Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process

Thomas Mikosch
University of Groningen

and

Cătălin Stărică
Chalmers University Gothenburg

Abstract

The asymptotic theory for the sample autocorrelations and extremes of a GARCH$(1,1)$ process is provided. Special attention is given to the case when the sum of the ARCH and GARCH parameters is close to one, i.e. when one is close to an infinite variance marginal distribution. This situation has been observed for various financial log–return series and led to the introduction of the IGARCH model. In such a situation the sample autocorrelations are unreliable estimators of their deterministic counterparts for the time series and its absolute values, and the sample autocorrelations of the squared time series have non–degenerate limit distributions. We discuss the consequences for a foreign exchange rate series.

AMS 1991 Subject Classification: Primary: 62P20 Secondary: 90A20 60G55 60J10 62F10 62F12 62G30 62M10
Key Words and Phrases. GARCH, sample autocorrelations, stochastic recurrence equation, Pareto tail, extremes, point processes, foreign exchange rates

1 Introduction

Log–returns $X_t = \ln P_t - \ln P_{t-1}$ of foreign exchange rates, stock indices and share prices P_t, $t = 1, 2, \ldots$, typically share the following features:

- The frequency of large and small values (relative to the range of the data) is rather high, suggesting that *the data do not come from a normal, but from a heavy–tailed distribution.*
- Exceedances of high thresholds occur in clusters, which indicates that *there is dependence in the tails.*
- Sample autocorrelations of the data are tiny whereas the sample autocorrelations of the absolute and squared values are significantly different from zero even for large lags. This behaviour suggests that *there is some kind of long–range dependence in the data.*

Various models have been proposed in order to describe these features. Among them, models of the type

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

have become particularly popular. Here (Z_t) is a sequence of iid symmetric random variables with $Z = Z_1$,**i dont know what that means** $EZ^2 = 1$. One often assumes the Z_ts to be standard normal. Moreover, the sequence (σ_t) consists of non–negative random variables such that Z_t and σ_t are independent for every fixed t. In what follows, we frequently refer to σ_t as the *stochastic volatility of X_t*. Models of this type include the ARCH and GARCH family; see for example Engle [16] for their definitions and properties. In what follows, we often write σ for a generic random variable with the distribution of σ_1, X for a generic random variable with the distribution of X_1, etc.

We restrict ourselves to one particular model which has very often been used in applications: the GARCH$(1,1)$ process. It is defined by specifying σ_t as follows:

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

The parameters α_0, α_1 and β_1 are non–negative.

The stationary GARCH$(1,1)$ process is believed to capture, despite its simplicity, various of the empirically observed properties of log–returns. (Stationarity is always understood as strict stationarity.) For example, the stationary GARCH$(1,1)$ processes can exhibit heavy–tailed marginal distributions of power law type and hence they could be appropriate tools to model the heavier-than-normal tails of the financial data. This follows from a classical result by Kesten [31]; see Theorem 2.2 below. Although this result does not seem to be well known, the fact that certain power moments of X need not exist has been known for a long time in the econometrics literature;

see for example Nelson [39]. The question of the extent to which the the tails of the estimated GARCH$(1,1)$ model do describe the tails of the empirical distribution was addressed in [45]. It is shown there that, when using normal innovations, the tails of the fitted GARCH$(1,1)$ models seem to be much thinner than the tails apparent in the data. Hence, even though the GARCH$(1,1)$ processes could display heavy tails, when estimated on the data they *do not* produce tails that match the empirical ones. The relationship between the tail index of a GARCH$(1,1)$ process, its coefficients and the distribution of the innovations is made clear in Section 2.2.

The tail behaviour of the fitted GARCH$(1,1)$ processes is important from another perspective also. The empirical fact that the GARCH$(1,1)$ models fitted to log–return data often satisfy the condition $\alpha_1 + \beta_1 \approx 1$ implies that one often deals with a class of models with $E|X|^{2+\delta} = \infty$ for δ close to zero. (In particular, X may have infinite variance.) For such models, the asymptotic behaviour of various classical time series tools such as the sample autocorrelations and the periodogram are not always well–understood and give rise to many theoretical questions. (The GARCH$(1,1)$ model with $\alpha_1 + \beta_1 = 1$ is called integrated GARCH$(1,1)$ or IGARCH(1,1); see Engle and Bollerslev [17].)

Another empirical finding concerns the behaviour of the sample autocorrelation function (sample ACF) of powers of absolute log–return data at large lags. It has been noticed that the mentioned sample autocorrelations decay to zero at a hyperbolic rate (" long–range dependence"). This seems to be in contradiction with the sample ACF behaviour of a GARCH$(1,1)$ model. The GARCH$(1,1)$ process has good mixing properties; it is strongly mixing with geometric rate, provided Z has a density and $E|Z|^\epsilon < \infty$ for some $\epsilon > 0$; see for example Davis et al. [3]. Hence the autocorrelations of the underlying process, its absolute values and squares, given these quantities are well defined, decrease to zero at an exponential rate.

However, as mentioned above, most often the fitted GARCH$(1,1)$ models for log–return data belong to the class of GARCH$(1,1)$ processes with very heavy tails, i.e. models which do not have a finite 4th moment, although their second moment may still exist. Hence, autocovariances and autocorrelations are either not defined (for the squares, third powers, etc.), or when they exist (for the time series itself and its absolute values) the standard theory for the sample autocorrelation function, i.e. Gaussian limit distributions and $\sqrt{n}$–rates of convergence, is not valid any more. We show that in these cases the sample autocorrelations have infinite variance distributional limits and the rates of convergence are extremely slow. As a result, the asymptotic confidence bands are much wider than in the classical asymptotically normal theory.

Under these circumstances one could hope that the confidence bands are perhaps wide enough to bound the apparently hyperbolically decaying sample autocorrelation function of the absolute values of log–returns. In other words, it is possible that the discrepancy we mentioned between the empirically observed hyperbolic decay rate in the sample autocorrelation function and the exponential decay of the autocorrelation function of the GARCH$(1,1)$ model could be explained through statistical uncertainty related to the estimation procedure and hence claimed to be insignificant. If this were true then, up to statistical uncertainty, the GARCH$(1,1)$ model could be said to explain at least the empirical sample autocorrelation function behaviour. One of the conclusions of the paper is that, even when the mentioned larger–than–usual statistical uncertainty is accounted for, the GARCH$(1,1)$ *cannot* explain the effect of almost constancy of the sample ACF of the absolute values of log–returns.

As another desirable property that would recommend the GARCH$(1,1)$model as a viable candidate that captures the already–mentioned common features of the financial log–returns, exceedances of very low/high thresholds by the GARCH(1,1) process tend to occur in clusters. Formally, this behaviour can be described by the weak convergence of the point processes of exceedances, associated with the time series, to a compound Poisson process. The cluster sizes of this limiting process determine the *extremal index* $\theta \in (0,1)$, $1/\theta$ being the expected size of the clusters. Section 4 is devoted to the extremal behaviour of the GARCH$(1,1)$. A comparison of the estimated extremal indices of simulated GARCH$(1,1)$ and foreign exchange rate (FX) data is given in Section 6. This analysis reveals that the GARCH$(1,1)$ model fit to the log–returns does not, once again, properly describe the observed features of the data. The expected cluster sizes of the exceedances of high/low thresholds of the FX log–returns are smaller than the expected cluster sizes of simulated GARCH$(1,1)$ processes whose parameters were estimated from the FX observations. This means that there is less dependence in the tails of real–life data than in the GARCH$(1,1)$ model.

Our results serve in our view a double finality. On one hand, they can be thought of as a tool for deciding to which extent the potentially useful features of the GARCH$(1,1)$ model (heavy tails, slowly decaying sample ACF in the case $\alpha_1 + \beta_1 \approx 1$, clustering of the extremes) do actually describe accurately the corresponding empirical behaviour. In this sense, we conclude that, although displaying useful features, the GARCH$(1,1)$ model does not seem to accurately describe neither the extremal behaviour nor the correlation structure captured by the sample ACF of the data set that we analysed in detail.

On the other hand, we think of our findings as contributions to the growing number of results that emphasise the serious differences between the behaviour of various statistical tools under light and heavy tails, when dependency is present; see Resnick [42] for a recent survey paper. In this direction, we showed that the sample ACFs of GARCH$(1,1)$ models, their absolute values, squares, third powers, etc., fitted to real–life FX log–returns, are either poor estimators of the ACFs (slow convergence rates) or meaningless (non–degenerate limit distributions). Hence, in the case of the GARCH$(1,1)$ modelling, the sample ACF can be an extremely problematic statistical instrument that has to be used with caution when making statistical statements.

The paper is organised as follows. In Section 2 we consider some basic theoretical properties of the GARCH$(1,1)$ model. The weak convergence of the point processes associated with the sequences $((X_t,\sigma_t))$, $((|X_t|,\sigma_t))$ and $((X_t^2,\sigma_t^2))$ is considered in Section 3. In Section 4 we use these results to study the extremal behaviour of a GARCH$(1,1)$ process, including the calculation of its extremal index, the weak convergence of the point processes of exceedances and the weak limits of the distributions of the extremes. In Section 5 we study the asymptotic behaviour of the sample autocovariances and autocorrelations of the σ_ts and X_ts, their squares and absolute values. Section 6 contains an empirical study of foreign exchange rates and simulated GARCH$(1,1)$. In particular, we check the appropriateness of the GARCH$(1,1)$ as a model for the observed data as regards their dependence structure described by the autocorrelation and autocovariance functions, tails and extremal behaviour. Section 7 concludes with a short discussion on possible implications of the theoretical results provided.

2 Basic properties of GARCH$(1,1)$

In what follows, we collect some facts about the probabilistic properties of the GARCH$(1,1)$. First we notice that the GARCH$(1,1)$ can be considered in the much wider context of stochastic recurrence (or difference) equations of type

$$\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{B}_t\,, \quad t\in\mathbb{Z}\,, \tag{2.1}$$

where $((\mathbf{A}_t,\mathbf{B}_t))$ is an iid sequence, for every t the vector $\mathbf{X}_{t-1}$ is independent of $(\mathbf{A}_t,\mathbf{B}_t)$, the $\mathbf{A}_t$s are iid random $d\times d$ matrices and the $\mathbf{B}_t$s are iid d–dimensional random vectors.

Indeed, write

$$\mathbf{X}_t = \begin{pmatrix} X_t^2 \\ \sigma_t^2 \end{pmatrix}, \quad \mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 & \beta_1 Z_t^2 \\ \alpha_1 & \beta_1 \end{pmatrix}, \quad \mathbf{B}_t = \begin{pmatrix} \alpha_0 Z_t^2 \\ \alpha_0 \end{pmatrix}. \tag{2.2}$$

Then $\mathbf{X}_t$ satisfies equation (2.1) with $d=2$. Also observe that σ_t^2 satisfies the recurrence equation

$$\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2(\alpha_1 Z_{t-1}^2 + \beta_1)\,, \quad t\in\mathbb{Z}\,, \tag{2.3}$$

which is of the same type as (2.1) for $d=1$, with $\mathbf{X}_t=\sigma_t^2$, $\mathbf{A}_t = \alpha_1 Z_{t-1}^2+\beta_1$ and $\mathbf{B}_t=\alpha_0$.

Equations of type (2.1) have been extensively studied.

2.1 Existence of a stationary solution

The first question regards the existence of a stationary solution to equations (2.1) and (2.3). Applying the theory of Furstenberg and Kesten [19] or Kesten [31] (see also Vervaat [46] or Brandt [6]) it follows that (2.1) has a stationary solution if $E\ln^+\|\mathbf{A}\|<\infty$, $E\ln^+|\mathbf{B}|<\infty$ and if the top Lyapunov exponent $\widetilde{\gamma}$ defined as

$$\widetilde{\gamma} = \inf\{n^{-1}E\ln\|\mathbf{A}_1\cdots\mathbf{A}_n\|\,, \quad n\in\mathbb{N}\}\,,$$

is negative. Here $|\cdot|$ is any norm in $\mathbb{R}^n$, and

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

is the corresponding operator norm. Moreover, these conditions are close to necessity; see Bougerol and Picard [7]. Bougerol and Picard [8] also study the stationarity of the squared stochastic volatility process σ_t^2 for a general GARCH(p,q) process; the case $p=q=1$ was treated in Nelson [39].

It is in general difficult to calculate the top Lyapunov exponent $\widetilde{\gamma}$. However, in the particular case (2.2) calculation yields

$$\mathbf{A}_n \cdots \mathbf{A}_1 = \mathbf{A}_n \prod_{t=1}^{n-1} (\alpha_1 Z_t^2 + \beta_1)\,, \tag{2.4}$$

and so $\widetilde{\gamma} = E\ln(\alpha_1 Z^2 + \beta_1)$, provided $E|\ln|Z|| < \infty$. Alternatively, one can use the one–dimensional equation (2.3) and conclude from the conditions and literature above that

$$\alpha_0 > 0 \quad \text{and} \quad E\ln(\alpha_1 Z^2 + \beta_1) < 0 \tag{2.5}$$

are necessary and sufficient for stationarity of (X_t^2, σ_t^2); see Babillot et al. [1]. Notice that stationarity of σ_t^2 implies stationarity of the sequence

$$(X_t^2, \sigma_t^2) = \sigma_t^2\,(Z_t^2, 1)\,, \quad t \in \mathbb{Z}\,.$$

By construction of the sequence (X_t), stationarity of the sequence $((X_t, \sigma_t))$ follows.

The case $\alpha_0 = 0$ has also been considered in the literature (for example Nelson [39] and Ding and Granger [13]). The process σ_t is then non–stationary. If $E\ln(\alpha_1 Z^2 + \beta_1) < 0$ it converges to zero at an exponential rate with probability 1, a property which is undesirable for the purposes of time series analysis and extreme value theory. For example, the sample autocorrelations of such a sequence would be meaningless since they do not estimate the theoretical autocorrelations.

In what follows, we always assume that condition (2.5) *is satisfied.* Then a stationary version of $((X_t, \sigma_t))$ exists.

Remark 2.1 In practice, one observes a sample $X_1, \ldots, X_n$. Also in this situation we assume that this vector comes from a stationary model. In particular, *we assume that* (X_0, σ_0) *has the stationary initial distribution.* (Notice that we never observe the values σ_0 and X_0.) A theoretical argument shows that the distribution of $\mathbf{X}_t$ in (2.1) is close to the stationary distribution whatever the distribution of $\mathbf{X}_0$, given the latter is independent of $(Z_t)_{t=1,2,\ldots}$ and (2.5) holds. Consider two sequences $(\mathbf{X}_t(\mathbf{X}_0))_{t=0,1,\ldots}$ and $(\mathbf{X}_t(\mathbf{Y}))_{t=0,1,\ldots}$ given by the same recursion (2.1) but with initial conditions $\mathbf{X}_0$ and $\mathbf{Y}$, where both vectors are independent of the future values $((\mathbf{A}_t, \mathbf{B}_t))_{t=1,2,\ldots}$. Also assume that $\mathbf{X}_0$ has the stationary distribution. For any initial value $\mathbf{Y}$ we have the recursion

$$\mathbf{X}_t(\mathbf{Y}) = \mathbf{A}_t \cdots \mathbf{A}_1 \mathbf{Y} + \sum_{j=1}^{t} \mathbf{A}_t \cdots \mathbf{A}_{j+1} \mathbf{B}_j\,, \quad t = 1, 2, \ldots\,.$$

Hence (see (2.4)) for every $\epsilon > 0$,

$$
\begin{aligned}
(2.6) \qquad E|\mathbf{X}_t(\mathbf{X}_0) - \mathbf{X}_t(\mathbf{Y})|^\epsilon &\le E|\mathbf{A}_t \cdots \mathbf{A}_1(\mathbf{X}_0 - \mathbf{Y})|^\epsilon \\
&= E|\mathbf{A}_1(\mathbf{X}_0 - \mathbf{Y})|^\epsilon \left(E|\alpha_1 Z^2 + \beta_1|^\epsilon\right)^{t-1} \\
&\le E\|\mathbf{A}_1\|^\epsilon E|\mathbf{X}_0 - \mathbf{Y}|^\epsilon \left(E|\alpha_1 Z^2 + \beta_1|^\epsilon\right)^{t-1} .
\end{aligned}
$$

The right–hand side is finite if $E|Z|^{2\epsilon} < \infty$, $E|\mathbf{X}_0|^\epsilon < \infty$ and $E|\mathbf{Y}|^\epsilon < \infty$. Under (2.5), $E|\alpha_1 Z^2 + \beta_1|^\epsilon < 1$ for some sufficiently small ϵ, and so the left–hand side of (2.6) decays to zero exponentially as $t \to \infty$.

2.2 The tails of X_t and σ_t

Kesten [31] developed a renewal theory for products of random matrices. In particular, his results yield a characterisation of the asymptotic behaviour of the tails of the random variable X.

Theorem 2.2 *Assume Z has a density with unbounded support,* (2.5) *holds,*

$$E|\alpha_1 Z^2 + \beta_1|^{p/2} \ge 1 \quad \text{and} \quad E|Z|^p \ln^+ |Z| < \infty \quad \text{for some } p > 0.$$

A) There exists a number $\kappa \in (0, p]$ which is the unique solution of the equation

$$(2.7) \qquad E(\alpha_1 Z^2 + \beta_1)^{\kappa/2} = 1\,,$$

and there exists a positive constant $c_0 = c_0(\alpha_0, \alpha_1, \beta_1)$ such that

$$P(\sigma > x) \sim c_0\, x^{-\kappa} \quad \text{as } x \to \infty.$$

B) If $E|Z|^{\kappa+\epsilon} < \infty$ for some $\epsilon > 0$, then

$$(2.8) \qquad P(|X| > x) \sim E|Z|^\kappa \, P(\sigma > x)\,,$$

and the vector (X, σ) is jointly regularly varying in the sense that

$$(2.9) \qquad \frac{P\left(|(X,\sigma)| > xt\,, \quad (X,\sigma)/|(X,\sigma)| \in \cdot\right)}{P\left(|(X,\sigma)| > t\right)} \stackrel{v}{\to} x^{-\kappa} P\left(\Theta \in \cdot\right)\,, \quad x > 0\,,$$

as $t \to \infty$, where $\stackrel{v}{\to}$ denotes vague convergence on the Borel σ–field of the unit sphere $\mathbb{S}^1$ of $\mathbb{R}^2$, relative to the norm $|\cdot|$, and Θ is a random vector with values in $\mathbb{S}^1$ and distribution

$$P(\Theta \in \cdot) = \frac{E|(Z,1)|^\kappa I_{\{(Z,1)/|(Z,1)| \in \cdot\}}}{E|(Z,1)|^\kappa}\,.$$

Remark 2.3 The exact value of the constant c_0 is given in Goldie [20] who also gave an alternative proof of Kesten's [31] results for $d = 1$.

Remark 2.4 Under the assumptions of Theorem 2.2, $\alpha_1 = 0$ is not a possible parameter choice.

Remark 2.5 Assume in addition to the the conditions of Theorem 2.2 that $EZ^2 = 1$, $E|Z|^{2+\epsilon} < \infty$ for some $\epsilon > 0$ and $\alpha_1 + \beta_1 = 1$. Then (2.7) has the unique solution $\kappa = 2$. This implies that $P(|X| > x) \sim c\, x^{-2}$ for some $c > 0$ and, in turn, that $EX^2 = \infty$. GARCH(1,1) models fitted to real log–returns frequently have parameters α_1 and β_1 such that $\alpha_1 + \beta_1$ is close to 1. This indicates that one deals with time series models with extremely heavy tails.

Proof. Part A follows from an application of Theorems 3 and 4 in Kesten [31] to the recurrence equation (2.3). Equation (2.8) is a consequence of a result by Breiman [9]: assume ξ and η are independent non–negative random variables such that $P(\xi > x) = L(x)x^{-\kappa}$ for some slowly varying function L and $E\eta^{\kappa+\epsilon} < \infty$ for some $\epsilon > 0$. Then

$$P(\eta\, \xi > x) \sim E\eta^{\kappa}\, P(\xi > x)\,, \quad x \to \infty\,. \tag{2.10}$$

Another application of Breiman's result yields for any Borel set $B \subset \mathbb{S}^1$ that

$$\begin{aligned} P\left(|(X,\sigma)| > xt\,,\ (X,\sigma)/|(X,\sigma)| \in B\right) &= P\left(\sigma|(Z,1)| > xt\,,\ (Z,1)/|(Z,1)| \in B\right) \\ &= P\left(\sigma|(Z,1)| I_{\{(Z,1)/|(Z,1)|\in B\}} > xt\right) \\ &\sim E|(Z,1)|^{\kappa} I_{\{(Z,1)/|(Z,1)|\in B\}}\, P(\sigma > xt) \\ &\sim E|(Z,1)|^{\kappa} I_{\{(Z,1)/|(Z,1)|\in B\}}\, x^{-\kappa} P(\sigma > t)\,. \end{aligned}$$

Moreover,

$$P\left(|(X,\sigma)| > t\right) = P(\sigma|(Z,1)| > t) \sim E|(Z,1)|^{\kappa} P(\sigma > t)\,.$$

This concludes the proof. □

The idea of the proof of Theorem 2.2 can be used to consider the joint tail behaviour of the random variables σ_t, σ_t^2, X_t, $|X_t|$ and X_t^2. This behaviour can be described by a multivariate regular variation condition similar to (2.9); see Resnick [40], [41] or Bingham et al. [5] for properties and applications of multivariate regular variation in various fields. A d–dimensional vector $\mathbf{Y}$ is said to be *regularly varying with index* $\kappa \geq 0$ *and spectral measure* P_Θ on the Borel σ–field of the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ if

$$\frac{P(|\mathbf{Y}| > xt\,, \mathbf{Y}/|\mathbf{Y}| \in \cdot)}{P(|\mathbf{Y}| > t)} \stackrel{v}{\to} x^{-\kappa} P_\Theta(\cdot)\,. \tag{2.11}$$

Here $\stackrel{v}{\to}$ denotes vague convergence on the Borel σ–field of $\mathbb{S}^{d-1}$ and P_Θ is the distribution of a certain random vector Θ with values in $\mathbb{S}^{d-1}$.

Write

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

For every $h \geq 0$, define

$$
\begin{aligned}
\mathbf{Y}_h &= (X_0, \sigma_0, \ldots, X_h, \sigma_h)\,, \\
\mathbf{Y}_h^{(i)} &= (|X_0|^i, \sigma_0^i, \ldots, |X_h|^i, \sigma_h^i)\,, \quad i = 1, 2\,, \\
\mathbf{Z}_h^{(2)} &= \left((Z_0^2, 1)\,, A_1\,(Z_1^2, 1)\,, \prod_{i=1}^{2} A_i\,(Z_2^2, 1), \ldots, \prod_{i=1}^{h} A_i\,(Z_h^2, 1) \right). \qquad (2.12)
\end{aligned}
$$

Theorem 2.6 *Let $h \geq 0$ and assume that the conditions of Theorem 2.2,B are satisfied. Let $|\cdot|$ denote the max-norm.*

A) $\mathbf{Y}_h^{(2)}$ is regularly varying with index $\kappa/2$ and spectral measure

$$
P_\Theta(\cdot) = \frac{E|\mathbf{Z}_h^{(2)}|^{\kappa/2} I_{\{\mathbf{Z}_h^{(2)}/|\mathbf{Z}_h^{(2)}| \in \cdot\}}}{E|\mathbf{Z}_h^{(2)}|^{\kappa/2}}\,.
$$

B) $\mathbf{Y}_h^{(1)}$ is regularly varying with index κ and spectral measure $P_{\Theta^{1/2}}$, where $\Theta = (\theta_0, \ldots, \theta_h)$ and

$$
\Theta^{1/2} = (\theta_0^{1/2}, \ldots, \theta_h^{1/2})\,.
$$

C) $\mathbf{Y}_h$ is regularly varying with index κ and spectral measure given by the distribution of the vector

$$
\left(r_0 \theta_0^{1/2}, \theta_0^{1/2}, \ldots, r_h \theta_h^{1/2}, \theta_h^{1/2} \right), \qquad (2.13)
$$

where (r_t) is a sequence of iid Bernoulli random variables such that $P(r = \pm 1) = 0.5$, independent of Θ.

Proof. We start with

$$
\begin{aligned}
\mathbf{Y}_h^{(2)} &= \left(\sigma_0^2 (Z_0^2, 1), \sigma_1^2 (Z_1^2, 1), \ldots, \sigma_h^2 (Z_h^2, 1)\right) \\
&= \left(\sigma_0^2 (Z_0^2, 1), (\alpha_0 + \sigma_0^2 A_1)(Z_1^2, 1), \ldots, (\alpha_0 + \sigma_{h-1}^2 A_h)(Z_h^2, 1)\right) \\
&= \left(\sigma_0^2 (Z_0^2, 1), \sigma_0^2 A_1 (Z_1^2, 1), \ldots, \sigma_{h-1}^2 A_h (Z_h^2, 1)\right) + \mathbf{R}_h \\
&= \mathbf{C}_h + \mathbf{R}_h\,.
\end{aligned}
$$

Under the assumptions of Theorem 2.2 on Z, each of the random variables σ_t^2 is regularly varying with index $\kappa/2$ and therefore the tail of $\mathbf{R}_h$ is small compared to the tail of $\mathbf{Y}_h^{(2)}$. Hence the tail of $\mathbf{Y}_h^{(2)}$ is determined only by the tail of $\mathbf{C}_h$. By the same argument and induction we may conclude that the tail of $\mathbf{Y}_h^{(2)}$ is determined by the tail of the vector

$$
\sigma_0^2 \mathbf{Z}_h^{(2)} = \sigma_0^2 \left((Z_0^2, 1), A_1\,(Z_1^2, 1), \prod_{i=1}^{2} A_i\,(Z_2^2, 1), \ldots, \prod_{i=1}^{h} A_i\,(Z_h^2, 1) \right).
$$

Hence for any Borel set $B \subset \mathbb{S}^{h-1}$ and in view of Breiman's result (2.10), as $t \to \infty$,

$$\frac{P\left(|\mathbf{Y}_h^{(2)}| > xt\,, \mathbf{Y}_h/|\mathbf{Y}_h| \in B\right)}{P\left(|\mathbf{Y}_h^{(2)}| > t\right)} \sim \frac{P\left(\sigma_0^2|\mathbf{Z}_h^{(2)}| > xt\,, \mathbf{Z}_h^{(2)}/|\mathbf{Z}_h^{(2)}| \in B\right)}{P\left(\sigma_0^2|\mathbf{Z}_h^{(2)}| > t\right)}$$

$$\sim x^{-\kappa/2}\frac{E|\mathbf{Z}_h^{(2)}|^{\kappa/2} I_{\{\mathbf{Z}_h^{(2)}/|\mathbf{Z}_h^{(2)}| \in B\}}}{E|\mathbf{Z}_h^{(2)}|^{\kappa/2}}$$

$$=: x^{-\kappa/2} P\left(\Theta \in B\right) .$$

Since $\mathbf{Y}_h^{(2)}$ is positive with probability 1, it follows from the results in the Appendix of Davis et al. [3] that $\mathbf{Y}_h^{(1)}$ is regularly varying with index κ and spectral measure $P_{\Theta^{1/2}}$. It remains to consider $\mathbf{Y}_h$. We can write

$$\mathbf{Y}_h = (\mathrm{sign}(Z_0)|X_0|, \sigma_0, \ldots, \mathrm{sign}(Z_h)|X_h|, \sigma_h)\,,$$

and, by symmetry of Z, we know that the sequence $(\mathrm{sign}(Z_t))$ is independent of the sequence $((|X_t|, \sigma_t))$. Therefore we can use the results in the Appendix of Davis et al. [3] to conclude that $\mathbf{Y}_h$ is regularly varying with index κ and spectral measure given by the distribution of the vector (2.13). □

3 Convergence of point processes

3.1 Preliminaries

We follow the point process theory in Kallenberg [30]. The state space of the point processes considered is $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$. Write $\mathcal{M}$ for the collection of Radon counting measures on $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ with null measure o. We consider a strictly stationary sequence $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ of random row vectors with values in $\mathbb{R}^d$. For simplicity, we write $\mathbf{X} = \mathbf{X}_0 = (X_1, \ldots, X_d)$. In what follows, we formulate results on the weak convergence of the point processes

$$N_n = \sum_{t=1}^n \varepsilon_{\mathbf{X}_t/a_n}\,, \quad n = 1, 2, \ldots\,,$$

where $\mathbf{X}$ is regularly varying in the sense of the defining property (2.11), and (a_n) is a sequence of positive numbers such that

$$n\,P(|\mathbf{X}| > a_n) \to 1\,, \quad n \to \infty\,. \tag{3.1}$$

The proofs follow from the results in Davis and Mikosch [12]; they are variations on results for $d = 1$ in Davis and Hsing [11]; see also Davis et al. [3].

Define

$$\widetilde{\mathcal{M}} = \left\{\mu \in \mathcal{M} :\ \mu(\{\mathbf{x} : |\mathbf{x}| > 1\}) = 0 \quad \text{and} \quad \mu(\{\mathbf{x} : \mathbf{x} \in \mathbb{S}^{d-1}\}) > 0\right\}\,, \tag{3.2}$$

and let $\mathcal{B}(\widetilde{\mathcal{M}})$ be the Borel σ–field of $\widetilde{\mathcal{M}}$.

Theorem 3.1 *Assume that the following conditions are satisfied:*

- *The stationary sequence* $(\mathbf{X}_t)$ *is strongly mixing with mixing rate* ϕ_n*.*
- *The following condition holds:*

$$\lim_{k\to\infty}\limsup_{n\to\infty} P\left(\bigvee_{k\le|t|\le r_n}|\mathbf{X}_t|>a_n y\,\middle|\,|\mathbf{X}_0|>a_n y\right)=0\,,\quad y>0\,, \tag{3.3}$$

  where $r_n, m_n\to\infty$ *are two integer sequences such that* $n\phi_{m_n}/r_n\to 0$*,* $r_n m_n/n\to 0$*.*

- *All finite–dimensional distributions of* $(\mathbf{X}_t)$ *are jointly regularly varying with index* $\kappa>0$*. To be specific, let* $(\boldsymbol{\theta}^{(-k)},\dots,\boldsymbol{\theta}^{(k)})$ *be the* $(2k+1)d$*–dimensional random row vector with values in the unit sphere* $\mathbb{S}^{(2k+1)d-1}$ *that appears in the definition of joint regular variation of* $(\mathbf{X}_{-k},\dots,\mathbf{X}_k)$*,* $k\ge 0$*.*

Then the limit

$$\gamma=\lim_{k\to\infty} E\left(|\boldsymbol{\theta}_0^{(k)}|^\kappa-\textstyle\bigvee_{j=1}^k|\boldsymbol{\theta}_j^{(k)}|^\kappa\right)_+\Big/\,E|\boldsymbol{\theta}_0^{(k)}|^\kappa \tag{3.4}$$

exists and is the extremal index of the sequence $(|\mathbf{X}_t|)$*.*

- *If* $\gamma=0$ *in* (3.4)*, then* $N_n\stackrel{d}{\to} o$*.*
- *If* $\gamma>0$*, then* $N_n\stackrel{d}{\to} N\neq o$*, where*

$$N\stackrel{d}{=}\sum_{i=1}^\infty\sum_{j=1}^\infty \varepsilon_{P_i\mathbf{Q}_{ij}}\,,$$

  $\sum_{i=1}^\infty\varepsilon_{P_i}$ *is a Poisson process on* $\mathbb{R}_+$ *with intensity measure*

$$\nu(dy)=\gamma\kappa y^{-\kappa-1}dy\,.$$

  This process is independent of the sequence of iid point processes $\sum_{j=1}^\infty\varepsilon_{\mathbf{Q}_{ij}}$*,* $i\ge 1$*, with joint distribution* Q *on* $(\widetilde{\mathcal{M}},\mathcal{B}(\widetilde{\mathcal{M}}))$*, where* Q *is the weak limit of*

$$E\left[\left(|\boldsymbol{\theta}_0^{(k)}|^\kappa-\textstyle\bigvee_{j=1}^k|\boldsymbol{\theta}_j^{(k)}|^\kappa\right)_+ I_{\cdot}\left(\textstyle\sum_{|t|\le k}\varepsilon_{\boldsymbol{\theta}_t^{(k)}}\right)\right]\Big/\,E\left(|\boldsymbol{\theta}_0^{(k)}|^\kappa-\textstyle\bigvee_{j=1}^k|\boldsymbol{\theta}_j^{(k)}|^\kappa\right)_+$$

  as $k\to\infty$ *which exists.*

3.2 Main result

The following theorem is our main result on weak convergence for point processes associated with a GARCH$(1,1)$ process.

Theorem 3.2 *Let* (X_t) *be the* GARCH$(1,1)$ *process defined in Section* 1 *and assume that the conditions of Theorem 2.2,B hold. For fixed* $h \geq 0$, *set* $\mathbf{X}_t = (X_t, \sigma_t, \ldots, X_{t+h}, \sigma_{t+h})$. *Let* (a_n) *be a sequence of constants such that* (3.1) *holds. Then the conditions of Theorem 3.1 are met, and hence*

$$N_n = \sum_{t=1}^{n} \varepsilon_{\mathbf{X}_t/a_n} \stackrel{d}{\to} N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{P_i \mathbf{Q}_{ij}} \,,$$

where (P_i) *and* $(\mathbf{Q}_{ij})$ *are defined in the statement of the theorem. We write*

$$\mathbf{Q}_{ij} = \left((Q_{ij,X}^{(m)}, Q_{ij,\sigma}^{(m)}), m = 0, \ldots, h \right) .$$

Proof. The joint regular variation of all finite–dimensional distributions follows from Theorem 2.6. In addition, the process is strongly mixing with geometric rate; so that the mixing condition of Theorem 3.1 is met; see Davis et al. [3].

It remains to show (3.3). By the definition of the sequence $(\mathbf{X}_t)$, it suffices to switch in condition (3.3) to the sequence $((X_t^2, \sigma_t^2))$. Recall that the latter sequence satisfies the recurrence equation (2.1). In this situation one can apply the techniques of the proof of Theorem 2.3 in Davis et al. [3] to conclude that (3.3) holds. □

Remark 3.3 Analogous results can be obtained for the vectors

$$\mathbf{X}_t^{(l)} = (|X_t|^l, \sigma_t^l, \ldots, |X_{t+h}|^l, \sigma_{t+h}^l) \,, \quad l = 1, 2 \,,$$

either by applying the same arguments of proof as above or by deriving the weak limit of the point processes from Theorem 3.2 in combination with a continuous mapping argument. Indeed, under the assumptions of Theorem 3.2,

$$N_n^{(l)} = \sum_{t=1}^{n} \varepsilon_{\mathbf{X}_t^{(l)}/a_n^l} \stackrel{d}{\to} N^{(l)} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{P_i^l \mathbf{Q}_{ij}^{(l)}} \,, \quad l = 1, 2 \,.$$

where (P_i) is the same as above and

$$\mathbf{Q}_{ij}^{(l)} = \left((|Q_{ij,X}^{(m)}|^l, |Q_{ij,\sigma}^{(m)}|^l), m = 0, \ldots, h \right) \,, \quad l = 1, 2 \,.$$

Remark 3.4 It is possible to extend the above results to two–dimensional point processes. For illustrational purposes we restrict ourselves to the processes

$$\widehat{N}_n = \sum_{t=1}^{n} \varepsilon_{(t/n, X_t/a_n)} \,.$$

The analogous results for the point processes $\widehat{N}_n^{(l)}$ for $(|X_t|^l)$ are also valid. The weak convergence of (N_n) implies the convergence of $(\widehat{N}_n)$ under the assumption of strong mixing; see Mori [38]. For fixed $x > 0$, the point process of exceedances of the threshold xa_n by the sequence (X_t) is defined as

$$\widetilde{N}_n(\cdot) = \sum_{t=1}^{n} \varepsilon_{t/n}(\cdot) I_{\{X_t > xa_n\}} = \widehat{N}_n(\cdot \times (x, \infty)) \, .$$

According to a result by Hsing et al. [28] (cf. Falk et al. [18]), the weak limit of $(\widetilde{N}_n)$ is compound Poisson with compounding probabilities (π_k) and probability generating function

$$\Pi(u) = \sum_{k=1}^{\infty} \pi_k u^k .$$

Specifically, in the limiting compound Poisson process events occur as an ordinary Poisson process, independent of the multiplicities (cluster sizes) of the events with compounding probabilities π_k, i.e. π_k is the probability that an event has multiplicity k. The intuitive content of this result in the case of the GARCH$(1,1)$ process is that excursions above (below) a high (low) threshold are independent and occur exponentially spaced in time. The lengths of these excursions are iid with probabilities π_k.

4 The extremal behaviour

The point process results of Section 3 enable one to study the extremal behaviour of the sequences (X_t) and (σ_t). In what follows, we restrict ourselves to the partial maxima and the extremal index. For any sequence of random variables (Y_n) define the partial maxima

$$M_{n,Y} = \max_{i=1,\dots,n} Y_i \, , \quad n = 1, 2 \dots \, .$$

We assume that the conditions of Theorem 2.2,B hold. If (Y_n) is iid with the same marginal distribution as X we may conclude that $P(Y > x) \sim cx^{-\kappa}$ for some $\kappa > 0$. Then, with (b_n) such that $nP(X > b_n) \sim 1$,

$$P(b_n^{-1} M_{n,Y} \le x) \to P(Y^{(\kappa)} \le x) \, , \quad x > 0 \, , \tag{4.1}$$

where $Y^{(\kappa)}$ has a standard Fréchet distribution Φ_κ (see for example Embrechts et al. [14], Chapter 3):

$$\Phi_\kappa(x) = P(Y^{(\kappa)} \le x) = \exp\{-x^{-\kappa}\} \, , \quad x > 0 \, .$$

For dependent sequences such as (X_t) and (σ_t) exceedances of high thresholds occur in clusters, and so we cannot expect that (4.1) remains valid for them. For stationary sequences (Y_t) with tail behaviour $P(Y > x) \sim cx^{-\kappa}$ the notion of *extremal index* θ_Y describes the clustering behaviour of the extremes (see Leadbetter [33]; cf. Embrechts et al. [14], Section 8.1):

$$P(b_n^{-1} M_{n,Y} \le x) \to P^{\theta_Y}(Y^{(\kappa)} \le x)\,, \quad x > 0\,.$$

The extremal index θ_Y assumes values in $[0,1]$ and can be interpreted as the reciprocal of the expected cluster size of high level exceedances of the normalised sequence (Y_t). For an iid sequence (Y_t) one has $\theta_Y = 1$.

Remark 4.1 In what follows, we define the sequence $(b_{n,Y})$ for an iid sequence (Y_t) as $n\,P(Y > b_{n,Y}) \sim 1$. In particular, $b_{n,\sigma}$, $b_{n,|X|}$ and $b_{n,X}$, up to a multiplicative constant, are asymptotically of the same order as $n^{1/\kappa}$.

Theorem 4.2 *Assume the conditions of Theorem 2.2,B are satisfied.*
A) The partial maxima of (σ_t) *satisfy the limit relation*

$$P(b_{n,\sigma}^{-1} M_{n,\sigma} \le x) \to P^{\theta_\sigma}(Y^{(\kappa)} \le x)\,, \quad x > 0\,,$$

with extremal index

$$\theta_\sigma = \int_1^\infty P\left(\sup_{t\ge 1} \prod_{j=1}^t A_j \le y^{-1}\right) \frac{\kappa}{2} y^{-(\kappa/2)-1}\, dy\,,$$

where $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$, $t = 1,2,\ldots$.
B) The partial maxima of $(|X_t|)$ *satisfy the limit relation*

$$P(b_{n,|X|}^{-1} M_{n,|X|} \le x) \to P^{\theta_{|X|}}(Y^{(\kappa)} \le x)\,, \quad x > 0\,, \tag{4.2}$$

with extremal index

$$\theta_{|X|} = \lim_{k\to\infty} E\left(|Z_1|^\kappa - \max_{j=2,\ldots,k+1} \left|Z_j^2 \prod_{i=2}^j A_i\right|^{\kappa/2}\right)_+ \Big/ E|Z_1|^\kappa\,.$$

C) The partial maxima of (X_t) *satisfy the limit relation*

$$P(b_{n,X}^{-1} M_{n,X} \le x) \to P^{\theta_X}(Y^{(\kappa)} \le x)\,, \quad x > 0\,,$$

with extremal index θ_X

$$\theta_X = 2\theta_{|X|}(1 - \widetilde{\Pi}(0.5))\,,$$

where $\widetilde{\Pi}$ *is the probability generating function of the compounding probabilities of a compound Poisson process. The latter is the weak limit of the point processes of exceedances of the thresholds* $b_{n,|X|}$ *by* $(|X_t|)$ *explained in Remark* 3.4.

Remark 4.3 The formulae for the extremal indices given above can be evaluated numerically or by Monte–Carlo techniques. An example how to proceed for an ARCH(1) process has been given in de Haan et al. [25]. We refrain here from numerically calculating the extremal indices and restrict ourselves to their statistical estimation for some simulated GARCH$(1,1)$ and foreign exchange log–returns; see Section 6.

Proof. A) The proof follows by an application of Theorem 2.1 in de Haan et al. [25] to the recurrence equation (2.3) for the sequence (σ_t^2).

B) The existence of the Frechét limit in (4.2) follows from Theorem 3.3.3 in Leadbetter et al. [34], the fact that (X_t) is strongly mixing and the Pareto–like tails of X. The existence of the extremal index $\theta_{|X|}$ is a consequence of Theorem 3.1. For the calculation of $\theta_{|X|} = \theta_{X^2}$ we follow the lines of the calculation of the extremal index of an ARCH(1) in Remark 4.3 of Davis and Mikosch [12]. From (3.4) and (2.12), one obtains θ_{X^2} as the limit for $k \to \infty$ of the quantities

$$E\left(\left|Z_{k+1}^2 \prod_{i=1}^{k+1} A_i\right|^{\kappa/2} - \max_{j=k+2,\ldots,2k+1}\left|Z_j^2 \prod_{i=1}^{j} A_i\right|^{\kappa/2}\right)_+ \Big/ E\left|Z_{k+1}^2 \prod_{i=1}^{k+1} A_i\right|^{\kappa/2}$$

$$= E\left(|Z_{k+1}|^\kappa - \max_{j=k+2,\ldots,2k+1}\left|Z_j^2 \prod_{i=k+2}^{j} A_i\right|^{\kappa/2}\right)_+ \Big/ E|Z_1|^\kappa .$$

C) Notice that, by symmetry of Z, $(X_t) = (r_t|X_t|)$, where the sequence of the $r_t = \text{sign}(X_t)$ is independent of $(|X_t|)$. As in de Haan et al. [25] for the ARCH(1), one can use this property to obtain the limit distribution of $(b_{n,X}^{-1} M_{n,X})$ and the extremal index θ_X by independent thinning from the point processes of exceedances for $(|X_t|)$. The weak convergence of the latter processes to a compound Poisson process has been described in Remark 3.4. Then proceed as in de Haan et al. [25], pp. 222–223. $\square$

5 Convergence of the sample autocorrelations

In what follows, we study the weak limit behaviour of the sample autocovariances and sample autocorrelations of the sequences (X_t) and (σ_t), their squares and absolute values. We assume that the conditions of Theorem 2.2,B hold. Then the vector (X_t, σ_t) is regularly varying with index $\kappa > 0$ and, by Theorem 3.2 and Remark 3.3, the point processes N_n, $N_n^{(1)}$ and $N_n^{(2)}$ generated by the vectors $\mathbf{X}_t$, $\mathbf{X}_t^{(1)}$ and $\mathbf{X}_t^{(2)}$, respectively, converge in distribution to the processes N, $N_n^{(1)}$ and $N_n^{(2)}$; see Section 3.2 for the notation. This is the basis for the weak convergence of the sample autocovariance function (ACVF)

$$\gamma_{n,X}(h) = \frac{1}{n}\sum_{t=1}^{n-h} X_t X_{t+h}\,, \quad h = 0, 1, \ldots,$$

and the corresponding versions for $(|X_t|)$ and (X_t^2):

$$\gamma_{n,|X|}(h) = \frac{1}{n}\sum_{t=1}^{n-h} |X_t||X_{t+h}|\,, \quad \gamma_{n,X^2}(h) = \frac{1}{n}\sum_{t=1}^{n-h} X_t^2 X_{t+h}^2\,, \quad h = 0,1,\ldots.$$

The sample autocorrelation function (ACF) of (X_t) is defined as

$$\rho_{n,X} = \gamma_{n,X}(h)/\gamma_{n,X}(0)\,, \quad h = 0,1,\ldots\,.$$

The sample ACVF and ACF for the sequences $(|X_t|)$, (X_t^2), (σ_t) and (σ_t^2) are defined analogously. The deterministic counterparts (ACVF, ACF) are denoted by

$$\gamma_X(h) = EX_0X_h\,, \quad \gamma_{|X|}(h) = E|X_0X_h|\,, \quad \gamma_{X^2}(h) = EX_0^2X_h^2\,, \quad \rho_X(h) = \gamma_X(h)/\gamma_X(0)\,, \quad \text{etc.}$$

Remark 5.1 Alternatively, one can consider the sample ACVF and ACF for the mean–corrected random variables $X_t - \overline{X}_n$ ($\overline{X}_n$ is the sample mean) and the corresponding versions for $|X_t|$, X_t^2, etc. The same arguments as below show that the limits do not change. In the heavy-tailed situation centring with the sample mean is not necessarily a natural choice; for example, the mean does not necessarily exist.

The following result from Davis and Mikosch [12] is the basis for the understanding of the asymptotic behaviour of sample ACVF and ACF:

Theorem 5.2 *Assume that (X_t) is a strictly stationary sequence of random variables such that $\mathbf{X}_t = (X_t,\ldots,X_{t+h})$ satisfied the conditions of Theorem 3.1 with $\gamma > 0$ and $d = h+1$. Then the following statements hold.*

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

$$\left(n\, a_n^{-2}\gamma_{n,X}(m)\right)_{m=0,\ldots,h} \xrightarrow{d} (V_m)_{m=0,\ldots,h}\,,$$
$$\left(\rho_{n,X}(m)\right)_{m=1,\ldots,h} \xrightarrow{d} (V_m/V_0)_{m=1,\ldots,h}\,,$$

where

$$V_m = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij}^{(0)} Q_{ij}^{(m)}\,, \quad m = 0,\ldots,h\,.$$

The vector $(V_0,\ldots,V_h)$ is jointly $\kappa/2$–stable in $\mathbb{R}^{h+1}$.

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

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\ \operatorname{var}\left(a_n^{-2}\sum_{t=1}^{n-m} X_tX_{t+m}I_{\{|X_tX_{t+m}|\le a_n^2\epsilon\}}\right) = 0\,, \tag{5.1}$$

then

$$(5.2)\qquad \left(n\, a_n^{-2}(\gamma_{n,X}(m) - \gamma_X(m))\right)_{m=0,\ldots,h} \stackrel{d}{\to} (V_m)_{m=0,\ldots,h}\,,$$

where $(V_0,\ldots,V_h)$ *is the distributional limit of*

$$\left(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij}^{(0)} Q_{ij}^{(m)} I_{(\epsilon,\infty]}(P_i^2|Q_{ij}^{(0)}Q_{ij}^{(m)}|) - \int_{B_{\epsilon,m}} x_0 x_m \tau(d\mathbf{x})\right)_{m=0,\ldots,h},$$

where

$$B_{\epsilon,m} = \left\{\mathbf{x}\in\mathbb{R}^{h+1}\,:\,\epsilon < |x_0 x_m|\right\},$$

and τ *is the limiting measure which appears in the relation* $nP(\mathbf{X}/a_n\in\cdot)\stackrel{v}{\to}\tau(\cdot)$, *and* $\stackrel{v}{\to}$ *denotes vague convergence in* $\overline{\mathbb{R}}^{h+1}\backslash\{\mathbf{0}\}$. *Moreover,*

$$(5.3)\qquad \left(na_n^{-2}(\rho_{n,X}(m) - \rho_X(m))\right)_{m=1,\ldots,h} \stackrel{d}{\to} \gamma_X^{-1}(0)\,(V_m - \rho_X(m)\,V_0)_{m=1,\ldots,h}\,.$$

Remark 5.3 The limits in (5.2) and (5.3) are $\kappa/2$–stable. The case $\kappa = 2$ can also be treated but one has to take care of the particular centring constants in this case. Therefore it is omitted. See Samorodnitsky and Taqqu [43] for definitions and properties of stable random vectors.

5.1 Convergence in distribution of the sample ACF

We first consider the cases when the sample ACFs of (X_t), $(|X_t|)$, (σ_t^2), (σ_t) have non–degenerate limits in distribution. Recall the definition of the vectors $\mathbf{X}_t$, $\mathbf{X}_t^{(1)}$ and $\mathbf{X}_t^{(2)}$ from Section 3.2. Assume that the conditions of Theorem 3.2 hold. Then $\mathbf{X}_t$ is regularly varying with index $\kappa > 0$ and

$$N_n \stackrel{d}{\to} N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\varepsilon_{P_i\mathbf{Q}_{ij}} \quad\text{and}\quad N_n^{(l)}\stackrel{d}{\to}N^{(l)}\,,\quad l=1,2\,,$$

where

$$\mathbf{Q}_{ij} = \left((Q_{ij,X}^{(m)}, Q_{ij,\sigma}^{(m)})\,, m=0,\ldots,h\right)$$

The case $\kappa\in(0,2)$**.** An application of Theorem 5.2(1) yields

$$(na_n^{-2}(\gamma_{n,X}(m),\gamma_{n,\sigma}(m)))_{m=0,\ldots,h} \stackrel{d}{\to} ((V_{m,X},V_{m,\sigma}))_{m=0,\ldots,h}\,,$$

$$((\rho_{n,X}(m),\rho_{n,\sigma}(m)))_{m=1,\ldots,h} \stackrel{d}{\to} \left(\left(\frac{V_{m,X}}{V_{0,X}}\right),\left(\frac{V_{m,\sigma}}{V_{0,\sigma}}\right)\right)_{m=1,\ldots,h},$$

where the vector $((V_{m,X},V_{m,\sigma}))_{m=1,\ldots,h}$ is $\kappa/2$–stable with point process representation

$$(5.4)\qquad V_{m,X} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij,X}^{(0)} Q_{ij,X}^{(m)}\,,\quad m=0,\ldots,h\,,$$

$$V_{m,\sigma} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij,\sigma}^{(0)} Q_{ij,\sigma}^{(m)}\,,\quad m=0,\ldots,h\,.$$

The analogous relations hold for the sample ACVF and ACF of the sequences $(|X_t|)$ and (σ_t). In this case, one has to replace $((V_{m,X}, V_{m,\sigma}))$ with $((V_{m,|X|}, V_{m,\sigma}))$, where $V_{m,|X|}$ is obtained by replacing the vectors $\mathbf{Q}_{ij}$ in the infinite series (5.4) with

$$\mathbf{Q}_{ij}^{(1)} = \Big((|Q_{ij,X}^{(m)}|, Q_{ij,\sigma}^{(m)}) \, , m = 0, \ldots, h \Big) \, .$$

The case $\kappa \in (0,4)$. We consider the sample ACVF and ACF of the squares (X_t^2) and (σ_t^2). The same argument as above yields that

$$\begin{aligned}
(na_n^{-4}(\gamma_{n,X^2}(m), \gamma_{n,\sigma^2}(m)))_{m=0,\ldots,h} &\stackrel{d}{\to} \big((V_{m,X^2}, V_{m,\sigma^2})\big)_{m=0,\ldots,h} \, , \\
((\rho_{n,X^2}(m), \rho_{n,\sigma^2}(m)))_{m=1,\ldots,h} &\stackrel{d}{\to} \left(\left(\frac{V_{m,X^2}}{V_{0,X^2}} \right), \left(\frac{V_{m,\sigma^2}}{V_{0,\sigma^2}} \right) \right)_{m=1,\ldots,h} \, ,
\end{aligned}$$

where the vector $((V_{m,X^2}, V_{m,\sigma^2}))_{m=1,\ldots,h}$ is $\kappa/4$–stable with point process representation

$$\begin{aligned}
V_{m,X^2} &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_i^4 (Q_{ij,X}^{(0)} Q_{ij,X}^{(m)})^2 \, , \quad m = 0, \ldots, h \, , \\
V_{m,\sigma^2} &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_i^4 (Q_{ij,\sigma}^{(0)} Q_{ij,\sigma}^{(m)})^2 \, , \quad m = 0, \ldots, h \, .
\end{aligned}$$

5.2 Rates of convergence for the sample ACF toward the ACF

In this section we assume that the covariances of X_t (respectively X_t^2) are finite. Then, by the ergodic theorem,

$$\gamma_{n,X}(h) \stackrel{\text{a.s.}}{\to} \gamma_X(h) \, , \quad \gamma_{n,|X|^l}(h) \stackrel{\text{a.s.}}{\to} \gamma_{|X|^l}(h) \, , \quad l = 1, 2 \, ,$$

and the analogous relations hold for the σ– and σ^2–sequences. It arises the question as to the rate of convergence in these results.

5.2.1 Convergence to the normal distribution

The Markov chain $((X_t^2, \sigma_t^2))$ is strongly mixing with geometric rate; see for example Davis et al. [3]. Hence the standard CLT for strongly mixing sequences applies provided suitable moment conditions hold; see for example Ibragimov and Linnik [29], Meyn and Tweedie [36].

The case $\kappa \in (8, \infty)$. Then $EX^8 < \infty$ and $E\sigma^8 < \infty$. The standard CLT applies to the sample ACVF and ACF of the X^2– and σ^2–sequences:

$$\begin{aligned}
& n^{1/2} \big((\gamma_{n,X^2}(m) - \gamma_{X^2}(m), \gamma_{n,\sigma^2}(m) - \gamma_{\sigma^2}(m)) \big)_{m=0,\ldots,h} \\
\stackrel{d}{\to} \; & \big((G_{m,X^2}, G_{m,\sigma^2}) \big)_{m=0,\ldots,h} \, , \\
& n^{1/2} \big((\rho_{n,X^2}(m) - \rho_{X^2}(m), \rho_{n,\sigma^2}(m) - \rho_{\sigma^2}(m)) \big)_{m=1,\ldots,h} \\
\stackrel{d}{\to} \; & \big((\gamma_{X^2}^{-1}(0)(G_{m,X^2} - \rho_{X^2}(m) G_{0,X^2}), \gamma_{\sigma^2}^{-1}(0)(G_{m,\sigma^2} - \rho_{\sigma^2}(m) G_{0,\sigma^2})) \big)_{m=1,\ldots,h} \, ,
\end{aligned}$$

where the limits are multivariate Gaussian with mean zero.

The case $\kappa \in (4,\infty)$. The analogous results hold for the X–, $|X|$– and σ–sequences. We omit details.

5.2.2 Convergence to infinite variance stable distributions

These results follow from part 2 of Theorem 5.2; their derivation can be quite delicate. We characterise the weak limits of the sample ACVF in terms of limiting point processes; the limit of the sample ACF can be derived by applying a simple continuous mapping argument. The weak limits are infinite variance stable random vectors. However, they are functionals of point processes and therefore it is difficult to determine the spectral measure of these stable vectors; the spectral measure determines the dependence structure of the random vector; see Samorodnitsky and Taqqu [43]. Therefore the results below are qualitative ones. At the moment we do not know how to use them for the construction of asymptotic confidence bands for the sample ACFs and ACVFs. The interpretation of these stable limit distributions needs further investigation.

The case $\kappa \in (4,8)$. We commence with the sequences (X_t^2) and (σ_t^2). Due to the difficulty of verifying condition (5.1) directly we establish joint convergence of the sample ACVF directly from the point process convergence

$$N_n^{(2)} = \sum_{t=1}^{\infty} \varepsilon_{a_n^{-2}(X_t^2,\sigma_t^2,\ldots,X_{t+h}^2,\sigma_{t+h}^2)} \xrightarrow{d} N^{(2)} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{P_i^2 \mathbf{Q}_{ij}^{(2)}}\,, \tag{5.5}$$

where

$$\mathbf{Q}_{ij}^{(2)} = \Big((|Q_{ij,X}^{(m)}|^2, |Q_{ij,\sigma}^{(m)}|^2)\,, m=0,\ldots,h \Big)\,.$$

We follow the ideas of proof in Section 4 of Davis and Mikosch [12].

We start with the σ^2–sequence. We only establish joint convergence of $(\gamma_{n,\sigma^2}(0),\gamma_{n,\sigma^2}(1))$, since the extension to arbitrary lags is a straightforward generalisation. Recall that $A_{t+1} = \alpha_1 Z_t^2 + \beta_1$. Now, using the representation (2.3) and the CLT for $(\sigma_t^2 A_{t+1})$ we obtain

$$\begin{aligned}
a_n^{-4}\sum_{t=1}^{n}(\sigma_{t+1}^4 - E\sigma^4) &= a_n^{-4}\sum_{t=1}^{n}\Big(\big(\alpha_0 + \sigma_t^2 A_{t+1}\big)^2 - E\sigma^4 \Big) \\
&= a_n^{-4}\sum_{t=1}^{n}\big(\sigma_t^4 A_{t+1}^2 - E\sigma^4 EA^2\big) + o_P(1) \\
&= a_n^{-4}\sum_{t=1}^{n}\sigma_t^4 (A_{t+1}^2 - EA^2) + a_n^{-4} EA^2 \sum_{t=1}^{n}\big(\sigma_t^4 - E\sigma^4\big) + o_P(1)\,.
\end{aligned}$$

We conclude that for every $\epsilon > 0$,

(5.6) $\quad (1-EA^2)\, n a_n^{-4}\big(\gamma_{n,\sigma^2}(0) - E\sigma^4\big)$

$$
\begin{aligned}
&= a_n^{-4} \sum_{t=1}^{n} \sigma_t^4 (A_{t+1}^2 - EA^2) + o_P(1) \\
&= a_n^{-4} \sum_{t=1}^{n} \sigma_t^4 (A_{t+1}^2 - EA^2) I_{\{\sigma_t > a_n \epsilon\}} + a_n^{-4} \sum_{t=1}^{n} \sigma_t^4 (A_{t+1}^2 - EA^2) I_{\{\sigma_t \le a_n \epsilon\}} + o_P(1) \\
&= \mathrm{II} + \mathrm{I} + o_P(1)\,.
\end{aligned}
$$

Now, by Karamata's theorem (see for example Bingham et al. [5]) as $n \to \infty$,

$$
\mathrm{var}(\mathrm{I}) = const\; n(\epsilon a_n)^{-8} E\sigma^8 I_{\{\sigma_t \le a_n \epsilon\}} \sim const\; \epsilon^{8-\alpha} \to 0 \quad \text{as } \epsilon \to 0. \tag{5.7}
$$

As for II, let $\mathbf{x}_t = (x_{t,X^2}^{(0)}, x_{t,\sigma^2}^{(0)}, \ldots, x_{t,X^2}^{(h)}, x_{t,\sigma^2}^{(h)}) \in \overline{\mathbb{R}}^{h+1} \backslash \{\mathbf{0}\}$ and define the mappings

$$
T_{m,\epsilon,X^2} : \mathcal{M} \to \overline{\mathbb{R}}
$$

by

$$
\begin{aligned}
T_{0,\epsilon,X^2}\left(\sum_{i=1}^{\infty} n_i \varepsilon_{\mathbf{x}_i}\right) &= \sum_{i=1}^{\infty} n_i (x_{i,X^2}^{(0)})^2 I_{\{|x_{i,X^2}^{(0)}| > \epsilon\}}\,, \\
T_{1,\epsilon,X^2}\left(\sum_{i=1}^{\infty} n_i \varepsilon_{\mathbf{x}_i}\right) &= \sum_{i=1}^{\infty} n_i (x_{i,X^2}^{(1)})^2 I_{\{|x_{i,X^2}^{(0)}| > \epsilon\}}\,, \\
T_{m,\epsilon,X^2}\left(\sum_{i=1}^{\infty} n_i \varepsilon_{\mathbf{x}_i}\right) &= \sum_{i=1}^{\infty} n_i x_{i,X^2}^{(0)} x_{i,X^2}^{(m-1)} I_{\{|x_{i,X^2}^{(0)}| > \epsilon\}}\,, \quad m \ge 2\,,
\end{aligned}
$$

and the mappings T_{m,ϵ,σ^2} are defined analogously. Since the set $\{\mathbf{x} \in \overline{\mathbb{R}}^{2(h+1)} \backslash \{\mathbf{0}\} : |x^{(l)}| > \epsilon\}$ for any $l \ge 0$ is bounded, the CLT and the convergence in (5.5) imply that

$$
\begin{aligned}
\mathrm{II} &= a_n^{-4} \sum_{t=1}^{n} \left(\left(\sigma_t^2 A_{t+1} + \alpha_0\right)^2 - \sigma_t^4\, EA^2 \right) I_{\{\sigma_t > a_n \epsilon\}} + o_P(1) \\
&= a_n^{-4} \sum_{t=1}^{n} \left(\sigma_{t+1}^4 - \sigma_t^4\, EA^2 \right) I_{\{\sigma_t > a_n \epsilon\}} + o_P(1) \\
&= T_{1,\epsilon,\sigma^2} N_n^{(2)} - EA^2\; T_{0,\epsilon,\sigma^2} N_n^{(2)} \stackrel{d}{\to} T_{1,\epsilon,\sigma^2} N^{(2)} - EA^2\; T_{0,\epsilon,\sigma^2} N^{(2)} + o_P(1) =: S(\epsilon,\infty)\,,
\end{aligned}
$$

where $ES(\epsilon, \infty) = 0$. Using again (5.7) and the argument in Davis and Hsing [11] on pp. 897–898, $S(\epsilon, \infty) \stackrel{d}{\to} V_0^*$, say, as $\epsilon \to 0$. Turning to (5.6), we finally obtain

$$
n a_n^{-4} \left(\gamma_{n,\sigma^2}(0) - E\sigma^4 \right) \stackrel{d}{\to} \frac{1}{1 - EA^2}\, V_0^* =: V_0\,.
$$

As for $\gamma_{n,\sigma^2}(1)$, we proceed as above and write

$$
a_n^{-4} \sum_{t=1}^{n} \left(\sigma_t^2 \sigma_{t+1}^2 - \gamma_{\sigma^2}(1) \right)
$$

$$
\begin{aligned}
&= a_n^{-4}\sum_{t=1}^{n}\left(\sigma_t^2\left(\alpha_0+\sigma_t^2 A_{t+1}\right)-\gamma_{\sigma^2}(1)\right)\\
&= a_n^{-4}\sum_{t=1}^{n}\left(\sigma_t^4 A_{t+1}-E\sigma_t^4 EA\right)+o_P(1)\\
&= a_n^{-4}\sum_{t=1}^{n}\sigma_t^4(A_{t+1}-EA)I_{\{\sigma_t>a_n\epsilon\}}+a_n^{-4}\sum_{t=1}^{n}\sigma_t^4(A_{t+1}-EA)I_{\{\sigma_t\le a_n\epsilon\}}\\
&\quad +EA\; na_n^{-4}\left(\gamma_{n,\sigma^2}(0)-\gamma_{\sigma^2}(0)\right)+o_P(1)\\
&= \mathrm{III}+\mathrm{IV}+\mathrm{V}+o_P(1)\,.
\end{aligned}
$$

As for I, $\lim_{\epsilon\to 0}\limsup_{n\to\infty}\mathrm{var}(\mathrm{IV})=0$. Moreover,

$$
\begin{aligned}
\mathrm{III}+\mathrm{V} &= a_n^{-4}\sum_{t=1}^{n}(\sigma_t^2\sigma_{t+1}^2-\sigma_t^4\, EA)I_{\{\sigma_t>a_n\epsilon\}}+EA\; na_n^{-4}\left(\gamma_{n,\sigma^2}(0)-\gamma_{\sigma^2}(0)\right)+o_P(1)\\
&= T_{2,\epsilon,\sigma^2}N_n^{(2)}-EA\; T_{1,\epsilon,\sigma^2}N_n^{(2)}+EA\; na_n^{-4}\left(\gamma_{n,\sigma^2}(0)-\gamma_{\sigma^2}(0)\right)+o_P(1)\\
&\stackrel{d}{\to} T_{2,\epsilon,\sigma^2}N^{(2)}-EA\; T_{1,\epsilon,\sigma^2}N^{(2)}+EA\; V_0 \quad \text{as } n\to\infty\\
&\stackrel{d}{\to} V_1^*+EA\; V_0=:V_1 \quad \text{as } \epsilon\to 0.
\end{aligned}
$$

It also follows from Davis and Hsing [11], pp. 897–898, that (V_0,V_1) is jointly $\kappa/4$–stable. The general result for a finite number of sample autocovariances can be obtained in an analogous way.

The weak convergence of $na_n^{-4}(\gamma_{n,X^2}(m)-\gamma_{X^2}(m), m=0,\dots,h)$ follows the same patterns and can indeed be reduced to the convergence of linear combinations of the sample ACVF of the σ^2–sequence. Notice that the condition $\alpha_1>0$ is necessary for the existence of a regularly varying tail for X with index κ provided $E|Z|^\kappa<\infty$; see Remark 2.4. We can write

$$
Z_t^2=\alpha_1^{-1}\left((\alpha_1 Z_t^2+\beta_1)-\beta_1\right)=\alpha_1^{-1}(A_{t+1}-\beta_1)\,, \tag{5.8}
$$

and so, using the CLT, for $m\ge 1$,

$$
\begin{aligned}
&na_n^{-4}(\gamma_{n,X^2}(m)-\gamma_{X^2}(m))\\
&= a_n^{-4}\sum_{t=1}^{n}\left[\sigma_t^2 Z_t^2\sigma_{t+m}^2 Z_{t+m}^2-\gamma_{X^2}(m)\right]\\
&= a_n^{-4}\alpha_1^{-2}\sum_{t=1}^{n}\left[\sigma_t^2\sigma_{t+m}^2(A_{t+1}-\beta_1)(A_{t+m+1}-\beta_1)-\alpha_1^2\gamma_{X^2}(m)\right]\\
&= a_n^{-4}\alpha_1^{-2}\sum_{t=1}^{n}\left[\left(\sigma_{t+1}^2\sigma_{t+m+1}^2-\beta_1\sigma_t^2\sigma_{t+m+1}^2-\beta_1\sigma_{t+1}^2\sigma_{t+m}^2+\beta_1^2\sigma_t^2\sigma_{t+m}^2\right)\right.
\end{aligned}
$$

$$
\begin{aligned}
& \quad -\big((1+\beta_1^2)E\sigma_0^2\sigma_m^2 - \beta_1 E\sigma_0^2\sigma_{m+1}^2 - \beta_1 E\sigma_0^2\sigma_{m-1}^2\big)\big] + o_P(1) \\
&= \quad n a_n^{-4}\alpha_1^{-2}\left[(1+\beta_1^2)(\gamma_{n,\sigma^2}(m) - \gamma_{\sigma^2}(m))\right. \\
& \quad -\beta_1(\gamma_{n,\sigma^2}(m+1) - \gamma_{\sigma^2}(m+1)) - \beta_1(\gamma_{n,\sigma^2}(m-1) - \gamma_{\sigma^2}(m-1))\big] + o_P(1) \\
&\stackrel{d}{\to} \quad \alpha_1^{-2}\left((1+\beta_1^2)V_m - \beta_1 V_{m+1} - \beta_1 V_{m-1}\right).
\end{aligned}
$$

For $m = 0$ one can get a similar expression for the limit variable. We omit details.

The case $\kappa \in (2,4)$. Here we consider the weak limit behaviour of the sample ACVF of the X–, $|X|$– and σ–sequences. The case (X_t) is particularly simple. Indeed, the symmetry of the random variables X_t and Karamata's theorem imply that as $n \to \infty$,

$$
\begin{aligned}
\operatorname{var}\left(a_n^{-2}\sum_{t=1}^{n-m} X_t X_{t+m} I_{\{|X_t X_{t+m}| \le a_n^2 \epsilon\}}\right) &= (n-m)a_n^{-4} E X_0^2 X_m^2 I_{\{|X_0 X_m| \le a_n^2 \epsilon\}} \\
&\sim \epsilon^{4-\kappa} \to 0 \quad \text{as } \epsilon \to 0.
\end{aligned}
$$

Hence (5.1) is satisfied and part 2 of Theorem 5.2 applies.

For the sake of illustration we only consider the weak limit behaviour of $(\gamma_{n,|X|}(0), \gamma_{n,|X|}(1))$. The other cases can be handled in a similar way. We follow the ideas of proof of Section 4 in Davis and Mikosch [12].

We have with (5.8) that

$$
(5.9) \qquad n a_n^{-2}\left(\gamma_{n,|X|}(0) - \gamma_{|X|}(0)\right) = a_n^{-2}\sum_{t=1}^{n} \sigma_t^2 (Z_t^2 - 1) + a_n^{-2}\sum_{t=1}^{n}(\sigma_t^2 - E\sigma^2).
$$

Moreover,

$$
\begin{aligned}
& n a_n^{-2}(\gamma_{n,\sigma}(0) - \gamma_\sigma(0)) \\
&= \alpha_1\, a_n^{-2}\sum_{t=1}^{n}\sigma_{t-1}^2(Z_{t-1}^2 - 1) + (\alpha_1 + \beta_1)\, a_n^{-2}\sum_{t=1}^{n}(\sigma_{t-1}^2 - E\sigma^2).
\end{aligned}
$$

Hence

$$
\begin{aligned}
& (1 - (\alpha_1 + \beta_1))\, n a_n^{-2}(\gamma_{n,\sigma}(0) - \gamma_\sigma(0)) \\
&= \alpha_1 a_n^{-2}\sum_{t=1}^{n}\sigma_t^2(Z_t^2 - 1) I_{\{\sigma_t > a_n \epsilon\}} + \alpha_1 a_n^{-2}\sum_{t=1}^{n}\sigma_t^2(Z_t^2 - 1) I_{\{\sigma_t \le a_n \epsilon\}} + o_P(1) \\
&= \mathrm{VI} + \mathrm{VII} + o_P(1).
\end{aligned}
$$

Using Karamata's theorem, one can show that $\lim_{\epsilon \to 0}\limsup_{n\to\infty}\operatorname{var}(\mathrm{VII}) = 0$. Moreover,

$$
\mathrm{VI} = a_n^{-2}\sum_{t=1}^{n}(\sigma_{t+1}^2 - (\alpha_1 + \beta_1)\sigma_t^2) I_{\{\sigma_t > a_n \epsilon\}} + o_P(1)
$$

$$
\begin{aligned}
&= T_{1,\epsilon,\sigma} N_n^{(1)} - (\alpha_1+\beta_1) T_{0,\epsilon,\sigma} N_n^{(1)} + o_P(1) \\
&\stackrel{d}{\to} T_{1,\epsilon,\sigma} N^{(1)} - (\alpha_1+\beta_1) T_{0,\epsilon,\sigma} N^{(1)} \,,
\end{aligned}
$$

and the limit of the right hand expression as $\epsilon \to 0$ is $\kappa/2$–stable. Similar arguments applied to (5.9) show that $na_n^{-2}(\gamma_{n,|X|}(0) - \gamma_{|X|}(0))$ converges in distribution to a $\kappa/2$–stable limit. Next we consider

$$
\begin{aligned}
& a_n^{-2} \sum_{t=1}^n (|X_t X_{t+1}| - E|X_0 X_1|) \\
&= a_n^{-2} \sum_{t=1}^n (|X_t|\sigma_{t+1}(|Z_{t+1}| - E|Z|) + a_n^{-2} E|Z| \sum_{t=1}^n (|X_t|\sigma_{t+1} - E|X_0|\sigma_1) \\
&= \mathrm{VIII} + \mathrm{IX}\,.
\end{aligned}
$$

Notice that there exists a constant $a \in (0,1)$ such that

$$
\begin{aligned}
(5.10)\qquad & \mathrm{var}\left(a_n^{-2} \sum_{t=1}^n \left(|X_t|(\sigma_{t+1} - \sigma_t A_{t+1}^{1/2}) - E[|X_0|(\sigma_1 - \sigma_0 A_1^{1/2})]\right)\right) \\
&= a_n^{-4} \sum_{t=1}^n \sum_{s=1}^n \mathrm{cov}\left(|X_t|(\sigma_{t+1} - \sigma_t A_{t+1}^{1/2}), |X_s|(\sigma_{s+1} - \sigma_s A_{s+1}^{1/2})\right) \\
&\le \mathit{const}\; n a_n^{-4} \sum_{h=1}^n a^h \to 0 \quad \text{as } n \to \infty.
\end{aligned}
$$

since the chain $((X_t, \sigma_t))$ is strongly mixing with geometric rate (see Davis et al. [3]), and there exists $\delta > 0$ such that

$$
E\left||X_0|(\sigma_1 - \sigma_0 A_1^{1/2})\right|^{2+\delta} < \infty\,.
$$

Indeed, in this case there exists a constant $K > 0$ such that

$$
\mathrm{cov}\left(|X_s|(\sigma_{s+1} - \sigma_s A_{s+1}^{1/2}), |X_t|(\sigma_{t+1} - \sigma_t A_{t+1}^{1/2})\right) \le K a^{|t-s|}\,.
$$

See for example Theorem 17.2.2 in Ibragimov and Linnik [29].

The expression VIII can be treated in the same way as VI+VII, using the point process convergence $N_n^{(1)} \stackrel{d}{\to} N^{(1)}$, and so, by (5.10), it remains to consider the limit of

$$
\begin{aligned}
& a_n^{-2} \sum_{t=1}^n \left(|X_t|\sigma_t A_{t+1}^{1/2} - E(|X_0|\sigma_0 A_1^{1/2})\right) \\
&= a_n^{-2} \sum_{t=1}^n \left(|X_t|\sigma_t (A_{t+1}^{1/2} - EA^{1/2})\right) + EA^{1/2}\, a_n^{-2} \sum_{t=1}^n (|X_t|\sigma_t - E(|X_0|\sigma_0)) \\
&= a_n^{-2} \sum_{t=1}^n \left(|X_t|\sigma_t (A_{t+1}^{1/2} - EA^{1/2})\right) + EA^{1/2}\, a_n^{-2} \sum_{t=1}^n \sigma_t^2 (|Z_t| - E|Z|)
\end{aligned}
$$

$$+EA^{1/2}E|Z|\, a_n^{-2} \sum_{t=1}^{n} \left(\sigma_t^2 - E\sigma^2\right) \,.$$

From the above discussion it is clear how to deal with the limits of the expressions on the right–hand side and that they are $\kappa/2$–stable. □

Remark 5.4 The same arguments of proof as for $(|X_t|)$ also apply to the sample ACFs of the sequences $(|X_t|^\delta)$ for any $\delta \in [1,2)$, i.e. these sample ACFs have weak limits in terms of stable distributions as well.

6 An empirical study of simulated data and foreign exchange rates

In this section we study the sample ACFs of foreign exchange (FX) log–returns and a fitted GARCH$(1,1)$ model. The tick–by–tick data of the exchange rate Japanese Yen–US Dollar between 1992 and 1996 were kindly provided to us by Olsen and Associates (Zürich). Since tick–by–tick data contain a strong daily seasonal effect, the seasonal component was estimated from the data which were then time transformed by using Olsen's θ–time algorithm; see Dacorogna et al. [10] for a description of this procedure. The time series we consider consists of 30–minute FX log—returns, where 30 minutes are measured in θ–time. Plots of these data and their estimated density are given in Figure 6.1. We fitted a GARCH$(1,1)$ to the FX log–returns, using quasi–maximum likelihood

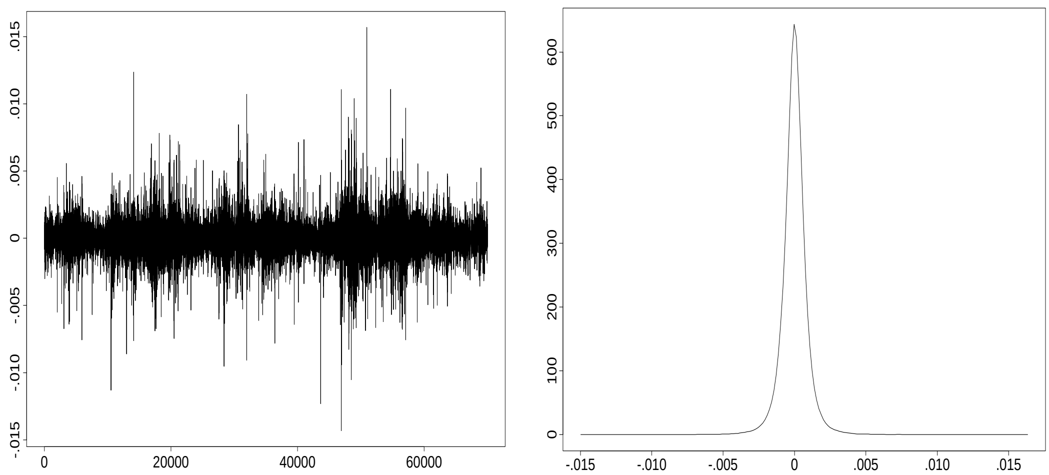

Figure 6.1 Left: *Plot of all* 70000 *JPY–USD FX log–returns.* Right: *Their density.*

estimation as for example explained in Gourieroux et al. [22], cf. Gourieroux [21]. The following parameters were obtained:

$$\alpha_0 = 10^{-7}\,, \quad \alpha_1 = 0.11\,, \quad \beta_1 = 0.88\,. \tag{6.1}$$

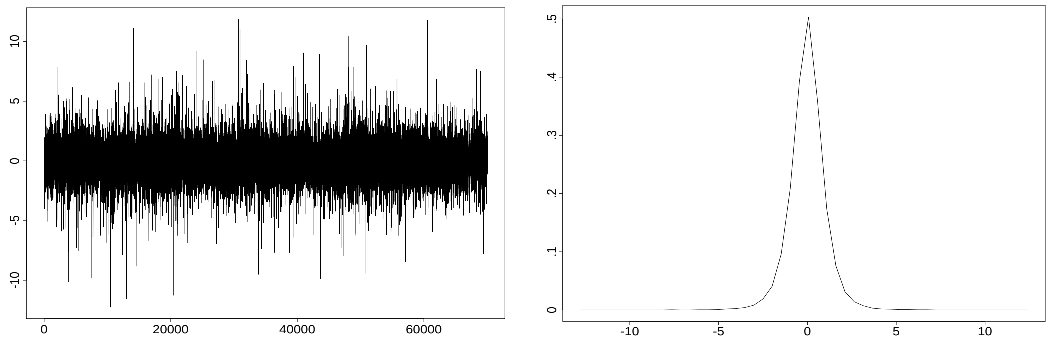

Figure 6.2 Left: *Plot of the residuals of the JPY–USD FX log–returns.* Right: *Their density.*

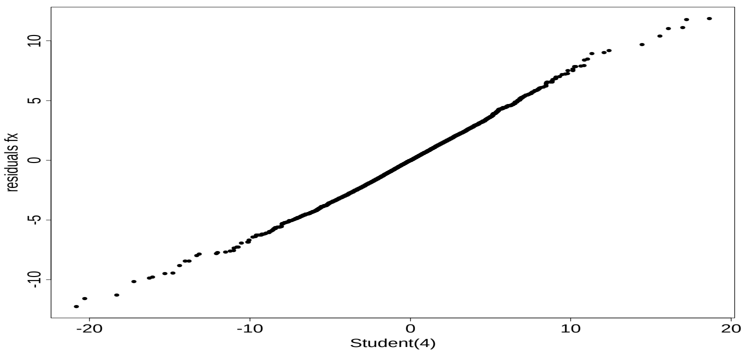

Figure 6.3 *QQ–plot of the residuals against the quantiles of a Student*(4) *distribution.*

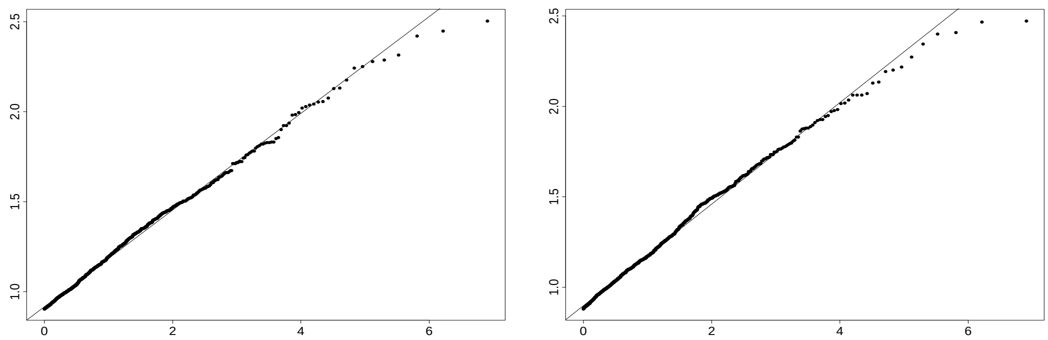

Figure 6.4 Left: *Plot of the logarithms of the* 1000 *largest residuals against the* $i/1000$*–quantiles of a standard exponential distribution;* $i = 1, \ldots, 1000$. *The fitted straight line has slope* 3.56 *indicating that the tail is Pareto–like with this tail parameter.* Right: *Plot of the absolute values of the logarithms of the* 1000 *smallest residuals against the* $i/1000$*–quantiles of a standard exponential.* (*Slope* 3.72).

Notice that $\alpha_1 + \beta_1 = 0.99$, a value which is very close to one. This is a typical situation for various financial time series; see for example Engle and Bollerslev [17], Baillie and Bollerslev [2] or Guillaume et al. [24].

The residuals

Given estimates $\widehat{\alpha}_1$ and $\widehat{\beta}_1$, one can calculate the *residuals*

$$\widehat{Z}_t = X_t / \widehat{\sigma}_t \,, \quad \text{where} \quad \widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \widehat{\sigma}_{t-1}^2 + \widehat{\beta}_1 X_{t-1}^2 \,.$$

Clearly, one has to choose an initial value $\widehat{\sigma}_1$. However, empirical evidence from simulations, paired with the theoretical Remark 2.1, show convincingly that the values $\widehat{\sigma}_t$ for large t are quite insensitive to the initial value.

For notational ease, we will not distinguish between the estimates $\widehat{\alpha}_1$, $\widehat{\beta}_1$, $\widehat{\sigma}_t$ and $\widehat{Z}_t$ and the true values α_1, β_1, σ_t and Z_t. Figure 6.2 presents the residuals of the FX log–returns with parameters (6.1) and their density. The latter shows that the distribution is roughly symmetric and very much peaked around zero ("leptokurtic").

Figure 6.3 displays the QQ—plot of the residuals against the quantiles of the standard Student(4) distribution. The plot indicates an overall good fit of the Student distribution to the estimated residuals. The tails of a Student(4) random variable S are Pareto-like, i.e. $P(S > x) \sim cx^{-4}$. The hypothesis of a non–Gaussian heavy–tailed distribution of the residuals is strongly supported also by the plot in Figure 6.4. The log–log plot estimation yields for the right tail a point estimate of 3.56 with a 95% asymptotic confidence interval [3.24,3.88] and for the left tail a point estimate of 3.72 with a confidence interval [3.39,4.05]. The theory for the asymptotic confidence intervals is given by Kratz and Resnick [32]. Based on this estimation procedure we cannot reject the hypothesis that the right and left tails are equally heavy. In the literature Student and generalised Pareto distributions, as heavy–tailed distributions, were fitted to the residuals; see for example Baillie and Bollerslev [2] and McNeil and Frey [35]. As regards the tails of the distribution of Z, the choice of a Student distribution is certainly closer to reality than the assumption of normality.

However, the tails of X and σ are determined by the centre *and* the tails of the distribution of Z; see the discussion below. For this reason, in this paper we do not give a precise parametric description of the distribution of the residuals of the FX log–returns. For our purposes, it is more realistic to work with the empirical distribution derived from the 70000 historical data under the hypothesis of a GARCH(1,1) with parameters (6.1).

The tails

In agreement with the theories of Sections 2 and 5, we calculated the tail parameter $\kappa = 2.25$ as the solution of the equation

$$E|\alpha_1 Z^2 + \beta_1|^{\kappa/2} = 1 \,, \tag{6.2}$$

where we replaced α_1 and β_1 with their estimated values and the expectation with its empirical version based on the residuals, i.e.

$$\frac{1}{n}\sum_{i=1}^{n}|\alpha_1 Z_i^2+\beta_1|^{\kappa/2}=1\,. \tag{6.3}$$

The theoretical basis for this approach has been provided by Pitts et al. [23]. There it is shown that the solution $\widehat{\kappa}$ to (6.3) is asymptotically normal with mean κ and variance

$$s^2=\frac{1}{n}\frac{E|A|^{\kappa}-1}{E(|A|^{\kappa/2}\ln|A|)}\,,\quad \text{where } A=\alpha_1 Z^2+\beta_1\,, \tag{6.4}$$

provided that $E|A|^{\kappa+\delta}<\infty$ for some $\delta>0$. If the latter condition is not satisfied one can still show consistency of $\widehat{\kappa}$ under the assumption $E|A|^{\kappa/2+\delta}<\infty$; see Embrechts and Mikosch [15], Lemma 3.3. Replacing in (6.4) the expectations by their sample analogues, we obtain the asymptotic confidence band $[2.25-0.02, 2.25+0.02]$ for κ. The latter confidence bands have to be treated with caution since the existence of the 4.5th moment of X (which is necessary for the above asymptotic theory for $\widehat{\kappa}$ does not seem to be guaranteed for the FX log–returns; cf. Figures 6.3 and 6.4.

Choosing in (6.3) a Student distribution for Z with d degrees of freedom, implied by the tail estimates of the residuals, we get tail indices close to 2.25: for $d=3.25, 3.5, 3.75, 4$ one has $\kappa=2.26, 2.34, 2.4, 2.46$, respectively. This provides another argument for the *overall good fit* of the Student distribution as a model of the noise.

For the GARCH$(1,1)$ process with parameters (6.1) estimated from the FX log–returns, we thus may conclude (see Section 2) that

$$P(\sigma>x)\sim c_0x^{-2.25}\,,\quad P(X>x)\sim E|Z|^{2.25}P(\sigma>x)\,,\quad x\to\infty\,.$$

This implies that $EX^2<\infty$, but $EX^4=\infty$.

The sample ACFs

Figure 6.5 shows the sample ACFs of the FX log–returns, their absolute values, squares and third powers at the first 300 lags. Confidence bands were derived from 1000 independent simulations of the sample ACFs at these lags. The underlying time series is a GARCH$(1,1)$ process with parameters (6.1). To be as close to the real–life FX log–returns as possible, the corresponding iid noise was generated from the empirical distribution of the residuals of the FX log–returns. The densities of the corresponding sample autocorrelations at lag 1 are presented in Figure 6.7. They are also obtained from the 1000 independent simulations.

The interpretation of the densities in Figure 6.7 and the sample ACFs in Figure 6.5 very much depends on how heavy the tails of the X_ts are. Using the above findings of a Pareto–like tail for

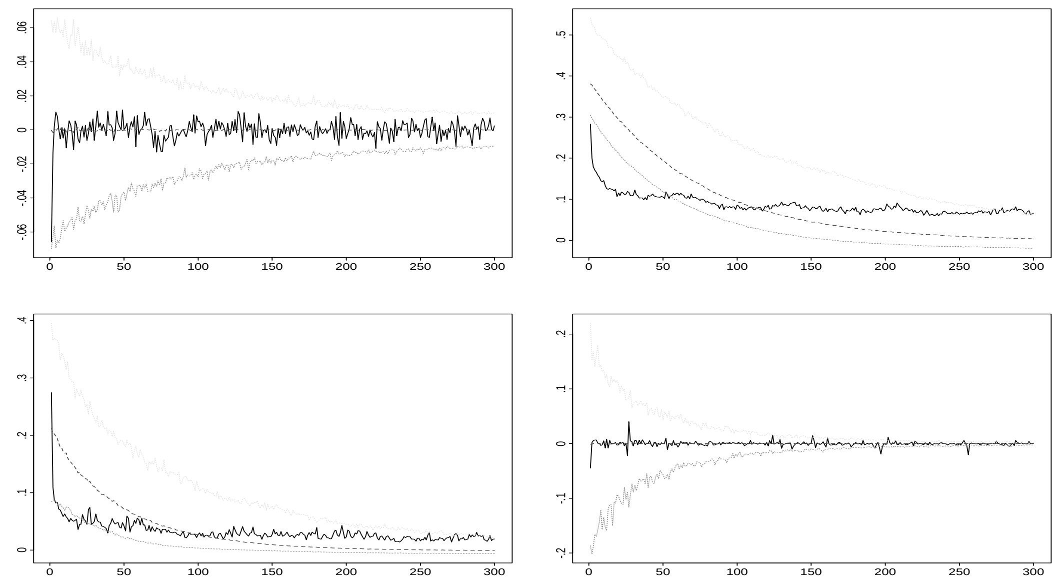

Figure 6.5 *The sample ACF of the FX log–returns* (top, left)*, their absolute values* (top, right)*, squares* (bottom, left) *and third powers. The upper and lower solid curves indicate the* 2.5%*– and* 97.5%*–quantiles of the distributions of the sample ACFs at a fixed lag. The dotted line corresponds to the median of those distributions. In the top left and bottom right figures the median curve and the sample ACF curve are almost indistinguishable.*

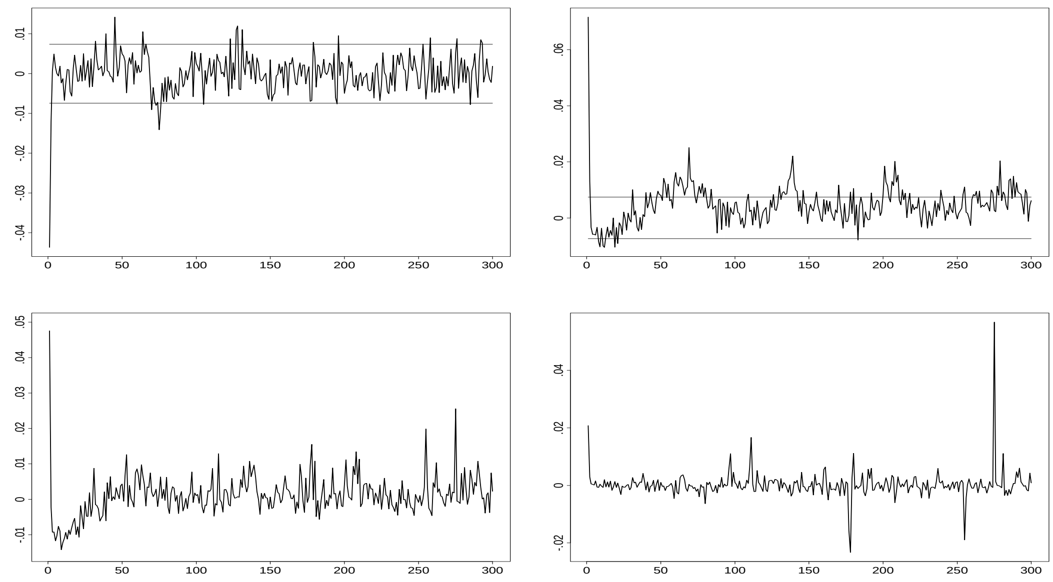

Figure 6.6 *The sample ACFs of the residuals of the FX log–returns* (top, left)*, their absolute values* (top, right)*, squares* (bottom, left) *and third powers. The straight lines in the two upper graphs indicate the* $\pm 1.96/\sqrt{n}$ *asymptotic confidence bands for an iid sequence with finite second moment. In the lower two graphs we refrain from giving* $\sqrt{n}$*–confidence bands because* Z *possibly has an infinite* 4*th moment. Compare with the sample ACFs of the FX log–returns; see Figure 6.5. Mind the difference in scale!*

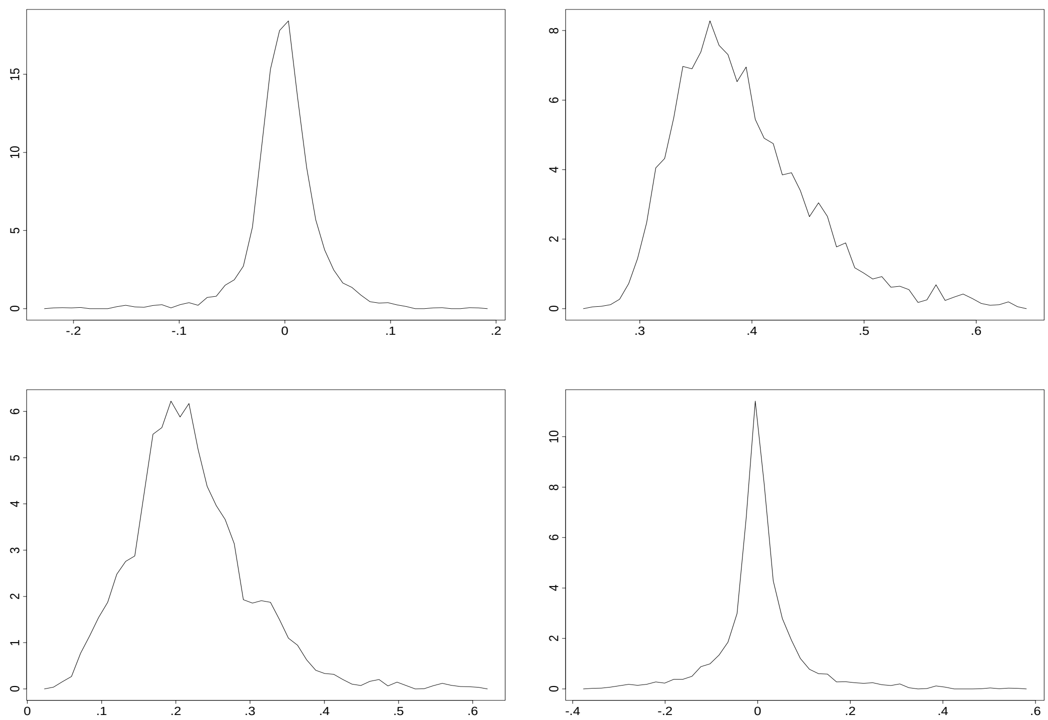

Figure 6.7 *The densities of the sample autocorrelation of a* GARCH(1, 1) *with parameters* (6.1) *at lag* 1*: for the* GARCH(1, 1) (top, left)*, its absolute values* (top, right)*, its squares* (bottom, left) *and third powers.*

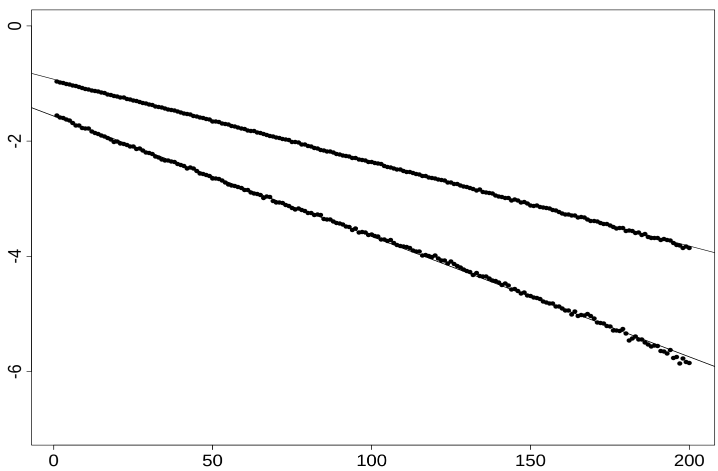

Figure 6.8 *The logarithms of the medians for the distributions of the sample ACFs of the* GARCH(1, 1) *with parameters* (6.1) *at the first* 200 *lags. The upper curve corresponds to the absolute values* (*fitted regression line with slope* $\ln 0.98$) *and the lower one to the squares* (*slope* $\ln 0.99$)*.*

X and the theory of Section 5, we conclude that the limit distribution of the sample ACFs of the X_ts and $|X_t|$s have infinite variance $\kappa/2$–stable limits with rate of convergence $(na_n^{-2})^{-1} \sim const\ n^{-1+\kappa/2} = const\ n^{0.125}$. Notice that $n^{0.125} = 70000^{0.125} = 4.03$. Thus, despite the large sample size, the asymptotic confidence bands for $\rho_{n,X}(h)$ and $\rho_{n,|X|}(h)$ are huge. This observation is supported by the bands in Figure 6.5, but even more by the densities in Figure 6.7. Notice that $\rho_{n,X}(h) \stackrel{\text{a.s.}}{\to} \rho_X(h)$ and $\rho_{n,|X|}(h) \stackrel{\text{a.s.}}{\to} \rho_{|X|}(h)$, fact which explains the peaks of the densities in Figure 6.7 (top). The slow rate of convergence of these estimators in combination with the extremely heavy tails of the limit distribution raises serious questions about the meaning and quality of these estimators. This remark applies even more to the sample ACFs of the squares and third powers. In those cases, both $\rho_{n,X^2}(h)$ and $\rho_{n,X^3}(h)$ converge in distribution, i.e. these statistics do not estimate anything.

The sample ACFs at the first 50 lags, say, of the absolute values of the FX log–returns do not fall within the 95% confidence bands for the corresponding sample ACFs of the GARCH$(1,1)$ process; see Figure 6.5. This means that, even when accounting for the statistical uncertainty due to the estimation procedure, the GARCH$(1,1)$ model does not describe the second order dependence structure of the FX log–returns sufficiently accurately.

On the other hand, a look at the sample ACFs for the residuals, their absolute values, squares and third powers (Figure 6.6) convinces one that they behave very much like the sample ACF of a finite variance iid sequence or of a moving average process with very small parameters. This compliance with the theoretical requirements of the model is a remarkable feature of the GARCH$(1,1)$ process and contributed greatly to its success.

As a conclusion, the GARCH$(1,1)$ process cannot explain the long–range dependence effect observed in the sample ACFs of the FX log–returns; see Figure 6.5. Even if we take into account that the sample ACFs of the squares and third powers are not meaningful, the sample ACF of the absolute values, despite its big statistical uncertainty, should decay to zero roughly at an exponential rate. (For reasons of comparison it would be desirable to explicitly calculate the autocorrelations of the absolute values but this seems extremely difficult even if Z is normal.) This fact follows from the strong mixing property of the GARCH$(1,1)$ with geometric rate. This theoretical fact is in agreement with our simulation study which shows that the medians of the distributions of the sample ACFs of the absolute values and squares of a GARCH$(1,1)$ decay at an exponential rate; see Figure 6.8.

The extremal index

In Section 4 we pointed out that the normalised partial maxima of the sequences (X_t), (σ_t) and $(|X_t|)$ have a Frechét limit distribution with parameter κ. Moreover, the reciprocal of the expected cluster size of high threshold exceedances is the extremal index. In what follows we estimate this index for both, simulated GARCH$(1,1)$ processes with parameters (6.1) and the FX log–returns. Various methods for estimating the extremal index have been proposed in the literature; see for example Hsing [26, 27] and Smith and Weissman [44]; cf. Embrechts et al. [14], Section 8.1. Below we use the so–called *blocks method*: divide the data into $k = k_n$ blocks of length $r = r_n$, where $n \approx r_n k_n$ and $k_n \to \infty$, $k_n/n \to 0$. Each block can be considered as a *cluster of exceedances.* Two quantities are of interest: the number K of blocks in which at least one exceedance of the threshold u_n occurs, and the total number N of exceedances of u_n.

We choose the following natural estimator of θ:

$$\widehat{\theta}_n = \frac{k}{n}\frac{\ln(1-K/k)}{\ln(1-N/n)} = \frac{1}{r}\frac{\ln(1-K/k)}{\ln(1-N/n)}.$$

Under general conditions, $\widehat{\theta}_n$ is consistent and asymptotically normal; cf. the aforementioned literature.

Figure 6.9 displays the results for estimating the extremal indices θ_X of the JPY–USD FX log–returns, θ_Z of their residuals and θ_σ of the volatility sequence, based on the GARCH$(1,1)$ fit with parameters (6.1). The 95% confidence bands and the median were obtained from 400 independent simulated GARCH$(1,1)$ with parameters (6.1) and the empirical distribution of the FX residuals as the distribution of the Z_ts. We choose the block size $r = 100$ and vary the threshold u_n (expressed in terms of the number of order statistics exceeding u_n) such that the first 1400 upper order statistics are taken into account for the calculation of $\widehat{\theta}_n$.

The estimates for θ_X and θ_σ of the FX log–returns lie above the 97.5% curve. This indicates that these extremal indices are larger than for the corresponding GARCH$(1,1)$ process. This means that the expected cluster size is smaller than for a GARCH$(1,1)$, i.e. there is less dependence in the tails for the returns than for the estimated —garch model. The corresponding estimates for θ_Z lie within the 95% bands for an iid sequence. This again seems to imply that the residuals very much behave like an iid sequence (with extremal index 1).

7 Concluding remarks

One of the aims of this paper was to show that, in the case of a GARCH$(1,1)$ model estimated from FX log–returns, the sample autocovariances and autocorrelations are extremely unreliable tools for model selection and validation. This is in contrast to the results of classical time series analysis. In many practical situations one has to deal with models that do not have a finite 4th moment, although their second moment may still exist. In this case, the standard theory for the sample autocorrelation function (Gaussian limit distributions and $\sqrt{n}$–rates of convergence) is not valid any more for the time series and its absolute values. The asymptotic confidence bands are much wider because of infinite variance stable limits and slow rates of convergence. In addition,

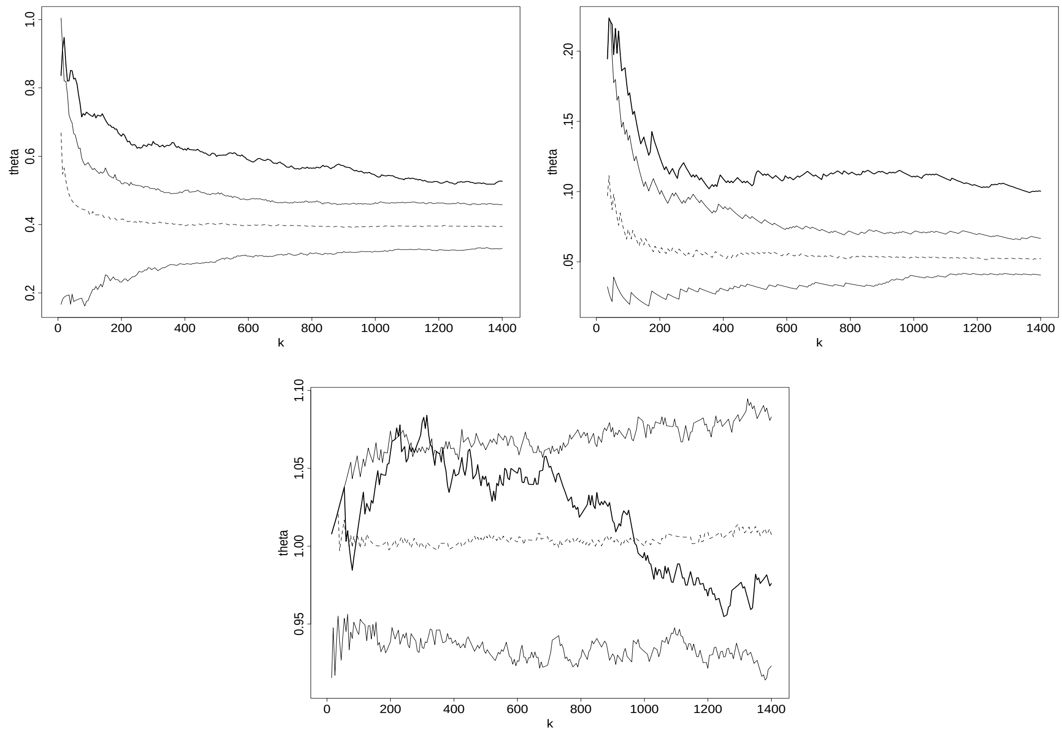

Figure 6.9 *Estimates of the extremal indices* θ_X (top left), θ_σ (top right) *and* θ_Z*, based on the blocks methods, as a function of the number k of upper order statistics. The latter corresponds to a threshold* u_n *such that exactly k order statistics of the sample are above* u_n*. In the two upper graphs, the* 3 *lower curves correspond to the* 97.5%–, 50%– *and* 2.5%*–quantiles of the estimators for a* GARCH(1,1)*. These confidence bands were obtained from* 400 *independent repetitions. The upper curve corresponds to the estimates for the FX log–returns. In the lower graph the estimates for* θ_Z *fall into the* 95% *confidence bands.*

the autocorrelations of the squares, third powers, etc., are not defined; the corresponding sample autocorrelations converge to non–degenerate limits with infinite variance.

Another conclusion of the paper is that, even when accounting for the mentioned larger–than–usual statistical uncertainty, the GARCH(1,1) (and more generally any GARCH model) cannot explain the effect of almost constant sample ACFs over various lags for the absolute values of log–returns ("long–range dependence" (LRD)). Despite the fact that there is no unique definition of the phenomenon LRD, one way to characterise it is given by $\sum_h |\rho_{|X|}(h)| = \infty$; see for example Beran [4]. The medians of the distributions of the sample autocorrelations $\rho_{n,|X|}(h)$ of a GARCH(1,1) with parameters (6.1) decay to zero at an exponential rate a^h for some $a \in (0,1)$, say. However, Figure 6.8 indicates that $a \approx 0.99$, which fact implies, for example, that $0.4a^{100} = .146$. This, to some extent, explains the observation that the function $\rho_{n,|X|}(h)$ for GARCH(1,1) models, when fitted to log–return data, decays to zero "slowly". However it does not fully explain the behaviour of the sample ACF of the absolute values of the FX log–returns; see Figure 6.5.

An alternative explanation of this phenomenon could be the non–stationarity of the underlying time series. Mikosch and Stărică [37] show that log–return data cannot be modelled by one particular GARCH model over a long period of time. There we give a change–point analysis of financial time series modelled by GARCH processes with parameters varying with time. Contrary to the common-hold belief that the LRD characteristic carries meaningful information about the price generating process, we show that the LRD behaviour could be just an artifact due to structural changes in the data. The effect of non–stationarity in real–life financial log–return time series can also be seen by considering the sample ACF of moving blocks of the same length. When taking statistical uncertainty into account, estimates of the ACF usually differ from block to block.

To summarise, the main conclusions of our study are the following:

- Although displaying potentially useful features, the GARCH$(1,1)$ model does not seem to accurately capture neither the extremal behaviour nor the correlation structure as described by the sample ACF of the data.
- The sample ACFs of GARCH$(1,1)$ models, their absolute values, squares, third powers, etc., fitted to real–life FX log–returns, are either poor estimators of the ACFs (slow convergence rates) or meaningless (non–degenerate limit distributions). **Avoid using the sample ACF to draw conclusions about the second order dependence structure of log–return data!**
- The behaviour of the sample ACF of the absolute values of real–life data (LRD effect) cannot be sufficiently explained by a GARCH process. The ACF of its absolute values decays to zero at an exponential rate. The sample ACF of FX log–returns at the first lags falls outside the 95% confidence bands for a fitted GARCH$(1,1)$ model.

In this study we restricted ourselves to the GARCH$(1,1)$ case. It is this model that appears most frequently in practical applications and that is believed to capture, despite its simplicity (3 parameters!) some of the main features of real–life log—returns. Our study sheds some light on the extent to which this belief is true. Parts of our study (convergence of point processes and sample autocorrelations) can be extended to GARCH(p,q) models, including multivariate ones, and, more generally, to solutions of certain stochastic recurrence equations; see Davis et al. [3]. However, the GARCH$(1,1)$ theory is particularly elegant because it allows one to calculate quantities such as κ explicitly. This turns out to be difficult in the general GARCH case.

Acknowledgements. We started this work when the first named author visited the Department of Mathematics at Chalmers University of Technology, Gothenburg. He would like to thank his colleagues at the Centre for Stochastic Processes, in particular Holger Rootzén, for stimulating discussions and generous financial support. The final version of this paper was written while the second named author visited the Department of Mathematics at University of Groningen. The financial support and the hospitality of the Department is acknowledged. Both authors are grateful to Olsen & Associates (Zürich) for making the FX data available to us.

References

[1] 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.

[2] BAILLIE, R.T. AND BOLLERSLEV, T. (1989) The message in daily exchange rates: a conditional-variance tale. *J. Business and Economic Statistics* **7**, 297–304.

[3] DAVIS, R.A., MIKOSCH, T. AND BASRAK, B. (1998) The sample ACF of solutions to multivariate stochastic recurrence equations. Technical Report.

[4] BERAN, J. (1994) *Statistics for Long–Memory Processes.* Monographs on Statistics and Applied Probability, No. 61. Chapman and Hall, New York.

[5] BINGHAM, N.H., GOLDIE, C.M. AND TEUGELS, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge.

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

[7] BOUGEROL, P. AND PICARD, N. (1992) Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.

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

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

[10] DACOROGNA, U.A. MÜLLER, U.A., NAGLER, R.J., OLSEN, R.B., PICTET, O.V. (1993) A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *J. Intern. Money and Finance* **12**, 413–438.

[11] DAVIS, R.A. AND HSING, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879–917.

[12] DAVIS, R.A. AND MIKOSCH, T. (1998) The sample autocorrelations of heavy–tailed processes with applications to ARCH. *Ann. Statist.*, to appear.

[13] DING, Z. AND GRANGER, C.W.J. (1996) Modeling volatility persistence of speculative returns: a new approach. *J. Econometrics* **73**, 185–215.

[14] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[15] EMBRECHTS, P. AND MIKOSCH, T. (1991) A bootstrap procedure for estimating the adjustment coefficient. *Insurance: Mathematics and Economics* **10**, 181–190.

[16] ENGLE, R.F. (ED.) (1995) *ARCH Selected Readings.* Oxford University Press, Oxford (U.K.).

[17] 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.

[18] FALK, M., HÜSLER, J. AND REISS, R.–D. (1994) *Laws of Small Numbers: Extremes and Rare Events.* Birkhäuser, Basel.

[19] FURSTENBERG, H. AND KESTEN, H. (1960) Products of random matrices. *Ann. Math. Statist.* **31**, 457–469.

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

[21] GOURIEROUX, C. (1997) *ARCH Models and Financial Applications.* Springer Series in Statistics. Springer, New York.

[22] GOURIEROUX, C., MONFORT, A. AND TROGNON, A. (1984) Pseudo maximum likelihood methods: theory. *Econometrica* **52**, 681–700.

[23] PITTS, S.M., GRÜBEL, R. AND EMBRECHTS, P. (1996) Confidence bounds for the adjustment coefficient. *Adv. in Appl. Probab.* **28** 802–827.

[24] GUILLAUME, D.M., PICTET O.V. AND DACOROGNA, M.M. (1995) On the intra-daily performance of GARCH processes. Olsen & Associates (Zürich). Internal Document DMG.1994-07-31.

[25] 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.

[26] HSING, T. (1991) Estimating the parameters of rare events. *Stoch. Proc. Appl.* **37**, 117–139.

[27] HSING, T. (1993) On some estimates based on sample behavior near high level excursions. *Probab. Theory Related Fields* **95**, 331–356.

[28] HSING, T., HÜSLER, J. AND LEADBETTER, M.R. (1988) On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* **78** 97–112.

[29] IBRAGIMOV, I.A. AND LINNIK, YU.V. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen.

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

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

[32] KRATZ, M. AND RESNICK, S.I. (1996) The QQ-estimator and heavy tails. *Commun. Statist. Stochastic Models* **12**, 699–724.

[33] LEADBETTER, M.R. (1983) Extremes and local dependence of stationary sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **65**, 291–306.

[34] LEADBETTER, M.R., LINDGREN, G. AND ROOTZÉN, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin.

[35] MCNEIL, A.J. AND FREY, R. (1998) Estimation of tail–related risk measures for heteroscedastic financial time series: an extreme value appproach. Technical Report. Departement Mathematik, ETHZ Zürich.

[36] MEYN, S.P. AND TWEEDIE, R.L. (1993) *Markov Chains and Stochastic Stability.* Springer, London.

[37] Mikosch, T. and Stărică, C. (1998) A change point analysis for foreign exchange data. Technical Report. University of Groningen. In preparation.

[38] Mori, T. (1977) Limit distributions of two–dimensional point processes generated by strong–mixing sequences. *Yokohama Math. J.* **25**, 155–168.

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

[40] Resnick, S.I. (1986) Point processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66–138.

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

[42] Resnick, S.I. (1997) Heavy tail modeling and teletraffic data. With discussion and a rejoinder by the author. *Ann. Statist.* **25**, 1805–1869.

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

[44] Smith, R.L. and Weissman, I. (1994) Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* **56**, 515–528.

[45] Stărică, C. and Pictet, O. (1997) *The tales the tails of* GARCH(1,1) *processes tell.* Technical report. Available at *www.math.chalmers.se/˜starica*

[46] Vervaat, W. (1979) On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. Appl. Probab.* **11**, 750–783.

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

Cătălin Stărică
Department of Mathematics
Chalmers University of Technology
S–412 96 Gothenburg
SWEDEN