doi:10.1111/j.1467-9892.2006.00467.x

ON A MIXTURE GARCH TIME-SERIES MODEL

By Zhiqiang Zhang, Wai Keung Li and Kam Chuen Yuen

East China Normal University and The University of Hong Kong

First Version received March 2005

Abstract. Recently, there has been a lot of interest in modelling real data with a heavy-tailed distribution. A popular candidate is the so-called generalized autoregressive conditional heteroscedastic (GARCH) model. Unfortunately, the tails of GARCH models are not thick enough in some applications. In this paper, we propose a mixture generalized autoregressive conditional heteroscedastic (MGARCH) model. The stationarity conditions and the tail behaviour of the MGARCH model are studied. It is shown that MGARCH models have tails thicker than those of the associated GARCH models. Therefore, the MGARCH models are more capable of capturing the heavy-tailed features in real data. Some real examples illustrate the results.

Keywords. GARCH; MGARCH; stochastic difference equation; tail behaviour; volatility clustering.
AMS 2000 Subject Classification. 62M10.

1. INTRODUCTION

Financial time series typically exhibit the features of heavy-tailed distribution and volatility clustering, and some kind of long-range dependence in the data. In the past two decades, various models have been proposed in order to describe these features. Among them, the generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986) has been proven to be a powerful one in capturing the first two empirical features. The GARCH(q, p) model is defined as

$$X_t = \varepsilon_t \sqrt{h_t},$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}, \quad t \in N,$$

where $q, p \geq 0$ are integers; $\alpha_0 > 0$, α_i, $\beta_j \geq 0$ are constants; and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) symmetric random variables with $\mathrm{var}(\varepsilon_1) = \sigma^2 < \infty$. The random variables ε_t are called the innovation and usually assumed to be standard-normal. The most simple and popular GARCH model is the GARCH(1,1) model, which has been extensively discussed in the finance literature. When $p = 0$, the model reduces to the ARCH model of Engle (1982).

Mikosch and Stărică (2000) pointed out 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'. In the statistics literature, mixtures of distributions have been widely used in modelling of heavy-tailed distributions (see, e.g. McLachlan and Peel, 2000). Extension of the classical mixture models to time series has been considered by Le *et al.* (1996) and Wong and Li (2000, 2001). In Wong and Li (2001), a mixture ARCH model was proposed and its application to financial data demonstrated. In the simplest two-component case, the conditional variance h_t of the process X_t has a probability p satisfying one ARCH specification and a probability $(1 - p)$ satisfying an alternative ARCH specification. Specifically, if both of the ARCH components are of order 1, we have

$$X_t = \varepsilon_t \sqrt{h_t}, \tag{1}$$

where ε_t is defined as above and

$$h_t = h_{1t} = \alpha_0 + \alpha_1 X_{t-1}^2, \quad \text{with probability } p,$$

$$h_t = h_{2t} = \alpha_0' + \alpha_1' X_{t-1}^2, \quad \text{with probability } (1-p),$$

where α_0, $\alpha_0' > 0$ and α_1, $\alpha_1' \geq 0$. Along the same lines, a mixture autoregressive process in the conditional mean of X_t satisfies

$$X_t = \phi_1 X_{t-1} + a_{1t}, \quad \text{with probability } p,$$

where a_{1t} has mean zero and conditional variance $h_{1t} = \alpha_0 + \alpha_1 a_{1t-1}^2$; or

$$X_t = \phi_1' X_{t-1} + a_{2t}, \quad \text{with probability } 1-p,$$

where a_{2t} has mean zero and conditional variance $h_{2t} = \alpha_0' + \alpha_1' a_{2t-1}^2$. This provides the simplest example of the mixture autoregressive ARCH (MAR-ARCH) process in Wong and Li (2001).

In this paper, we consider a natural extension of Wong and Li (2001) by including the past values of h_t in the specification of the conditional variance, resulting in a mixture of GARCH model. For simplicity, we will consider processes with only a mixture conditional variance structure in this paper. That is, X_t satisfies eqn (1) with a mixture GARCH (MGARCH) structure. The MGARCH model allows a much heavier tail than that of the normal GARCH model. Another example of non-Gaussian innovation models is given by Zhang (2002), where the tail behaviour of an ARCH(1) process with mixture-normal innovation was discussed. In general, the K-component MGARCH model under consideration has the form

$$X_t = \varepsilon_t \sqrt{h_t},$$

$$h_t = \xi_{0,t} + \sum_{i=1}^{q} \xi_{i,t} X_{t-i}^2 + \sum_{j=1}^{p} \eta_{j,t} h_{t-j}, \quad t \in N, \tag{2}$$

where q, $p \geq 0$ are integers; $\{\varepsilon_t,\ t \in N\}$ is a sequence of i.i.d. random variables with a common standard normal distribution; and $\{(\xi_{0,t}, \ldots, \xi_{q,t}, \eta_{1,t}, \ldots, \eta_{p,t})\}$ is independent of $\{\varepsilon_t\}$ and is a sequence of i.i.d. random vectors having joint probability

$$P(\xi_{i,t} = \alpha_{i,k}, \eta_{j,t} = \beta_{j,k}, 0 \leq i \leq q, 1 \leq j \leq p) = p_k, \quad 1 \leq k \leq K, \tag{3}$$

with $K \geq 1$ being an integer, $p_k > 0$ for $1 \leq k \leq K$, $\sum_{k=1}^{K} p_k = 1$, and all the $\alpha_{0,k}$, $\alpha_{i,k}$, $\beta_{j,k}$ being non-negative constants. We label this MGARCH model as the MGARCH(K; q, p) model in this paper. Specifically, if $p = q = 1$ and $K = 2$, then h_t in eqn (1) satisfies

$$h_t = \alpha_{0,1} + \alpha_{1,1}X_{t-1}^2 + \beta_{1,1}h_{t-1}, \quad \text{with probability } p_1,$$

or

$$h_t = \alpha_{0,2} + \alpha_{1,2}X_{t-1}^2 + \beta_{1,2}h_{t-1}, \quad \text{with probability } p_2 = 1 - p_1.$$

Note that the MGARCH(K; q, 0) model is sometimes referred to as the MARCH(K; q) model. To avoid the trivial case, we assume that for each $1 \leq k \leq K$, not all of the $\alpha_{i,k}(1 \leq i \leq q)$ and $\beta_{j,k}$ $(1 \leq j \leq p)$ are zero. Moreover, it is assumed that for each $1 \leq i \leq q$, not all of the $\alpha_{i,k}(1 \leq k \leq K)$ are zero and that for each $1 \leq j \leq p$, not all of the $\beta_{j,k}(1 \leq k \leq K)$ are zero. If $p = 0$ and $\alpha_{i,k} = 0$ for $q_k + 1 \leq i \leq q$ and $1 \leq k \leq K$, then the MGARCH(K; q, p) model is reduced to the MAR-ARCH(K; $0, \ldots, 0$; $q_1, \ldots, q_k$) model of Wong and Li (2001). In other words, it is a mixture ARCH(K; $q_1, \ldots, q_k$) model. If $K = 1$, that is, $\xi_{i,t}$ and $\eta_{j,t}$ are constants, then the MGARCH(1; q, p) model is reduced to the GARCH(q, p) model.

Section 2 summarizes some useful results on stochastic difference equations and obtains a stationary solution of the MGARCH process. Section 3 contains the main results on stationarity and tail behaviour for three different MGARCH models, namely the MARCH(K; 1) model, the MGARCH(K; 1, 1) model, and the general MGARCH(K; q, p) model. Moments of the MGARCH(K; 1, 1) process including the autocorrelations for the squared process are derived in Section 4. Some simulated MGARCH processes and the autocorrelations of the squared processes are provided. Section 5 discusses the estimation of the MGARCH model via the EM algorithm and considers some real examples. In particular, the tail exponent of the fitted models are calculated and compared with those obtained by using the popular Hill estimator (Embrechts *et al.*, 1997, Ch. 6) and the GARCH model.

2. PRELIMINARY RESULTS ON STOCHASTIC DIFFERENCE EQUATIONS

In this section, we present some preliminary results on stochastic difference equations, which plays a key role in proving the probabilistic properties of model eqn (2) in the Section 3.

Define

$$\mathbf{Z}_t = (X_t^2, \ldots, X_{t-q+1}^2, h_t, \ldots, h_{t-p+1})',$$

$$\mathbf{M}_t = \begin{pmatrix} \xi_{1,t}\varepsilon_t^2 & \cdots & \xi_{q-1,t}\varepsilon_t^2 & \xi_{q,t}\varepsilon_t^2 & \eta_{1,t}\varepsilon_t^2 & \cdots & \eta_{p-1,t}\varepsilon_t^2 & \eta_{p,t}\varepsilon_t^2 \\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \xi_{1,t} & \cdots & \xi_{q-1,t} & \xi_{q,t} & \eta_{1,t} & \cdots & \eta_{p-1,t} & \eta_{p,t} \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$
$$= \begin{pmatrix} \mathbf{M}_t(1,1) & \mathbf{M}_t(1,2) \\ \mathbf{M}_t(2,1) & \mathbf{M}_t(2,2) \end{pmatrix},$$

$$\mathbf{Q}_t = (\xi_{0,t}\varepsilon_t^2, 0, \ldots, 0, \xi_{0,t}, 0, \ldots, 0)',$$

where $(\cdots)'$ denotes the transpose of matrix, and $\mathbf{M}_t(1, 1)$, $\mathbf{M}_t(1, 2)$, $\mathbf{M}_t(2, 1)$ and $\mathbf{M}_t(2, 2)$ are sub-matrices of order $q \times q$, $q \times p$, $p \times q$ and $p \times p$ respectively. Then, the MGARCH(q, p) model of eqn (2) can be rewritten as the following stochastic difference equation,

$$\mathbf{Z}_t = \mathbf{M}_t\mathbf{Z}_{t-1} + \mathbf{Q}_t, \quad t \in N. \tag{4}$$

By definition, $\{(\mathbf{M}_t, \mathbf{Q}_t), t \in N\}$ is a sequence of i.i.d. non-negative random matrices, and for each t, the vector $\mathbf{Z}_{t-1}$ is independent of $(\mathbf{M}_t, \mathbf{Q}_t)$.

In general, given an initial value $\mathbf{Z}_0$, the solution for the stochastic difference equation (4) is given by

$$\mathbf{Z}_t = \sum_{k=0}^{t-1} \mathbf{Q}_{t-k} \prod_{j=0}^{k-1} \mathbf{M}_{t-j} + \mathbf{Z}_0 \prod_{j=0}^{t-1} \mathbf{M}_{t-j}, \quad t \in N. \tag{5}$$

Under certain conditions on $\mathbf{M}_t$, $\mathbf{Q}_t$ and the following condition on the initial value $\mathbf{Z}_0$

$$\mathbf{Z}_0 \stackrel{\mathrm{d}}{=} \mathbf{R} \widehat{=} \sum_{k=1}^{\infty} \mathbf{Q}_k \prod_{j=1}^{k-1} \mathbf{M}_j,$$

where '$\stackrel{\mathrm{d}}{=}$' denotes the identity in distribution, the solution (5) is strictly stationary and $\mathbf{Z}_t \stackrel{\mathrm{d}}{=} \mathbf{R}$ for all $t \geq 1$.

In order to introduce the properties of stochastic difference equation (4), we need some notations. Throughout the paper, $\mathbf{x} = (x_1, \ldots, x_{p+q})$ stands for a generic row vector. Put $|\mathbf{x}| = \max\{|x_i|; 1 \leq i \leq p + q\}$ and let

$$S_+ = \{\mathbf{x} \in R^{p+q} : |\mathbf{x}| = 1, x_i \geq 0, 1 \leq i \leq p+q\}.$$

For a non-zero vector $\mathbf{x}$, define $\tilde{\mathbf{x}} = (\mathbf{x})^{\sim} = |\mathbf{x}|^{-1}\mathbf{x}$, and hence $|\tilde{\mathbf{x}}| = 1$. Let $\mathbf{m} = (m_{ij})_{1 \leq i,j \leq p+q}$ be a matrix of order $p + q$, and define

$$\|\mathbf{m}\| = \max_{|\mathbf{x}|=1} |\mathbf{xm}|.$$

For $\mathbf{m} > 0$, we use $\rho(\mathbf{m})$ to denote its largest positive eigenvalue, i.e. the so-called Frobenius eigenvalue. Here, the notation $\mathbf{m} \geq 0$ (>0) means that $m_{ij} \geq 0$ (>0) for $1 \leq i, j \leq d$.

Theorem 1 is a variation of the result of Kesten (1973) and is a combined and revised version of Theorem 2.1, Corollaries 2.2 and 2.3 of Zhang and Tong (2004).

THEOREM 1. *Suppose $\{\mathbf{Z}_t, t \in N\}$ is a random series defined by stochastic difference equation* (4). *Denote the common distribution of $\{\mathbf{M}_t\}$ by μ. Assume that*

$$P(\mathbf{M}_1 \geq 0) = 1, P(\mathbf{M}_1 \textit{ has a zero row}) = 0, E(\ln^+\|\mathbf{M}_1\|) < \infty, \tag{6}$$

and that the additive group generated by

$$\{\ln\, \rho(\pi) : \pi = \mathbf{m}_1 \cdots \mathbf{m}_n \textit{ for some } n \textit{ and } \mathbf{m}_i \in \mathrm{supp}(\mu), \pi > 0\}$$

is dense in R where supp(μ) *denotes the support of probability measure μ. Then, there exists a constant $\alpha < +\infty$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \ln\, \|\mathbf{M}_1 \cdots \mathbf{M}_n\| = \lim_{n\to\infty} \frac{1}{n} \ln\, |\mathbf{xM}_1 \cdots \mathbf{M}_n| = \alpha,$$

for all $\mathbf{x} \in S_+$. If $\alpha < 0$, then $\{\mathbf{Z}_t\}$ is strictly stationary. Assume, in addition, that there exists a $\kappa_0 > 0$ for which

$$E(|\mathbf{xM}_1|^{\kappa_0}) \geq 1, \quad \textit{for all } \mathbf{x} \in S_+, \tag{7}$$

and

$$E(\|\mathbf{M}_1\|^{\kappa_0} \ln{}^+\|\mathbf{M}_1\|) < \infty. \tag{8}$$

Then, there exists a $\kappa_1 \in (0, \kappa_0]$ and a continuous and strictly positive function $r(\cdot)$ on S_+ such that

$$r(\mathbf{x}) = E(|\mathbf{xM}_1|^{\kappa_1}) r((\mathbf{xM}_1)^{\sim}), \quad \mathbf{x} \in S_+,$$

and

$$t^{\kappa_1} \mathrm{P}(\max\{|\mathbf{xM}_1 \cdots \mathbf{M}_n|; n \in N\} > t) \to C_1 r(\mathbf{x}), \quad t \to \infty,$$

where $0 < C_1 < \infty$ is a constant. Moreover, assume that

$$P(\mathbf{Q}_1 = 0) < 1, \quad \mathrm{E}(|\mathbf{Q}_1|^{\kappa_1}) < \infty, \quad \textit{and} \quad P(\mathbf{Q}_1 \geq 0) = 1.$$

Then, for each $\mathbf{x} \in S_+$,

$$\lim_{t\to\infty} t^{\kappa_1} \mathrm{P}(\mathbf{xR} \geq t) = C_2 r(\mathbf{x}), \tag{9}$$

where $0 < C_2 < \infty$ *is a constant. Finally, if the above* κ_1 *exists, then the following equation holds for all* $\mathbf{x} \in S_+$

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathrm{E}(\|\mathbf{M}_1 \cdots \mathbf{M}_n\|^{\kappa_1}) = \lim_{n\to\infty} \frac{1}{n} \ln \mathrm{E}(|\mathbf{xM}_1 \cdots \mathbf{M}_n|^{\kappa_1}) = 0.$$

Remark. Theorem 1 was first proven by Kesten (1973) with a condition slightly stronger than condition (7), and lately modified by Zhang and Tong (2004). In both papers, the norm of $\mathbf{x}$ was chosen as $|\mathbf{x}| = \sqrt{x_1^2 + \cdots + x_{p+q}^2}$. It is also remarked in Bougerol and Picard (1992) that the norm of $\mathbf{x}$ in R^{p+q} can be arbitrary chosen. For our purpose, we choose the maximum norm in this paper. The results of this paper also hold for the Euclidean norm (see also, Kesten and Spitzer, 1984).

3. PROBABILISTIC PROPERTIES

In this section, we apply Theorem 1 to derive the conditions for stationarity and the tail behaviour of the MGARCH model (2). From now on, we may use ξ_i, η_j, ε and $\mathbf{M}$ to replace $\xi_{i,t}$, $\eta_{j,t}$, ε_t and $\mathbf{M}_t$, respectively, for notational convenience.

3.1. *MARCH*(*K;* 1)

The MARCH(K; 1) model illustrates the general idea and is considered first. For this model, that is, $q = 1$ and $p = 0$ in model (2), the stochastic difference equation (4) becomes

$$X_t^2 = \xi_{1,t}\varepsilon_t^2 X_{t-1}^2 + \xi_{0,t}\varepsilon_t^2. \tag{10}$$

By letting $Y_t = X_t^2$, $A_t = \xi_{1,t}\varepsilon_t^2$ and $B_t = \xi_{0,t}\varepsilon_t^2$, eqn (10) can be written as a univariate and non-negative stochastic difference equation of the form

$$Y_t = A_t + B_t Y_{t-1}, \quad t \in N, \quad Y_0 \geq 0, \tag{11}$$

where $\{(A_t, B_t), t \in N\}$ is a sequence of i.i.d. non-negative random pairs, and for each fixed t, Y_{t-1} is independent of (A_t, B_t). We also write, for convenience, $(A, B) = (A_t, B_t)$. Theorem 2 gives the conditions for stationarity and the tail behaviour for such a univariate stochastic difference equation. As in Embrechts *et al.* (1997), the stationarity of X_t^2 implies the stationarity of X_t. Theorem 2 summarizes some of the results in Embrechts *et al.* (1997, Ch. 8) with slight modifications.

Theorem 2. *Let* $\{Y_t\}$ *be the stochastic process defined in eqn* (11). *Assume that* $E(\ln^+ A) < \infty$ *and* $E(\ln B) < 0$. *Then,* $Y_t \xrightarrow{d} Y$ *for some random variable* Y ('$\xrightarrow{d}$' *denotes convergence in distribution*) *and* Y *satisfies the identity in law*

$$Y \stackrel{d}{=} \sum_{m=1}^{\infty} A_m \prod_{j=1}^{m-1} B_j \stackrel{d}{=} A + BY, \tag{12}$$

where Y *and* (A, B) *are independent. If we choose* $Y_0 \stackrel{d}{=} Y$ *in model* (11), *then the process* $\{Y_t\}$ *is strictly stationary.*

Moreover, assume that $E(A^p) < \infty$ *and* $E(B^p) < 1$ *for some* $p \in [1, \infty)$. *Then,* $E(Y_0^p) < \infty$ *and* $\{Y_t\}$ *converges to* Y *in the pth mean. Furthermore, the moments* $E(Y^m)$ *are uniquely determined by the equations*

$$E(Y^m) = \sum_{k=0}^{m} \binom{m}{k} E(B^k A^{m-k}) E(Y^k), \quad m = 1, \ldots, [p],$$

where $[p]$ *denotes the integer part of* p.

Finally, suppose that there is a constant $\kappa > 0$ *with*

$$E(A^\kappa) < \infty, \quad E(B^\kappa) = 1, \quad E(B^\kappa \ln^+ B) < \infty,$$

and that $A/(1 - B)$ *is non-degenerate and the conditional distribution of* $\ln B$ *given* $B \neq 0$ *is non-lattice. Then, we have, as* $x \to \infty$,

$$P(Y > x) \sim cx^{-\kappa}, \tag{13}$$

and

$$c = \frac{\mathrm{E}((A + BY)^\kappa - (BY)^\kappa)}{\kappa \mathrm{E}(B^\kappa \ln B)} \tag{14}$$

is a positive constant.

In view of eqns (9) and (13), the constants κ_1 in Theorem 1 and κ in Theorem 2 are usually called the tail exponents of $\mathbf{xR}$ and Y respectively.

Theorem 3. *Suppose that* $\{X_t\}$ *is a* MARCH(K; 1) *time series defined in eqn* (10). *Then, we have the following results:*

(1) *Suppose that*

$$\sum_{k=1}^{K} p_k \ln \alpha_{1,k} < \ln 2 + \gamma, \tag{15}$$

where $\gamma \approx 0.5772$ *is the Euler's constant. Then,* $\{X_t\}$ *is strictly stationary with initial value* X_0 *independent of* $\{\xi_{i,t}\}$ *and* $\{\varepsilon_t\}$, *and*

$$X_0 \stackrel{\mathrm{d}}{=} \delta \left(\sum_{m=1}^{\infty} \xi_{0,m} \varepsilon_m^2 \prod_{j=1}^{m-1} (\xi_{1,j} \varepsilon_j^2) \right)^{1/2}, \tag{16}$$

where δ *is a random variable also independent of* $\{\xi_{i,t}\}$ *and* $\{\varepsilon_t\}$ *satisfying* $P(\delta = 1) = \mathrm{P}(\delta = -1) = 1/2$.

(2) *For any integer* $m \geq 1$, *if*

$$\sum_{k=1}^{K} p_k \alpha_{i,k}^m < \frac{1}{(2m-1)}!! \quad \textit{where } (2m-1)!! = \frac{(2m)!}{(2^m m!)},$$

then $\{X_t\}$ *is weakly stationary of order* $2m$. *Moreover, the moments* $\mu_{2j} \widehat{=} E(X_t^{2j}), j = 1, \ldots, m$, *are uniquely determined by the equations*

$$\mu_{2j} = \sum_{k=1}^{K} p_k \alpha_{0,k}^{j-l} \alpha_{1,k}^l \sum_{l=0}^{j} \binom{j}{l} (2j-1)!! \mu_{2l}. \tag{17}$$

(3) *Suppose that* $\{X_t\}$ *is strictly stationary. Then, there exists a unique constant* $\kappa > 0$ *such that*

$$\sum_{k=1}^{K} p_k \alpha_{1,k}^{\kappa} = \frac{\sqrt{\pi}}{2^{\kappa} \Gamma(\kappa + \frac{1}{2})}.$$

Moreover, if ξ_0/ξ_1 *is non-degenerate, then for any* $t \in N$, *as* $x \to \infty$,

$$P(|X_t| > x) = c x^{-2\kappa}, \tag{18}$$

where

$$c = \frac{E((\xi_0 \varepsilon^2 + \xi_1 \varepsilon^2 X_0^2)^{\kappa} - (\xi_1 \varepsilon^2 X_0^2)^{\kappa})}{\kappa E(\xi_1^{\kappa} \varepsilon^{2\kappa} \ln(\xi_1 \varepsilon^2))}$$

is a positive constant.

Proof. It suffices to check those conditions in Theorem 2 with $Y_t = X_t^2$, $A_t = \xi_{0,t} \varepsilon_t^2$ and $B_t = \xi_{1,t} \varepsilon_t^2$.

For (1) of Theorem 3, $\mathrm{E}(\ln {}^{+}A) = E(\ln^{+}(\xi_0 \varepsilon^2)) < \infty$ is obvious. Moreover,

$$E(\ln B) = E(\ln \xi_1 + \ln \varepsilon^2) = \sum_{k=1}^{K} p_k \ln \alpha_{1,k} - \ln 2 - \gamma.$$

Therefore, condition (15) implies $E(\ln B) < 0$. Notice that $P(X_0 > 0) = P(X_0 < 0) = 1/2$, eqn (16) follows from (12) directly.

For (2) of Theorem 3, $E(A^m) < \infty$ for any integer $m \geq 1$ is obvious. Moreover,

$$E(B^m) = E(\xi_1^m) E(\varepsilon^{2m}) = \sum_{k=1}^{K} p_k \alpha_{i,k}^m < \frac{1}{(2m-1)!!}.$$

Hence, (2) follows from Theorem 2.

For (3) of Theorem 3, the existence of κ is similar to that in Embrechts *et al.* (1997). The limit (18) comes from the last part of Theorem 2 because

$$E(B^{\kappa}) = E(\xi_i^{\kappa})\mathrm{E}(\varepsilon^{2\kappa}) = \sum_{k=1}^{K} p_k \alpha_{1,k}^{\kappa} \frac{2^{\kappa}}{\sqrt{\pi}} \Gamma\left(\kappa + \frac{1}{2}\right).$$

The case $m = 1$ *in* (2) *of Theorem* 3 gives the condition for second-order stationarity. *Equation (17) gives formulae for the even moments of* X_t *if they exist. These results are extensions of those in Wong and Li* (2001). □

3.2. *MGARCH*(K; 1, 1)

For the MGARCH(K; 1, 1) model, we have

$$h_t = (\xi_{1,t}\varepsilon_{t-1}^2 + \eta_{1,t})h_{t-1} + \xi_{0,t}. \tag{19}$$

Like eqn (10), eqn (19) can be written as a univariate stochastic difference equation (11). As in eqn (10), in order to find the weak or strict stationarity properties of $\{X_t\}$, it suffices to find those of $\{h_t\}$. The same applies to the study of tail behaviour. Therefore, we can apply Theorem 2 to stochastic difference equation (19).

THEOREM 4. *Suppose* $\{X_t\}$ *is a* MGARCH(K; 1, 1) *time series defined in eqn* (19). *Then, we have the following results:*

(1) *If*

$$\sum_{k=1}^{K} p_k E(\ln(\alpha_{1,k}\varepsilon^2 + \beta_{1,k})) < 0, \tag{20}$$

then $\{X_t\}$ *is strictly stationary with the initial value* X_0 *independent of* $\{\xi_{i,t},\ i = 0, 1\}$, $\{\eta_{1,t}\}$ *and* $\{\varepsilon_t\}$, *and distributed as*

$$X_0 \overset{\mathrm{d}}{=} \varepsilon_0 \left(\sum_{m=1}^{\infty} \xi_{0,m} \prod_{j=1}^{m-1} (\xi_{1,j}\varepsilon_j^2 + \eta_{1,j}) \right)^{1/2}.$$

(2) *For any integer* $m \geq 1$, *if*

$$\sum_{k=1}^{K} p_k \sum_{j=0}^{m} \binom{m}{j} (2j-1)!!\alpha_{1,k}^{j}\beta_{1,k}^{m-j} < 1,$$

then $\{X_t\}$ *is weakly stationary of order* $2m$. *Moreover, denote* $\mu_j \,\widehat{=}\, E(h_t^j)$. *Then,* $E(X_t^{2j}) = (2j-1)!!\mu_j$, *and* μ_j *are uniquely determined by the equations*

$$\mu_j = \sum_{k=1}^{K} p_k \sum_{l=0}^{j} \binom{j}{l} \sum_{i=0}^{l} \binom{l}{i} \alpha_{0,k}^{j-l} \alpha_{1,k}^{i} \beta_{1,k}^{l-i} (2i-1)!!\mu_l, \quad j = 1, \ldots, m.$$

(3) *Suppose that* $\{X_t\}$ *is strictly stationary. Then, there exists a unique constant* $\kappa > 0$ *such that*

$$E((\xi_1 \varepsilon^2 + \eta_1)^{\kappa}) = \sum_{k=1}^{K} p_k E((\alpha_{1,k}\varepsilon^2 + \beta_{1,k})^{\kappa}) = 1. \tag{21}$$

Moreover, for any $t \in N$, *as* $x \to \infty$,

$$P(h_t > x) \sim cx^{-\kappa},$$
$$P(|X_t| > x) \sim cE(|\varepsilon|^{2\kappa})x^{-2\kappa},$$

where c is a positive constant in the form of eqn (14) *with* $A = \xi_0$, $B = \xi_1\varepsilon^2 + \eta_1$ *and*

$$Y \stackrel{\mathrm{d}}{=} \sum_{m=1}^{\infty} \xi_{0,m} \prod_{j=1}^{m-1} (\xi_{1,j}\varepsilon_j^2 + \eta_{1,j})$$

but independent of A and B.

Proof. The proof is similar to that of Theorem 3 except for the last equation, which is a consequence of the result of Breiman (1965). (See also Mikosch and Stărică, 2000.) □

Theorem 4 extends the stationarity and moment results for the MARCH case of Wong and Li (2001) to the MGARCH case. Equation (21) will be useful in determining the tail exponent of the MGARCH(1, 1) process. Note also that because $0 < p_k < 1$ in eqn (20), some components can be allowed to have $E(\ln(\alpha_{1,k}\varepsilon^2 + \beta_{1,k})) > 0$ but still eqn (20) is satisfied. In other words, some components can be allowed to be non-stationary but X_t is still strictly stationary overall.

3.3. *MGARCH*(*K*; *q, p*)

Finally, we consider the general MGARCH(K; q, p) model (2). Applying Theorem 1, we have the following results.

Theorem 5. *Suppose that* $\{X_t\}$ *is a* MGARCH(K; q, p) *time series defined in eqns* (2) *and* (3). *Then, there exists a constant* $\alpha < +\infty$ *such that*

$$\lim_{n\to\infty} \frac{1}{n} \ln\|\mathbf{M}_1 \cdots \mathbf{M}_n\| = \lim_{n\to\infty} \frac{1}{n} \ln|\mathbf{x}\mathbf{M}_1 \cdots \mathbf{M}_n| = \alpha, \quad a.s.,$$

for all $\mathbf{x} \in S_+$. *If* $\alpha < 0$, *then* $\{X_t\}$ *is strictly stationary.*

If $\{X_t\}$ *is strictly stationary and either one of the following conditions hold:* (1) $p = 0$, *or* (2) $p > 0$ *and* $P(\max\{\xi_i, \eta_j; 1 \le i \le q, 1 \le j \le p\} > 1) > 0$, *then there exist two positive constants* $\kappa > 0$ *and* $C > 0$ *such that for each* $t \in N$,

$$\lim_{x\to\infty} P(|X_t| \ge x) = Cx^{-2\kappa}. \tag{22}$$

Finally, the following equation holds for all $\mathbf{x} \in S_+$,

$$\lim_{n\to\infty} \frac{1}{n} \ln E(\|\mathbf{M}_1 \cdots \mathbf{M}_n\|^\kappa) = \lim_{n\to\infty} \frac{1}{n} \ln E(|\mathbf{x}\mathbf{M}_1 \cdots \mathbf{M}_n|^\kappa) = 0. \tag{23}$$

Remark. At first glance, one might doubt that condition (2) in Theorem 5 implies that at least one of the K components of the MGARCH(K; q, p) model must not be second-order stationary. However, if we allow that $E(\varepsilon^2) = \sigma^2 \neq 1$, then even for a second-order stationary component, i.e. a second-order stationary GARCH(q, p) model, it might happen that $\alpha_i > 1$ for some $1 \le i \le q$. Furthermore, since $\beta_j < 1$ for all $1 \le j \le p$ is necessary for a GARCH(q, p) model to be second-order stationary, the GARCH(0, p) model and therefore the MGARCH(K; 0, p) model are excluded from condition (2) (see also Mikosch and Stărică, 2000). However, this should not be seen as a restriction since the kurtosis and hence the tail of the GARCH process depends on the ARCH parameter α_j. In other words, a heavy-tailed model requires at least some $\alpha_j \neq 0$.

To prove Theorem 5, we check whether the conditions of Theorem 1 are satisfied. One can do this by means of the following lemmas. The first of these can be proved by direct computation.

Lemma 1. *The characteristic polynomial for* $\mathbf{M}$ *is*

$$\begin{aligned} g(\lambda) &= \det(\lambda I_{p+q} - \mathbf{M}) \\ &= \lambda^p(\lambda^q - \xi_1\varepsilon^2\lambda^{q-1} - \cdots - \xi_q\varepsilon^2) - \lambda^q\eta_p(\lambda^{p-1} + \lambda^{p-2} + \cdots + 1), \end{aligned} \tag{24}$$

where $\det(\cdot)$ *denotes the determinant of a matrix and* I_{p+q} *is the identity matrix of order* $p + q$. *Moreover, the equation* $g(\lambda) = 0$ *has at least one positive solution almost surely, and therefore, the Frobenius eigenvalue of* $\mathbf{M}$, $\rho(\mathbf{M})$, *exists almost surely.*

Lemma 2. *For any constant* $r \ge 0$,

$$E(\|\mathbf{M}\|^r \ln^+ \|\mathbf{M}\|) < \infty. \tag{25}$$

Proof. By the definition of $\|\mathbf{M}\|$ and noticing that the MARCH(K; 1) and MGARCH(K; 1, 1) models are excluded from Theorem 5, we have $p + q \ge 2$, and

$$\|\mathbf{M}\| \leq \sum_{i,j=1}^{p+q} |m_{ij}| = [p+q-2+(\xi_1+\cdots+\xi_q+\eta_1+\cdots+\eta_p)(1+\varepsilon^2)].$$

From the definitions of ξ_i and η_j, there exists a constant c such that $\xi_1 + \cdots + \xi_q + \eta_1 + \cdots + \eta_p \leq c$. Since $\varepsilon \sim N(0, 1)$, simple calculations yield inequality (25). □

Lemma 3. *Assume that either condition* (1) *or* (2) *of Theorem* 5 *holds. Then, there exists a positive constant* $\kappa_0 > 0$ *such that*

$$E|\mathbf{xM}|^{\kappa_0} \geq 1, \quad \textit{for all } \mathbf{x} \in S_+. \tag{26}$$

Proof. We prove this lemma only for the case that condition (2) of Theorem 5 holds. If condition (1) holds, the proof is similar. Denote $\zeta = \max\{\xi_i, \eta_j; 1 \leq i \leq q, 1 \leq j \leq p\}$. For any $\mathbf{x} = (x_1, \ldots, x_q, x_{q+1}, \ldots, x_{q+p}) \in S_+$, if $x_{q+1} = 1$, then $|\mathbf{xM}| \geq \zeta$. By the assumption that $P(|\mathbf{xM}| > 1) > 0$, inequality (26) holds for some positive constant κ_1. If $x_1 = 1$, then $|\mathbf{xM}| \geq \zeta\varepsilon^2$. From the definition of ε, inequality (26) holds for some positive constant κ_2. If $x_i = 1$ for some $2 \leq i \leq q$ or $q + 2 \leq i \leq q + p$, then $|\mathbf{xM}| \geq 1$. Hence, eqn (26) holds for any positive constant κ_0. Thus, under condition (2), (26) holds for some positive constant κ_0 and any $\mathbf{x} \in S_+$. □

Remark. We see from the proof that even if ε is not normal, the unit-variance assumption, $\text{var}(\varepsilon) = 1$, implies that $P(\varepsilon^2 \geq 1) > 0$, $P(\zeta\varepsilon^2 > 1) > 0$, and κ_2 still exists. Furthermore, if we allow $\text{var}(\varepsilon) = \sigma^2 \neq 1$, an additional condition that $P(\zeta\varepsilon^2 > 1) > 0$ should be added to both conditions (1) and (2) of Theorem 5.

Proof of Theorem 5. To prove the theorem, we only need to check those conditions in Theorem 1.

The existence of the Frobenius eigenvalue of $\mathbf{M}$ is given by Lemma 1. Since $\rho(\mathbf{M})$ is a continuous function of parameters of the polynomial $g(\lambda)$ and $\varepsilon \sim N(0, 1)$, all possible values of $\ln \rho(\mathbf{M})$ must contain at least an interval in R. Therefore, the density of the associated additive group in Theorem 1 is ensured, and hence the equality for the Lyapunov exponent α holds. Moreover, condition (7) follows from Lemma 3; conditions (6) and (8) are satisfied by Lemma 2; and all other conditions in Theorem 1 are satisfied in a trivial manner. Therefore, tail probability (22) holds with C equal to $C_2 r(1, 0, \ldots, 0)$ where C_2 and r are given in Theorem 1. □

4. MOMENTS FOR MGARCH(K; 1, 1)

As indicated in many studies on GARCH models and their applications, the most popular GARCH model is the GARCH(1, 1) model. As a result, one may suspect that the most useful MGARCH model might be the MGARCH(K; 1, 1) model.

Furthermore, the moments of a time-series model, especially the autocorrelations, play an important role in both theoretical and practical works. Therefore, we study the moments of the MGARCH(K; 1, 1) process in detail in this section.

Here, $\{X_t\}$ is assumed to be a MGARCH(K; 1, 1) time series given by

$$\begin{aligned} X_t &= \varepsilon_t \sqrt{h_t}, \\ h_t &= \xi_{0,t} + \xi_{1,t} X_{t-1}^2 + \eta_{1,t} h_{t-1}, \quad t \in N, \end{aligned} \tag{27}$$

where $\xi_{0,t}$, $\xi_{1,t}$ and $\eta_{1,t}$ are defined in Section 1. Define $\theta = (1, \xi_0, \xi_1, \eta_1)'$ and denote the first- and second-order moments of θ by

$$E(\theta\theta') = (m_{ij})_{0 \le i,j \le 3}, \tag{28}$$

that is, $E(\xi_0) = m_{01}$, $E(\xi_1^2) = m_{22}$, $E(\xi_0\xi_1) = m_{12}$, and so on.

Obviously, $E(X_t) = 0$ whenever the first-order moment exists. Moreover, we have the following results on the second-order moments.

Proposition 1. *If $E(\xi_1\varepsilon^2 + \eta_1) = m_{02} + m_{03} < 1$, then $\{X_t\}$ is second-order stationary. Moreover*

$$E(X_t^2) = \mathrm{E}(h_t) = \frac{m_{01}}{1 - m_{02} - m_{03}}, \quad \textit{and} \quad E(X_t X_{t-l}) = 0, \ \ l > 0. \tag{29}$$

More explicity, the right-hand side of eqn (29) can be written as $\sum_{k=1}^K p_k\alpha_{0k}/(1 - \sum_{k=1}^K p_k(\alpha_{1k} + \beta_{1k}))$. As shown in Wong and Li (2001), the condition allows some of the components to have $\alpha_{1,k} + \beta_{1,k} > 1$. Hence, it is able to have a covariance-stationary process even though some of the components are not covariance-stationary themselves.

As the useful moments of the MGARCH(K; 1, 1) process might be the fourth-order moments including the autocorrelations of the squared series $\{X_t^2\}$, we give some results on these moments in proposition 2.

Proposition 2. *If $E(\xi_1\varepsilon^2 + \eta_1)^2 = 3m_{22} + 2m_{23} + m_{33} < 1$, then $\{X_t\}$ is fourth-order stationary. Denote the autocorrelations of the squared series $\{X_t^2\}$ by $\{\rho_l\}$. Then,*

$$E(X_t^4) = 3E(h_t^2) = \frac{3m_{11}(1 - m_{02} - m_{03}) + 6m_{01}(m_{12} + m_{13})}{(1 - m_{02} - m_{03})(1 - 3m_{22} - 2m_{23} - m_{33})}, \tag{30}$$

$$\rho_l = (m_{02} + m_{03})^{l-1}\rho_1, \quad l \ge 2. \tag{31}$$

Note that the condition for fourth-order stationarity reduces to $\sum_{k=1}^K p_k\alpha_{i,k}^2 < 1/3$ when $\beta_{1k} = 0$. This result was given in Wong and Li (2001).

Proof. The condition for fourth-order stationarity follows from Theorem 4. Suppose $\{X_t\}$ is fourth-order stationary. Then, $E(X_t^4) = E(\varepsilon_t^4 h_t^2) = 3E(h_t^2)$, and

$$\begin{aligned} E(X_t^2X_{t-l}^2) &= E(\varepsilon_t^2 h_t \varepsilon_{t-l}^2 h_{t-l}) = E(h_t \varepsilon_{t-l}^2 h_{t-l}) \\ &= E(\xi_{0,t}\varepsilon_{t-l}^2 h_{t-l} + (\xi_{1,t}\varepsilon_{t-1}^2 + \eta_{1,t})h_{t-1}\varepsilon_{t-l}^2 h_{t-l}) \\ &= \begin{cases} m_{01}\mathrm{E}(h_t) + (3m_{02} + m_{03})E(h_t^2), & l = 1, \\ m_{01}\mathrm{E}(h_t) + (m_{02} + m_{03})E(h_{t-1}\varepsilon_{t-l}^2 h_{t-l}), & l \geq 2. \end{cases} \end{aligned} \tag{32}$$

By definition and stationarity,

$$\begin{aligned} E(h_t^2) &= E(\xi_{0,t}^2 + 2\xi_{0,t}(\xi_{1,t}\varepsilon_{t-1}^2 + \eta_{1,t})h_{t-1} + (\xi_{1,t}\varepsilon_{t-1}^2 + \eta_{1,t})^2 h_{t-1}^2) \\ &= m_{11} + 2(m_{12} + m_{13})E(h_t) + (3m_{22} + 2m_{23} + m_{33})\mathrm{E}(h_t^2). \end{aligned}$$

Substituting eqn (29) into the above equation yields eqn (30).

Denote the autocovariances of the the squared process $\{X_t\}$ by $\gamma_l = E(X_t^2X_{t-l}^2) - (E(X_t^2))^2$ and $g_l = E(X_t^2X_{t-l}^2) = E(h_t\varepsilon_{t-l}^2 h_{t-l})$ for $l \geq 2$. Then,

$$\gamma_l = g_l - \left(\frac{m_{01}}{1 - m_{02} - m_{03}}\right)^2. \tag{33}$$

By eqn (32) and stationarity, we have

$$g_l = \frac{m_{01}^2}{1 - m_{02} - m_{03}} + (m_{02} + m_{03})g_{l-1}, \quad l \geq 2.$$

Plugging eqn (33) into this difference equation gives

$$\begin{aligned} \gamma_l &= (m_{02} + m_{03})\gamma_{l-1} \\ &= \gamma_{l-1}\sum_{k=1}^{K} p_k(\alpha_{1k} + \beta_{1k}), \end{aligned}$$

for $l \geq 2$, which implies eqn (31). □

Remark. By eqn (32), we can actually find the autocovariances $\{\gamma_l\}$ in terms of the m_{ij} values for all lag $l \geq 0$ and also the lag 1 autocorrelation ρ_1. In view of their complex forms, we do not include them in this paper. From eqn (31), we see that even though some of the components of the MGARCH(K; 1, 1) process are not second- or fourth-order stationary, for example, the IGARCH(1, 1) processes, the autocorrelations of the squared process can still have the short-memory property when they exist.

Finally, we are interested in comparing the MGARCH(K; 1, 1) process $\{X_t\}$ of eqn (27) with the following GARCH(1, 1) process $\{Y_t\}$,

$$\begin{aligned} Y_t &= \varepsilon_t\sqrt{v_t}, \\ v_t &= m_{01} + m_{02}Y_{t-1}^2 + m_{03}v_{t-1}, \quad t \in N, \end{aligned} \tag{34}$$

where the m_{ij} values are constants defined in eqn (28) and the innovation $\{\varepsilon_t\}$ is the same as that in model (27). We call $\{Y_t\}$ the GARCH(1, 1) process related to the MGARCH(K; 1, 1) process $\{X_t\}$. It is well known that the strict stationarity condition for $\{Y_t\}$ is $E(\ln(m_{02}\varepsilon^2 + m_{03})) < 0$, and that the second-order stationarity condition is $m_{02} + m_{03} < 1$. Note that the latter is the same as that for $\{X_t\}$. Furthermore, the second-order moments and the autocorrelations of $\{Y_t^2\}$ (if exist) also have the forms of eqns (29) and (31).

For the tail behaviour of $\{Y_t\}$, we have (see Mikosch and Stărică, 2000)

$$P(v_t > x) \sim c_0 x^{-\kappa_0},$$
$$P(|Y_t| > x) \sim c_0 E(|\varepsilon|^{2\kappa_0}) x^{-2\kappa_0},$$

where c_0 is a positive constant and has the form of eqn (14) with $A = m_{01}$, $B = m_{02}\varepsilon^2 + m_{03}$, and κ replaced by κ_0, the smallest positive constant satisfying

$$E(m_{02}\varepsilon^2 + m_{03})^{\kappa_0} = 1. \tag{35}$$

Comparing the tail behaviour of $\{X_t\}$ with that of $\{Y_t\}$, we obtain Theorem 6.

Theorem 6. *Suppose that* $m_{02} + m_{03} < 1$. *Then,*

$$\kappa_0 \geq \kappa > 1, \tag{36}$$

where κ *and* κ_0 *are the tail exponents defined in eqns* (21) *and* (35), *respectively. In other words, when both the MGARCH(K;* 1, 1) *process* $\{X_t\}$ *and the GARCH*(1, 1) *process* $\{Y_t\}$ *are second-order stationary, the former has a thicker tail.*

Proof. Define two functions $f(u)$ and $g(u)$ on $[0, \infty)$ as

$$f(u) = E(\xi_1\varepsilon^2 + \eta_1)^u, \quad g(u) = E(m_{02}\varepsilon^2 + m_{03})^u.$$

By definition and stationarity, it is not difficult to see that

$$f(0) = g(0) = 1, \quad f'(0) < 0, \quad g'(0) < 0, \quad f''(u) > 0, \quad g''(u) > 0.$$

These imply that both $f(u)$ and $g(u)$ are convex functions, starting from 1, decreasing at some neighbourhood of 0, and tending to infinity as $u \to \infty$. Therefore, $\kappa > 0$ and $\kappa_0 > 0$ exist, and are uniquely determined by $f(\kappa) = 1$ and $g(\kappa_0) = 1$ respectively.

In contrast, $m_{02} + m_{03} < 1$ implies that $f(1) < 1$. By the definition of κ, we have $\kappa > 1$. Then, using Jensen's inequality of conditional expectation, we get

$$\begin{aligned} 1 &= E(\xi_1\varepsilon^2 + \eta_1)^\kappa = E(E((\xi_1\varepsilon^2 + \eta_1)^\kappa|\varepsilon)) \\ &\geq E(E(\xi_1\varepsilon^2 + \eta_1|\varepsilon))^\kappa = E(m_{02}\varepsilon^2 + m_{03})^\kappa = g(\kappa). \end{aligned}$$

This together with the properties of $g(u)$ and κ_0 imply that $\kappa_0 \geq \kappa$. □

Remark. When $\{X_t\}$ is strictly stationary but not second-order stationary, that is, $m_{02} + m_{03} \geq 1$. Theorem 4 implies that there still exists a unique constant

$\kappa > 0$ with $f(\kappa) = 1$ but $\kappa \leq 1$ this time, and that inequality (36) does not hold any longer. Furthermore, similar results can be obtained for the MARCH(K;1) process and the related ARCH(1) process analogously. We expect that similar results also hold for the general MGARCH(K; q, p) process and the related GARCH(q, p) process. However, there seems to be no convexity for either $f(u) = \ln E(\|\mathbf{M}_1 \cdots \mathbf{M}_n\|^u)$ or $g(u) = \ln E(|\mathbf{x}\mathbf{M}_1 \cdots \mathbf{M}_n|^u)$ (see eqn (23)), and thus the above discussion fails in this case.

As an illustration, we simulate some MGARCH(K; 1, 1) processes of eqn (27) and their related GARCH(1, 1) processes of eqn (34). We choose the following three MGARCH(K; 1, 1) models.

Model 1. A fourth-order stationary MGARCH(2; 1, 1) process $\{X_t\}$ is generated using $p_1 = 0.75$, $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (1, 0.4, 0.2)$, $p_2 = 0.25$ and $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (1, 0, 1.2)$. Therefore, $m_{01} = 1$, $m_{02} = 0.3$ and $m_{03} = 0.45$ in the related GARCH(1, 1) process $\{Y_t\}$.

Model 2. A second-order but not fourth-order stationary MGARCH(2; 1, 1) process $\{X_t\}$ is generated using $p_1 = 0.6$, $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (1, 0.7, 0.2)$, $p_2 = 0.4$ and $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (1, 0.8, 0.3)$. Therefore, $m_{01} = 1$, $m_{02} = 0.74$ and $m_{03} = 0.24$.

Model 3. A second-order but not fourth-order stationary MGARCH(3; 1, 1) process $\{X_t\}$ is generated using $p_1 = 0.5$, $(\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}) = (1, 0.4, 0.2)$, $p_2 = 0.3$, $(\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}) = (1, 0.8, 0.3)$, $p_3 = 0.2$ and $(\alpha_{0,3}, \alpha_{1,3}, \beta_{1,3}) = (1, 0.2, 1.1)$. Therefore, $m_{01} = 1$, $m_{02} = 0.48$ and $m_{03} = 0.41$.

In all models, we simulate 1500 time units and choose the last 500 units as the simulated samples.

Figure 1 shows that the time plots of the simulated MGARCH(2; 1, 1) series $\{X_t\}$ for model 1. The simulated series appears to be stationary with volatility clustering around time $t = 400$. The sample autocorrelations of the squared process $\{X_t^2\}$ are shown in Figure 2.

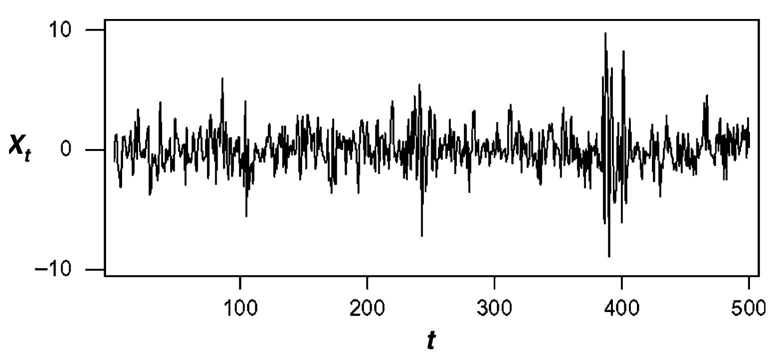

FIGURE 1. The simulated MGARCH(2; 1, 1) process $\{X_t\}$ for model 1.

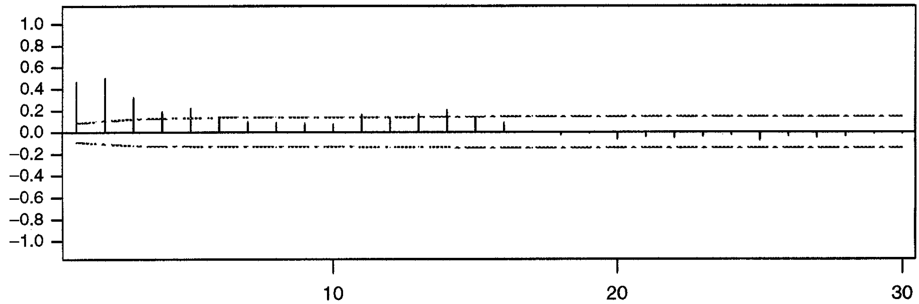

FIGURE 2. Sample auto-correlations of $\{X_t^2\}$ for Model 1.

TABLE I

THE TAIL EXPONENTS κ AND κ_0

	Model 1	Model 2	Model 3
κ	2.418	1.055	1.314
κ_0	2.868	1.060	1.529

Table I gives the tail exponents κ's and κ_0's for the three models. Generally, eqns (21) and (35) do not have closed-form solutions for κ and κ_0 respectively. However, by the properties of $f(u)$ and $g(u)$ in the proof of Theorem 6, it is not difficult to find the numerical solutions. Simulation results show that the process $\{X_t\}$ has a tail thicker than that of the process $\{Y_t\}$ for each of the three models.

Figure 3 shows the time plots of the conditional variances of six simulated processes. The plots on the left-hand side are the conditional variances of $\{X_t\}$ for models 1, 2 and 3, and the plots on the right-hand side are those of their related GARCH(1, 1) processes. These plots also show the volatility clustering properties for both the MGARCH processes and their related GARCH processes. Moreover, we see that for each model, the MGARCH process $\{X_t\}$ has conditional variances larger than those of its corresponding GARCH process $\{Y_t\}$, especially the extreme values of the conditional variances.

5. ESTIMATION RESULTS

The MGARCH model can be easily estimated using the EM algorithm given K. Suppose that the observation $\mathbf{X} = (x_1, \ldots, x_n)'$ satisfies model (2). Let $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be the unobserved random variable where $\mathbf{z}_t$ is a K-dimensional vector with component k equal to one if x_t comes from the kth component and to zero otherwise.

Let $\boldsymbol{\pi} = (p_1, \ldots, p_K)'$, $\boldsymbol{\theta}_k = (\alpha_{0,k}, \ldots, \alpha_{q,k}, \beta_{1,k}, \ldots, \beta_{p,k})'$ and $\boldsymbol{\theta} = (\boldsymbol{\pi}', \boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_K)'$. Following McLachlan and Krishnan (1997) and Wong and Li (2000), the conditional log-likelihood l given $\mathbf{z}$ can be written

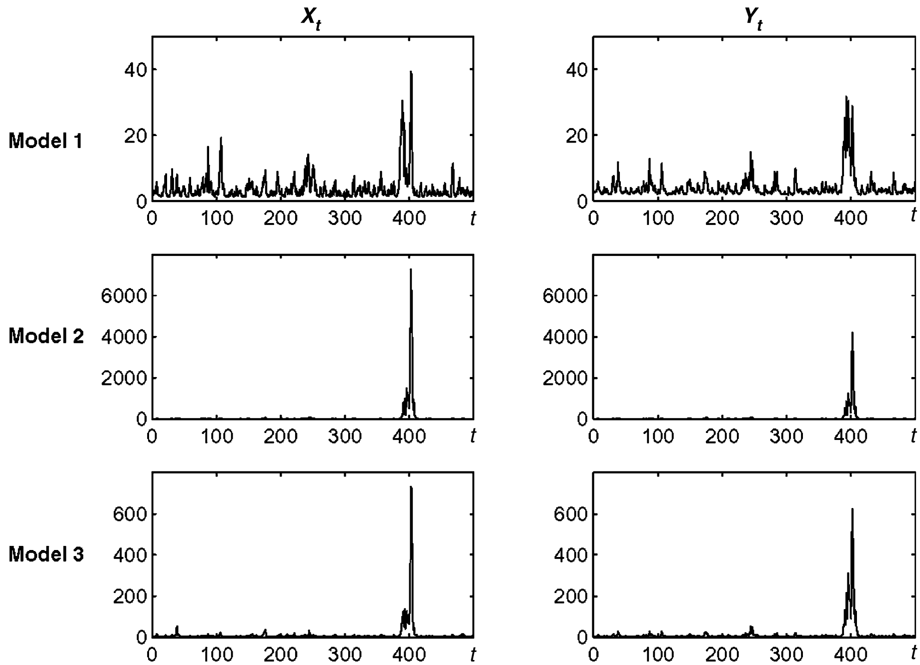

FIGURE 3. The conditional variances for MGARCH(K; 1, 1) processes $\{X_t\}$'s (left hand panel) and their related GARCH(1, 1) processes $\{Y_t\}$ (right hand panel). The horizontal axis is the time axis.

$$l = \sum_{t=\max(p,q)+1}^{n} l_t,$$

where

$$l_t = \sum_{k=1}^{K} z_{t,k} \log p_k - \frac{1}{2}\sum_{k=1}^{K} z_{t,k} \log h_{t,k} - \frac{\sum_{k=1}^{K} z_{t,k} x_t^2}{2h_{t,k}},$$

$z_{t,k}$ is the kth component of $\mathbf{z}_t$, and $h_{t,k}$ is the value of h_t when the GARCH parameters are from the kth component. The vector $\mathbf{Z}$ is not observed and is estimated by its conditional expectation given the parameters and $\mathbf{X}$. Denote by $\tau_{t,k}$ the conditional expectation of the kth component of $\mathbf{Z}_t$. Given values of $\boldsymbol{\theta}$ and $\mathbf{X}$, $\tau_{t,k}$ can be expressed as

$$\tau_{t,k} = \frac{p_k(1/\sqrt{h_{t,k}})\phi(x_t/\sqrt{h_{t,k}})}{\sum_{k=1}^{K} p_k(1/\sqrt{h_{t,k}})\phi(x_t/\sqrt{h_{t,k}})},$$

for $k = 1, \ldots, K$ where $\phi(\cdot)$ is the probability density function of the standard normal random variable. Given $\tau_{t,k}$, the conditional log-likelihood can then be maximized with respect to the model parameter $\boldsymbol{\theta}$ by a standard maximization routine. The $\tau_{t,k}$ can then be re-computed using this new estimate of $\boldsymbol{\theta}$ and the maximization step for $\boldsymbol{\theta}$ can be invoked again. The entire procedure can be re-iterated until the parameters have converged according to some criteria.

Three real data sets were considered as an illustration of the usefulness of the MGARCH model in practice. The three data sets are the daily closing NASDAQ index, 2 January 1997 to 31 December 2003; the daily NIKKEI index, 6 January 1997 to 31 December 2003; and the TAIWAN weighted index, 2 July 1997 to 31 December 2003. For simplicity, we consider the MGARCH(2;1, 1) model for the log returns of the series. Because a major motivation of using MGARCH models is to capture the heavy-tailed feature in the data, we also calculated the tail exponent of the fitted models using eqn (21). As a comparison, we fitted the corresponding GARCH(1, 1) model to the data and computed the corresponding tail exponent using (35). As a benchmark, we also computed the tail exponent of the log returns using the Hill estimator (Embrechts *et al.*, 1997).

The Hill estimator is a standard tool for estimating the tail exponent. It was originally derived for i.i.d. observations. Extension to the dependent case and, in particular, processes defined in eqn (11) has been established by Resnick and Stărică (1998). Specifically, if $X_1, \ldots, X_n$ is a realization of the time series, and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the corresponding order statistics, then the Hill estimator $\hat{\boldsymbol{\theta}}_k$ is given by

$$\hat{\boldsymbol{\theta}}_k = \frac{1}{l}\sum_{i=1}^{l} \log \frac{\mathbf{X}_{(i)}}{\mathbf{X}_{(l+1)}}, \tag{37}$$

where $l \to \infty$ in an appropriate rate.

The results are reported in Table II. The Hill estimate reported is the average of the Hill estimates for the left- and right-hand tails of the log returns. The GARCH(1, 1) models were estimated using the SAS software package. The value of l in eqn (37) is chosen by the method suggested in Tsay (2002, Ch. 7) and has a value of around 200. It can be seen that except for NASDAQ, the GARCH(1, 1) model gives tails that are too thin when compared with the Hill estimates. In contrast, the MGARCH(2; 1, 1) model seems capable of giving tails that are much closer to that of the Hill estimates. In the case of TAIWAN, the match is indeed very close. In the case of NASDAQ, the MGARCH(2; 1, 1) also gives a result that is slightly closer to that of the Hill estimate (0.215 versus 0.24) than the

TABLE II

SUMMARY OF TAIL EXPONENT ESTIMATES

Index	Dates	Length	Hill estimate	MGARCH	GARCH
NASDAQ	02/01/1997–31/12/2003	1761	1.5880	1.373	1.828
NIKKEI	06/01/1997–31/12/2003	1720	2.5905	1.639	6.037
TAIWAN	02/07/1997–31/12/2003	1591	2.0255	2.002	4.332

GARCH(1, 1) model. These results seem capable of sheding some light on the potential usefulness of MGARCH models in modelling heavy tails in financial time series.

6. CONCLUSION

The proposed MGARCH model provides an extension of the GARCH model and the MARCH model. One interesting feature of the mixture model is that some of its components can be non-stationary but yet the entire process is stationary. Furthermore, its tail behaviour is also more flexible and can be thicker than that of the GARCH model. The MGARCH model may be worth considering if accurate modelling of the tail is important in such applications like the estimation of value-at-risk.

ACKNOWLEDGEMENTS

W.K. Li thanks the Croucher Foundation for the award of a senior research fellowship in 2003–2004. Z. Zhang thanks the National Science Foundation of China Grant No. 10371042 for partial support. The authors would like to thank Mr S.T. Wong for his help in computer work. The authors thank a referee for suggestions that led to improvement of the paper.

REFERENCES

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–27.

Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–27.

Breiman, L. (1965) On some limit theorems similar to the arc sin law. *Theory of Probability and its Applications* 10, 323–31.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events*. Springer-Verlag: Berlin.

Engle, R. F. (1982) Autogressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987–1008.

Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* 131, 207–48.

Kesten, H. and Spitzer, F. (1984) Convergence in distribution of products of random matrices. *Zeitschrift für Wahrscheinlichkeitstheorieund verwandte Gebiete* 67, 363–86.

Le, N. D., Martin, R. D. and Raftery, A. Z. (1996) Modeling flat stretches, bursts and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association* 91, 1504–14.

McLachlan, G. and Peel, D. (2000) *Finite Mixture Models*. New York: Wiley.

McLachlan, G. and Krishnan, T. (1997) *The EM Algorithm and Extensions*. New York: Wiley.

Mikosch, T. and Stărică, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics* 28, 1427–51.

Resnick, S. and Stărică, C. (1998) Tail index estimation for dependent data. *Annals of Applied Probability* 8, 1156–83.

Tsay, R.S. (2002) *Analysis of Financial Time Series*. New York: Wiley.

Wong, C. S. and Li, W. K. (2000) On a mixture autoregressive model. *Journal of Royal Statistical Society* B 62, 95–115.

Wong, C. S. and Li, W. K. (2001) On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association* 96, 982–95.

Zhang, Z. Q. (2002) On tail behaviour and extremal values of some non-negative time series models. Ph.D. thesis, The University of Hong Kong.

Zhang, Z. Q. and Tong, H. (2004) A note on stochastic difference equations and its application to GARCH models. *Chinese Journal of Applied Probability and Statistics* 20, 259–69.