Extremes of Stochastic Volatility Models

Richard A. Davis[1] and Thomas Mikosch[2]

[1] Department of Statistics, Colorado State University, Fort Collins, CO 80523, U.S.A., www.stat.colostate.edu/∼rdavis, rdavis@stat.colostate.edu

[2] Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, www.math.ku.dk/∼mikosch, mikosch@math.ku.dk

The simple stochastic volatility process $(X_t)_{t\in\mathbb{Z}}$ is given by the equation

$$X_t = \sigma_t\, Z_t\,, \quad t \in \mathbb{Z}\,, \tag{1}$$

where (Z_t) is iid, $(\sigma_t)_{t\in\mathbb{Z}}$ is the log-linear Gaussian process given by

$$2\log\sigma_t = \sum_{j=0}^{\infty}\psi_j\eta_{t-j}\,,$$

with $\sum_{j=0}^{\infty}\psi_j^2 < \infty$, and the sequence (η_t) is iid $\mathrm{N}(0,\tau^2)$ and independent of (Z_t). If $\mathrm{var}(Z_t) < \infty$, then it is customary to assume that (Z_t) is iid with mean 0 and variance 1. In this article, we describe the limiting behavior of the sample maxima,

$$M_n = \max(X_1,\ldots,X_n)\,,$$

of the strictly stationary stochastic volatility sequence (X_t) in the cases that the noise (Z_t) has either a light- or heavy-tailed distribution.

In Section 1, we describe the tail behavior of the marginal distribution of X_1. Point process convergence based on the normalized process is described in Section 2. This provides the key result from which limiting behavior of the extremes of (X_t) can be determined.

Interestingly, and unlike the situation for GARCH processes (see Davis and Mikosch [5]), there is no extremal clustering for stochastic volatility processes in both the light- and heavy-tailed cases. That is, large values of the processes do not come in clusters. More precisely, the large sample behavior of M_n is the same as that of the maxima of the associated iid sequence $(\widehat{X}_t)$, where $\widehat{X} \stackrel{d}{=} X$.

1 The tail behavior of the marginal distribution

1.1 The light-tailed case

For the model given by (1), assume further that the noise (Z_t) is iid N(0,1). Notice that the log of the squares of the process, i.e.,

$$Y_t = \log X_t^2 \tag{2}$$

is the superposition of a linear Gaussian process with iid log-χ_1^2 distributed noise. Since $\log Z_t^2$ is distributed as the log of a χ_1^2 random variable, its cumulant generating function is given by

$$\begin{aligned}\log E\exp\left\{\lambda \log \chi_1^2\right\} &= \lambda \log 2 + \log \Gamma(1/2+\lambda) - \log \Gamma(1/2)\\ &= \lambda\log\lambda + \lambda(\log 2 - 1) + (\log 2)/2 + h_0(\lambda)\,,\end{aligned}$$

where the remainder $h_0(\lambda) = O(1/\lambda)$ as $\lambda \to \infty$. The cumulant generating function of Y_t is

$$\begin{aligned}\kappa(\lambda) = \log E\mathrm{e}^{\lambda Y_t} &= \lambda^2\widetilde{\sigma}^2/2 + \lambda\log 2 + \log\Gamma(1/2+\lambda) - \log\Gamma(1/2)\\ &= \lambda^2\widetilde{\sigma}^2/2 + \lambda\log\lambda + \lambda(\log 2 - 1) + (\log 2)/2 + O(1/\lambda),\end{aligned}$$

where

$$\widetilde{\sigma}^2 = \operatorname{var}\left(\sum_{j=0}^{\infty}\psi_j\eta_{t-j}\right) = \tau^2\sum_{j=0}^{\infty}\psi_j^2\,. \tag{3}$$

The following proposition, due to Breidt and Davis [1], describes the tail behavior of Y_t. The proof of this result is based on the asymptotic normality of the Esscher transform of the distribution of Y which can be viewed as the saddlepoint approximation to the cumulant generating function, see Jensen [10]. This technique was also used by Feigin and Yashchin [8] and Davis and Resnick [3].

Proposition 1. *For the log-squared volatility process* (Y_t) *defined in* (2), *we have*

$$P(Y_t > x) \sim \frac{\widetilde{\sigma}^2}{\sqrt{\pi}}\exp\left\{-\frac{x^2}{2\widetilde{\sigma}^2} + \frac{x\log x}{\widetilde{\sigma}^2} + \frac{(k-1)x}{\widetilde{\sigma}^2} - \frac{(k+\widetilde{\sigma}^2)\log x}{\widetilde{\sigma}^2} - \frac{\log^2 x}{2\widetilde{\sigma}^2} - \frac{k^2}{2\widetilde{\sigma}^2} + O\left(\frac{\log^2 x}{x}\right)\right\},$$

as $x \to \infty$, *where* $k = \log(2/\widetilde{\sigma}^2)$.

By symmetry of X_t,

$$P(X_t > x) = \frac{1}{2}\,P(|X_t| > x) = \frac{1}{2}\,P(\log X_t^2 > 2\log x)\,,$$

and then the asymptotic behavior of $P(X_t > x)$ as $x \to \infty$ is straightforward from Proposition 1.

1.2 The heavy-tailed case

Now assume that Z_t has regularly varying tail probabilities with index $\alpha > 0$. This means that the distribution of $|Z_t|$ is regularly varying with index α, i.e.,

$$P(|Z_t| > x) = L(x)\, x^{-\alpha} \tag{4}$$

as $x \to \infty$, where $L(\cdot)$ is a slowly varying function at infinity (see Section 4 of Davis and Mikosch [5]), and that the tail balancing condition,

$$\lim_{x\to\infty} \frac{P(Z_t > x)}{P(|Z_t| > x)} = p \quad \text{and} \quad \lim_{x\to\infty} \frac{P(Z_t \le -x)}{P(|Z_t| > x)} = q\,, \tag{5}$$

where $p + q = 1$ for some $p \in [0, 1]$, holds. Then, by virtue of Breiman's result (see equation (16) in [5]), the distribution of X_t inherits the same tail behavior as Z_t.

Proposition 2. *Under the regularly varying and tail-balancing assumptions* (4) *and* (5)*, we have*

$$P(X_t > x) \sim E(\sigma_t^\alpha)\, P(Z_t > x) \quad \textit{and} \quad P(X_t \le -x) \sim E(\sigma_t^\alpha)\, P(Z_t \le -x)\,,$$

as $x \to \infty$.

Proposition 2 remains valid even if $\log \sigma_t$ is not a linear Gaussian process. In order to apply Breiman's result, one only needs that σ_t is independent of Z_t and $E(\sigma_t^{\alpha+\epsilon}) < \infty$ for some $\epsilon > 0$.

2 Point process convergence

2.1 Background

The theory of point processes plays a central role in extreme value theory. For example, the limiting distribution of order statistics, such as the kth largest, is often easy to derive from the convergence of a particular sequence of point processes. To illustrate this notion, suppose $(\widehat{X}_t)$ is an iid sequence of random variables with the same common distribution function F as X_t. Further assume that there exist sequences of constants $a_n > 0$ and b_n such that

$$P(a_n^{-1}(\widehat{M}_n - b_n) \le x) = F^n(a_n x + b_n) \to G(x) \tag{6}$$

for all x, where $\widehat{M}_n = \max(\widehat{X}_1, \dots, \widehat{X}_n)$ and G is a nondegenerate distribution function. Then by the extremal types theorem, G has to be an extreme value distribution of which there are only three types, see Leadbetter et al. [11]. Moreover, by taking logarithms and using a Taylor series expansion, (6) holds if and only if for any $x \in \mathbb{R}$,

$$n\,(1 - F(a_n x + b_n)) \to -\log G(x)\,.$$

or,[3] equivalently, if for any $x \in \mathbb{R}$,

$$n\,P(a_n^{-1}(\widehat{X}_1 - b_n) > x) \to -\log G(x)\,. \tag{7}$$

Now (7) can be strengthened to the statement,

$$n\,P(a_n^{-1}(\widehat{X}_1 - b_n) \in B) \to \nu(B) \tag{8}$$

for all suitably chosen Borel sets B, where the measure ν is defined by its value on intervals of the form $(a, b]$ as

$$\nu(a, b] = \log G(b) - \log G(a)\,. \tag{9}$$

The convergence in (8) can be connected with the convergence in distribution of a sequence of point processes. For a bounded Borel set B in the product space $(0, \infty) \times \mathbb{R}$, define the sequence of point processes $(\widehat{N}_n)$ by

$$\widehat{N}_n(B) = \#\{(j/n, a_n^{-1}(\widehat{X}_j - b_n)) \in B\,, j = 1, 2, \ldots\}\,.$$

If B is the rectangle $(a, b] \times (c, d]$ with $0 \le a < b < \infty$ and $-\infty < c < d < \infty$, then since the $\widehat{X}_j$ are iid, $\widehat{N}_n(B)$ has a binomial distribution with number of trials $[nb] - [na]$ ($[s]$ = integer part of s), and probability of success

$$p_n = P(a_n^{-1}(\widehat{X}_1 - b_n) \in (c, d])\,.$$

Provided $\nu(c, d] < \infty$, it follows from (8) that $\widehat{N}_n(B)$ converges in distribution to a Poisson random variable $N(B)$ with mean $\mu(B) = (b - a)\,\nu(c, d]$. In fact, we have the stronger point process convergence,

$$\widehat{N}_n \xrightarrow{d} N\,, \tag{10}$$

where N is a Poisson process on $(0, \infty) \times \mathbb{R}$ with mean measure $\mu(dt, dx) = dt \times \nu(dx)$ and $\xrightarrow{d}$ denotes convergence in distribution of point processes. For our purposes, $\xrightarrow{d}$ for point processes means that for any collection of *bounded*[4] Borel sets $B_1, \ldots, B_k$ for which $P(N(\partial B_j) > 0) = 0$, $j = 1, \ldots, k$, we have

[3] If $G(x) = 0$ we interpret $-\log G(x)$ as ∞.

[4] In some cases, especially the heavy-tailed case, the state space of the point process is often defined to be $(0, \infty) \times ([-\infty, \infty] \setminus \{0\})$. On the second product space, the roles of zero and infinity have been interchanged so that bounded sets are now those sets which are bounded away from 0. With this convention, a bounded set on the product space is contained in the rectangle $[0, c] \times ([-\infty, -d] \cup [d, \infty])$ for some positive and finite constants c and d. Under this topology, the intensity measure for the Poisson process defined in Theorem 2 is ensured to be finite on all bounded Borel sets.

$$(\widehat{N}_n(B_1), \ldots, \widehat{N}_n(B_k)) \xrightarrow{d} (N(B_1), \ldots, N(B_k))$$

on $\mathbb{R}^k$, see [7, 11, 12].

As an application of (10), define $\widehat{M}_{n,2}$ to be the second largest among $\widehat{X}_1, \ldots, \widehat{X}_n$. Since the event $\{a_n^{-1}(\widehat{M}_{n,2} - b_n) \le y\}$ is the same as $\{\widehat{N}_n((0,1] \times (y,\infty)) \le 1\}$, we conclude from (10) that

$$\begin{aligned} P(a_n^{-1}(\widehat{M}_{n,2} - b_n) \le y) &= P(\widehat{N}_n((0,1] \times (y,\infty)) \le 1) \\ &\to P(N((0,1] \times (y,\infty)) \le 1)) \\ &= G(y)\,(1 - \log G(y))\,. \end{aligned}$$

Similarly, the joint limiting distribution of $(\widehat{M}_n, \widehat{M}_{n,2})$ can be calculated by noting that for $y \le x$, $\{a_n^{-1}(\widehat{M}_n - b_n) \le x, a_n^{-1}(\widehat{M}_{n,2} - b_n) \le y\} = \{\widehat{N}_n((0,1] \times (x,\infty)) = 0, \widehat{N}_n((0,1] \times (y,x]) \le 1\}$. Hence,

$$\begin{aligned} &P(a_n^{-1}(\widehat{M}_n - b_n) \le x, a_n^{-1}(\widehat{M}_{n,2} - b_n) \le y) \\ &= P(\widehat{N}_n((0,1] \times (x,\infty)) = 0, \widehat{N}_n((0,1] \times (y,x]) \le 1) \\ &\to P(N((0,1] \times (x,\infty)) = 0, N((0,1] \times (y,x]) \le 1)) \\ &= G(y)(1 + \log G(x) - \log G(y))\,. \end{aligned}$$

2.2 Application to stochastic volatility models

The point process convergence in (10) can be extended to general stationary time series provided a mixing condition and a local dependence condition (such as D and D' in Leadbetter et al. [11]) hold. The mixing condition governs how fast a certain class of events become independent as their time separation increases. Typically, many time series models, including stochastic volatility processes, satisfy a mixing condition such as strong mixing. (For stochastic volatility processes, see the discussion for strong mixing given in Section 2 of Davis and Mikosch [5].) On the other hand, the dependence condition D' restricts the clustering of extremes. That is, given an observation at time t is large, the probability that any of its neighboring observations are also large is quite small. The stochastic volatility processes (X_t) given in (1) with either light- or heavy-tailed noise satisfies generalized versions of conditions D and D'; see [1, 4]. Thus the point process convergence in (10) holds. This result is recorded in the following two theorems whose proofs can be found in Breidt and Davis [1] for the light-tailed case (Theorem 1) and in Davis and Mikosch [4] for the heavy-tailed case (Theorem 2).

The light-tailed case

Theorem 1. *Suppose* (X_t) *is the stochastic volatility process defined in* (1)*, where the noise* (Z_t) *is iid* $\mathrm{N}(0,1)$ *and the autocorrelation function*

$$\rho(h) = \operatorname{corr}(\log \sigma_t^2, \log \sigma_{t+h}^2)$$

decays at the rate $\rho(h) = o(1/\log h)$ *as* $h \to \infty$. *Let the constants* a_n *and* b_n *be defined by*

$$a_n = \widetilde{\sigma}\ (2\log n)^{-1/2} = (2/\widetilde{\sigma}^2)^{1/2}/d_n \tag{11}$$

where $d_n = (\log n)^{1/2}$, $\widetilde{\sigma}^2$ *is given in* (3), *and*

$$b_n = c_1 d_n + \log d_n + c_2 + c_3 \frac{\log d_n}{d_n} + c_4 \frac{1}{d_n}, \tag{12}$$

where

$$c_1 = \left(2\widetilde{\sigma}^2\right)^{1/2},\ c_2 = \frac{3}{2}\log 2 - \frac{1}{2}\log \widetilde{\sigma}^2 - 1,\ c_3 = -\frac{\widetilde{\sigma}}{\sqrt{2}},$$

and

$$c_4 = -\frac{1}{2(2\widetilde{\sigma}^2)^{1/2}}\left(1 + \widetilde{\sigma}^2 \log(2\pi)\right).$$

Then, with $Y_t = \log X_t^2$, *the limit in* (8) *holds with* $\widehat{X}_1$ *replaced with* Y_1 *and* $G(x) = \exp\{-\exp\{-x\}\}$ *in* (9). *Moreover,* $N_n \xrightarrow{d} N$, *where* N_n *is the point process defined by*

$$N_n(B) = \#\{(j/n, a_n^{-1}(Y_j - b_n)) \in B,\ j = 1, 2, \ldots\},$$

and N *is a Poisson point process on* $(0,\infty)\times(-\infty,\infty)$ *with intensity measure* $dt \times \nu(dx)$.

The theorem shows that for a wide class of stochastic volatility models driven with normal noise, the extremes of (Y_t) can be normalized independently of the covariance structure in $(\log \sigma_t^2)$, and the same limiting distribution is obtained in all cases. In finite samples, however, the degree of dependence in this linear process does affect the goodness-of-fit of the limiting distribution (see Figure 1 of [1]).

Defining the maximum of the log-squared volatility sequence by $M_n^Y = \max(Y_1, \ldots, Y_n)$, the limit distribution of the maxima can be determined directly from the theorem in the way explained in Section 2.1 and is given by

$$\begin{aligned} P(a_n^{-1}(M_n^Y - b_n) \le x) &= P(N_n((0,1] \times (x,\infty)) = 0) \\ &\to P(N((0,1] \times (x,\infty)) = 0) \\ &= \mathrm{e}^{-\mathrm{e}^{-x}}, \quad x \in \mathbb{R}. \end{aligned} \tag{13}$$

The limit is the Gumbel distribution. It is one of the extreme value distributions, see [11].

The limiting distribution for the maxima $M_n^{|X|} = \max(|X_1|, \ldots, |X_n|)$ of the absolute values of the original (untransformed) volatility process (X_t) can also be found from (13). Indeed, observe that for any $x \in \mathbb{R}$,

$$
\begin{aligned}
&P(a_n^{-1}(M_n^Y - b_n) \le x)\\
&= P(M_n^{|X|} \le \mathrm{e}^{0.5\,(a_n\,x+b_n)})\\
&= P(\mathrm{e}^{-0.5\,b_n}\,(0.5\,a_n)^{-1}\,(M_n^{|X|} - \mathrm{e}^{0.5\,b_n}) \le\, x + o(1))\,,
\end{aligned}
$$

where we used the Taylor expansion argument $\exp\{0.5\,a_n\,x\} = 1 + 0.5\,a_n\,x + o(a_n)$. Another Taylor expansion yields as $n \to \infty$

$$
\mathrm{e}^{\,0.5\,b_n}\,0.5\,a_n \sim 2^{-3/4}\,\widetilde{\sigma}^{-3/2}\,\mathrm{e}^{\,-0.5}\,(\log d_n)^{1/2}\mathrm{e}^{\,\widetilde{\sigma}\,d_n/\sqrt{2}} = \widetilde{a}_n\,. \tag{14}
$$

Combining the arguments above and recalling that the Gumbel distribution is continuous, we may conclude the following.

Corollary 1. *Under the conditions of Theorem 1,*

$$
P(\widetilde{a}_n^{-1}\,(M_n^{|X|} - \mathrm{e}^{\,0.5\,b_n}) \le x) \to \mathrm{e}^{\,-\mathrm{e}^{\,-x}}\,, \quad x \in \mathbb{R}\,,
$$

where $\widetilde{a}_n$ *and* b_n *are defined in* (14) *and* (12)*, respectively.*

One can also recover the limit distribution for the maxima $M_n^X = \max(X_1,\ldots,X_n)$ of the original series. The proof, which we sketch here, follows the argument given in de Haan et al. [9] as adapted by Breidt and Davis [1]. First note that $X_t = |X_t|\,r_t$, where $(r_t) = (\mathrm{sign}(X_t))$ is an iid sequence with $P(r_t = 1) = P(r_t = -1) = 0.5$ and independent of $(|X_t|)$. For x fixed, set $B_n = N_n((0,1] \times (x,\infty))$, $u_n = a_n x + b_n$, and $v_n^2 = \exp\{u_n\}$. If $1 \le \tau_1 < \tau_2 < \cdots$ denote the times at which (X_t^2) exceeds v_n^2, then

$$
\begin{aligned}
P(M_n^X \le v_n) &= \sum_{k=0}^{\infty} P\left(B_n = k, M_n^X \le v_n\right)\\
&= \sum_{k=0}^{\infty} P\left(B_n = k, r_{\tau_1} = -1,\ldots,r_{\tau_k} = -1\right),
\end{aligned} \tag{15}
$$

because the event $\{B_n = k, M_n^X \le v_n\}$ corresponds to the event that there are exactly k exceedances of v_n^2 by $X_1^2,\ldots,X_n^2$, where each such exceedance corresponds to a negative sign of the respective noise term. Since B_n is independent of the signs of the X_t and the random times τ_i depend only on $|X_{\tau_i}|$ and are independent of (r_{τ_i}), the right-hand side of (15) is equal to

$$
\begin{aligned}
\sum_{k=0}^{\infty} P\left(B_n = k\right) 2^{-k} &\to \sum_{k=0}^{\infty} P\left(N((0,1] \times (x,\infty)) = k\right) 2^{-k}\\
&= \sum_{k=0}^{\infty} \frac{(\mathrm{e}^{\,-x}/2)^k\,\mathrm{e}^{\,-\mathrm{e}^{\,-x}}}{k!}\\
&= \mathrm{e}^{\,-\mathrm{e}^{\,-x}/2}\,,
\end{aligned}
$$

where the first equality follows from dominated convergence. Using a Taylor series expansion on v_n as above, we obtain the following limit result for M_n^X.

Corollary 2. *Under the conditions of Theorem 1,*

$$P(\widetilde{a}_n^{-1}\,(M_n^X - \mathrm{e}^{\,0.5\,b_n}) \le x) \to \mathrm{e}^{\,-\mathrm{e}^{\,-x}}\,, \quad x \in \mathbb{R}\,,$$

where $\widetilde{a}_n$ *and* b_n *are defined in* (14) *and* (12)*, respectively.*

The heavy-tailed case

Theorem 2. *Suppose* (X_t) *is the stochastic volatility process given by* (1)*, where* Z_t *satisfies* (4) *and* (5)*. Let* a_n *be the* $(1-n^{-1})$*-quantile of* $|X_t|$*, i.e.,* $a_n = \inf\{x:\, P(|X_t| > x) \le n^{-1}\}$ *and define the point process* N_n *by*

$$N_n(B) = \#\{(j/n, a_n^{-1}X_j) \in B,\, j = 1,2,\ldots\}\,.$$

Then $N_n \xrightarrow{d} N$*, where* N *is a Poisson point process on* $(0,\infty)\times(-\infty,\infty)$ *with intensity measure* $dt \times \nu(dx)$*, and*

$$\nu(dx) = \big(p\,\alpha\,x^{-\alpha-1}1_{(0,\infty)}(x) + q\,\alpha\,(-x)^{-\alpha-1}1_{(-\infty,0)}(x)\big)\,dx\,.$$

Moreover,

$$P(a_n^{-1}\,M_n \le x) \to \mathrm{e}^{\,-p\,x^{-\alpha}}\,, \quad x > 0\,,$$

i.e., the limit is the Fréchet distribution which is one of the extreme value distributions, see [7, 11, 12].

For a stationary process (X_t) that satisfies a general mixing condition, one can often show the existence of a $\theta \in (0,1]$ such that

$$P(a_n^{-1}(M_n - b_n) \le x) \to G^{\theta}(x)\,,$$

where the marginal distribution of the process satisfies (7). The parameter θ is called the *extremal index* and measures the level of clustering of extremes for stationary processes. One can interpret $1/\theta$ as the mean cluster size of exceedances above a high threshold. For $\theta = 1$, there is no clustering and so the maxima behave asymptotically the same as the corresponding maxima of the iid sequence with the same marginal distribution. For the stochastic volatility process with either light- or heavy-tailed noise, it follows from Corollary 2 and Theorem 2 that the extremal index is always 1. In contrast, the extremal index for the GARCH process is always less than one; see Davis and Mikosch [6]. So while both stochastic volatility and GARCH processes exhibit volatility clustering, only the GARCH has clustering of extremes.

References

1. Breidt, F.J. and Davis, R.A. (1998) Extremes of stochastic volatility models. *Ann. Appl. Probab* **8**, 664–675.

2. Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition. Springer, Berlin, Heidelberg, New York.
3. Davis, R.A. and Resnick, S.I. (1991) Extremes of moving averages of random variables with finite endpoint. *Ann. Probab.* **19**, 312–328.
4. Davis, R.A. and Mikosch, T. (2001) Point process convergence of stochastic volatility processes with application to sample autocorrelation. *J. Appl. Probab* **38A**, 93–104.
5. Davis, R.A. and Mikosch, T. (2007a) Probabilistic properties of stochastic volatility models. This volume, pp. ???.
6. Davis, R.A. and Mikosch, T. (2007b) Extreme value theory for GARCH modelling with applications. This volume, pp. ???
7. Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.
8. Feigin, P.D. and Yashchin, E. (1983) On a strong Tauberian result. *Z. Wahrsch. Verw. Gebiete* **65**, 35–48.
9. Haan, L. de, Resnick, S.I., Rootzén, H. and Vries, C.G. de (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stoch. Proc. Appl.* **32**, 213–224.
10. Jensen, J.L. (1995) *Saddlepoint Approximations,* Oxford University Press, Oxford.
11. Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes,* Springer-Verlag, New York.
12. Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.