HARCH processes are heavy tailed

Paul Embrechts (embrechts@math.ethz.ch)
Department of Mathematics, ETHZ, CH–8092 Zürich, Switzerland

Rudolf Grübel (rgrubel@stochastik.uni-hannover.de)
Institut für Mathematische Stochastik, Universität Hannover, D–30060 Hannover, Germany

Abstract. Heterogeneous Auto-Regressive Conditional Heteroskedastic processes were introduced by Müller, Dacarogna, Davé, Olsen, Pictet and von Weizsäcker in order to improve traditional ARCH-type models describing financial time series. In a later paper Embrechts, Samorodnitsky, Dacorogna and Müller asked how heavy the tails of stationary processes of this type are. We provide a partial answer to this question, using mainly monotonicity arguments to compare HARCH processes to other processes with a simpler recursive structure.

Keywords: Financial time series, Pareto tails, random recursions.

AMS 1991 subject classification: 60K30,90A09

1. Introduction and results.

Suppose that $\{S_n : n = 0, 1, 2, \ldots\}$ denotes the market prices of a financial asset (e.g. stock, exchange rate, commodity, ...) at given discrete time points. In empirical finance, rather than considering S_n, one looks at the so–called (log–)return process $\{R_n = \log(S_n/S_{n-1}) : n = 1, 2, \ldots\}$. When plotted, the latter series at first very much looks like white noise. A more careful analysis however reveals important extra structure. The following, so–called stylised facts of (R_n) have been deduced in numerous econometric publications; see for instance Taylor (1986):

- (SF1) the R_n's are uncorrelated;
- (SF2) $(|R_n|)$ as well as $\left(R_n^2\right)$ exhibit significant correlation;
- (SF3) the unconditional distribution $P(R > x)$ is heavy–tailed, or indeed leptokurtic;
- (SF4) the volatility of the returns (R_n) changes stochastically, and
- (SF5) the data show long memory.

The above observations are clearly linked: for instance (SF4) induces, through variance mixing, heavy–tailedness, i.e. (SF3). Also (SF1–2) can partly be explained through a careful choice of stochastic volatility in (SF4). Volatility is the main driving force behind trading in financial markets. Money is gained or lost in high–volatile markets (in this brief discussion, we do not distinguish between implied or historical volatility). To give an idea of volatility changes, whereas for most markets, volatility on a yearly basis is about 15–20%, in periods of crisis, values may jump to (and indeed persist at) much higher levels. During the 1987 crash, volatility jumped to well over 150%, during the Kuweit crisis volatility persisted for longer periods at about 40%, and in the wake of the LTCM debacle peaks of about 80% were reached. As such, of key importance is the modelling of the volatility process $\{\sigma_n : n = 1, 2, \ldots\}$ in $R_n = \sigma_n \epsilon_n$, where (ϵ_n) is a specified innovation process. Clearly, the plain vanilla Black–Scholes model where $\sigma_n \equiv \sigma$ and the (ϵ_n) are iid $N(0,1)$ is definitely out. Early competitors for more realistic models were of the so–called ARCH type: for instance, an ARCH(k) process satisfies the following recursive equation,

$$R_n = \sigma_n \epsilon_n \,, \quad \sigma_n^2 = c_0 + \sum_{j=1}^{k} c_j R_{n-j}^2 \,,$$

where the (ϵ_j) are iid innovations and the constants (c_j) so defined that a stationary solution to the above equations exists. In line with classical time series analysis, one may introduce a moving average term in the σ_n^2 and/or include differencing. Such processes exhibit many of the properties found in empirical data. In particular, (SF3) can be deduced using a powerful result of Kesten (1973) on products of random matrices. The key point is that the stationary solution of the defining recursive equations turns out to be the distributional fixpoint of a random map. For an example on this for ARCH(1) processes, see Embrechts, Klüppelberg and Mikosch (1997), Section 8.4. For processes more general than the ARCH-GARCH type, such a fixpoint approach may not exist. In the present paper such a class of processes is discussed and its distributional tail properties investigated using a stochastic inequality argument. We believe that this approach may be more widely applicable. For an excellent review on stochastic volatility models, see Frey (1997).

A HARCH(k) process $(R_n)_{n\in\mathbb{N}_0}$ with parameters $c_0, \ldots, c_k$ satisfies the recursive relation

$$R_n = \sigma_n \epsilon_n, \quad \sigma_n^2 \;=\; c_0 \;+\; \sum_{j=1}^{k} c_j \left(\sum_{i=1}^{j} R_{n-i} \right)^2 , \tag{1}$$

where we generally assume that $c_0 > 0$, $c_k > 0$ and $c_j \geq 0$ for $j = 1, \ldots, k-1$. The innovation sequence $(\epsilon_n)_{n\in\mathbb{N}_0}$ consists of independent and identically distributed random variables; ϵ_n is independent of σ_n.

Typically, the distribution of the ϵ-variables is assumed to be normal with mean 0 and variance $\sigma^2 > 0$, which we abbreviate by $\epsilon \sim N(0, \sigma^2)$, or the ϵ_n's have a (central) Student t–distribution.

HARCH processes were introduced by Müller et al. (1995) in order to improve the traditional ARCH–type models, especially with a view towards long memory of volatility ((SF5)) and possible asymmetries between volatilities with different degrees of time resolution. The main difference between the HARCH and ARCH definition is that in the former, a cancelling effect for successive returns is possible because of the $\sum_{i=1}^{j} R_{n-i}$ term. It is especially the latter which allows for the distinction between so–called intra–day and long–term traders; see Müller et al. (1995). In Embrechts et al. (1998) the question was asked how heavy the tails of the stationary distribution of such models are ((SF3)). An answer might imply, for example, the existence or non–existence of certain moments, which in turn is of considerable importance for the applicability of standard statistical techniques such as central–limit–theorem based asymptotic confidence intervals for the model parameters; see Embrechts, Klüppelberg and Mikosch (1997) for an in–depth discussion of such aspects.

Embrechts et al. (1998) give conditions under which a k-step Markov chain $(R_n)_{n \in \mathbb{N}_0}$ with dynamics given by (1) converges in distribution; in this case we write R_∞ for the (distributional) limit of R_n as $n \to \infty$. General Markov chain theory can be used to show that the law of R_∞ does not depend on the initial values of $R_0, \ldots, R_{k-1}$. We will assume throughout this note that the parameters $c_0, \ldots, c_k$ of the model and the innovation distribution are such that we indeed are in this situation and refer to Embrechts et al. (1998) for further details. Under these conditions we have the following results.

THEOREM 1. *If* $c_1 > 0$ *and* $P\left(\epsilon^2 > c_1^{-1}\right) > 0$, *then*

$$\liminf_{r \to \infty} \frac{\log P\left(|R_\infty| > r\right)}{\log r} \; > \; -\infty \, . \tag{2}$$

THEOREM 2. *If the distribution of the innovations* ϵ_i, $i \in \mathbb{N}_0$, *is symmetric about 0 and if* $P\left(\epsilon^2 > \max\{c_1, \ldots, c_k\}^{-1}\right) > 0$, *then (2) holds.*

As a corollary we obtain that under quite general conditions not all moments of the stationary distribution exist. For symmetric innovation distributions the distribution of R_∞ is also symmetric so that both individual tails decrease at a rate which is at best polynomial ('Pareto tails'). The proofs, which we give in the next section, can also be used to obtain explicit asymptotic lower bounds. Together with some other extensions these are discussed in Section 3.

2. Proofs.

Both proofs rely on a comparison to a suitably chosen simpler model. Our results deal with distributions rather than individual random variables. In order to be able to compare distributions we need a notion of stochastic ordering: If μ_i, $i = 1, 2$, are probability distributions on the (Borel subsets of) the real line $\mathbb{R}$, then we call μ_1 smaller than or equal to μ_2 and write $\mu_1 \preceq \mu_2$, if $\mu_1((x,\infty)) \leq \mu_2((x,\infty))$ for all $x \in \mathbb{R}$. We will occasionally use the sloppy notation $X_1 \preceq X_2$ if $\mathcal{L}(X_1) \preceq \mathcal{L}(X_2)$ where, generally, $\mathcal{L}(X)$ denotes the distribution of the random variable X. The key to heavy–tailedness is the following observation, interesting in its own right.

LEMMA 3. *If X is a non–negative random variable with $P(X > 1) > 0$ and if the non–negative random variable Y satisfies the stochastic inequality*

$$Y \succeq X \cdot Y, \quad \text{with } X \text{ and } Y \text{ independent}, \tag{3}$$

then either $P(Y = 0) = 1$ or

$$\liminf_{y\to\infty} \frac{\log P(Y > y)}{\log y} > -\infty \,.$$

Proof of Lemma 3. If $Y = 0$ almost surely then there is nothing to prove, so let $c > 0$ be such that $p := P(Y \geq c) > 0$. By assumption there exists an $a > 1$ such that $q := P(X \geq a) > 0$. Clearly, we may assume that $q < 1$, as otherwise Y would be equal to $+\infty$ with non–vanishing probability. We claim that

$$P(Y \geq c \cdot a^n) \geq p \cdot q^n \quad \text{for all } n \in \mathbb{N}_0. \tag{4}$$

For $n = 0$ this is obvious. If (4) holds for some $n \in \mathbb{N}_0$ then we obtain on using (3)

$$\begin{aligned} P\left(Y \geq ca^{n+1}\right) &\geq P\left(X \cdot Y \geq ca^{n+1}\right) \\ &\geq P\left(aY \geq ca^{n+1},\, X \geq a\right) \\ &= P\left(Y \geq ca^n\right) P\left(X \geq a\right) \\ &\geq p\, q^n\, q\,, \end{aligned}$$

which completes the induction step so that (4) is proved. We now use a standard trick to apply the tail bound (4) in connection with the continuous limit $y \to \infty$: Any $y \geq c$ determines a unique $n = n(y) \in \mathbb{N}_0$

such that $ca^n \leq y < ca^{n+1}$. Note that $n \leq (\log a)^{-1}(\log y - \log c)$ and that $\log q < 0$. Therefore,

$$\begin{aligned}\log P(Y \geq y) &\geq \log P\left(Y \geq ca^{n+1}\right) \\ &\geq \log p + (n+1)\, \log q \\ &\geq \log p + \left(\frac{\log y - \log c}{\log a} + 1\right) \log q\,,\end{aligned}$$

which yields the asymptotic lower bound

$$\liminf_{y \to \infty} \frac{\log P(Y \geq y)}{\log y} \;\geq \frac{\log q}{\log a} > -\infty\,.$$

Proof of Theorem 1. The model equation (1) and the non–negativity assumption on the parameters together imply

$$R_n^2 \geq c_1 \epsilon_n^2\, R_{n-1}^2\,. \tag{5}$$

Stochastic ordering is compatible with weak convergence in the following sense: if $\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x)$ for all continuity points x of the distribution function of X and $\lim_{n\to\infty} P(Y_n \leq y) = P(Y \leq y)$ for all continuity points y of the distribution function of Y, and if $X_n \preceq Y_n$ for all $n \in \mathbb{N}$, then $X \preceq Y$. Hence, if R_n converges to R_∞ in distribution then the right hand side of (5) converges to $c_1\epsilon^2 R_\infty^2$, with ϵ independent of R_∞ and $\mathcal{L}(\epsilon) = \mathcal{L}(\epsilon_i)$ (see Billingsley (1968), Theorem 3.2). Putting this together we see that we can apply the lemma with $Y = R_\infty^2$, $X = c_1\epsilon^2$.

Proof of Theorem 2. Let

$$\eta_n := \begin{cases} 1, & \epsilon_n > 0\,, \\ 0, & \epsilon_n = 0\,, \\ -1, & \epsilon_n < 0\,, \end{cases}$$

be the sign of ϵ_n and let

$$A_{n,m} := \{\eta_{n+k} = \eta_n \quad \text{for } k = 1, \ldots, m\}$$

be the event that the m innovations following ϵ_n have the same sign as ϵ_n. Because of the symmetry of the innovation distribution we have that $A_{n,m}$ is independent of all ϵ- and R–values with index less than or equal to n.

Now suppose that $(R_n)_{n\in\mathbb{N}_0}$ satisfies (1), let $j \in \{1,\dots,k\}$ be fixed. On $A_{n-j,j}$ we have

$$c_j \left(\sum_{i=1}^{j} R_{n-i}\right)^2 \geq c_j R_{n-j}^2 \,,$$

so that

$$R_n^2 \geq c_j \epsilon_n^2 1_{A_{n-j,j}} R_{n-j}^2 \,, \tag{6}$$

where 1_A denotes the indicator function of the set A. With (6) we can proceed as with (5) in the proof of Theorem 1.

3. Comments.

(a) In Embrechts et al. (1998) explicit conditions for the existence of moments were obtained. These results rely on the special algebraic structure of the model equation. Here our results are less explicit, but Theorem 1, for example, can easily be transferred to processes with a different dynamic evolution. Using comparison arguments rather than algebraic manipulations increases the range of applicability, but yields less explicit results.

(b) Our results can in principle be used to obtain quantitative statements on the asymptotic lower bounds. In the situation of Theorem 1, for example, any pair of values $a > c_1^{-1}$, $q > 0$ with $P\left(\epsilon^2 \geq a\right) \geq q$ yields

$$\liminf_{r\to\infty} \frac{\log P\left(|R_\infty| > r\right)}{\log r} \geq \frac{2\log q}{\log a + \log c_1} \,.$$

For a given innovation distribution we can determine the best lower bound by varying a.

(c) A simple fact such as the above lemma often underlies the phenomenon that light–tailed model input can generate heavy–tailed model output. For a similar reasoning in the context of random affine recurrences see Goldie and Grübel (1996), Theorem 4.1. Even the famous Cramér results on the probability of ruin in the Sparre–Anderson model of risk theory can be seen in this light: If M denotes the maximum of a random walk with generic step X, then $M =_{\text{distr}} \max(0, M+X)$. If $P(X>0)>0$ then, upon taking exponentials in the stochastic inequality we see that the lemma can be used to explain the fact that, in a random walk, even a step distribution with bounded support will generate a distribution of the global maximum that has at best an exponentially decreasing tail.

Acknowledgements

This work has been supported by the Mathematical Research Institute of the ETH Zürich. The second author is grateful to the ETH, for hospitality and support.

References

Billingsley, P. (1968) *Weak Convergence of Probability Measures.* Wiley, New York.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

Embrechts, P., Samorodnitsky, G., Dacorogna, M.M. and Müller, U.A. (1998) How heavy are the tails of a stationary HARCH(k) process? A study of the moments. In: *Stochastic Processes and Related Topics. In Memory of Stamatis Cambanis 1943–1995.* I. Karatzas, B.S. Rajput, M.S. Taqqu (Eds.), 69–103.

Frey, R. (1997) Derivative asset analysis in models with level–dependent and stochastic volatlity. *CWI Quaterly* **10**, 1–35.

Goldie, C.M. and Grübel, R. (1996) Perpetuities with thin tails. *Adv. in Applied Prob.* **28**, 463–480.

Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.

Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V. and von Weizsäcker, J.E. (1995) Volatilities of different time resolutions–analyzing the dynamics of market components. *Journal of Empirical Finance* **4**, 213–240.

Taylor, S.J. (1986) *Modelling Financial Time Series.* Wiley, Chichester.