2007/97

■

Mixed exponential power asymmetric conditional heteroskedasticity

Mohammed Bouaddi and Jeroen V.K. Rombouts

CORE DISCUSSION PAPER
2007/97

Mixed exponential power asymmetric conditional heteroskedasticity

Mohammed BOUADDI [1] and Jeroen V.K. ROMBOUTS[2]

December 2007

Abstract

To match the stylized facts of high frequency financial time series precisely and parsimoniously, this paper presents a finite mixture of conditional exponential power distributions where each component exhibits asymmetric conditional heteroskedasticity. We provide stationarity conditions and unconditional moments to the fourth order. We apply this new class to Dow Jones index returns. We find that a two-component mixed exponential power distribution dominates mixed normal distributions with more components, and more parameters, both in-sample and out-of-sample. In contrast to mixed normal distributions, all the conditional variance processes become stationary. This happens because the mixed exponential power distribution allows for component-specific shape parameters so that it can better capture the tail behaviour. Therefore, the more general new class has attractive features over mixed normal distributions in our application: Less components are necessary and the conditional variances in the components are stationary processes. Results on NASDAQ index returns are similar.

Keywords: finite mixtures, exponential power distributions, conditional heteroskedasticity, asymmetry, heavy tails, value at risk.

JEL Classification: C11, C22, C52

---

[1] Institute of Applied Economics at HEC Montréal, CIRPEE, Canada. E-mail: mohammed.bouaddi@hec.ca

[2] Institute of Applied Economics at HEC Montréal, CIRANO, CIRPEE, Canada and CORE, Université catholique de Louvain, Belgium. Email: jeroen.rombouts@hec.ca

The authors thank Luc Bauwens and Vanessa Sumo for their comments. Mohammed Bouaddi acknowledges financial support from IFM2 of Montréal.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

Finite mixture models are becoming a standard tool in econometrics. They are attractive because of the flexibility they provide in model specification, which gives them a semiparametric flavour. Finite mixture textbooks are for example McLachlan and Peel (2000) and Frühwirth-Schnatter (2006). Early applications are Kon (1984) and Kim and Kon (1994) who investigate the statistical properties of stock returns using mixture models. Boothe and Glassman (1987), Tucker and Pond (1988) and Pan, Chan, and Fok (1995) use mixtures of normals to model exchange rates. Recent examples are Geweke and Keane (2005) in microeconometrics using panel data and Bauwens and Rombouts (2007a) and Frühwirth-Schnatter and Kaufmann (2008) for clustering purposes.

In this paper, we model the conditional distribution of time series of financial returns. Substantial research has been put into the refinement of the dynamic specification of the conditional variance equation, for which the benchmark is the linear GARCH specification of Bollerslev (1986). A survey on GARCH type models is given by Bollerslev, Engle, and Nelson (1994). The conditional distribution of the innovations is in most applicatons either normal, Student-t, skewed versions of these distributions, and the GED distribution. These extensions are often based on Azzalini (1985), Nelson (1991), Fernández and Steel (1998) and Jones and Feddy (2003). A stable GARCH process is considered in Mittnik, Paolella, and Rachev (2002). The GARCH type models fit the most important stylized facts of financial returns, which are volatility clustering and fat tails. However, for relatively long high frequency time series a typical result of the estimation of GARCH type models is that the conditional variance process is nearly integrated of order one. Diebold (1986) and Mikosch and Starica (2004) suggest that this is due to structural changes. To cope with this issue, finite mixtures of conditional distributions or, in our context, mixture GARCH models have been recently developed using normal distributions for the components. Building on the finite mixtures with autoregressive means and variances of Wong and Li (2000) and Wong and Li (2001), Haas, Mittnik, and Paolella (2004a) develop a mixture of normals coupled with the GARCH specification to capture, for example, conditional kurtosis and skewness as documented in Harvey and Siddique (1999), Harvey and Siddique (2000) and Brooks, Burke, Heravi, and Persand (2005). In an application to daily NASDAQ returns, they find that the best model contains three components, two of which are driven by nonstationary GARCH processes.

Other applications of mixture GARCH models are Alexander and Lazar (2005) and Haas, Mittnik, and Paolella (2006).

We propose a flexible mixture family based on exponential power distributions, also known as GED distributions, that nests the mixture of normals and that allows for leptokurtic as well as platikurtic components thanks to component specific shape parameters. The model is termed a mixed exponential power asymmetric conditional heteroskedasticity model (MEP-AGARCH) because the model is based on Engle and Ng (1993) to include the leverage effect in the component variances. There is an interesting tradeoff between the flexibility of the component distribution and the number of components. In our application to Dow Jones index returns, we find that a two-component MEP-AGARCH model dominates mixed normal distributions with more components (and more parameters) both in-sample and out-of-sample. In contrast to mixed normal distributions, all the conditional variance processes in the MEP-AGARCH model become stationary. While the former distribution needs nonstationary components to match the characteristics of the data, the latter can handle this also through its extra component specific shape parameters.

The rest of the paper is organized as follows. In section 2, we define the MEP-AGARCH model. Section 3 states the stationarity condition, the unconditional moments, and the autocorrelation function of the squared process. An application of the MEP-AGARCH model to Dow Jones index returns and a study of the accuracy and the relative performance of the model both in-sample and out-of-sample are provided in Section 4. Section 5 concludes. The Appendix contains the proof for proposition 1 of Section 3.

2 The model

We let y_t denote a univariate time series of interest and define $\varepsilon_t = y_t - E(y_t|\mathcal{F}_t)$, where $\mathcal{F}_t$ is the information set up to time $t-1$, and assume that the conditional mean does not depend on the components of the mixture. We say that ϵ_t follows a mixed exponential power asymmetric conditional heteroskedasticity model (MEP-AGARCH) if its conditional cdf is given by

$$F(\varepsilon_t \mid \mathcal{F}_t) = \sum_{n=1}^{N} \pi_n EP\left(\frac{\varepsilon_t - \mu_n}{\sqrt{h_{n,t}}}\right), \tag{1}$$

where

$$EP\left(x\right)=\frac{\lambda_n}{2\sqrt{2}\Gamma(\frac{1}{\lambda_n})}\int_{-\infty}^{x}\exp(-\left|\frac{z}{\sqrt{2}}\right|^{\lambda_n})dz. \tag{2}$$

The component mean μ_n is a real parameter, λ_n is a shape parameter defined on the positive line and π_n is the mixture weight for component n such that $0 \leqslant \pi_n \leqslant 1$ $\forall n = 1, ..., N$ and $\sum_{n=1}^{N}\pi_n = 1$, $\Gamma(\cdot)$ is the gamma function and

$$h_t = \sigma + \sum_{p=1}^{P}\psi_p(\iota\varepsilon_{t-p} - \delta_p) \odot (\iota\varepsilon_{t-p} - \delta_p) + \sum_{q=1}^{Q}\beta_q h_{t-q}, \tag{3}$$

where $h_t = (h_{1,t}, ..., h_{N,t})^T$, $\sigma = (\sigma_1, ..., \sigma_N)^T$, $\delta_p = (\delta_{1,p}, ..., \delta_{N,p})^T$, $\psi_p = diag(\alpha_p)$, $\alpha_p = (\alpha_{1,p}, ..., \alpha_{N,p})^T$, ι is a N-vector of ones, β_q are $N \times N$ matrices ($p = 1, ..., P$ and $q = 1, ..., Q$) and $\odot$ is the Hadamard product. The model is based on the Engle and Ng (1993) model to include the asymmetry effect on the component variances $h_{n,t}$. The effect of negative shocks on volatility is captured by $\delta_{n,p}$. When $\delta_{n,p}$ is positive, then negative shocks have a higher effect on the component volatility $h_{n,t}$ than positive shocks. Other models could be considered that allow for asymmetric news effects, for example, the GJR-GARCH model of Glosten, Jagannathan, and Runkle (1993) and the EGARCH model of Nelson (1991). Outside the mixture framework, the exponential power, or GED, distribution is used, for example, in financial econometrics by Nelson (1991), Liesenfeld and Jung (2000) and Hardouvelis and Theodossiou (2002). Komunjer (2007) presents an asymmetric extension of the exponential power distribution with applications to risk management.

To ensure that the volatility processes in the components are positive, we impose that $\sigma_n > 0$, $\alpha_p \geqslant 0$, and $\beta_q \geqslant 0$. As ϵ_t has zero mean we also have the restriction

$$\mu_N = -\sum_{n=1}^{N-1}\frac{\pi_n}{\pi_N}\mu_n. \tag{4}$$

Several special cases arise from the MEP-AGARCH model. The first one is the diagonal MEP-AGARCH model in which $\beta(L)$ is diagonal, implying that each component has an univariate AGARCH structure

$$h_{n,t} \quad = \quad \sigma_n + \sum_{p=1}^{P}\alpha_{n,p}(\varepsilon_{t-p} - \delta_{n,p})^2 + \sum_{q=1}^{Q}\beta_{nn,q}h_{n,t-q}. \tag{5}$$

We will use this diagonal model in the empirical illustration. The model becomes the mixed normal GARCH of Haas, Mittnik, and Paolella (2004a) when $\lambda_1 = ... = \lambda_N = 2$ and $\delta_{n,p} = 0$

($n = 1, ..., N$ and $p = 1, ..., P$). One can also consider having some components with constant variances, or with the same conditional variance apart from a constant as in Vlaar and Palm (1993).

Conditional moments of the data are combinations of the component moments. It can be shown that the K^{th} conditional centered moment of y_t is given by

$$E_{t-1}(\varepsilon_t^K) = \sum_{n=1}^{N} \frac{\pi_n \sum_{k=0}^{K} \binom{K}{k} \Gamma(\frac{k+1}{\lambda_n})(1+(-1)^k)(2h_{n,t})^{\frac{k}{2}} \mu_n^{K-k}}{2\Gamma(\frac{1}{\lambda_n})}. \quad (6)$$

For example, the conditional variance of y_t is

$$\begin{aligned} \sigma_t^2 &= E_{t-1}(\varepsilon_t^2) = \sum_{n=1}^{N} \pi_n \mu_n^2 + \sum_{n=1}^{N} \frac{2\pi_n \Gamma(\frac{3}{\lambda_n})}{\Gamma(\frac{1}{\lambda_n})} h_{n,t} \\ &= \pi^T \mu^{(2)} + \Delta^T h_t, \end{aligned} \quad (7)$$

the conditional third moment is

$$\begin{aligned} E_{t-1}(\varepsilon_t^3) &= \sum_{n=1}^{N} \pi_n \mu_n^3 + \sum_{n=1}^{N} \frac{6\pi_n \Gamma(\frac{3}{\lambda_n})}{\Gamma(\frac{1}{\lambda_n})} h_{n,t} \mu_n \\ &= \pi^T \mu^{(3)} + (\Upsilon \odot \mu^{(1)})^T h_t, \end{aligned} \quad (8)$$

and the conditional fourth moment is

$$\begin{aligned} E_{t-1}(\varepsilon_t^4) &= \sum_{n=1}^{N} \pi_n \mu_n^4 + \sum_{n=1}^{N} \frac{12\pi_n \Gamma(\frac{3}{\lambda_n})\mu_n^2}{\Gamma(\frac{1}{\lambda_n})} h_{n,t} + \sum_{n=1}^{N} \frac{4\pi_n \Gamma(\frac{5}{\lambda_n})}{\Gamma(\frac{1}{\lambda_n})} h_{n,t}^2 \\ &= \pi^T \mu^{(4)} + (\Xi \odot \mu^{(2)})^T h_t + trace(D \odot h_t h_t^T), \end{aligned} \quad (9)$$

where $\pi = (\pi_1, ..., \pi_N)$, $\Delta = \left(\frac{2\pi_1 \Gamma(\frac{3}{\lambda_1})}{\Gamma(\frac{1}{\lambda_1})}, ..., \frac{2\pi_N \Gamma(\frac{3}{\lambda_N})}{\Gamma(\frac{1}{\lambda_N})} \right)^T$, $\Upsilon = \left(\frac{3\pi_1 \Gamma(\frac{3}{\lambda_1})}{\Gamma(\frac{1}{\lambda_1})}, ..., \frac{3\pi_N \Gamma(\frac{3}{\lambda_N})}{\Gamma(\frac{1}{\lambda_N})} \right)^T$, $\Xi = \left(\frac{12\pi_1 \Gamma(\frac{3}{\lambda_1})}{\Gamma(\frac{1}{\lambda_1})}, ..., \frac{12\pi_N \Gamma(\frac{3}{\lambda_N})}{\Gamma(\frac{1}{\lambda_N})} \right)^T$, $\mu^{(k)} = (\mu_1^k, ..., \mu_N^k)$, $D = diag\left(\frac{4\pi_n \Gamma(\frac{5}{\lambda_n})}{\Gamma(\frac{1}{\lambda_n})} \right)$ is an $n \times n$ diagonal matrix and $trace(A)$ is the sum of the diagonal elements of the square matrix A.

3 Stationarity condition and unconditional moments

An interesting property is that the model allows for some variance components to be non stationary. However, the process remains globally stationary if the weights of the nonstationary components are sufficiently small, as shown in this section. For the theoretical properties it is convenient to write (3) as

$$(I_N - \beta(L))\, h_t = (\sigma + \sum_{p=1}^{P} \psi_p \delta_p^{(2)}) + \alpha(L)\varepsilon_t^2 - 2\,[\psi\delta]\,(L)\varepsilon_t, \quad (10)$$

where $\delta_p^{(2)} = (\delta_{1,p}^2, ..., \delta_{N,p}^2)^T$, $\alpha(L) = \sum_{p=1}^{P} \alpha_p L^p$, $[\psi\delta](L) = \sum_{p=1}^{P} (\alpha_p \odot \delta_p) L^p$, $\beta(L) = \sum_{q=1}^{Q} \beta_q L^q$ and L is the lag operator. If $E(h_t)$ exists, then by the law of iterated expectations and using (4) and (10) one can show that

$$E(h_t) = \left(I_N - \beta(1) - \alpha(1)\Delta^T\right)^{-1} \left(\sigma + \sum_{p=1}^{P} \psi_p \delta_p^{(2)} + \alpha(1)\mu^{(2)}\right), \tag{11}$$

and by (4) we get

$$\sigma^2 = E(\varepsilon_t^2) = \pi^T \mu^{(2)} + \Delta^T \left(I_N - \beta(1) - \alpha(1)\Delta^T\right)^{-1} \left(\sigma + \sum_{p=1}^{P} \psi_p \delta_p^{(2)} + \alpha(1)\pi^T \mu^{(2)}\right). \tag{12}$$

Therefore, the process is second-order stationary if and only if

$$\det\left(I_N - \beta(1) - \alpha(1)\Delta^T\right) > 0. \tag{13}$$

Proving this stationarity condition is similar to the proof in Haas, Mittnik, and Paolella (2004a). In the diagonal case, (12) reduces to

$$\begin{aligned}\sigma^2 \quad = \quad & \left(\sum_{n=1}^{N} \frac{\pi_n \left(1 - \sum_{q=1}^{Q} \beta_{n,q} - \frac{2\Gamma(\frac{3}{\lambda_n})}{\Gamma(\frac{1}{\lambda_n})} \sum_{p=1}^{P} \alpha_{n,p}\right)}{1 - \sum_{q=1}^{Q} \beta_{n,q}}\right)^{-1} \\ & \left(\sum_{n=1}^{N} \pi_n \mu_n^2 + \sum_{n=1}^{N} \pi_n \frac{2\Gamma(\frac{3}{\lambda_n})}{\Gamma(\frac{1}{\lambda_n})} \frac{\sigma_n + \sum_{p=1}^{P} \alpha_{n,p} \delta_{n,p}^2}{1 - \sum_{q=1}^{Q} \beta_{n,q}}\right),\end{aligned} \tag{14}$$

and second order stationarity is satisfied if and only if

$$\left(\sum_{n=1}^{N} \frac{\pi_n \left(1 - \sum_{q=1}^{Q} \beta_{n,q} - \frac{2\Gamma(\frac{3}{\lambda_n})}{\Gamma(\frac{1}{\lambda_n})} \sum_{p=1}^{P} \alpha_{n,p}\right)}{1 - \sum_{q=1}^{Q} \beta_{n,q}}\right) > 0. \tag{15}$$

The persistence of the volatility process can be measured by the largest eigenvalue of the matrix

$$M_{11} = \begin{pmatrix} \beta_1 + \alpha_1 \Delta^T & \beta_2 + \alpha_2 \Delta^T & \cdots & \beta_{N-1} + \alpha_{N-1}\Delta^T & \beta_N + \alpha_N \Delta^T \\ I_N & 0_N & \cdots & 0_N & 0_N \\ 0_N & I_N & \ddots & \vdots & 0_N \\ \vdots & \vdots & \ddots & 0_N & \vdots \\ 0_N & 0_N & \cdots & I_N & 0_N \end{pmatrix}. \tag{16}$$

We now concentrate on skewness, kurtosis and the autocorrelation function of the squared data.

Proposition 1 *If $E(h_t)$ and $E(h_t h_t^T)$ exist then the unconditional third moment is*

$$E(\varepsilon_t^3) = \pi^T \mu^{(3)} + (\Upsilon \odot \mu^{(1)})^T E(h_t). \tag{17}$$

The unconditional fourth moment is

$$\begin{aligned} E(\varepsilon_t^4) &= \pi^T \mu^{(4)} + (\Xi \odot \mu^{(2)})^T E(h_t) + trace(D \odot E(h_t h_t^T)) \\ &= \pi^T \mu^{(4)} + (\Xi \odot \mu^{(2)})^T E(h_t) + vec(D)^T E(vec(h_t h_t^T)), \end{aligned} \tag{18}$$

with

$$E(h_t) = (I - M_{11})^{-1} c_1, \tag{19}$$

$$E(vec(h_t h_t^T)) = (I - M_{22})^{-1} M_{21} (I - M_{22})^{-1} c_1 + (I - M_{22})^{-1} c_2, \tag{20}$$

and where

$$\begin{aligned} c_1 &= \sigma + \alpha \odot \delta \odot \delta + \alpha \pi^T \mu^{(2)}, \\ c_2 &= \sigma^* \otimes \sigma^* + (\alpha \otimes \sigma^* + \sigma^* \otimes \alpha + \Lambda \otimes \Lambda) \pi^T \mu^{(2)} \\ &\quad + (\Lambda \otimes \alpha + \alpha \otimes \Lambda) \pi^T \mu^{(3)} + (\alpha \otimes \alpha) \pi^T \mu^{(4)}, \\ \sigma^* &= \sigma + \alpha \odot \delta \odot \delta, \\ \Lambda &= -2\alpha \odot \delta, \end{aligned}$$

and

$$\begin{aligned} M_{11} &= \beta + \alpha \Delta^T \\ M_{21} &= (\alpha \Delta^T) \otimes \sigma^* + \sigma^* \otimes (\alpha \Delta^T) + (\Lambda \otimes (\Lambda \Delta^T)) \\ &\quad + (\Lambda \otimes \alpha)(\Upsilon \odot \mu^{(1)})^T + (\alpha \otimes \Lambda)(\Upsilon \odot \mu^{(1)})^T + (\beta \otimes \alpha + \alpha \otimes \beta) \pi^T \mu^{(2)} \\ &\quad + (\alpha \otimes \alpha)(\Xi \odot \mu^{(2)})^T + \beta \otimes \sigma^* + \sigma^* \otimes \beta, \\ M_{22} &= (\alpha \otimes \alpha) vec(D)^T + (\alpha \Delta^T) \otimes \beta + \beta \otimes (\alpha \Delta^T) + \beta \otimes \beta. \end{aligned}$$

The autocovariance function for the squared process is

$$\begin{aligned} \gamma(\tau) &= \gamma(-\tau) = E(\varepsilon_t^2 \varepsilon_{t-\tau}^2) - E^2(\varepsilon_t^2) = cov(\varepsilon_t^2, \varepsilon_{t-\tau}^2) \\ &= \Delta^T (\alpha \Delta^T + \beta)^{\tau - 1} \left\{ \sigma^* E(\varepsilon_t^2) + \alpha E(\varepsilon_t^4) - 2 (\alpha \odot \delta) E(\varepsilon_t^3) \right. \\ &\quad \left. + \beta \left(\pi^T \mu^{(2)} E(h_t) + E(h_t h_t^T) \Delta \right) - E(h_t) E(\varepsilon_t^2) \right\}. \end{aligned} \tag{21}$$

Proof: See the Appendix.

From the Appendix we also learn that the fourth unconditional moment exists when the largest eigenvalue of the following matrix is less than one:

$$M = \begin{pmatrix} M_{11} & 0_{N \times N^2} \\ M_{21} & M_{22} \end{pmatrix}.$$

In the application, we will compare the theoretical moments implied by the parameter estimates with the empirical moments.

4 Empirical results

4.1 Data

From Datastream we have daily Dow Jones index returns based on closing prices from January 3, 1950 to March 22, 2006, implying a sample of 14,231 observations. See Figure 1 for the sample path and Table 1 for some descriptive statistics.

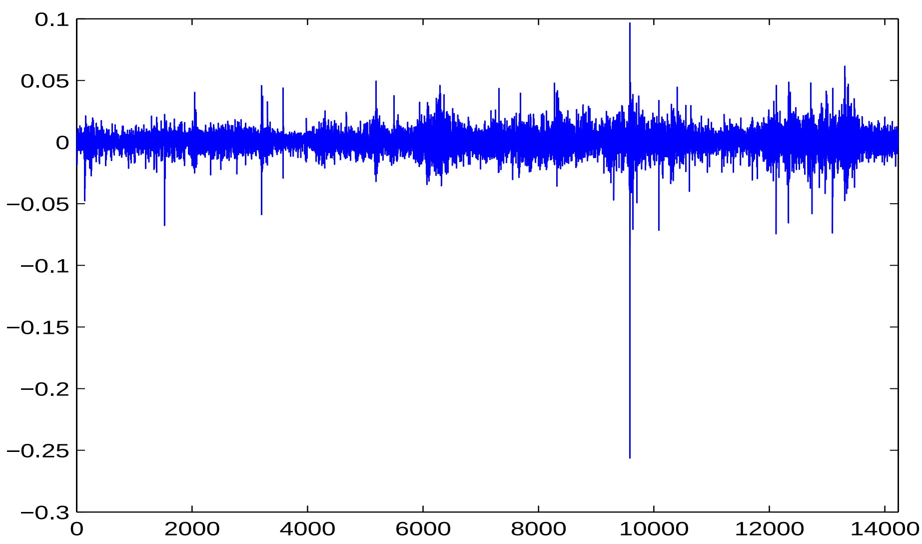

Figure 1: Dow Jones returns

Table 1: Descriptive statistics for Dow Jones index returns

Mean	0.000284	Maximum	0.0967
Standard deviation	0.009101	Minimum	-0.2563
Skewness	-1.67487	Kurtosis	52.63

Sample period: January 3, 1950 to March 22, 2006 (14,231 observations)

4.2 Model selection and in-sample fit

After fitting an ARMA(1,1) model for the conditional mean, we consider twenty-eight candidate models, with one to three components, to fit the Dow Jones returns. Fourteen models are estimated with a GARCH(1,1) specification for the component specific variance processes and another fourteen with asymmetric GARCH(1,1) specifications (AGARCH). The models that are termed MNs(i) and MN(i) are the symmetric and asymmetric mixed normal models with i components, where a symmetric mixture has $\mu_1 = \mu_2 = 0$. Similarly, MEPs(i;λ) and MEP(i;λ) are the symmetric and asymmetric mixed exponential power models with the same, but not fixed, shape parameter. Finally, MEPs(i;λ_i) and MEP(i;λ_i) represent those with different shape parameters. All the models in the application are estimated by maximum likelihood (ML) estimation. The loglikelihood function is given by

$$\sum_{t=1}^{T} \log \left(\sum_{n=1}^{N} \pi_n \frac{\lambda_n}{2\Gamma(\frac{1}{\lambda_n})\sqrt{2h_{n,t}}} \exp \left(- \left| \frac{\varepsilon_t - \mu_n}{\sqrt{2h_{n,t}}} \right|^{\lambda_n} \right) \right), \tag{22}$$

and is maximized under the constraint $\pi_1 \geqslant \pi_2 \geqslant ... > \pi_N$ to circumvent the label switching problem. Bayesian inference could also be done as explained in Bauwens and Rombouts (2007b). But given the large sample size and the fact that we estimate an important amount of models, we prefer ML estimation.

To determine the best in-sample fit among the models, we use the Bayesian information criterion (BIC), some goodness-of-fit tests on the normalized residuals, and compare empirical with implied theoretical moments according to the results in Section 3. Table 2 reports the goodness-of-fit results based on the BIC criterion for the models with the GARCH variance processes. The BIC selects the asymmetric three-component mixed-normal, i.e. MN(3), as the best model of all normal mixed models, which is a similar result to that obtained in Haas, Mittnik, and Paolella (2004a). Meanwhile, when each component of the mixture has its own shape parameter, the models of mixed exponential power with flexible shape behaviour outperform all the mixed normal models. The BIC selects the asymmetric mixed exponential power model with two components and different shape parameter for each component, i.e. MEP($2,\lambda_i$), as the best of all fourteen models. The last two columns of Table 2 give the

Table 2: In sample fit (models without asymmetry effect)

Model	n-par	Loglik	BIC	$\rho_{\max}(M_{11})$	$\rho_{\max}(M_{22})$
MN(1)	6	48722.71	-97388	0.9880	0.9874
MNs(2)	10	54029.11	-107963	0.9594	0.9222
MN(2)	11	54032.79	-107960	0.9600	0.9234
MNs(3)	14	54073.11	-108011	0.9617	0.9273
MN(3)	16	54082.41	**-108012**	0.9614	0.9269
MEP(1)	7	49038.37	-98010	0.9900	0.9939
MEPs($2;\lambda$)	11	54075.78	-108046	0.9906	0.9972
MEP($2;\lambda$)	12	54079.03	-108043	0.9907	0.9960
MEPs($2;\lambda_i$)	12	54077.71	-108041	0.9915	1.0061
MEP($2;\lambda_i$)	13	54086.27	**-108048**	0.9917	0.9997
MEPs($3;\lambda$)	15	54093.28	-108043	0.9960	0.9968
MEP($3;\lambda$)	17	54101.48	-108040	0.9956	0.9953
MEPs($3;\lambda_i$)	17	54098.57	-108035	0.9967	1.0003
MEP($3;\lambda_i$)	19	54107.05	-108032	0.9967	0.9991

In the second column, n-par denotes the number of the parameters in the model. The last two columns give the maximum eigenvalue of the matrix M_{11} and M_{22}.

values of $\rho_{\max}(M_{11})$ and $\rho_{\max}(M_{22})$ that are necessary to evaluate for the existence of the second and fourth moments. All models show that $\rho_{\max}(M_{11})$ is less than one in modulus suggesting that the return series is second-order stationary. Also, the results show that the unconditional fourth moment exists except in two out of the fourteen cases: MEPs($2;\lambda_i$) and MEPs($3;\lambda_i$) for which $\rho_{\max}(M_{22})$ is slightly higher than unity. We find the same conclusions in Table 3, which summarizes the models with AGARCH component variances. The best model is still the MEP($2,\lambda_i$). In addition, all the models now indicate the existence of fourth moments. Regarding the values of the BIC, the models with asymmetry effect dominate their counterparts in Table 2.

Table 3: In sample fit (models with asymmetry effect)

Model	n-par	Loglik	BIC	$\rho_{\max}(M_{11})$	$\rho_{\max}(M_{22})$
MN(1)	7	48796.33	-97526	0.9812	0.9723
MNs(2)	12	54118.54	-108122	0.9566	0.9165
MN(2)	13	54121.62	-108119	0.9566	0.9165
MNs(3)	17	54136.56	-108111	0.9599	0.9239
MN(3)	19	54159.89	**-108138**	0.9591	0.9224
MEP(1)	8	49100.47	-98124	0.9843	0.9812
MEPs(2;λ)	13	54149.57	-108175	0.9853	0.9796
MEP(2;λ)	14	54157.71	-108182	0.9858	0.9808
MEPs(2;λ_i)	14	54158.46	-108183	0.9854	0.9791
MEP(2;λ_i)	15	54166.89	**-108190**	0.9863	0.9821
MEPs(3;λ)	18	54160.93	-108150	0.9857	0.9791
MEP(3;λ)	20	54171.83	-108152	0.9898	0.9943
MEPs(3;λ_i)	20	54173.03	-108155	0.9874	0.9819
MEP(3;λ_i)	22	54192.21	-108174	0.9945	0.9897

In the second column, n-par denotes the number of parameters in the model. The last two columns give the maximum eigenvalue of the matrix M_{11} and M_{22}.

To test the distributional assumption, we use (1) to compute the residual $\hat{u}_t = F(\hat{\epsilon}_t \mid \mathcal{F}_t)$, which we transform, following Vlaar and Palm (1993), into $z_t = \Phi^{-1}(\hat{u}_t)$, where $\Phi^{-1}(.)$ is the quantile function of the normal distribution. Testing if z_t is normally distributed can be done using classical tests like the Cramer-von Mises, Anderson-Darling, Watson empirical distribution and Jarque-Bera tests. The results of these tests indicate that one-component models systematically reject normality (results not reported here). For the two-component models the normal mixture rejects and the exponential power mixtures do not reject. However, we do not reject normality using a three-component normal mixture. The LM test of heteroskedasticity indicates that there is no evidence of autocorrelation in the squares of the normalized residuals except in the case of one-component models that do not include the asymmetry effect.

We now focus on the implied theoretical moments according to the results in Section 3 for an informal comparison with the sample moments. Table 4 displays the empirical mean, variance, skewness and kurtosis together with the theoretical moments based on the ML estimates using the full sample for the most promising models with AGARCH component variances. We observe that the mean and variance are matched equally well for the models

Table 4: Sample versus implied moments

	Sample	MN(2)	MN(3)	MEPL(2;λ_i)
Mean	2.84E-04	2.92E-04	2.31E-04	2.92E-04
Variance	8.28E-05	1.04E-04	1.05E-04	1.04E-04
Skewness	-1.67477	-0.2683	-1.6305	-1.4086
Kurtosis	52.63699	10.483	31.3476	48.7634

under consideration. With respect to skewness, only the two-component MEP-AGARCH and the three-component normal GARCH model perform well. Only the two-component MEP-AGARCH is able the match the sample kurtosis.

4.3 Normal versus exponential power components

Using the whole sample period, Tables 5 and 6 report the model parameter estimates for the GARCH and AGARCH variance specifications, respectively (*** means significant at the 1 percent level, ** and * at 5 and 10 percent respectively). The parameter estimates for the symmetric mixtures are not reported since they underpeform (see the previous section).

For the mixed normal models, we observe in Table 5 that when the component mean μ_n decreases, the response of the component volatilities $h_{n,t}$ to the unexpected return ε_t increases (α_n increases strongly) and β_n decreases. Also, the variance components with the smallest μ_n are explosive ($\alpha_n + \beta_n > 1$) and have small mixing probabilities π_n. For the MEP models, the estimated shape parameters λ_n are significantly different from 2, hence the normality hypothesis is rejected for all the components. More precisely, for the two-component mixture MEP(2,λ_i), $\hat{\lambda}_1 = 1.65$ and $\hat{\lambda}_2 = 0.78$, meaning that both components have fat tails. In contrast to the normal mixture models, all the component-specific variance processes become now stationary ($\alpha_n + \beta_n < 1$). The component of the mixture with the negative mean and the lowest mixing probability still exhibits the highest reaction of its variance to shocks, though this reaction remains moderate (small α's) compared with the mixed normal models. The mixed exponential power models with the same shape parameter, MEP(i,λ), are not flexible enough to prevent this effect. Including the asymmetry effect in the variance components (δ_n), the results in Table 6 illustrate, moreover, that the effect of bad shocks relative to good shocks on the component volatilities is higher in the regime with the high mixing probability.

4.4 Out-of-sample performance

The out of sample performance is evaluated by one step ahead daily value at risk (VaR) forecasts obtained using parameter estimates estimated by a moving data window of 10,654 observations. Doing so, we obtain 3,576 (January 15, 1992 to March 22, 2006) VaR predictions at the 1, 2.5 and 5 percent levels. Among the mixture models, we only consider the best, which are the three-component mixed normal model and the two component mixed exponential power model with different shape parameters and including the asymmetry effect. The one component models are also included in the comparison.

We use three tests based on Christoffersen (1998), see also for example Kuester, Mittnik, and Paolella (2006). Let I_t^α be 1 when $y_t < VaR_t(\alpha)$ and 0 otherwise, where $VaR_t(\alpha)$ is the α-th quantile of the conditional distribution under study. We compute three tests using the estimated I_t^α's. The unconditional coverage test checks if the failure rate, defined by $F_\alpha = \sum_t \hat{I}_t^\alpha / 3576$, is equal to the pre-specified level α. Independence is tested in a Markovian framework, by verifying whether the first column in the transition probability matrix are equal. The conditional coverage test combines the two previous tests. The three tests are asymptotically Chi-squared distributed under the null hypothesis (one degree of freedom for the first two tests and two for the combined test). Table 7 presents failure rates and p-values of the VaR prediction tests for the three VaR levels. The failure rates show that both mixture models are equally close to the 5% and 2.5% target levels. At the 1% level, only the mixed exponential power model is accurate. These findings are also confirmed in the unconditional coverage tests. Also, as expected, both the normal and the exponential power AGARCH one component models systematically overestimate the failure rates. Except for the two mixture models at the 5% VaR level, the independence test does not reject. Based on

Table 5: Parameter estimates for the no asymmetry effect models

	MN(1)	MEP(1)	MN(2)	MN(3)	MEP(2;λ)	MEP(3;λ)	MEP(2;λ_i)	MEP(3;λ_i)
μ_1			$9.28E^{-05**}$ ($5.63E^{-05}$)	0.0004^{***} (0.0001)	$6.48E^{-05*}$ ($4.84E^{-05}$)	0.0007^{***} (0.0002)	0.0003^{***} ($4.52E^{-05}$)	0.0007^{***} (0.0002)
σ_1	$1.08E^{-06}$ ($6.05E^{-08}$)	$5.12E^{-07***}$ ($5.70E^{-08}$)	$2.53E^{-07***}$ ($3.50E^{-08}$)	$1.52E^{-07***}$ ($5.30E^{-08}$)	$4.28E^{-07***}$ ($6.35E^{-08}$)	$8.53E^{-08}$ ($1.50E^{-07}$)	$2.85E^{-07***}$ ($6.66E^{-08}$)	$5.13E^{-08}$ ($1.22E^{-07}$)
α_1	0.0751^{***} (0.0013)	0.0410^{***} ($5.70E^{-08}$)	0.0253^{***} (0.0015)	0.0191^{***} (0.0027)	0.0424^{***} (0.0029)	0.0683^{***} (0.0090)	0.0409 (0.0029)	0.0564^{***} (0.0069)
β_1	0.9129^{***} (0.0019)	0.9223^{***} (0.0034)	0.9336^{***} (0.0037)	0.9289^{***} (0.0083)	0.9338^{***} (0.0039)	0.9092^{***} (0.0093)	0.9375^{***} (0.0038)	0.9165^{***} (0.0082)
λ_1	2	1.4099^{***} (0.0117)	2	2	1.6263^{***} (0.0329)	1.6805^{***} (0.0426)	1.6469^{***} (0.0374)	1.5899^{***} (0.0633)
π_1	1	1	0.9691^{***} (0.0048)	0.5934^{***} (0.1124)	0.9924^{***} (0.0028)	0.6658^{***} (0.1072)	0.9527^{****} (0.0151)	0.6845^{***} (0.1653)
$\alpha_1+\beta_1$	0.9880	0.9633	0.9589	0.9480	0.9762	0.9776	0.9784	0.9729
μ_2			-0.0029^{***} (0.0012)	-0.0006^{*} (0.0004)	-0.0085^{**} (0.0045)	-0.0013^{***} (0.0005)	-0.0067 (0.0006)	-0.0010^{**} (0.0004)
σ_2			$1.31E^{-05**}$ ($5.96E^{-06}$)	$4.67E^{-07***}$ ($1.28E^{-07}$)	0.0001 ($8.90E^{-05}$)	$1.86E^{-07***}$ ($7.34E^{-08}$)	$1.31E^{-06}$ ($1.49E^{-06}$)	$1.79E^{-07**}$ ($7.75E^{-08}$)
α_2			0.3927^{***} (0.0700)	0.0426^{***} (0.0055)	2.0229^{**} (1.1171)	0.0073^{***} (0.0024)	0.0492 (0.0425)	0.0080^{***} 0.0026
β_2			0.7861^{***} (0.0645)	0.9344^{***} (0.0069)	0.5120^{*} (0.3347)	0.9862^{***} (0.0038)	0.6840^{***} (0.1416)	0.9900^{***} (0.0026)
λ_2			2	2	1.6263^{***} (0.0329)	1.6805^{***} (0.0426)	0.7774^{***} (0.1010)	2.4149^{***} (0.3806)
π_2			0.0309^{***} (0.0050)	0.4035^{***} (0.0700)	0.0076^{***} (0.0028)	0.3285^{***} (0.0644)	0.0473^{***} (0.0158)	0.2542^{***} (0.0729)
$\alpha_2+\beta_2$			1.1778	0.9770	2.5350	0.9934	0.7331	0.9980
μ_3				-0.0103^{*} (0.0073)		-0.0080 (0.0528)		-0.0033 (0.0034)
σ_3				0.0002 (0.0002)		0.0002 (0.0002)		$4.37E^{-07}$ ($6.30E^{-07}$)
α_3				2.6709 (2.2954)		2.8945^{*} (1.9256)		0.0150 (0.0149)
β_3				0.3391 (0.7061)		0.4007 (0.4843)		0.7568^{***} (0.1445)
λ_3						1.6805^{***} (0.0426)		0.6729^{***} (0.0905)
π_3				0.0032^{***} (0.0010)		0.0056^{***} (0.0023)		0.0613^{***} (0.0189)
$\alpha_3+\beta_3$				3.0101		3.2952		0.7718

Table 6: Parameter estimates for the asymmetry effect models

	MN(1)	MEP(1)	MN(2)	MN(3)	MEP(2;λ)	MEP(3;λ)	MEP(2;λ_i)	MEP(3;λ_i)
μ_1			$7.16E^{-05}$	0.0004^{***}	$7.81E^{-05}$	0.0004^{**}	0.0002^{***}	$3.86E^{-05}$
			$(7.68E^{-05})$	(0.0001)	$(9.69E^{-05})$	(0.0002)	$(7.36E^{-05})$	(0.0003)
σ_1	$6.49E^{-07***}$	$1.88E^{-07**}$	$1.68E^{-13}$	$1.17E^{-11}$	$7.25E^{-12}$	$5.21E^{-12}$	$1.17E^{-11}$	$9.93E^{-12}$
	$(7.38E^{-08})$	$(9.14E^{-08})$	$(3.41E^{-09})$	$(9.64E^{-08})$	$(9.81E^{-08})$	$(1.77E^{-08})$	$(9.91E^{-08})$	$(6.23E{-}08)$
α_1	0.0691^{***}	0.0400^{***}	0.0247^{***}	0.0190^{***}	0.0433^{***}	0.0503^{***}	0.0410^{***}	0.0574^{***}
	(0.0016)	(0.0023)	(0.0015)	(0.0029)	(0.0030)	(0.0074)	(0.0030)	(0.0075)
β_1	0.9121^{***}	0.9195^{***}	0.9314^{***}	0.9227^{***}	0.9309^{***}	0.9001^{***}	0.9358^{***}	0.8989^{***}
	(0.0004)	(0.0007)	(0.0037)	(0.0093)	(0.0040)	(0.0169)	(0.0040)	(0.0094)
δ_1	0.0035^{***}	0.0037^{***}	0.0040^{***}	0.0043^{***}	0.0039^{***}	0.0047^{***}	0.0035^{***}	0.0043^{***}
	(0.0002)	(0.0004)	(0.0004)	(0.0008)	(0.0004)	(0.0010)	(0.0004)	(0.000573)
λ_1	2	1.4255^{***}	2	2	1.6841^{***}	1.7845^{***}	1.6932^{***}	1.6304^{***}
		(0.0117)			(0.0363)	(0.0704)	(0.0392)	(0.069308)
π_1	1	1	0.9767^{***}	0.6065^{***}	0.9902^{***}	0.6152^{***}	0.9469^{***}	0.7331^{***}
			(0.0038)	(0.1568)	(0.0039)	(0.2113)	(0.0156)	(0.0576)
$\alpha_1 + \beta_1$	0.9812	0.9595	0.9561	0.9417	0.9742	0.9504	0.9768	0.9563
μ_2			-0.0030^{**}	-0.0004^{**}	-0.0079^{***}	-0.0004	-0.0043^{***}	0.0003
			(0.0016)	(0.0002)	(0.0045)	(0.0004)	(0.0007)	(0.0005)
σ_2			$2.13E^{-05}$	$1.21E^{-08}$	$6.02E^{-06}$	$2.25E^{-08}$	$8.42E^{-09}$	$2.54E^{-09}$
			$(1.79E^{-05})$	$(1.61E^{-07})$	$(7.66E^{-05})$	$(2.10E^{-07})$	$(1.79E^{-06})$	$(8.34E{-}08)$
α_2			0.4487^{***}	0.0414^{***}	0.5246	0.0454^{***}	0.0355	0.0111^{***}
			(0.1416)	(0.0059)	(0.3250)	(0.0105)	(0.0293)	(0.0032)
β_2			0.7069^{***}	0.9349^{***}	0.8187^{***}	0.9554^{***}	0.6339^{***}	0.9883^{***}
			(0.0912)	(0.0068)	(0.1310)	(0.0090)	(0.1060)	(0.0023)
δ_2			0.0054	0.0035^{***}	0.0085	0.0026^{**}	0.0089^{**}	0.0004
			(0.0036)	(0.0007)	(0.0113)	(0.0012)	(0.0039)	(0.0027)
λ_2			2	2	1.6841^{***}	1.7845^{***}	0.7773	2.2696^{***}
					(0.0363)	(0.0704)	(0.1046)	(0.3511)
π_2			0.0233^{***}	0.3903^{***}	0.0098^{***}	0.3805^{***}	0.0531^{***}	0.2535^{***}
			(0.0039)	(0.0866)	(0.0039)	(0.1552)	(0.0164)	(0.0722)
$\alpha_2 + \beta_2$			1.1556	0.9763	1.3432	1.0008	0.6694	0.9995
μ_3				-0.0182		-0.0153		-0.0087^{**}
				(0.0471)		(0.0624)		(0.0042)
σ_3				0.0001		0.0002		$8.63E^{-06}$
				(0.0002)		(0.0002)		$(4.29E{-}05)$
α_3				2.8615		4.3983^{*}		0.1983
				(2.6248)		(3.4017)		(0.4714)
β_3				0.3656		0.3690		0.3920
				(0.7738)		(0.5341)		(0.3840)
δ_3				-0.0023		-0.0021		0.0110
				(0.0034)		(0.0026)		(0.0108)
λ_3				2		1.7845^{***}		0.9415^{**}
						(0.0704)		(0.4481)
π_3				0.0032^{**}		0.0043^{**}		0.0134^{**}
				(0.0016)		(0.0018)		(0.0079)
$\alpha_3 + \beta_3$				3.2272		4.7673		0.5903

these results, we conclude that the two componenent exponential power AGARCH mixture performs best in this out of sample performance exercise.

Table 7: Failure rates and p-values for VaR tests

	MN(1)	MEP(1)	MEP(2;λ_i)	MN(3)
	$\alpha = 1\%$			
Failure rate	0.0453	0.0224	0.0108	0.0185
Unconditional Coverage	0.0000	0.0000	0.6384	0.0000
Independence	0.7762	0.8683	0.4330	0.5078
Conditional Coverage	0.0000	0.0000	0.6585	0.0000
	$\alpha = 2.5\%$			
Failure rate	0.0763	0.0475	0.0277	0.0280
Unconditional Coverage	0.0000	0.0000	0.3054	0.2559
Independence	0.5372	0.5690	0.0423	0.1327
Conditional Coverage	0.0000	0.0000	0.0753	0.1694
	$\alpha = 5\%$			
Failure rate	0.1202	0.0886	0.0459	0.0445
Unconditional Coverage	0.0000	0.0000	0.2498	0.1218
Independence	0.5665	0.3972	0.0002	0.0001
Conditional Coverage	0.0000	0.0000	0.0006	0.0001

4.5 NASDAQ returns

To compare with Haas, Mittnik, and Paolella (2004a), we repeat the same exercise as above, results not reported here, to daily NASDAQ returns from February 1971 to June 2001 (7,681 observations). From the estimates of the three-components mixed normal and the two component mixed exponential power models we find the same conclusions as in our application to Dow Jones returns: The three-component mixed-normal has two explosive component variances, while all the variance components of the preferred two-component mixed exponential power model are stationary.

5 Conclusion

In this paper, we develop a finite mixture of conditional exponential power distributions where each component exhibits asymmetric conditional heteroskedasticity. We provide weak stationarity conditions and unconditional moments to the fourth order for this mixture. The mixture is more flexible than a normal mixture because the components have shape-specific parameters. Thanks to the extra shape parameters, an exponential power mixture with two components is found to be flexible enough to accommodate financial time series characteristics as in our application to Dow Jones and NASDAQ daily return series. Another attractive feature of the mixed exponential power mixture that we find in the application is that, in contrast to mixed normal distributions, all the conditional variance processes become stationary. One extension of this paper is to allow for dependent states in the mixture distribution as Haas, Mittnik, and Paolella (2004b). Another extension is the generalization to the multivariate case, as Bauwens, Hafner, and Rombouts (2007) did for the univariate normal GARCH mixture.

Appendix: Proof of Proposition 1

The proof follows the same idea as in Haas, Mittnik, and Paolella (2004a). From (3) we obtain the diagonal MEP-AGARCH(1,1)

$$h_t = \sigma^* + \alpha\varepsilon_{t-1}^2 + \Lambda\varepsilon_{t-1} + \beta h_{t-1}, \tag{23}$$

where $\sigma^* = \sigma + \alpha \odot \delta \odot \delta$, $\Lambda = -2\alpha \odot \delta$, $P = Q = 1$ and β ($\beta_1 = \beta$) is a diagonal matrix. It follows that

$$\begin{aligned} h_t h_t^T \quad = \quad & \sigma^*\sigma^{*T} + \sigma^*\alpha^T\varepsilon_{t-1}^2 + \sigma^*\Lambda^T\varepsilon_{t-1} + \sigma^* h_{t-1}^T\beta + \alpha\sigma^{*T}\varepsilon_{t-1}^2 + \alpha\alpha^T\varepsilon_{t-1}^4 \\ & +\alpha\Lambda^T\varepsilon_{t-1}^3 + \alpha h_{t-1}^T\varepsilon_{t-1}^2\beta + \Lambda\sigma^{*T}\varepsilon_{t-1} + \Lambda\alpha^T\varepsilon_{t-1}^3 + \Lambda\Lambda^T\varepsilon_{t-1}^2 \\ & +\Lambda h_{t-1}^T\varepsilon_{t-1}\beta + \beta h_{t-1}\sigma^{*T} + \beta h_{t-1}\varepsilon_{t-1}^2\alpha^T + \beta h_{t-1}\varepsilon_{t-1}\Lambda^T \\ & +\beta h_{t-1}h_{t-1}^T\beta. \end{aligned} \tag{24}$$

We note that $W_t = vec(h_t, h_t h_t^T) = \left(h_t^T, vec(h_t h_t^T)^T\right)^T$, and using (7) to (9) we get [1],

$$vec(\sigma^* \sigma^{*T}) = \sigma^* \otimes \sigma^*,$$

$$E_{t-2}(vec(\sigma^* \alpha^T \varepsilon_{t-1}^2)) = (\alpha \otimes \sigma^*)\, \pi^T \mu^{(2)} + \left((\alpha \Delta^T) \otimes \sigma^*\right) h_{t-1},$$

$$E_{t-2}\left(vec(\sigma^* \Lambda^T \varepsilon_{t-1})\right) = (\Lambda \otimes \sigma^*)\, E_{t-2}(\varepsilon_{t-1}) = 0,$$

$$E_{t-2}\left(vec(\sigma^* h_{t-1}^T \beta)\right) = (\beta \otimes \sigma^*)\, h_{t-1},$$

$$E_{t-2}(vec(\alpha \varepsilon_{t-1}^2 \sigma^{*T})) = (\sigma^* \otimes \alpha)\, \pi^T \mu^{(2)} + \left(\sigma^* \otimes (\alpha \Delta^T)\right) h_{t-1},$$

$$\begin{aligned} E_{t-2}(vec(\alpha \alpha^T \varepsilon_{t-1}^4)) &= (\alpha \otimes \alpha)\, \pi^T \mu^{(4)} + (\alpha \otimes \alpha)\, (\Xi \odot \mu^{(2)})^T h_{t-1} \\ &\quad + (\alpha \otimes \alpha)\, vec(D)^T vec(h_{t-1} h_{t-1}^T), \end{aligned}$$

$$E_{t-2}(vec(\alpha \Lambda^T \varepsilon_{t-1}^3)) = (\Lambda \otimes \alpha)\, \pi^T \mu^{(3)} + \Big(\Lambda \otimes (\alpha (\Upsilon \odot \mu^{(1)})^T)\Big) h_{t-1},$$

$$E_{t-2}(vec(\alpha h_{t-1}^T \varepsilon_{t-1}^2 \beta)) = (\beta \otimes \alpha)\, \pi^T \mu^{(2)} h_{t-1} + \left(\beta \otimes \alpha \Delta^T\right) vec(h_{t-1} h_{t-1}^T),$$

$$E_{t-2}\left(vec(\Lambda \sigma^{*T} \varepsilon_{t-1})\right) = (\sigma^* \otimes \Lambda)\, E_{t-2}(\varepsilon_{t-1}) = 0,$$

$$E_{t-2}\left(vec(\Lambda \alpha^T \varepsilon_{t-1}^3)\right) = (\alpha \otimes \Lambda)\, \pi^T \mu^{(3)} + \Big((\alpha (\Upsilon \odot \mu^{(1)})^T) \otimes \Lambda\Big) h_{t-1},$$

$$E_{t-2}\left(vec(\Lambda \Lambda^T \varepsilon_{t-1}^2)\right) = (\Lambda \otimes \Lambda)\, \pi^T \mu^{(2)} + \left(\Lambda \otimes (\Lambda \Delta^T)\right) h_{t-1},$$

$$E_{t-2}\left(vec(\Lambda h_{t-1}^T \varepsilon_{t-1} \beta)\right) = (\beta \otimes \Lambda)\, h_{t-1} E_{t-2}(\varepsilon_{t-1}) = 0,$$

$$E_{t-2}(vec(\beta h_{t-1} \sigma^{*T})) = (\sigma^* \otimes \beta)\, h_{t-1},$$

$$E_{t-2}(vec(\beta h_{t-1} \varepsilon_{t-1}^2 \alpha^T)) = (\alpha \otimes \beta)\, \pi^T \mu^{(2)} h_{t-1} + \left((\alpha \Delta^T) \otimes \beta\right) vec(h_{t-1} h_{t-1}^T),$$

$$E_{t-2}\left(vec(\beta h_{t-1} \varepsilon_{t-1} \Lambda^T)\right) = (\Lambda \otimes \beta)\, h_{t-1} E_{t-2}(\varepsilon_{t-1}) = 0$$

---

[1] We use the properties of vec operator: $vec(xy^T) = y \otimes x$ and $vec(ABC) = (C^T \otimes A) vec(B)$, where x and y are vectors with the same order and A, B and C are matrices with appropriate dimensions. $vec(A)$ is the operator that stacks the columns of the matrix A.

and

$$E_{t-2}(vec(\beta h_{t-1} h_{t-1}^T \beta)) = (\beta \otimes \beta)\, vec(h_{t-1} h_{t-1}^T).$$

Then it follows that

$$E_{t-2}(W_t) = c + M W_{t-1}, \tag{25}$$

where

$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

$$c_1 = \sigma^* + \alpha \pi^T \mu^{(2)},$$

$$\begin{aligned} c_2 \quad &= \quad \sigma^* \otimes \sigma^* + (\alpha \otimes \sigma^* + \sigma^* \otimes \alpha + \Lambda \otimes \Lambda) \pi^T \mu^{(2)} \\ &\quad + (\Lambda \otimes \alpha + \alpha \otimes \Lambda)\, \pi^T \mu^{(3)} + (\alpha \otimes \alpha) \pi^T \mu^{(4)}, \end{aligned}$$

and

$$M = \begin{pmatrix} M_{11} & 0_{N \times N^2} \\ M_{21} & M_{22} \end{pmatrix},$$

where

$$M_{11} = \beta + \alpha \Delta^T,$$

$$\begin{aligned} M_{21} \quad = \quad & (\alpha \Delta^T) \otimes \sigma^* + \sigma^* \otimes (\alpha \Delta^T) + (\Lambda \otimes (\Lambda \Delta^T)) \\ & + (\Lambda \otimes \alpha)(\Upsilon \odot \mu^{(1)})^T + (\alpha \otimes \Lambda)(\Upsilon \odot \mu^{(1)})^T + (\beta \otimes \alpha + \alpha \otimes \beta) \pi^T \mu^{(2)} \\ & + (\alpha \otimes \alpha)(\Xi \odot \mu^{(2)})^T + \beta \otimes \sigma^* + \sigma^* \otimes \beta, \end{aligned}$$

$$M_{22} = (\alpha \otimes \alpha) vec(D)^T + (\alpha \Delta^T) \otimes \beta + \beta \otimes (\alpha \Delta^T) + \beta \otimes \beta.$$

By the law of iterated expectations we have

$$E_{t-h-1}(W_t) = \sum_{i=1}^{h-1} M^i c + M^h W_{t-h}. \tag{26}$$

As h goes to infinity, the limit exists and does not depend on t if and only if all the eigenvalues of M lie inside the unit circle, i.e., all the eigenvalues of M_{11} and M_{22} lie inside the unit circle:

$$\lim_{h \longrightarrow +\infty} E_{t-h-1}(W_t) = E(W_t) = (I - M)^{-1} c. \tag{27}$$

We deduce that the process is covariance stationary if all the eigenvalues of M_{11} lie inside the unit circle, and the fourth moment exists if all the eigenvalues of M_{11} and M_{22} lie inside the unit circle.

We focus next on the autocorrelations for the squared process. Consider the diagonal MEP-AGARCH(1,1) process, then from (27)

$$E(h_t) = (I - \beta - \alpha\Delta^T)^{-1}(\sigma^* + \alpha\pi^T\mu^{(2)}), \tag{28}$$

and the two-step ahead forecast of the variance vector is

$$\begin{aligned} E_{t-1}(h_{t+1}) &= \sigma^* + \alpha E_{t-1}(\varepsilon_t^2) - 2\alpha \odot \delta E_{t-1}(\varepsilon_t) + \beta h_t \\ &= (\sigma^* + \alpha\pi^T\mu^{(2)}) + (\alpha\Delta^T + \beta)h_t \\ &= E(h_t) + (\alpha\Delta^T + \beta)(h_t - E(h_t)). \end{aligned} \tag{29}$$

By recursive substitution we get the τ-step ahead forecast of h_t

$$E_{t-1}(h_{t+\tau}) = E(h_t) + (\alpha\Delta^T + \beta)^\tau (h_t - E(h_t)). \tag{30}$$

If the process has a finite fourth moment, then

$$\begin{aligned} E(\varepsilon_t^2\varepsilon_{t-\tau}^2) &= E(\varepsilon_{t-\tau}^2 E_{t-\tau}(\varepsilon_t^2)) \\ &= E(\varepsilon_{t-\tau}^2 E_{t-\tau}(\pi^T\mu^{(2)} + \Delta^T h_t)) \\ &= \pi^T\mu^{(2)}E(\varepsilon_t^2) + \Delta^T E(\varepsilon_{t-\tau}^2 E_{t-\tau}(h_t)). \end{aligned} \tag{31}$$

Using (30) and (23) we get

$$\begin{aligned} E(\varepsilon_t^2\varepsilon_{t-\tau}^2) &= \pi^T\mu^{(2)}E(\varepsilon_t^2) + \Delta^T E(h_t)E(\varepsilon_t^2) \\ &\quad + \Delta^T(\alpha\Delta^T + \beta)^{\tau-1}\left[\sigma^* E(\varepsilon_t^2) + \alpha E(\varepsilon_t^4) + \Lambda E(\varepsilon_t^3)\right. \\ &\quad \left. + \beta\left(\pi^T\mu^{(2)}E(h_t) + E(h_t h_t^T)\Delta\right) - E(h_t)E(\varepsilon_t^2)\right] \\ &= E^2(\varepsilon_t^2) + \Delta^T(\alpha\Delta^T + \beta)^{\tau-1}\left[\sigma E(\varepsilon_t^2) + \alpha E(\varepsilon_t^4)\right. \\ &\quad \left. + \beta\left(\pi^T\mu^{(2)}E(h_t) + E(h_t h_t^T)\Delta\right) - E(h_t)E(\varepsilon_t^2)\right]. \end{aligned} \tag{32}$$

Therefore by (28) and (4) we get

$$\begin{aligned} cov(\varepsilon_t^2, \varepsilon_{t-\tau}^2) &= \Delta^T(\alpha\Delta^T + \beta)^{\tau-1}\left\{\sigma^* E(\varepsilon_t^2) + \alpha E(\varepsilon_t^4) + \Lambda E(\varepsilon_t^3)\right. \\ &\quad \left. + \beta\left(\pi^T\mu^{(2)}E(h_t) + E(h_t h_t^T)\Delta\right) - E(h_t)E(\varepsilon_t^2)\right\}. \end{aligned} \tag{33}$$

End of proof ■

References

Alexander, C., and E. Lazar (2005): "Normal Mixture GARCH(1,1): Application to Exchange Rate Modelling," *Journal of Applied Econometrics*, 20, 1–30.

Azzalini, A. (1985): "A Class of Distributions which Includes the Normal ones," *Scandinavian Journal of Statistics*, 12, 171–178.

Bauwens, L., C. Hafner, and J. Rombouts (2007): "Multivariate Mixed Normal Conditional Heteroskedasticity," *Computational Statistics and Data Analysis*, 51, 3551–3566.

Bauwens, L., and J. Rombouts (2007a): "Bayesian Clustering of Many GARCH Models," *Econometric Reviews*, 26, 365–386.

——— (2007b): "Bayesian Inference for the Mixed Conditional Heteroskedasticity Model," *Econometrics Journal*, 10, 408–425.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., R. Engle, and D. Nelson (1994): "ARCH Models," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, chap. 4, pp. 2959–3038. North Holland Press, Amsterdam.

Boothe, P., and D. Glassman (1987): "The Statistical Distribution of Exchange Rates," *Journal of International Economics*, 22, 297–319.

Brooks, C., S. Burke, S. Heravi, and G. Persand (2005): "Autoregressive conditional Kurtosis," *Journal of Financial Econometrics*, 3, 399–421.

Christoffersen, P. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

Diebold, F. (1986): "Comment on Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 51–56.

Engle, R., and V. Ng (1993): "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

Fernández, C., and M. Steel (1998): "On Bayesian Modelling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359–371.

Frühwirth-Schnatter, S. (2006): *Finite Mixture and Markov Switching Models*. Springer, New York.

Frühwirth-Schnatter, S., and s. Kaufmann (2008): "Model-Based Clustering of Multiple Time Series," *Journal of Business and Economic Statistics*, 26, 78–89.

Geweke, J., and M. Keane (2005): "Smoothly Mixing Regressions," Working Paper, University of Iowa. Forthcoming in *Journal of Econometrics*.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Haas, M., S. Mittnik, and M. Paolella (2004a): "Mixed Normal Conditional Heteroskedasticity," *Journal of Financial Econometrics*, 2, 211–250.

——— (2004b): "A New Approach to Markov-Switching GARCH Models," *Journal of Financial Econometrics*, 2, 493–530.

——— (2006): "Modelling and Predicting Market Risk with Laplace-Gaussian Mixture Distributions," *Applied Financial Economics*, 16, 1145–1162.

Hardouvelis, G., and P. Theodossiou (2002): "The Asymmetric Relation Between Initial Margin Requirements and Stock Market Volatility Across Bull and Bear Markets," *The Review of Financial Studies*, 15, 1525–1559.

Harvey, C., and A. Siddique (1999): "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34, 465–487.

——— (2000): "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263–1295.

Jones, M., and M. Feddy (2003): "A Skew Extension of the t-Distribution, with Applications," *Journal of the Royal Statistical Society, series B*, 65, 159–174.

Kim, D., and S. Kon (1994): "Alternative Models for the Conditional Heteroscedasticity of Stock Returns," *Journal of Business*, 67, 563–598.

Komunjer, I. (2007): "Asymmetric Power Distribution: Theory and Applications to Risk Measurement," *Journal of Applied Econometrics*, 22, 891–921.

Kon, S. (1984): "Models of Stock Returns - A Comparison," *Journal of Finance*, 39, 147–165.

Kuester, K., S. Mittnik, and M. Paolella (2006): "Value-at-Risk Prediction: A Comparison of Alternative Strategies," *Journal of Financial Econometrics*, 4, 53–89.

Liesenfeld, R., and R. Jung (2000): "Stochastic Volatility Models: Conditional Normality versus Heavy-Tailed Distributions," *Journal of Applied Econometrics*, 15, 137–160.

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

Mikosch, T., and C. Starica (2004): "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects," *Review of Economics and Statistics*, 86, 378–390.

Mittnik, S., M. Paolella, and S. Rachev (2002): "Stationarity of Stable Power-GARCH Processes," *Journal of Econometrics*, 106, 97–107.

Nelson, D. (1991): "Conditional Heteroskedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 349–370.

Pan, M., K. Chan, and C. Fok (1995): "Currency Futures Price Changes: A Two-Piece Mixture of Normals Approach," *International Review of Economics and Finance*, 4, 69–78.

Tucker, A., and L. Pond (1988): "The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes," *The Review of Economics and Statistics*, 70, 638–647.

Vlaar, P., and F. Palm (1993): "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity, and Jumps," *Journal of Business and Economic Statistics*, 11, 351–360.

Wong, C., and W. Li (2000): "On a Mixture Autoregressive Model," *Journal of the Royal Statistical Society, Series B*, 62, 95–115.

——— (2001): "On a Mixture Autoregressive Conditional Heteroscedastic Model," *Journal of the American Statistical Association*, 96, 982–995.