Applied Probability Trust (10 February 2016)

THE EXTREMOGRAM AND THE CROSS-EXTREMOGRAM FOR A BIVARIATE GARCH(1, 1) PROCESS

MUNEYA MATSUI,* *Nanzan University*

THOMAS MIKOSCH,** *University of Copenhagen*

Abstract

We derive asymptotic theory for the extremogram and cross-extremogram of a bivariate GARCH(1, 1) process. We show that the tails of the components of a bivariate GARCH(1, 1) process may exhibit power-law behavior but, depending on the choice of the parameters, the tail indices of the components may differ. We apply the theory to 5-minute return data of stock prices and foreign-exchange rates. We judge the fit of a bivariate GARCH(1, 1) model by considering the sample extremogram and cross-extremogram of the residuals. The results are in agreement with the independent and identically distributed hypothesis of the two-dimensional innovations sequence. The cross-extremograms at lag zero have a value significantly distinct from zero. This fact points at some strong extremal dependence of the components of the innovations.

Keywords: Regular variation; bivariate GARCH(1, 1); Kesten's theorem; extremogram; extremal dependence

2010 Mathematics Subject Classification: Primary 60G70
Secondary 62M10; 60G10; 91B84

1. The extremogram and the cross-extremogram

In this paper we conduct an empirical study of extremal serial dependence in a bivariate return series. Our main tools for describing extremal dependence will be the *extremogram* and the *cross-extremogram.* For the sake of argument and for simplicity, we restrict ourselves to bivariate series $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$, $t \in \mathbb{Z}$. We assume that $(\mathbf{X}_t)$ has the following structure:

$$\mathbf{X}_t = \Sigma_t \mathbf{Z}_t, \qquad t \in \mathbb{Z}, \tag{1.1}$$

where $(\mathbf{Z}_t)$ constitutes an independent and identically distributed (iid) bivariate noise sequence and

$$\Sigma_t = \mathrm{diag}(\sigma_{1,t}, \sigma_{2,t}), \quad t \in \mathbb{Z},$$

* Postal address: Department of Business Administration, Nanzan University, 18 Yamazato-cho Showa-ku Nagoya, 466–8673, Japan

* Email address: mmuneya@nanzan-u.ac.jp

** Postal address: University of Copenhagen, Department of Mathematics, Universitetsparken 5, DK–2100 Copenhagen, Denmark

** Email address: mikosch@math.ku.dk

where $\sigma_{i,t}$ is the (non-negative) volatility of $X_{i,t}$. We will assume that $(\mathbf{X}_t)$, (Σ_t) constitute strictly stationary sequences and that Σ_t is predictable with respect to the filtration generated by $(\mathbf{Z}_s)_{s\le t}$. We also assume that $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t})'$ has mean zero and covariance matrix (standardized to correlations)

$$P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{1.2}$$

where $\rho = \mathrm{corr}(Z_{1,t}, Z_{2,t})$. Later, we will choose parametric models for $(\mathbf{X}_t)$ such as univariate GARCH(1,1) models both for $X_{i,t}$, $i = 1, 2$, or a vector GARCH(1,1) model; see Section 2.2 for model descriptions. In the context of these parametric models, the choice of the covariance matrix P as a correlation matrix is a matter of identifiability of the model since one can always swap a positive constant multiplier between Σ_t and $\mathbf{Z}_t$.

The extremogram and cross-extremogram for a stationary sequence $(\mathbf{X}_t)$ were introduced in Davis and Mikosch [13] and further developed in Davis et al. [14, 15]. These quantities can be defined in different ways: they are proportional to each other (such as the covariance function and the correlation function of a stationary process). In this paper we define the extremogram and cross-extremogram of a bivariate sequence $(\mathbf{X}_t)$ in standardized form such that they assume values in $[0, 1]$:

$$\begin{pmatrix} \rho_{11}(h) & \rho_{12}(h) \\ \rho_{21}(h) & \rho_{22}(h) \end{pmatrix}, \qquad h = 0, 1, 2, \ldots,$$

where

$$\begin{pmatrix} \rho_{11}(h) \\ \rho_{22}(h) \\ \rho_{12}(h) \\ \rho_{21}(h) \end{pmatrix} = \lim_{x\to\infty} \begin{pmatrix} \mathbb{P}(X_{1,h} \in xA \mid X_{1,0} \in xA) \\ \mathbb{P}(X_{2,h} \in xB \mid X_{2,0} \in xB) \\ \mathbb{P}(X_{2,h} \in xB \mid X_{1,0} \in xA) \\ \mathbb{P}(X_{1,h} \in xA \mid X_{2,0} \in xB) \end{pmatrix}. \tag{1.3}$$

Here A, B are sets bounded away from zero and we assume that these limits exist. Typically, we choose intervals $(1, \infty)$, $(-\infty, -1)$ for A, B and we also suppress the dependence on A, B in the ρ_{ij}-notation. Notice that $xA = \{xy : y \in A\}$ has interpretation as an extreme event if x is sufficiently large. Thus, the extremogram $\rho_{ii}(h)$ describes the likelihood of an extreme event at lag h given there is an extreme event in the ith component at time zero. Correspondingly, the cross-extremogram $\rho_{ij}(h)$ for $i \neq j$ describes the likelihood of an extreme event at time-lag h in the jth component given there is an extreme event at time zero in the ith component. In general, $\rho_{12}(h) \neq \rho_{21}(h)$. The limits ρ_{ij} can be understood as generalizations of the (upper) tail dependence coefficient $\rho = \lim_{q\uparrow 1} \mathbb{P}(Y > F_Y^{\leftarrow}(q) \mid X > F_X^{\leftarrow}(q))$ for a bivariate vector (X, Y) to the time series context. Here $F_X^{\leftarrow}(q)$, $F_Y^{\leftarrow}(q)$ are the q-quantiles of the distributions of X, Y, respectively. The tail-dependence coefficients have been proposed for measuring extremal dependence in a bivariate vector in the context of quantitative risk management; see e.g. McNeil et al. [25].

Moreover, each of the quantities $\rho_{ij}(h)$ has interpretation as a limiting covariance or cross-covariance function. For example,

$$\begin{aligned}\rho_{11}(h) &= \lim_{x\to\infty} \frac{\operatorname{cov}(\mathbf{1}(X_{1,0}\in xA), \mathbf{1}(X_{1,h}\in xA)) + [\mathbb{P}(X_{1,0}\in xA)]^2}{\mathbb{P}(X_{1,0}\in xA)} \\ &= \lim_{x\to\infty} \frac{\operatorname{cov}(\mathbf{1}(X_{1,0}\in xA), \mathbf{1}(X_{1,h}\in xA))}{\mathbb{P}(X_{1,0}\in xA)} \\ &= \lim_{x\to\infty} \mathbb{P}(X_{1,h}\in xA \mid X_{1,0}\in xA).\end{aligned}$$

The limits $\rho_{ij}(h)$ in (1.3) do not automatically exist. A convenient theoretical assumption for their existence is the condition of *regular variation of the time series* $(\mathbf{X}_t)$. This notion is explained in Section 2.1. Its close relationship with GARCH models is investigated in Section 2.2. Return series are often heavy-tailed and therefore it is attractive to model them by a regularly varying model. For example, under mild conditions the GARCH model automatically ensures that sufficiently high moments of the series are infinite. In particular, univariate and multivariate GARCH models exhibit power laws. This will be explained in Section 2.2. In Section 3 we investigate the tail behavior of a bivariate GARCH$(1,1)$ process. Exploiting Kesten's [21] theory for stochastic recurrence equations, we find that these processes have power-law tail behavior, possibly with distinct tail indices in each component. In Section 4 we apply this theory to the extremogram and cross-extremogram of bivariate GARCH$(1,1)$ processes. In particular, we show that these processes have exponentially decaying extremograms and, in this sense, "short serial extremal dependence". In Section 5 we apply the results to the empirical (cross-)extremograms for simulated bivariate GARCH$(1,1)$ processes. Assuming that bivariate return data have GARCH$(1,1)$ structure, we also apply the empirical (cross-)extremograms to the data and their residuals. While the data themselves exhibit some serial extremal dependence the (cross-)extremograms of the residuals (after fitting an AR-GARCH$(1,1)$ model) are in agreement with the iid hypothesis. However, the noise variables show some clear extremal dependence between the components.

2. Some preliminaries

2.1. Regularly varying time series

We say that an $\mathbb{R}^d$-valued strictly stationary time series $(\mathbf{X}_t)$ is *regularly varying with index* $\alpha > 0$ if its finite-dimensional distributions are regularly varying in the following sense: for every $h \ge 0$, the following limits in distribution exist

$$\mathbb{P}(x^{-1}(\mathbf{X}_0,\dots,\mathbf{X}_h)\in\cdot \mid |\mathbf{X}_0|>x) \xrightarrow{\mathrm{w}} \mathbb{P}((\mathbf{Y}_0,\dots,\mathbf{Y}_h)\in\cdot), \qquad x\to\infty,$$

where the limit vector $(\mathbf{Y}_0,\dots,\mathbf{Y}_h)$ has the same distribution as $|\mathbf{Y}_0|(\mathbf{\Theta}_0,\dots,\mathbf{\Theta}_h)$, the distribution of $|\mathbf{Y}_0|$ is given by $\mathbb{P}(|\mathbf{Y}_0|>y) = y^{-\alpha}$, $y>1$, and $|\mathbf{Y}_0|$ and $(\mathbf{\Theta}_0,\dots,\mathbf{\Theta}_h)$ are independent. Of course, the distribution of $\mathbf{\Theta}_0$, the *spectral measure*, is concentrated on the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x}\in\mathbb{R}^d : |\mathbf{x}| = 1\}$. The spectral measure describes the likelihood of the directions of extreme values of the lagged vector $\mathbf{X}_0$. Here $\xrightarrow{\mathrm{w}}$ denotes weak convergence and $|\cdot|$ denotes any norm in $\mathbb{R}^d$; from now on we choose the Euclidean one. The aforementioned definition of a regularly varying time series is based on work by Basrak and Segers [3] which yields a convenient description of the topic. Davis and Hsing [12] introduced the notion of a regularly varying time series which is attractive for describing serial extremal dependence in the presence of heavy tails. They used an alternative definition of multivariate regular variation which is equivalent to the definition above.

A direct consequence of the regular variation of a time series is that

$$\mathbb{P}(|\mathbf{X}_0| > x) = x^{-\alpha} L(x), \qquad x > 0, \qquad \text{for a slowly varying function } L, \tag{2.1}$$

i.e., L is a positive function on $(0,\infty)$ such that $L(cx)/L(x) \to 1$ as $x \to \infty$ for any $c > 0$. Then we also have

$$\mathbb{P}(\mathbf{X}_0/|\mathbf{X}_0| \in \cdot \mid |\mathbf{X}_0| > x) \xrightarrow{w} \mathbb{P}(\mathbf{\Theta}_0 \in \cdot), \qquad x \to \infty. \tag{2.2}$$

Regular variation of the marginal distribution of the time series is equivalent to the set of relations (2.1) and (2.2). A further consequence is that $\mathbb{P}(\mathbf{s}'\mathbf{X}_0 > x)/\mathbb{P}(|\mathbf{X}_0| > x) \to e_\alpha(\mathbf{s})$ as $x \to \infty$ for any choice of $\mathbf{s} \in \mathbb{S}^{d-1}$ and some function e_α such that $e_\alpha(\mathbf{s}_0) \neq 0$ for some $\mathbf{s}_0 \in \mathbb{S}^{d-1}$ and $e_\alpha(t\mathbf{s}) = t^{-\alpha} e_\alpha(\mathbf{s})$, $t > 0$. For proofs of the aforementioned properties and further reading on regular variation, we refer to Bingham et al. [4] and Resnick [28, 29] in the univariate and multivariate cases, respectively.

A particular consequence of the property of regular variation of a time series $(\mathbf{X}_t)$ is the fact that the limits in (1.3), leading to the extremogram and cross-extremogram, are well defined. For this reason, we will assume that $(\mathbf{X}_t)$ is regularly varying or we will assume that a deterministic monotone-increasing transformation of the components $X_{i,t}$, $i = 1, 2$, of $\mathbf{X}_t$ results in a regularly varying time series. Such transformations can be necessary, for example, if both components are not regularly varying or if both components have rather different tail behavior. These cases are relevant for real-life time series. As an example, assume that $(\mathbf{X}_t)$ is a bivariate strictly stationary Gaussian time series. This is not a regularly varying time series. However, the extremogram and cross-extremogram of this sequence exist for various sets A, B, for example, if $A = B = (1,\infty)$ (a corresponding remark applies if A or B is the set $(-\infty,-1)$). Denote the marginal distribution functions of the components $X_{i,t}$, $i = 1, 2$ by $F_{X_{i,0}}$ respectively. If G denotes the distribution function of a t-distribution with α degrees of freedom then calculation yields that

$$\big(G^{\leftarrow}(F_{X_{1,0}}(X_{1,t})), G^{\leftarrow}(F_{X_{2,0}}(X_{2,t}))\big), \qquad t \in \mathbb{Z}, \tag{2.3}$$

has G-distributed marginals and one can indeed show that the transformed time series is regularly varying with index α. The same transformation arguments apply to a non-Gaussian time series but, in contrast to a Gaussian time series, in general one cannot ensure that the resulting time series is regularly varying in the sense defined above. Given that a transformation of the type (2.3) yields a regularly varying time series, one can modify the cross-extremogram e.g. for the sets $A = B = (1,\infty)$ in the following way:

$$\begin{aligned}\rho_{12}(h) &= \lim_{q \uparrow 1} \mathbb{P}(X_{2,h} > F^{\leftarrow}_{X_{2,0}}(q) \mid X_{1,0} > F^{\leftarrow}_{X_{1,0}}(q)) \\ &= \lim_{x \to \infty} \mathbb{P}(G^{\leftarrow}(F_{X_{2,0}}(X_{2,h})) > x \mid G^{\leftarrow}(F_{X_{1,0}}(X_{1,0})) > x).\end{aligned}$$

For practical purposes, we will replace the high quantiles $F^{\leftarrow}_{X_{i,0}}(q)$, $i = 1, 2$, by their empirical versions, such as the 97%, 98%, ... component-wise empirical quantiles, depending on the sample size available.[1]

Regular variation of a time series is a convenient theoretical property but it cannot be tested on data. In what follows, we will assume a GARCH model for $(\mathbf{X}_t)$. This model ensures regular variation of the sequence.

2.2. Univariate GARCH(1,1) models

From Bollerslev [5] recall the notion of a univariate GARCH(1,1) model

$$X_t = \sigma_t Z_t, \qquad t \in \mathbb{Z}, \tag{2.4}$$

where (Z_t) is an iid unit-variance mean-zero sequence and (σ_t) is a positive volatility sequence whose dynamics are given by the causal non-zero solution to the stochastic recurrence equation

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2, \qquad t \in \mathbb{Z}. \tag{2.5}$$

Here $\alpha_0 > 0$, $\alpha_1 > 0$, and $\beta_1 \geq 0$ are constants. The probabilistic structure of (σ_t^2) can be investigated in the context of solutions to the general stochastic recurrence equation

$$Y_t = A_t Y_{t-1} + B_t, \qquad t \in \mathbb{Z}, \tag{2.6}$$

where (A_t, B_t), $t \in \mathbb{Z}$, constitutes an $\mathbb{R}_+^2$-valued iid sequence. Indeed, (σ_t^2) satisfies this equation with $B_t = \alpha_0$ and $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$. Based on the theory for these equations (see Bougerol and Picard [8]), we conclude that a strictly stationary positive solution (σ_t^2) to (2.5) exists if and only if

$$\mathbb{E}\log(\alpha_1 Z_0^2 + \beta_1) < 0 \qquad \text{and} \qquad \alpha_0 > 0. \tag{2.7}$$

In view of Jensen's inequality and since $\mathbb{E}Z_0^2 = 1$, $\mathbb{E}\log(\alpha_1 Z_0^2 + \beta_1) \leq \log\mathbb{E}(\alpha_1 Z_0^2 + \beta_1) = \log(\alpha_1 + \beta_1)$. Therefore the condition $\alpha_1 + \beta_1 < 1$ ensures strict stationarity as well as second-order stationarity of (σ_t) and (X_t), but the condition (2.7) is much more general and also allows for certain choices of α_1, β_1 such that $\alpha_1 + \beta_1 \geq 1$; see Nelson [27], Bougerol and Picard [8]. In the latter cases, $\mathbb{E}[X_0^2] = \infty$.

The solution to (2.6) has a rather surprising property which was discovered by Kesten [21]; see also Goldie [18]. Under mild conditions, there exists a positive constant c_0 such that $\mathbb{P}(Y_0 > x) \sim c_0 x^{-\alpha}$ for some $\alpha > 0$. We apply the aforementioned theory to (2.5) and get the following result which can be found in Goldie's [18] paper as regards the marginal distributions. Mikosch and Stărică [26] proved that (σ_t) and (X_t) are regularly varying time series.

Proposition 1. *Assume that $\alpha_0 > 0$, Z_0 has Lebesgue density and there exists $\alpha > 0$ such that*

$$\mathbb{E}(\alpha_1 Z_0^2 + \beta_1)^{\alpha/2} = 1, \tag{2.8}$$

and $\mathbb{E}\big[(\alpha_1 Z_0^2 + \beta_1)^{\alpha/2}\log^+(\alpha_1 Z_0^2 + \beta_1)\big] < \infty$. Then there exists a unique strictly stationary causal non-zero solution to (2.5) *and* (2.4)*, and there exists a constant $c_0 > 0$ such that*

$$\mathbb{P}(\sigma_0 > x) \sim c_0 x^{-\alpha}, \qquad x \to \infty. \tag{2.9}$$

[1] We also experimented with the corresponding high quantiles underlying the theoretical (in the case of simulations) or fitted (in the case of real-life data) GARCH models. The results for the empirical (cross-)extremograms were essentially the same as for using the empirical quantiles but the computational efforts for calculating the theoretical quantiles were substantial.

Moreover, as $x \to \infty$,

$$\mathbb{P}(X_0 > x) \sim \mathbb{E}[(Z_0^+)^\alpha]\mathbb{P}(\sigma_0 > x) \quad \text{and} \quad \mathbb{P}(X_0 \le -x) \sim \mathbb{E}[(Z_0^-)^\alpha]\mathbb{P}(\sigma_0 > x), \tag{2.10}$$

where $x^\pm = \max(0, \pm x)$*. In addition, the sequences* (σ_t) *and* (X_t) *are regularly varying with index* α*.*

Relation (2.10) is an immediate consequence of (2.9) and a result of Breiman [10] about the tails of products of independent random variables; cf. Jessen and Mikosch [20].

3. Bivariate GARCH(1,1) processes and their properties

Our next goal is to consider multivariate extensions of the GARCH$(1,1)$ model of the type described in (1.1). A simple way of doing this is by assuming that both component sequences $(X_{i,t})$, $i = 1, 2$, constitute univariate GARCH$(1,1)$ processes, i.e., $(\mathbf{X}_t)$ in (1.1) is specified via the vector recursion

$$\begin{aligned}\begin{pmatrix}\sigma_{1,t}^2\\ \sigma_{2,t}^2\end{pmatrix} &= \begin{pmatrix}\alpha_{01}\\ \alpha_{02}\end{pmatrix} + \begin{pmatrix}\alpha_{11} & 0\\ 0 & \alpha_{22}\end{pmatrix}\begin{pmatrix}X_{1,t-1}^2\\ X_{2,t-1}^2\end{pmatrix} + \begin{pmatrix}\beta_{11} & 0\\ 0 & \beta_{22}\end{pmatrix}\begin{pmatrix}\sigma_{1,t-1}^2\\ \sigma_{2,t-1}^2\end{pmatrix}\\ &= \begin{pmatrix}\alpha_{01}\\ \alpha_{02}\end{pmatrix} + \begin{pmatrix}\alpha_{11}Z_{1,t-1}^2 + \beta_{11} & 0\\ 0 & \alpha_{22}Z_{2,t-1}^2 + \beta_{22}\end{pmatrix}\begin{pmatrix}\sigma_{1,t-1}^2\\ \sigma_{2,t-1}^2\end{pmatrix},\end{aligned}$$

and $(\mathbf{Z}_t)$ is an iid sequence with covariance matrix P given in (1.2). We can apply the univariate theory to the components $(\sigma_{i,t}^2)$, $i = 1, 2$. There exist unique strictly stationary solutions $(\sigma_{i,t}^2)$, $i = 1, 2$, if and only if $\alpha_{0i} \ne 0$ and $\mathbb{E}\log^+(\alpha_{ii}Z_{i,0}^2 + \beta_{ii}) < 0$ for $i = 1, 2$, and the resulting unique bivariate processes (Σ_t) and $(\mathbf{X}_t)$ are strictly stationary. Notice that the dependence structure between the univariate component processes is then completely determined by the dependence structure of the components of the noise $(\mathbf{Z}_t)$. We can also apply Proposition 1 to get conditions for power-law tails and regular variation of the component processes of $(\mathbf{X}_t)$.

Remark 1. The crucial condition for the component-wise tail behavior is (2.8). Since the distributions of $Z_{i,0}$, $i = 1, 2$, and the parameter sets $(\alpha_{ii}, \beta_{ii})$, $i = 1, 2$, may be distinct, $X_{1,t}$ and $X_{2,t}$ will in general have different tail indices α_1 and α_2, respectively. This fact can be considered an advantage when studying multivariate return series because there is empirical evidence that the components of these series have distinct tail indices.

There exist various extensions of a univariate GARCH model to the multivariate case. We stick here to the *constant conditional correlation model* of Bollerslev [6] and Jeantheau [19]. It is the model (1.1) with specification

$$\begin{aligned}\begin{pmatrix}\sigma_{1,t}^2\\ \sigma_{2,t}^2\end{pmatrix} &= \begin{pmatrix}\alpha_{01}\\ \alpha_{02}\end{pmatrix} + \begin{pmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{21} & \alpha_{22}\end{pmatrix}\begin{pmatrix}X_{1,t-1}^2\\ X_{2,t-1}^2\end{pmatrix} + \begin{pmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{pmatrix}\begin{pmatrix}\sigma_{1,t-1}^2\\ \sigma_{2,t-1}^2\end{pmatrix}\\ &= \begin{pmatrix}\alpha_{01}\\ \alpha_{02}\end{pmatrix} + \begin{pmatrix}\alpha_{11}Z_{1,t-1}^2 + \beta_{11} & \alpha_{12}Z_{2,t-1}^2 + \beta_{12}\\ \alpha_{21}Z_{1,t-1}^2 + \beta_{21} & \alpha_{22}Z_{2,t-1}^2 + \beta_{22}\end{pmatrix}\begin{pmatrix}\sigma_{1,t-1}^2\\ \sigma_{2,t-1}^2\end{pmatrix}.\end{aligned} \tag{3.1}$$

Writing $\mathbf{W}_t = (\sigma_{1,t}^2, \sigma_{2,t}^2)'$,

$$\mathbf{B}_t = \begin{pmatrix}\alpha_{01}\\ \alpha_{02}\end{pmatrix}, \quad \text{and} \quad \mathbf{A}_t = \begin{pmatrix}\alpha_{11}Z_{1,t-1}^2 + \beta_{11} & \alpha_{12}Z_{2,t-1}^2 + \beta_{12}\\ \alpha_{21}Z_{1,t-1}^2 + \beta_{21} & \alpha_{22}Z_{2,t-1}^2 + \beta_{22}\end{pmatrix},$$

we see that we are again in the framework of the stochastic recurrence equation (2.6), but this time for vector-valued $\mathbf{B}_t$ and matrix-valued $\mathbf{A}_t$:

$$\mathbf{W}_t = \mathbf{A}_t \mathbf{W}_{t-1} + \mathbf{B}_t, \qquad t \in \mathbb{Z}. \tag{3.2}$$

Kesten [21] also provided the corresponding theory for stationarity and tails in this case. Stărică [30] dealt with the corresponding problems for vector GARCH(1, 1) processes, making use of the theory in Kesten [21], Bougerol and Picard [8] and its specialization to the tails of GARCH models in Basrak et al. [2]. In the bivariate GARCH(1, 1) case the theory in Stărică [30] can be written in a more compact form due to the representation (3.1); in the case of higher-order GARCH models (3.1) has to be written as an equation for vectors involving both σ^2- and X^2-terms at more than 1 lag.

According to Bougerol and Picard [8], (3.1) has a unique strictly stationary solution if the *top Lyapunov exponent* γ associated with the sequence $(\mathbf{A}_t)$ is negative, i.e.,

$$\gamma = \lim_{n\to\infty} n^{-1} \log\|\mathbf{A}_1 \cdots \mathbf{A}_n\| < 0, \tag{3.3}$$

where $\|\cdot\|$ denotes the matrix norm and the limit on the right-hand side exists a.s. In view of the Remark on p. 122 in [8], a sufficient condition for $\gamma < 0$ is that the matrix

$$\mathbb{E}\mathbf{A}_1 = \begin{pmatrix} \alpha_{11} + \beta_{11} & \alpha_{12} + \beta_{12} \\ \alpha_{21} + \beta_{21} & \alpha_{22} + \beta_{22} \end{pmatrix} =: \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{3.4}$$

has spectral radius smaller than 1. We assume that all entries of $\mathbb{E}\mathbf{A}_1$ are positive. Then, by the Perron–Frobenius theorem (see Lancaster [22], Section 9.2), $\mathbb{E}\mathbf{A}_1$ has a dominant single positive eigenvalue. Keeping this fact in mind, the largest positive solution to the characteristic equation $\det(\lambda I - \mathbb{E}\mathbf{A}_1) = 0$ yields the sufficient condition

$$\frac{a_{11} + a_{22}}{2} + \sqrt{\left(\frac{a_{11} - a_{22}}{2}\right)^2 + a_{12}a_{21}} < 1. \tag{3.5}$$

Next we give sufficient conditions for the regular variation of a bivariate GARCH(1, 1) process $(\mathbf{X}_t)$. The proof is based on Kesten's fundamental results [21], in particular Theorem 4. Stărică [30] gave a similar result, referring to Basrak et al. [2] for a related proof in the situation of a univariate GARCH(p, q) process (see also Fernández and Muriel [16]). We give a proof by verifying Kesten's conditions.

Proposition 2. *Consider the bivariate* GARCH(1, 1) *model and assume the following conditions:*

1. *condition* (3.3)*;*
2. $\mathbf{Z}_0$ *has Lebesgue density in* $\mathbb{R}^2$*;*
3. *there exists* $p > 0$ *such that*

$$\mathbb{E}[|\mathbf{Z}_0|^p \log^+|\mathbf{Z}_0|] < \infty \quad \text{and} \quad \mathbb{E}\left[\min_{i=1,2}\left(\sum_{j=1}^{2}(\alpha_{ij} Z_{j,0}^2 + \beta_{ij})\right)^p\right] \geq 2^{p/2}; \tag{3.6}$$

4. *all entries of* $\mathbf{A}_0$ *are positive a.s.,* $\alpha_{0i} > 0$*,* $i = 1, 2$*, and not all values* α_{ij}*,* $1 \leq i, j \leq 2$*, vanish.*

Then there exists a unique $\alpha \in (0, 2p]$ *such that*

$$0 = \lim_{n \to \infty} n^{-1} \log \mathbb{E}\big[\|\mathbf{A}_1 \cdots \mathbf{A}_n\|^{\alpha/2}\big], \tag{3.7}$$

there exists a strictly stationary causal non-zero solution $(\mathbf{X}_t)$ *to* (1.1) *with specification* (3.1)*, and* $(\mathbf{X}_t)$ *is regularly varying with index* α*. In particular, for every* $n \geq 1$*, there exists a non-null Radon measure* μ_n *on* $\overline{\mathbb{R}}^{2n} \backslash \{\mathbf{0}\}$, $\overline{\mathbb{R}} = \{-\infty, \infty\} \cup \mathbb{R}$*, such that*

$$x^{\alpha} \mathbb{P}\big(x^{-1}(\mathbf{X}_1, \ldots, \mathbf{X}_n) \in \cdot\big) \xrightarrow{\text{v}} \mu_n(\cdot), \qquad x \to \infty.$$

Here $\xrightarrow{\text{v}}$ *denotes vague convergence and the limit measures have the property* $\mu_n(t\cdot) = t^{-\alpha}\mu_n(\cdot)$, $t > 0$.

Proof. According to Kesten [21], Theorem 4, there exist

- a unique strictly stationary solution $(\mathbf{W}_t)$ to the equation (3.2),
- a positive value α and a non-negative function e_α on $\mathbb{S}^1$ such that

$$\lim_{x \to \infty} x^{\alpha/2} \mathbb{P}(\mathbf{u}'\mathbf{W}_0 > x) = e_\alpha(\mathbf{u}), \qquad \mathbf{u} \in \mathbb{S}^1, \tag{3.8}$$

  and the function e_α is positive for $\mathbf{u} \in \mathbb{S}^1$ such that $\mathbf{u} \geq \mathbf{0}$,

if the following conditions hold:

1. $\mathbf{A}_0 \geq \mathbf{0}$ and $\mathbf{B}_0 \geq \mathbf{0}$ and $\mathbf{B}_0 \neq \mathbf{0}$, where $\mathbf{C} \geq \mathbf{0}$ (respectively $> \mathbf{0}$) means all entries in $\mathbf{C}$ are non-negative (respectively positive);
2. the additive group generated by the numbers $\log \rho(\mathbf{a}_1 \cdots \mathbf{a}_n)$ is dense in $\mathbb{R}$, where $\mathbf{a}_i$ are elements in the support of the distribution of $\mathbf{A}_0$ such that $\mathbf{a}_1 \cdots \mathbf{a}_n$ has positive entries and ρ is the spectral radius;
3. condition (3.3) holds;
4. there exists $\alpha > 0$ such that (3.7) holds;
5. $\mathbb{E}[\|\mathbf{A}_0\|^{\alpha/2} \log^+ \|\mathbf{A}_0\|] < \infty$ and $\mathbb{E}[|\mathbf{B}_0|^{\alpha/2}] < \infty$.

Condition 1 holds in view of the assumptions $\mathbf{A}_0 > \mathbf{0}$ a.s. and $\mathbf{B}_0 > \mathbf{0}$.

Condition 2: we assume that $\mathbf{A}_0 > \mathbf{0}$ a.s. Therefore $\mathbf{a}_1 \cdots \mathbf{a}_n > \mathbf{0}$ for any $n \geq 1$ and any $\mathbf{a}_i$ in the support of $\mathbf{A}_0$. Since we assume a Lebesgue density for $\mathbf{Z}_0$ there exists an open set in $\mathbb{R}^2$ where this density is positive. Therefore and since not all values α_{ij} vanish, there exists a continuum of values $\rho(\mathbf{a}_1)$ for $\mathbf{a}_1$ in the support of $\mathbf{A}_0$.

Conditions 3 and 5 follow from the assumptions.

Condition 4: the existence of such an α follows from the existence of $p > 0$ such that (3.6) holds. Then $\alpha \leq 2p$.

Thus Kesten's Theorem 4 can be applied. In particular, (3.8) holds. Due to results in Boman and Lindskog [7] and since $\mathbf{W}_0$ is positive, (3.8) implies that $\mathbf{W}_0$ is regularly varying with index $\alpha/2$ in the sense of Section 2.1.

Next we show that the finite-dimensional distributions of $(\mathbf{W}_t)$ are regularly varying. By induction,

$$\mathbf{W}_t = \mathbf{\Pi}_t \mathbf{W}_0 + \mathbf{R}_t,$$

where $\mathbf{\Pi}_t = \mathbf{A}_t \cdots \mathbf{A}_1$, $\mathbf{R}_t = \sum_{i=1}^{t-1} \mathbf{A}_t \cdots \mathbf{A}_{t-i+1}\mathbf{B}_{t-i} + \mathbf{B}_t$ for $t \ge 1$, and all vectors are interpreted as column vectors. With this interpretation we write

$$(\mathbf{W}_1, \dots, \mathbf{W}_t) = (\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_t)\mathbf{W}_0 + (\mathbf{R}_1, \dots, \mathbf{R}_t), \qquad t \ge 1, \tag{3.9}$$

where $(\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_t)$, $(\mathbf{R}_1, \dots, \mathbf{R}_t)$ have moments of order $\frac{1}{2}\alpha$ with respect to the corresponding matrix norms and are independent of $\mathbf{W}_0$. Now an application of the multivariate Breiman theorem in Basrak et al. [1] yields the regular variation of the finite-dimensional distributions of the bivariate series $(\mathbf{W}_t)$ with index $\frac{1}{2}\alpha$, due to the regular variation of $\mathbf{W}_0$ with the same index. Hence $(\Sigma_t) = ((\operatorname{diag}(\mathbf{W}_t))^{1/2})$ inherits regular variation with index α. Here $\mathbf{x}^{1/2}$ for any matrix or vector $\mathbf{x}$ refers to taking square roots for all entries.

It remains to show that $(\mathbf{X}_t)$ is regularly varying with index α. We write $\widetilde{\Sigma}_t = (\operatorname{diag}(\mathbf{W}_t - \mathbf{R}_t))^{1/2}$. It is not difficult to see that $|(\Sigma_t - \widetilde{\Sigma}_t)\mathbf{Z}_t|$ is dominated by $c|\mathbf{R}_t|^{1/2}|\mathbf{Z}_t|$ for some constant c and this bound has finite α moment. Therefore

$$\lim_{x\to\infty} x^\alpha \mathbb{P}\big(|(\Sigma_1\mathbf{Z}_1, \dots, \Sigma_t\mathbf{Z}_t) - (\widetilde{\Sigma}_1\mathbf{Z}_1, \dots, \widetilde{\Sigma}_t\mathbf{Z}_t)| > x\big) = 0.$$

Since $\mathbf{W}_0$ is regularly varying with index $\frac{1}{2}\alpha$ an application of the multivariate Breiman result (see Basrak et al. [1]) shows that $(\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_t)\mathbf{W}_0$ is regularly varying with index $\frac{1}{2}\alpha$ as well. Combining these facts, we conclude that

$$(\widetilde{\Sigma}_1\mathbf{Z}_1, \dots, \widetilde{\Sigma}_t\mathbf{Z}_t) \quad \text{and} \quad (\Sigma_1\mathbf{Z}_1, \dots, \Sigma_t\mathbf{Z}_t)$$

have the same tail behavior and are regularly varying with index α; cf. Jessen and Mikosch [20]. In particular, we have

$$\begin{aligned}&\mathbb{P}\big(x^{-1/2}\big((\operatorname{diag}(\mathbf{\Pi}_1\mathbf{W}_0))^{1/2}\mathbf{Z}_1, \dots, (\operatorname{diag}(\mathbf{\Pi}_t\mathbf{W}_0))^{1/2}\mathbf{Z}_t\big) \in \cdot \mid |\mathbf{W}_0| > x\big)\\ &\quad \xrightarrow{\mathrm{w}} \mathbb{P}\big(Y_0\big((\operatorname{diag}(\mathbf{\Pi}_1\mathbf{\Theta}_0))^{1/2}\mathbf{Z}_1, \dots, (\operatorname{diag}(\mathbf{\Pi}_t\mathbf{\Theta}_0))^{1/2}\mathbf{Z}_t\big) \in \cdot\big),\end{aligned}$$

where $\mathbb{P}(Y_0 > x) = x^{-\alpha}$ for $x > 1$, Y_0 is independent of $\mathbf{\Theta}_0, \mathbf{Z}_1, \dots, \mathbf{Z}_t$, and $\mathbf{\Theta}_0$ has the spectral distribution of $\mathbf{W}_0$. □

Remark 2. In view of Kesten's result, relation (3.8) holds for any $\mathbf{u} \in \mathbb{S}^1$ and $e_\alpha(\mathbf{u}) \neq 0$ for $\mathbf{u} \ge \mathbf{0}$. In particular, for $\mathbf{u}_1 = (0, 1)$ and $\mathbf{u}_2 = (1, 0)$ we conclude that $\mathbb{P}(\sigma_{i,0} > x) \sim c_i x^{-\alpha}$ as $x \to \infty$, where both constants c_i are positive. In turn, Breiman's result [10] ensures that

$$\mathbb{P}(X_{i,0}^\pm > x) \sim \mathbb{E}[(Z_{i,0}^\pm)^\alpha]\mathbb{P}(\sigma_{i,0} > x), \qquad x \to \infty, \qquad i = 1, 2.$$

This means that the right and left tails of the distribution of $\mathbf{X}_0$ are equivalent and they have the same tail index α. Equivalent tail behavior of the component series is not necessarily an advantage as regards realistic modeling of the extremes of multivariate return models: there is statistical evidence that univariate return series have distinct tail indices. This case is more easily modeled under the assumption $\alpha_{ij} = \beta_{ij} = 0$ for $i \neq j$ (see Remark 1), where the components of $\mathbf{X}_t$ may have different tail behavior, providing more flexibility for component-wise extremes.

The crucial condition in Proposition 2 which makes the difference from Proposition 1 is the assumption that all entries of $\mathbf{A}_0$ must be positive and random. This condition is also satisfied if $\alpha_{ii} > 0$ for $i = 1, 2$, and $\alpha_{ij} = 0$ and $\beta_{ij} > 0$ for $i \neq j$, i.e., the off-diagonal elements in the matrix $\mathbf{A}_0$ may be positive constant.

The case when $\mathbf{A}_0$ is an upper or lower triangular matrix is not covered by Proposition 2. For example, assume that $\alpha_{21} = \beta_{21} = 0$. Then we have the GARCH$(1, 1)$ equation

$$\sigma_{2,t}^2 = \alpha_{02} + (\alpha_{22} Z_{2,t-1}^2 + \beta_{22})\sigma_{2,t-1}^2, \qquad t \in \mathbb{Z},$$

which can be solved and, under the conditions of Proposition 1, the solution has tail index $\frac{1}{2}\alpha_2 > 0$. Writing $C_t = \alpha_{01} + (\alpha_{12} Z_{2,t-1}^2 + \beta_{12})\sigma_{2,t-1}^2$, we get

$$\sigma_{1,t}^2 = C_t + (\alpha_{11} Z_{1,t-1}^2 + \beta_{11})\sigma_{1,t-1}^2, \qquad t \in \mathbb{Z}.$$

This is again a 1-dimensional recurrence equation but now the coefficients $(C_t, \alpha_{11} Z_{1,t-1}^2 + \beta_{11})$, $t \in \mathbb{Z}$, constitute a dependent strictly stationary sequence. Appealing to Brandt [9], a unique causal solution to this equation exists but its theoretical properties are not straightforward due to the dependence of the coefficient sequence. However, the tail of $\sigma_{1,0}^2$ is asymptotically at least as heavy as the tail of $\sigma_{2,0}^2$. Indeed, as $x \to \infty$,

$$\begin{aligned}\mathbb{P}(\sigma_{1,t}^2 > x) &\geq \mathbb{P}(C_t > x)\\ &\geq \mathbb{P}\big((\alpha_{12} Z_{2,t-1}^2 + \beta_{12})\sigma_{2,t-1}^2 > x\big)\\ &\sim \mathbb{E}\big[(\alpha_{12} Z_{2,t-1}^2 + \beta_{12})^{\alpha_2/2}\big]\mathbb{P}(\sigma_{2,t}^2 > x).\end{aligned}$$

In the last step we applied Breiman's theorem and used stationarity.

4. The extremogram and cross-extremogram for a bivariate GARCH$(1,1)$ process

Davis and Mikosch [13] showed for a univariate GARCH$(1, 1)$ process under the conditions of Proposition 1 that

$$\begin{aligned}\rho_\sigma(h) &= \lim_{x\to\infty} \mathbb{P}(\sigma_h > x \mid \sigma_0 > x)\\ &= \lim_{x\to\infty} \mathbb{P}(\sigma_h^2 > x \mid \sigma_0^2 > x) = \mathbb{E}[\min(1, \Pi_h^{\alpha/2})], \qquad h \geq 1,\end{aligned}$$

where $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$, $t \in \mathbb{Z}$ and $\Pi_h = A_h \cdots A_1$. While the value of these quantities is not known it is possible to determine their asymptotic order for large h. By convexity of the function $g(h) = \mathbb{E}[A_0^h]$ and since $g(\alpha/2) = 1$ we have $\mathbb{E}[A_0^p] < 1$ for $p < \frac{1}{2}\alpha$. Therefore

$$\rho_\sigma(h) \leq \mathbb{E}[\min(1, \Pi_h^p)] \leq \mathbb{E}[\Pi_h^p] = (\mathbb{E}[A_0^p])^h, \qquad h \geq 1,$$

and the right-hand side converges to zero exponentially fast. The extremogram of the X-sequence inherits this rate. Written in the form $\sigma_h^2 = \alpha_0 + A_h\sigma_{h-1}^2 = \Pi_h\sigma_0^2 + R_h$, since $R_h^{0.5} Z_h^+$ has a finite α moment, multiple use of Breiman's result yields

$$\begin{aligned}
\rho_X(h) &= \lim_{x\to\infty} \mathbb{P}(\sigma_h Z_h > x \mid \sigma_0 Z_0 > x) \\
&= \lim_{x\to\infty} \frac{\mathbb{P}(\min(\sigma_h Z_h^+, \sigma_0 Z_0^+) > x)}{\mathbb{P}(\sigma_0 Z_0^+ > x)} \\
&\le \limsup_{x\to\infty} \frac{\mathbb{P}(\sigma_0 \min(\Pi_h^{0.5} Z_h^+, Z_0^+) > x/2)}{\mathbb{P}(\sigma_0 Z_0^+ > x)} + \limsup_{x\to\infty} \frac{\mathbb{P}(R_h^{0.5} Z_h^+ > x/2)}{\mathbb{P}(\sigma_0 Z_0^+ > x)} \\
&= \text{const.} \frac{\mathbb{E}\big[\big(\min(\Pi_h^{0.5} Z_h^+, Z_0^+)\big)^\alpha\big]}{\mathbb{E}[(Z_0^+)^\alpha]} \\
&\le \text{const.} \, (\mathbb{E}[A_0^p])^h.
\end{aligned}$$

This means that $\rho_X(h)$ inherits the exponential rate from $\rho_\sigma(h)$. Both rates indicate that the σ- and X-sequences have rather "short extremal memory". This fact is in agreement with the empirical results of Section 5.

Similar calculations can be done in the bivariate case. We restrict ourselves to the σ-sequences. We assume the conditions of Proposition 2; in this case both components $\sigma_{i,t}^2$, $i = 1, 2$, of the vector $\mathbf{W}_t$ in (3.2) have the same tail index. Using relation (3.9), we see that

$$\begin{aligned}
\rho_{ij}(h) &= \lim_{x\to\infty} \mathbb{P}(\sigma_{j,h}^2 > x \mid \sigma_{i,0}^2 > x) \\
&= \lim_{x\to\infty} \frac{\mathbb{P}(\sigma_{j,h}^2 > x, \sigma_{i,0}^2 > x)}{\mathbb{P}(\sigma_{i,0}^2 > x)} \\
&\le \lim_{x\to\infty} \frac{\mathbb{P}(|\mathbf{W}_h| > x, |\mathbf{W}_0| > x)}{\mathbb{P}(|\mathbf{W}_0| > x)} \times \frac{\mathbb{P}(|\mathbf{W}_0| > x)}{\mathbb{P}(\sigma_{i,0}^2 > x)}.
\end{aligned}$$

The limit of the latter ratio converges to a constant by virtue of regular variation. Thus the extremograms ρ_{ij} are bounded by the extremogram $\rho_{|\mathbf{W}|}$ of $(|\mathbf{W}_t|)$ times this constant. However, (3.9) and the independence of $\mathbf{W}_0$ and $\mathbf{R}_h$ imply that for $p < \frac{1}{2}\alpha$,

$$\begin{aligned}
\rho_{|\mathbf{W}|}(h) &= \lim_{x\to\infty} \frac{\mathbb{P}(|\mathbf{W}_h| > x, |\mathbf{W}_0| > x)}{\mathbb{P}(|\mathbf{W}_0| > x)} \\
&\le \limsup_{x\to\infty} \frac{\mathbb{P}(\|\mathbf{\Pi}_h\| \, |\mathbf{W}_0| > x/2, |\mathbf{W}_0| > x)}{\mathbb{P}(|\mathbf{W}_0| > x)} + \lim_{x\to\infty} \mathbb{P}(|\mathbf{R}_h| > \tfrac{1}{2}x) \\
&= \text{const.} \, \mathbb{E}[\min(1, \|\mathbf{\Pi}_h\|^{\alpha/2})] \\
&\le \text{const.} \, \mathbb{E}[\min(1, \|\mathbf{\Pi}_h\|^p)] \\
&\le \text{const.} \, \mathbb{E}[\|\mathbf{\Pi}_h\|^p], \qquad h \ge 1.
\end{aligned}$$

The right-hand side converges to zero at an exponential rate in view of $\mathbb{E}[\|\mathbf{\Pi}_{h_0}\|^p] < 1$ for a sufficiently large h_0.

5. An empirical study of the extremogram and the cross-extremogram

5.1. Estimation of the extremogram and cross-extremogram

Davis and Mikosch [13] and Davis et al. [14] proposed estimators of the quantities $\rho_{ij}(h)$, $h \in \mathbb{Z}$, for given sets A, B bounded away from zero:

$$\widehat{\rho}_{ij}(h) = \frac{\sum_{t=1}^{n-h} \mathbf{1}(X_{j,t+h} \in \widehat{F}^{\leftarrow}_{X_{j,0}}(1-1/m) \times B, X_{i,t} \in \widehat{F}^{\leftarrow}_{X_{i,0}}(1-1/m) \times A)}{\sum_{t=1}^{n} \mathbf{1}(X_{i,t} \in \widehat{F}^{\leftarrow}_{X_{i,0}}(1-1/m) \times A)} \tag{5.1}$$

for some sequence $m = m_n \to \infty$ such that $m = o(n)$ as $n \to \infty$. In order to ensure standard asymptotic properties such as consistency and asymptotic normality, [13, 14] assumed the strong mixing condition and regular variation for the sequence $(\mathbf{X}_t)$, possibly after a monotone transformation of its components as explained in Section 2.1. The aforementioned growth conditions on the sequence (m_n) are standard in extreme-value statistics and cannot be avoided. They ensure that sufficiently high thresholds $F^{\leftarrow}_{X_{i,0}}(1-1/m)$, $i = 1, 2$, are chosen. These thresholds guarantee that a certain fraction of the data is taken which may be considered extreme as regards distance from the origin. For practical purposes, we take the corresponding empirical $(1-1/m)$-quantiles of the components. Although we do not have a theoretical justification for this replacement, we have simulation evidence that this approach works. In the GARCH$(1,1)$ case the theoretical quantiles can be determined by Monte-Carlo simulation using estimated GARCH$(1,1)$ parameters for the model. The results for the sample extremogram based on the theoretical and empirical quantiles were essentially the same, where we neglect the uncertainty of estimating parameters.

Although central limit theory can be shown for $\widehat{\rho}_{ij}$ at a finite number of lags h, the asymptotic covariance structure is not tractable. Davis et al. [14] propose two methods for the construction of credible confidence bands: the stationary bootstrap and random permutations. In this paper, we stick to the latter procedure. It is based on the simple idea that, if the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ were iid, random permutations of the sample would not change its dependence structure, hence the distributions of the empirical extremogram and cross-extremogram would not change under permutations. In what follows, we calculate the (cross-)extremograms based on 100 random permutations of the sample. First we calculate the 100 extremogram values at each lag. Then we choose the 96% empirical quantile at each lag and finally take the maximum of empirical quantiles over the lags of interest. This value is shown as a solid horizontal line in the corresponding graphs. This procedure is quick and clean: if the sample (cross-)extremogram at a given lag is above the horizontal line this is an indication of disagreement with the iid hypothesis.[2]

5.2. Simulated GARCH$(1,1)$ data

We provide a brief study of the sample (cross-)extremograms of simulated bivariate GARCH$(1,1)$ processes and their residuals. We choose bivariate GARCH$(1,1)$ models with iid bivariate t-distributed innovations $(\mathbf{Z}_t)$ with 10 degrees of freedom and covariance matrix P given in (1.2). We simulate from the model (3.1) with parameters

[2] Alternatively, one could plot the curve connecting the 96% empirical quantiles at each lag. However, this line could vary from lag to lag. We prefer to choose the line representing the maximum of all lag-wise 96% empirical quantiles, which presents a conservative confidence band.

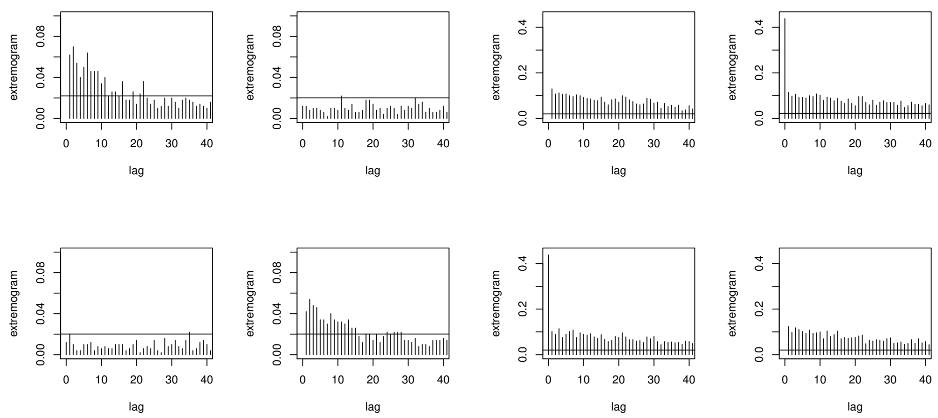

FIGURE 1: Cross-extremograms of Examples (1) (left 2×2 graphs, ref. (5.3)) and (2) (right 2×2 graphs). In both cases we observe serial extremal dependence in extremograms, whereas extremal dependence in cross-terms is observed only in (2). The reason is that (1) exhibits component-wise independence, while (2) does not by virtue of the setting $\alpha_{ij}, \beta_{ij} \neq 0$ for $i \neq j$, and $\rho \neq 0$. In the latter case, we also observe large spikes at lag zero caused by $\rho \neq 0$.

$(\alpha_{01}, \alpha_{02}) = (10^{-6}, 10^{-6})$ (the magnitude of these parameters is in agreement with values estimated from return data) and specified matrices and correlations

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix} \quad \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{12} & \beta_{22} \end{pmatrix}, \ \rho. \tag{5.2}$$

We start by considering examples with respective symmetric parameter matrices (5.2):

$$(1) \quad \begin{pmatrix} .1 & 0 \\ 0 & .1 \end{pmatrix}, \begin{pmatrix} .8 & 0 \\ 0 & .8 \end{pmatrix}, \ 0; \qquad (2) \quad \begin{pmatrix} .1 & .02 \\ .02 & .1 \end{pmatrix}, \begin{pmatrix} .8 & .04 \\ .04 & .8 \end{pmatrix}, \ .7.$$

Here we always choose small α-values while the diagonal β-values are close to 1. This is in agreement with estimated parameters on return data. We generate samples of size $n = 50,000$, using the R package 'ccgarch'[3], and calculate the (cross-)extremograms $\widehat{\rho}_{ij}(h)$ in (5.1) with $A = B = (1, \infty)$. The simulation results for Examples (1) and (2) are given in Figure 1. In each figure, (cross-)extremograms are given by 2×2 graphs as functions of time-lag h,

$$\begin{pmatrix} \widehat{\rho}_{11}(h) & \widehat{\rho}_{12}(h) \\ \widehat{\rho}_{21}(h) & \widehat{\rho}_{22}(h) \end{pmatrix}. \tag{5.3}$$

These figures indicate that small changes in the α- or β-values may lead to major changes in the extremal dependence structure. In Example (2) we also observe large

[3]Note that estimation with "ccgarch" requires choosing initial values. In most cases, we first examine component-wise univariate GARCH(1, 1) fits by the R package "fGarch" and then we choose these estimates as initial values. If the univariate estimation does not converge we try several initial values on a grid of size 0.1. In this case, the estimates sometimes differ by attaining local minima. Judging from the residuals, the eigenvalues of the estimated parameters (3.4) and the values of the likelihood functions, we choose an "optimal" estimator. Except for one case of stock-return data (see Section 5.4), this procedure works.

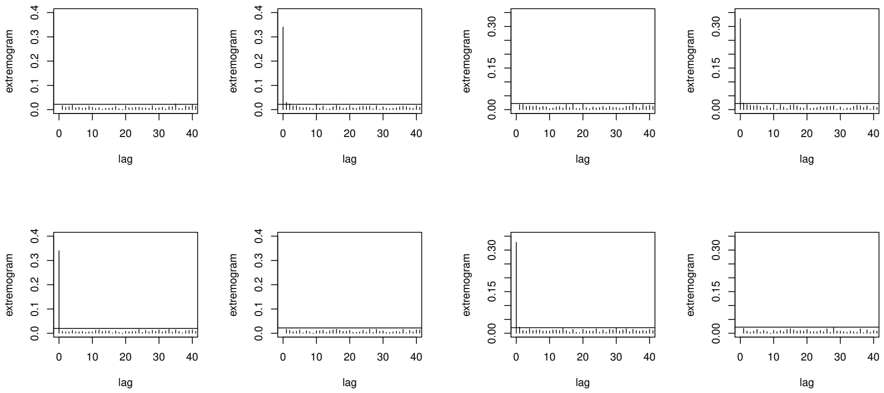

FIGURE 2: Left 2×2 graphs, ref. (5.3): (cross-)extremogram of residuals based on a bivariate QMLE fit. Right 2×2 graphs: (cross-)extremogram of residuals from component-wise MLE fits. There is no difference in both cases, where serial extremal dependencies have been removed except for lag 0.

spikes in the cross-extremograms at lag zero due to $\rho \neq 0$. This is in contrast to Example (1) with $\rho = 0$.

Our next goal is to show (cross-)extremograms of the residuals of simulated bivariate GARCH$(1,1)$ models. Although we know the innovations sequence in this case, we want to illustrate how standard maximum-likelihood estimation (MLE) techniques work. In particular, we expect that the empirical extremograms of the residuals should be close to zero. The estimation is done in two ways: (1) we fit component-wise univariate GARCH$(1,1)$ models, applying MLE and assuming Student t-distributions for the innovations; (2) following Ling and McAleer [23] (see also Francq and Zakoïan [17]), we apply bivariate Gaussian quasi-MLE (QMLE). We consider the model (3.1) with given parameter $(\alpha_{01}, \alpha_{02}) = (10^{-6}, 10^{-6})$ and parameter matrices (5.2) as follows:

$$(3) \qquad \begin{pmatrix} .1 & 0 \\ .05 & .1 \end{pmatrix}, \begin{pmatrix} .8 & .03 \\ 0 & .8 \end{pmatrix}, \ .7.$$

Component-wise univariate MLE (left) and bivariate QMLE (right) respectively yield the following estimation results:

	$\hat{\alpha}_i$	$\hat{\beta}_i$	degree for t
$i = 1$	.137	.831	9.81
$i = 2$	.169	.802	10.00

$$\begin{pmatrix} .130 & 0 \\ .056 & .125 \end{pmatrix}, \begin{pmatrix} .778 & .025 \\ .039 & .790 \end{pmatrix}, \ .7.$$

Despite the misspecification of a bivariate GARCH$(1,1)$ model, univariate estimation leads to reasonable estimation results. (Cross-)extremograms of residuals in Figure 2 indicate that extremal cross-serial dependence is not present in the residuals of both bivariate GARCH$(1,1)$ fit (Fig. 2: left) and component-wise univariate fits (Fig. 2: right). However, another example in Matsui and Mikosch [24] shows that univariate fits do not remove all cross-dependencies from the residuals (in this case the degrees of freedom were not correctly estimated). In [24] we experimented with distinct

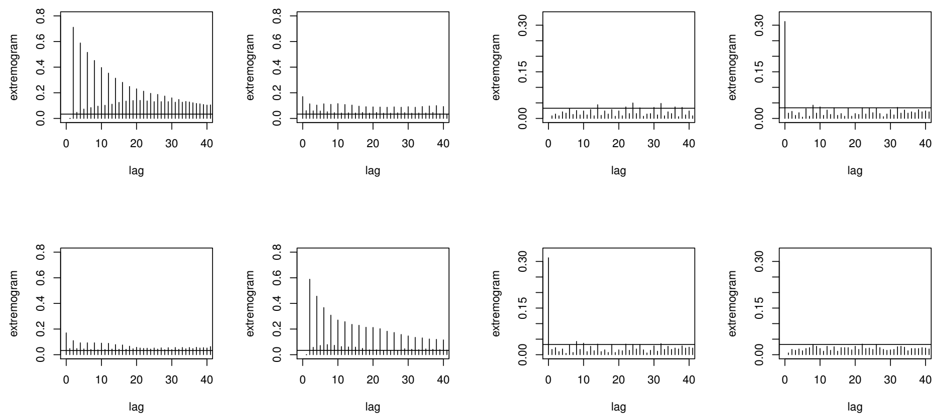

FIGURE 3: Five-minute returns of USD-DEM and USD-FRF foreign exchange rates. Left 2×2 graphs, ref. (5.3): (cross-)extremograms of the original data. The extremograms oscillate strongly while the cross-extremograms show little extremal serial dependence. Right 2×2 matrix: (cross-)extremograms of the residuals after an AR-GARCH fit. Except for lag zero, serial extremal dependence has been removed.

parameter sets close to the true ones and we also replaced univariate MLE by univariate Gaussian QMLE. In all cases, one cannot remove all cross-dependencies of the residuals. Therefore bivariate GARCH$(1,1)$ fitting is recommended if one suspects dependence in the noise sequence.

5.3. An analysis of foreign exchange rates

We analyze a bivariate high-frequency time series, consisting of 35,135 five minute returns of USD-DEM and USD-FRF foreign exchange rates. Throughout this subsection we choose the 98% component-wise sample quantiles as the threshold for the sample (cross-)extremograms. In each plot the horizontal line shows the 96% quantile obtained from 100 random permutations of the data.

The data exhibit rather strong cross-correlations and autocorrelations ([24, Figure 5]). So it is not unexpected that we also observe dependence of the extreme values of the two series. This is apparent in the extremograms of Figure 3 (left). After fitting a bivariate vector AR model of order 19 to the data (chosen by the Schwarz criterion or Bayesian information criterion, see e.g. [11, Section 9.3]), we fit a bivariate GARCH$(1,1)$ model to the residuals, by employing bivariate QMLE. The estimated matrices (5.2) are as follows:

$$\begin{pmatrix} .214 & .013 \\ .110 & .223 \end{pmatrix}, \begin{pmatrix} .697 & .008 \\ .280 & .663 \end{pmatrix}, .372.$$

which satisfy the sufficient condition for stationarity of a bivariate GARCH$(1,1)$ model; see (3.5). After the AR fit, the cross-extremograms of the residuals do not vanish although their values are small. After fitting a bivariate GARCH$(1,1)$ model to the residuals of the AR model, the residuals of the resulting AR-GARCH model exhibit extremal cross-dependence only at time-lag 0; see Figure 3 (right). This means that the components of the innovations $\mathbf{Z}_t$ exhibit extremal dependence. QQ-plots for the residuals of the vector AR and AR-GARCH models show that t-distributions with 2.5 and 3 degrees of freedom (respectively), give a good fit to the residuals.

5.4. An analysis of stock returns

We consider log-return series of three stock prices from the NY Stock Exchange: "Caterpillar Inc.", "FedEx Corporation" and "Exxon Mobil Corporation" ("cat", "fdx" and "xom" for short).[4] In each series, the raw tick-by-tick trade data have been processed into 5-minute grid data by taking the last realized trade price in each interval. Prices have been restricted to exchange trading hours 9:30 a.m. to 4:00 p.m., Monday to Friday, so that 78 data per day have been collected in the time period from 2009-02-18 9:30 to 2013-12-31 16:00.

The sample (cross-)extremograms of the log-returns of the stock prices are shown in Figure 4, where we choose the empirical 0.99 quantiles of the returns as the threshold. Although we observe typical GARCH$(1,1)$ (cross-)extremograms close to lag 0, there is a clear seasonal component in these plots, appearing as spikes at lag 78, corresponding to the beginning and end of the days. A GARCH$(1,1)$ model (bivariate or component-wise univariate) cannot explain the seasonal extremal components in the data. However, the (cross-)extremograms of the residuals after a bivariate GARCH$(1,1)$ fit show that most of the serial dependence has been removed from the data, although the seasonal component is also present in the residuals; see Figure 5. The extremograms of the log-returns are shown in Figure 4, where the thresholds are chosen as the component-wise sample quantiles.

We fit a bivariate GARCH$(1,1)$ model to each pair of stock prices, i.e., (cat, fdx), (fdx, xom) and (cat, xom). The estimated values of the bivariate QMLE (5.2) for (cat, fdx), (fdx, xom), (cat, xom), respectively, are

$$\begin{pmatrix} .215 & .210 \\ .029 & .287 \end{pmatrix}, \begin{pmatrix} .666 & .144 \\ .002 & .668 \end{pmatrix}, .55, \qquad \begin{pmatrix} .178 & .000 \\ .006 & .250 \end{pmatrix}, \begin{pmatrix} .712 & .115 \\ .007 & .666 \end{pmatrix}, .484,$$
$$\begin{pmatrix} .094 & .153 \\ .009 & .278 \end{pmatrix}, \begin{pmatrix} .789 & .094 \\ .007 & .650 \end{pmatrix}, .567.$$

The estimators of the combination (cat, fdx) are unstable and take the boundary value[5] of the sufficient condition (3.5) while the estimates for (fdx, xom) and (cat, xom) satisfy (3.5). The obvious seasonal component of the data (corresponding to the end of a trading day at lag 78) probably violates the stationary condition. Nevertheless, the standardized residuals appear "de-volatilized" in all (cross-)extremograms modulo the fact that the seasonal component is always present.

[4]We would like to thank Martin Anders Jönsson for arranging for us the stock-price data.

[5]In this case, the univariate GARCH$(1,1)$ fit does not converge. Therefore we examine several initial values for "ccgarch" on a grid of size 0.1 and choose an "optimal" value based on their likelihoods. We also tried several optimization methods included in "ccgarch". Then we calculated the eigenvalues of (3.4) from the estimates, including the "optimal" ones. However, the largest eigenvalues are very close to one in all cases. Since "ccgarch" finds the optimal value under the sufficient condition (3.5), the real optima would certainly violate (3.5).

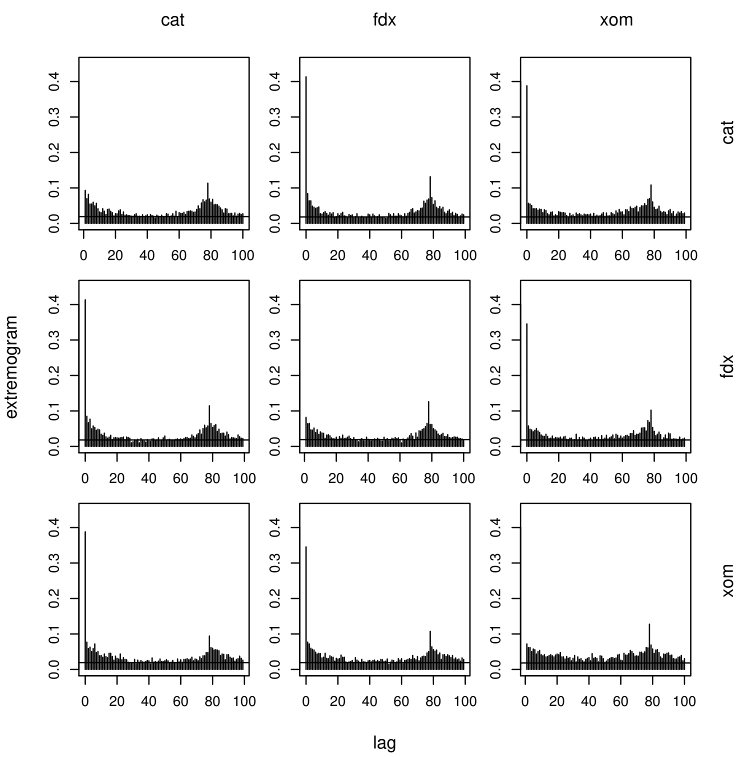

FIGURE 4: (Cross-)extremograms of log-return series of three stock prices (cat, fdx, xom). Graphs show strong serial extremal dependence in each series together with strong extremal dependence between the three series. Other than large spikes at lag 0 in cross-extremograms, we observe spikes at lag 78, which show seasonal fluctuation in a day. Moreover, extremal data around the beginning, 9:30 a.m., and the end, 4:00 p.m., may exhibit strong dependence.

Acknowledgements

Thomas Mikosch met Nick Bingham at ETH Zurich for the first time in 1992. He loves Nick's elegant work on summability and has long admired the great book Bingham et al. [4] which has proven most influential on Thomas Mikosch's work in extreme value theory, time series and risk analysis. Both authors wish Nick many more years of fruitful work and creative thinking.

Both authors thank the reviewer for careful reading and constructive criticism.

Research of MM was partly supported by the JSPS Research Fellowship for Young Scientists, the JSPS Grant-in-Aid for Young Scientists B (25870879) and Nanzan University Pache Research Subsidy I–A–2 for the 2015 academic year. Research of TM was partly supported by a Grant of the Danish Council for Free Research No. DFF–4002–00435.

References

[1] BASRAK, B., DAVIS, R. A. AND MIKOSCH. T. (2002). A characterization of multivariate regular variation. *Ann. Appl. Probab.* **12**, 908–920.

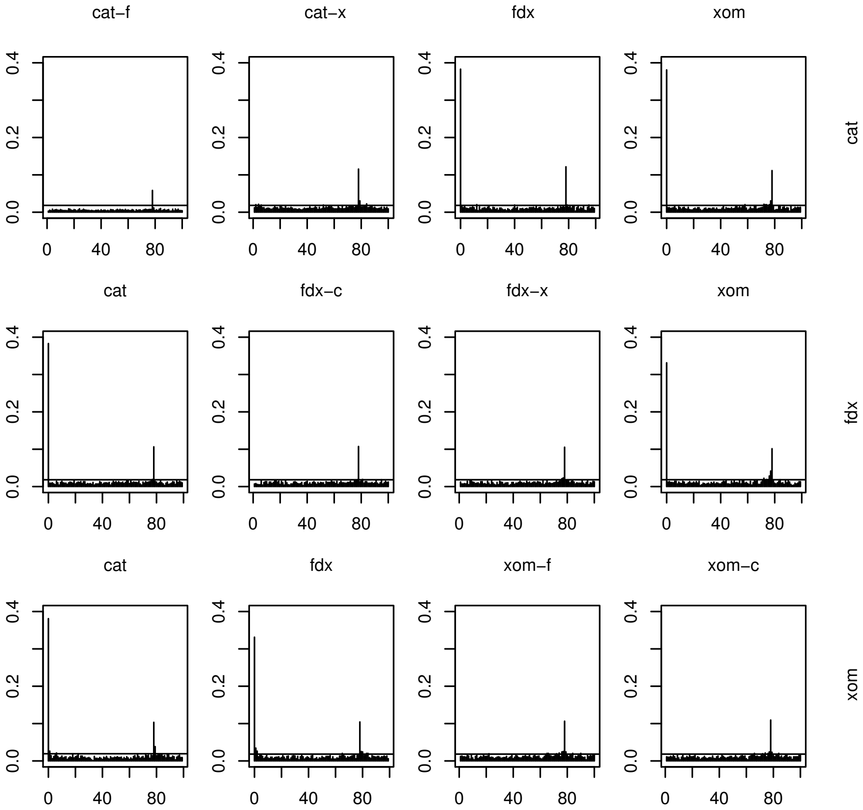

FIGURE 5: (Cross-)extremograms for residuals of bivariate GARCH$(1,1)$ fits to combinations (cat, fdx), (fdx, xom) and (xom, cat), so that we have two extremograms in each row, i.e. cat-f and cat-x are those for residuals of cat components respectively from bivariate QMLE of (cat, fdx) and (xom, cat). Other elements are cross-extremograms for residuals of (cat, fdx, xom) against (cat, fdx, xom). Although residuals show less extremal dependence except for large spikes at 0, we could not remove the seasonal component at lag 78.

[2] Basrak, B., Davis, R. A. and Mikosch, T. (2002). Regular variation of GARCH processes. *Stoch. Proc. Appl.* **99**, 95–115.

[3] Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stoch. Proc. Appl.* **119**, 1055–1080.

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

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

[6] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Rev. Economics & Statistics* **72**, 498–505.

[7] Boman, J. and Lindskog, F. (2009). Support theorems for the Radon transform and Cramér-Wold theorems. *J. Theoret. Probab.* **22**, 683–710.

[8] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some non-negative 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. Prob.* **18**, 211–220.

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

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

[12] Davis, R. A. and Hsing, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Prob.* **23**, 879–917.

[13] Davis, R. A. and Mikosch, T. (2009). The extremogram: a correlogram for extreme events. *Bernoulli* **15**, 977–1009.

[14] Davis, R. A., Mikosch, T. and Cribben, I. (2012). Towards estimating extremal serial dependence via the boostrapped extremogram. *J. Econometrics* **170**, 142–152.

[15] Davis, R. A., Mikosch, T. and Zhao, Y. (2013). Measures of serial extremal dependence and their estimation. *Stoch. Proc. Appl.* **123**, 2575–2602.

[16] Fernández, B. and Muriel, N. (2009). Regular variation and related results for the multivariate GARCH(p, q) model with constant conditional correlations. *J. Multivar. Anal.* **100**, 1538–1550.

[17] Francq, C. and Zakoïan, J. M. (2012). QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* **28**, 179–206.

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

[19] Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**, 70–86.

[20] Jessen, A. H. and Mikosch, T. (2006). Regularly varying functions. *Publ. Inst. Math. (Beograd) (N.S.)* **80(94)**, 171–192.

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

[22] Lancaster, P. (1969). *Theory of Matrices.* Academic Press, New York.

[23] Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19**, 280–310.

[24] Matsui, M. and Mikosch, T. (2015). The extremogram and the cross-extremogram for a bivariate GARCH(1, 1) process. `ArXiv:1505.05385`.

[25] McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools.* Revised edn. Princeton Univ. Press, Princeton (NJ).

[26] Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28**, 1427–1451.

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

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

[29] Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

[30] Stărică, C. (1999). Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance* **6**, 515–553.