Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 3551–3566
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Multivariate mixed normal conditional heteroskedasticity

L. Bauwens[a], C.M. Hafner[b], J.V.K. Rombouts[c,*]

[a] *CORE and Department of Economics, Université catholique de Louvain, Belgium*
[b] *Institut de statistique and CORE, Université catholique de Louvain, Belgium*
[c] *HEC Montréal, CIRANO, CIRPEE, CORE and CREF, HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7*

Available online 7 November 2006

Abstract

A new multivariate volatility model where the conditional distribution of a vector time series is given by a mixture of multivariate normal distributions is proposed. Each of these distributions is allowed to have a time-varying covariance matrix. The process can be globally covariance stationary even though some components are not covariance stationary. Some theoretical properties of the model such as the unconditional covariance matrix and autocorrelations of squared returns are derived. The complexity of the model requires a powerful estimation algorithm. A simulation study compares estimation by maximum likelihood with the EM algorithm. Finally, the model is applied to daily US stock returns.

Keywords: Multivariate volatility; Finite mixture; EM algorithm

1. Introduction

Several authors have argued in favor of adding flexibility to the family of GARCH models by using the idea of mixture models. For example, extending the model of Wong and Li (2000, 2001), Haas et al. (2004a) propose a mixed normal conditional heteroskedastic model where the conditional distribution of returns is a mixture of normal distributions, each of which has a regime specific conditional variance specified as a GARCH equation. In this way, they avoid the problem of path-dependence of the conditional variance of regime-switching GARCH models outlined by Gray (1996) and solved by Bauwens et al. (2006b). Other related papers are those of Haas et al. (2004b) and Alexander and Lazar (2006). All these articles deal with a univariate setting.

Multivariate mixture models have been frequently used in an iid context, but not, to the best of our knowledge, for time series models of conditional volatility, in particular multivariate GARCH models. In this paper, we try to fill this gap by extending the univariate model of Haas et al. (2004a) to the multivariate case. Mixing two or more conditionally normal and heteroskedastic components can generate quite complex stochastic behavior, similar to the one often observed in financial time series. For example, it may be that a component is covariance stationary, another is not, but mixing them might again generate a covariance stationary process. It is possible that mixing many components, of which some are non-stationary, produces behavior similar to processes with long memory, but we have not investigated this issue further.

* Corresponding author. Tel.: +1 5143406466; fax: +1 5143406469.
E-mail address: jeroen.rombouts@hec.ca (J.V.K. Rombouts).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.10.012

Note that our approach is different from the regime-switching model of Pelletier (2006), where the unobserved state variable follows a Markov chain and where within a regime correlations are constant.

The paper is organized as follows. In Section 2, we define the model and derive its properties. In Section 3, we present the estimation methods. In Section 4, we illustrate the estimation methods on simulated data, and in Section 5, we present an application using daily data for two stocks. Proofs are relegated in an Appendix.

2. The model

Consider an N-dimensional vector time series $\{\varepsilon_t, t \in \mathbb{N}\}$. A flexible model for the distribution of ε_t conditional on the information set $\mathscr{F}_{t-1}$ is given by

$$\varepsilon_t|\mathscr{F}_{t-1} \sim f(\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_k, \Sigma_{1t}, \ldots, \Sigma_{kt}) \tag{1}$$

$$= \sum_{j=1}^{k} \lambda_j f(\varepsilon_t|\mu_j, \Sigma_{jt}), \tag{2}$$

where $\lambda_j > 0$, $j = 1, \ldots, k$, $\sum_{j=1}^{k} \lambda_j = 1$ and $f(\varepsilon_t|\mu_j, \Sigma_{jt})$ is a multivariate density with mean vector μ_j and variance–covariance matrix Σ_{jt}. Note that λ_j is the probability of being in state j, characterized by the density $f(\varepsilon_t|\mu_j, \Sigma_{jt})$, and λ_j is constant over time. Similarly, the means of each state density, $\mu_1, \ldots, \mu_k$, are assumed constant over time. If ε_t is an error term, one would like to impose a restriction on the μ_j such that the conditional mean of ε_t is zero. For example, one such condition is

$$\mu_k = -\sum_{j=1}^{k-1} (\lambda_j/\lambda_k)\mu_j. \tag{3}$$

The first two conditional moments of ε_t are then be given by

$$E[\varepsilon_t \mid \mathscr{F}_{t-1}) = 0 \tag{4}$$

and

$$\mathrm{Var}[\varepsilon_t \mid \mathscr{F}_{t-1}] = \sum_{j=1}^{k} \lambda_j \Sigma_{jt}. \tag{5}$$

The process ε_t is conditionally heteroskedastic as every Σ_{jt} is allowed to depend on the information set. We model this dependence using multivariate GARCH (MGARCH) specifications. In particular, we assume that Σ_{jt} is a function of ε_{t-1} and of $\Sigma_{j,t-1}$, which can be called a 'diagonality' restriction since the conditional variance of state j depends only on its own past. In principle, any MGARCH model (VEC, BEKK, DCC, ..., see Bauwens et al., 2006a) can be used, but we focus here on the VEC model. Each matrix Σ_{jt} is a VEC model, such that

$$h_{jt} = \mathrm{vech}(\Sigma_{jt}), \tag{6}$$

has the dynamic structure

$$h_{jt} = \omega_j + A_j \eta_{t-1} + B_j h_{j,t-1}, \tag{7}$$

where ω_j is a vector of $N^* = N \times (N+1)/2$ parameters, A_j and B_j are square matrices of order N^*, and

$$\eta_t = \mathrm{vech}(\varepsilon_t \varepsilon_t'). \tag{8}$$

In words, we have k VEC models with common shocks that are a function of ε_t. We can write the model compactly as

$$h_t = \omega + A\eta_{t-1} + Bh_{t-1}, \tag{9}$$

where

$$h_t = \begin{pmatrix} h_{1t} \\ h_{2t} \\ \vdots \\ h_{kt} \end{pmatrix}_{kN^* \times 1}, \quad \omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{pmatrix}_{kN^* \times 1}, \quad A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{pmatrix}_{kN^* \times N^*} \tag{10}$$

and

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_k \end{pmatrix}_{kN^* \times kN^*}. \tag{11}$$

We refer to the process defined by Eqs. (1)–(10) as the MN-MGARCH (VEC), for mixed normal MGARCH (in VEC version), model.

For later reference, we provide the uncentered conditional second moment of ε_t,

$$E[\eta_t \mid \mathscr{F}_{t-1}] = \Lambda' h_t + c, \tag{12}$$

where $c = \sum_{i=1}^{k} \lambda_i \operatorname{vech}(\mu_i \mu_i')$, and

$$\Lambda = \begin{pmatrix} \lambda_1 I_{N^*} \\ \vdots \\ \lambda_k I_{N^*} \end{pmatrix}_{kN^* \times N^*}. \tag{13}$$

Theorem 1. *The process* $\{\varepsilon_t\}$ *defined by* (1)–(10) *is covariance stationary if and only if the eigenvalues of the matrix*

$$C = A\Lambda' + B, \tag{14}$$

are smaller than one in modulus. In that case,

$$h = E[h_t] = (I_{kN^*} - C)^{-1}(\omega + Ac), \tag{15}$$

and the unconditional covariance matrix is given by

$$E[\eta_t] = \Lambda'(I_{kN^*} - C)^{-1}(\omega + Ac) + c. \tag{16}$$

The crucial matrix to check is therefore C, written explicitly

$$C = \begin{pmatrix} \lambda_1 A_1 + B_1 & \lambda_2 A_1 & \cdots & \lambda_k A_1 \\ \lambda_1 A_2 & \lambda_2 A_2 + B_2 & \cdots & \lambda_k A_2 \\ \vdots & & \ddots & \\ \lambda_1 A_k & \cdots & & \lambda_k A_k + B_k \end{pmatrix}_{kN^* \times kN^*}.$$

We can get results on the fourth moment structure of the model by assuming that the densities of the individual states are spherical. For simplicity we assume that they are Gaussian with mean zero. Some more notation is necessary. Denote by $\tilde{\Lambda}$ the $N^{*2} \times kN^{*2}$ matrix

$$\tilde{\Lambda} = (\lambda_1 I_{N^{*2}}, \ldots, \lambda_k I_{N^{*2}}).$$

Furthermore, let P_{kq} be the $kq^2 \times (kq)^2$ selection matrix such that for any $kq \times kq$ matrix A, P_{kq} vec $A = (\operatorname{vec}(A_1)', \ldots, \operatorname{vec}(A_k)')'$, where A_j is the jth $q \times q$ matrix on the block-diagonal of A.

Theorem 2. *For the process defined by* (1)–(10), *assume that* $f(\varepsilon_t \mid \mu_j, \Sigma_{jt}) = N(0, \Sigma_{jt})$. *Then a necessary and sufficient condition for finite fourth moments of* ε_t *is that the eigenvalues of the matrix*

$$Z = (A \otimes A)G_N \tilde{\Lambda} P_{kN^*} + B \otimes B + B \otimes A\Lambda' + A\Lambda' \otimes B, \tag{17}$$

have modulus smaller than one, where

$$G_N = 2\{(D_N^+ \otimes D_N^+)(I_N \otimes C_{NN} \otimes I_N)(D_N \otimes D_N) + I_{N^{*2}}\},$$

C_{NN} is the commutation matrix, D_N the duplication matrix and D_N^+ its generalized inverse. In that case, the unconditional fourth moments of ε_t are given by

$$\mathrm{vec}(\Sigma_\eta) = G_N \tilde{\Lambda} P_{kN^*} (I_{N^{*2}} - Z)^{-1} \gamma, \tag{18}$$

where

$$\gamma = \mathrm{vec}(\omega\omega' + \omega h' \Lambda A' + \omega h' B + A\Lambda' h\omega' + B' h\omega'),$$

and h is given by (15). *Moreover, the autocovariance function of η_t, $\Gamma(\tau) = E[\eta_t \eta_{t-\tau}] - E[\eta_t]E[\eta_t]'$ is given by*

$$\Gamma(\tau) = \Lambda' C^{\tau-1} \left\{ A\Sigma_\eta + B\Sigma_h \Lambda - C(I_{kN^*} - C)^{-1} \omega\omega' (I_{kN^*} - C)^{-1} \Lambda \right\}, \tag{19}$$

where $\Sigma_h = E[h_t h_t']$.

The matrix D_N denotes the $N^2 \times N(N+1)/2$ duplication matrix, defined by the property $D_N\,\mathrm{vech}(A) = \mathrm{vec}(A)$ for any symmetric $(N \times N)$ matrix A, $D_N^+ = (D_N' D_N)^{-1} D_N'$, and $C_{NN} = 2D_N D_N^+ - I_{N^2}$. More details can be found in Lütkepohl, 1996. Theorem 2 generalizes the results of Haas et al. (2004a) to the multivariate case. Note that (19) can be used to calculate auto- and cross-correlations of squared errors. Moreover, existence of fourth moments is often assumed in deriving the asymptotic distribution of the maximum likelihood (ML) estimator in GARCH-type models. Our condition in Theorem 2 for finite fourth moments of the given process is easy to check, and we assume in the following that it holds.

3. Estimation

We describe how we perform estimation by the ML method (Section 3.1) and by the expectation–maximization (EM) algorithm (Section 3.2) of Dempster et al. (1977). We assume that T observation vectors y_t, for $t = 1$ to T, are available for estimation. The link between y_t and ε_t in (1) is given by $\varepsilon_t = y_t - E(y_t|\mathcal{F}_{t-1})$. We suppose for ease of the presentation that the conditional mean is either known or estimated consistently in a first step, so that the residuals ε_t are available for estimation of the parameters of the MN-MGARCH (VEC) model in the second step. We denote by ε the vector of observations $(\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_T')'$. We do not write explicitly the observations before $t = 1$, which are used as initial conditions where they should appear. The complete parameter vector, called Ψ, regroups the parameters λ_j, μ_j, and θ_j for $j = 1, \ldots, k$, where θ_j' is the row vector containing all the parameters of ω_j, A_j and B_j, see Eq. (7). Thus, $\Psi = (\mu', \theta', \lambda')'$, where $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_k)$, $\mu' = (\mu_1', \mu_2', \ldots, \mu_k')$, and $\theta' = (\theta_1', \theta_2', \ldots, \theta_k')$.

3.1. ML estimation

The log-likelihood of ε for the MN-MGARCH(VEC) model is given by

$$\mathcal{L}(\Psi; \varepsilon) = \sum_{t=1}^{T} \log \left(\sum_{j=1}^{k} \lambda_j \phi(\varepsilon_t | \mu_j, \Sigma_{jt}(\theta_j)) \right), \tag{20}$$

where $\phi(\cdot|\mu_j, \Sigma_{jt}(\theta_j))$ denotes a multivariate normal density with mean μ_j and variance–covariance matrix denoted by $\Sigma_{jt}(\theta_j)$, see Eq. (6).

Numerical methods are needed to obtain $\hat{\Psi} = \arg\max \mathcal{L}(\Psi; \varepsilon)$. To avoid the problem of label switching, we impose the identifying restrictions

$$\lambda_1 > \lambda_2 > \cdots > \lambda_k. \tag{21}$$

Because of these restrictions, we use the FSQP algorithm of Lawrence and Tits (2001) which allows optimization subject to constraints. We impose here the same restrictions as in Haas et al. (2004a). In the context of financial time

series this restriction seems to work fine, as shown in the simulation and the application of this paper. Alternative restrictions on the other parameters are also possible, see Hamilton et al. (2005) for examples on mixtures and other econometric models. More details on the label-switching problem in finite mixtures for dynamic models can be found in Frühwirth-Schnatter (2001).

It is a well known fact that the likelihood function can be unbounded in finite mixture models, see McLachlan and Peel (2000). Fortunately, for the ML estimation this is only a theoretical issue because the pole at which this happens is virtually impossible in our context of multivariate heteroskedastic time series.

3.2. EM algorithm

In the EM framework, the observed data vector ε is considered as incomplete since we do not know from which component of the mixture each observation is generated. This information is given by the latent variable $z_t = (z_{t1}, z_{t2}, \ldots, z_{tK})'$ where z_{tk} is a dichotomous variable taking the value 1 if ε_t comes from the kth mixture component, and 0 otherwise. The complete data log-likelihood is given by

$$\mathscr{L}_c(\Psi; \varepsilon) = \sum_{t=1}^{T} \sum_{j=1}^{k} z_{tj} [\log \lambda_j + \log \phi_j(\varepsilon_t | \mu_j, \theta_j)]. \tag{22}$$

This simplifies the expression of the log-likelihood in (20) because we do not take the logarithm over the entire sum but a sum of logarithms. Because z_t is not observed, we proceed in two steps.

E-step: Suppose that Ψ is known and equal to $\Psi^{(i)}$. We compute the expectation of the unobserved variable z_{tj} given all the observations y. This is given by

$$E(z_{tj} | \varepsilon, \Psi^{(i)}) = \tau_{tj}(\varepsilon; \Psi^{(i)}) = \frac{\lambda_j^{(i)} \phi(\varepsilon_t | \mu_j^{(i)}, \theta_j^{(i)})}{\sum_{j=1}^{k} \lambda_j^{(i)} \phi(\varepsilon_t | \mu_j^{(i)}, \theta_j^{(i)})}. \tag{23}$$

Next, we substitute the latter for z_{tk} in (22). This yields the observed complete data log-likelihood:

$$Q(\Psi, \Psi^{(i)}; \varepsilon) = \sum_{t=1}^{T} \sum_{j=1}^{k} \tau_{tj}(\varepsilon; \Psi^{(i)}) [\log \lambda_j + \log \phi(\varepsilon_t | \mu_j, \theta_j)]. \tag{24}$$

M-step: We maximize numerically $Q(\Psi, \Psi^{(i)}; \varepsilon)$ with respect to Ψ to get updated estimates of the parameters, denoted by $\Psi^{(i+1)}$. Notice that we have to impose the constraints (21) and (3), so that the maximization has to be done numerically with respect to all the parameters, including the probabilities.

The E-step and M-step are alternated repeatedly until convergence, see McLachlan and Peel (2000) for a detailed description of the application of the EM algorithm to mixture models.

4. Illustration with simulated data

We illustrate the estimation methods on two bivariate two component data generating processes for which we simulate one data set each. The first one has one stable component with high probability and one unstable component. The parameters are given by

*DGP*1:

$$\lambda_1 = 0.8, \quad \mu_1 = \begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix}, \quad \omega_1 = \begin{pmatrix} 0.001 \\ 0.005 \\ 0.02 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0.05 & 0.0 & 0.0 \\ 0.0 & 0.04 & 0.0 \\ 0.0 & 0.0 & 0.06 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0.92 & 0.0 & 0.0 \\ 0.0 & 0.9 & 0.0 \\ 0.0 & 0.0 & 0.85 \end{pmatrix},$$

$$\lambda_2 = 0.2, \quad \mu_2 = \begin{pmatrix} -0.4 \\ -0.2 \end{pmatrix}, \quad \omega_2 = \begin{pmatrix} 0.015 \\ 0.01 \\ 0.05 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.25 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0.85 & 0.0 & 0.0 \\ 0.0 & 0.75 & 0.0 \\ 0.0 & 0.0 & 0.8 \end{pmatrix}.$$

The largest eigenvalue of the matrix C in (14) is 0.96162 which is smaller than 1 so the overall process is stationary, even if for example $A_{2,11} + B_{2,11}$ is larger than 1. The implied unconditional standard deviations for the first and second series are, respectively, 0.648 and 0.662 and the unconditional correlation is 0.305.

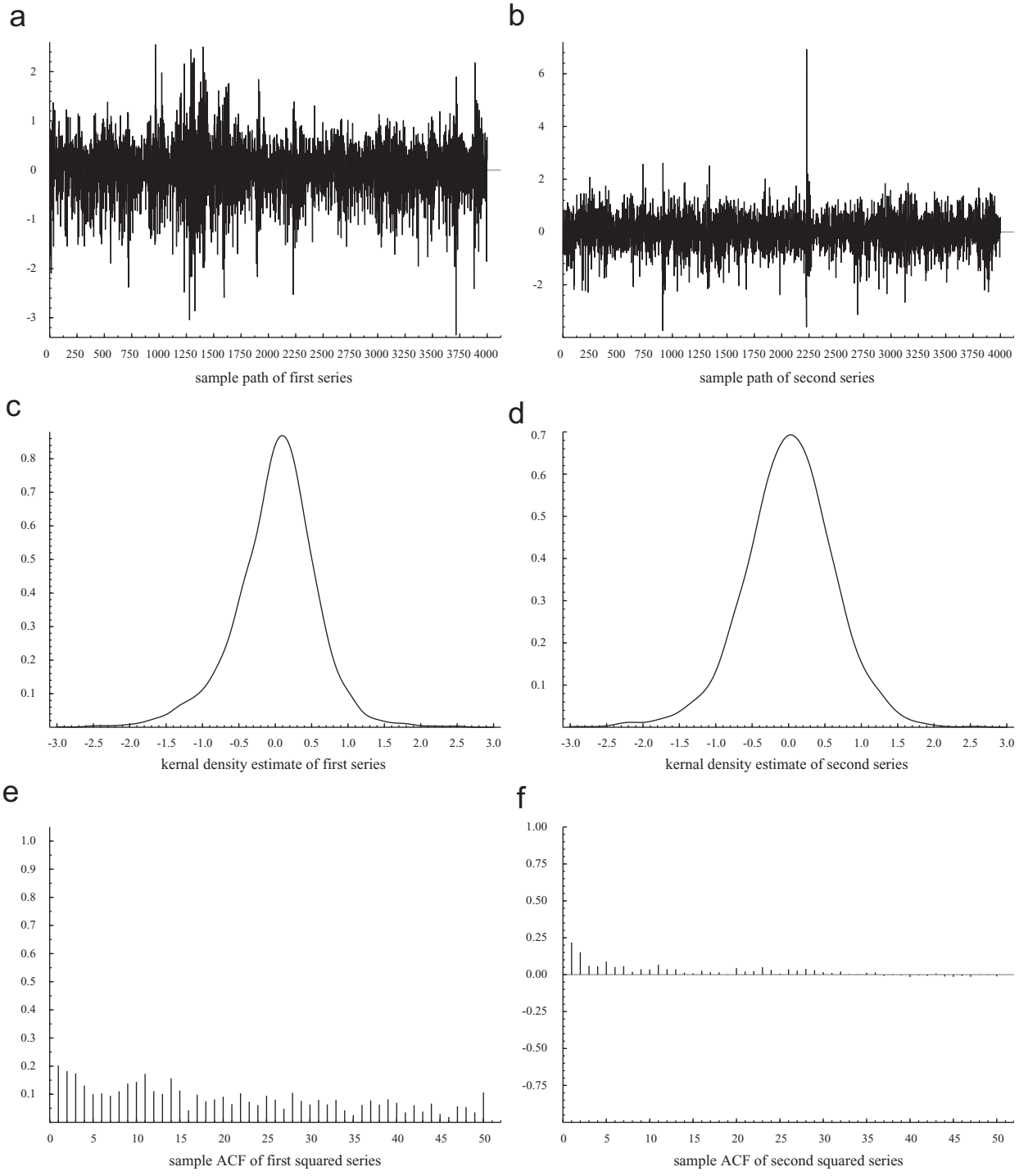

Fig. 1. Sample paths, kernel density estimates and sample autocorrelation functions of the data simulated using DGP 1 (4000 observations): (a) sample path of first series; (b) sample path of second series; (c) kernel density estimate of first series; (d) kernel density estimate of second series; (e) sample ACF of first squared series; (f) sample ACF of second squared series.

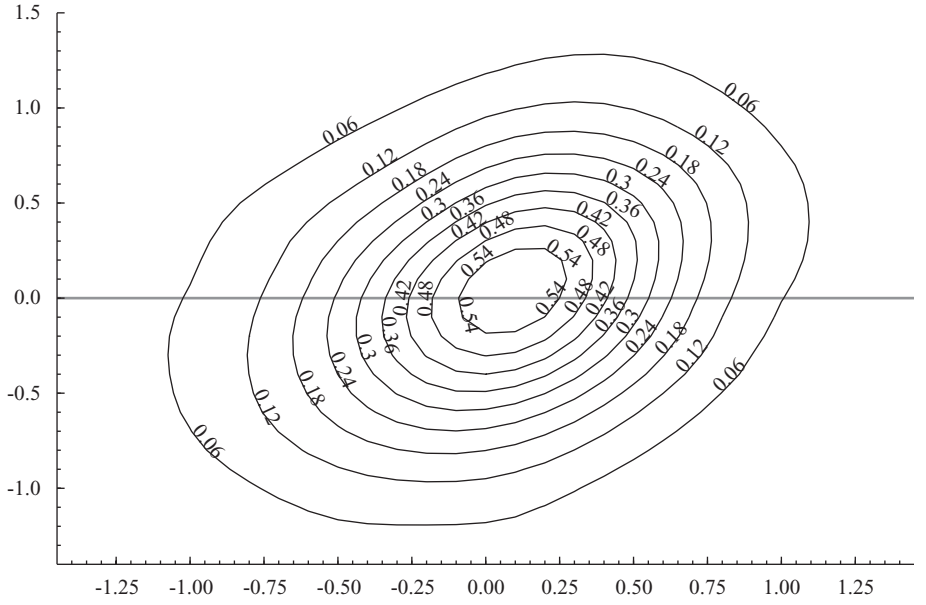

Fig. 2. Contour plot of the data simulated using DGP 1 (4000 observations).

The second DGP has the same first component as DGP1 but the second component is now less persistent than the first one. This is done by lowering the values in A_2 and B_2. The parameters are given by

*DGP*2:

$$\lambda_1 = 0.8, \quad \mu_1 = \begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix}, \quad \omega_1 = \begin{pmatrix} 0.001 \\ 0.005 \\ 0.02 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0.05 & 0.0 & 0.0 \\ 0.0 & 0.04 & 0.0 \\ 0.0 & 0.0 & 0.06 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0.92 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.85 \end{pmatrix},$$

$$\lambda_2 = 0.2, \quad \mu_2 = \begin{pmatrix} -0.4 \\ -0.2 \end{pmatrix}, \quad \omega_2 = \begin{pmatrix} 0.015 \\ 0.01 \\ 0.05 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.15 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.2 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0.45 & 0.0 & 0.0 \\ 0.0 & 0.35 & 0.0 \\ 0.0 & 0.0 & 0.5 \end{pmatrix}.$$

The largest eigenvalue of the matrix C in (14) is given by 0.96021 which is smaller than 1 so that the overall process is stationary, which is not surprising here since both components are stable. The implied unconditional standard deviations for the first and second series are, respectively, 0.353 and 0.477 and the unconditional correlation is 0.316.

We simulated $T = 4000$ observations for DGP1 and DGP2. The sample paths, marginal kernel density estimates and sample autocorrelation functions of the data simulated using DGP 1 are given in Fig. 1. A contour plot is given in Fig. 2. From the graphs we see that the sample autocorrelations for the squared data decay faster for the second series as expected given the DGP1 parameter values, and that there is a higher negative skewness in the first series than in the second. This is indeed confirmed by the summary statistics given in Table 1. The estimated kurtosis coefficient is higher for the second series, though this is likely due to the high maximum in that series. Note that the empirical second moments match the theoretical second moments reasonably well, for example the estimated and theoretical correlation are, respectively, given by 0.314 and 0.305.

In Table 2, we report the parameter estimates of DGP1 obtained by using the EM algorithm, see Section 3 for details. We assume that asymptotic normality holds and that standard errors can be computed via the Hessian of (19) evaluated at the estimates. The estimates are reasonably close to the true parameter values given the standard errors. Note that the standard errors of the parameters in the second component are drastically higher compared to those of the first component. This corresponds to the intuition that there is less information in the data about the low probability regime than about the other one. The EM estimates are almost identical to the ML estimates. The EM standard errors are computed in the same way as for the ML estimates so they also hardly differ.

We now turn to DGP2. The sample paths, marginal kernel density estimates and sample autocorrelation functions of the simulated data are given in Fig. 3. A contour plot is given in Fig. 4. Descriptive statistics are given in Table 3.

Table 1
DGP 1 summary statistics

	First series	Second series
Mean	0.046	0.042
Standard deviation	0.570	0.638
Maximum	2.548	6.922
Minimum	−3.349	−3.738
Skewness	−0.453	−0.037
Kurtosis	5.298	7.862

Descriptive statistics of the data simulated using DGP 1 ($T = 4000$). The estimated correlation coefficient is 0.31433.

Table 2
Estimation results for two components models

	DGP1	Estimate	Std. error	DGP2	Estimate	Std. error
λ_1	0.80	0.812	0.0205	0.80	0.811	0.0302
λ_2	0.20	0.188	0.0205	0.20	0.189	0.0302
$\mu_{1,1}$	0.10	0.105	0.0080	0.10	0.092	0.0125
$\mu_{1,2}$	0.05	0.058	0.0092	0.05	0.054	0.0120
$\mu_{2,1}$	−0.40	−0.453	0.0860	−0.40	−0.394	0.1316
$\mu_{2,2}$	−0.20	−0.250	0.0590	−0.20	−0.232	0.0934
$\omega_{1,11}$	0.001	0.001	0.0005	0.001	0.001	0.0006
$\omega_{1,22}$	0.005	0.004	0.0010	0.005	0.002	0.0008
$\omega_{1,33}$	0.02	0.015	0.0038	0.02	0.027	0.0069
$A_{1,11}$	0.05	0.044	0.0047	0.05	0.057	0.0068
$A_{1,22}$	0.04	0.033	0.0052	0.04	0.031	0.0088
$A_{1,33}$	0.06	0.052	0.0088	0.06	0.078	0.0143
$B_{1,11}$	0.92	0.930	0.0075	0.92	0.911	0.0103
$B_{1,22}$	0.90	0.916	0.0148	0.80	0.908	0.0300
$B_{1,33}$	0.85	0.876	0.0210	0.85	0.798	0.0388
$\omega_{2,11}$	0.015	0.020	0.0102	0.015	0.012	0.0056
$\omega_{2,22}$	0.01	0.003	0.0097	0.01	0.005	0.0049
$\omega_{2,33}$	0.05	0.097	0.0384	0.05	0.050	0.0193
$A_{2,11}$	0.25	0.234	0.0521	0.15	0.170	0.0406
$A_{2,22}$	0.20	0.194	0.0569	0.10	0.096	0.0364
$A_{2,33}$	0.30	0.302	0.0739	0.20	0.238	0.0584
$B_{2,11}$	0.85	0.824	0.0388	0.45	0.443	0.1398
$B_{2,22}$	0.75	0.703	0.0966	0.35	0.353	0.1854
$B_{2,33}$	0.80	0.728	0.0682	0.50	0.458	0.1227

The lower autocorrelations in the squared data compared to DGP1 are not surprising given the much less persistent second component in the mixture. The standard deviations are also smaller compared to DGP1 because we keep the same values in DGP2 for ω_1 and ω_2. Estimation results for DGP2 (by the EM algorithm) are given in Table 2. The parameter estimates are again reasonably close to the DGP values.

To confirm that our algorithms seem correct, we generated some extra sample paths of the same sample size for both DGP1 and DGP2 and we estimated the model parameters again. The results, which are not reported, were of the same type as described in this section.

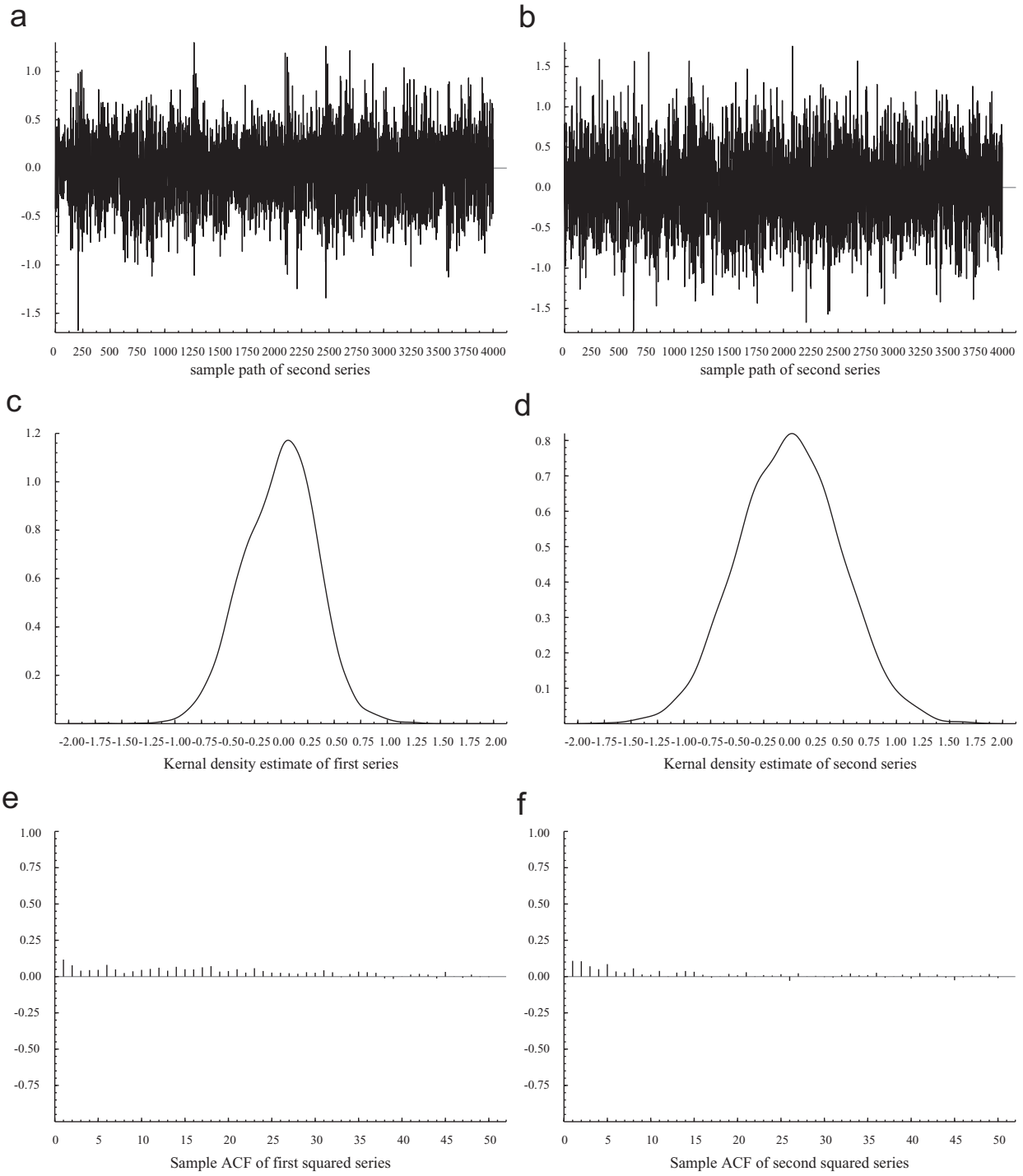

Fig. 3. Sample paths, kernel density estimates and sample autocorrelation functions of the data simulated using DGP 2 (4000 observations): (a) sample path of first series; (b) sample path of second series; (c) kernel density estimate of first series; (d) kernel density estimate of second series; (e) sample ACF of first squared series; (f) sample ACF of second squared series.

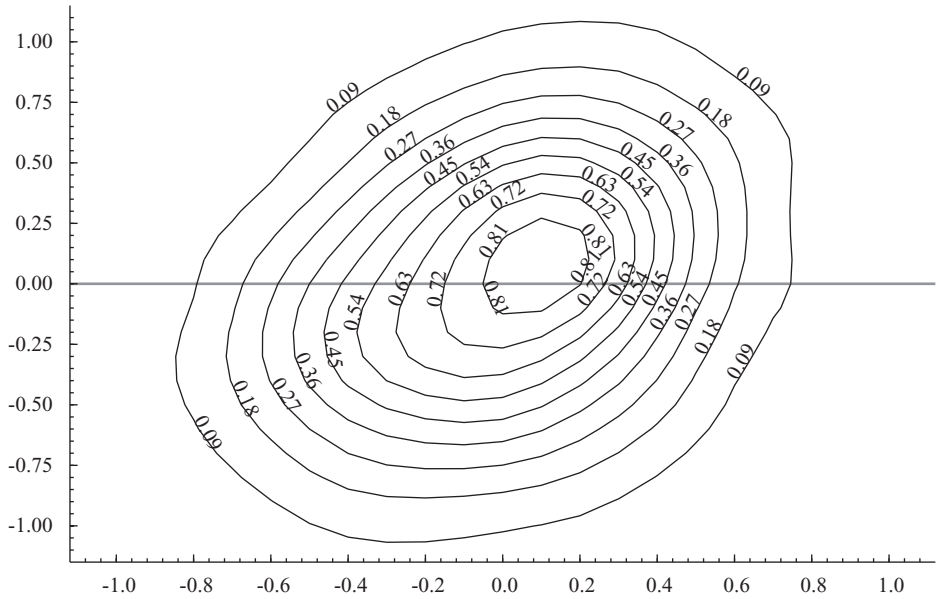

Fig. 4. Contour plot of the data simulated using DGP 2 (4000 observations).

Table 3
DGP 2 summary statistics

	First series	Second series
Mean	−0.010	−0.005
Standard deviation	0.344	0.481
Maximum	1.297	1.752
Minimum	−1.677	−1.777
Skewness	−0.122	0.012
Kurtosis	3.206	3.025

Descriptive statistics of the data simulated using DGP 2 ($T = 4000$). The estimated correlation coefficient is 0.32746.

5. Application

We model daily return data from the Bank of America and Boeing stocks using a sample from 01/01/1980 to 30/07/2003 implying 6152 observations downloaded from Datastream. Daily returns are measured by log-differences of closing prices multiplied by 100. The sample paths, marginal kernel density estimates and sample autocorrelation functions of the data are given in Fig. 5. A contour plot is given in Fig. 6. Both companies share similar summary statistics which are given in Table 4. Some important events between 1980 and 2003 gave rise to several extreme values for both companies. These values are not discarded from the sample. We start by fitting univariate one and two component models to learn more about the individual time series dynamics of both companies and also to get an idea of good starting values for the multivariate mixture model. The ML estimates for the univariate models are reported in Table 5. The one component model, or the usual GARCH model, estimates for both Bank of America and Boeing imply stationary but highly persistent processes. The two component mixture model parameter estimates reveal indeed that for both companies the second component is not stable with probabilities belonging to that component, respectively, given by 0.165 and 0.079.

The estimation results, obtained by using the EM algorithm, for the bivariate one and two component models are given in Table 6. The ML estimation results hardly differ and are not reported here. The largest eigenvalue of the estimated matrix C in (14) is given by 0.98435 which implies a stationary process. The implied estimated unconditional standard deviations for Bank of America and Boeing are, respectively, given by 1.8653 and 1.9716 and the unconditional

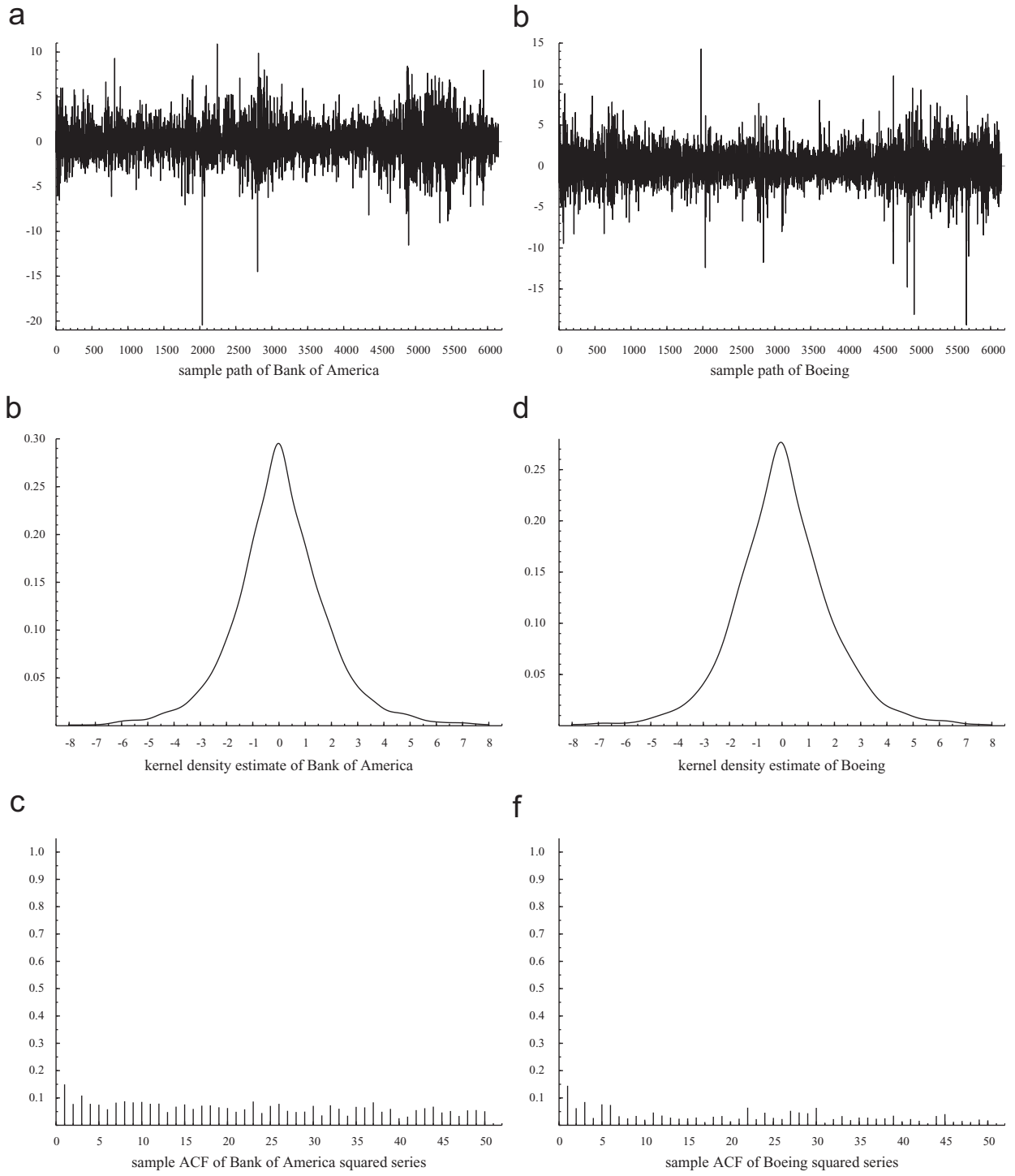

Fig. 5. Sample paths, kernel density estimates and sample autocorrelation functions for the Bank of America - Boeing data: (a) sample path of Bank of America; (b) sample path of Boeing; (c) kernel density estimate of Bank of America; (d) kernel density estimate of Boeing; (e) sample ACF of Bank of America squared series; (f) sample ACF of Boeing squared series.

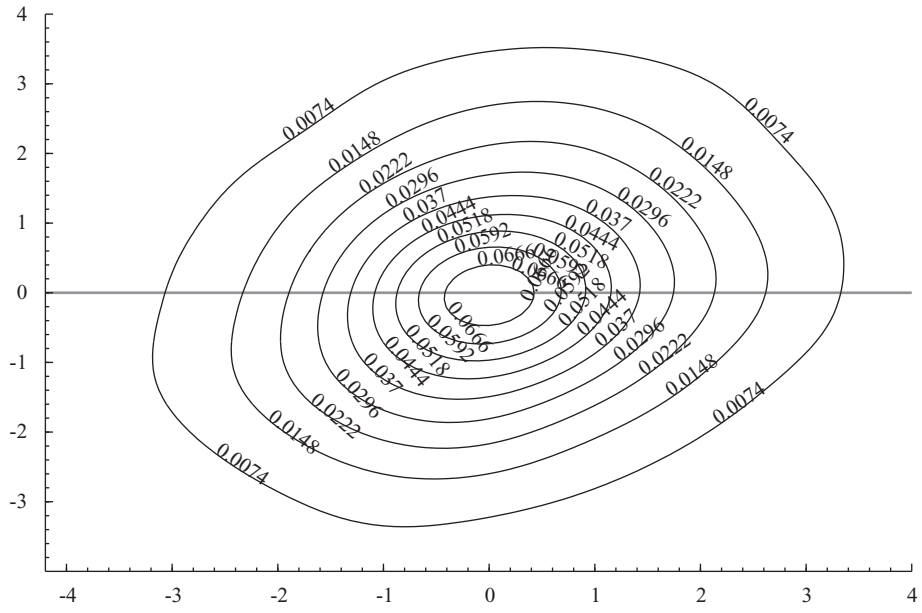

Fig. 6. Contour plot for the Bank of America—Boeing data.

Table 4
Bank of America—Boeing summary statistics

	Bank of America	Boeing
Mean	0.052	0.0304
Standard deviation	1.892	1.973
Maximum	10.903	14.278
Minimum	−20.458	−19.389
Skewness	−0.176	−0.281
Kurtosis	8.057	9.104

Descriptive statistics for the Bank of America—Boeing data. Data from 01/01/1980–30/07/2003 or $T = 6152$. The estimated correlation coefficient is 0.25448.

Table 5
Univariate estimation results

	Bank of America				Boeing			
	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
λ_1	—	—	0.836	0.028	—	—	0.921	0.016
μ_1	—	—	0.022	0.020	—	—	−0.006	0.017
ω_1	0.135	0.030	0.031	0.013	0.055	0.002	0.029	0.009
$A_{1,11}$	0.082	0.012	0.047	0.009	0.040	0.009	0.030	0.005
$B_{1,11}$	0.881	0.019	0.918	0.017	0.946	0.016	0.948	0.009
ω_2	—	—	3.130	0.944	—	—	2.653	1.218
$A_{2,11}$	—	—	0.698	0.183	—	—	0.511	0.202
$B_{2,11}$	—	—	0.416	0.125	—	—	0.730	0.090
$\mathscr{L}(\hat{\Psi})$	−12231.65		−12019.42		−12608.33		−12339.15	

Results for the one component (first two columns for each company) and two component (last two columns for each company) univariate mixture GARCH(1,1) models. All the parameters are estimated by ML.

Table 6
Bank of America—Boeing results

	Estimate	Std. error	Estimate	Std. error
λ_1	—	—	0.848	0.012
λ_2	—	—	0.152	0.012
$\mu_{1,1}$	—	—	0.037	0.018
$\mu_{1,2}$	—	—	−0.013	0.019
$\mu_{2,1}$	—	—	−0.206	0.108
$\mu_{2,2}$	—	—	0.073	0.112
$\omega_{1,11}$	0.146	0.029	0.030	0.010
$\omega_{1,22}$	0.026	0.008	0.004	0.002
$\omega_{1,33}$	0.073	0.017	0.028	0.009
$A_{1,11}$	0.084	0.011	0.045	0.008
$A_{1,22}$	0.026	0.005	0.011	0.003
$A_{1,33}$	0.041	0.005	0.027	0.004
$B_{1,11}$	0.876	0.018	0.926	0.012
$B_{1,22}$	0.940	0.013	0.975	0.005
$B_{1,33}$	0.940	0.009	0.948	0.009
$\omega_{2,11}$	—	—	3.348	0.953
$\omega_{2,22}$	—	—	0.412	0.272
$\omega_{2,33}$	—	—	2.149	0.777
$A_{2,11}$	—	—	0.590	0.150
$A_{2,22}$	—	—	0.128	0.069
$A_{2,33}$	—	—	0.254	0.077
$B_{2,11}$	—	—	0.381	0.130
$B_{2,22}$	—	—	0.785	0.117
$B_{2,33}$	—	—	0.723	0.076

Results for the one component (first two columns) and two component (last two columns) bivariate mixture model. The value of the log-likelihood function evaluated at the EM estimates of the one and two component models are, respectively, given by −24663.714 and −24177.475.

correlation is 0.209, which is close to the summary statistics reported in Table 4. Our model assumes that the state probabilities λ_j are time invariant. To evaluate this, we compute the estimated probabilities via (23) the estimated autocorrelations of which are displayed in Fig. 7, with 95% confidence interval limits. The first estimated autocorrelation (=0.0468) is found to be significant given the high sample size. The higher order autocorrelations are not significant.

Comparing the univariate one component estimates with their equivalents in the bivariate one component model, or the usual diagonal VEC model, we see that they differ only marginally as expected. Generally speaking, this is also true for the bivariate two component model but to a lesser extent so for the second component which is now stable. The large difference in the log-likelihood function values evaluated at their ML estimates between the one and the two component models allows to reject easily a likelihood ratio test in favor of the more general model.

6. Conclusion

The multivariate mixture model we have proposed in this paper can be extended in several ways. One can use other multivariate GARCH models for the components than the VEC formulation. We refer to the survey of Bauwens et al. (2006a) for other multivariate GARCH models. One advantage of the VEC specification is the ease with which moments can be derived. One could also think of using non-normal distributions, but this may not be worth the effort since a mixture of normal distributions allows for a lot of flexibility. The most important challenge at this stage is to improve upon the estimation algorithms and to test them with time series of higher dimension. Another topic for future research is to evaluate the models on statistical and economic criteria, in comparison with one-component models. An advantage of the mixture model is that in high dimensions, simple models with few parameters could be mixed to

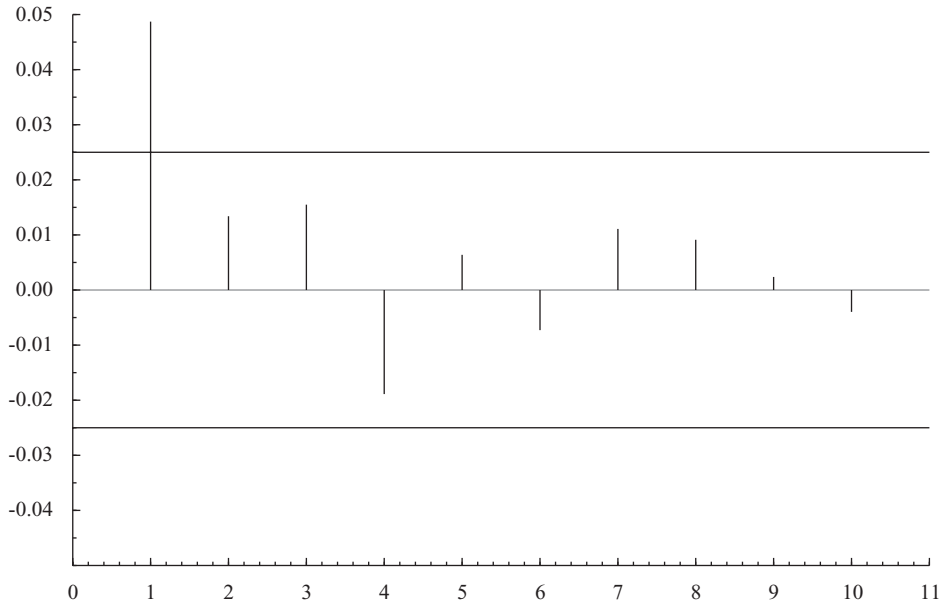

Fig. 7. Sample ACF for the estimated probabilities computed via (23) with asymptotic pointwise confidence intervals.

obtain more flexibility than specifying a complex one-component model. Finally, formal criteria could be employed for the choice of k as in Haas et al. (2004a).

Acknowledgments

We thank John Geweke, Razvan Sufana and three anonymous referees for useful comments. Bauwens' work was supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, MICFINMA. Both Bauwens and Hafner acknowledge financial support by an FSR Grant of Université catholique de Louvain. This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. Rombouts' work was supported by a HEC Montréal Fonds de démarrage and by the Centre for Research on e-Finance. The scientific responsibility is assumed by the authors.

Appendix

Proof of Theorem 1. Let $u_t = \eta_t - \Lambda' h_t - c$ and note that $E[u_t \mid I_{t-1}] = 0$. Write (9) as

$$\begin{aligned} h_t &= \omega + A(\Lambda' h_{t-1} + c + u_{t-1}) + Bh_{t-1}, \\ &= (\omega + Ac) + (A\Lambda' + B)h_{t-1} + Au_{t-1}. \end{aligned}$$

Denoting the lag operator by L and $C = A\Lambda' + B$, this can be written as

$$(I_{kN^*} - CL)h_t = (\omega + Ac) + Au_{t-1}. \tag{25}$$

The linear operator $(I_{kN^*} - CL)$ is invertible if and only if all eigenvalues of C have modulus smaller than one. In that case we can write $h_t = (I_{kN^*} - C)^{-1}(\omega + Ac) + (I_{kN^*} - CL)^{-1}Au_{t-1}$, which is a VMA($\infty$) representation of $\{h_t\}$ from which we directly deduce $h = E[h_t] = (I_{kN^*} - C)^{-1}(\omega + Ac)$. Premultiplying both sides of (25) by the adjoint, $(I_{kN^*} - CL)^*$, we obtain

$$\det(I_{kN^*} - CL)h_t = (I_{kN^*} - C)^*(\omega + Ac) + (I_{kN^*} - CL)^* Au_{t-1}.$$

Premultiplying by Λ' and using $\Lambda' h_t = \eta_t - u_t - c$ gives

$$\det(I_{kN^*} - CL)(\eta_t - u_t - c) = \Lambda'(I_{kN^*} - C)^*(\omega + Ac) + \Lambda'(I_{kN^*} - CL)^* A u_{t-1}.$$

The process is stable if and only if all roots of the characteristic equation $\det(I_{kN^*} - Cz) = 0$ lie outside the unit circle or, equivalently, all eigenvalues of C have modulus smaller than one. Finally, dividing both sides by $\det(I_{kN^*} - CL)$ and rearranging yields

$$\eta_t = \Lambda'(I_{kN^*} - C)^{-1}(\omega + Ac) + c + \Lambda'(I_{kN^*} - CL)^{-1} A u_{t-1} + u_t.$$

This is the VMA(∞) representation of $\{\eta_t\}$ and we deduce directly the unconditional variance of $\{\varepsilon_t\}$, i.e.

$$\operatorname{vech}(\operatorname{Var}(\varepsilon_t)) = E[\eta_t] = \Lambda'(I_{kN^*} - C)^{-1}(\omega + Ac) + c. \qquad \square$$

Proof of Theorem 2. First, $\operatorname{vec}(E[\eta_t \eta_t' \mid \mathscr{F}_{t-1}]) = G_N \sum_{j=1}^{k} \lambda_j \operatorname{vec}(h_{jt} h_{jt}')$ by application of Hafner (2003, Theorem 1). Taking the expectation operator on both sides yields

$$\operatorname{vec}(\Sigma_\eta) = G_N \tilde{\Lambda} P_{kN^*} \operatorname{vec}(\Sigma_h), \tag{26}$$

where $\Sigma_\eta = E[\eta_t \eta_t']$ and $\Sigma_h = E[h_t h_t']$. Substituting the model for h_t in Σ_h, one obtains

$$\begin{aligned}
\operatorname{vec}(\Sigma_h) &= \operatorname{vec}(\omega\omega' + \omega h' \Lambda A' + \omega h' B + A\Lambda' h \omega' + B' h \omega') \\
&\quad + (A \otimes A) \operatorname{vec}(E[\eta_{t-1}\eta_{t-1}']) + (B \otimes B) \operatorname{vec}(E[h_{t-1} h_{t-1}']) \\
&\quad + (B \otimes A) \operatorname{vec}(E[\eta_{t-1} h_{t-1}']) + (A \otimes B) \operatorname{vec}(E[h_{t-1}\eta_{t-1}']) \\
&= \gamma + (A \otimes A) \operatorname{vec}(\Sigma_\eta) + (B \otimes B) \operatorname{vec}(\Sigma_h) \\
&\quad + (B \otimes A) \operatorname{vec}(E[E(\eta_{t-1} h_{t-1}' \mid \mathscr{F}_{t-1})]) + (A \otimes B) \operatorname{vec}(E[E(h_{t-1}\eta_{t-1} \mid \mathscr{F}_{t-1})]), \\
&= \gamma + (A \otimes A) G_N \tilde{\Lambda} P_{kN^*} \operatorname{vec}(\Sigma_h) + (B \otimes B) \operatorname{vec}(\Sigma_h) \\
&\quad + (B \otimes A) \operatorname{vec}(E[\Lambda' h_{t-1} h_{t-1}']) + (A \otimes B) \operatorname{vec}(E[h_{t-1} h_{t-1}' \Lambda]) \\
&= \gamma + (A \otimes A) G_N \tilde{\Lambda} P_{kN^*} \operatorname{vec}(\Sigma_h) + (B \otimes B) \operatorname{vec}(\Sigma_h) \\
&\quad + (B \otimes A\Lambda') \operatorname{vec}(\Sigma_h) + (A\Lambda' \otimes B) \operatorname{vec}(\Sigma_h) \\
&= \gamma + Z \operatorname{vec}(\Sigma_h).
\end{aligned}$$

Rearranging gives the result provided that $I_{N^{*2}} - Z$ is invertible, which is the case if and only if all eigenvalues of Z have modulus smaller than one. Finally, application of (26) yields the desired result for Σ_η.

For the second part of the theorem, note that

$$\begin{aligned}
E[h_t \mid \mathscr{F}_{t-\tau}] &= (I_{kN^*} + C + \cdots + C^{\tau-1})\omega + C^{\tau-1}(A\eta_{t-\tau} + B h_{t-\tau}), \\
&= (I_{kN^*} - C^\tau)(I_{kN^*} - C)^{-1}\omega + C^{\tau-1}(A\eta_{t-\tau} + B h_{t-\tau}).
\end{aligned}$$

Now,

$$\begin{aligned}
E[\eta_t \eta_{t-\tau}] &= E[E(\eta_t \mid \mathscr{F}_{t-1})\eta_{t-\tau}'], \\
&= E[\Lambda' h_t \eta_{t-\tau}'], \\
&= E[\Lambda' E(h_t \mid \mathscr{F}_{t-\tau})\eta_{t-\tau}'], \\
&= E[\Lambda'\{(I_{kN^*} - C^\tau)(I_{kN^*} - C)^{-1}\omega + C^{\tau-1}(A\eta_{t-\tau} + B h_{t-\tau})\}\eta_{t-\tau}'], \\
&= \Lambda'(I_{kN^*} - C^\tau)(I_{kN^*} - C)^{-1}\omega\omega'(I_{kN^*} - C')^{-1}\Lambda + \Lambda' C^{\tau-1}(A\Sigma_\eta + B\Sigma_h \Lambda).
\end{aligned}$$

Subtracting $E[\eta_t]E[\eta_t]' = \Lambda'\{(I_{kN^*} - C)^{-1}\omega\omega'(I_{kN^*} - C')^{-1}\Lambda$, the result for $\Gamma(\tau)$ is obtained. $\square$

References

Alexander, C., Lazar, E., 2006. Normal mixture GARCH(1,1). J. Appl. Econometrics 21, 307–336.

Bauwens, L., Laurent, S., Rombouts, J., 2006a. Multivariate GARCH models: a survey. J. Appl. Econometrics 21, 79–109.

Bauwens, L., Preminger, A., Rombouts, J., 2006b. Regime-switching GARCH models. Revision of CORE Discussion Paper 2006-11.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. Ser. B 39, 1–38.

Frühwirth-Schnatter, S., 2001. Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models. J. Amer. Statist. Assoc. 96, 194–209.

Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. J. Financial Economics 42, 27–62.

Haas, M., Mittnik, S., Paolella, M., 2004a. Mixed normal conditional heteroskedasticity. J. Financial Econometrics 2, 211–250.

Haas, M., Mittnik, S., Paolella, M., 2004b. A new approach to markov-switching GARCH models. J. Financial Econometrics 2, 493–530.

Hafner, C., 2003. Fourth moment structure of multivariate GARCH models. J. Financial Econometrics 1, 26–54.

Hamilton, J., Waggoner, D., Zha, T., 2005. Normalization in econometrics. Working paper.

Lawrence, C., Tits, A., 2001. A computationally efficient feasible sequential quadratic programming algorithm. SIAM J. Optim. 11, 1092–1118.

Lütkepohl, H., 1996. Handbook of Matrices. Wiley, New York.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. Wiley Interscience, New York.

Pelletier, D., 2006. Regime switching for dynamic correlations. J. Econometrics 131, 445–473.

Wong, C., Li, W., 2000. On a mixture autoregressive model. J. Roy. Statist. Soc. Ser. B 62, 95–115.

Wong, C., Li, W., 2001. On a mixture autoregressive conditional heteroscedastic model. J. Amer. Statist. Assoc. 96, 982–995.