations, S . For each distribution we generated S independent samples with n observations each and estimate several parameters. The results from applying the iterative bootstrap procedure to the Student- $t_{(5)}$ are presented in Table 2 and the final results for the Student- $t_{(5)}$ are presented in Table 3. Due to space limitations we leave out detailed results for the other distributions and only report the alpha and quantile forecast estimates in Table 4 for a sample of size $3n$. The true value is not known for the ARCH process. Other values of interest are the initial α estimates, the bootstrap optimal number of order statistics, the β estimate and the number of order statistics used in the estimation of α . A box plot is used to show the range of the α estimates for each distribution. See Figs. 1–3.

A.2. Estimation with the Olson data set

We used the Olsen forex data set to estimate the tail index and forecast quantiles. The quotes were transformed into returns by using the difference in log averages of bids and asks. We transformed the data into ten minute and one second intervals. Summary statistics for the transformed data are reported in Table 5.

We applied our procedure to each data set, and present the results in Table 6. Results were obtained for both the lower and upper tails of the sample data. We only report the initial tail index estimate, α_0 , the bootstrap estimate of β_1 , the optimal number of order statistics, m^* , and the final tail index estimate, α .

We used the extreme quantile estimator Eq. (8) to obtain quantile forecasts for a range of probabilities, for the 10 minute data set. See Table 7. The results for the 1 second data set are similar and can be obtained from the authors. The probabilities are represented as a multiple of the sample size, n . The number k indicates that we will forecast the expected maxima in a sample of size $k \cdot n$. We know the empirical quantile when $k \leq 1$ and report those numbers.

References

- Arzac, E.R., Bawa, V.S., 1977. Portfolio choice and equilibrium in capital markets with safety-first investors. *Journal of Financial Economics* 4, 277–288.

- Dacorogna, M.M., Muller, U.A., Pictet O.V., de Vries, C.G., 1995. The distribution of extremal foreign exchange rate returns in extremely large data sets. Tinbergen Institute Working Paper, TI 95-70.
- Danielsson, J., Jansen, D.W., de Vries, C.G., 1996. Beyond the sample: Extreme quantile and probability estimation with application to financial data. Mimeo. Tinbergen Institute, Rotterdam.
- De Haan, L., Jansen, D.W., Koedijk, K., de Vries, C.G., 1994. Safety first portfolio selection, extreme value theory and long run asset risks. In: Galambos, J. (Ed.), *Proceedings from a Conference on Extreme Value Theory and Applications*. Kluwer Press, pp. 471–487.
- De Haan, L., Resnick, S.I., Rootzen, H., De Vries, C.G., 1989. Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Process and their Applications* 32, 213–224.
- De Haan, L., Stadtmuller, U., 1996. Generalized regular variation of second order. *Journal of the Australian Mathematical Society, Series A*, 61, pp. 381–395.
- De Vries, C.G., 1994. Stylized facts of the nominal exchange rate returns In: van der Ploeg, F. (Ed.), *The Handbook of International Macroeconomics*. Basil Blackwell, pp. 348–389.
- Feller, W. 1971. *An Introduction to Probability Theory and its Applications*, vol. II, 2nd ed. Wiley, New York.
- Goldie, C.M., Smith, R.L., 1987. Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics* 38, 45–71, Oxford 2nd series.
- Hall, P., 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B*, 42, pp. 37–42.
- Hall, P., 1990. Using the bootstrap to estimate mean square error and select smoothing parameter in nonparametric problem. *Journal of Multivariate Analysis* 32, 177–203.
- Jansen, D.W., Koedijk, C.G., de Vries, C.G. 1995. Operationalizing safety first portfolio selection using extreme value theory. Tinbergen Institute Working Paper, TI 95-84.
- Leadbetter, M.R., Lindgren, G., Rootzen, H., 1983. *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics. Springer-Verlag, New York.
- Muller, U.A., Dacorogna, M.M., Olsen, R.B., Pictet, O.V., Schwarz, M., Morgeneegg, C., 1990. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law and intra day analysis. *Journal of Banking and Finance*, 14, 1189–1208.
- Roy, A.D., 1952. Safety first and the holding of assets. *Econometrica* 20, 431–449.
- Shao, J., Ti, W., 1995. *The Jackknife and Bootstrap*. Springer Verlag, New York.