ARTICLE IN PRESS
JID: ECOSTA [m3Gsc;August 4, 2023;4:21]
Econometrics and Statistics xxx (xxxx) xxx
Contents lists available at ScienceDirect
Econometrics and Statistics
journal homepage: www.elsevier.com/locate/ecosta

Estimation of Extreme Risk Measures for Stochastic Volatility Models with Long Memory and Heavy Tails

Clémonell Bilayi-Biakana[a], Gail Ivanoff[b], Rafał Kulik[c,*]

[a] Department of Mathematics and Statistics, University of Guelph, 50 Stone Road East, N1G 2W1, Guelph, ON, Canada
[b] Department of Mathematics and Statistics, University of Ottawa, 550 Cumberland St, K1N 6N5, Ottawa, ON, Canada

A R T I C L E I N F O

Article history:
Received 14 October 2021
Revised 17 July 2023
Accepted 18 July 2023
Available online xxx

Keywords:
Long Range Dependence
Heavy Tails
Hill Estimator
Stochastic Volatility
Value-at-Risk
Expected Shortfall

A B S T R A C T

Financial data, such as returns on investments, typically exhibit some non-standard features: long memory or long range dependence (LRD) and heavy tails. Therefore, any mathematical model approximating the evolution of asset price should be able to generate these properties. This can be achieved through the use of a long memory stochastic volatility (LMSV) model. The focus is on estimation of Value-at-Risk (VaR) and Expected Shortfall (ES) for such models. While long memory has no effect on the Hill estimator of the tail index, in contrast it is shown that long memory affects the rates of convergence and asymptotic behaviour of the estimators of VaR and ES.

1. Introduction

The goal is to establish the asymptotic behaviour of estimators of extreme risk measures for a financial asset whose returns are heavy-tailed and exhibit long memory. The Value-at-Risk (VaR) and the Expected Shortfall (ES) are commonly used measures of risk. For a continuous random variable X with distribution function F_X, the Value-at-Risk of X at level $p \in (0,1)$ is defined as

$$\mathrm{VaR}_p(X) = F_X^{\leftarrow}(1-p) =: Q_X(1/p),$$

where $F_X^{\leftarrow}$ is the left-continuous inverse of F_X and $Q_X(t) = F_X^{\leftarrow}(1-1/t),\ t > 1$. Likewise, the Expected Shortfall at level p is defined as

$$\mathrm{ES}_p(X) = E\Big(X | X > Q_X(1/p)\Big).$$

Given stationary data $X_1, \dots, X_n$, the objective is to estimate both $\mathrm{VaR}_p(X)$ and $\mathrm{ES}_p(X)$. If the distribution of X is known (e.g. normal, Student t), then analytical formulas are usually available and it remains only to estimate the model parameters. Alternatively, one can estimate both risk measures using nonparametric estimators. Estimation of Value-at-Risk is equivalent to quantile estimation. For this, asymptotic theory for order statistics or kernel quantile estimators exists for both independent and weakly dependent data (see e.g. Sheather and Marron (1990) or Chen (2005)). Similarly, different versions of estimators of the Expected Shortfall have been proposed, again based on (smoothed) order statistics, for both i.i.d. and time series data

* Corresponding author.
E-mail addresses: cbilayib@uoguelph.ca (C. Bilayi-Biakana), givanoff@uottawa.ca (G. Ivanoff), rkulik@uottawa.ca (R. Kulik).

https://doi.org/10.1016/j.ecosta.2023.07.004

Please cite this article as: C. Bilayi-Biakana, G. Ivanoff and R. Kulik, Estimation of Extreme Risk Measures for Stochastic Volatility Models with Long Memory and Heavy Tails, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2023.07.004

(see Scaillet (2005), Chen (2008) or Brazauskas et al. (2008)). In all of these cases, under appropriate assumptions one can establish asymptotic normality with explicit formulas for the limiting variance.

The articles above deal with estimation of the risk measures for a fixed $p \in (0,1)$. However, from a regulatory perspective, one is interested in *extreme risk measures* in which case the value of p is very small. When $p = p_n \to 0$ as the sample size n increases, then the aforementioned limiting theories do not apply and tools adapted from extreme value theory are needed. At the same time, since financial data typically exhibit heavy tails, an important line of research is to model log-returns by means of random variables with distributions with regularly varying tails (for example, Pareto or Student t). A random variable X is regularly varying at infinity (with index α) if for all $t > 0$

$$\lim_{x\to\infty} \frac{P(X > tx)}{P(X > x)} = t^{-\alpha}.$$

As will be seen, an estimator of the Value-at-Risk can be defined as

$$X_{(n-k)} \left(\frac{k}{np}\right)^{\widehat{\gamma}},$$

where $X_{(1)} \leqslant \cdots \leqslant X_{(n-k)} \leqslant \cdots \leqslant X_{(n)}$ are the order statistics corresponding to the data set, $k = k_n$ is a sequence of integers such that $k_n \to \infty$ and $k_n/n \to 0$, and $\widehat{\gamma}$ is an estimator of the reciprocal of the tail index $\gamma = 1/\alpha$. This form of the estimator is due to Weissman (1978). Similarly, an estimator of the Expected Shortfall can be constructed as

$$\frac{1}{1-\widehat{\gamma}} X_{(n-k)} \left(\frac{k}{np}\right)^{\widehat{\gamma}}.$$

It is readily seen that the asymptotic theory for both estimators involves the asymptotic theory for the *intermediate* order statistics $X_{(n-k)}$ as well as the asymptotic theory for the estimators of the tail index. For i.i.d. and weakly dependent data, the limiting theory for the above estimators and their modifications is readily available (see e.g. Weissman (1978), de Haan and Ferreira (2006), Necir et al. (2010) among others).

In addition to heavy tails, log-returns exhibit long range dependence of absolute values. In order to reflect the effects of long memory, Long Memory Stochastic Volatility (LMSV) models were proposed in Breidt et al. (1998). An overview of these models is given in Deo et al. (2006) and Deo et al. (2009). However, these particular models cannot capture heavy-tailed behaviour. In order to model long memory and heavy tails simultaneously, another LMSV model was proposed in Kulik and Soulier (2011) (see also Bilayi-Biakana et al. (2019) for a model with leverage). It was shown there that long memory does not influence the asymptotic behaviour of estimators of the tail index. Based on these findings one could expect that estimation of extremal risk measures is again not affected by long memory. However, as is shown in this article, this is not the case. In fact, the asymptotic behaviour of the estimators of the extreme risk measures VaR and ES depends on a fine interplay between the number of order statistics k used in the estimators and the strength of long memory.

The rest of the article is organized as follows. In Section 2 the model is introduced and all relevant assumptions are stated. The assumptions stem from Bilayi-Biakana et al. (2019), where additional discussion can be found. In Section 3 the joint limiting behaviour of the tail empirical process and the order statistics $X_{(n-k)}$ is reviewed. These results are key to establishing the asymptotic behaviour of estimators of extreme risk measures. In Section 4 weak convergence of integral functionals of the tail empirical process is reviewed. This provides a unified approach to central limit theorems for estimators of the tail index. All the results in Sections 3 and 4 are proven in Bilayi-Biakana et al. (2019).

The main contributions appear in Section 5: the asymptotic behaviour of estimators of Value-at-Risk and Expected Shortfall for the LMSV model. This is done in Section 5.1 and 5.2, respectively. Theorems 5.1 and 5.3 are the principal results, illustrating the non-trivial influence of both the tail index and long memory on the asymptotic behaviour of the estimators. Simulation studies appear in Section 6. In Section 7, a summary of results and directions for further research are provided. The proofs of all new results are found in Section 8. Properties of regular variation are addressed in Section 8.1 and technical details on the models used in the simulations may be found in Section 8.2.

2. Long Memory Stochastic Volatility Model

The following real-valued long memory stochastic volatility (LMSV) model is considered:

$$X_j = \phi(Y_j) Z_j\,, \quad j \in \mathbb{Z}\,. \tag{1}$$

The variables Z_j represent noise and are assumed to be non-negative, while volatility comes from the scaling variables $\phi(Y_j)$. The goal is to incorporate both long memory and heavy tails in the model. Heavy tails come from the i.i.d. noise variables Z_j, while long memory is built into the volatility $\phi(Y_j)$, which may be interpreted as random scaling applied to heavy tailed noise. The assumption of non-negative noise is made to avoid superfluous technical detail, but in practice is not necessary since it is only the upper tail behaviour that is of interest. For further discussion, see Remark 2.2.

The following assumptions A(i)-A(v) are made. Additional explanation and discussion can be found in Bilayi-Biakana et al. (2019). To proceed, recall that a real-valued function ℓ is slowly varying if

$$\lim_{x\to\infty} \ell(tx)/\ell(x) = 1,$$

for all $t > 0$. Examples include for instance non-negative constants and $\log(x)$.

A(i) The sequence $\{Y_j\}$ is strictly stationary and ergodic long memory Gaussian, that is

$$Y_j = \sum_{i=1}^{\infty} a_i \epsilon_{j-i} \,,$$

where $\{\epsilon_j\}$ is a sequence of i.i.d. standard normal random variables and

$$a_i = i^{d-1}\ell_a(i), \; \sum_{i=1}^{\infty} a_i^2 = 1,$$

where $d \in (0, 1/2)$ and $\ell_a(\cdot)$ is slowly varying. As a consequence, $\gamma_Y(j) = \mathrm{Cov}(Y_0, Y_j) \sim j^{2d-1}\ell_Y(j)$. Note that ℓ_a and ℓ_Y are slowly varying functions at infinity such that:

$$\ell_Y(j) = \ell_a^2(j)B(1-2d, d),$$

where $B(a, b)$ stands for the Beta function with parameters a and b. The parameter $0 < d < 1/2$ is referred to as the long memory parameter (for details, see Beran et al. (2013)). Furthermore, assume that $\{(\epsilon_j, Z_j)\}$ is a sequence of i.i.d. random vectors. For each fixed j, ϵ_j and Z_j may be dependent, but due to the construction above, the random variables Y_j and Z_j are independent. However, there can be dependence between the sequences $\{Z_j\}$ and $\{Y_j\}$, allowing for *leverage* in the model.

A(ii) The random variables Z_j are i.i.d. with tail distribution function $\bar{F}_Z$ that is second-order regular varying at infinity with parameters $-\alpha, -\kappa$ and rate function η^*. This means that for $x > 0$,

$$\bar{F}_Z(x) = c^* x^{-\alpha} \exp\left(\int_1^x \frac{\eta^*(u)}{u}\, du\right),$$

where $\alpha, \kappa, c^* > 0$ and η^* is either non-negative or non-positive, regularly varying at infinity with index $-\kappa$: i.e. there exists a function ℓ^* slowly varying at infinity such that $\eta^*(x) = x^{-\kappa}\ell^*(x)$. Further, η^* is bounded - that is, there exists $\beta > 0$ such that $|\eta^*(x)| \le \beta$. In particular, $\bar{F}_Z$ is regularly varying at infinity with index $-\alpha$, that is

$$\lim_{x\to\infty} \bar{F}_Z(tx)/\bar{F}_Z(x) = t^{-\alpha} \text{ for } t > 0.$$

A(iii) The function ϕ is a non-negative measurable function and $\phi(Y_0)$ is not equal to 0 with probability one. It is assumed that ϕ^α is square integrable and m denotes its Hermite rank. The rank determines the asymptotic behaviour of sums of long memory sequences. (For more details on the Hermite rank, see Beran et al. (2013) pg. 108.)

A(iv) Let $k_n \to \infty$ be an increasing sequence of positive integers such that $k_n/n \to 0$ and let $u_n \to \infty$ be defined by $u_n = \bar{F}_X^{\leftarrow}(k/n)$ where $\bar{F}_X^{\leftarrow}$ is the left continuous inverse function of the tail distribution function $\bar{F}_X$ of X. (As will be argued below, $\bar{F}_X$ is continuous). For ease of notation, it is convenient to suppress dependence of k_n on n, which is the standard practice in the extreme value literature. Moreover, consider the sequence $(b_{n,m})_n$ defined as follows:

$$b_{n,m} := n^{1-m(1/2-d)}\sqrt{m!\delta_m \ell_Y^m(n)}\,, \tag{2a}$$

$$\delta_m = \frac{2}{[(2d-1)m+1][(2d-1)m+2]}\,, \tag{2b}$$

where m is the Hermite rank as introduced in A(iii). Assume that

$$\lim_{n\to\infty} a_{n,m}\eta^*\left(\bar{F}_X^{\leftarrow}(k/n)\right) = 0, \tag{3}$$

where

$$a_{n,m} := \left(\sqrt{k} + \frac{n}{b_{n,m}}\right)\mathbb{1}_{\{m(1-2d)<1\}} + \sqrt{k}\mathbb{1}_{\{m(1-2d)>1\}}.$$

A(v) For all $\epsilon > 0$, and $\alpha, \beta, \kappa > 0$ as above,

$$E\left((\phi(Y))^{2\alpha+2\beta}\right) + E\left((\phi(Y))^{2\alpha-2\beta}\right) < \infty, \tag{4a}$$

$$E\left((\phi(Y))^{\alpha+\kappa+\epsilon}\right) + E\left((\phi(Y))^{\alpha+\kappa-\epsilon}\right) < \infty. \tag{4b}$$

Remark 2.1. Under the assumptions A(i)-A(v), it can be shown that this model captures heavy tails, long-memory and stochastic volatility. The moment assumptions imposed in A(v) imply also that $\bar{F}_X$ is regularly varying at infinity with index $-\alpha$. In fact, by Breiman's lemma (see Lemma 8.1),

$$\lim_{x\to\infty} \frac{\bar{F}_X(x)}{\bar{F}_Z(x)} = E(\phi^\alpha(Y)) \ .$$

As indicated in the limiting behaviour above, the tail behaviour is not affected by (long memory) random scaling. Furthermore, under the stated assumptions, $\bar{F}_X$ is second-order regularly varying at infinity with parameters $-\alpha, -\kappa$ and the rate function $\tilde{\eta}$ such that:

$$\tilde{\eta}(x) = \frac{E\big(\phi^\alpha(Y)\eta^*(x/\phi(Y))\ell^*(x/\phi(Y))\big)}{E\big(\phi^\alpha(Y)\ell^*(x/\phi(Y))\big)} \underset{x\to\infty}{\sim} \frac{E\big(\phi^{\alpha+\kappa}(Y)\big)}{E\big(\phi^\alpha(Y)\big)}\eta^*(x). \tag{5}$$

See (Kulik and Soulier, 2011, p. 117).

Remark 2.2. In principle, for the theoretical results it is not necessary to assume that the Z_j's (and hence the X_j's) are non-negative, since it is only the upper tail of the distribution that is of interest. However, it shall be seen that estimators of the upper tail parameter α and the risk measures (Value-at-Risk and Expected Shortfall) make sense only for non-negative values of the intermediate order statistic $X_{(n-k)}$. In practice, this is handled by ignoring negative observations of X_j or by taking absolute values if it can be assumed that the lower tail is not heavier than the upper.

Remark 2.3. Some of the assumptions warrant additional explanation.

- A(ii) This definition of second order regular variation differs from that seen in de Haan and Ferreira (2006) (Definition 2.3.1). The version used here stems from Drees (1998) (see also Kulik and Soulier (2011)) and is appropriate for establishing the asymptotic behaviour of empirical tail distributions. In contrast, the definition in de Haan and Ferreira (2006) is better adapted to the quantile function - see Theorem 2.3.9 of de Haan and Ferreira (2006). The precise relationships between different notions of second order regular variation can be found in Bilayi-Biakana (2019).
- A(iii) The Hermite rank m plays a crucial role in the asymptotic behaviour of long memory sequences, and can lead to non-standard scaling (see $b_{n,m}$ above) and non-standard limiting processes. In particular, if $m > 1$, the limiting processes that appear are so-called Hermite-Rosenblatt-type: i.e. multiple stochastic integrals with respect to a Gaussian white noise - see Definition 8.1.
- A(iv) Condition (3) deals with the bias introduced by the tail empirical process. If the X_j's are i.i.d., the condition becomes $\lim_{n\to\infty}\sqrt{k}\eta^*\big(\bar{F}_X^{\leftarrow}(k/n)\big) = 0$. For the more general LMSV model, the rate $n/b_{n,m}$ is included to control the long memory behaviour. The same approach was used in Kulik and Soulier (2011).

3. Limit Theorems for Tail Empirical Processes (TEP) and Order Statistics

In this section, the joint weak convergence of the tail empirical process (TEP) and the order statistics $X_{(n-k)}$ is reviewed. For details, see Bilayi-Biakana et al. (2019). Interestingly, it will be seen that while the asymptotic behaviour of the order statistics can be influenced by long memory, that of the TEP is unaffected. These results are critical for determining the asymptotic behaviour of the estimators of risk measures.

The TEP is used to investigate the estimators of the tail index α of X. Let the sequences (k_n) and (u_n) be as defined in A(iv). If

$$T_u(t) := \frac{\bar{F}_X(ut)}{\bar{F}_X(u)} \ , \tag{6}$$

then

$$T_{u_n}(t) = \frac{\bar{F}_X(u_n t)}{k/n} \ , \quad \lim_{n\to\infty} T_{u_n}(t) = t^{-\alpha} =: T(t) \ .$$

An estimator of $T(t)$ is used to estimate the tail index α. However, two issues arise. The first is that $T(t)$ cannot be estimated directly; instead one estimates $T_{u_n}(t)$. As discussed in Bilayi-Biakana et al. (2019), this introduces bias which can be controlled thanks to the second order regular variation of $\bar{F}_X$ and (3): for any $\tau_0 > 0$,

$$a_{n,m} \sup_{t>\tau_0} \left|T_{u_n}(t) - T(t)\right| \underset{n\to\infty}{\to} 0.$$

(See Bilayi-Biakana et al. (2019). equation (10).)

The second issue is that while the empirical distribution can be used to estimate the unknown distribution F_X, the upper quantiles u_n are also unknown. Thus, a data based tail empirical process must be introduced. Let $F_{n,X}$ be the usual empirical distribution function and $\bar{F}_{n,X}(x) = 1 - F_{n,X}(x)$. Let $X_1, \ldots, X_n$ be a sample from the stochastic volatility model defined in (1) and let $X_{(1)} \leqslant \cdots \leqslant X_{(i)} \leqslant \cdots \leqslant X_{(n)}$ be the corresponding order statistics.

Since $\bar{F}_X$ is continuous, $u_n = \bar{F}_X^{\leftarrow}\big(k/n\big)$ for $k = k_n$ and $\bar{F}_{n,X}^{\leftarrow}\big(k/n\big) = X_{(n-k)}$, and so it is natural to approximate $\bar{F}_X$ with $\bar{F}_{n,X}$ and u_n with $X_{(n-k)}$.

Accordingly, the *tail empirical distribution function* of $\{X_j\}$ is defined as:

$$\widehat{T}_n(t) := \frac{\bar{F}_{n,X}(X_{(n-k)}t)}{k/n} = \frac{1}{k}\sum_{j=1}^{n} \mathbb{1}_{\{X_j > X_{(n-k)}t\}}, \quad t > 0. \tag{7}$$

The *tail empirical process* of $\{X_j\}$ is

$$\widehat{S}_n(t) := k\big(\widehat{T}_n(t) - T(t)\big).$$

The next theorem specifies the joint asymptotic behaviour of the TEP and the corresponding order statistics, suitably scaled. It is an immediate consequence of Lemma 3.7 and Theorem 3.8 of Bilayi-Biakana et al. (2019). The asymptotic behaviour depends on the interplay between the memory parameter d, the Hermite rank m and the number of order statistics k.

Theorem 3.1. *Let $\{X_j\}$ be the long memory stochastic volatility model given in (1). Then,*

$$\frac{\widehat{S}_n(t)}{\sqrt{k}} \xrightarrow[n\to\infty]{d} W\big(T(t)\big) - T(t)W(1),$$

in $D(0,\infty)$ equipped with the Skorokhod J_1 topology, where $W(\cdot)$ is a standard Brownian motion on [0,1]. The limiting process $W\big(T(\cdot)\big) - T(\cdot)W(1)$ is a centered time-changed Brownian bridge on $[1,\infty)$.

Furthermore,

- *If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to 0$, or $m(1-2d) > 1$, then as $n \to \infty$,*

$$\left(\frac{\widehat{S}_n(t)}{\sqrt{k}}, \sqrt{k}\left(\frac{X_{(n-k)}}{u_n} - 1\right)\right) \xrightarrow[n\to\infty]{d} \left(W\big(T(t)\big) - T(t)W(1), \frac{W(1)}{\alpha}\right).$$

- *If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to \infty$, then as $n \to \infty$,*

$$\left(\frac{n}{kb_{n,m}}\widehat{S}_n(t), \frac{n}{b_{n,m}}\left(\frac{X_{(n-k)}}{u_n} - 1\right)\right) \xrightarrow[n\to\infty]{d} \left(0, \frac{\mu_{\phi,\alpha}(m)}{\alpha m! E(\phi^\alpha(Y))}\xi_{m,d+1/2}(1)\right),$$

  where $\xi_{m,d+1/2}(1)$ is Hermite-Rosenblatt (cf. Definition 8.1) with the Hurst parameter $d+\frac{1}{2}$ and $\mu_{\phi,\alpha}(m) = E(H_m(Y)\phi^\alpha(Y))$. ($H_m$ denotes the Hermite polynomial of order m.)

These two joint weak convergences hold in $D(0,\infty)\times\mathbb{R}$.

Comment 3.1. There are a few remarks to be made about the preceding theorem. First, the limiting behaviour of the TEP is unaffected by long memory and converges at rate $\sqrt{k}$. On the other hand, the limiting behaviour of the scaled order statistics is dichotomous and depends either on the tail behaviour of the distribution or on long memory. In the first case, the rate of convergence is the same for the TEP and the order statistics. Furthermore, it should be noted that the limiting distributions are independent since the limiting covariances are 0.

It is the strength of long memory (i.e. the value of the parameter d appearing in the denition of $b_{n,m}$) that determines which limiting behaviour prevails for the order statistics. In the long memory case, the limiting random variable, $\xi_{m,d+1/2}(1)$, is non-Gaussian unless $m = 1$. See Beran et al. [2013] for more details.

4. Tail Index Estimation

The power of weak convergence theory comes from the fact that many diverse results emerge as corollaries of a basic convergence theorem. As will be seen in Theorem 4.1, the main convergence Theorem 3.1 can be extended to integral functionals of the tail empirical process. This in turn yields a unified approach to establishing weak convergence of estimators of the tail index. In what follows, r denotes an arbitrary non-negative integer.

Theorem 4.1. *Let $\{X_j\}$ be the long memory stochastic volatility model in (1).*

If $\alpha > 2(1-r)$, then

$$\sqrt{k}\int_1^\infty \frac{\widehat{T}_n(t) - T(t)}{t^r}dt = \frac{1}{\sqrt{k}}\int_1^\infty \frac{\widehat{S}_n(t)}{t^r}\,dt \xrightarrow[n\to\infty]{d} \int_1^\infty \frac{W\big(T(t)\big) - W(1)T(t)}{t^r}\,dt.$$

The proof can be found in Bilayi-Biakana et al. (2019). Notice that

$$\int_1^\infty \frac{W\big(T(t)\big) - W(1)T(t)}{t^r}\,dt \stackrel{d}{=} \frac{\alpha^{1/2}}{(\alpha+r-1)(\alpha+2r-2)^{1/2}}\mathcal{N},$$

where $\mathcal{N}$ is a standard normal random variable.

Since the tail distribution of X is regularly varying with index $-\alpha$, then this raises the question of estimating the index of regular variation α. For this purpose, attention is restricted to the family of tail index estimators, $\widehat{\gamma}_{r,k}$, known as *harmonic moment estimators* (HME) of order r of $\gamma := 1/\alpha$. The goal is to prove their asymptotic normality. Since $T(t) = t^{-\alpha}$, then for $r \geqslant 0$,

$$\zeta_r := \int_1^\infty \frac{T(t)}{t^r}\,dt = \frac{1}{\alpha + r - 1}. \tag{8}$$

If $\widehat{\zeta}_{r,k}$ denotes an estimator of ζ_r, then the plug-in method and (7) yield

$$\widehat{\zeta}_{r,k} = \int_1^\infty \frac{\widehat{T}_n(t)}{t^r}\,dt = \int_1^\infty \frac{1}{k}\sum_{j=1}^n 1\!\!1_{\{X_{(j)} > X_{(n-k)}t\}} \frac{dt}{t^r} = \frac{1}{k}\sum_{j=1}^n \int_1^\infty 1\!\!1_{\left\{\frac{X_{(j)}}{X_{(n-k)}} > t\right\}} \frac{dt}{t^r}.$$

Furthermore, since $t \geqslant 1$, then the following hold:

$$\begin{aligned}\widehat{\zeta}_{r,k} &= \frac{1}{k}\sum_{j=1}^k \int_1^{\frac{X_{(n-j+1)}}{X_{(n-k)}}} \frac{dt}{t^r} \\ &= \begin{cases} \frac{1}{r-1}\left(1 - \frac{1}{k}\sum_{j=1}^k \left(\frac{X_{(n-k)}}{X_{(n-j+1)}}\right)^{r-1}\right) & \text{if } r \neq 1, \\ \frac{1}{k}\sum_{j=1}^k \ln\left(\frac{X_{(n-j+1)}}{X_{(n-k)}}\right) & \text{if } r = 1. \end{cases}\end{aligned}$$

To derive the estimators of $\gamma = 1/\alpha$, solving for $1/\alpha$ in (8) yields

$$\zeta_r = \frac{1}{\alpha + r - 1} \Rightarrow \frac{1}{\alpha} = \frac{\zeta_r}{1 + (1-r)\zeta_r}.$$

The plug-in method produces the HMEs below:

$$\begin{aligned}\widehat{\gamma}_{r,k} &= \frac{\widehat{\zeta}_{r,k}}{1 + (1-r)\widehat{\zeta}_{r,k}} \\ &= \begin{cases} \frac{1}{r-1}\left(\left(\frac{1}{k}\sum_{j=1}^k \left(\frac{X_{(n-k)}}{X_{(n-j+1)}}\right)^{r-1}\right)^{-1} - 1\right) & \text{if } r \neq 1\ , \\ \frac{1}{k}\sum_{j=1}^k \ln\left(\frac{X_{(n-j+1)}}{X_{(n-k)}}\right) & \text{if } r = 1\ . \end{cases}\end{aligned} \tag{9}$$

For i.i.d. regularly varying random variables, the asymptotic theory for harmonic moment estimators was first developed in Beran et al. (2014), where it is shown that although the Hill estimator has the smallest asymptotic variance among HMEs, other HMEs can be more efficient if the data are biased or contaminated by noise.

- The HME that corresponds to $r = 1$ is the Hill estimator of $\gamma = 1/\alpha$.
- The HME that corresponds to $r = 2$ is the t-Hill estimator of γ, that is

$$\widehat{\gamma}_{2,k} = \left(\frac{1}{k}\sum_{j=1}^k \frac{X_{(n-k)}}{X_{(n-j+1)}}\right)^{-1} - 1\ .$$

The main result of this section, the asymptotic normality of $\widehat{\gamma}_{r,k}$, is a simple application of the delta method and Theorem 4.1. It is noticeable that long memory does not affect the limiting behaviour of the estimators of $1/\alpha$. A heuristic explanation is that long memory enters the model through (random) scaling, while at the same time, the index of regular variation is scale invariant.

Theorem 4.2 (cf. Bilayi-Biakana et al. (2019)). *Let $\{X_j\}$ be the long memory stochastic volatility model in (1). If $\alpha > 2(1-r)$, then*

$$\sqrt{k}\left(\widehat{\gamma}_{r,k} - \gamma\right) \xrightarrow[n\to\infty]{d} \frac{(\alpha + r - 1)}{\left(\alpha^3(\alpha + 2r - 2)\right)^{1/2}}\mathcal{N}\ , \tag{10}$$

where $\mathcal{N}$ is a standard normal random variable.

Recalling Theorem 3.1 and nhhd5 3.1, the following Corollary holds. Although long memory does not affect HMEs, it does play a role when quantiles are considered. Since quantiles are not scale invariant, this is unsurprising.

Corollary 4.1. *Let $\{X_j\}$ be the long memory stochastic volatility model given in (1). If $\alpha > 2(1-r)$, then*

- *If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to 0$, or $m(1-2d) > 1$, then as $n \to \infty$,*

$$\sqrt{k}\left((\widehat{\gamma}_{r,k} - \gamma) \, , \left(\frac{X_{(n-k)}}{u_n} - 1 \right) \right) \xrightarrow[n\to\infty]{d} \left(\frac{(\alpha + r - 1)}{\left(\alpha^3(\alpha + 2r - 2)\right)^{1/2}} \mathcal{N} \, , \frac{1}{\alpha} \tilde{\mathcal{N}} \right),$$

where $\mathcal{N}$ and $\tilde{\mathcal{N}}$ are independent standard normal random variables.

- *If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to \infty$ then as $n \to \infty$,*

$$\frac{n}{b_{n,m}}\left((\widehat{\gamma}_{r,k} - \gamma) \, , \left(\frac{X_{(n-k)}}{u_n} - 1 \right) \right) \xrightarrow[n\to\infty]{d} \left(0, \, \frac{\mu_{\phi,\alpha}(m)}{\alpha m! E(\phi^\alpha(Y))} \xi_{m,d+1/2}(1) \right),$$

where $\xi_{m,d+1/2}(1)$ is Hermite-Rosenblatt with the Hurst parameter $d + \frac{1}{2}$ and $\mu_{\phi,\alpha}(m) = E(H_m(Y)\phi^\alpha(Y))$. ($H_m$ denotes the Hermite polynomial of order m.)

5. Estimation of Extreme Risk Measures

This section deals with statistics of extreme financial risk of a financial asset. Three approaches are usually considered. The first approach stems from portfolio-theory, in which an investor relies heavily on the expected return and the magnitude of the standard deviation. This is not of interest in this article. The second approach is based on Value-at-Risk (VaR), a measure of financial risk that captures not only the volatility of an asset but also the maximum of the likely loss. The third approach is based on coherent risk measures, which capture the size of a potential loss of a financial institution. In short, a coherent risk measure ρ satisfies $\rho(0) = 0$ and must be decreasing, subadditive, positively homogeneous and transitionally invariant. These financial risk measures were introduced in Artzner et al. (1999) in the late nineties and are alternatives to Value-at-Risk which is not subadditive. One such coherent risk measure is the so-called Expected Shortfall (ES) (cf. Acerbi and Tasche (2002a)). In this section, estimation of Value-at-Risk and Expected Shortfall is investigated under the assumption that returns are heavy-tailed long memory sequences as defined in (1).

5.1. Estimation of Value-at-Risk

Recall the LMSV model defined in (1): $X_j = \phi(Y_j)Z_j, \;\; j \in \mathbb{Z}$. The goal is to estimate Value-at-Risk at level $p \in (0,1)$, that is

$$\mathrm{VaR}_p(X) = F_X^{\leftarrow}(1-p) = Q_X(1/p) \; ,$$

where $F_X^{\leftarrow}$ is the left-continuous inverse of F_X and $Q_X(t) = F_X^{\leftarrow}(1 - 1/t)$, $t > 1$. Let $X_{(1)} \leqslant \cdots \leqslant X_{(i)} \leqslant \cdots \leqslant X_{(n)}$ be the order statistics of $(X_1, \ldots, X_n)$. Notice

$$E\left(\sum_{j=1}^{n} \mathbb{1}_{\{X_j > Q_X(1/p)\}} \right) = np,$$

and so np, the expected number of observations exceeding $Q_X(1/p)$, is small if p is very small. In the sequel, it is assumed that p depends on n and $p = p_n \to 0$, as the sample size n gets large.

If $F_{n,X}$ denotes the empirical distribution function of the sample $(X_1, \ldots, X_n)$, then the usual empirical estimate of $Q_X(1/p)$ is

$$F_{n,X}^{\leftarrow}(1-p) = X_{(n-[np])} \; . \tag{11}$$

However, for a very small value of p, this procedure is not very reliable since for $p_1 \neq p_2$ such that $[np_1] = [np_2]$, the values $Q_X(1/p_1)$ and $Q_X(1/p_2)$ may differ significantly, but both values will be estimated by the same order statistic $X_{(n-[np_1])}$. In particular, for all $p < 1/n$, $Q_X(1/p)$ will be always estimated by $X_{(n)}$.

To address this, let $k = k_n$ be an intermediate sequence such that $k \to \infty$ and $k/n \to 0$. Since $\bar{F}_X$ is strictly decreasing on the range of X, then $1/\bar{F}_X$ is strictly increasing. Further, the fact $\bar{F}_X$ is regularly varying at infinity with index $-\alpha$ implies that $U_X = \left(1/\bar{F}_X\right)^{\leftarrow}$ is regularly varying with index $\gamma := 1/\alpha$, (cf. Proposition B.1.9 in de Haan and Ferreira (2006)). Thanks to the continuity of $\bar{F}_X$, U_X and Q_X coincide. Therefore, as a result of the representation theorem (cf. (Bingham et al., 1987, p. 21)), $Q_X(x) = x^\gamma \tilde{\ell}(x)$, where $\tilde{\ell}$ is a slowly varying function at infinity. Thus,

$$\frac{Q_X(1/p)}{Q_X(n/k)} = \frac{Q_X\left((k/np)n/k\right)}{Q_X(n/k)} = \frac{(k/np)^\gamma \tilde{\ell}((k/np)n/k)}{\tilde{\ell}(n/k)} \underset{n\to\infty}{=} \left(\frac{k}{np}\right)^\gamma (1 + o(1)), \tag{12}$$

as long as (k/np) is bounded away from 0 and ∞ or $k/(np) \to \infty$. The former case is more relevant from a practical point of view.

By (11), $Q_X(n/k)$ can be estimated by $X_{(n-k)}$. With $\widehat{\gamma}$ denoting an HME estimator $\widehat{\gamma}_{r,k}$ of $\gamma = 1/\alpha$, (12) suggests the following estimator of Value-at-Risk $Q_X(1/p)$:

$$\widehat{Q}_X(1/p) = X_{(n-k)} \left(\frac{k}{np}\right)^{\widehat{\gamma}}. \tag{13}$$

Next, the limiting behaviour of this estimator of Value-at-Risk is investigated under the assumptions of the long memory stochastic volatility model. For this, it is essential to use Corollary 4.1, which defines the joint limiting behaviour of the intermediate order statistics and the estimators of $1/\alpha$.

Two limiting schemes are considered, either $k/(np) \to \infty$ or $k/(np) \to \nu \in (0, \infty)$. The first is the classical situation of intermediate order statistics and the limiting scheme that appears in statistics for extremes. The second has received less attention in the extreme value literature, but it is more relevant for practical purposes. Indeed, it can explain the results of some finite sample simulation studies.

The following no-bias condition is needed, which is satisfied if (3) holds (see Lemma 8.4):

$$\limsup_{n\to\infty} a_{n,m} \left| \frac{(k/np)^{\gamma} Q_X(n/k)}{Q_X(1/p)} - 1 \right| = 0. \tag{14}$$

Note that the no-bias condition (14) is always valid for Pareto distributions.

Theorem 5.1. *Let $\{X_j\}$ be the long memory stochastic volatility model as in (1) and let $\widehat{\gamma}$ denote the HME estimator $\widehat{\gamma}_{r,k}$ in (12). Assume that $\alpha > 2(1-r)$ and let $\mathcal{N}$ denote a standard normal variable and $0 < \nu < \infty$.*

1. *Assume that $k/np \to \nu$. If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to 0$ or $m(1-2d) > 1$, then*

$$\sqrt{k}\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\sqrt{\frac{1}{\alpha^2} + \frac{(\alpha+r-1)^2 \ln^2 \nu}{\alpha^3(\alpha+2r-2)}}. \tag{15}$$

2. *Assume that $k/np \to \nu$. If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to \infty$, then*

$$\frac{n}{b_{n,m}}\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right) \xrightarrow[n\to\infty]{d} \frac{\mu_{\phi,\alpha}(m)\xi_{m,d+1/2}(1)}{\alpha m! E(\phi^\alpha(Y))}. \tag{16}$$

3. *Assume that $k/np \to \infty$. If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to 0$ or $m(1-2d) > 1$, then*

$$\frac{\sqrt{k}}{\ln(k/np)}\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right) \xrightarrow[n\to\infty]{d} \frac{(\alpha+r-1)}{\sqrt{\alpha^3(\alpha+2r-2)}}\mathcal{N}. \tag{17}$$

4. *Assume that $k/np \to \infty$. If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to \infty$, then*

$$\frac{n}{b_{n,m}\ln(k/np)}\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right) \xrightarrow[n\to\infty]{d} 0. \tag{18}$$

Remark 5.1. Note that for $r = 1$, (17) is in agreement with the result of Weissman (1978) in the i.i.d. case (see also (de Haan and Ferreira, 2006, Theorem 4.3.8)). On the other hand, the convergence in (18) is driven by long memory and this result appears to be new. Also note that these results differ from the asymptotic behaviour obtained under weak dependence; see Drees (2000). Indeed, there are two extreme forces driving the asymptotic behaviour of the LMSV model: either the i.i.d. heavy tails, or the light-tailed long memory part. The following remark elaborates on this.

Remark 5.2. It turns out that when returns are assumed to exhibit heavy tails, long range dependence and leverage, the estimation of VaR is influenced either by heavy tails or long memory. Since the estimation of the tail index does not depend on long memory, it plays a role only when the rescaled order statistics converge at rate $\sqrt{k}$. When long memory predominates, only order statistics affect the limiting behaviour of Value at Risk. This is demonstrated in the following scenarios:

1. In case of (15) the limiting behaviour is affected by both order statistics and estimation of the tail index.
2. In case of (16) the limiting behaviour is affected by the limiting behaviour of order statistics only. Estimation of the tail index does not play any role.
3. In case of (17) the limiting behaviour is affected by estimation of the tail index only.
4. In case of (18) the limiting behaviour is degenerate. This is addressed in the following result by imposing more detailed conditions.

Theorem 5.2. *Let $\{X_j\}$ be the long memory stochastic volatility model as in (1) and let $\widehat{\gamma}$ denote the HME estimator $\widehat{\gamma}_{r,k}$ in (12). Assume that $k/np \to \infty$. If $m(1-2d) < 1$, $\alpha > 2(1-r)$ and*

$$\frac{b_{n,m}}{n}\sqrt{k}/\ln(k/np) \to \infty, \tag{19}$$

then

$$\frac{n}{b_{n,m}}\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)}-1\right)\xrightarrow[n\to\infty]{d}\frac{\mu_{\phi,\alpha}(m)\xi_{m,d+1/2}(1)}{\alpha m!E(\phi^\alpha(Y))}.$$

Remark 5.3. Note that (19) automatically implies $\frac{b_{n,m}}{n}\sqrt{k}\to\infty$. However, it is unclear what happens if

$$\frac{b_{n,m}}{n}\sqrt{k}\to\infty \text{ but } \frac{b_{n,m}}{n}\sqrt{k}/\ln(k/np)\to 0.$$

5.2. Estimation of Expected Shortfall

As previously discussed, a number of deficiencies make VaR unsatisfactory as a financial risk measure. Since VaR is simply a quantile, it lacks subadditivity, a serious limitation. To get around the limitations of VaR, in Artzner et al. (1999) the authors introduced coherent measures as alternatives to VaR. It turns out that Expected Shortfall (ES) is the natural coherent alternative to VaR (Acerbi and Tasche (2002a)). The coherence of Expected Shortfall is extensively discussed in Acerbi and Tasche (2002b). Expected Shortfall can be simply viewed as the excess mean function - that is the average value of all values exceeding a certain threshold of the VaR. The ES is also called the Conditional Tail Expectation (CTE) when the distribution of returns is continuous. The expected shortfall of a random variable X at level $p\in(0,1)$ is

$$\mathrm{ES}_p(X)=E\big(X|X>Q_X(1/p)\big)=:\theta_X(p)\ .$$

The goal of this section is to estimate the expected shortfall of $X_j=\phi(Y_j)Z_j$ for the long range stochastic volatility model (1).

By integration by parts and Karamata's theorem (cf. (Bingham et al., 1987, p.26; p.27)), for $\alpha>1$, it is straightforward that

$$E(X/x|X>x)\xrightarrow[x\to\infty]{}\frac{\alpha}{\alpha-1}.$$

Notice that $Q_X(1/p)\to\infty$ if and only if $p\to 0$. Therefore, as $p\to 0$,

$$\theta_X(p)=E\big(X|X>Q_X(1/p)\big)\sim\frac{1}{1-\gamma}Q_X(1/p). \tag{20}$$

This suggests that an estimator of ES can be defined as follows:

$$\widehat{\theta}_X(p)=\frac{1}{1-\widehat{\gamma}}\widehat{Q}_X(1/p). \tag{21}$$

Once again, two limiting schemes are considered, when $k/(np)\to\infty$ and when $k/(np)\to\nu\in(0,\infty)$. As before, a no-bias condition is required:

$$\lim_{n\to\infty}a_{n,m}\left(\frac{Q_X(1/p)}{(1-\gamma)\theta_X(p)}-1\right)=0. \tag{22}$$

It is shown in Lemma 8.5 that the condition is satisfied under second-order regular variation and (3).

Theorem 5.3. *Let $\{X_j\}$ be the long memory stochastic volatility model as in (1) and let $\widehat{\gamma}$ denote the HME estimator $\widehat{\gamma}_{r,k}$ in (21). Assume that $\alpha>1$ if $r\geqslant 1$ and $\alpha>2$ if $r=0$.*

Let $\mathcal{N}$ denote a standard normal variable and $0<\nu<\infty$.

1. *Assume that $k/np\to\nu$. If $m(1-2d)<1$ and $\frac{b_{n,m}}{n}\sqrt{k}\to 0$ or $m(1-2d)>1$, then*

$$\sqrt{k}\left(\frac{\widehat{\theta}_X(p)}{\theta_X(p)}-1\right)\xrightarrow[n\to\infty]{d}\mathcal{N}\sqrt{\frac{1}{\alpha^2}+\frac{(\alpha+r-1)^2}{\alpha^3(\alpha+2r-2)}\left(\frac{\alpha}{\alpha-1}+\ln\nu\right)^2}. \tag{23}$$

2. *Assume that $k/np\to\nu$. If $m(1-2d)<1$ and $\frac{b_{n,m}}{n}\sqrt{k}\to\infty$, then*

$$\frac{n}{b_{n,m}}\left(\frac{\widehat{\theta}_X(p)}{\theta_X(p)}-1\right)\xrightarrow[n\to\infty]{d}\frac{\mu_{\phi,\alpha}(m)\xi_{m,d+1/2}(1)}{\alpha m!E(\phi^\alpha(Y))}. \tag{24}$$

3. *Assume that $k/np\to\infty$. If $m(1-2d)<1$ and $\frac{b_{n,m}}{n}\sqrt{k}\to 0$ or $m(1-2d)>1$, then*

$$\frac{\sqrt{k}}{\ln\big(k/np\big)}\left(\frac{\widehat{\theta}_X(p)}{\theta_X(p)}-1\right)\xrightarrow[n\to\infty]{d}\frac{(\alpha+r-1)}{\sqrt{\alpha^3(\alpha+2r-2)}}\mathcal{N}. \tag{25}$$

4. *Assume that* $k/np \to \infty$. *If* $m(1-2d) < 1$ *and* $\frac{b_{n,m}}{n}\sqrt{k} \to \infty$, *then*

$$\frac{n}{b_{n,m} \ln\left(k/np\right)}\left(\frac{\widehat{\theta}_X(p)}{\theta_X(p)} - 1\right) \xrightarrow[n\to\infty]{d} 0. \tag{26}$$

Remark 5.4. This remark parallels Remark 5.1. For $r = 1$ and the i.i.d. case, the asymptotics in (25) can be deduced from Hoga (2019). The general results for weakly dependent sequences do not apply here. On the other hand, (26) is new and cannot be compared to anything that exists in the literature.

Remark 5.5. As was the case with VaR, estimation of ES is influenced in a dichotomous way either by heavy tails or long memory, and the scenarios described in Remark 5.2 still apply. One additional observation is made: in all scenarios except the first, it is seen that the limiting behaviour is the same for Value-at-Risk and Estimated Shortfall. This is a consequence of the norming. Let C_n denote the norming constant in (24), (25) or (26) and note that in each case $C_n/\sqrt{k_n} \to 0$. Since

$$\frac{\widehat{\theta}_X(p)}{\theta_X(p)} \sim \frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} \cdot \frac{1-\gamma}{1-\widehat{\gamma}} \tag{27}$$

and $\sqrt{k_n}\left(\frac{1-\gamma}{1-\widehat{\gamma}} - 1\right)$ is stochastically bounded (Corollary 4.1), it follows that $C_n\left(\frac{1-\gamma}{1-\widehat{\gamma}} - 1\right) \xrightarrow[n\to\infty]{d} 0$ and the asymptotic behaviour of Value-at-Risk determines that of the product above. The same observation applies to the following Theorem. Details may be found in the proofs.

Theorem 5.4. *Let* $\{X_j\}$ *be the long memory stochastic volatility model as in* (1) *and let* $\widehat{\gamma}$ *denote the HME estimator* $\widehat{\gamma}_{r,k}$ *in* (12). *Assume that* $k/np \to \infty$, $\alpha > 1$ *if* $r \geqslant 1$ *and* $\alpha > 2$ *if* $r = 0$. *Let* $\mathcal{N}$ *denote a standard normal variable and* $0 < \nu < \infty$. *If* $m(1-2d) < 1$ *and*

$$\frac{b_{n,m}}{n}\sqrt{k}/\ln(k/np) \to \infty,$$

then

$$\frac{n}{b_{n,m}}\left(\frac{\widehat{\theta}_X(p)}{\theta_X(p)} - 1\right) \xrightarrow[n\to\infty]{d} \frac{\mu_{\phi,\alpha}(m)\xi_{m,d+1/2}(1)}{\alpha m! E(\phi^\alpha(Y))}. \tag{28}$$

6. Simulation Studies

[1] This section is devoted to some numerical studies of estimation of Value-at-Risk and Expected Shortfall. The following examples of the LMSV model will be considered: the noise variables $\{Z_j\}$, $j = 1, \ldots, n$ are i.i.d. Pareto, mixed Pareto, or the absolute value of Student t random variables. The mixed Pareto has the following tail distribution function:

$$\bar{F}_Z(x) = \begin{cases} \frac{1}{2}\left(x^{-\alpha} + x^{-2\alpha}\right), & x \geqslant 1, \\ 1 & 0 < x < 1. \end{cases}$$

Note that the Pareto case theoretically does not yield bias (formally, $\kappa = \infty$). The mixed Pareto is second order regular varying with $\kappa = \alpha$ (cf. Example 5.1 of Bilayi-Biakana et al. (2019)). Finally, the Student t with b degrees of freedom is second order regularly varying with $\alpha = b$ and $\kappa = 2$ for all b. See Example 3 of Hua and Joe (2011). Simulations are constructed to illustrate

- the effect of bias;
- the differing effects of the strength of long memory;
- the asymptotic behaviour of the estimators as $n \to \infty$, $k = k_n \to \infty$, $k/n \to 0$, $p = p_n \to 0$, $k/np \to \nu$.

With these points in mind, n observations from the long memory stochastic volatility model are simulated:

$$X_j = \exp(\sigma Y_j)Z_j, \quad j = 1, \ldots, n \tag{29}$$

where

1) Z_j is a regularly varying sequence of random variables with index $-\alpha$ (Pareto, mixed Pareto or the absolute value of Student t, $\alpha = 2$ or 4).
2) Y_j is a fractional Gaussian noise sequence, that is, (Y_j) *is a stationary sequence of standard normal random variables and*

$$\mathrm{Cov}(Y_0, Y_j) \sim d(2d+1)j^{2d-1},$$

with the long memory parameter $d \in (0, 1/2)$. *Further,* $\ell_Y(j) \sim d(2d+1)$ *(cf. A(i))*. This is simulated using the R command `fracdiff` for $d = .1$ and $d = .4$.

[1] R codes are available upon request from the authors.

Table 1
Theoretical Standard Deviation.

	n	Tail Index	VaR	ES
	500	0.2115	0.1287	0.3038
$\alpha = 2,\ d = 0.1$	1000	0.1778	0.1082	0.2554
	10000	0.1000	0.0608	0.1436
	500	0.2115	0.0537	0.0537
$\alpha = 2,\ d = 0.4$	1000	0.1778	0.0501	0.0501
	10000	0.1000	0.0398	0.0398
	500	0.2115	0.0644	0.1195
$\alpha = 4,\ d = 0.1$	1000	0.1778	0.0541	0.1005
	10000	0.1000	0.0304	0.0565
	500	0.2115	0.0537	0.0537
$\alpha = 4,\ d = 0.4$	1000	0.1778	0.0501	0.0501
	10000	0.1000	0.0398	0.0398

Table 2
Monte Carlo Results - Pareto.

Pareto	n	Tail Index				VaR				ES			
		Mean	Median	RMSE	Std	Mean	Median	RMSE	Std	Mean	Median	RMSE	Std
	500	1.048	1.009	0.247	0.242	1.005	0.998	0.130	0.130	1.097	0.985	0.493	0.484
$\alpha = 2,\ d = 0.1$	1000	1.034	1.012	0.199	0.196	1.007	0.997	0.116	0.116	1.048	0.983	0.328	0.324
	10000	1.010	1.009	0.104	0.103	1.007	1.001	0.063	0.062	1.023	0.998	0.162	0.160
	500	1.057	1.029	0.246	0.239	1.013	1.005	0.152	0.151	1.095	0.981	0.516	0.507
$\alpha = 2,\ d = 0.4$	1000	1.035	1.013	0.200	0.197	1.016	1.007	0.142	0.142	1.076	0.998	0.379	0.372
	10000	1.009	0.999	0.103	0.103	1.018	1.012	0.089	0.087	1.035	1.018	0.176	0.172
	500	1.039	1.010	0.233	0.230	1.020	1.014	0.071	0.068	1.032	1.018	0.136	0.133
$\alpha = 4,\ d = 0.1$	1000	1.035	1.018	0.194	0.191	1.019	1.016	0.060	0.057	1.024	1.010	0.110	0.108
	10000	1.015	1.009	0.104	0.103	1.019	1.018	0.035	0.030	1.019	1.017	0.060	0.057
	500	1.056	1.020	0.249	0.243	1.030	1.024	0.110	0.106	1.037	1.018	0.164	0.160
$\alpha = 4,\ d = 0.4$	1000	1.034	1.014	0.198	0.195	1.032	1.027	0.098	0.093	1.038	1.022	0.138	0.133
	10000	1.014	1.002	0.103	0.102	1.038	1.037	0.075	0.065	1.038	1.036	0.091	0.082

3) The variability parameter in (29) is $\sigma = 0.1$.
4) The following sample sizes n are considered: $n = 500$, $n = 1000$ and $n = 10,000$. The corresponding number of top order statistics is $k = \sqrt{n}$, so that $k \to \infty$ and $k/n \to 0$.
5) The value of p is $p = k/2n$ (i.e. $\nu = 2$).
6) The estimators of VaR and ES are as defined in (13) and (21), respectively. The estimator $\widehat{\gamma}$ is the Hill estimator $\widehat{\gamma}_{1,k}$ and the estimator of the tail index is $\widehat{\alpha} := (\widehat{\gamma})^{-1}$.
7) The estimates in 6) are obtained from $B = 1000$ Monte Carlo simulations.

Section 8.2 contains detailed remarks about the choice of parameters, justifying the statements below.

- $b_{n,m} = b_{n,1} \sim n^{.5+d}$; $a_{n,m} = a_{n,1} \sim n^{.25} + n^{.5-d}$.
- $b_{n,m}\sqrt{k}/n = b_{n,1}/n^{.75} \sim n^{d-.25}$.

Conditions A(i)-A(v) are satisfied by all examples, *except for the student t, $\alpha = 4$ which does not satisfy the no-bias condition (3) (cf. Section 8.2). The $t(4)$ example is included to illustrate the effect of the choice of k_n on bias.* In all other cases the following are true:

- If $d = 0.1$, then $b_{n,m}\sqrt{k}/n \longrightarrow 0$ and the limiting behaviour of VaR and ES is given by (15) and (23), respectively. Long memory does not affect the Gaussian limits, which depend only on the tail behaviour of the noise. ES has a larger limiting variance than VaR.
- If $d = 0.4$, then $b_{n,m}\sqrt{k}/n \longrightarrow \infty$ and the limiting behaviour of VaR and ES is given by (16) and (24), respectively. Long memory dominates and yields the same Gaussian limit for both VaR and ES.
- The limiting behaviour of $\widehat{\alpha}$ is unaffected by the value of d. In particular, by Theorem 4.2,

$$\sqrt{k}\left(\frac{\widehat{\alpha}}{\alpha} - 1\right) \xrightarrow[n\to\infty]{d} \mathcal{N}.$$

Below tables for the ratios $\widehat{\alpha}/\alpha$, $\widehat{Q}_X(1/p)/Q_X(1/p)$, and $\widehat{\theta}_X(p)/\theta_X(p)$ are presented for the simulated data described above. For each scenario, the summary data in Tables 2 to 4 for the respective ratios include the mean, median, standard deviation and root mean square error (RMSE).

Table 3
Monte Carlo Results - Mixed Pareto.

Pareto	n	Tail Index				VaR				ES			
		Mean	Median	RMSE	Std	Mean	Median	RMSE	Std	Mean	Median	RMSE	Std
$\alpha = 2,\ d = 0.1$	500	1.042	1.009	0.232	0.229	1.021	1.006	0.132	0.130	1.114	1.006	0.478	0.465
	1000	1.024	1.000	0.188	0.186	1.031	1.030	0.112	0.108	1.090	1.036	0.348	0.337
	10000	1.007	0.999	0.101	0.101	1.004	1.001	0.063	0.063	1.022	1.000	0.156	0.155
$\alpha = 2,\ d = 0.4$	500	1.034	1.004	0.233	0.231	1.032	1.022	0.155	0.152	1.131	1.023	0.472	0.454
	1000	1.025	1.004	0.189	0.187	1.040	1.035	0.132	0.126	1.097	1.045	0.360	0.346
	10000	1.008	1.001	0.101	0.101	1.014	1.008	0.086	0.085	1.030	1.006	0.169	0.166
$\alpha = 4,\ d = 0.1$	500	1.045	1.020	0.233	0.229	1.020	1.014	0.067	0.064	1.028	1.011	0.130	0.127
	1000	1.025	1.002	0.192	0.191	1.023	1.021	0.061	0.057	1.033	1.020	0.116	0.111
	10000	1.015	1.009	0.104	0.103	1.017	1.016	0.035	0.031	1.017	1.014	0.059	0.057
$\alpha = 4,\ d = 0.4$	500	1.041	1.002	0.232	0.229	1.030	1.026	0.106	0.102	1.040	1.023	0.157	0.152
	1000	1.027	1.005	0.193	0.191	1.039	1.034	0.100	0.092	1.048	1.034	0.140	0.131
	10000	1.016	1.011	0.103	0.102	1.032	1.028	0.073	0.066	1.032	1.027	0.087	0.081

Table 4
Monte Carlo Results - $|t(\alpha)|$.

Pareto	n	Tail Index				VaR				ES			
		Mean	Median	RMSE	Std	Mean	Median	RMSE	Std	Mean	Median	RMSE	Std
$\alpha = 2,\ d = 0.1$	500	1.016	0.971	0.250	0.250	0.983	0.974	0.129	0.127	1.124	1.004	0.600	0.587
	1000	1.007	0.988	0.182	0.182	0.990	0.989	0.110	0.109	1.069	0.995	0.352	0.345
	10000	1.006	0.997	0.098	0.098	1.002	1.000	0.060	0.060	1.019	1.003	0.149	0.148
$\alpha = 2,\ d = 0.4$	500	1.012	0.975	0.228	0.228	0.990	0.979	0.151	0.151	1.121	1.016	0.535	0.522
	1000	1.004	0.980	0.182	0.182	1.001	0.992	0.133	0.133	1.078	1.021	0.335	0.326
	10000	1.005	1.001	0.097	0.097	1.011	1.007	0.083	0.083	1.028	1.016	0.160	0.158
$\alpha = 4,\ d = 0.1$	**500**	**0.833**	**0.814**	**0.239**	**0.170**	**1.399**	**1.396**	**0.414**	**0.111**	**1.543**	**1.519**	**0.590**	**0.229**
	1000	**0.863**	**0.848**	**0.208**	**0.157**	**1.425**	**1.425**	**0.435**	**0.092**	**1.536**	**1.515**	**0.567**	**0.186**
	10000	**0.914**	**0.909**	**0.125**	**0.091**	**1.505**	**1.506**	**0.507**	**0.051**	**1.562**	**1.560**	**0.571**	**0.097**
$\alpha = 4,\ d = 0.4$	**500**	**0.832**	**0.815**	**0.243**	**0.176**	**1.412**	**1.407**	**0.442**	**0.159**	**1.562**	**1.530**	**0.621**	**0.266**
	1000	**0.860**	**0.839**	**0.205**	**0.149**	**1.442**	**1.434**	**0.462**	**0.133**	**1.555**	**1.541**	**0.593**	**0.210**
	10000	**0.912**	**0.903**	**0.126**	**0.090**	**1.524**	**1.522**	**0.534**	**0.101**	**1.584**	**1.580**	**0.599**	**0.134**

Table 5
Theoretical Standard Deviation: $k_n = \sqrt{n}$.

	Tail Index	VaR	ES
$\alpha = 2,\ d = 0.1$	$\frac{1}{n^{1/4}}$	$\frac{(1+\ln^2 2)^{1/2}}{2n^{1/4}}$	$\frac{(1+(2+\ln 2)^2)^{1/2}}{2n^{1/4}}$
$\alpha = 2,\ d = 0.4$	$\frac{1}{n^{1/4}}$	$\frac{1}{10n^{0.1}}$	$\frac{1}{10n^{0.1}}$
$\alpha = 4,\ d = 0.1$	$\frac{1}{n^{1/4}}$	$\frac{(1+\ln^2 2)^{1/2}}{4n^{1/4}}$	$\frac{(1+(\frac{4}{3}+\ln 2)^2)^{1/2}}{4n^{1/4}}$
$\alpha = 4,\ d = 0.4$	$\frac{1}{n^{1/4}}$	$\frac{1}{10n^{0.1}}$	$\frac{1}{10n^{0.1}}$

- The mean of each simulated ratio illustrates bias; *with the exception of $t(4)$ noise, in each case it should converge to 1 as $n \to \infty$.*
- The median is indicative of the presence of skewness. If the mean is less (respectively, more) than the median, the data is skewed to the left (respectively, right). Since all limiting distributions are symmetric, any indication of skewness should decrease as n increases.
- Standard deviation reflects variability of the estimators, and should decrease as n increases.
- The RMSE combines both bias and variability.

What does the theory predict? As noted above, whenever (3) is satisfied, there is no asymptotic bias or skewness and the RMSE should reflect primarily variability. As a basis of comparison with the empirical results in Tables 2 to 4, the nominal values of the standard deviation for each ratio, sample size and model appear in Theorems 1 below. These are based on the asymptotic variances in Theorems 5.1 and 5.3 and are calculated from the general formulas found in Section 8.2, Table 5.

What is observed in the simulations? To address this question, the models satisfying (3) are first considered and the following comments apply to:

All models except for $|t(4)|$ noise:

In all the models satisfying (3), there is little evidence of bias or skewness in the estimators of α, $Q_X(1/p)$ and $\theta_X(p)$, and so in all cases the standard deviation and RMSE are similar. The following points focus on comparisons of the sample standard deviation and RMSE with the theoretical values in Tabel 1.

Tail Index: As predicted by the theory, in all models the empirical results for the tail index are roughly the same for $d = 0.1$ and $d = 0.4$, illustrating that long memory has no effect on estimation of α. In all cases, the sample standard deviation and RMSE are close to the predicted value of $n^{-1/4}$.

Value-at-Risk:

- Short memory ($d = 0.1$): In all cases, the standard deviation and RMSE are close to the predicted values in Table 1.
- Long memory ($d = 0.4$): In contrast to the short memory case, the sample standard deviation is consistently greater than the nominal value in Table 1, although the disparity decreases with increasing n and is less marked when $\alpha = 4$. This illustrates the slower rate of convergence in the presence of long memory.

Expected Shortfall:

- Short memory ($d = 0.1$): The standard deviations are slightly larger than the predicted values in Table 1, but converge as expected as n increases. The discrepancies are smaller when $\alpha = 4$.
- Long memory ($d = 0.4$): When long memory prevails, although the asymptotic distributions of VaR and ES are the same, the empirical standard deviations differ. In finite samples this is unsurprising and reflects the decomposition in (27). In all cases, this disparity decreases as n increases and is less when $\alpha = 4$.

$|t(4)|$ Noise:

In the case of $t(4)$ noise, the simulations clearly reflect the effect of bias in the model. While there is little evidence of skewness, the ratios for the tail index and risk measures do not converge to 1. The variability of the estimates decreases as n increases, but RMSE remains high due to bias. This is due to the choice $k_n = \sqrt{n}$, as shown in Section 8.2.

The simulations demonstrate the theoretical results. While long memory does not play any role in the limiting behaviour of the Hill estimator, it does influence the asymptotics of the estimators of Value-At-Risk and Expected Shortfall.

7. Conclusion

7.1. Summary

The object has been estimation of extreme risk measures for returns of an asset in a portfolio that exhibit heavy tails and long memory. Estimators for both Value-at-Risk and Expected Shortfall have been derived. The estimators of Value-at-Risk are functions of both intermediate order statistics and Harmonic Moment Estimators (HME) of the index of regular variation. Their asymptotic behaviour is established in Theorems 5.1 and 5.2 using the weak convergence of intermediate order statistics and asymptotic normality of HMEs. In turn, the estimators of Expected Shortfall are functions of estimators of VaR and HMEs. This in turn leads to the limiting behaviour of the estimators of ES, which is investigated in Theorems 5.3 and 5.4.

In summary, when returns are assumed to exhibit heavy tails, long memory and leverage, estimators of Value-at-Risk and Expected Shortfall are not affected by the leverage effect. However, heavy tails and long memory do influence their estimators in a nontrivial and dichotomous way.

7.2. Directions for Future Research

There are several interesting questions that are beyond the scope of this article.

- A serious issue is bias. From the theoretical perspective, bias can be controlled via the condition (3). From a practical perspective, in the context of dependent data, to the best of our knowledge, there is no satisfactory way of systematically dealing with bias (see de Haan et al. (2016) for an approach based on estimation of the second order parameter). The use of HMEs other than the Hill estimator might prove to be more efficient. Further, it would be of interest to explore the suitability of other estimators of the tail index for the LMSV model.
- A promising approach to bias is based on bootstrap techniques. In Ivanoff et al. (2021) the authors prove validity of the nonparametric bootstrap for the Hill estimator in the case of regularly varying, i.i.d. random variables. Bias can be reduced through using a bootstrapped variance estimator of α.
- More generally, both the issue of the dichotomous limiting behaviour and bias for the LMSV models could be potentially addressed through bootstrap techniques. This is currently under investigation.
- A multivariate LMSV model should be developed, in particular for the representation of multiple stocks making up an individual portfolio.
- Backtesting: Large financial institutions may be required to backtest certain risk models, according to Basel financial regulations. Backtesting uses large historical data sets to estimate and compare the performance of risk modelling strategies. This is an important question, but well beyond the scope of the present article.

8. Proofs

This section begins with the definition of a Hermite-Rosenblatt process.

Definition 8.1 ((Beran et al., 2013, p.194)). Let $B(\cdot)$ denote a standard Brownian motion on $\mathbb{R}$, m a positive integer and $h \in \mathbb{R}_+$ be such that $1 - 1/2m < h < 1$. A stochastic process $\{\xi_{m,h}(u) : u \geqslant 0\}$ defined by

$$\xi_{m,h}(u) = \omega(m,h) \int_{-\infty}^{\infty} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{m-1}} \left(\int_0^u \prod_{j=1}^{m} (s - x_i)_+^{h-\frac{3}{2}} ds \right) dB(x_m) \cdots dB(x_1),$$

where $\omega(m,h) > 0$ satisfies

$$\omega^2(m,h) = \frac{m!(2m(h-1)+1)(m(h-1)+1)}{\left(\int_0^{\infty} [x(x+1)]^{h-\frac{3}{2}} dx\right)^m}$$

is called a **Hermite-Rosenblatt process**. Note that $x_+ := \max(0, x)$.

Note: The constant $\omega(m,h)$ ensures that $E\left(\xi_{m,h}^2(1)\right) = 1$.

Proof. *Proof of Theorem 5.1*. To begin, return to Corollary 4.1, and let $\nu_n := k/(np)$ and $\tilde{\nu} = \lim \nu_n$. (Note that $\tilde{\nu} = \nu$ in Cases 1 and 2 and $\tilde{\nu} = \infty$ in Cases 3 and 4.) It is now immediate that if $\tilde{\nu} \neq 1$ then the following hold for $\{X_j\}$ the long memory stochastic volatility model given in (1). If $\alpha > 2(1-r)$, then

- If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to 0$, or $m(1-2d) > 1$, then as $n \to \infty$,

$$\frac{\sqrt{k}}{\ln \nu_n} \left(\ln \nu_n (\widehat{\gamma}_{r,k} - \gamma) \, , \, \left(\frac{X_{(n-k)}}{u_n} - 1 \right) \right) \xrightarrow[n\to\infty]{d} \left(\frac{(\alpha + r - 1)}{\left(\alpha^3(\alpha + 2r - 2)\right)^{1/2}} \mathcal{N} \, , \, \frac{1}{\ln \tilde{\nu} \, \alpha} \tilde{\mathcal{N}} \right), \tag{30}$$

  where $\mathcal{N}$ and $\tilde{\mathcal{N}}$ are independent standard normal random variables.
- If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to \infty$ then as $n \to \infty$,

$$\frac{n}{b_{n,m} \ln \nu_n} \left(\ln \nu_n (\widehat{\gamma}_{r,k} - \gamma) \, , \, \left(\frac{X_{(n-k)}}{u_n} - 1 \right) \right) \xrightarrow[n\to\infty]{d} \left(0, \, \frac{\mu_{\phi,\alpha}(m)}{\ln \tilde{\nu} \, \alpha m! E(\phi^{\alpha}(Y))} \xi_{m,d+1/2}(1) \right). \tag{31}$$

In what follows, for simplicity write $\widehat{\gamma}$ for $\widehat{\gamma}_{r,k}$. Since $k_n = n\bar{F}_n(u_n)$, then $Q_X(n/k) = u_n$. Now let c_n denote the normalizing multiplier in (30) or (31) and observe that in each case $c_n < a_{n,m}/\ln \nu_n$ and either $c_n = \sqrt{k}/\ln \nu_n$ or $c_n \ln \nu_n/\sqrt{k} \to 0$. The following decomposition holds:

$$\begin{aligned} c_n \left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1 \right) &= c_n \left(\frac{X_{n:n-k}}{u_n} \left(\frac{k}{np} \right)^{\widehat{\gamma}} \frac{Q_X(n/k)}{Q_X(1/p)} - 1 \right) \\ &= c_n \left(\frac{X_{n:n-k}}{u_n} \left(\frac{k}{np} \right)^{(\widehat{\gamma}-\gamma)} - 1 \right) \frac{Q_X(n/k)}{Q_X(1/p)} \left(\frac{k}{np} \right)^{\gamma} + c_n \left(\frac{Q_X(n/k)}{Q_X(1/p)} \left(\frac{k}{np} \right)^{\gamma} - 1 \right) \\ &= c_n \left(\frac{X_{n:n-k}}{u_n} e^{\ln \nu_n (\widehat{\gamma}-\gamma)} - 1 \right) (1 + o(1)) + o(1), \end{aligned}$$

by the no-bias condition (14).

Thus, in each case the asymptotic behaviour of $c_n\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1 \right)$ is the same as that of $c_n\left(\frac{X_{n:n-k}}{u_n} e^{\ln \nu_n (\widehat{\gamma}-\gamma)} - 1 \right)$. All of the cases now follow in a routine manner from (30) and (31) by applying the delta method to the map $g(u,v) = ue^v$, with Hadamard derivative at (1,0) equal to $g'_{(1,0)}(a,b) = a + b$, letting $u = X_{n:n-k}/u_n$ and $v = \ln \nu_n(\widehat{\gamma} - \gamma)$.

To be precise, the representations in Cases 1 and 2 ($-\infty < \tilde{\nu} = \nu < \infty$, $\nu \neq 1$) are obtained by noting that $\ln \nu/\ln \nu_n \to 1$ and multiplying (30) and (31), respectively, by $\ln \nu$. Cases 3 and 4 follow directly from (30) and (31), observing that $1/\ln \tilde{\nu} = 0$.

It remains to consider the case $\tilde{\nu} = 1$. If $\alpha > 2(1-r)$, then

- If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to 0$, or $m(1-2d) > 1$, then as $n \to \infty$,

$$\sqrt{k} \left(\ln \nu_n (\widehat{\gamma}_{r,k} - \gamma) \, , \, \left(\frac{X_{(n-k)}}{u_n} - 1 \right) \right) \xrightarrow[n\to\infty]{d} \left(0 \, , \, \frac{1}{\alpha} \mathcal{N} \right), \tag{33}$$

  where $\mathcal{N}$ is a standard normal random variable.
- If $m(1-2d) < 1$ and $\frac{b_{n,m}}{n}\sqrt{k} \to \infty$ then as $n \to \infty$,

$$\frac{n}{b_{n,m}} \left(\ln \nu_n (\widehat{\gamma}_{r,k} - \gamma) \, , \, \left(\frac{X_{(n-k)}}{u_n} - 1 \right) \right) \xrightarrow[n\to\infty]{d} \left(0, \, \frac{\mu_{\phi,\alpha}(m)}{\alpha m! E(\phi^{\alpha}(Y))} \xi_{m,d+1/2}(1) \right). \tag{34}$$

Exactly as above, use the decomposition in (32) with the normalizing multiplier c_n as in (33) or (34), and apply the delta method to obtain Cases 1 and 2 for $\nu = 1$. □

Proof of Theorem 5.2. If $m(1-2d) < 1$, $k/np = \nu_n \to \infty$ and $\frac{b_{n,m}}{n}\sqrt{k}/\ln \nu_n \to \infty$ then as $n \to \infty$,

$$\frac{n}{b_{n,m}}\left(e^{\ln \nu_n(\widehat{\gamma}-\gamma)}, \left(\frac{X_{(n-k)}}{u_n} - 1\right)\right) \xrightarrow[n\to\infty]{d} \left(0, \frac{\mu_{\phi,\alpha}(m)}{\alpha m! E(\phi^\alpha(Y))}\xi_{m,d+1/2}(1)\right). \tag{35}$$

As in the proof of Theorem 5.1, the following decomposition holds:

$$\frac{n}{b_{n,m}}\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right) = \frac{n}{b_{n,m}}\left(\frac{X_{n:n-k}}{u_n}e^{\ln \nu_n(\widehat{\gamma}-\gamma)} - 1\right)(1+o(1)) + o(1).$$

The result follows from (35) by an application of the delta method exactly as in the preceding proof. □

Proof of 5.3. Denote $\vartheta = 1/(1-\gamma) = \alpha/(\alpha-1)$ and $\widehat{\vartheta} = \widehat{\vartheta}_{r,k} = 1/(1-\widehat{\gamma}_{r,k}) = 1/(1-\widehat{\gamma})$. It is a straightforward consequence of Theorem 4.2 and the delta method that

$$\sqrt{k}\left(\widehat{\gamma} - \gamma, \frac{\widehat{\vartheta}}{\vartheta} - 1\right) \xrightarrow[n\to\infty]{d} \left(1, \frac{\alpha}{\alpha-1}\right)\frac{(\alpha+r-1)}{\sqrt{\alpha^3(\alpha+2r-2)}}\mathcal{N}. \tag{36}$$

Next, letting c_n denote the normalizing multiplier appearing in (23)-(26),

$$\begin{aligned} c_n\left(\frac{\widehat{\theta}(p)}{\theta(p)} - 1\right) &= c_n\left(\frac{\widehat{\vartheta}}{\vartheta}\cdot\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right)\frac{\vartheta Q_X(1/p)}{\theta(p)} \\ &\quad + c_n\left(\frac{\vartheta\, Q_X(1/p)}{\theta(p)} - 1\right), \end{aligned}$$

and so by the no-bias condition (22),

$$c_n\left(\frac{\widehat{\theta}(p)}{\theta(p)} - 1\right) \overset{p}{\sim} c_n\left(\frac{\widehat{\vartheta}}{\vartheta}\cdot\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right). \tag{37}$$

In each case, apply the delta method to the product map $g(u,v) = uv$, with Hadamard derivative at (1,1) equal to $g'_{(1,1)}(a,b) = a+b$. Letting $u = \widehat{\vartheta}/\vartheta$ and $v = \widehat{Q}_X(1/p)/Q_X(1/p)$ yields the representations in (23)-(26) from (36) and (15)-(18), respectively. □

Proof of 5.4. Since

$$\begin{aligned} \frac{n}{b_{n,m}}\left(\frac{\widehat{\theta}_X(p)}{\theta_X(p)} - 1\right) &= \frac{n}{b_{n,m}}\left(\frac{\widehat{Q}_X(1/p)}{Q_X(1/p)} - 1\right) \\ &\quad + \frac{1}{(1-\widehat{\gamma})}\frac{n}{b_{n,m}\sqrt{k}}\sqrt{k}(\widehat{\gamma}-\gamma) + \frac{n}{b_{n,m}}\left(\frac{Q_X(1/p)}{(1-\gamma)\theta_X(p)} - 1\right), \end{aligned}$$

then Theorems 5.2 and 4.2 and Lemma 8.5 yield (28). □

8.1. Regular and Second-Order Regular Variation

This subsection starts with two classical results.

Lemma 8.1 (Breiman's Lemma (Soulier, 2009, p.49)). *Let V and W be two independent non-negative random variables such that $\bar{F}_V$ is regularly varying with index $-\gamma$. If there exists $\epsilon > 0$ such that $E(W^{\gamma+\epsilon}) < \infty$, then*

$$\frac{P(VW > x)}{P(V > x)} \underset{x\to\infty}{\rightarrow} E(W^\gamma). \tag{38}$$

Lemma 8.2 (Potter's bound (Bingham et al., 1989, p. 25)). *For all $C > 1$, $\epsilon > 0$, there exists $\delta = \delta(C;\epsilon) \geqslant 0$ such that for $x \geqslant \delta$, $t > 0$,*

$$T_x(t) \leqslant C\left(t^{-(\alpha+\epsilon)} \vee t^{-(\alpha-\epsilon)}\right). \tag{39}$$

The following result is a re-statement of Theorem 2.3.9 in de Haan and Ferreira (2006).

Lemma 8.3. *Let U_W be a real-valued function such that*

$$\frac{U_W(xt)/U_W(x) - t^{1/\alpha}}{\eta^\dagger(x)} \underset{x\to\infty}{\rightarrow} t^{1/\alpha}\frac{1-t^{-\rho^\dagger}}{\rho^\dagger}. \tag{40}$$

Then for all $\epsilon, \delta > 0$ *there exists* $x_0 = x_0(\epsilon, \delta) > 1$ *such that for all* $x > x_0$, $t > x_0/x$,

$$\left| \frac{U_W(xt)/U_W(x) - t^{1/\alpha}}{D_0(x)} - t^{1/\alpha}\frac{1-t^{-\rho^\dagger}}{\rho^\dagger} \right| \leqslant \epsilon t^{1/\alpha-\rho^\dagger} \max\left(t^{\delta}, t^{-\delta}\right),$$

where

$$D_0(x) = \rho^\dagger \left(\frac{x^{1/\alpha}}{U_W(x)} \lim_{s\to\infty} \frac{U_W(s)}{s^{1/\alpha}} - 1 \right).$$

By (Kulik and Soulier, 2011, p. 117) the second-order regular variation of $\bar{F}_Z$ assumed in A(ii) implies second order-regular variation of $\bar{F}_X$ (see also Lemma 3.3.7 in Bilayi-Biakana (2019)). This, in addition to Remark 2.4.18 and Lemma 2.4.23 in Bilayi-Biakana (2019), implies that (40) holds for Q_X with $\rho^\dagger = \kappa/\alpha$ and $\eta^\dagger(x) = \tilde{\eta}(\bar{F}_X^{\leftarrow}(x))/\alpha^2$.

In the following two lemmas, it is seen that (3) implies the no-bias conditions (14) and (22).

Lemma 8.4. *Let* $\{X_j\}$ *be the long memory stochastic volatility model as in (1). If* $k/(np)$ *is bounded away from 0, then*

$$\limsup_{n\to\infty} a_{n,m}\left(\left(\frac{k}{np}\right)^{-\gamma} \frac{Q_X(1/p)}{Q_X(n/k)} - 1 \right) = 0. \tag{41}$$

Proof. Proof of Lemma 8.4. Applying Lemma 8.3 with Q_X, $x = n/k$, $t = k/(np)$ ensures that for $n/k > x_0$, $1/p > x_0$,

$$\left| a_{n,m}\frac{Q_X(1/p)/Q_X(n/k) - (k/(np))^{\gamma}}{a_{n,m}D_0(n/k)} - \left(\frac{k}{np}\right)^{\gamma} \frac{1-(k/np)^{-\rho^\dagger}}{\rho^\dagger} \right|$$
$$\leqslant \epsilon \left(\frac{k}{np}\right)^{\gamma-\rho^\dagger} \left(\left(\frac{k}{np}\right)^{\delta} \vee \left(\frac{k}{np}\right)^{-\delta} \right), \tag{42}$$

where

$$D_0(x) = \rho^\dagger \left(\frac{x^{\gamma}}{Q_X(x)} \lim_{s\to\infty} \frac{Q_X(s)}{s^{\gamma}} - 1 \right).$$

In the notation used here, the first conclusion of Theorem 2.3.6 in de Haan and Ferreira (2006) is that $D_0(n/k) \sim \eta^\dagger(n/k)/\rho = \tilde{\eta}(\bar{F}_X^{\leftarrow}(n/k))/(\rho\alpha^2)$. From (5) it follows that

$$a_{n,m}D_0(n/k) = O(a_{n,m}\eta^*(\bar{F}_X^{\leftarrow}(n/k))$$

and so by (3), $a_{n,m}D_0(n/k) \to 0$ as $n \to \infty$.

Choose $\delta < \rho^\dagger$. Divide both sides of (42) by $(k/(np))^{\gamma}$ to get

$$\left| \frac{a_{n,m}\left((k/(np))^{-\gamma}Q_X(1/p)/Q_X(n/k) - 1\right)}{a_{n,m}D_0(n/k)} - \frac{1-(k/np)^{-\rho^\dagger}}{\rho^\dagger} \right|$$
$$\leqslant \epsilon \left(\frac{k}{np}\right)^{-\rho^\dagger} \left(\left(\frac{k}{np}\right)^{\delta} \vee \left(\frac{k}{np}\right)^{-\delta} \right).$$

Since k/np is bounded away from 0, there exists $C = C(\delta)$ such that for any ϵ, the right hand side above is bounded by $C\epsilon$ for n/k and $1/p$ sufficiently large. Now consider the left hand side. Since the second term remains bounded as $n \to \infty$, so does the first term. But since $a_{n,m}D_0(n/k) \to 0$, then the numerator must satisfy (41). This completes the proof. □

Lemma 8.5. *Let* $\{X_j\}$ *be the long memory stochastic volatility model as in (1) and assume that* $\alpha > 1$. *If* k/np *is bounded away from 0, then*

$$\lim_{n\to\infty} a_{n,m}\left(\frac{(1-\gamma)\theta_X(p)}{Q_X(1/p)} - 1 \right) = 0. \tag{43}$$

Proof. *Proof of Lemma 8.5.* With T_u defined in (6), the following holds:

$$\frac{1}{u}E\left(X|X > u\right) = \int_1^\infty xT_u(dx).$$

Taking $u = Q_X(1/p)$ yields

$$\frac{(1-\gamma)\theta_X(p)}{Q_X(1/p)} = (1-\gamma)\int_1^\infty xT_{Q_X(1/p)}(dx).$$

Note that if X is Pareto then $T_u(t) = t^{-\alpha} = T(t)$ for all u whenever $u, ut > 1$. Since $p \to 0$, it can be assumed that $Q_X(1/p) > 1$. Thus, in the Pareto case

$$(1-\gamma)\int_1^\infty x T_{Q_X(1/p)}(dx) = (1-\gamma)\alpha\int_1^\infty x x^{-\alpha-1}dx = \frac{(1-\gamma)\alpha}{\alpha-1} = 1\ .$$

Hence, in the Pareto case there is no bias. Also, note that since $\alpha > 1$, then $\lim_{x\to\infty} xT(x) = \lim_{x\to\infty} xT_{Q_X(1/p)}(x) = 0$. The first statement is obvious, while the latter follows from Potter's bounds (see Lemma 8.2). Thus, in the general case, integration by parts gives

$$\left|(1-\gamma)\int_1^\infty x T_{Q_X(1/p)}(dx) - (1-\gamma)\int_1^\infty x T(dx)\right|$$
$$\leqslant (1-\gamma)\left|T_{Q_X(1/p)}(1) - T(1)\right| + (1-\gamma)\int_1^\infty |T_{Q_X(1/p)}(x) - T(x)|dx.$$

Recall (cf. (Bilayi-Biakana et al., 2019, Eq. (29b))) that the following bound holds under the second-order regular variation assumption: for arbitrary $\epsilon > 0$,

$$|T_u(t) - T(t)| \leqslant C\left(t^{-(\alpha+\kappa)+\epsilon} + t^{-(\alpha+\kappa)-\epsilon}\right)\tilde{\eta}(u)\ . \tag{44}$$

Next, apply this with $u = Q_X(1/p)$ and $t = 1$. Furthermore, use the asymptotic equivalence between $\tilde{\eta}$ and η^* stated in (5). This yields

$$a_{n,m}\left|T_{Q_X(1/p)}(1) - T(1)\right| = O(a_{n,m})\frac{\tilde{\eta}(Q_X(1/p))}{\tilde{\eta}(Q_X(n/k))}\tilde{\eta}(Q_X(n/k))$$
$$= O\left(a_{n,m}\eta^*(F_X^{\leftarrow}(k/n))\right)\frac{\tilde{\eta}(Q_X(1/p))}{\tilde{\eta}(Q_X(n/k))}.$$

Moreover, regular variation of $\tilde{\eta}$ and Q_X implies regular variation of the composition with index $\kappa\gamma$. Therefore, for arbitrary $\epsilon > 0$, Potter's bounds give

$$a_{n,m}\left|T_{Q_X(1/p)}(1) - T(1)\right| = O\left(a_{n,m}\eta^*(F_X^{\leftarrow}(k/n))\left(\frac{k}{np}\right)^{-\kappa\gamma\pm\epsilon}\right).$$

Thus, whenever $k/(np)$ is bounded away from zero, from (3) it follows that

$$a_{n,m}\left|T_{Q_X(1/p)}(1) - T(1)\right| = o(1).$$

Likewise, application of (44) gives

$$a_{n,m}\int_1^\infty |T_{Q_X(1/p)}(x) - T(x)|dx$$
$$= O(a_{n,m}\tilde{\eta}(Q_X(1/p)))\int_1^\infty \left(x^{-(\alpha+\kappa)+\epsilon} + x^{-(\alpha+\kappa)-\epsilon}\right)dx.$$

This shows $O(a_{n,m}\tilde{\eta}(Q_X(1/p))) = o(1)$, while the integral is finite since $\epsilon > 0$ is arbitrary. □

8.2. Simulations: Technical Details

Consider the long memory stochastic volatility model

$$X_j = \exp(\sigma Y_j)Z_j, \quad j = 1, \ldots, n \tag{45}$$

where Y_j is a fractional Gaussian noise sequence and $\ell_Y(j) \sim d(2d+1)$. The noise Z_j is i.i.d. second order regular varying with parameters $-\alpha$, $-\kappa$ and rate function η^*.

Remark 8.1. The Hermite rank of $\phi^\alpha(\cdot) = \exp(\alpha\sigma\cdot)$ is $m = 1$. Therefore:

- The process $\xi_{1,d+1/2}(\cdot)$ is fractional Brownian motion and $\xi_{1,d+1/2}(1)$ is a standard normal random variable;
- $E\left(\phi^\alpha(Y)\right) = E(\exp(\alpha\sigma Y)) = \exp((\alpha\sigma)^2/2)$;
- $\mu_{\phi,\alpha}(1) = E(Y\phi^\alpha(Y)) = E(Y\exp(\alpha\sigma Y)) = \alpha\sigma\exp((\alpha\sigma)^2/2)$.

Therefore, if $k_n = \sqrt{n}$, the choice of the various parameters yields the following:

- $b_{n,m} = b_{n,1} = n^{.5+d}\sqrt{\ell_Y(n)/(d(2d+1))} \sim n^{.5+d}$;
  $a_{n,m} = a_{n,1} = \left(n^{.25} + n^{.5-d}\sqrt{(d(2d+1))/\ell_Y(n)}\right) \sim n^{.25} + n^{.5-d}$.
- $b_{n,m}\sqrt{k}/n = b_{n,1}/n^{.75} \sim n^{d-.25}$.

Recall that $m = 1$, $\nu = 2$, and $\sigma = 0.1$ and the estimator of α^{-1} is the HME of order $r = 1$. The resulting asymptotic variances in Theorems 5.1 and 5.3 are:

$$\frac{1}{\alpha^2} + \frac{(\alpha + r - 1)^2 \ln^2 \nu}{\alpha^3(\alpha + 2r - 2)} = \frac{1 + \ln^2 2}{\alpha^2} \text{ cf. (15)} \tag{46}$$

$$\frac{1}{\alpha^2} + \frac{(\alpha + r - 1)^2}{\alpha^3(\alpha + 2r - 2)}\left(\frac{\alpha}{\alpha - 1} + \ln \nu\right)^2 = \frac{1 + \left(\frac{\alpha}{\alpha-1} + \ln 2\right)^2}{\alpha^2} \text{ cf. (23)} \tag{47}$$

$$E\left(\left(\frac{\mu_{\phi,\alpha}(1)\xi_{1,d+1/2}(1)}{\alpha E(\phi^\alpha(Y))}\right)^2\right) = \sigma^2 = .01 \text{ cf. (16), (24).} \tag{48}$$

In turn, since

$$k^{-1/2} = n^{-1/4} \text{ and } \frac{b_{n,1}}{n} \sim n^{-(.5-d)},$$

the corresponding theoretical sample standard deviations of the ratios for the various parameters (tail index, VaR, ES) and scenarios ($\alpha = 2,\ 4$; $d = 0.1,\ 0.4$) for samples of size n are given in Table 5, noting that (46) and (47) apply to the short memory case ($d = 0.1$), and (48) applies to the long memory case ($d = 0.4$).

Turning now to the noise variables Z_j, the following observations are made about the effect of the choice of k_n on the no-bias assumption (3) in the case of each of the three distributions:

- Pareto: There is no bias, and Theorem 5.1 and 5.3 hold for any sequence $k_n = o(n)$.
- Mixed Pareto: It is shown in Example 5.1 of Bilayi-Biakana et al. (2019) that for $\kappa = \alpha$, the following restrictions on the choice of k ensure that (3) is satisfied:

  $$k = o\left(n^{2/3}\right), \quad k = o\left(n^{.5+d}\right).$$

  The sequence $k = k_n = \sqrt{n}$ satisfies both restrictions for any $d \in (0, 1/2)$.
- $|t(\alpha)|$: Let c denote a non-negative generic constant that can differ at each appearance. From the discussion in Example 5.1 of Bilayi-Biakana et al. (2019), it follows that $\bar{F}_X^{-1}(y) \sim cy^{-1/\alpha}$ as $y \to 0$ and that (3) is satisfied if

  $$\left(\sqrt{k} + n^{(.5-d)}\right)\eta^*\left(\bar{F}_X^{-1}(k/n)\right) \longrightarrow 0. \tag{49}$$

  For Z the absolute value of a $t(\alpha)$ random variable, refer to Example 3 and Proposition 6 of Hua and Joe (2011) from which it may be concluded that η^* is regularly varying with index -2 and $\eta^*(x) \sim cx^{-2}$ as $x \to \infty$. Thus, (49) is equivalent to

  $$\left(\sqrt{k} + n^{(.5-d)}\right)\left(\frac{k}{n}\right)^{2/\alpha} \longrightarrow 0.$$

  For $k = \sqrt{n}$ and $d = 0.1$ or 0.4, it is easily seen that (49) holds if $\alpha = 2$ but not if $\alpha = 4$.

One final technical detail must be addressed. The value of $\alpha = 1/\gamma$ (2 or 4) is known for all the models and $\theta_X(p) \sim \frac{1}{1-\gamma} Q_X(1/p)$. However, $Q_X(1/p) = F_X^{\leftarrow}(1-p) = \bar{F}_X^{\leftarrow}(p)$ must be determined in each case for $p = k/2n = 1/2\sqrt{n}$ ($n = 500,\ 1,000,\ 10,000$). From the argument that follows, it is observed that as $p \to 0$,

$$\bar{F}_X^{\leftarrow}(p) \approx \bar{F}_Z^{\leftarrow}(p), \tag{50}$$

and so it suffices to use the upper quantiles of the noise variable Z. In the case of Pareto or $|t(\alpha)|$ noise, quantiles are readily available in R. For the two mixture distributions, the quantiles were estimated empirically from 100,000 simulated observations of Z.

To justify (50), appeal to Lemma 8.1, recalling that $Y \sim \mathcal{N}$ and $\sigma = 0.1$. Therefore, as $x \to \infty$,

$$\bar{F}_X(x) \sim E\left(e^{0.1\alpha Y}\right)\bar{F}_Z(x) = e^{(.01\alpha^2/2)}\bar{F}_Z(x).$$

From this, it follows that $\bar{F}_X^{\leftarrow}(p) \approx \bar{F}_Z^{\leftarrow}(p/c_\alpha)$, where $c_\alpha = \exp(.01\alpha^2/2)$. Since $c_2 = 1.02$ and $c_4 = 1.08$, (50) follows.

Acknowledgement

Research of all authors was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

Acerbi, C., Tasche, D., 2002. Expected shortfall: A natural coherent alternative to value at risk. Economic Notes 31 (2), 379–388.
Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26 (7), 1487–1503.
Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9 (3), 203–228.
Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long Memory Processes: Probabilistic Properties and Statistical Methods. Springer.
Beran, J., Schell, D., Stehlik, M., 2014. The harmonic moment tail index estimator: asymptotic distribution and robustness. Ann.Inst.Stat.Math. 66, 93–220.
Bilayi-Biakana, C., 2019. Regularly varying time series with long memory: Probabilistic properties and estimation. PhD thesis, University of Ottawa.
Bilayi-Biakana, C., Ivanoff, G., Kulik, R., 2019. The tail empirical process for long memory stochastic volatility models with leverage. Electron. J. Statist. 13 (2), 3453–3484.
Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. Regular variation, first Press Syndicate of the University of Cambridge.
Bingham, N.H., Goldie, C.M., Teugels, J.L., 1989. Regular variation. Cambridge University Press, Cambridge.
Brazauskas, V., Jones, B.L., Puri, M.L., Zitikis, R., 2008. Estimating conditional tail expectation with actuarial applications in view. Journal of Statistical Planning and Inference 138 (11), 3590–3604.
Breidt, F.J., Crato, N., de Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83 (1-2), 325–348.
Chen, S.X., 2005. Nonparametric inference of value-at-risk for dependent financial returns. Journal of Financial Econometrics 3 (2), 227–255.
Chen, S.X., 2008. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6 (1), 87–107.
Deo, R., Hsieh, M.-C., Hurvich, C.M., Soulier, P., 2006. Long memory in nonlinear processes. In: Dependence in probability and statistics. In: Lecture Notes in Statist., Vol. 187. Springer, New York, pp. 221–244.
Deo, R., Hurvich, C.M., Soulier, P., Wang, Y., 2009. Conditions for the propagation of memory parameter from durations to counts and realized volatility. Econometric Theory 25 (3), 764–792.
Drees, H., 1998. Optimal rates of convergence for estimates of the extreme value index. The Annals of Statistics 26 (1), 434–448.
Drees, H., 2000. Weighted approximations of tail processes for β-mixing random variables. The Annals of Applied Probability 10 (4), 1274–1301.
de Haan, L., Ferreira, A., 2006. Extreme value theory. Springer, New York. An introduction
de Haan, L., Mercadier, C., Zhou, C., 2016. Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. Finance and Stochastics 20(2), 321–354.
Hoga, Y., 2019. Extreme conditional tail moment estimation under serial dependence. Journal of Financial Econometrics 17, 587–615.
Hua, L., Joe, H., 2011. Second order regular variation and conditional tail expectation of multiple risks. Insurance: Mathematics and Economics 49 (3), 537–546.
Ivanoff, G., Kulik, R., Loukrati, H., 2021. Integral functionals and the bootstrap for the tail empirical process.
Kulik, R., Soulier, P., 2011. The tail empirical process for long memory stochastic volatility sequences. Stochastic Processes and their Applications 121 (1), 109–134.
Necir, A., Rassoul, A., Zitikis, R., 2010. Estimating the conditional tail expectation in the case of heavy-tailed losses. Journal of probability and statistics 2010 (2010), 1–17. https://doaj.org/article/7b5eb78a0c6d49e3ab4a5a2a419bf118.
Scaillet, O., 2005. Nonparametric estimation of conditional expected shortfall. Assurances et Gestion des Risques 72 (4), 639–660. http://search.proquest.com/docview/217465456/.
Sheather, S.J., Marron, J.S., 1990. Kernel quantile estimators. Journal of the American Statistical Association 85 (410), 410–416. http://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10476214.
Soulier, P., 2009. Some applications of regular variation in probability and statistics. XXII Escuela Venezolana de Mathematicas.
Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association 73 (364), 812–815. http://www.tandfonline.com/doi/abs/10.1080/01621459.1978.10480104.