The sample autocorrelations of financial time series models

Richard A. Davis[1]
Colorado State University

Thomas Mikosch
University of Groningen and EURANDOM

Abstract

In this chapter we review some of the limit theory for the sample autocorrelation function (ACF) of linear and non-linear processes $f(x)$ with regularly varying finite-dimensional distributions. We focus in particular on non-linear process models which have attracted the attention for modeling financial time series.

In the first two parts we give a short overview of the known limit theory for the sample ACF of linear processes and of solutions to stochastic recurrence equations (SRE's), including the squares of GARCH processes. In the third part we concentrate on the limit theory of the sample ACF for stochastic volatility models. The limit theory for the linear process and the stochastic volatility models turns out to be quite similar; they are consistent estimators with rates of convergence faster than $\sqrt{n}$, provided that the second moment of the marginal distributions is infinite. In contrast to these two kinds of processes, the sample ACF of the solutions to SRE's can be very slow: the closer one is to an infinite second moment the slower the rate, and in the case of infinite variance the sample ACF has a non-degenerate limit distribution without any normalization.

[1]This research supported in part by NSF DMS Grant No. DMS-9972015.
AMS 1991 Subject Classification: Primary: 62M10 Secondary: 62G20 60G55 62P05 60G10 60G70
Key Words and Phrases. Point process, vague convergence, multivariate regular variation, mixing condition, stationary process, heavy tail, sample autocovariance, sample autocorrelation, GARCH, stochastic volatility model, linear process, financial time series.

1 Introduction

Over the past few years heavy-tailed phenomena have attracted the interest of various researchers in time series analysis, extreme value theory, econometrics, telecommunications, and various other fields. The need to consider time series with heavy-tailed distributions arises from the observation that traditional models of applied probability theory fail to describe jumps, bursts, rapid changes and other erratic behavior of various real-life time series.

Heavy-tailed distributions have been considered in the financial time series literature for some time. This includes the GARCH processes whose marginal distributions can have surprisingly heavy (Pareto-like) tails. There is plenty of empirical evidence (see for example Embrechts et al. [23] and the references cited therein) that financial log-return series of stock indices, share prices, exchange rates, etc., can be reasonably modeled by processes with infinite 3rd, 4th or 5th moments. In order to detect non-linearities, the econometrics literature often recommends to consider not only the time series itself but also powers of the absolute values. This leads to some serious problems: if we accept that the underlying time series has infinite 2nd, 3rd, 4th,... moments we have to think about the meaning of the classical tools of time series analysis. Indeed, the sample autocovariance function (sample ACVF), sample autocorrelation function (sample ACF) and the periodogram are meaningful estimators of their deterministic counterparts ACVF, ACF and spectral density only if the second moment structure of the underlying time series is well defined. If we detect that a log-return series has infinite 4th moment it is questionable to use the sample ACF of the squared time series in order to make statements about the dependence structure of the underlying stationary model. For example consider plots of the sample ACF of the squares of the Standard & Poor's index for the periods 1961-1976, and 1977-1993 displayed in Figure 1.1. For the first half of the data, the ACF of the squares appears to decay slowly, while for the second half the ACF is not significant past lag 9. The discrepancies in the appearance in the two graphs suggest that either the process is non-stationary or that the process exhibits heavy tails.

The same drawback for the sample ACF is also present for the periodogram. The latter estimates the spectral density, a quantity that does not exist for the squared process if the fourth moments are infinite. Thus one should exercise caution in the interpretation of the periodogram of the squares for heavy-tailed data.

Since it has been realized that heavy tails are present in many real-life situations the research on heavy-tailed phenomena has intensified over the years. Various recent publications and monographs such as Samorodnitsky and Taqqu [40] on infinite variance stable processes, Embrechts et al. [23] on extremes in finance and insurance, and Adler et al. [1] on heavy tails, demonstrate the emerging interest and importance of the topic.

It is the aim of this chapter to re-consider some of the theory for the sample ACVF and sample

Figure 1.1 *Sample ACF of the squares of the S& P index for the periods (a) 1961-1976 and (b) 1977-1993.*

ACF of some classes of heavy-tailed processes. These include linear processes with regularly varying tails, solutions to stochastic recurrence equations (SRE's) and stochastic volatility models. The latter two classes are commonly used for the econometric modeling of financial time series in order to describe the following empirically observed facts: non-linearity, dependence and heavy tails. We also included the class of linear processes because of its enormous practical importance for applications but also because heavy tails and linear processes do actually interact in an "optimal" way. This means that the sample ACF still estimates some notion of a *population* ACF, even if the variance of the underlying time series is infinite, and the rate of convergence is faster than the classical $\sqrt{n}$ asymptotics. The situation can change abruptly for non-linear processes. In this case, the sample ACF can have a non-degenerate limit distribution — a fact which makes the interpretation of the sample ACF impossible — or the rates of convergence to the ACF can be extremely slow even when it exists. Such cases include GARCH processes and, more generally, solutions to SRE's. However, not all non-linear models exhibit unpleasant behavior of their sample ACF's. A particularly "good" example in this context is the class of stochastic volatility models whose behavior of the sample ACF is close to the linear process case.

Fundamental to the study of all these heavy-tailed processes is the fact that their finite-dimensional distributions are multivariate regularly varying. Therefore we start in Section 2 with a short introduction to this generalization of power law tails to the multivariate setting. We also define stable random distributions which constitute a well-studied class of infinite variance distributions with multivariate regularly varying tails. In Section 3 we consider the sample ACF of linear processes, followed by the sample ACF of solutions to SRE's in Section 4 and stochastic volatility models in Section 5. The interplay between the tails and the dependence is crucial for the understanding of the asymptotic behavior of the sample ACF. Therefore we first introduce in every section the corresponding model and discuss some of its basic properties. Then we explain where the heavy tails in the process come from and, finally, we give the theory for the sample ACF of these processes. One may distinguish between two types of models. The first type consists of models with a heavy-tailed input (noise) resulting in a heavy-tailed output. This includes the linear and stochastic volatility models. The second type consists of models where light- or heavy-tailed input results in heavy-tailed output. Solutions to SRE's belong to the latter type. They are mathematically more interesting in the sense that the occurrence of the heavy tails has to be explained by a deeper understanding of the nonlinear filtering mechanism.

2 Preliminaries

2.1 Multivariate regular variation

Recall that a non-negative function f on $(0,\infty)$ is said to be regularly varying at infinity if there exists an $\alpha \in \mathbb{R}$ such that

$$f(x) = x^{\alpha}\, L(x)\,,$$

and L is slowly varying, i.e.

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1\,, \quad \forall\, c > 0\,.$$

We refer to Bingham et al. [5] for an encyclopedic treatment of regular variation.

For many applications in probability theory we need to define regular variation of random variables and random vectors.

Definition 2.1 *We say that the random vector $\mathbf{X} = (X_1,\ldots,X_d)$ with values in $\mathbb{R}^d$ and its distribution are* regularly varying in $\mathbb{R}^d$ *if there exist $\alpha \ge 0$ and a probability measure P_Θ on the Borel σ-field of the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ such that the following limit exists for all $x > 0$:*

$$\frac{P(|\mathbf{X}| > t\,x\,,\, \mathbf{X}/|\mathbf{X}| \in \cdot\,)}{P(|\mathbf{X}| > t)} \stackrel{v}{\to} x^{-\alpha}\, P_{\mathbf{\Theta}}(\cdot)\,, \quad t \to \infty\,, \tag{2.1}$$

where $\stackrel{v}{\to}$ denotes vague convergence on the Borel σ-field of $\mathbb{S}^{d-1}$. The distribution P_Θ is called the spectral measure *of $\mathbf{X}$, and α is the* index *of $\mathbf{X}$.*

We refer to Kallenberg [32] for a detailed treatment of vague convergence of measures. We also mention that (2.1) can be expressed in various equivalent ways. For example, (2.1) holds if and only if there exists a sequence of positive constants a_n and a measure μ such that

$$n\, P(a_n\, \mathbf{X} \in \cdot) \stackrel{v}{\to} \mu(\cdot)$$

on the Borel σ-field of $\mathbb{R}^d$. In this case, one can choose (a_n) and μ such that, for every Borel set B and $x > 0$,

$$\mu((x,\infty) \times B) = x^{-\alpha}\, P_\Theta(B)\,.$$

For $d = 1$, we see that X is regularly varying with index α if and only if

$$P(X > x) \sim p\, x^{-\alpha}\, L(x) \quad \text{and} \quad P(X \le -x) \sim q\, x^{-\alpha}\, L(x)\,, \quad x \to \infty\,, \tag{2.2}$$

where p, q are non-negative numbers such that $p + q = 1$ and L is slowly varying. Notice that the spectral measure is just a two-point distribution on $\{-1, 1\}$. Condition (2.2) is usually referred to as the *tail balancing condition*.

Note that regular variation of $\mathbf{X}$ in $\mathbb{R}^d$ implies regular variation of $|\mathbf{X}|$ and of any linear combination of the components of X. The measure P_Θ can be concentrated on a lower-dimensional subset of $\mathbb{S}^{d-1}$. For example, if the random variable X is regularly varying with index α then $\mathbf{X} = (X, 1, \ldots, 1)$ with values in $\mathbb{R}^d$ is regularly varying. If $\mathbf{X}$ has independent components it is easily seen that the spectral measure P_Θ has support on the intersections with the axes. For further information on multivariate regular variation we refer to de Haan and Resnick [27] or Resnick [39]. We also refer to the Appendix of Davis et al. [16] for some useful results about equivalent definitions of regular variation in $\mathbb{R}^d$ and about functions of regularly varying vectors.

In what follows, we will frequently make use of a result by Breiman [12] about the regular variation of products of independent non-negative random variables ξ and η. Assume ξ is regularly varying with index $\alpha > 0$ and $E\eta^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$. Then $\xi\eta$ is regularly varying with index α:

$$P(\xi\eta > x) \sim E\eta^\alpha\, P(\xi > x)\,, \quad x \to \infty\,. \tag{2.3}$$

A multivariate version of Breiman's result can be found in Davis et al. [16].

2.2 Stable distributions

For further use we introduce the notion of α-stable distribution. The following definition is taken from Samorodnitsky and Taqqu [40] which we recommend as a general reference on stable processes and their properties.

Definition 2.2 *Let $0 < \alpha < 2$. Then $\mathbf{X} = (X_1, \ldots, X_d)$ is an* α-stable random vector *in $\mathbb{R}^d$ if there exists a finite measure Γ on the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ and a vector $\mathbf{x}_0$ in $\mathbb{R}^d$ such that:*

1. *If $\alpha \neq 1$, then $\mathbf{X}$ has characteristic function*

$$E\exp\{i\,(\mathbf{y},\mathbf{X})\} = \exp\left\{-\int_{\mathbb{S}^{d-1}} |(\mathbf{y},\mathbf{x})|^\alpha\,(1 - i\,\mathrm{sign}((\mathbf{y},\mathbf{x}))\,)\,\Gamma(d\mathbf{x}) + i\,(\mathbf{y},\mathbf{x}_0)\right\}\,.$$

2. If $\alpha = 1$, *then* $\mathbf{X}$ *has characteristic function*

$$E\exp\{i\ (\mathbf{y},\mathbf{X})\} = \exp\left\{-\int_{\mathbb{S}^{d-1}} |(\mathbf{y},\mathbf{x})| \left(1 + i\ \frac{2}{\pi}\mathrm{sign}((\mathbf{y},\mathbf{x}))\ \log|(\mathbf{y},\mathbf{x})|\right)\Gamma(d\mathbf{x}) + i\ (\mathbf{y},\mathbf{x}_0)\right\}\ .$$

It can be *thatthepair*$(\Gamma, \mathbf{x}_0)$ is unique. Moreover, the vector $\mathbf{X}$ is regularly varying with index α, and the measure Γ determines the form of the spectral measure P_Θ.

The characteristic function of an α-stable vector $\mathbf{X}$ is particularly simple if it is *symmetricinthethat*$\mathbf{X}$*and-*$\mathbf{X}$*havethesamedistribution and*$\mathbf{x}_0 = 0$. In this case, we say that $\mathbf{X}$ has a *symmetric α-stable distribution* (SαS). The characteristic function of a SαS vector $\mathbf{X}$ is particularly simple:

$$E\exp\{i\ (\mathbf{y},\mathbf{X})\} = \exp\left\{-\int_{\mathbb{S}^{d-1}} |(\mathbf{y},\mathbf{x})|^\alpha \Gamma(d\mathbf{x})\right\}\ .$$

For $d = 1$, this formula is even simpler:

$$E\mathrm{e}^{\,i\,y\,X} = \mathrm{e}^{\,-\sigma_\alpha\,|y|^\alpha} \quad \text{for some } \sigma_\alpha > 0.$$

3 The linear process

3.1 Definition

Recall the definition of a *linear process*:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}\,, \quad t \in \mathbb{Z}\,, \tag{3.1}$$

where (Z_t) is an iid sequence of random variables, usually called the *noise* or *innovations* sequence, and (ψ_t) is a sequence of real number. For the a.s. convergence of the infinite series (3.1) one has to impose special conditions on the sequence (ψ_j) which depend on the distribution of Z. The formulation of these conditions will be specified below. It is worth noting that stationary ARMA and fractionally integrated ARMA processes have such a linear representation. We refer to Brockwell and Davis [13] as a general reference on the theory and statistical estimation of linear processes.

Linear processes, in particular the ARMA models, constitute perhaps the best studied class of time series models. Their theoretical properties are well understood and estimation techniques are covered by most standard texts on the subject. By choosing appropriate coefficients ψ_j, the ACF of a linear process can approximate the ACF of any stationary ergodic process, a property that helps explain the popularity and modeling success enjoyed by linear processes. Moreover, the tails of a linear process can be made as heavy as one wishes by making the tails of the innovations heavy. The latter property is an attractive one as well; the coefficients ψ_j occur in the tails only as a scaling factor. Thus the tails and the ACF behavior of a linear process can be modelled almost independently of each other: the tails are essentially determined by the tails of the innovations, whereas the ACF only depends on the choice of the coefficients. This will be made precise in what follows.

3.2 Tails

The distribution of X can have heavy tails only if the innovations Z_t have heavy tails. This follows from some general results for regularly varying and subexponential Z_t's; see Appendix A3.3 in Embrechts et al. [23]. For the sake of completeness we state a result from Mikosch and Samorodnitsky [34] which requires the weakest conditions on (ψ_j) known in the literature.

Proposition 3.1 **1)** *Assume that Z satisfies the tail balancing conditition* (2.2) (*with* $X = Z$) *for some* $p \in (0,1]$ *and* $\alpha > 0$. *If* $\alpha > 1$ *assume* $EZ = 0$. *If the coefficients* ψ_j *satisfy*

$$\begin{cases} \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty & \text{for } \alpha > 2 \\ \sum_{j=-\infty}^{\infty} |\psi_j|^{\alpha-\epsilon} < \infty \quad \text{for some } \epsilon > 0 & \text{for } \alpha \le 2\,, \end{cases}$$

then the infinite series (3.1) *converges a.s. and the following relation holds*

$$\frac{P(X > x)}{P(|Z| > x)} \sim \sum_{j=-\infty}^{\infty} |\psi_j|^\alpha [p\, I_{\{\psi_j > 0\}} + q\, I_{\{\psi_j < 0\}}] =: \|\psi\|_\alpha^\alpha\,. \tag{3.2}$$

2) *Assume Z satisfies the tail balancing condition* (2.2) *for some* $p \in (0,1]$ *and* $\alpha \in (0,2]$, *that the infinite series* (3.1) *converges a.s.,*

$$\sum_{j=-\infty}^{\infty} |\psi_j|^\alpha < \infty\,,$$

and one of the conditions

$$L(\lambda_2) \le c\, L(\lambda_1) \text{ for } \lambda_0 < \lambda_1 < \lambda_2, \text{ some constants } c, \lambda_0 > 0.$$

or

$$L(\lambda_1 \lambda_2) \le c\, L(\lambda_1)\, L(\lambda_2) \text{ for } \lambda_1, \lambda_2 \ge \lambda_0 > 0, \text{ some constants } c, \lambda_0 > 0$$

is satisfied. Then relation (3.2) *holds.*

This proposition implies that heavy-tailed input (regularly varying noise (Z_t)) results in heavy-tailed output. Analogously, one can show that light-tailed input forces the linear process to be light-tailed as well. For example, if the Z_t's are iid Gaussian, then the output time series (X_t) is Gaussian. This is clearly due to the linearity of the process: an infinite sum of independent random variables cannot have lighter tails than any of its summands.

It can be shown, using an extension of the proof of Proposition 3.1, that the finite-dimensional distributions of the process (X_t) are also regularly varying with index α. This means that the vectors $(X_0, \dots, X_d)$, $d \ge 1$, are regularly varying in $\mathbb{R}^d$ with index α and spectral measure determined by the coefficients ψ_j.

3.3 Limit theory for the sample ACF

The limit theory for the sample ACVF and ACF of linear processes with infinite variance was derived in Davis and Resnick [17, 18, 19]. The limit theory for finite variance linear processes is very much the same as for Gaussian processes; see for example Brockwell and Davis [13]. For the sake of simplicity and for ease of presentation we restrict ourselves to the case of infinite variance symmmetric α-stable (SαS) noise (Z_t); see Section 2.2. In this case, one can show that Z has Pareto-like behavior in the sense that

$$P(Z > x) \sim \text{const}\, x^{-\alpha}\,, \quad x \to \infty\,.$$

Define the sample ACF as follows:

$$\widetilde{\rho}_{n,X}(h) := \frac{\sum_{t=1}^{n-h} X_t X_{t+h}}{\sum_{t=1}^{n} X_t^2}\,, \quad h = 1, 2, \ldots\,. \tag{3.3}$$

If (Z_t) was an iid Gaussian $N(0, \sigma^2)$ noise sequence with the same coefficient sequence (ψ_j), (X_t) would be a Gaussian linear process with ACF

$$\rho_X(h) := \frac{\sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=-\infty}^{\infty} \psi_j^2}\,, \quad h = 1, 2, \ldots\,.$$

If (X_t) is generated from iid SαS noise it is by no means clear that $\widetilde{\rho}_{n,X}(h)$ is even a consistent estimator of $\rho_X(h)$. However, from the following surprising result of Davis and Resnick [19] we find that $\widetilde{\rho}_{n,X}(h)$ is not only consistent but has other good properties as an estimator of $\rho_X(h)$. (The following theorem can also be found in Brockwell and Davis [13], Theorem 13.3.1.)

Theorem 3.2 *Let (Z_t) be an iid sequence of SαS random variables and let (X_t) be the stationary linear process* (3.1)*, where*

$$\sum_{j=-\infty}^{\infty} |j|\, |\psi_j|^\delta < \infty \quad \textit{for some } \delta \in (0, \alpha) \cap [0, 1].$$

Then for any positive integer h,

$$\left(\frac{n}{\log n}\right)^{1/\alpha} \left(\widetilde{\rho}_{n,X}(1) - \rho_X(1), \ldots, \widetilde{\rho}_{n,X}(h) - \rho_X(h)\right) \stackrel{d}{\to} (Y_1, \ldots, Y_h)\,,$$

where

$$Y_k = \sum_{j=1}^{\infty} \left[\rho_X(k+j) + \rho_X(k-j) - 2\rho_X(j)\rho_X(k)\right] \frac{S_j}{S_0}\,, \tag{3.4}$$

and $S_0, S_1, \ldots$ are independent stable random variables, S_0 is positive stable with characteristic function

$$E e^{\,i\lambda S_0} = \exp\left\{-\Gamma(1 - \alpha/2)\, \cos(\pi\alpha/4)\, |u|^{\alpha/2}\, (1 - \text{sign}(u)\, \tan(\pi\alpha/4))\right\}$$

and $S_1, S_2, \ldots$ are iid $S\alpha S$ with characteristic function $E\mathrm{e}^{iyS_1} = \mathrm{e}^{-\sigma_\alpha |y|^\alpha}$, where

$$\sigma_\alpha = \begin{cases} \Gamma(1-\alpha)\cos(\pi\alpha/2) & \alpha \neq 1\,, \\ \dfrac{\pi}{2} & \alpha = 1 \end{cases} .$$

Remark 3.3 If $\alpha > 1$ the theorem remains valid for the mean corrected sample ACF, i.e. when $\widetilde{\rho}_{n,X}(h)$ is replaced by

$$\widehat{\rho}_{n,X}(h) := \frac{\sum_{t=1}^{n-h} \left(X_t - \overline{X}_n\right)\left(X_{t+h} - \overline{X}_n\right)}{\sum_{t=1}^{n} \left(X_t - \overline{X}_n\right)^2}\,, \tag{3.5}$$

where $\overline{X}_n = n^{-1}\sum_{t=1}^n X_t$ denotes the sample mean.

Remark 3.4 It follows at once that

$$\widetilde{\rho}_{n,X}(h) - \rho_X(h) = O_P([n/\log n]^{-1/\alpha})\,.$$

This rate of convergence to zero compares favourably with the slower rate, $O_P(n^{-1/2})$, for the difference $\widetilde{\rho}_{n,X}(h) - \rho_X(h)$ in the finite variance case.

Remark 3.5 If $EZ^2 < \infty$ and $EZ = 0$, a modification of Theorem 3.2 holds with the S_j's, $j \geq 1$, replaced by iid $N(0,1)$ random variables and S_0 by the constant 1. Notice that the structure of relation (3.4) is the reason for the so-called Bartlett formula; see Brockwell and Davis [13]. Thus (3.4) is an analogue to Bartlett's formula in the infinite variance case.

The proof of this result depends heavily on point process convergence results. However, in order to give some intuition for why $\widetilde{\rho}_{n,X}(h)$ is a consistent estimator of $\rho_X(h)$, consider the simplest case of a linear process as given by the MA(1) process

$$X_t = \theta Z_{t-1} + Z_t\,, \quad t \in \mathbb{Z}\,.$$

The limit behavior of the sample ACF is closely connected with the large sample behavior of the corresponding sample ACVF. Define

$$\widetilde{\gamma}_{n,X}(h) := \frac{1}{n}\sum_{t=1}^{n-h} X_t X_{t+h}\,, \quad h = 0, 1, \ldots\,, \tag{3.6}$$

and choose the sequences (a_n) and (b_n) such that

$$P(|Z| > a_n) \sim n^{-1} \quad \text{and} \quad P(|Z_1 Z_2| > b_n) \sim n^{-1}\,, \quad n \to \infty\,.$$

A simple calculation shows that $a_n \sim c_1 n^{1/\alpha}$ and $b_n \sim c_2 (n \log n)^{1/\alpha}$ for certain constants c_i, where we have made use of the fact that

$$P(Z_1 Z_2 > x) \sim c_3 x^{-\alpha} \log x \tag{3.7}$$

Now, a point process convergence result shows that

$$n\ \left(a_n^{-2}\widetilde{\gamma}_{n,Z}(0), b_n^{-1}\widetilde{\gamma}_{n,Z}(1), b_n^{-1}\widetilde{\gamma}_{n,Z}(2)\right) \stackrel{d}{\rightarrow} c_4\,(S_0, S_1, S_2)\ , \tag{3.8}$$

for some constant c_4, where S_0, S_1, S_2 are independent stable as described above. Now, consider the difference

$$\Delta_n := \widetilde{\rho}_{n,X}(1) - \rho_X(1) = \frac{\sum_{t=1}^n X_t X_{t-1} - \rho(1)\sum_{t=1}^n X_t^2}{\sum_{t=1}^n X_t^2}\,.$$

Recalling that $\rho_X(1) = \theta/(1+\theta^2)$ and (3.8), it is not difficult to see that

$$\begin{aligned}
&\left(\frac{n}{\log n}\right)^{1/\alpha}\ \Delta_n \\
&= \left(\frac{n}{\log n}\right)^{1/\alpha} \frac{\left(\left[\theta\sum_{t=1}^n Z_t^2 + (1+\theta^2)\sum_{t=1}^n Z_t Z_{t+1} + \theta\sum_{t=1}^n Z_{t-1}Z_{t+1}\right] - \theta\sum_{t=1}^n Z_t^2\right)}{(1+\theta^2)\sum_{t=1}^n Z_t^2} + o_P(1) \\
&= \left(\frac{n}{\log n}\right)^{1/\alpha} \frac{(1+\theta^2)\sum_{t=1}^{n-1} Z_t Z_{t+1} + \theta\sum_{t=1}^{n-2} Z_t Z_{t+2}}{(1+\theta^2)\sum_{t=1}^n Z_t^2} + o_P(1) \\
&\stackrel{d}{\rightarrow}\ S_0^{-1}\,[S_1 + \rho_X(1)\ S_2] = Y_1\,.
\end{aligned}$$

From this limit relation one can see that the consistency of the estimator $\widetilde{\rho}_{n,X}(1)$ is due to a special cancellation that eliminates all terms involving $\sum_{t=1}^n Z_t^2$ in the numerator which, otherwise, would determine the rate of convergence. Since the summands $Z_t Z_{t+1}$ and $Z_t Z_{t+2}$ have tails lighter than those of Z_t^2 (see (3.7)) the faster rate of convergence follows from (3.8) and the continuous mapping theorem.

Clearly, the cancellation effect described above is due to the particular structure of the linear process. For general stationary sequences (X_t) such extraordinary behavior cannot be expected. This will become clear in the following section.

Despite their flexibility for modeling tails and ACF behavior, linear processes are not considered good models for log-returns. Indeed, the sample ACF of the S&P index for the years 1961-1993 suggests that this log-return series might be well modelled by an MA(1) process. However the innovations from such a fitted model could not be iid since the sample ACF of the absolute log-returns and their squares (see Figure 1.1) suggest dependence well beyond lag 1. This kind of sample ACF behavior shows that the class of *standard linear models* are not appropriate for describing the dependence of log-return series and therefore various non-linear models have been proposed in the literature. In what follows, we focus on two standard models, the GARCH and the stochastic volatility processes. We investigate their tails and sample ACF behavior.

The latter two models are multiplicative noise models that have the form $X_t = \sigma_t\ Z_t$, where (σ_t) is referred to as the stochastic volatility process and is assumed to be independent of the noise

(Z_t). The sequence (Z_t) is often assumed to be iid with $EZ = 0$ and $EZ^2 = 1$. GARCH models take σ_t to be a function of the "past" of the process, whereas one specifies a stochastic model for (σ_t) in the case of a stochastic volatility model.

We start by investigating the GARCH model in the more general context of stochastic recurrence equations (SRE's).

4 Stochastic recurrence equations

4.1 Definition

In what follows, we consider processes which are given by a SRE of the form

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t\,, \quad t \in \mathbb{Z}\,, \tag{4.1}$$

where $((\mathbf{A}_t, \mathbf{B}_t))$ is an iid sequence ($\mathbf{A}_t$ and $\mathbf{B}_t$ can be dependent), the $\mathbf{A}_t$'s are $d \times d$ random matrices and the random vectors $\mathbf{B}_t$ assume values in $\mathbb{R}^d$. For ease of presentation, we often use the convention that $\mathbf{A} = \mathbf{A}_1$, $\mathbf{B} = \mathbf{B}_1$, $\mathbf{Y} = \mathbf{Y}_1$, etc.

Example 4.1 (ARCH(1) process)
An important example of a process (Y_t) satisfying (4.1) is given by the squares (X_t^2) of an ARCH(1) process (autoregressive conditionally heteroscedastic processes of order 1). It was introduced by Engle [24] as an econometric model for log-returns of speculative prices (foreign exchange rates, stock indices, share prices, etc.). Given non-negative parameters α_0 and α_1, (X_t) is defined as

$$X_t = \sigma_t\, Z_t\,, \quad t \in \mathbb{Z}\,, \tag{4.2}$$

where (Z_t) is an iid sequence, and

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2\,, \quad t \in \mathbb{Z}\,.$$

Clearly,

$$Y_t = X_t^2\,, \quad A_t = \alpha_1 Z_t^2\,, \quad B_t = \alpha_0 Z_t^2\,,$$

satisfy the SRE (4.1).

An ARCH process (X_t) of order p (ARCH(p)) is defined in an analogous way: it satisfies the equation (4.2) with σ_t^2 given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i\, X_{t-i}^2\,, \quad t \in \mathbb{Z}\,,$$

where $\alpha_i \geq 0$ are certain parameters and $\alpha_p \neq 0$. We would like to stress that, for $p > 1$, neither (X_t^2) nor (σ_t^2) can be given as solutions to a one-dimensional SRE but have to be embedded in a multivariate SRE; see Example 4.3 below.

Example 4.2 (GARCH(1,1) process)

Since the fit of ARCH processes to log-returns was not completely satisfactory (a good fit to real-life data requires a large number of parameters α_j), Bollerslev [6] introduced a more parsimonious family of models, the GARCH (generalised ARCH) processes. A GARCH(1,1) (GARCH of order (1,1)) process (X_t) is given by relation (4.2), where

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \,, \quad t \in \mathbb{Z}\,. \tag{4.3}$$

The process (X_t^2) cannot be written in the form (4.1) for one-dimensional Y_t's. However, an iteration of (4.3) yields

$$\sigma_t^2 \;=\; \alpha_0 + \alpha_1 \sigma_{t-1}^2 Z_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + \sigma_{t-1}^2[\alpha_1 Z_{t-1}^2 + \beta_1]\,,$$

and so the sequence (σ_t^2) satisfies (4.1) with

$$Y_t = \sigma_t^2\,, \quad A_t = \alpha_1 Z_{t-1}^2 + \beta_1\,, \quad B_t = \alpha_0\,, \quad t \in \mathbb{Z}\,.$$

The GARCH(1,1) model is capable of capturing the main distinguishing features of log-returns of financial assets and, as a result, has become one of the mainstays of econometric models. In addition to the model's flexibility in describing certain types of dependence structure, it is also able to model tail heaviness, a property often present in observed data. A critical discussion of the GARCH model, and the GARCH(1,1) in particular, is given in Mikosch and Stărică [35, 36].

Example 4.3 (GARCH(p,q) process)

A GARCH(p,q) process (GARCH of order (p,q)) is defined in a similar way. It is given by (4.2) with

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2\,, \quad t \in \mathbb{Z}\,, \tag{4.4}$$

where the integers $p, q \geq 0$ determine the order of the process. Write

$$\mathbf{Y}_t = (X_t^2, \ldots, X_{t-p+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+1}^2)'\,. \tag{4.5}$$

This process satisfies (4.1) with matrix-valued $\mathbf{A}_t$'s and vector-valued $\mathbf{B}_t$'s:

$$\mathbf{A}_t \;=\; \begin{pmatrix} \alpha_1 Z_t^2 & \cdots & \alpha_{p-1} Z_t^2 & \alpha_p Z_t^2 & \beta_1 Z_t^2 & \cdots & \beta_{q-1} Z_t^2 & \beta_q Z_t^2 \\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_1 & \cdots & \alpha_{p-1} & \alpha_p & \beta_1 & \cdots & \beta_{q-1} & \beta_q \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{4.6}$$

(4.7) $$\mathbf{B}_t = (\alpha_0 Z_t^2, 0, \ldots, 0, \alpha_0, 0 \ldots, 0)' .$$

Example 4.4 (The simple bilinear process)
The simple bilinear process

$$X_t = aX_{t-1} + bX_{t-1}Z_{t-1} + Z_t , \quad t \in \mathbb{Z} ,$$

for positive a, b and an iid sequence (Z_t) can be embedded in the framework of a SRE of type (4.1). Indeed, notice that $X_t = Y_{t-1} + Z_t$, where (Y_t) satisfies (4.1) with

$$A_t = a + bZ_t \quad \text{and} \quad B_t = A_t Z_t .$$

This kind of process has been treated in Basrak et al. [3].

One of the crucial problems is to find conditions for the existence of a strictly stationary solution to (4.1). These conditions have been studied for a long time, even under less restrictive assumptions than $((\mathbf{A}_t, \mathbf{B}_t))$ being iid; see for example Brandt [9], Kesten [33], Vervaat [41], Bougerol and Picard [7]. The following result gives some conditions which are close to necessity; see Babillot et al. [2].

Recall the notion of operator norm of a matrix $\mathbf{A}$ with respect to a given norm $|\cdot|$:

$$\|\mathbf{A}\| = \sup_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}| .$$

For an iid sequence $(\mathbf{A}_n)$ of iid $d \times d$ matrices,

(4.8) $$\gamma = \inf \left\{ \frac{1}{n} E \log \|\mathbf{A}_1 \cdots \mathbf{A}_n\| , \quad n \in \mathbb{N} \right\}$$

is called the *top Lyapunov exponent associated with* $(\mathbf{A}_n)$. If $E \log^+ \|\mathbf{A}_1\| < \infty$, it can be shown (see Furstenberg and Kesten [26]) that

(4.9) $$\gamma = \lim_{n \to \infty} \frac{1}{n} \log \|\mathbf{A}_1 \cdots \mathbf{A}_n\| \quad \text{a.s.}$$

With a few exceptions (including the ARCH(1,1) and GARCH(1,1) cases) one cannot calculate γ explicitly.

Theorem 4.5 *Assume $E \log^+ \|\mathbf{A}_1\| < \infty$, $E \log^+ |\mathbf{B}_1| < \infty$ and $\gamma < 0$. Then the series*

(4.10) $$\mathbf{Y}_n = \mathbf{B}_n + \sum_{k=1}^{\infty} \mathbf{A}_n \cdots \mathbf{A}_{n-k+1} \mathbf{B}_{n-k}$$

converges a.s., and the so-defined process $(\mathbf{Y}_n)$ is the unique causal strictly stationary solution of (4.1).

Notice that $\gamma < 0$ holds if $E \log \|\mathbf{A}_1\| < 0$. The condition on γ in Theorem 4.5 is particularly simple in the case $d = 1$ since then

$$\frac{1}{n} E \log |A_1 \cdots A_n| = E \log |A_1| = \gamma \,.$$

Corollary 4.6 *Assume $d = 1$, $-\infty \le E \log |A_1| < 0$ and $E \log^+ |B_1| < \infty$. Then the unique stationary solution of* (4.1) *is given by* (4.10).

Example 4.7 (Conditions for stationarity)

1) The process (σ_t^2) of an ARCH(1) process has a stationary version if $\alpha_0 > 0$ and $E \log(\alpha_1 Z^2) < 0$. If Z is $N(0,1)$, one can choose a positive $\alpha_1 < 2\mathrm{e}^{\gamma_0} \approx 3.568...$, where γ_0 is Euler's constant. See Goldie [29]; cf. Section 8.4 in Embrechts et al. [23]. Notice that the stationarity of (σ_t^2) also implies the stationarity of the ARCH(1) process (X_t).

2) The process (σ_t^2) of a GARCH(1,1) process has a stationary version if $\alpha_0 > 0$ and $E \log(\alpha_1 Z^2 + \beta_1) < 0$. Also in this case, stationarity of (σ_t^2) implies stationarity of the GARCH(1,1) process (X_t).

We mention at this point that it is very difficult to make any statements about the stationarity of solutions to general SRE's and GARCH(p,q) processes in particular. For general GARCH(p,q) processes, precise necessary and sufficient conditions for $\gamma < 0$ in terms of explicit and calculable conditions on the the parameters α_j, β_k and the distribution of Z are not known; see Bougerol and Picard [8] for the most general sufficient conditions which amount to certain restrictions on the distribution of Z, the following assumptions on the parameters

$$\alpha_0 > 0 \quad \text{and} \quad \sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k \le 1 \,, \tag{4.11}$$

and some further technical conditions. We also mention that the X_t's have a second finite moment if $EZ = 0$, $EZ^2 = 1$ and one has strict inequality in (4.11). See Davis et al. [16] for further discussion and details. In the latter reference it is mentioned that the case of multivariate GARCH processes could be treated in an analogous way, but the theoretical difficulties are then even more significant.

4.2 Tails

Recall the definition of multivariate regular variation from Section 2.1. It is quite surprising that the stationary solutions to SRE's have finite-dimensional distributions with multivariate regularly varying tails under very general conditions on $((\mathbf{A}_t, \mathbf{B}_t))$. This is due to a deep result on the renewal theory of products of random matrices given by Kesten [33] in the case $d \ge 1$. The one-dimensional case was considered by Goldie [29]. We state a modification of Kesten's fundamental result (Theorems 3 and 4 in [33]; the formulation of Theorem 4.8 below is taken from Davis et al. [16]). In these results, $\|\cdot\|$ denotes the operator norm defined in terms of the Euclidean norm $|\cdot|$.

Theorem 4.8 *Let $(\mathbf{A}_n)$ be an iid sequence of $d \times d$ matrices with non-negative entries satisfying:*

- *For some $\epsilon > 0$, $E\|\mathbf{A}\|^\epsilon < 1$.*
- *$\mathbf{A}$ has no zero rows a.s.*
- *The group generated by*

  (4.12) $$\{\log \rho(\mathbf{a}_n \cdots \mathbf{a}_1) : \mathbf{a}_n \cdots \mathbf{a}_1 > 0 \text{ for some } n \text{ and } \mathbf{a}_i \in \operatorname{supp}(P_{\mathbf{A}})\}$$

  is dense in $\mathbb{R}$, where $\rho(\mathbf{C})$ is the spectral radius of the matrix $\mathbf{C}$, $\mathbf{C} > 0$ means that all entries of this matrix are positive, $P_{\mathbf{A}}$ is the distribution of $\mathbf{A}$, and $\operatorname{supp}(P_{\mathbf{A}})$ its support.
- *There exists a $\kappa_0 > 0$ such that*

  (4.13) $$E\left(\min_{i=1,\ldots,d} \sum_{j=1}^{d} A_{ij}\right)^{\kappa_0} \geq d^{\kappa_0/2}$$

  and

  (4.14) $$E\left(\|\mathbf{A}\|^{\kappa_0} \log^+ \|\mathbf{A}\|\right) < \infty\,.$$

Then there exists a unique solution $\kappa_1 \in (0, \kappa_0]$ to the equation

(4.15) $$0 = \lim_{n\to\infty} \frac{1}{n} \log E\|\mathbf{A}_n \cdots \mathbf{A}_1\|^{\kappa_1}\,.$$

If $(\mathbf{Y}_n)$ is the stationary solution to the SRE in (4.1) with coefficient matrices $(\mathbf{A}_n)$ satisfying the above conditions and $\mathbf{B}$ has non-negative entries with $E|\mathbf{B}|^{\kappa_1} < \infty$, then $\mathbf{Y}$ is regularly varying with index κ_1. Moreover, the finite-dimensional distributions of the stationary solution $(\mathbf{Y}_t)$ of (4.1) are regularly varying with index κ_1.

A combination of the general results for SRE's (Theorems 4.5 and 4.8) specified to GARCH(p, q) processes yields the following result which is given in Davis et al. [16].

Theorem 4.9 *Consider the SRE (4.1) with $\mathbf{Y}_t$ given by (4.5), $\mathbf{A}_t$ by (4.6) and $\mathbf{B}_t$ by (4.7).*

(A) (Existence of stationary solution)
Assume that the following condition holds:

(4.16) $$E \log^+ |Z| < \infty \text{ and the Lyapunov exponent } \gamma < 0.$$

Then there exists a unique causal stationary solution of the SRE (4.1).

(B) (Regular variation of the finite-dimensional distributions)
Let $|\cdot|$ denote the Euclidean norm and $\|\cdot\|$ the corresponding operator norm. In addition to the Lyapunov exponent γ (see (4.8)) being less than 0, assume the following conditions:

1. *Z has a positive density on $\mathbb{R}$ such that either $E|Z|^h < \infty$ for all $h > 0$ or $E|Z|^{h_0} = \infty$ for some $h_0 > 0$ and $E|Z|^h < \infty$ for $0 \leq h < h_0$.*

2. *Not all of the parameters α_j and β_k vanish.*

Then there exists a $\kappa_1 > 0$ such that $\mathbf{Y}$ is regularly varying with index κ_1.

A consequence of the theorem is the following:

Corollary 4.10 *Let (X_t) be a stationary GARCH(p, q) process. Assume the conditions of part B of Theorem 4.9 hold. Then there exists a $\kappa > 0$ such that the finite-dimensional distributions of the process $((\sigma_t, X_t))$ are regularly varying with index κ.*

Example 4.11 (ARCH(1) and GARCH(1,1))
For these two models we can give an explicit equation for the value of κ. Indeed, (4.15) for $d = 1$ degenerates to $E|A|^{\kappa_1} = 1$. Recall from Example 4.1 that $A_t = \alpha_1 Z_t^2$. Hence the tail index κ of X is given by the solution to the equation $E(\alpha_1 Z^2)^{\kappa/2} = 1$. Similarly, in the GARCH(1,1) case of Example 4.2 we have $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$ which gives the tail index κ for σ by solving $E(\alpha_1 Z^2 + \beta_1)^{\kappa/2} = 1$. Then, by Breiman's results (2.3) it follows that

$$P(|X| > x) = P(|Z|\sigma > x) \sim \text{const } P(\sigma > x) \sim \text{const } x^{-\kappa} .$$

Unfortunately, these are the *only* two cases where one can give an explicit formula for κ in terms of the parameters of the GARCH process and the distribution of the noise.

The above results show that there is quite an intriguing relation between the parameters of a GARCH(p, q) process, the distribution of the noise (Z_t) and the tails of the process. In particular, it is rather surprising that the finite-dimensional distributions are regularly varying. Indeed, although the input noise (Z_t) may have light tails (exponential, normal) the resulting output (X_t) has Pareto-like tails. This is completely different from the linear process case where we discovered that the tails and the ACF behavior are due to totally different sources: the coefficients ψ_j and the tails of the noise. In the GARCH(p, q) case the parameters α_j, β_k and the whole distribution of Z, not only its tails, contribute to the heavy tailedness of marginal distribution of the process.

The squares of a GARCH(p, q) process can be written as the solution to an ARMA equation with a martingale difference sequence as noise *provided the second moment of X_t is finite.* However, the analogy between an ARMA and GARCH process can be quite misleading especially when discussing conditions for stationarity and the tail behavior of the marginal distribution. The source of the heavy tails of GARCH process does not come directly from the martingale difference sequence, but rather the nonlinear mechanism that connects the output with the input.

The interaction between the parameters of the GARCH(p, q) process and its tails is illustrated in the form of the invariant distribution of the process which contains products of the matrices

$\mathbf{A}_t$ in front of the "noise" $\mathbf{B}_t$ (see (4.10)). This is in contrast to a linear process (3.1) where the coefficents in front of the innovations Z_t are constants. Notice that it is the presence of sums of products of an increasing number of $\mathbf{A}_t$'s which causes the heavy tails of the distribution of $\mathbf{Y}_t$. For example, if one assumes that (Z_t) is iid Gaussian noise in the definition of a GARCH(p,q) process and considers the corresponding $\mathbf{Y}_t$'s, $\mathbf{A}_t$'s and $\mathbf{B}_t$'s (see (4.5)–(4.7)), then it is readily seen that a truncation of the infinite series (4.10) yields a random variable which has all finite power moments.

The interaction between the tails and dependence structure, in particular the non-linearity of the process, is also responsible for the sample ACF behavior of solutions to SRE's. In contrast to the linear process case of Section 3.3, we show in the next section that the cancellation effect which was explained in Section 3.3 does not occur for this class of processes. This fact makes the limit theory of the sample ACF for such processes more difficult to study.

4.3 Limit theory for the sample ACF

The limit theory for the sample ACF, ACVF and cross-correlations of solutions to SRE's heavily depends on point process techniques. We refrain here from discussing those methods and refer to Davis et al. [16] for details. As mentioned earlier, because of the non-linearity of the processes, we cannot expect that a theory analogous to linear processes holds, in particular we may expect complications for the sample ACF behavior if the tail index of the marginal distribution is small. This is the content of the following results.

We start with the sample autocovariances of the first component process (Y_t) say of $(\mathbf{Y}_t)$; the case of sample cross-correlations and the joint limits for the sample autocorrelations of different component processes can be derived as well.

Recall the definition of the sample ACVF $\widetilde{\gamma}_{n,Y}$ from (3.6) and the corresponding sample ACF from (3.3). We also write

$$\gamma_Y(h) = EY_0Y_h \quad \text{and} \quad \rho_Y(h) = \gamma_Y(h)/\gamma_Y(0)\,, \quad h \geq 0\,,$$

for the autocovariances and autocorrelations, respectively, of the sequence (Y_t), when these quantities exist. Also recall the notion of an infinite variance stable random vector from Section 2.2.

Theorem 4.12 *Assume that $(\mathbf{Y}_t)$ is a solution to (4.1) satisfying the conditions of Theorem 4.8.*

(1) If $\kappa_1 \in (0,2)$, then

$$\left(n^{1-2/\kappa_1}\gamma_{n,Y}(h)\right)_{h=0,\ldots,m} \stackrel{d}{\rightarrow} (V_h)_{h=0,\ldots,m}\,,$$

$$(\rho_{n,Y}(h))_{h=1,\ldots,m} \stackrel{d}{\rightarrow} (V_h/V_0)_{h=1,\ldots,m}\,,$$

where the vector $(V_0,\ldots,V_m)$ is jointly $\kappa_1/2$-stable in $\mathbb{R}^{m+1}$.

(2) If $\kappa_1 \in (2,4)$ and for $h = 0, \ldots, m$,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \operatorname{var} \left(n^{-2/\kappa_1} \sum_{t=1}^{n-h} Y_t Y_{t+h} I_{\{|Y_t Y_{t+h}| \le a_n^2 \epsilon\}} \right) = 0 \,, \tag{4.17}$$

then

$$\left(n^{1-2/\kappa_1} (\gamma_{n,Y}(h) - \gamma_Y(h)) \right)_{h=0,\ldots,m} \stackrel{d}{\to} (V_h)_{h=0,\ldots,m} \,, \tag{4.18}$$

$$\left(n^{1-2/\kappa_1} (\rho_{n,X}(h) - \rho_X(h)) \right)_{h=1,\ldots,m} \stackrel{d}{\to} \gamma_X^{-1}(0) \, (V_h - \rho_X(h) \, V_0)_{h=1,\ldots,m} \,, \tag{4.19}$$

where $(V_0, \ldots, V_m)$ is jointly $\kappa_1/2$-stable in $\mathbb{R}^{m+1}$.

(3) If $\kappa_1 > 4$ then (4.18) and (4.19) hold with normalization $n^{1/2}$, where $(V_1, \ldots, V_m)$ is multivariate normal with mean zero and covariance matrix $[\sum_{k=-\infty}^{\infty} \operatorname{cov}(Y_0 Y_i, Y_k Y_{k+j})]_{i,j=1,\ldots,m}$ and $V_0 = E(Y_0^2)$.

The limit random vectors in parts (1) and (2) of the theorem can be expressed in terms of the limiting points of appropriate point processes. For more details, see Davis and Mikosch [14] where the proofs of (1) and (2) are provided and also Davis et al. [16]. Part (3) follows from a standard central limit theorem for strongly mixing sequences; see for example Doukhan [21].

The distributional limits of the sample ACF and ACVF of GARCH(p,q) processes (X_t) do not follow directly from Theorem 4.12 since only the the squares of the process satisfy the SRE (4.1). However, an application of the point process convergence in Davis and Mikosch [14] guarantees that similar results can be proved for the processes (X_t), $(|X_t|)$ and (X_t^2) or any power $(|X_t|^p)$ for some $p > 0$. The limit results of Theorem 4.12 remain qualitatively the same for $Y_t = X_t, |X_t|, X_t^2, \ldots$, but the parameters of the limiting stable laws have to be changed. See Davis et al. [16] for details.

Theorems 3.2 and 4.12 demonstrate quite clearly the differences between the limiting behavior of the ACF for linear and non-linear processes. In the linear case, the rate of convergence as determined by the normalising constants is faster the heavier the tails. In the nonlinear case, the rate of convergence of the sample ACF to their deterministic counterpart is slower the heavier the tails, and if the underlying time series has infinite variance, the sample autotcorrelations have non-degenerate limit laws.

Since it is generally believed that log-returns have heavy tails in the sense that they are Pareto-like with tail parameter between 2 and 5 (see for example Müller et al. [37] or Embrechts et al. [23], in particular Chapters 6 and 7), Theorem 4.12 indicates that the sample ACF of such data has to be treated with some care because it could mean nothing or that the classical $\pm 1.96/\sqrt{n}$ confidence bands are totally misleading. Clearly, for GARCH processes the form of the limit distribution and the growth of the scaling constants of the sample ACF depend critically on the values of the model's parameters. We will see in the next section that the sample ACF of stochastic volatility models behaves quite differently. Its limiting behavior is more in line with that for a linear process.

5 Stochastic volatility models

5.1 Definition

As evident from the preceding discussion, the theoretical development of the basic probabilistic properties of GARCH processes is thorny: conditions for stationarity are difficult to formulate and verify, the tail behavior is complicated and little is known about the dependence structure. On the other hand, estimation for GARCH processes is relatively easy by using conditional maximum likelihood based on the iid assumption of the noise; see for example Gourieroux [30] and the references therein. The latter property is certainly one of the attractions for this kind of model and has contributed to its popularity.

Over the last few years, another kind of econometric time series has attracted some attention: the *stochastic volatility processes*. Like GARCH models, these processes are multiplicative noise models, i.e.

$$X_t = \sigma_t \, Z_t \,, \quad t \in \mathbb{Z} \,, \tag{5.1}$$

where (Z_t) is an iid sequence of random variables which is completely independent of another strictly stationary sequence (σ_t) of non-negative random variables. The independence of the two sequences (Z_t) and (σ_t) allows one to easily derive the basic probabilistic properties of stochastic volatility processes. For example, the dependence structure of the process is determined via the dependence in the volatility sequence (σ_t). For our purposes, we shall assume that

$$\sigma_t = \mathrm{e}^{\,Y_t} \,, \quad t \in \mathbb{Z} \,, \tag{5.2}$$

where (Y_t) is a linear process

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \,, \quad t \in \mathbb{Z} \,, \tag{5.3}$$

with coefficients ψ_j satisfying

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty \tag{5.4}$$

and an iid noise sequence (ε_t). For ease of presentation we assume that ε is $N(0,1)$ which, together with (5.4), ensures that the defining sum for Y_t in (5.3) is convergent a.s. The condition of Gaussianity of the ε_t's can be relaxed at the cost of more technical conditions, see Davis and Mikosch [15] for details.

Notice that the assumption (5.4) is the weakest possible; it allows one to use any non-deterministic Gaussian stationary time series as a model for (Y_t), in particular one can choose (Y_t) as a stationary ARMA or a FARIMA process for modeling any kind of long or short range dependence in (Y_t); see Brockwell and Davis [13] for an extensive discussion of ARMA and FARIMA processes and

Samorodnitsky and Taqqu [40] for a discussion and mathematical aspects of long range dependence. Hence one can achieve any kind of ACF behavior in (σ_t) as well as in (X_t) (due to the independence of (Y_t) and (ε_t)). This latter property gives the stochastic volatility models a certain advantage over the GARCH models. The latter are strongly mixing with geometric rate under very general assumptions on the paramaters and the noise sequence; see Davis et al. [16] for details. As a consequence of the mixing property, if the ACF of these processes is well defined, it decays to zero at an exponential rate, hence long range dependence effects (in the sense that the ACF is not absolutely summable) cannot be achieved for a GARCH process or any of its powers. Since it is believed in parts of the econometrics community that log-return series might exhibit long range dependence, the stochastic volatility models are quite flexible for modelling this behavior; see for example Breidt et al. [11].

In what follows, we show that the tails and the sample ACF of these models also have more attractive properties than the GARCH models even in the infinite second and fourth moment cases. Ultimately, it is hoped that the dichotomy in behavior of the ACF for GARCH and stochastic volatility models will be of practical significance in choosing between the two models. This will be the subject of future research. In contrast to the well behaved properties of the ACF, estimation of the parameters in stochastic volatility models tends to be more complicated than that for GARCH processes. Often one needs to resort to simulation based methods to calculate efficient estimates. Quasi-likelihood estimation approaches are discussed in Harvey et al. [28], Breidt and Carriquiry [10], and Breidt et al. [11] while simulation based methods can be found in Jacquier, Polson, and Rossi [31] and Durbin and Koopman [22].

5.2 Tails

By virtue of Breiman's result (2.3), we know that

$$P(X > x) \sim E\sigma^\alpha\, P(Z > x) \quad \text{and} \quad P(X > x) \sim E\sigma^\alpha\, P(Z \le -x)\,, \quad x \to \infty\,, \tag{5.5}$$

provided $E\sigma^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$ and Z is regularly varying with index $\alpha > 0$ and tail balancing condition

$$P(Z > x) = p\, x^{-\alpha}\, L(x) \quad \text{and} \quad P(Z > x) = q\, x^{-\alpha}\, L(x)\,, \tag{5.6}$$

where L is slowly varying and $p + q = 1$ for some $p \in [0,1]$. In what follows, we assume that (5.6) holds, and we also require

$$E|Z|^\alpha = \infty\,. \tag{5.7}$$

Then $Z_1 Z_2$ is also regularly varying with index α satisfying (see equations (3.2) and (3.3) in Davis and Resnick [19])

$$\frac{P(Z_1 Z_2 > x)}{P(|Z_1 Z_2| > x)} \to \widetilde{p} := p^2 + (1-p)^2 \qquad \text{as } x \to \infty\,. \tag{5.8}$$

Another application of (2.3) implies that $X_1 X_h$ is regularly varying with index α:

$$(5.9) \qquad \begin{cases} P(X_1X_h > x) & = P(Z_1Z_2\sigma_1\sigma_h > x) \quad \sim E[\sigma_1\sigma_h]^\alpha \, P(Z_1Z_2 > x)\,, \\ P(X_1X_h \le -x) & = P(Z_1Z_2\sigma_1\sigma_h \le -x) \quad \sim E[\sigma_1\sigma_h]^\alpha \, P(Z_1Z_2 \le -x)\,, \end{cases}$$

provided $E[\sigma_1\sigma_h]^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$. Since we assumed the exponential structure (5.2) for the σ_t's and that the Y_t's are Gaussian, the σ_t's are log-normal and therefore the latter moment condition holds for every $\epsilon > 0$.

An application of a multivariate version of Breiman's result (see the Appendix in Davis et al. [16]) ensures that the finite-dimensional distributions of (X_t) are regularly varying with the same index α. We refrain from giving details.

5.3 Limit theory for the sample ACF

In order to describe the limiting behavior of the sample ACF of a stochastic volatility process in the heavy-tailed case, two sequences of constants (a_n) and (b_n) which figure into the normalizing constants must be defined. Specifically, let (a_n) and (b_n) be the respective $(1-n^{-1})$-quantiles of $|Z_1|$ and $|Z_1Z_2|$ defined by

$$(5.10) \qquad a_n = \inf\{x : P(|Z_1| > x) \le n^{-1}\} \quad \text{and} \quad b_n = \inf\{x : P(|Z_1Z_2| > x) \le n^{-1}\}\,.$$

Using point process techniques and arguments similar to the ones given in [19], the weak limit behavior for the sample ACF can be derived for stochastic volatility processes. These results are summarized in the following theorem.

Theorem 5.1 *Assume (X_t) is the stochastic volatility process satisfying (5.1)–(5.3) where Z satisfies conditions (5.6) and (5.7). Let $\widetilde{\gamma}_{n,X}(h)$ and $\widetilde{\rho}_{n,X}(h)$ denote the sample ACVF and ACF of the process as defined in (3.6) and (3.3) and assume that either*

(i) $\alpha \in (0,1)$,

(ii) $\alpha = 1$ *and Z_1 has a symmetric distribution,*

or

(iii) $\alpha \in (1,2)$ *and Z_1 has mean 0.*

Then

$$n\, \left(a_n^{-2}\widetilde{\gamma}_{n,X}(0), b_n^{-1}\widetilde{\gamma}_{n,X}(1), \ldots, b_n^{-1}\widetilde{\gamma}_{n,X}(r)\right) \quad \stackrel{d}{\to} \quad (V_h)_{h=0,\ldots,r}\,,$$

where $(V_0, V_1, \ldots, V_r)$ are independent random variables, V_0 is a non-negative stable random variable with exponent $\alpha/2$ and $V_1, \ldots, V_r$ are identically distributed as stable with exponent α. In addition, we have for all three cases that,

$$\left(a_n^2 b_n^{-1}\widetilde{\rho}_{n,X}(h)\right)_{h=1,\ldots,r} \quad \stackrel{d}{\to} \quad (V_h/V_0)_{h=1,\ldots,r}\,.$$

Remark 5.2 By choosing the volatility process (σ_t) to be identically 1, we can recover the limiting results obtained in Davis and Resnick [19] for the autocovariances and autocorrelations of the (Z_t) process. If $(S_0, S_1, \ldots, S_r)$ denotes the limit random vector of the sample autocovariances based on (Z_t), then there is an interesting relationship between S_k and V_k, namely,

$$(V_0, V_1, \ldots, V_r) \stackrel{d}{=} (\|\sigma_1\|_\alpha^2 S_0, \|\sigma_1\sigma_2\|_\alpha S_1, \ldots, \|\sigma_1\sigma_{1+r}\|_\alpha S_r),$$

where $\|\cdot\|_\alpha$ denotes the L_α-norm. It follows that

$$\left(a_n^2 b_n^{-1} \widetilde{\rho}_{n,X}(h)\right)_{h=1,\ldots,r} \stackrel{d}{\to} \left(\frac{\|\sigma_1\sigma_{h+1}\|_\alpha}{\|\sigma_1\|_\alpha^2} \frac{S_h}{S_0}\right)_{h=1,\ldots,r}.$$

Remark 5.3 The conclusion of (iii) of the theorem remains valid if $\widetilde{\rho}_{n,X}(h)$ is replaced by the mean-corrected version of the ACF given by (3.5).

5.3.1 Other powers

It is also possible to investigate the sample ACVF and ACF of the processes $(|X_t|^\delta)$ for any power $\delta > 0$. We restrict ourselves to the case $\delta = 1$ in order to illustrate the method.

Notice that $|X_t| = |Z_t|\sigma_t$, $t = 1, 2, \ldots$, has a structure similar to the original process (X_t). Hence Theorem 5.1 applies directly to the ACF of $|X_t|$ when $\alpha < 1$ and to the ACF of the stochastic volatility model with noise $|Z_t| - E|Z_t|$ when $\alpha \in (1, 2)$.

In order to remove the centering of the noise in the $\alpha \in (1, 2)$ case, we use the following decomposition for $h \geq 1$ with $\gamma_{|X|} = E|X_0 X_k|$, $\widetilde{Z}_t = |Z_t| - E|Z|$ and $\widetilde{X}_t = \widetilde{Z}_t\sigma_t$:

$$\begin{aligned} n\left(\widetilde{\gamma}_{n,|X|}(h) - \widetilde{\gamma}_{|X|}(h)\right) &= \sum_{t=1}^{n-h} \widetilde{Z}_t\widetilde{Z}_{t+h}\sigma_t\sigma_{t+h} + E|Z| \sum_{t=1}^{n-h} \widetilde{Z}_t\sigma_t\sigma_{t+h} + E|Z| \sum_{t=1}^{n-h} \widetilde{Z}_{t+h}\sigma_t\sigma_{t+h} \\ &\quad -(E|Z|)^2 \sum_{t=1}^{n-h} (\sigma_t\sigma_{t+h} - E\sigma_0\sigma_h) \\ &= I_1 + I_2 + I_3 + I_4\,. \end{aligned}$$

Since

$$n^{-1} I_1 = \widetilde{\gamma}_{n,\widetilde{X}}(h)\,,$$

and $E\widetilde{Z} = 0$, Theorem 5.1 (iii) is directly applicable to $(\widetilde{X}_t)$. Also notice that $n a_n^{-2} \widetilde{\gamma}_{n,|X|}(0)$ converges weakly to an $\alpha/2$-stable distribution, for the same reasons as given for (X_t). It remains to show that

$$b_n^{-1} I_j \stackrel{P}{\to} 0\,, \quad j = 2, 3, 4\,. \tag{5.11}$$

Point process arguments can be used to show that $a_n^{-1} I_j$, $j = 2, 3$ converge to an α-stable distribution, and since $a_n/b_n \to 0$, (5.11) holds for $j = 2, 3$. It is straightforward to show that

$\text{var}(b_n^{-1} I_4) \to 0$ for cases when the linear process in (5.3) has absolutely summable coefficients or when the coefficients are given by a fractionally integrated model. Thus $b_n^{-1} I_4 \xrightarrow{P} 0$ and the limit law for $\widetilde{\gamma}_{n,|X|}(h)$ is as specified in Theorem 5.1.

5.3.2 A brief simulation comparison

To illustrate the differences in the asymptotic theory for the ACF of GARCH and stochastic volatility models, a small simulation experiment was conducted. One thousand replicates of lengths 10,000 and 100,000 were generated from a GARCH(1,1) time series model with parameter values in the stochastic volatility recursion (4.3) given by

$$\alpha_0 = 8.6 \times 10^{-6}, \quad \alpha_1 = .110 \quad \text{and} \quad \beta_1 = .698 \,.$$

The noise was generated from a student's t distribution with 4 degrees of freedom—normalized to have variance 1. With this choice of parameter values and noise distribution the marginal distribution has Pareto tails with approximate exponent 3. The sampling behavior of the ACF of the data and its squares are depicted using box plots in Figures 5.4(a),(c) and 5.5(a),(c) for samples of size 10,000 and 100,000. As seen in these figures, the sample ACF of the data appears to be converging to 0 as the sample size increases (note the differences in the magnitude of the vertical scaling on the graphs). On the other hand, the sampling distributions for the ACF of the squares (Figure 5.5(a),(c)) appear to be the same for the two samples sizes $n = 10,000$ and $100,000$ reflecting the limit theory as specified by Theorem 4.12 (1). These box plots can be interpreted as estimates of the limiting distributions of the ACF.

Samples paths of sizes $10,000$ and $100,000$ were also generated from a stochastic volatility model, where the stochastic volatility process (σ_t) satisfies the model

$$\log \sigma_t = .85 \log \sigma_{t-1} + \epsilon_t \,,$$

(ϵ_t) is a sequence of iid $\text{N}(0,1)$ random variables, and the noise (Z_t) was taken to be iid with a t-distribution on 3 degrees of freedom. (The noise was normalized to have variance 1.) We chose this noise distribution in order to match the tail behavior of the GARCH(1,1) process. The sampling behavior of the ACF of the data and its squares are shown in Figures 5.4(b),(d) and 5.5(b),(d). These plots demonstrate the *weak consistency* of the estimates to zero as n increase. Notice that the autocorrelations of the squares are more concentrated around 0 as predicted by the theory. Finally, these graphs illustrate the generally good performance of the sample ACF of stochastic volatility models in terms of convergence in probability to their *population* counterparts—especially when compared to the behavior found in the GARCH model.

Figure 5.4 *Boxplots based on* 1000 *replications of the sample ACF* (*at lags* 1 *to* 20) *for data generated from a* GARCH(1,1) *model and from a stochastic volatility model.* (*a*) *GARCH model with sample size* $n = 10,000$*;* (*b*) *stochastic volatility model with* $n = 10,000$*;* (*c*) *GARCH model with* $n = 100,000$*; and* (*d*) *stochastic volatility model with* $n = 100,000$.

Figure 5.5 *Boxplots based on* 1000 *replications of the sample ACF* (*at lags* 1 *to* 20) *of the squares from a* GARCH(1,1) *model and from a stochastic volatility model.* (*a*) *GARCH model with sample size* $n = 10,000$*;* (*b*) *stochastic volatility model with* $n = 10,000$*;* (*c*) *GARCH model with* $n = 100,000$*; and* (*d*) *stochastic volatility model with* $n = 100,000$.

References

[1] Adler, R., Feldman, R. and Taqqu, M.S. (Eds.) (1998) *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy-Tailed Distributions.* Birkhäuser, Boston.

[2] Babillot, M., Bougerol, P. and Elie, L. (1997) The random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. *Ann. Probab.* **25** 478–493.

[3] Basrak, B., Davis, R.A. and Mikosch, T. (1999) The sample ACF of a simple bilinear process. *Stoch. Proc. Appl.* **83** (1999), 1–14.

[4] Billingsley, P. (1968) *Convergence of Probability Measures* Wiley, New York.

[5] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge.

[6] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

[7] Bougerol, P. and Picard, N. (1992a) Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.

[8] Bougerol, P. and Picard, N. (1992b) Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.

[9] Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.* **18**, 211–220.

[10] Breidt, F.J. and Carriquiry, A.L. (1996) Improved quasi-maximum likelihood estimation for stochastic volatility models. In: *Modelling and Prediction. Proceedings Hsinchu, 1994,* pp. 228–247. Springer, New York.

[11] Breidt, F.J., Crato, N. and Lima, P. de (1998) The detection and estimation of long memory in stochastic volatility. *J. Econom.* **83**, 325–348.

[12] Breiman, L. (1965) On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10**, 323–331.

[13] Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition. Springer, New York.

[14] Davis, R.A. and Mikosch, T. (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049–2080.

[15] Davis, R.A. and Mikosch, T. (2000) The sample ACF of stochastic volatility models. In preparation.

[16] Davis, R.A., Mikosch, T. and Basrak, B. (1999) Limit theory for the sample autocorrelations of solutions to stochastic recurrence equations with applications to GARCH processes. Preprint. Available at www.math.rug.nl/∼mikosch.

[17] Davis, R.A. and Resnick, S.I. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13**, 179–195.

[18] Davis, R.A. and Resnick, S.I. (1985) More limit theory for the sample correlation function of moving averages. *Stoch. Proc. Appl.* **20**, 257–279.

[19] Davis, R.A. and Resnick, S.I. (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533–558.

[20] Davis, R.A. and Resnick, S.I. (1988) Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution. *Stoch. Proc. Appl.* **30**, 41–68.

[21] Doukhan, P. (1994) *Mixing. Properties and Examples.* Lecture Notes in Statistics **85**. Springer Verlag, New York.

[22] Durbin, J. and Koopman, S. J. (1997) Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**, 669–684.

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

[24] Engle, R.F. (1982) Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

[25] Engle, R.F. and Bollerslev, T. (1986) Modelling the persistence of conditional variances. With comments and a reply by the authors. *Econometric Rev.* **5**, 1–87.

[26] Furstenberg, H. and Kesten, H. (1960) Products of random matrices. *Ann. Math. Statist.* **31**, 457–469.

[27] Haan, L. de and Resnick, S.I. (1977) Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **40**, 317–337.

[28] Harvey, A.C., Ruiz, E., Shephard, N. (1994) Mulitvariate stochastic variance models. *Review of Economic Studies* **61**, 247–264.

[29] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126–166.

[30] Gouriéroux, C. (1997) ARCH Models and Financial Applications. Springer Series in Statistics. Springer-Verlag, New York.

[31] Jacquier, E., Polson, N.G., and Rossi, P.E. (1994) Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Econom. Statist.* **12**, 371–417.

[32] Kallenberg, O. (1983) *Random Measures,* 3rd edition. Akademie–Verlag, Berlin.

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

[34] Mikosch, T. and Samorodnitsky, G. (2000): The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Appl. Probab.*, to appear. Available as walk.ps at www.orie.cornell.edu/~/gennady/techreports.

[35] Mikosch, T. and Stărică, C. (1998) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Technical Report. University of Groningen. Available at www.math.rug.nl/~mikosch.

[36] Mikosch, T. and Stărică, C. (1999) Change of sructure in financial time series, long–range dependence and the GARCH model. Technical Report. University of Groningen. Available at www.math.rug.nl/~mikosch.

[37] Müller, U.A., Dacorogna, M.M. and Pictet, O.V. (1996) Heavy tails in high–frequency financial data. Preprint. Olsen & Associates, Zürich.

[38] Nelson, D.B. (1990) Stationarity and persistence in the GARCH$(1,1)$ model. *Econometric Theory* **6**, 318–334.

[39] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[40] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman and Hall, London.

[41] Vervaat, W. (1979) On a stochastic differential equation and a representation of a non–negative infinitely divisible random variables. *Adv. Appl. Prob.* **11**, 750–783.

Richard A. Davis
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523–1877
U.S.A.

Thomas Mikosch
Department of Mathematics
P.O. Box 800
University of Groningen
NL-9700 AV Groningen
and
EURANDOM
Eindhoven
THE NETHERLANDS