No. 2008/07

Asymmetric Multivariate Normal Mixture GARCH

Markus Haas, Stefan Mittnik,
and Marc S. Paolella

Center for Financial Studies
an der Universität Frankfurt ▪ Mertonstraße 17-21 ▪ D-60325 Frankfurt am Main

Telefon: +49 (0)69 798-30050
Fax: +49 (0)69 798-30077 ▪ http://www.ifk-cfs.de ▪ E-Mail: info@ifk-cfs.de

Center for Financial Studies

The *Center for Financial Studies* is a nonprofit research organization, supported by an association of more than 120 banks, insurance companies, industrial corporations and public institutions. Established in 1968 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia.

The CFS Working Paper Series presents the result of scientific research on selected topics in the field of money, banking and finance. The authors were either participants in the Center´s Research Fellow Program or members of one of the Center´s Research Projects.

If you would like to know more about the *Center for Financial Studies*, please let us know of your interest.

Prof. Dr. Jan Pieter Krahnen

Prof. Volker Wieland, Ph.D.

Center for Financial Studies
an der Universität Frankfurt ▪ Mertonstraße 17-21 ▪ D-60325 Frankfurt am Main
Telefon: +49 (0)69 798-30050
Fax: +49 (0)69 798-30077 ▪ http://www.ifk-cfs.de ▪ E-Mail: info@ifk-cfs.de

CFS Working Paper No. 2008/07

Asymmetric Multivariate Normal Mixture GARCH

Markus Haas[1], Stefan Mittnik[2],
and Mark S. Paolella[3]

January 18, 2008

Abstract:
An asymmetric multivariate generalization of the recently proposed class of normal mixture GARCH models is developed. Issues of parametrization and estimation are discussed. Conditions for covariance stationarity and the existence of the fourth moment are derived, and expressions for the dynamic correlation structure of the process are provided. In an application to stock market returns, it is shown that the disaggregation of the conditional (co)variance process generated by the model provides substantial intuition. Moreover, the model exhibits a strong performance in calculating out–of–sample Value–at–Risk measures.

JEL Classification: C32, C51, G10, G11

Keywords: Conditional Volatility, Finite Normal Mixtures, Multivariate GARCH, Leverage Effect

1 Corresponding Author: Institute of Statistics, University of Munich, D-80799 Munich, Germany; E-mail: haas@stat.uni-muenchen.de; Phone: +49 (0) 89 21 80 - 25 70; Fax: +49 (0) 89 21 80 - 50 44.

2 Department of Statistics, University of Munich, Center for Financial Studies, Frankfurt, and Ifo Institute for Economic Research, Munich

3 Swiss Banking Institute, University of Zurich, Switzerland

1 Introduction

Dynamic mixture models for the volatility of financial variables are gaining popularity, partly because they often provide a plausible disaggregation of the conditional variance process, and partly because they have been shown to deliver accurate out–of–sample predictive densities, which is important for risk management applications such as the computation of Value–at–Risk. A finite mixture of a few normal distributions, say two or three, is capable of capturing the skewness and kurtosis detected in both conditional and unconditional return distributions, and can, when coupled with GARCH–type equations for the component variances, exhibit quite complex dynamics, as often observed in financial markets. For example, there may be components driven by nonstationary dynamics, while the overall process is still stationary. This corresponds to the observation that markets are stable most of the time, but, occasionally, subject to severe, short–lived fluctuations. A general univariate normal mixture GARCH model, generalizing earlier specifications such as Vlaar and Palm (1993) and Wong and Li (2001), has been proposed by Haas et al. (2004) and Alexander and Lazar (2006) and further investigated by Alexander and Lazar (2005), Ausin and Galeano (2007), Bertholon et al. (2006), Haas et al. (2006a), Bauwens and Rombouts (2007), Wu and Lee (2007), and Giannikis et al. (2008).

All of the papers cited above are confined to univariate processes. Many problems in finance, however, are inherently multivariate and require us to understand the dependence structure between assets. For example, in applications to portfolio management, correlations between assets are often of predominant interest. Quite recently, in order to cope with such situations, Bauwens et al. (2007) proposed a multivariate version of the normal mixture GARCH model developed in Haas et al. (2004) and Alexander and Lazar (2006), investigated its fourth–moment structure and demonstrated its practicability in an application to a bivariate stock return series.

In this paper, we extend the work of Bauwens et al. (2007) in several ways. First, we enrich the model's structure by allowing for leverage effects, i.e., the "stylized fact" that, for stock returns, past negative shocks have a deeper impact on volatility than positive shocks. As this asymmetry is a robust feature of stock return series, we expect that its inclusion into the model will in many instances enhance its performance in density and volatility forecasting. Secondly, we provide a more complete characterization of the fourth–moment structure of the model, where we allow both for dynamic asymmetries, i.e., leverage effects, as well as for asymmetry of the conditional mixture density. Bauwens et al. (2007) account for the second type of asymmetry in the definition and the application of their model, but the fourth–moment

matrix as well as the autocorrelation matrices of the squares of the process are derived only for the symmetric case. However, skewness is frequently observed in stock return distributions, so that results for the more general specification are highly desirable. Moreover, to the best of our knowledge, no results on the fourth–moment structure of multivariate GARCH models with leverage effects exist in the literature so far. Finally, concerning the application of our model, we consider the bivariate volatility dynamics of the Dow Jones Industrial Average (DJIA) and NASDAQ indices, including the computation and backtesting of out–of–sample measures of Value–at–Risk.

The paper is organized as follows. In Section 2, we define the model and discuss estimation issues and theoretical properties, such as the existence of unconditional moments and the dynamic autocorrelation structure of the squared process. Section 3 provides an application to a bivariate stock return series, along with the computation and backtesting of out–of–sample Value–at–Risk measures. Section 4 concludes and identifies issues for further research, where we focus on possible remedies for the curse of dimensionality that will emerge in applications to time series of high dimension. Technical details are gathered in a set of appendices.

2 The Model and its Properties

In this section, we define the multivariate normal mixture GARCH process, discuss estimation issues, and present some theoretical properties.

2.1 Finite Mixtures of Multivariate Normal Distributions

An M–dimensional random vector X is said to have a k–component multivariate finite normal mixture distribution, or, in short, MNM(k), if its density is given by

$$f(x) = \sum_{j=1}^{k} \lambda_j \phi(x; \mu_j, H_j), \tag{1}$$

where $\lambda_j > 0$, $j = 1, \ldots, k$, $\sum_j \lambda_j = 1$, are the *mixing weights*, and

$$\phi(x; \mu_j, H_j) = \frac{1}{(2\pi)^{M/2}\sqrt{|H_j|}} \exp\left\{-\frac{1}{2}(x-\mu_j)' H_j^{-1}(x-\mu_j)\right\}, \quad j = 1, \ldots, k, \tag{2}$$

are the *component densities*. The normal mixture random vector has finite moments of all orders, with expected value and covariance matrix given by (see, e.g., McLachlan and Peel, 2000)

$$\mathrm{E}(X) = \sum_{j=1}^{k} \lambda_j \mu_j, \tag{3}$$

and

$$\mathrm{Cov}(X) = \sum_{j=1}^{k} \lambda_j H_j + \sum_{j=1}^{k} \lambda_j (\mu_j - \mathrm{E}(X))(\mu_j - \mathrm{E}(X))', \tag{4}$$

respectively. We will also make use of the third and fourth moments of a multivariate normal mixture distribution, which are given in Appendix B.

It is well–known that the class of finite normal mixture distributions exhibits an enormous flexibility with respect to distributional shape. For example, for univariate mixtures, Bertholon et al. (2006) show that even the class of two–component normal mixtures spans the feasible set of skewness–kurtosis combinations, $\mathcal{D} = \{(\gamma, \kappa) : \kappa \geq \gamma^2 + 1\}$, where γ and κ are the usual moment–based measures of skewness and kurtosis, respectively, i.e., $\gamma = m_3/m_2^{3/2}$, and $\kappa = m_4/m_2^2$, where m_i, $i = 2, 3, 4$, denotes the ith central moment of any random variable with finite fourth moment (cf. Wilkins, 1944). See also Cohen (1967) for related results in the context of estimation by the method of moments. This illustrates the capability of the normal mixture to capture a broad range of distributional shapes, although a note of caution is always in order when interpreting the widely used moment–based measures γ and κ as indicators of shape.

A question that naturally arises in the estimation of mixture distributions is identifiability. Obviously, a lack of identification always arises as a consequence of label switching, but this can be ruled out by restricting the parameter space such that no duplication appears, e.g., by imposing $\lambda_1 > \lambda_2 > \cdots > \lambda_k$. However, there is a more fundamental problem when the class of density functions to be mixed is linearly dependent (Yakowitz and Spragins, 1968). Fortunately, the class of multivariate finite normal mixtures is identifiable, as has been shown by Yakowitz and Spragins (1968), who generalized Teicher's (1963) result for univariate finite normal mixtures.

An issue which has not been satisfactorily resolved so far is the empirical determination of the number of mixture components, i.e., the choice of k in (1). It is well–known that standard test theory breaks down in this context (McLachlan and Peel, 2000). However, there is some evidence that, at least for unconditional mixture models, the Bayesian information criterion (BIC) of Schwarz (1978) provides a reasonably good indication for the number of components (see McLachlan and Peel, 2000, Ch. 6, for a survey and further references). According to Kass and Raftery (1995), a BIC difference of less than two corresponds to "not worth more than a bare mention", while differences between two and six imply positive evidence, differences between six and ten give rise to strong evidence, and differences greater than ten invoke very strong evidence. However, in the context of multivariate dynamic mixture models, for reasons

of parsimony, it will usually be reasonable to a priori restrict the number of components to be rather small, e.g., $k = 2$ in (1).

2.2 Multivariate Normal Mixture GARCH Processes

The M–dimensional time series $\{\epsilon_t\}$ is said to be generated by a k–component multivariate normal mixture GARCH(p, q) process, or, in short, MNM(k)–GARCH(p, q), if its *conditional* distribution is a k–component multivariate normal mixture (1)–(2), denoted as

$$\epsilon_t|\Psi_{t-1} \sim \text{MNM}(\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_k, H_{1t}, \ldots, H_{kt}), \tag{5}$$

where Ψ_t is the information set at time t. By imposing $\mu_k = -\sum_{j=1}^{k-1}(\lambda_j/\lambda_k)\mu_j$ on the mean of the kth component it is, by (3), guaranteed that ϵ_t in (5) has zero mean. Furthermore, stack the $N := M(M+1)/2$ independent elements of the covariance matrices and the "squared" ϵ_t (i.e., $\epsilon_t\epsilon_t'$) in $h_{jt} := \text{vech}(H_{jt})$, $j = 1, \ldots, k$, and $\eta_t := \text{vech}(\epsilon_t\epsilon_t')$, respectively. Then the component covariance matrices evolve according to

$$h_{jt} = A_{0j} + \sum_{i=1}^{q} A_{ij}\tilde{\eta}_{ij,t-i} + \sum_{i=1}^{p} B_{ij}h_{j,t-i}, \quad j = 1, \ldots, k, \tag{6}$$

where $\tilde{\eta}_{ij,t} = \text{vech}[(\epsilon_t - \theta_{ij})(\epsilon_t - \theta_{ij})']$; θ_{ij}, $i = 1, \ldots, q$, and A_{0j} are columns of length M and N, respectively; and A_{ij}, $i = 1, \ldots, q$, and B_{ij}, $i = 1, \ldots, p$, are $N \times N$ matrices, $j = 1, \ldots, k$. The θ_{ij}'s are introduced in order to allow for the leverage effect in applications to stock market returns, i.e., the strong negative correlation between equity returns and future volatility. In the univariate GARCH literature, various specifications of the leverage effect exist; see, e.g., Ané (2006) and Broto and Ruiz (2006) for recent investigations of such models. The specification in (6) can be viewed as a multivariate generalization of one of the earliest versions, namely Engle's (1990) asymmetric GARCH (AGARCH) model. In the univariate framework, this model has been coupled with the normal mixture GARCH structure by Alexander and Lazar (2005), who demonstrate, in an application to European stock indices, its superior fit when compared to the normal mixture GARCH process with symmetric variance dynamics. We will denote the asymmetric MNM(k)–GARCH(p, q) as MNM(k)–AGARCH(p, q). We also note that, for $p = q = 1$, Engle's (1990) specification coincides with the quadratic GARCH (QGARCH) model of Sentana (1995), so that, in this case, specification (6) can also be interpreted as a MNM(k)–QGARCH$(1, 1)$ model. Finally, in some applications, a symmetric conditional density will be appropriate, so that, in (5), $\mu_1 = \cdots = \mu_k = 0$. We will denote this restricted symmetric version as MNM$_S(k)$–(A)GARCH(p, q). An overview of the different model specifications is provided in Table 1.

Table 1: Variants of MNM–GARCH models.

Model	Conditional Density	Leverage Effect
$\text{MNM}_S(k)$–GARCH(p,q)	symmetric	no
$\text{MNM}_S(k)$–AGARCH(p,q)	symmetric	yes
MNM(k)–GARCH(p,q)	possibly asymmetric	no
MNM(k)–AGARCH(p,q)	possibly asymmetric	yes

A symmetric conditional density is enforced by restricting the component means in (5) to zero, i.e., $\mu_1 = \cdots = \mu_k = 0$. The absence of a leverage effect is imposed by restricting the θ_{ij}'s in (6) to zero, i.e., $\theta_{ij} = 0$, $j = 1, \ldots, k$, $i = 1, \ldots, q$.

To compactify the notation and facilitate the theoretical analysis of the model, note that, by (A.3) in Appendix A, $\text{vech}(\epsilon_{t-i}\theta_{ij}' + \theta_{ij}\epsilon_{t-i}') = 2D_M^+\text{vec}(\theta_{ij}\epsilon_{t-i}') = 2D_M^+(I_M \otimes \theta_{ij})\epsilon_{t-i}$. Then we rewrite (6) as

$$h_{jt} = \tilde{A}_{0j} + \sum_{i=1}^{q} A_{ij}\eta_{t-i} - \sum_{i=1}^{q} \Theta_{ij}\epsilon_{t-i} + \sum_{i=1}^{p} B_{ij}h_{j,t-i}, \quad j = 1, \ldots, k, \tag{7}$$

where $\tilde{A}_{0j} := A_{0j} + \sum_{i=1}^{q} A_{ij}\text{vech}(\theta_{ij}\theta_{ij}')$, and $\Theta_{ij} := 2A_{ij}D_M^+(I_M \otimes \theta_{ij})$, $j = 1, \ldots, k$, $i = 1, \ldots, q$. Let $h_t := (h_{1t}', \ldots, h_{kt}')'$; $\tilde{A}_0 := (\tilde{A}_{01}', \ldots, \tilde{A}_{0k}')'$; $\Theta_i := (\Theta_{i1}', \ldots, \Theta_{ik}')'$, $A_i := (A_{i1}', \ldots, A_{ik}')'$, $i = 1, \ldots, q$; and $B_i := \bigoplus_{j=1}^{k} B_{ij}$, $i = 1, \ldots, p$, where $\bigoplus$ denotes the matrix direct sum. Using these definitions, we have

$$h_t = \tilde{A}_0 + \sum_{i=1}^{q} A_i\eta_{t-i} - \sum_{i=1}^{q} \Theta_i\epsilon_{t-i} + \sum_{i=1}^{p} B_i h_{t-i}. \tag{8}$$

For estimation purposes, the general formulation as given in (6) is not directly applicable, and parameter constraints are required in order to guarantee positive definiteness of all conditional covariances matrices. A particular restriction of the vech form (6) of the multivariate GARCH process serving this purpose is implied by the BEKK model of Engle and Kroner (1995) which specifies the covariance matrices as

$$H_{jt} = A_{0j}^\star A_{0j}^{\star\prime} + \sum_{\ell=1}^{L}\sum_{i=1}^{q} A_{ij,\ell}^\star(\epsilon_{t-i} - \theta_{ij})(\epsilon_{t-i} - \theta_{ij})'A_{ij,\ell}^{\star\prime} + \sum_{\ell=1}^{L}\sum_{i=1}^{p} B_{ij,\ell}^\star H_{j,t-i}B_{ij,\ell}^{\star\prime}, \quad j = 1, \ldots, k, \tag{9}$$

where $A_{0j}^\star$, $j = 1, \ldots, k$, are lower triangular matrices. As shown by Engle and Kroner (1995), each BEKK model implies a unique vech representation (the converse is not true), and, once a BEKK representation (9) is estimated, the matrices A_{ij} and B_{ij} of the vech model (6) can be recovered via

$$A_{ij} = \sum_{\ell=1}^{L} D_M^+(A_{ij,\ell}^\star \otimes A_{ij,\ell}^\star)D_M, \quad i = 1, \ldots, q, \quad j = 1, \ldots, k, \tag{10}$$

and analogously for the B_{ij}, where D_M and D_M^+ denote the duplication matrix and its Moore–Penrose inverse, respectively, both of which we briefly review in Appendix A. Thus, all results derived for the vech model are also applicable to the BEKK model. In practical applications, $L = 1$ is the standard choice, as well as $p = q = 1$. For this specification, it follows from Proposition 2.1 of Engle and Kroner (1995) that the model is identified if the diagonal elements of $A_{0j}^\star$, as well as the top left elements of matrices $A_{1j}^\star$ and $B_{1j}^\star$, $j = 1, \ldots, k$, are restricted to be positive. In addition, while, for $L = 1$, the BEKK model already involves fewer parameters than the unrestricted vech form, further simplifications can be obtained by imposing that $A_{ij}^\star$ and $B_{ij}^\star$, $j = 1, \ldots, k$, are diagonal matrices, giving rise to a diagonal BEKK specification. The latter parametrization is parsimonious enough to be applicable to a relatively large number of assets, and, as noted by Bauwens et al. (2006), although diagonal BEKK models are, due to the inherent restrictions on the cross dynamics, not suitable if volatility transmission is the object under study, "they usually do a good job in representing the dynamics of variances and covariances." Moreover, in the last paragraph of Section 2.3, we will make precise a statement of Bauwens et al. (2007), namely, that "an advantage of the mixture model is that in high dimensions, simple models with few parameters could be mixed to obtain more flexibility than specifying a complex one–component model". In applications to very high–dimensional time series, however, even the diagonal BEKK model for the component covariance matrices will be too heavily parameterized, and techniques for dimensionality reduction, such as the use of factor structures, will be called for; see Section 4 for a brief discussion of these issues and possible starting points for further research in this direction. In the following discussion of the vech specification we will always assume that positive definite covariances matrices are guaranteed, without further specifying the constraints employed for achieving this.

2.3 Existence of Moments and Autocorrelation Structure

It is clear that, for practical purposes, the most important MNM(k)–AGARCH(p, q) process is the specification where $p = q = 1$, which is defined by (5) and

$$h_t = \tilde{A}_0 + A_1 \eta_{t-1} - \Theta_1 \epsilon_{t-1} + B_1 h_{t-1}. \tag{11}$$

For later reference, we summarize the dynamic properties of the process given by (5) and (11) in Proposition 1. The corresponding results for the MNM(k)–GARCH(p, q) specification, which are of less relevance for the applications, are provided in an earlier version of this paper (Haas et al., 2006b).

We denote as $\rho(A)$ the largest eigenvalue in modulus of a square matrix A, i.e.,

$$\rho(A) := \max\{|z| : z \text{ is an eigenvalue of } A\}, \tag{12}$$

and define the vector of mixing weights $\lambda := (\lambda_1, \ldots, \lambda_k)'$. Following the classic papers of Engle (1982) and Bollerslev (1986), we assume for simplicity that the process starts indefinitely far in the past with finite fourth moments.

Proposition 1 *The MNM(k)–AGARCH(1,1) process given by (5) and (11) is covariance stationary if and only if $\rho(C_{11}) < 1$, where the $kN \times kN$ matrix C_{11} is defined by*

$$C_{11} = \lambda' \otimes A_1 + B_1. \tag{13}$$

Moreover, the unconditional fourth moment $E(\eta_t \eta_t')$ exists if and only if, in addition, $\rho(C_{22}) < 1$, where C_{22} is the $(kN)^2 \times (kN)^2$ matrix given by

$$C_{22} = (A_1 \otimes A_1) G_M (I_N \otimes vec(\Lambda)' \otimes I_N)(K_{Nk} \otimes I_{kN}) + 2\tilde{N}_{kN}(B_1 \otimes \lambda' \otimes A_1) + B_1 \otimes B_1. \tag{14}$$

In (14), G_M is the $N^2 \times N^2$ matrix defined in (B.13) in Appendix B.2, $\Lambda = diag(\lambda_1 \ldots, \lambda_k)$, K_{mn} is the commutation matrix defined in Appendix A, and $\tilde{N}_n = (I_{n^2} + K_{nn})/2$. The unconditional covariance matrix follows from (4) and expression (C.22) in Appendix C.1, and the fourth–moment matrix can be obtained from expressions (B.15) and (C.23) in Appendices B.2 and C.1, respectively.

If $\rho(C_{22}) < 1$ holds, the multidimensional autocovariance function of the squared process, $\Gamma(\tau) := E(\eta_t \eta_{t-\tau}') - E(\eta_t)E(\eta_t)'$, is given by

$$\Gamma(\tau) = (\lambda' \otimes I_N) C_{11}^{\tau - 1} Q, \quad \tau \geq 1, \tag{15}$$

where Q is a constant matrix given in (C.24) in Appendix C.2.

Note that $(I_N \otimes \mathrm{vec}(\Lambda)' \otimes I_N)(K_{Nk} \otimes I_{kN})$ in (14) is the explicit expression for the matrix $\tilde{\Lambda} P_{kN^\star}$ defined only implicitly in Theorem 2 of Bauwens et al. (2007). This makes the fourth–moment condition more practicable. Also note that, analogously to Sentana's (1995) results for the QGARCH(1,1) model, the leverage parameters do not affect the second– and fourth–moment conditions. The results of Proposition 1 are derived in Appendices B and C. From (15), the autocorrelation matrices, R_τ, can be calculated in the usual way. I.e., if $D = I_N \odot \Gamma(0)$, where $\Gamma(0) = \mathrm{E}(\eta_t \eta_t') - \mathrm{E}(\eta_t)\mathrm{E}(\eta_t)'$, then

$$R(\tau) = D^{-1/2} \Gamma(\tau) D^{-1/2}. \tag{16}$$

The term determining the rate of decay of $\Gamma(\tau)$ is C_{11}^τ. Thus, under covariance stationarity, the largest eigenvalue in magnitude of the matrix C_{11} defined in (13) can be used as a measure for the persistence of shocks to volatility. Furthermore, the stationarity condition $\rho(C_{11}) < 1$ allows some components to be nonstationary, in the sense that the covariance stationarity

condition for single–component multivariate GARCH(1,1) processes, i.e., $\rho(A_{1j} + B_{1j}) < 1$ (Bollerslev and Engle, 1993), is not satisfied for some components. Nevertheless, the overall process can still be stationary, as long as the corresponding mixing weights are sufficiently small. This has also been noted by Bauwens et al. (2007) and parallels the situation in the univariate case (see Haas et al., 2004; and Alexander and Lazar, 2006).

As mentioned at the end of Section 2.2, in applications to a large number of assets, the diagonal BEKK model, which implies a restricted diagonal vech model, provides a parsimonious parametrization for the dynamics of variances and covariances. It is worthwhile to point out that this specification, when enriched with a normal mixture GARCH structure, can generate much more complex dynamics of the second moments than those achievable by the corresponding single–component GARCH(1,1) model. To illustrate, consider the *diagonal* two–component MNM–GARCH(1,1) model with $\mu_1 = \mu_2 = 0_{M\times 1}$, and $\Theta_1 = 0_{2N\times M}$. Then we have $h_{jt} = A_{0j} + A_{1j}\eta_{t-1} + B_{1j}h_{j,t-1}$, and, provided that $\max\{\rho(B_{11}), \rho(B_{12})\} < 1$, $h_{jt} = (I_N - B_{1j})^{-1}A_{0j} + (I_N - B_{1j}L)^{-1}A_{1j}\eta_{t-1}$, $j = 1, 2$, where L is the lag operator, i.e., $L^\tau x_t = x_{t-\tau}$. Therefore, from (4) and the diagonality of matrices A_{1j} and B_{1j}, $j = 1, 2$, the dynamics of $\text{vech}[\text{cov}(\epsilon_t|\Psi_{t-1})] = \lambda_1 h_{1t} + \lambda_2 h_{2t} =: \overline{h}_t$ are described by

$$\begin{aligned}
\overline{h}_t &= \lambda_1(I_N - B_{11})^{-1}A_{01} + \lambda_2(I_N - B_{12})^{-1}A_{02} \\
&\quad +[\lambda_1(I_N - B_{11}L)^{-1}A_{11} + \lambda_2(I_N - B_{12}L)^{-1}A_{12}]\eta_{t-1} \\
&= \lambda_1(I_N - B_{11})^{-1}A_{01} + \lambda_2(I_N - B_{12})^{-1}A_{02} \\
&\quad +(I_N - B_{11}L)^{-1}(I_N - B_{12}L)^{-1}[\lambda_1(I_N - B_{12}L)A_{11} + \lambda_2(I_N - B_{11}L)A_{12}]\eta_{t-1},
\end{aligned} \tag{17}$$

which implies a GARCH(2,2) structure for the conditional covariance matrix, i.e.,

$$\begin{aligned}
\overline{h}_t = \overline{A}_0 &+ (\lambda_1 A_{11} + \lambda_2 A_{12})\eta_{t-1} - (\lambda_1 B_{12}A_{11} + \lambda_2 B_{11}A_{12})\eta_{t-2} \\
&+ (B_{11} + B_{12})\overline{h}_{t-1} - B_{11}B_{12}\overline{h}_{t-2},
\end{aligned} \tag{18}$$

where $\overline{A}_0 = \lambda_1(I_N - B_{12})A_{01} + \lambda_2(I_N - B_{11})A_{02}$. Thus, and in sharp contrast to the single–component model, even if the parameter matrices A_{1j} and B_{1j}, $j = 1, 2$, are diagonal, as may be required in high–dimensional problems, the overall conditional variances and covariances in $\overline{h}_t$ will have a (restricted) GARCH(2,2) structure, allowing for a rich set of possible autocorrelation structures of the squared process. In particular, Equation (18) bears some resemblance to the GARCH(2,2) representation of the (univariate) component GARCH model of Ding and Granger (1996), which often captures the autocorrelation structure of squared returns much better than the GARCH(1,1) specification (see, e.g., Maheu, 2005; Bauwens and Storti, 2007; and Haas, 2007). The reasoning above can easily be generalized to the diagonal

MNM–GARCH(1,1) process with k components, resulting in a GARCH(k,k) structure for the overall covariance matrix, $\overline{h}_t$. We finally note that (17) and (18) are not generally valid for models with nondiagonal parameter matrices. However, in this case, $\overline{h}_t$ has the ARCH(∞) representation

$$\overline{h}_t = \lambda_1(I_N - B_{11})^{-1}A_{01} + \lambda_2(I_N - B_{12})^{-1}A_{02} + \sum_{i=1}^{\infty}(\lambda_1 B_{11}^{i-1}A_{11} + \lambda_2 B_{12}^{i-1}A_{12})\eta_{t-i}, \quad (19)$$

which is still evocative of the corresponding representation of the conditional variance in Ding and Granger's (1996) model, as given in Equation (4.7) of their paper. In addition, by taking unconditional expectations on both sides of (19), this ARCH(∞) representation can be used to obtain an explicit expression for $\mathrm{E}(\overline{h}_t)$ in terms of the original model parameters, which may, as suggested by a referee, be used for covariance targeting, so that the model–implied unconditional covariance matrix matches its sample analogue.

3 Application to Stock Market Returns

We investigate the bivariate time series of daily returns of the Dow Jones Industrial Average (DJIA) and the NASDAQ indices from January 1990 to September 2007, a sample of $T = 4,474$ observations. The data were obtained from Yahoo Finance. Continuously compounded percentage returns are considered, i.e., $r_{it} = 100 \times \log(P_{it}/P_{i,t-1})$, $i = 1, 2$, where P_{it} denotes the level of index i at time t. We denote the return vector at time t by $r_t = (r_{1t}, r_{2t})'$, where r_{1t} and r_{2t} are the time–t returns of the DJIA and the NASDAQ, respectively.

We first estimate the model over the first ten years of data, i.e., over the period from 1990–1999, accounting for the first 2,527 observations. The remaining observations are retained for computation and backtesting of out–of–sample Value–at–Risk measures. The return series are shown in the top panel of Figure 1, and a few descriptive statistics for the in–sample period are provided in Table 2. To specify the mean equation, we calculate the sample autocorrelation (SACF) and sample partial autocorrelation functions (SPACF) over the in–sample period, as shown in the middle and bottom panels of Figure 1. While there are no significant first–order dependencies in the returns of the DJIA, both the SACF and SPACF of the NASDAQ are significant at lag one and cut off after the first lag, which does not correspond to any standard textbook pattern. However, the residuals from a first–order autoregression of the NASDAQ returns fail to exhibit any significant spikes, and, therefore, we model returns as

$$r_t = \nu + F r_{t-1} + \epsilon_t, \quad \text{where } F = \begin{pmatrix} 0 & 0 \\ 0 & f_{22} \end{pmatrix}, \quad (20)$$

Table 2: Descriptive statistics of DJIA/NASDAQ returns over the in–sample period, 1990–1999.

	mean	covariance/ correlation matrix		skewness	kurtosis	JB
		DJIA	NASDAQ			
DJIA	0.056	0.795	0.728	–0.410	8.201	2919.2 (0.000)
NASDAQ	0.086	0.723	1.241	–0.540	7.692	2441.2 (0.000)

The top right entry of the "covariance/correlation matrix" is the correlation coefficient, and the bottom left entry is the covariance. "skewness" denotes the moment–based coefficient of skewness, $\gamma = m_3/m_2^{3/2}$, and "kurtosis" the moment–based coefficient of kurtosis, $\kappa = m_4/m_2^2$, where $m_i = T^{-1}\sum_t (r_t - \bar{r})^i$, $i = 2, 3, 4$, and $\bar{r} = T^{-1}\sum_t r_t$. JB is the Jarque–Bera test for normality, based on the result that, under normality, $\text{JB} = T\gamma^2/6 + T(\kappa - 3)^2/24 \overset{asy}{\sim} \chi^2(2)$. p–values are given in parentheses.

ν is a 2×1 vector of constants, and ϵ_t follows a GARCH process in BEKK form as given by (9), with $p = q = L = 1$. All parameters are estimated simultaneously by maximum likelihood.

3.1 Estimation Results

Several versions of the general mixture GARCH model (5)–(6) with $p = q = 1$ have been estimated. Namely, the single–component model, which corresponds to $k = 1$ in (1), and which is just the standard Normal–GARCH process, has been estimated with and without imposing a symmetric reaction to negative and positive shocks. The first of these models, where $\theta_{11} = 0$ in (6), will be denoted by Normal–GARCH(1,1), and the second by Normal–AGARCH(1,1). Also, two–component models are considered with and without symmetric conditional mixture densities, i.e., with and without imposing $\mu_1 = \mu_2 = 0$ in (5), as well as with and without leverage effects. To refer to these different models, we will use the typology of Table 1.

Table 3 reports likelihood–based goodness–of–fit measures for the models and their rankings with respect to each of these criteria, i.e., the value of the maximized log–likelihood function, and the AIC and BIC criteria of Akaike (1973) and Schwarz (1978), respectively. While it is not surprising that the Normal–GARCH model is the worst performer with respect to each of these criteria, several additional observations are worth mentioning. First, the normal mixture specifications allowing for asymmetric conditional densities, i.e., admitting nonzero component means in (5), are always favored against their symmetric counterparts. This is not the case when we consider the dynamic asymmetry, i.e., leverage effects. The improvement in log–likelihood is much larger when passing from the symmetric $\text{MNM}_S(2)$–GARCH(1,1) to the $\text{MNM}_S(2)$–AGARCH(1,1) model (difference in log–likelihood: 23.7) than when passing from

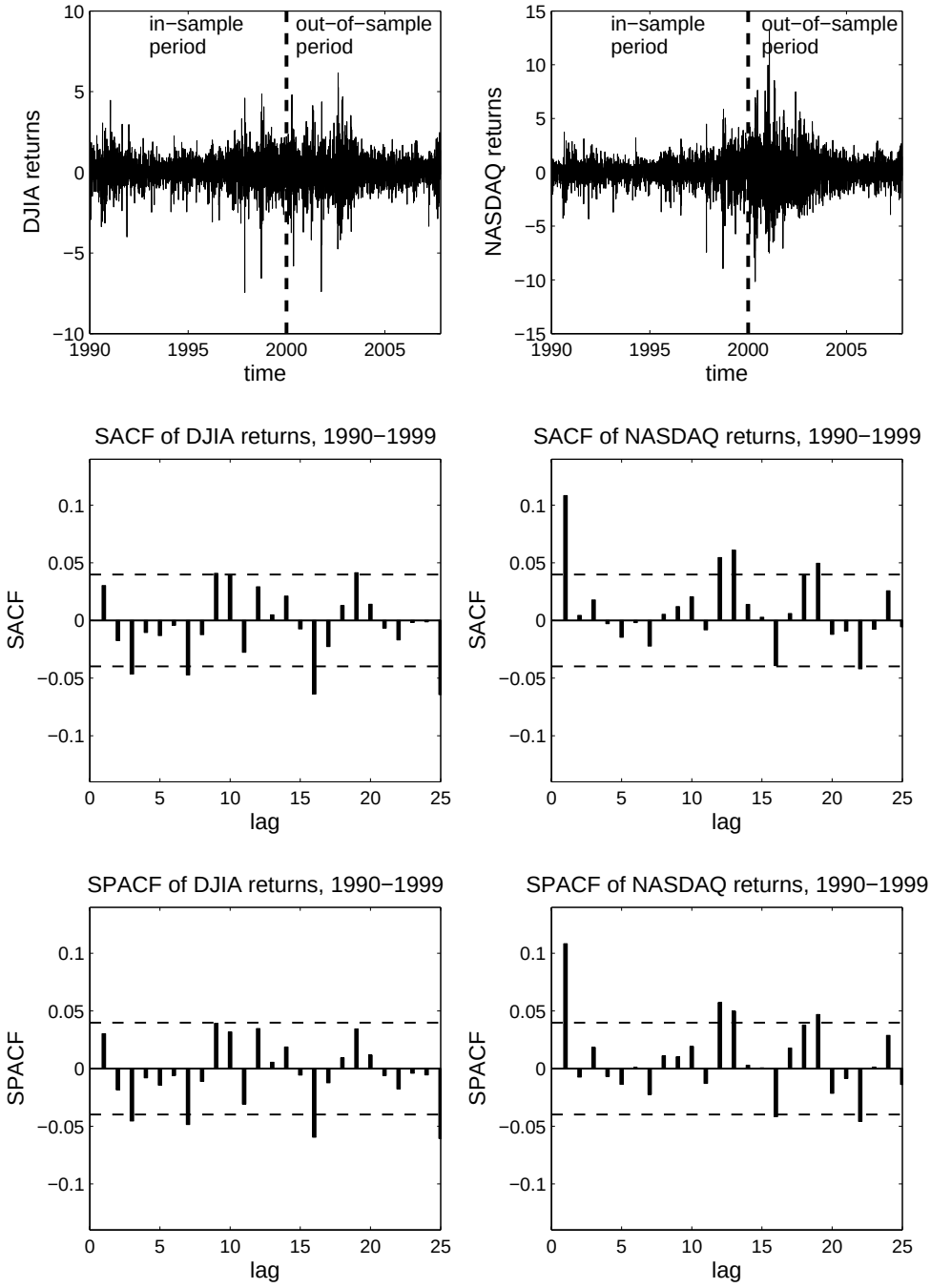

Figure 1: The top panel shows the percentage returns of the DJIA (left) and the NASDAQ (right). The middle and bottom panels show the sample autocorrelation (SACF) and partial autocorrelation functions (SPACF) over the period from 1990 to 1999 (in–sample period), respectively. Dashed lines represent approximate 95% one–at–a–time confidence intervals.

Table 3: Likelihood–based goodness of fit.

Distributional		L		AIC		BIC	
Model	K	Value	Rank	Value	Rank	Value	Rank
Normal–GARCH(1,1)	14	–5606.9	6	11241.8	6	11323.5	6
$\text{MNM}_S(2)$–GARCH(1,1)	26	–5504.2	4	11060.5	4	11212.2	4
MNM(2)–GARCH(1,1)	28	–5482.6	3	11021.3	3	**11184.7**	**1**
Normal–AGARCH(1,1)	16	–5592.5	5	11217.0	5	11310.4	5
$\text{MNM}_S(2)$–AGARCH(1,1)	30	–5480.5	2	11021.0	2	11196.1	3
MNM(2)–AGARCH(1,1)	32	**−5467.6**	**1**	**10999.2**	**1**	11185.9	2

The leftmost column states the type of volatility model fitted to the bivariate NASDAQ/DJIA returns. The column labeled K reports the number of parameters of a model (including the mean equation); L is the log–likelihood; $\text{AIC} = -2L + 2K$; and $\text{BIC} = -2L + K \log T$, where T is the number of observations. For each of the three criteria the criterion value and the ranking of the models are shown. Boldface entries indicate the best model for the particular criterion.

the asymmetric MNM(2)–GARCH(1,1) process to its AGARCH(1,1) counterpart (difference in log–likelihood: 15.1). As a consequence, the MNM(2)–GARCH(1,1) specification performs best overall according to the BIC. We note, however, that the difference in BIC for the latter two models is insignificant according to the Kass and Raftery–recommendation mentioned at the end of Section 2.1. Also, a closer inspection of the parameter estimates will reveal that the leverage effect may be an exclusive feature of the high–volatility component, so that the difference in the number of parameters between these models shrinks from four to two, which would reverse the models' ranking. Moreover, a likelihood ratio test for $\theta_1 = \theta_2 = 0$, with associated test statistic $\text{LRT} = 2 \times (5482.6 - 5467.6) = 30.1$, would reject at conventional critical values given by the asymptotically valid χ^2 distribution with four degrees of freedom, thus favoring the model with leverage effects.

The maximum likelihood estimates (MLEs) are reported in Tables 4 and 5 for the models without and with leverage effects, respectively. The function `fminunc` in Matlab (version 6.5) was used to find the MLEs. We did not encounter convergence problems, and the estimates were robust with respect to different sets of starting values. As our focus is on volatility dynamics, the parameters of the mean equation (20) are not reported. Shown are the parameter matrices $A_{0j}^{\star}$, $A_{1j}^{\star}$, and $B_{1j}^{\star}$, $j = 1, 2$, of the BEKK representation (9). In addition, we report the component–specific persistence measures, i.e., the largest eigenvalues of the matrices $A_{1j} + B_{1j}$, $j = 1, 2$, where these matrices have been recovered from the BEKK representation using (10), as well as the largest eigenvalues of the matrices C_{11} and C_{22} defined in Proposition 1. The two–component models have been ordered such that $\lambda_1 > \lambda_2$. Furthermore, the implied unconditional overall and component–specific covariance matrices and their associated correlation coefficients are shown in Table 6.

Table 4: MNM–GARCH(1,1) parameter estimates for DJIA/NASDAQ returns

	Normal–GARCH(1,1)	$\text{MNM}_S(2)$–GARCH(1,1)	MNM(2)–GARCH(1,1)
$A^\star_{01}$	$\begin{pmatrix} 0.055 & 0 \\ (0.021) & \\ 0.114 & 0.070 \\ (0.027) & (0.022) \end{pmatrix}$	$\begin{pmatrix} 0.007 & 0 \\ (0.020) & \\ -0.016 & 0 \\ (0.035) & (-) \end{pmatrix}$	$\begin{pmatrix} 0.006 & 0 \\ (0.019) & \\ -0.022 & 0 \\ (0.026) & (-) \end{pmatrix}$
$A^\star_{11}$	$\begin{pmatrix} 0.100 & 0.090 \\ (0.021) & (0.018) \\ -0.137 & 0.385 \\ (0.034) & (0.033) \end{pmatrix}$	$\begin{pmatrix} 0.060 & 0.060 \\ (0.020) & (0.015) \\ -0.139 & 0.290 \\ (0.031) & (0.030) \end{pmatrix}$	$\begin{pmatrix} 0.074 & 0.055 \\ (0.019) & (0.015) \\ -0.111 & 0.270 \\ (0.029) & (0.026) \end{pmatrix}$
$B^\star_{11}$	$\begin{pmatrix} 1.006 & -0.030 \\ (0.006) & (0.006) \\ 0.044 & 0.916 \\ (0.012) & (0.014) \end{pmatrix}$	$\begin{pmatrix} 1.002 & -0.017 \\ (0.003) & (0.004) \\ 0.025 & 0.954 \\ (0.006) & (0.008) \end{pmatrix}$	$\begin{pmatrix} 1.000 & -0.016 \\ (0.003) & (0.004) \\ 0.022 & 0.956 \\ (0.006) & (0.008) \end{pmatrix}$
$\rho(A_{11}+B_{11})$	0.997	0.995	0.994
θ_{11}	–	–	–
λ_1	1	$\underset{(0.041)}{0.817}$	$\underset{(0.031)}{0.835}$
μ_1	–	–	$\underset{(0.016)}{0.053}, \underset{(0.019)}{0.111}\Big)'$
$A^\star_{02}$	–	$\begin{pmatrix} 0.340 & 0 \\ (0.111) & \\ 0.484 & 0.244 \\ (0.119) & (0.108) \end{pmatrix}$	$\begin{pmatrix} 0.393 & 0 \\ (0.086) & \\ 0.449 & 0 \\ (0.081) & (-) \end{pmatrix}$
$A^\star_{12}$	–	$\begin{pmatrix} 0.353 & 0.200 \\ (0.121) & (0.117) \\ 0.043 & 0.711 \\ (0.158) & (0.151) \end{pmatrix}$	$\begin{pmatrix} 0.370 & 0.198 \\ (0.117) & (0.115) \\ -0.011 & 0.736 \\ (0.142) & (0.145) \end{pmatrix}$
$B^\star_{12}$	–	$\begin{pmatrix} 0.975 & -0.080 \\ (0.078) & (0.077) \\ 0.082 & 0.728 \\ (0.134) & (0.129) \end{pmatrix}$	$\begin{pmatrix} 0.915 & -0.033 \\ (0.058) & (0.057) \\ -0.032 & 0.830 \\ (0.072) & (0.076) \end{pmatrix}$
$\rho(A_{12}+B_{12})$	–	1.158	1.185
θ_{12}	–	–	–
λ_2	0	$\underset{(0.041)}{0.183}$	$\underset{(0.031)}{0.165}$
μ_2	–	–	$\underset{(0.091)}{-0.267}, \underset{(0.115)}{-0.563}\Big)'$
$\rho(C_{11})$	0.997	0.995	0.996
$\rho(C_{22})$	0.995	0.995	0.996

Approximate standard errors are given in parentheses. If parameters with nonnegativity restrictions were extremely close to the boundary, we reestimated the model with these parameters set to zero, so that their standard errors are not reported. This applies to the lower diagonal element of $A^\star_{01}$ for model $\text{MNM}_S(2)$–GARCH(1,1), as well as to the lower diagonal elements of $A^\star_{01}$ and $A^\star_{02}$ for model MNM(2)–GARCH(1,1). Note that matrices $A^\star_{0j}$, $A^\star_{1j}$, and $B^\star_{1j}$, $j = 1, 2$, correspond to the BEKK representation (9) of the model, while matrices $A_{1j} + B_{1j}$, $j = 1, 2$, the maximal eigenvalues of which are reported, are associated with the vech representation (6). $\rho(C_{11})$ and $\rho(C_{22})$ denote the largest eigenvalues of the matrices C_{11} and C_{22}, defined in Proposition 1, which determine whether the unconditional second and fourth moments, respectively, exist.

Table 5: MNM–AGARCH(1,1) parameter estimates for DJIA/NASDAQ returns

	Normal–AGARCH(1,1)	$\text{MNM}_S(2)$–AGARCH(1,1)	MNM(2)–AGARCH(1,1)
$A_{01}^\star$	$\begin{pmatrix} 0.054 & 0 \\ (0.027) & \\ 0.116 & 0.075 \\ (0.031) & (0.025) \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ (-) & \\ 0 & 0 \\ (-) & (-) \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ (-) & \\ 0 & 0 \\ (-) & (-) \end{pmatrix}$
$A_{11}^\star$	$\begin{pmatrix} 0.111 & 0.095 \\ (0.023) & (0.019) \\ -0.129 & 0.398 \\ (0.036) & (0.036) \end{pmatrix}$	$\begin{pmatrix} 0.061 & 0.056 \\ (0.020) & (0.015) \\ -0.138 & 0.284 \\ (0.031) & (0.029) \end{pmatrix}$	$\begin{pmatrix} 0.069 & 0.051 \\ (0.020) & (0.015) \\ -0.112 & 0.261 \\ (0.028) & (0.026) \end{pmatrix}$
$B_{11}^\star$	$\begin{pmatrix} 1.004 & -0.035 \\ (0.007) & (0.008) \\ 0.044 & 0.906 \\ (0.014) & (0.017) \end{pmatrix}$	$\begin{pmatrix} 1.000 & -0.014 \\ (0.003) & (0.004) \\ 0.023 & 0.958 \\ (0.006) & (0.007) \end{pmatrix}$	$\begin{pmatrix} 0.999 & -0.013 \\ (0.003) & (0.003) \\ 0.017 & 0.962 \\ (0.005) & (0.006) \end{pmatrix}$
$\rho(A_{11}+B_{11})$	0.996	0.997	0.996
θ_{11}	$\underset{(0.080)\ (0.062)}{(0.257, 0.318)}'$	$\underset{(0.099)\ (0.083)}{(-0.163, -0.124)}'$	$\underset{(0.100)\ (0.078)}{(-0.136, -0.098)}'$
λ_1	1	$\underset{(0.036)}{0.754}$	$\underset{(0.033)}{0.763}$
μ_1	–	–	$\underset{(0.018)\ (0.021)}{(0.047, 0.099)}'$
$A_{02}^\star$	–	$\begin{pmatrix} 0.060 & 0 \\ (0.075) & \\ -0.162 & 0 \\ (0.126) & (-) \end{pmatrix}$	$\begin{pmatrix} 0.088 & 0 \\ (0.081) & \\ -0.092 & 0 \\ (0.122) & (-) \end{pmatrix}$
$A_{12}^\star$	–	$\begin{pmatrix} 0.322 & 0.173 \\ (0.072) & (0.082) \\ 0.029 & 0.632 \\ (0.095) & (0.100) \end{pmatrix}$	$\begin{pmatrix} 0.331 & 0.115 \\ (0.076) & (0.081) \\ -0.013 & 0.587 \\ (0.099) & (0.093) \end{pmatrix}$
$B_{12}^\star$	–	$\begin{pmatrix} 0.985 & -0.108 \\ (0.040) & (0.052) \\ 0.101 & 0.679 \\ (0.078) & (0.081) \end{pmatrix}$	$\begin{pmatrix} 0.970 & -0.072 \\ (0.040) & (0.051) \\ 0.076 & 0.737 \\ (0.055) & (0.058) \end{pmatrix}$
$\rho(A_{12}+B_{12})$	–	1.023	1.017
θ_{12}	–	$\underset{(0.157)\ (0.134)}{(0.613, 0.778)}'$	$\underset{(0.164)\ (0.126)}{(0.664, 0.860)}'$
λ_2	0	$\underset{(0.036)}{0.246}$	$\underset{(0.033)}{0.237}$
μ_2	–	–	$\underset{(0.061)\ (0.071)}{(-0.153, -0.321)}'$
$\rho(C_{11})$	0.996	0.995	0.995
$\rho(C_{22})$	0.993	0.991	0.992

Approximate standard errors are given in parentheses. If parameters with nonnegativity restrictions were extremely close to the boundary, we reestimated the model with these parameters set to zero, so that their standard errors are not reported. This applies, for models $\text{MNM}_S(2)$–AGARCH(1,1) and MNM(2)–AGARCH(1,1), to the diagonal elements of $A_{01}^\star$ and to the lower diagonal element of $A_{02}^\star$. Note that, when both diagonal elements of $A_{01}^\star$ are set to zero, the sign of the bottom left element is not identified, and, consequently, given its closeness to the boundary (zero), it was likewise fixed to zero. See the legend of Table 4 for further explanations.

Table 6: Unconditional (component–specific) covariance matrices and implied correlations.

Model	Normal–GARCH(1,1)	$\text{MNM}_S(2)$–GARCH(1,1)	MNM(2)–GARCH(1,1)
$\text{E}(\epsilon_t\epsilon_t')$	$\begin{pmatrix} 0.900 & 0.723 \\ 0.781 & 1.297 \end{pmatrix}$	$\begin{pmatrix} 0.793 & 0.704 \\ 0.686 & 1.197 \end{pmatrix}$	$\begin{pmatrix} 0.782 & 0.696 \\ 0.671 & 1.190 \end{pmatrix}$
$\text{E}(H_{1t})$	–	$\begin{pmatrix} 0.536 & 0.624 \\ 0.435 & 0.907 \end{pmatrix}$	$\begin{pmatrix} 0.549 & 0.624 \\ 0.433 & 0.876 \end{pmatrix}$
$\text{E}(H_{2t})$	–	$\begin{pmatrix} 1.942 & 0.822 \\ 1.808 & 2.490 \end{pmatrix}$	$\begin{pmatrix} 1.877 & 0.800 \\ 1.698 & 2.401 \end{pmatrix}$
Model	Normal–AGARCH(1,1)	$\text{MNM}_S(2)$–AGARCH(1,1)	MNM(2)–AGARCH(1,1)
$\text{E}(\epsilon_t\epsilon_t')$	$\begin{pmatrix} 0.854 & 0.716 \\ 0.734 & 1.230 \end{pmatrix}$	$\begin{pmatrix} 0.701 & 0.689 \\ 0.583 & 1.023 \end{pmatrix}$	$\begin{pmatrix} 0.695 & 0.685 \\ 0.547 & 0.918 \end{pmatrix}$
$\text{E}(H_{1t})$	–	$\begin{pmatrix} 0.470 & 0.607 \\ 0.375 & 0.812 \end{pmatrix}$	$\begin{pmatrix} 0.463 & 0.600 \\ 0.338 & 0.685 \end{pmatrix}$
$\text{E}(H_{2t})$	–	$\begin{pmatrix} 1.408 & 0.796 \\ 1.220 & 1.668 \end{pmatrix}$	$\begin{pmatrix} 1.412 & 0.786 \\ 1.157 & 1.535 \end{pmatrix}$

The table reports the unconditional overall and component–specific covariance matrices of the error term ϵ_t, as implied by the parameter estimates given in Tables 4 and 5. The associated correlation coefficients are shown in upper triangular parts of the respective matrices.

In discussing the parameter estimates, we first draw attention to a common characteristic of all mixture models, irrespective of their allowance for asymmetry and/or leverage: All these models identify two components with distinctly different volatility dynamics. More precisely, the first component, i.e., the component with the larger mixing weight, is stationary in the sense that $\rho(A_{11}+B_{11}) < 1$, and it has less weight on the reaction parameters in A_{11} and more weight on the persistence parameters in B_{11}, relative to the second component. An inspection of Table 6 also shows that Components 1 and 2 can be characterized as low– and high–volatility components, respectively. The latter is nonstationary in the sense that $\rho(A_{12}+B_{12}) > 1$, and it has considerably more weight on the reaction and less on the persistence parameters. This implies that the high–volatility component reacts more strongly to shocks, but has a shorter memory. However, all estimated mixture models are stationary in the aggregate with finite fourth unconditional moments, because, for all models, the largest eigenvalues of the matrices C_{11} and C_{22}, defined in (13) and (14), respectively, are less than unity. Another observation arising from Table 6 is that the correlations are higher in turbulent markets, i.e., in the high–volatility component, a phenomenon that has recently been investigated, among others, by Ang and Chen (2002) and Patton (2004). An informal comparison of Table 6 with columns 3–4 of Table 2 also shows that all models fit the unconditional covariance/correlation structure reasonably well, although the mixture models with a leverage effect do slightly worse in this regard.

If nonzero component means are allowed for, we observe that, both for the MNM(2)–

GARCH(1,1) model in Table 4 and the MNM(2)–AGARCH(1,1) model in Table 5, the low–volatility component is associated with positive means, and the high–volatility component is associated with statistically significant negative means for both indices, implying that the low– and high–volatility components can be interpreted as bull and bear markets, respectively. A similar finding holds for the leverage effects, i.e., the dynamic asymmetries in the GARCH structure, as reported in Table 5. For both mixture AGARCH models, a leverage effect seems to be present mainly in the high–volatility, bear market component. The leverage parameters in the first component, θ_{11}, are negative, and thus seem to indicate a "reverse" leverage effect, but they are also insignificant statistically. On the other hand, the leverage parameters of the nonstationary component, θ_{12}, are rather large, compared to those of the fitted Normal–AGARCH model, indicating a very strong negative relation between current returns and future volatility. It is also interesting to note that the introduction of the leverage effects reduces the persistence measure of the high–volatility component somewhat, i.e., $\rho(A_{12} + B_{12})$ decreases. (Note, however, that the interpretation of $\rho(A_{12} + B_{12})$ as a persistence measure is a little awkward when $\rho(A_{12}+B_{12}) > 1$.) However, at the same time, its mixing weight, λ_2, increases, so that the overall persistence of the model, as measured by $\rho(C_{11})$, remains approximately unchanged.

To assess the models' fit of the unconditional distribution, Figures 2 and 3 present the empirical densities of the residuals for the DJIA and the NASDAQ, respectively, as obtained via kernel density estimation (see, e.g., Silverman, 1986), along with kernel estimates of simulated samples of length 1,000,000 from the estimated models. The kernel estimator is given by $\widehat{f}_i(x) = (Th)^{-1} \sum_{t=1}^{T} K[(x - \epsilon_{it})/h]$, $i =$ DJIA, NASDAQ, where we use a Gaussian kernel, i.e., $K(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}$, and $h = 1.06\widehat{\sigma}_i T^{-1/5}$, where $\widehat{\sigma}_i$ is the respective sample standard deviation. While it is usually difficult to see the fatter tails in such figures, it is apparent that the empirical density is remarkably more peaked than the unconditional distribution implied by the single–regime Normal–GARCH(1,1) process, while the mixture models provide a much closer approximation to the empirical densities. Recall that the leptokurtosis observed in financial time series includes both peakedness and tailedness. In fact, both features reflect the same phenomenon, because, as noted by Ruppert (1987), if one moves probability mass from the shoulders of a distribution to the tails, then to keep the scale fixed one must also move mass from the shoulders to the center.

Finally, Figures 4 and 5 show the empirical autocorrelations of the squared residuals for the two series, along with their theoretical counterparts implied by the six estimated GARCH models. As often observed in the literature since Ding et al. (1993) and Ding and Granger (1996), the empirical autocorrelations decay rapidly at the beginning and then decrease rather

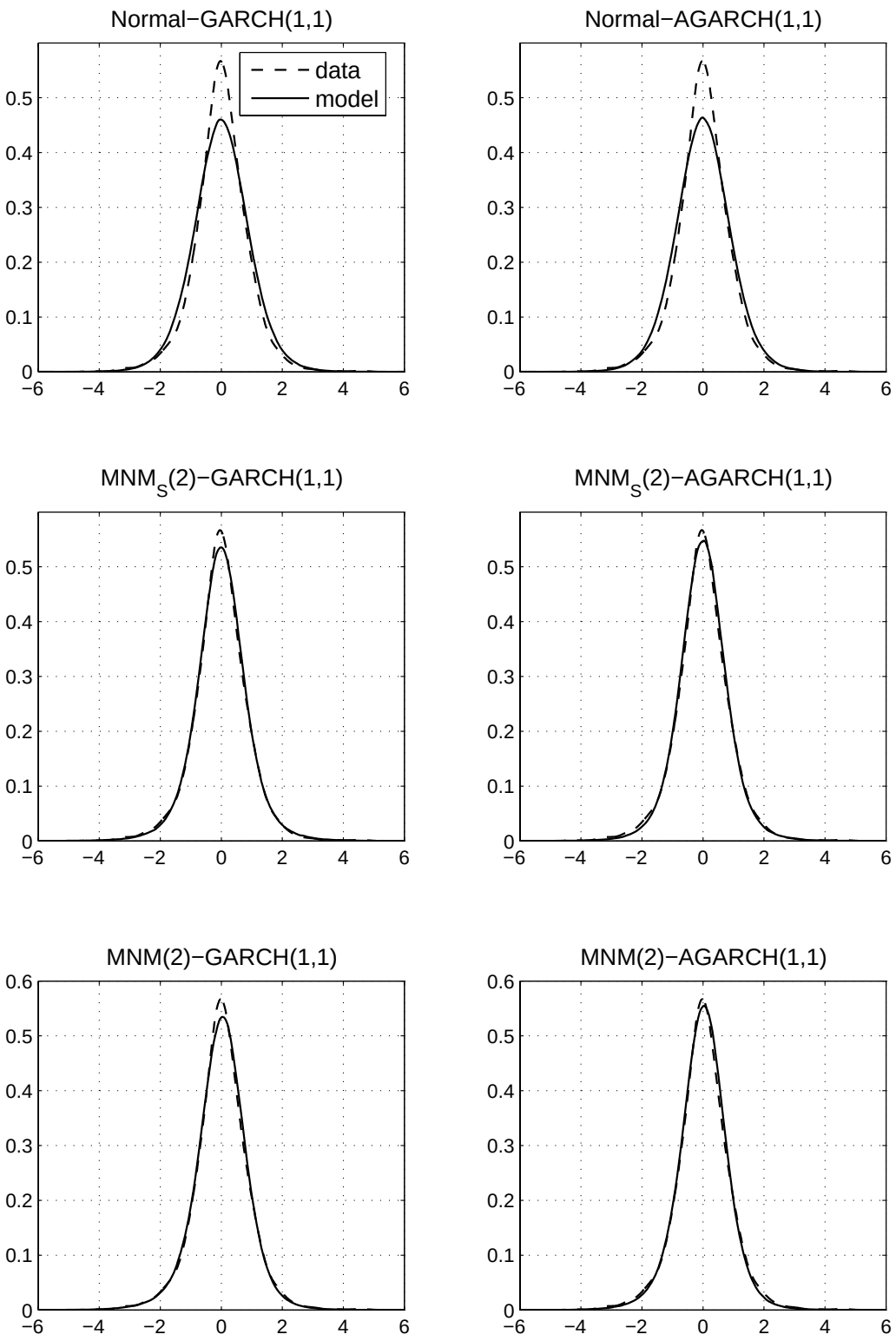

Figure 2: Kernel density estimates for the DJIA errors. Shown are, for each fitted GARCH model, kernel density estimates for the estimated empirical DJIA errors in (20) (dashed line), along with kernel estimates for the error distributions implied by the respective models (solid line), as obtained from simulated samples of length 1,000,000.

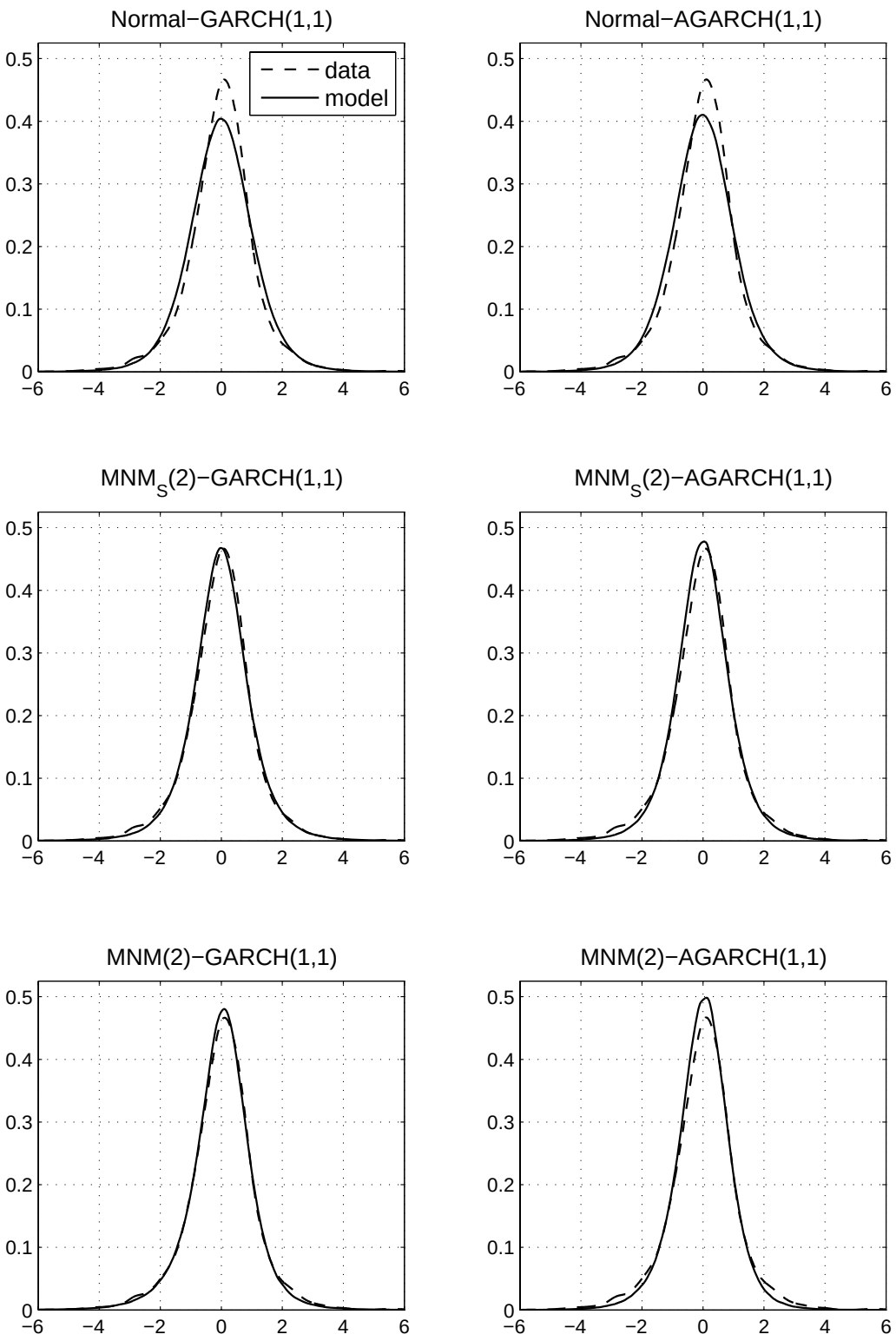

Figure 3: Kernel density estimates for the NASDAQ errors. Shown are, for each fitted GARCH model, kernel density estimates for the estimated empirical NASDAQ errors in (20) (dashed line), along with kernel estimates for the error distributions implied by the respective models (solid line), as obtained from simulated samples of length 1,000,000.

slowly, with the NASDAQ exhibiting more significant lags than the DJIA. While the single–regime models fail to capture this pattern, the mixture models tend to do better in this regard. However, the mixture models with leverage effects, with the exception of $\text{MNM}_S(2)$–AGARCH(1,1) in case of the NASDAQ, suffer from the autocorrelations being much too small at the beginning. Overall, models $\text{MNM}_S(2)$–GARCH(1,1) and MNM(2)–GARCH(1,1) provide the best fit to the empirical autocorrelations of the squares, which can presumably be explained by the reasoning at the end of Section 2.3.

3.2 Application to Value–at–Risk

In this section, we evaluate the models' capability to accurately measure the out–of–sample Value–at–Risk (VaR) of portfolios formed from the stock indices under investigation. In Section 3.2.1 we discuss methods for evaluating the VaR measures provided by the respective models, and Section 3.2.2 presents the empirical results.

3.2.1 Backtesting Value–at–Risk Measures

VaR is a widely employed tool in risk management (e.g., Christoffersen and Pelletier, 2004), and it can briefly be defined as follows. For a given model, the VaR at level ξ for period t, denoted by $\text{VaR}_t(\xi)$, is implicitly defined by $\widehat{F}(\text{VaR}_t(\xi)|\Psi_{t-1}) = \xi$, where $\widehat{F}(\cdot|\Psi_{t-1})$ is the conditional cumulative distribution function (cdf) of the portfolio return, $r_{p,t}$, implied by the model under consideration. A *violation* or *hit* is said to occur at time t if $r_{p,t} < \text{VaR}_t(\xi)$. To test the models' suitability for calculating accurate ex–ante VaR measures, we define the binary sequence

$$I_t = \begin{cases} 1, & \text{if} \quad r_{p,t} < \text{VaR}_t, \\ 0, & \text{if} \quad r_{p,t} \geq \text{VaR}_t. \end{cases} \tag{21}$$

Then the empirical shortfall probability is $\widehat{\xi} = x/T$, where $x = \sum_{t=1}^{T} I_t$ is the number of observed violations, and T is the number of forecasts evaluated. Two tests on the sequence (21) will be conducted, which can be characterized as tests for correct *unconditional* and *conditional* coverage, respectively.

For the first test, based on ideas of Kupiec (1995), we note that, from both the risk management and the regulatory perspective, the main interest is often whether a model's actual shortfall probability is *greater* than the target probability ξ. Therefore, the check whether $\widehat{\xi}$ is significantly larger than ξ is conducted using a one–sided binomial test, where

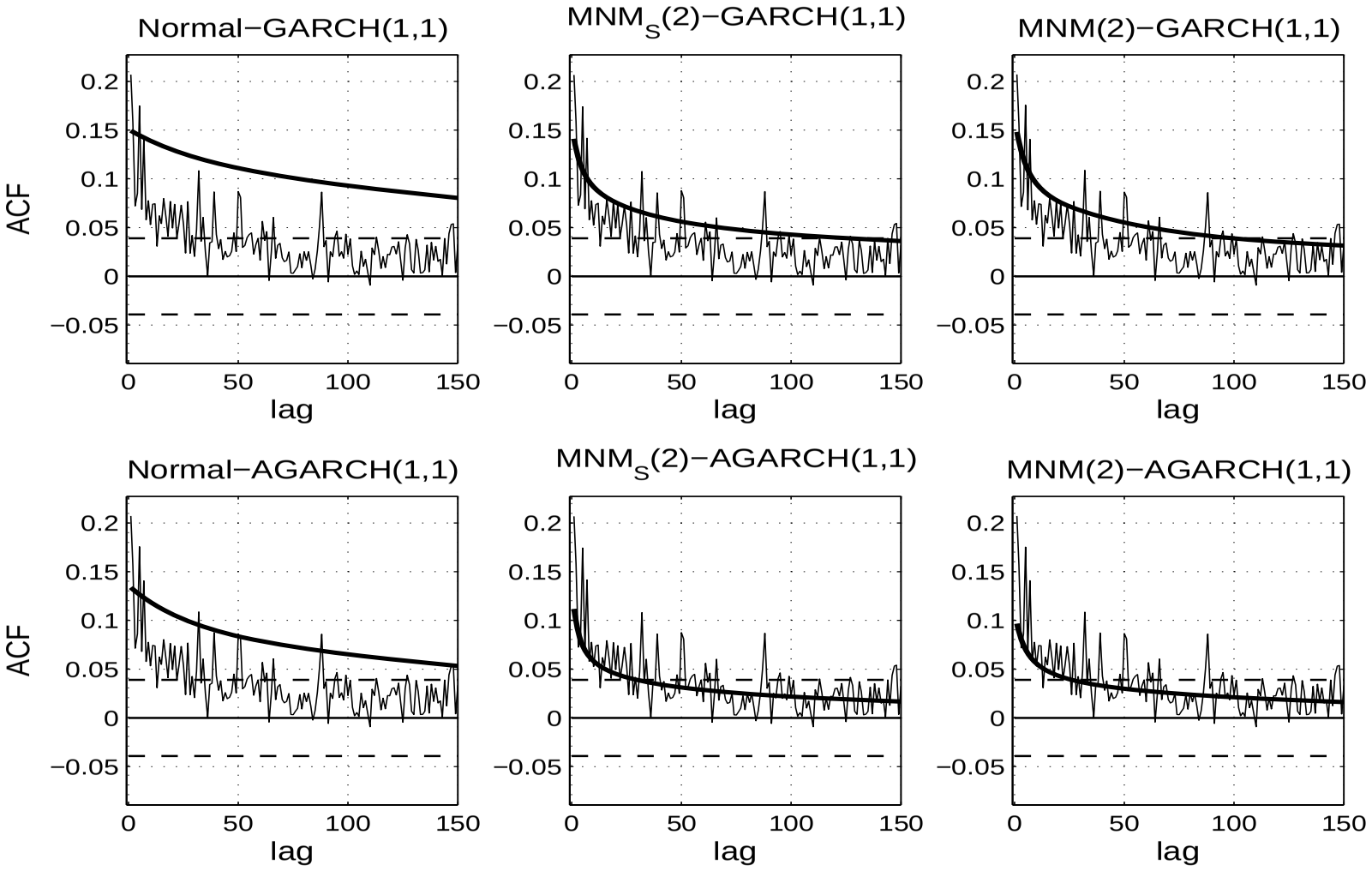

Figure 4: Shown are empirical and model–implied theoretical (bold line) autocorrelations of the squared errors for the DJIA. Dashed lines represent approximate 95% one–at–a–time confidence intervals.

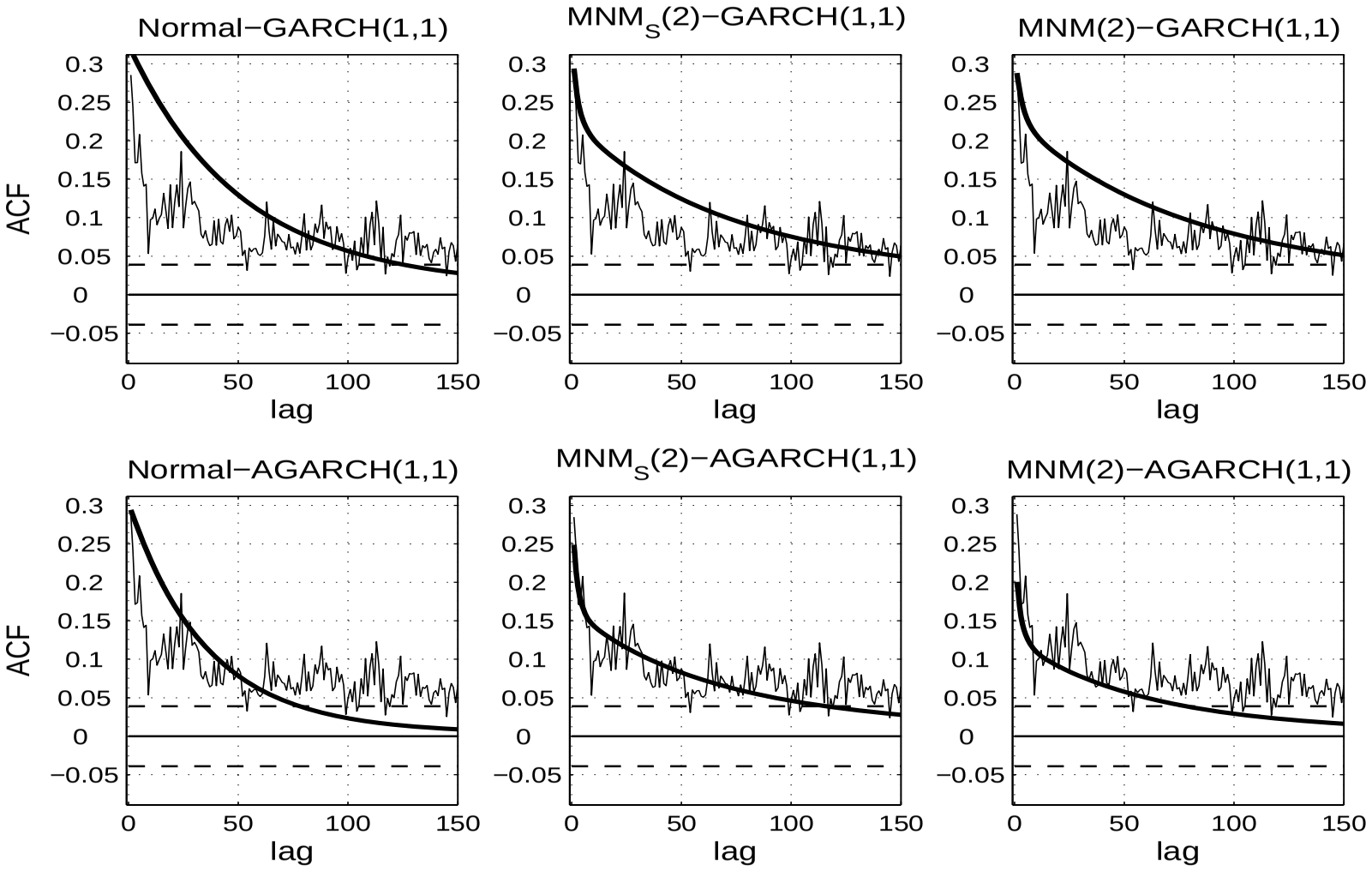

Figure 5: Shown are empirical and model–implied theoretical (bold line) autocorrelations of the squared errors for the NASDAQ. Dashed lines represent approximate 95% one–at–a–time confidence intervals.

the p–values are calculated by

$$p = \sum_{i=x}^{T} \binom{T}{i} \xi^i (1-\xi)^{T-i}. \tag{22}$$

If, according to (22), $\widehat{\xi}$ is significantly larger than ξ, then the model under investigation *on average* tends to *underestimate* the risk of the financial position. However, as stressed by Christoffersen (1998) and Lopez (1999), a satisfactory backtesting method should be able to detect both deviations from the unconditional nominal shortfall probability, ξ, as well as violation clustering. For example, a VaR model that fails to appropriately account for higher–order dynamics in the return density (e.g., ARCH effects) may be correct on average (have *unconditional* shortfall probability ξ), but in any given period will have uncorrect probability of violation, leading to violation clustering. See, however, Jorion (2002) for a skeptical discussion of the *economic* significance of violation clustering.

A duration–based backtesting approach which allows to detect rather general deviations from independence of the sequence (21) has recently been developed by Christoffersen and Pelletier (2004). Statistically, correct *conditional* coverage implies that the sequence $\{I_t\}$ defined in (21) is a random sample from a Bernoulli distribution with probability (of violation) parameter ξ, which in turn implies that the number of days between two violations is geometric. More formally, define the duration of time in days between two violations as

$$D_i = t_i - t_{i-1}, \tag{23}$$

where t_i denotes the day of violation number i. Then, for a correctly specified VaR model, the probability density function of the duration is given by

$$f_G(d;\xi) = (1-\xi)^{d-1}\xi, \quad d \in \mathbb{N}. \tag{24}$$

The geometric distribution is characterized unambiguously by its "lack of memory" property (cf. Rohatgi, 1976, p. 191), which means that the probability of observing a hit today does not depend on the number of days elapsed since the last violation. The statistical concept for characterizing the memory of a lifetime distribution is the hazard function, $\lambda(d)$, which, in the discrete framework, is defined to be the conditional probability of a violation on day d given that $d-1$ days have passed without a violation, that is,

$$\lambda(d) := \Pr(D = d | D \geq d) = \frac{\Pr(D=d)}{\Pr(D \geq d)} = \frac{f(d)}{\sum_{j=d}^{\infty} f(j)} = \frac{f(d)}{S(d)}, \tag{25}$$

where $S(d) := \Pr(D \geq d)$ denotes the survivor function. The "lack of memory" property of the geometric distribution (24) is associated with a constant hazard function, i.e., $\lambda_G(d) = \xi$

for all $d \geq 1$. In contrast, violation clustering corresponds to a decreasing hazard function (or *negative duration dependence*), implying that the probability of a no–hit spell ending shortly decreases as the spell increases in length. Christoffersen and Pelletier (2004) propose to test the iid–ness of the binary sequence (21) via the "lack of memory" property of the sequence of durations defined in (23). The approach is to specify a lifetime distribution with a flexible hazard function that nests the geometric, so that the "lack of memory" property can be tested by means of likelihood ratio (LR) tests. See Kiefer (1988) and Christoffersen and Pelletier (2004) for a discussion of how to construct the likelihood function in the case of censored spells.

In the applications of their approach, Christoffersen and Pelletier (2004) use the continuous analogue of (24), i.e., the exponential distribution, which, for testing, can be nested in the continuous Weibull distribution. As shown by Haas (2006), however, tests based on a discrete analogue of the continuous Weibull nesting the geometric (24) have (often considerably) more power to detect violation clustering, and, therefore, we employ the discrete Weibull distribution of Nakagawa and Osaki (1975), given by the probability density function

$$f_{DW}(d; a, b) = \exp\{-a^b(d-1)^b\} - \exp\{-a^b d^b\}, \quad a, b > 0, \quad d \in \mathbb{N}, \tag{26}$$

with distribution, survivor, and hazard functions given by $F_{DW}(d; a, b) = 1 - \exp\{-a^b d^b\}$, $S_{DW}(d; a, b) = \exp\{-a^b(d-1)^d\}$, and $\lambda_{DW}(d) = 1 - \exp\{-a^b[d^b - (d-1)^b]\}$, respectively. The geometric (24) is nested in (26) for $b = 1$ and $\xi = 1 - \exp\{-a\}$, and (26) has decreasing (increasing) hazard if $b < 1$ ($b > 1$). Thus, the hypothesis of a correct conditional (cc) shortfall probability ξ implies a simultaneous test of

$$H_{0,cc} : b = 1 \quad \text{and} \quad a = -\log(1 - \xi). \tag{27}$$

As pointed out by Christoffersen and Pelletier (2004), although the large–sample properties of the LR test are known, they may not lead to reliable inference in particular for small VaR levels, because, even if the return series is reasonably long, the associated series of durations will be rather short due to the scarcity of violations. Thus, for controlling the size of the tests, the Monte Carlo technique of Dufour (2006) is adopted for calculating p–values. To implement this technique, we first generate N independent realizations of the LR test statistic, LR_i, $i = 1, \ldots, N$, under the null hypothesis, i.e., using durations constructed from independent Bernoulli hit sequences, where we use $N = 9,999$. We denote by LR_0 the value of the test statistic obtained for the original sample. As there are no nuisance parameters under the null hypothesis, the only complication is that the test statistics derived from binary sequences such as (21) are discrete random variables, i.e., it may happen that $LR_i = LR_0$ for some i,

$1 \leq i \leq N$. Thus, we need a rule to break ties between the test value obtained from the original sample and those obtained from Monte Carlo simulation under the null hypothesis. As shown by Dufour (2006), in this situation, the Monte Carlo p–values can be calculated as follows. For each test statistic, LR_i, $i = 0, \ldots, N$, draw a random variable, U_i, $i = 0, \ldots, N$, which is independently uniformly distributed over the interval $(0, 1)$. The Monte Carlo p–value, $p_N(LR_0)$, is then given by

$$p_N(LR_0) = \frac{NG_N(LR_0) + 1}{N + 1},$$

where

$$G_N(LR_0) = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(LR_i \leq LR_0) + \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(LR_i = LR_0)\mathbf{1}(U_i \geq U_0),$$

and $\mathbf{1}(A)$ is the indicator function associated with statement A, i.e., $\mathbf{1}(A) = 1$ if A is true and $\mathbf{1}(A) = 0$ otherwise.

3.2.2 Empirical Results

In our application, we calculate one–step–ahead out–of–sample VaR measures and consider the VaR levels $\xi = 0.0025, 0.005, 0.01, 0.025$, and 0.05. The parameter estimates are updated (approximately) every month (i.e., 20 trading days) employing a moving window of data, i.e., using the most recent 2,527 observations in the sample. In this manner, we obtain, for each model, 1,947 one–step–ahead out–of–sample VaR measures.

In addition to the six GARCH models considered above, we also include the RiskMetrics model into the comparison, which, as a benchmark, has gained some popularity among risk management practitioners (JP Morgan, 1996). This model assumes that the conditional return distribution is normal with a covariance matrix H_t driven by an exponentially weighted moving average of past shocks,

$$H_t = \lambda H_{t-1} + (1 - \lambda)\epsilon_{t-1}\epsilon'_{t-1} = (1 - \lambda)\sum_{i=1}^{\infty} \lambda^{i-1}\epsilon_{t-i}\epsilon'_{t-i}, \tag{28}$$

where λ is fixed at 0.94 for daily data. To make the models comparable, we couple (28) with an AR(1) process for the conditional mean as in (20), where the parameters are estimated via a simple least squares regression.

To select economically reasonable portfolios, we assume that the preferences of the investor can be characterized by an exponential expected utility function of the form

$$U(r_{p,t}) = -\exp\{-cr_{p,t}\}, \quad c > 0, \tag{29}$$

where c is the coefficient of constant absolute risk aversion, and $r_{p,t}$ is the portfolio return at time t, i.e., $r_{p,t} = w_t r_{1t} + (1 - w_t) r_{2t}$, where w_t is the portfolio weight of the DJIA at time t. We note that, due to our use of continuously compounded returns, the linear relation between the returns of the individual indices and the portfolio return is only an approximation. For daily returns, however, this approximation is usually rather accurate and standard practice; for discussion, see, e.g., Fama (1976, Ch. 1). For a Gaussian investor with predictive density $r_t|\Psi_{t-1} \sim \mathrm{N}(\mu_t, H_t)$, where $\mu_t = (\mu_{1t}, \mu_{2t})'$, and $H_t = (h_{ij,t})_{i,j=1,2}$, the optimal portfolio weight in period t is given by

$$w_t = \frac{h_{22,t} - h_{12,t}}{h_{11,t} + h_{22,t} - 2h_{12,t}} + \frac{1}{c} \frac{\mu_{1t} - \mu_{2t}}{h_{11,t} + h_{22,t} - 2h_{12,t}}. \tag{30}$$

Note that the first term on the right–hand side of (30) represents the global minimum variance portfolio (GMVP). Expected utility of a mixture investor with predictive density $r_t|\Psi_{t-1} \sim \lambda_{1t}\mathrm{N}(\mu_{1t}, H_{1t}) + \lambda_{2t}\mathrm{N}(\mu_{2t}, H_{2t})$ is given by

$$\mathrm{E}[U(r_{p,t})|\Psi_{t-1}] = -\lambda_{1t} \exp\left\{-c\tilde{w}_t'\mu_{1t} + \frac{c^2}{2}\tilde{w}_t' H_{1t} \tilde{w}_t\right\} - \lambda_{2t} \exp\left\{-c\tilde{w}_t'\mu_{2t} + \frac{c^2}{2}\tilde{w}_t' H_{2t} \tilde{w}_t\right\}, \tag{31}$$

where $\tilde{w}_t = (w_t, 1 - w_t)'$. As the portfolio problem of the mixture investor does not admit a closed–from solution, we use the Newton–Raphson method to find the portfolio weight which maximizes (31). To account for different risk attitudes, we do the computations for values of c in (29) ranging from 0.1 to 1.5. Further increasing c did not result in any notable differences compared to $c = 1.5$.

The results are reported in Tables 7 and 8 for the tests for unconditional and conditional coverage, respectively. In Table 7, for each value of risk aversion, c, and VaR level, ξ, we show the empirical *percentage* shortfall probability $100 \times \widehat{\xi}$ of the respective models, as well as the mean absolute error (MAE) over the different c–values, given by $\mathrm{MAE}(\xi) = (1/6)\sum_{i=1}^{6} |\xi - \widehat{\xi}(c_i)|$, where $\widehat{\xi}(c_i)$ is the empirical shortfall probability associated with the ith value of c, $i = 1, \ldots, 6$. In Table 8, as the parameter a of the discrete Weibull distribution (26) is not easily interpretable for $b \neq 1$, we report, along with the estimated memory parameter b, the quantity $100/\mathrm{E}(D)$, where $\mathrm{E}(D)$ is the mean duration implied by the fitted discrete Weibull, i.e., $\mathrm{E}(D) = \sum_{d=1}^{\infty} d f_{DW}(d; a, b) = \sum_{d=0}^{\infty}(1 - F_{DW}(d; a, b)) = \sum_{d=0}^{\infty} \exp\{-a^b d^b\}$, which may serve as an estimate for the unconditional percentage shortfall probability.

Tables 7 and 8 show, in accordance with earlier results (e.g., Diebold et al., 1999), that the RiskMetrics model (28) is clearly not appropriate, as it significantly underestimates the VaR at all levels. The single–component GARCH(1,1) models, although much better than the RiskMetrics specification, are likewise inadequate in particular for the lower VaR levels, i.e.,

the more extreme risks. They do reasonably well for the higher levels $\xi = 0.025$ and 0.05. This is reconcilable with the occasionally expressed view that normality may be an appropriate assumption for everyday risks, given that, for example, the VaR at level 0.05 is expected to be violated once every month.

The best results with respect to unconditional coverage, as reported in Table 7, are obtained for the mixture model with an asymmetric conditional density and without leverage, i.e., model MNM(2)–GARCH(1,1), although model MNM(2)–AGARCH(1,1) also performs reasonably well. It thus appears that, in the context of the mixture models, capturing the asymmetries in the conditional density is much more important than accounting for dynamic asymmetries in the conditional variance. This may appear somewhat surprising in view of the in–sample significance of the leverage terms reported in Table 3, but the finding is similar to earlier results such as those of Loudon et al. (2000), who compare a number of both symmetric and asymmetric univariate GARCH models when applied to British stock returns. They find that the parameters governing the asymmetric response to negative and positive shocks are all highly significant in–sample, but the out–of–sample performance of symmetric and asymmetric models is fairly similar. Note, however, that an at least moderate improvement when allowing for leverage effects is observed within the class of single–component Normal–GARCH(1,1) models.

A comparison of the results for the duration–based tests in Table 8 with those in Table 7 reveals that violation clustering, in general, seems to be not a serious problem, although, for all models, the estimated memory parameter, b, tends to be (often slightly) below unity, thus indicating mild deviations from the geometric distribution. As before, models MNM(2)–GARCH(1,1) and MNM(2)–AGARCH(1,1) exhibit the best fit. However, while model MNM(2)–GARCH(1,1) passes the test for correct unconditional coverage for all (c, ξ)–combinations, the hypothesis of correct conditional coverage is now rejected in two cases. In particular, the estimated value of $b = 0.61$ for $c = 0.1$ and $\xi = 0.01$ indicates a relatively strong clustering of violations, leading to a rejection at the 1% level. Similarly, significant violation clustering is detected for several c–values at the 5% VaR level for the Normal–AGARCH(1,1) model, where the duration–based test, in contrast to the results in Table 7, rejects the hypothesis of a correctly specified VaR model.

Table 7: Evaluation of Value–at–Risk (VaR) measures: Unconditional coverage ($100 \times \widehat{\xi}$).

Risk aversion, c	0.1	0.25	0.5	0.75	1	1.5	MAE(α)
RiskMetrics							
VaR(0.0025)	0.87***	0.92***	0.98***	0.98***	1.03***	1.03***	0.0072
VaR(0.005)	1.23***	1.18***	1.23***	1.23***	1.23***	1.23***	0.0072
VaR(0.01)	1.69***	1.85***	2.00***	2.00***	2.00***	2.00***	0.0093
VaR(0.025)	3.85***	3.80***	3.80***	3.80***	3.85***	3.80***	0.0132
VaR(0.05)	6.57***	6.68***	6.78***	6.73***	6.68***	6.63***	0.0168
Normal–GARCH(1,1)							
VaR(0.0025)	0.41	0.56**	0.62***	0.62***	0.67***	0.67***	0.0034
VaR(0.005)	0.77*	0.87**	0.92**	0.98***	0.98***	0.98***	0.0042
VaR(0.01)	1.49**	1.34*	1.44**	1.44**	1.39*	1.34*	0.0040
VaR(0.025)	3.24**	2.93	2.93	3.03*	2.98	3.08*	0.0053
VaR(0.05)	5.34	5.39	5.39	5.44	5.29	5.39	0.0038
Normal–AGARCH(1,1)							
VaR(0.0025)	0.46*	0.56**	0.67***	0.67***	0.67***	0.62***	0.0036
VaR(0.005)	0.72	0.82**	0.92**	0.98***	0.98***	0.98***	0.0040
VaR(0.01)	1.34*	1.23	1.39*	1.39*	1.39*	1.44**	0.0036
VaR(0.025)	2.98	2.88	2.77	2.82	2.88	2.82	0.0036
VaR(0.05)	4.83	5.14	5.34	5.29	5.34	5.34	0.0027
$\text{MNM}_S(2)$–GARCH(1,1)							
VaR(0.0025)	0.26	0.31	0.36	0.36	0.31	0.31	0.0007
VaR(0.005)	0.46	0.51	0.51	0.51	0.67	0.72	0.0008
VaR(0.01)	1.28	1.18	1.13	1.18	1.13	1.13	0.0017
VaR(0.025)	3.49***	2.88	2.93	3.08*	3.03*	3.03*	0.0057
VaR(0.05)	6.11**	5.91**	6.11**	5.96**	5.86**	5.75*	0.0095
$\text{MNM}_S(2)$–AGARCH(1,1)							
VaR(0.0025)	0.26	0.46*	0.51**	0.56**	0.56**	0.51**	0.0023
VaR(0.005)	0.56	0.67	0.72	0.77*	0.77*	0.82**	0.0022
VaR(0.01)	1.39*	1.13	1.23	1.23	1.23	1.23	0.0024
VaR(0.025)	2.98	2.82	2.82	2.82	2.88	2.82	0.0036
VaR(0.05)	5.91**	5.91**	5.65	5.44	5.55	5.65	0.0068
MNM(2)–GARCH(1,1)							
VaR(0.0025)	0.26	0.36	0.36	0.26	0.26	0.26	0.0004
VaR(0.005)	0.41	0.46	0.46	0.51	0.51	0.56	0.0004
VaR(0.01)	0.92	0.77	0.87	0.87	0.92	0.92	0.0012
VaR(0.025)	2.57	2.26	2.41	2.52	2.52	2.62	0.0009
VaR(0.05)	5.39	5.39	5.44	5.29	5.29	5.19	0.0033
MNM(2)–AGARCH(1,1)							
VaR(0.0025)	0.31	0.51**	0.46*	0.36	0.36	0.36	0.0014
VaR(0.005)	0.51	0.62	0.67	0.72	0.77*	0.77*	0.0018
VaR(0.01)	1.28	0.98	1.18	1.18	1.18	1.18	0.0017
VaR(0.025)	2.72	2.67	2.47	2.52	2.52	2.52	0.0008
VaR(0.05)	5.55	5.44	5.24	5.19	5.24	5.29	0.0032

Shown are the results of the tests for correct *unconditional* coverage of out–of–sample Value–at–Risk (VaR) measures. "VaR(ξ)" refers to the VaR measures for a nominal shortfall probability ξ implied by the respective models. Reported are the empirical percentage shortfall probabilities, $100 \times \widehat{\xi} = 100 \times x/T$, observed for a nominal VaR level ξ, $\xi = 0.0025, 0.005, 0.01, 0.025, 0.05$, where x is the empirical shortfall frequency, and T is the number of forecasts evaluated. Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively, as obtained from the one–sided binomial test (22). For each model and each nominal VaR level, ξ, "MAE(ξ)" is the mean absolute error (MAE) over the different levels of risk aversion, c, i.e., $\text{MAE}(\xi) = (1/6) \sum_{i=1}^{6} |\xi - \widehat{\xi}(c_i)|$, where $(c_1, c_2, c_3, c_4, c_5, c_6) = (0.1, 0.25, 0.5, 0.75, 1, 1.5)$.

Table 8: Evaluation of Value–at–Risk (VaR) measures: Duration–based tests $(100/\mathrm{E}(D), b)$.

Risk aversion, c	0.1	0.25	0.5	0.75	1	1.5
RiskMetrics						
VaR(0.0025)	$(0.84, 1.22)^{***}$	$(0.88, 1.02)^{***}$	$(0.93, 1.05)^{***}$	$(0.93, 1.05)^{***}$	$(0.98, 1.05)^{***}$	$(0.98, 1.05)^{***}$
VaR(0.005)	$(1.18, 0.96)^{***}$	$(1.12, 0.92)^{***}$	$(1.17, 0.92)^{***}$	$(1.17, 0.92)^{***}$	$(1.17, 0.92)^{***}$	$(1.17, 0.92)^{***}$
VaR(0.01)	$(1.61, 0.79)^{***}$	$(1.79, 0.87)^{***}$	$(1.94, 0.81)^{***}$	$(1.94, 0.81)^{***}$	$(1.94, 0.81)^{***}$	$(1.94, 0.81)^{***}$
VaR(0.025)	$(3.78, 0.89)^{***}$	$(3.74, 0.95)^{***}$	$(3.75, 0.99)^{***}$	$(3.75, 0.99)^{***}$	$(3.80, 0.98)^{***}$	$(3.75, 0.98)^{***}$
VaR(0.05)	$(6.52, 0.96)^{***}$	$(6.61, 0.92)^{***}$	$(6.69, 0.86)^{***}$	$(6.65, 0.87)^{***}$	$(6.60, 0.89)^{***}$	$(6.56, 0.90)^{***}$
Normal–GARCH(1,1)						
VaR(0.0025)	$(0.35, 0.82)$	$(0.49, 0.77)^{**}$	$(0.55, 0.89)^{**}$	$(0.55, 0.89)^{**}$	$(0.61, 0.92)^{***}$	$(0.61, 0.92)^{***}$
VaR(0.005)	$(0.68, 0.72)^{*}$	$(0.80, 0.83)^{*}$	$(0.86, 0.88)^{*}$	$(0.92, 0.94)^{**}$	$(0.92, 0.94)^{**}$	$(0.92, 0.94)^{**}$
VaR(0.01)	$(1.39, 0.77)^{**}$	$(1.26, 0.81)$	$(1.37, 0.83)^{*}$	$(1.37, 0.83)^{*}$	$(1.32, 0.87)$	$(1.28, 0.97)$
VaR(0.025)	$(3.18, 0.96)$	$(2.87, 0.91)$	$(2.85, 0.86)$	$(2.96, 0.86)$	$(2.91, 0.84)$	$(3.01, 0.83)^{*}$
VaR(0.05)	$(5.27, 0.86)$	$(5.33, 0.87)$	$(5.33, 0.87)$	$(5.39, 0.89)$	$(5.23, 0.89)$	$(5.34, 0.93)$
Normal–AGARCH(1,1)						
VaR(0.0025)	$(0.41, 0.97)$	$(0.49, 0.74)^{**}$	$(0.61, 0.89)^{***}$	$(0.61, 0.89)^{***}$	$(0.61, 0.89)^{***}$	$(0.55, 0.79)^{**}$
VaR(0.005)	$(0.65, 0.86)$	$(0.75, 0.84)$	$(0.87, 0.98)^{*}$	$(0.93, 1.03)^{**}$	$(0.93, 1.03)^{**}$	$(0.93, 1.03)^{**}$
VaR(0.01)	$(1.27, 0.92)$	$(1.20, 1.18)$	$(1.35, 1.19)$	$(1.35, 1.19)$	$(1.36, 1.28)^{*}$	$(1.41, 1.27)^{*}$
VaR(0.025)	$(2.90, 0.87)$	$(2.80, 0.90)$	$(2.67, 0.83)$	$(2.73, 0.84)$	$(2.79, 0.86)$	$(2.73, 0.80)^{*}$
VaR(0.05)	$(4.75, 0.89)$	$(5.06, 0.88)$	$(5.26, 0.84)^{*}$	$(5.18, 0.79)^{**}$	$(5.24, 0.79)^{***}$	$(5.24, 0.79)^{***}$
$\mathrm{MNM}_S(2)$–GARCH(1,1)						
VaR(0.0025)	$(0.21, 1.17)$	$(0.23, 0.72)$	$(0.30, 0.88)$	$(0.30, 0.88)$	$(0.27, 1.61)$	$(0.27, 1.61)$
VaR(0.005)	$(0.38, 0.67)$	$(0.47, 1.09)$	$(0.48, 1.21)$	$(0.48, 1.21)$	$(0.63, 1.25)$	$(0.67, 1.10)$
VaR(0.01)	$(1.21, 0.87)$	$(1.12, 0.86)$	$(1.09, 1.11)$	$(1.13, 1.06)$	$(1.08, 1.02)$	$(1.08, 1.05)$
VaR(0.025)	$(3.45, 1.06)^{**}$	$(2.82, 0.95)$	$(2.86, 0.88)$	$(3.02, 0.90)$	$(2.97, 0.91)$	$(2.96, 0.84)^{*}$
VaR(0.05)	$(6.05, 0.89)^{**}$	$(5.85, 0.89)^{*}$	$(6.05, 0.87)^{**}$	$(5.89, 0.87)^{**}$	$(5.80, 0.88)^{*}$	$(5.70, 0.90)$

continued

Table 8: Evaluation of Value–at–Risk (VaR) measures: Duration–based tests $(100/\mathrm{E}(D), b)$ (continued).

Risk aversion, c	0.1	0.25	0.5	0.75	1	1.5
MNM$_S$(2)–AGARCH(1,1)						
VaR(0.0025)	$(0.21, 1.11)$	$(0.38, 0.65)^{**}$	$(0.44, 0.71)^{**}$	$(0.50, 0.83)^{**}$	$(0.50, 0.83)^{**}$	$(0.45, 0.86)$
VaR(0.005)	$(0.50, 0.79)$	$(0.60, 0.79)$	$(0.66, 0.91)$	$(0.72, 1.07)$	$(0.72, 1.07)$	$(0.78, 1.11)$
VaR(0.01)	$(1.32, 0.87)$	$(1.07, 0.88)$	$(1.18, 0.95)$	$(1.18, 0.95)$	$(1.18, 0.95)$	$(1.18, 0.95)$
VaR(0.025)	$(2.92, 0.93)$	$(2.77, 0.93)$	$(2.76, 0.84)$	$(2.76, 0.84)$	$(2.81, 0.85)$	$(2.75, 0.83)$
VaR(0.05)	$(5.85, 0.92)$	$(5.86, 0.95)$	$(5.60, 0.94)$	$(5.39, 0.93)$	$(5.49, 0.93)$	$(5.59, 0.91)$
MNM(2)–GARCH(1,1)						
VaR(0.0025)	$(0.21, 1.17)$	$(0.30, 0.88)$	$(0.30, 0.88)$	$(0.21, 1.41)$	$(0.21, 1.41)$	$(0.21, 1.41)$
VaR(0.005)	$(0.36, 0.92)$	$(0.40, 0.83)$	$(0.40, 0.83)$	$(0.46, 0.93)$	$(0.46, 0.93)$	$(0.51, 1.00)$
VaR(0.01)	$(0.85, 0.61)^{***}$	$(0.72, 1.02)$	$(0.83, 1.04)$	$(0.83, 1.04)$	$(0.88, 1.09)$	$(0.88, 1.09)$
VaR(0.025)	$(2.51, 0.93)$	$(2.19, 0.85)$	$(2.34, 0.81)$	$(2.45, 0.82)$	$(2.45, 0.82)$	$(2.56, 0.81)$
VaR(0.05)	$(5.32, 0.89)$	$(5.32, 0.83)^{**}$	$(5.38, 0.92)$	$(5.22, 0.89)$	$(5.21, 0.85)$	$(5.12, 0.89)$
MNM(2)–AGARCH(1,1)						
VaR(0.0025)	$(0.26, 1.07)$	$(0.44, 0.74)^{*}$	$(0.38, 0.65)^{**}$	$(0.29, 0.74)$	$(0.29, 0.74)$	$(0.29, 0.74)$
VaR(0.005)	$(0.43, 0.73)$	$(0.55, 0.86)$	$(0.61, 0.88)$	$(0.66, 0.91)$	$(0.72, 0.95)$	$(0.72, 0.95)$
VaR(0.01)	$(1.21, 0.82)$	$(0.92, 0.86)$	$(1.13, 0.91)$	$(1.13, 0.91)$	$(1.13, 0.91)$	$(1.13, 0.91)$
VaR(0.025)	$(2.67, 0.96)$	$(2.62, 0.92)$	$(2.43, 0.84)$	$(2.48, 0.82)$	$(2.48, 0.82)$	$(2.48, 0.81)^{*}$
VaR(0.05)	$(5.49, 0.90)$	$(5.40, 0.99)$	$(5.19, 0.97)$	$(5.14, 0.96)$	$(5.19, 0.95)$	$(5.24, 0.93)$

Shown are the results of the duration–based tests (27) for correct *conditional* coverage of out–of–sample Value–at–Risk (VaR) measures. "VaR(ξ)" refers to the VaR measures for a nominal shortfall probability ξ implied by the respective models. Reported are the pairs $(100/\mathrm{E}(D), b)$, where $\mathrm{E}(D) = \sum_{d=0}^{\infty} \exp\{-a^b d^b\}$ is the mean duration implied by the estimated parameters a and b of the discrete Weibull distribution (26), applied to the sequence of durations between observed shortfalls (23) derived from the hit sequence (21), and b is the (estimated) parameter monitoring the memory of the duration process. Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively, as obtained from Monte Carlo likelihood ratio tests, as described at the end of Section 3.2.1.

4 Conclusions

Several extensions and modifications of the analysis conducted in this paper are worth exploring. Most importantly, the unrestricted BEKK parametrization employed herein will not be suitable when the number of assets under consideration is large, because the number of parameters increases quadratically with the dimension of the return vector. The curse of dimensionality plagues multivariate GARCH models in general, but it will appear even more burdensome in the current framework, because we have as many covariance matrices to parameterize as we have mixture components. As noted in Section 2.2, the diagonal BEKK may be appropriate in situations with a relatively large number of assets; a recent application to a relatively high–dimensional problem in the framework of dynamic conditional correlation models is Cappiello et al. (2006). A perhaps more promising approach, however, which may be useful even in problems of rather high dimension, is to combine the present approach with the principal component GARCH model proposed in Alexander and Chibumba (1997) and Alexander (2001, 2002). In this context, a two–step estimation procedure suggests itself, where, on the second step, as the number of factors retained should be small, a relatively low–dimensional normal mixture GARCH model could be fitted to the factors which have been extracted on the first step as the conventional principal components. Another issue for further research is the development of easily implementable techniques for risk management and portfolio selection accommodating features such as regime–specific correlation structures and leverage effects, as documented in Section 3 of the present paper.

Acknowledgements

We are grateful for constructive comments and suggestions from two anonymous referees and the associate editor, which led to significant improvements of the paper. We also thank Matteo Bonato and participants of the 2005 NBER/NSF Satellite Workshop "Financial Risk and Time Series Analysis" in Munich and the 13th Annual Meeting of the German Finance Association in Oestrich–Winkel, 2006, for helpful discussions. The research of M. Haas was supported by the *Deutsche Forschungsgemeinschaft* (DFG). Part of the research of M. S. Paolella has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the *Swiss National Science Foundation*.

Appendix

In the Appendix, we derive the conditions for the moments of the MNM(k)–GARCH(1,1) model. We also provide expressions for these moments and the autocorrelation structure of the process.

A Notation

To conveniently write down the unconditional moments of the multivariate normal mixture GARCH model, use of several patterned matrices is rather advantageous, and we define them here. A detailed discussion of (as well as explicit expressions for) these matrices can be found in Magnus (1988). The first of these matrices is the *commutation matrix*, K_{mn}, which is the $mn \times mn$ matrix with the property that $K_{mn}\text{vec}(A) = \text{vec}(A')$ for every $m \times n$ matrix A. We will use the fact that the commutation matrix allows us to transform the vec of a Kronecker product into the Kronecker product of the vecs (Magnus, 1988, Theorem 3.6). More precisely, for an $m \times n$ matrix A and an $p \times q$ matrix B, it is true that

$$\text{vec}(A \otimes B) = (I_n \otimes K_{qm} \otimes I_p)(\text{vec}A \otimes \text{vec}B). \tag{A.1}$$

The *elimination matrix*, L_n, is the $n(n+1)/2 \times n^2$ matrix that takes away the redundant elements of a symmetric $n \times n$ matrix, i.e., for every $n \times n$ matrix A, we have $L_n\text{vec}(A) = \text{vech}(A)$. In contrast, the *duplication matrix*, D_n, is the $n^2 \times n(n+1)/2$ matrix with the property that $D_n\text{vech}(A) = \text{vec}(A)$ for every symmetric $n \times n$ matrix A. Its Moore–Penrose inverse, D_n^+, is given by $D_n^+ = (D_n'D_n)^{-1}D_n'$ (Magnus, 1988, Theorem 4.1).

To compactify the expressions for the moments of our model, we will also made extensive use of the matrix $\tilde{N}_n = (I_{n^2} + K_{nn})/2$, which is discussed in Section 3.10 of Magnus (1988), and which has the property that, for every $n \times n$ matrix A,

$$2\tilde{N}_n\text{vec}(A) = \text{vec}(A + A'). \tag{A.2}$$

Note that the matrix D_n^+ has a similar property. Namely, because of $D_n^+ = L_n\tilde{N}_n$ (Magnus, 1988, p. 80), we have

$$2D_n^+\text{vec}(A) = \text{vech}(A + A'). \tag{A.3}$$

B The Third and Fourth Moments of an Asymmetric Multivariate Normal Mixture Distribution

In this Appendix, we provide convenient expressions for the expectations of vec[vech(xx')x'] and vec[vech(xx')vech(xx')$'$], when x has a multivariate normal mixture distribution with (possibly) nonzero means, as defined in (1) and (2). These expressions will be useful for computing the unconditional moments of the multivariate mixed normal GARCH process in Appendix C.

To derive the expressions given in this Appendix, we draw on results of Magnus and Neudecker (1979), Balestra and Holly (1990), and Hafner (2003). We state the central results as Lemmas 2–4 for the third, and Lemmas 5–8 for the fourth moment. Details of the derivations are presented only for the third moment, because those for the fourth moment are similar. Detailed derivations are available on request from the authors.

B.1 The Third Moment

To find a compact expression for E{vec[vech(xx')x']}, which is needed due to the inclusion of the leverage terms, we make use of a formula of Balestra and Holly (1990) which we state as Lemma 2.

Lemma 2 *(Balestra and Holly, 1990) For an M–dimensional random vector x, which is normally distributed with mean μ and covariance matrix H, we have*

$$E[(x \otimes x)x'] = vec(H)\mu' + 2\tilde{N}_M(\mu \otimes H) + (\mu \otimes \mu)\mu'. \tag{B.4}$$

We are interested in E{vec[vech(xx')x']} as a linear function in h, where $h = \text{vech}(H)$. Such an expression is provided next.

Lemma 3 *For an M–dimensional random vector x, which is normally distributed with mean μ and covariance matrix H, we have*

$$E\{vec[vech(xx')x']\} = (I_M \otimes L_M)[\tilde{G}_M(\mu \otimes D_M)h + \mu \otimes \mu \otimes \mu], \tag{B.5}$$

where $h = vech(H)$, and

$$\tilde{G}_M = I_{M^3} + 2(I_M \otimes \tilde{N}_M)(K_{MM} \otimes I_M). \tag{B.6}$$

Proof. By Lemma 2, and using $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, we have

$$\begin{aligned} \text{E}\{\text{vec}[\text{vech}(xx')x']\} &= \text{E}\{\text{vec}[L_M\text{vec}(xx')x']\} = (I_M \otimes L_M)\text{E}\{\text{vec}[(x \otimes x)x']\} \\ &= (I_M \otimes L_M)\text{vec}[\text{vec}(H)\mu' + 2\tilde{N}_M(\mu \otimes H) + (\mu \otimes \mu)\mu']. \end{aligned}$$

Furthermore, $\text{vec}[2\tilde{N}_M(\mu \otimes H)] = 2(I_M \otimes \tilde{N}_M)\text{vec}(\mu \otimes H)$, and (A.1) implies that $\text{vec}(\mu \otimes H) = (K_{MM} \otimes I_M)(\mu \otimes \text{vec}(H))$. Finally, as $y \otimes x = \text{vec}(xy')$ for vectors x and y, we have $\mu \otimes \text{vec}(H) = \text{vec}[\text{vec}(H)\mu'] = \text{vec}(D_M h \mu') = (\mu \otimes D_M)h$, and thus (B.5). ■

Next, we consider the case of a normal mixture distribution.

Lemma 4 *Assume that* $x \sim MNM(\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_k, H_1, \ldots, H_k)$. *Let* $\lambda = (\lambda_1, \ldots, \lambda_k)'$, $\Lambda = diag(\lambda)$; $h_j = vech(H_j)$, $j = 1, \ldots, k$; $h = (h_1', \ldots, h_k')'$; $\Upsilon = (\mu_1, \ldots, \mu_k)$; $\mu = vec(\Upsilon) = (\mu_1', \ldots, \mu_k')'$; $\tilde{\mu}_j = vech(\mu_j \mu_j')$, $j = 1, \ldots, k$; $\tilde{\Upsilon} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$; *and* $\tilde{\mu} = vec(\tilde{\Upsilon}) = (\tilde{\mu}_1', \ldots, \tilde{\mu}_k')'$. *Then,*

$$\begin{aligned} &E\{vec[vech(xx')x']\} \\ &\quad = (I_M \otimes L_M)\tilde{G}_M(\Upsilon\Lambda \otimes D_M)h + (I_M \otimes vec(\Lambda)' \otimes I_N)(K_{Mk} \otimes I_{kN})vec(\tilde{\mu}\mu'), \end{aligned} \tag{B.7}$$

where $N = M(M+1)/2$, *and* $\tilde{G}_M$ *is defined in (B.6).*

Proof. Lemma 4 follows from the fact that the third moment of the mixture is just the weighted average of the component–specific moments as given in (B.5), i.e., for x mixed normal as defined in Lemma 4, we have

$$\text{E}\{\text{vec}[\text{vech}(xx')x']\} = (I_M \otimes L_M)\left\{\tilde{G}_M \sum_{j=1}^{k} \lambda_j(\mu_j \otimes D_M)h_j + \sum_{j=1}^{k} \lambda_j(\mu_j \otimes \mu_j \otimes \mu_j)\right\}. \tag{B.8}$$

Let e_j be the jth unit vector in $\mathbb{R}^k$. Then, for the first sum on the right–hand side of (B.8), we have that

$$\begin{aligned} \sum_{j=1}^{k} \lambda_j(\mu_j \otimes D_M)h_j &= \left\{\sum_{j=1}^{k} \lambda_j(e_j' \otimes \mu_j \otimes D_M)\right\} h = \left\{\left(\sum_{j=1}^{k} \lambda_j \mu_j e_j'\right) \otimes D_M\right\} h \\ &= (\Upsilon\Lambda \otimes D_M)h, \end{aligned} \tag{B.9}$$

where, in the last equation in the first line in (B.9), we have used that $y' \otimes x = xy'$. For the second sum on the right–hand side of (B.8), we find

$$\begin{aligned} \sum_j \lambda_j(\mu_j \otimes \mu_j \otimes \mu_j) &= \sum_j \lambda_j \text{vec}[(\mu_j \otimes \mu_j)\mu_j'] = (I_M \otimes D_M)\sum_j \lambda_j \text{vec}(\tilde{\mu}_j \mu_j') \\ &= (I_M \otimes D_M)\sum_j \lambda_j \text{vec}[(e_j' \otimes I_N)(\tilde{\mu}\mu')(e_j \otimes I_M)] \\ &= (I_M \otimes D_M)\sum_j \lambda_j (e_j' \otimes I_M \otimes e_j' \otimes I_N)\text{vec}(\tilde{\mu}\mu') \\ &= (I_M \otimes D_M)\sum_j \lambda_j (I_M \otimes e_j' \otimes e_j' \otimes I_N)(K_{Mk} \otimes I_{kN})\text{vec}(\tilde{\mu}\mu') \\ &= (I_M \otimes D_M)\sum_j \lambda_j (I_M \otimes \text{vec}(e_j e_j')' \otimes I_N)(K_{Mk} \otimes I_{kN})\text{vec}(\tilde{\mu}\mu') \\ &= (I_M \otimes D_M)(I_M \otimes \text{vec}(\Lambda)' \otimes I_N)(K_{Mk} \otimes I_{kN})\text{vec}(\tilde{\mu}\mu'), \end{aligned} \tag{B.10}$$

where we have used the identity $(A \otimes b')K_{np} = b' \otimes A$ for $m \times n$ matrix A and $p \times 1$ vector b (Magnus, 1988, p. 36). Finally, because $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ if AC and BD exist, we have $(I_M \otimes L_M)(I_M \otimes D_M) = (I_M \otimes L_M D_M)$, and, by Theorem 5.5 of Magnus (1988), $L_M D_M = I_N$, $N = M(M+1)/2$, so we get (B.7). ■

B.2 The Fourth Moment

For the fourth moment, we build on results of Magnus and Neudecker (1979) and Hafner (2003) which we state as Lemmas 5 and 6, respectively.

Lemma 5 *(Magnus and Neudecker, 1979, Theorem 4.3) For an M–dimensional random vector x, which is normally distributed with mean μ and covariance matrix H, we have*

$$\begin{aligned} E[(x \otimes x)(x \otimes x)'] &= 2D_M D_M^+ (H \otimes H) + vec(H)vec(H)' \\ &\quad + 2D_M D_M^+ (H \otimes \mu\mu' + \mu\mu' \otimes H) \\ &\quad + vec(H)vec(\mu\mu')' + vec(\mu\mu')vec(H)' + vec(\mu\mu')vec(\mu\mu')'. \end{aligned} \tag{B.11}$$

For the result in Lemma 5, see also Magnus (1988, Ch. 10).

We are interested in $\mathrm{E}[\mathrm{vech}(xx')\mathrm{vech}(xx')']$. Using the identity $\mathrm{vec}(xx') = x \otimes x$ and the definition of the elimination matrix L_M, this can be written as $L_M \mathrm{E}[(x \otimes x)(x \otimes x)']L_M'$, which is a simple transformation of (B.11). The case of a normal distribution with zero mean was considered by Hafner (2003), who considered the more general class of spherical distributions.

Lemma 6 *(Hafner, 2003, Theorem 1) For an M–dimensional normally distributed random vector x with zero mean and covariance matrix H, we have*

$$vec\{E[vech(xx')vech(xx')']\} = G_M vec(hh'), \tag{B.12}$$

where $h = vech(H)$, and

$$G_M = 2(L_M \otimes D_M^+)(I_M \otimes K_{MM} \otimes I_M)(D_M \otimes D_M) + I_{N^2}, \tag{B.13}$$

and $N := M(M+1)/2$ is the number of independent elements in H.

Our first step is to generalize (B.12) to the case of nonzero means, i.e., to consider the terms in the second and third line of (B.11).

Lemma 7 *For an M–dimensional normally distributed random vector x with mean μ and covariance matrix H, we have*

$$vec\{E[vech(xx')vech(xx')']\} = G_M vec(hh') + 2G_M \tilde{N}_N(\tilde{\mu} \otimes I_N)h + vec(\tilde{\mu}\tilde{\mu}'), \tag{B.14}$$

where G_M is defined in (B.13), $h = vech(H)$, $\tilde{\mu} = vech(\mu\mu')$, and $N = M(M+1)/2$.

The proof of Lemma 7 can be carried out along similar lines as the proof of Theorem 1 in Hafner (2003). The case of a multivariate normal mixture distribution is considered next. We make use of the notation introduced in Lemma 4.

Lemma 8 *Assume that* $x \sim MNM(\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_k, H_1, \ldots, H_k)$. *Then,*

$$\begin{aligned} & vec\{E[vech(xx')vech(xx')']\} \\ & \quad = G_M(I_N \otimes vec(\Lambda)' \otimes I_N)(K_{Nk} \otimes I_{kN})vec(hh') + 2G_M\tilde{N}_N(\tilde{\Upsilon}\Lambda \otimes I_N)h \\ & \quad +(I_N \otimes vec(\Lambda)' \otimes I_N)(K_{Nk} \otimes I_{kN})vec(\tilde{\mu}\tilde{\mu}'). \end{aligned} \tag{B.15}$$

Lemma 8 is obtained by combining the results of Lemma 7 with the fact that the fourth moment of the mixture is just the weighted average of the component–specific moments as given in (B.14), quite similar to equation (B.8) for the third moment, and by using arguments similar to those in the derivation of Lemma 4. For example, to show that

$$\sum_{j=1}^{k} \lambda_j \mathrm{vec}(h_j h_j') = (I_N \otimes \mathrm{vec}(\Lambda)' \otimes I_N)(K_{Nk} \otimes I_{kN})\mathrm{vec}(hh'), \tag{B.16}$$

we essentially repeat the argument in (B.10).

C The Moments of the MNM(k)–AGARCH(1,1) Model

In this Appendix, we use the results of Appendix B to derive the unconditional second and fourth moments of the asymmetric multivariate mixed normal GARCH(1,1) model as given in equation (11), as well as the conditions for their existence.

C.1 Moment Conditions

We will use the notation introduced in Section 2 and Lemmas 4 and 8. Also, as defined in (12), $\rho(A)$ denotes the largest eigenvalue in modulus of a square matrix A. Let $W_t = (h_t', \mathrm{vec}(h_t h_t')')'$. We have $\mathrm{E}(\eta_{t-1}|\Psi_{t-2}) = (\lambda' \otimes I_N)(h_{t-1} + \tilde{\mu})$, so that, using $A_1(\lambda' \otimes I_N) = (1 \otimes A_1)(\lambda' \otimes I_N) = \lambda' \otimes A_1$,

$$\mathrm{E}(h_t|\Psi_{t-2}) = \tilde{A}_0 + A_1(\lambda' \otimes I_N)\tilde{\mu} + (\lambda' \otimes A_1 + B_1)h_{t-1}. \tag{C.17}$$

Moreover, using the matrix $\tilde{N}_n$, and in particular its basic property (A.2), we have

$$\begin{aligned} \mathrm{vec}(h_t h_t') &= \tilde{A}_0 \otimes \tilde{A}_0 + 2\tilde{N}_{kN}\mathrm{vec}[\tilde{A}_0(\eta_{t-1}'A_1' + h_{t-1}'B_1')] + 2\tilde{N}_{kN}\mathrm{vec}(A_1\eta_{t-1}h_{t-1}'B_1') \\ &\quad +(A_1 \otimes A_1)\mathrm{vec}(\eta_{t-1}\eta_{t-1}') + (B_1 \otimes B_1)\mathrm{vec}(h_{t-1}h_{t-1}') \\ &\quad +\mathrm{vec}(\Theta_1\epsilon_{t-1}\epsilon_{t-1}'\Theta_1') - 2\tilde{N}_{kN}\mathrm{vec}[(\tilde{A}_0 + A_1\eta_{t-1} + B_1h_{t-1})\epsilon_{t-1}'\Theta_1']. \end{aligned} \tag{C.18}$$

Using Lemmas 4 and 8, and combining (C.17) and (C.18), it is now straightforward to derive the recursion

$$\mathrm{E}(W_t|\Psi_{t-2}) = d + CW_{t-1}, \tag{C.19}$$

where

$$d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad C = \begin{pmatrix} C_{11} & 0_{kN\times(kN)^2} \\ C_{21} & C_{22} \end{pmatrix},$$

and

$$\begin{aligned}
d_1 &= \tilde{A}_0 + A_1(\lambda' \otimes I_N)\tilde{\mu} \\
d_2 &= \tilde{A}_0 \otimes \tilde{A}_0 + 2\tilde{N}_{kN}(\lambda' \otimes A_1 \otimes \tilde{A}_0)\tilde{\mu} + (A_1 \otimes A_1)(I_N \otimes \mathrm{vec}(\Lambda)' \otimes I_N)(K_{Nk} \otimes I_{kN})\mathrm{vec}(\tilde{\mu}\tilde{\mu}') \\
&\quad +(\Theta_1 \otimes \Theta_1)D_M(\lambda' \otimes I_N)\tilde{\mu} - 2\tilde{N}_{kN}(\Theta_1 \otimes A_1)(I_M \otimes \mathrm{vec}(\Lambda)' \otimes I_N)(K_{Mk} \otimes I_{kN})\mathrm{vec}(\tilde{\mu}\mu'), \\
C_{11} &= \lambda' \otimes A_1 + B_1, \\
C_{21} &= 2\tilde{N}_{kN}(\lambda' \otimes A_1 + B_1) \otimes \tilde{A}_0 + 2\tilde{N}_{kN}[B_1 \otimes (\lambda' \otimes A_1)\tilde{\mu}] + 2(A_1 \otimes A_1)G_M\tilde{N}_N(\tilde{\Upsilon}\Lambda \otimes I_N) \\
&\quad +(\Theta_1 \otimes \Theta_1)D_M(\lambda' \otimes I_N) - 2\tilde{N}_{kN}(\Theta_1 \otimes A_1)(I_M \otimes L_M)\tilde{G}_M(\Upsilon\Lambda \otimes D_M), \\
C_{22} &= (A_1 \otimes A_1)G_M(I_N \otimes \mathrm{vec}(\Lambda)' \otimes I_N)(K_{Nk} \otimes I_{kN}) + 2\tilde{N}_{kN}(B_1 \otimes \lambda' \otimes A_1) + B_1 \otimes B_1.
\end{aligned}$$

Iterating (C.19), we obtain

$$\mathrm{E}(W_t|\Psi_{t-\tau-1}) = \sum_{i=0}^{\tau-1} C^i d + C^\tau W_{t-\tau}. \tag{C.20}$$

From the block–triangular structure of C, we have, from (C.20), that

$$\mathrm{E}(h_t|\Psi_{t-\tau-1}) = \sum_{i=0}^{\tau-1} C_{11}^i d_1 + C_{11}^\tau h_{t-\tau}. \tag{C.21}$$

Thus, as we have assumed that the process starts indefinitely far in the past with finite fourth moments, the unconditional expectation $\mathrm{E}(h_t)$ exists and is given by the limit as $\tau \to \infty$, i.e.,

$$\mathrm{E}(h_t) = \lim_{\tau\to\infty} \mathrm{E}(h_t|\Psi_{t-\tau-1}) = \sum_{i=0}^{\infty} C_{11}^i d_1 = (I_{kN} - C_{11})^{-1} d_1 \tag{C.22}$$

if and only if $\rho(C_{11}) < 1$, as stated in (13). By the same line of reasoning, $\mathrm{E}(W_t)$ exists if and only if, in addition, $\rho(C_{22}) < 1$, as stated in (14). In this case, by partitioned inversion of C,

$$\mathrm{E}[\mathrm{vec}(h_t h_t')] = (I_{(kN)^2} - C_{22})^{-1}(d_2 + C_{21}(I_{kN} - C_{11})^{-1} d_1). \tag{C.23}$$

C.2 Autocovariance Function of the Squares

To find the autocovariance matrices, i.e., $\Gamma(\tau) = \mathrm{E}(\eta_t \eta_{t-\tau}') - \mathrm{E}(\eta_t)\mathrm{E}(\eta_t)'$, we first note that (C.21) in Appendix C.1 implies

$$\mathrm{E}(h_t|\Psi_{t-\tau}) = \sum_{i=0}^{\tau-2} C_{11}^i d_i + C_{11}^{\tau-1} h_{t-\tau+1} = \mathrm{E}(h_t) + C_{11}^{\tau-1}[h_{t-\tau+1} - \mathrm{E}(h_t)].$$

Hence,

$$\begin{aligned}
\mathrm{E}(\eta_t \eta_{t-\tau}') &= \mathrm{E}[\mathrm{E}(\eta_t|\Psi_{t-\tau})\eta_{t-\tau}'] \\
&= \mathrm{E}\{(\lambda' \otimes I_N)[\mathrm{E}(h_t|\Psi_{t-\tau}) + \tilde{\mu}]\eta_{t-\tau}'\} \\
&= (\lambda' \otimes I_N)\mathrm{E}\{[\mathrm{E}(h_t) + \tilde{\mu} + C_{11}^{\tau-1}(h_{t-\tau+1} - \mathrm{E}(h_t))]\eta_{t-\tau}'\} \\
&= \mathrm{E}(\eta_t)\mathrm{E}(\eta_t)' + (\lambda' \otimes I_N)C_{11}^{\tau-1}\mathrm{E}\left\{[\tilde{A}_0 + A_1\eta_{t-\tau} - \Theta_1\epsilon_{t-\tau} + B_1 h_{t-\tau} - \mathrm{E}(h_t)]\eta_{t-\tau}'\right\}.
\end{aligned}$$

Thus we have (15) with

$$Q = \mathrm{E}\left\{[\tilde{A}_0 + A_1\eta_t - \Theta_1\epsilon_t + B_1 h_t - \mathrm{E}(h_t)]\eta_t'\right\}. \tag{C.24}$$

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In Petrov, B. N. and Csaki, F., editors, *2nd International Symposium on Information Theory*, pages 267–281, Akademiai Kiado, Budapest.

Alexander, C. (2001). *Orthogonal GARCH*, volume 2 of *Mastering Risk*, pages 21–38. FT Prentice Hall, London.

Alexander, C. (2002). Principal Component Models for Generating Large GARCH Covariance Matrices. *Economic Notes*, 31:337–359.

Alexander, C. and Chibumba, A. M. (1997). Multivariate Orthogonal Factor GARCH. Mimeo, University of Sussex.

Alexander, C. and Lazar, E. (2005). Asymmetries and Volatility Regimes in the European Equity Market. ICMA Centre Discussion Papers in Finance 2005–14, The Business School for Financial Markets at the University of Reading.

Alexander, C. and Lazar, E. (2006). Normal Mixture GARCH(1,1). Applications to Exchange Rate Modelling. *Journal of Applied Econometrics*, 21:307–336.

Ané, T. (2006). An Analysis of the Flexibility of Asymmetric Power GARCH Models. *Computational Statistics and Data Analysis*, 51:1293–1311.

Ang, A. and Chen, J. (2002). Asymmetric Correlations of Equity Portfolios. *Journal of Financial Economics*, 63:443–494.

Ausin, M. C. and Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics and Data Analysis*, 51:2636–2652.

Balestra, P. and Holly, A. (1990). A General Kronecker Formula for the Moments of the Multivariate Normal Distribution. Cahiers de recherches économiques 9002, Département d'économétrie et d'économie politique, Université de Lausanne.

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

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics*, 21:79–109.

Bauwens, L. and Rombouts, J. V. K. (2007). Bayesian Inference for the Mixed Conditional Heteroskedasticity Model. *Econometrics Journal*, 10:408–425.

Bauwens, L. and Storti, G. (2007). A Component GARCH Model with Time-varying Weights. CORE Discussion Paper 2007/19, Center for Operations Research and Econometrics, Université Catholique de Louvain.

Bertholon, H., Monfort, A., and Pegoraro, F. (2006). Pricing and Inference with Mixtures of Conditionally Normal Processes. Working paper, Centre de Recherche en Économie et Statistique (CREST), Laboratoire der Finance-Assurance.

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

Bollerslev, T. and Engle, R. F. (1993). Common Persistence in Conditional Variances. *Econometrica*, 61:167–186.

Broto, C. and Ruiz, E. (2006). Unobserved Component Models with Asymmetric Conditional Variances. *Computational Statistics and Data Analysis*, 50:2146–2166.

Cappiello, L., Engle, R. F., and Sheppard, K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, 4:537–572.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 4:841–862.

Christoffersen, P. F. and Pelletier, D. (2004). Backtesting Value–at–Risk: A Duration–Based Approach. *Journal of Financial Econometrics*, 2:84 – 108.

Cohen, A. C. (1967). Estimation in Mixtures of Two Normal Distributions. *Technometrics*, 9:15–28.

Diebold, F. X., Hahn, J., and Tay, A. S. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-frequency Returns on Foreign Exchange. *Review of Economics and Statistics*, 81:661–673.

Ding, Z. and Granger, C. W. J. (1996). Modeling Volatility Persistence of Speculative Returns: A New Approach. *Journal of Econometrics*, 73:185–215.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106.

Dufour, J.-M. (2006). Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite–sample Inference and Nonstandard Asymptotics. *Journal of Econometrics*, 133:443–477.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50:987–1008.

Engle, R. F. (1990). Stock Volatility and the Crash of '87: Discussion. *Review of Financial Studies*, 3:103–106.

Engle, R. F. and Kroner, K. F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11:122–150.

Fama, E. F. (1976). *Foundations of Finance*. Basic Books, New York.

Giannikis, D., Vrontos, I. D., and Dellaportas, P. (2008). Modelling Nonlinearities and Heavy Tails via Threshold Normal Mixture GARCH Models. *Computational Statistics and Data Analysis*, 52:1549–1571.

Haas, M. (2006). Improved Duration–based Backtesting of Value–at–Risk. *Journal of Risk*, 8:17–38.

Haas, M. (2007). Volatility Components and Long Memory-Effects Revisited. *Studies in Nonlinear Dynamics & Econometrics*, 11(2):Article 3.

Haas, M., Mittnik, S., and Mizrach, B. (2006a). Assessing Central Bank Credibility During the EMS Crises: Comparing Option and Spot Market-Based Forecasts. *Journal of Financial Stability*, 2:28–54.

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

Haas, M., Mittnik, S., and Paolella, M. S. (2006b). Multivariate Normal Mixture GARCH. CFS Working Paper 2006/9, Center for Financial Studies.

Hafner, C. M. (2003). Fourth Moment Structure of Multivariate GARCH Models. *Journal of Financial Econometrics*, 1:26–54.

Jorion, P. (2002). Fallacies about the Effects of Market Risk Management Systems. *Journal of Risk*, 5:75–96.

JP Morgan (1996). *RiskMetrics—Technical Document*. New York.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90:773–795.

Kiefer, N. M. (1988). Economic Duration Data and Hazard Functions. *Journal of Economic Literature*, 26:646–679.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives*, 3:73–84.

Lopez, J. A. (1999). Regulatory Evaluation of Value–at–Risk Models. *Journal of Risk*, 1:37–64.

Loudon, G. F., Watt, W. H., and Yadav, P. K. (2000). An Empirical Analysis of Alternative Parametric ARCH Models. *Journal of Applied Econometrics*, 15:117–136.

Magnus, J. R. (1988). *Linear Structures*. Griffin, London.

Magnus, J. R. and Neudecker, H. (1979). The Commutation Matrix: Some Properties and Applications. *Annals of Statistics*, 7:381–394.

Maheu, J. (2005). Can GARCH Models Capture Long-Range Dependence? *Studies in Nonlinear Dynamics & Econometrics*, 9(4):Article 1.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons, New York.

Nakagawa, T. and Osaki, S. (1975). The Discrete Weibull Distribution. *IEEE Transactions on Reliability*, 24:300–301.

Patton, A. J. (2004). On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation. *Journal of Financial Econometrics*, 2:130–168.

Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley and Sons, New York.

Ruppert, D. (1987). What is Kurtosis? An Influence Function Approach. *American Statistician*, 41:1–5.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6:461–464.

Sentana, E. (1995). Quadratic ARCH Models. *Review of Economic Studies*, 62:639–661.

Silverman, B. W. (1986). *Density estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Teicher, H. (1963). Identifiability of Finite Mixtures. *Annals of Mathematical Statistics*, 34:1265–1269.

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

Wilkins, J. E. (1944). A Note on Skewness and Kurtosis. *Annals of Mathematical Statistics*, 15:333–335.

Wong, C. S. and Li, W. K. (2001). On a Mixture Autoregressive Conditional Heteroscedastic Model. *Journal of the American Statistical Association*, 96:982–985.

Wu, C. and Lee, J. C. (2007). Estimation of a Utility-based Asset Pricing Model using Normal Mixture GARCH(1,1). *Economic Modelling*, 24:329–349.

Yakowitz, S. J. and Spragins, J. D. (1968). On the Identifiability of Finite Mixtures. *Annals of Mathematical Statistics*, 39:209–214.

CFS Working Paper Series:

No.	Author(s)	Title
2008/06	**Charles Grant** **Christos Koulovatianos** **Alexander Michaelides** **Mario Padula**	Evidence on the Insurance Effect of Marginal Income Taxes
2008/05	**Dimitris Christelis** **Dimitris Georgarakos** **Michael Haliassos**	Economic Integration and Mature Portfolios
2008/04	**Elena Carletti** **Philipp Hartmann** **Steven Onega**	The Economic Impact of Merger Control Legislation
2008/03	**Annamaria Lusardi** **Olivia S. Mitchell**	Planning and Financial Literacy: How Do Women Fare?
2008/02	**Bannier** **Hirsch**	The Economics of Rating Watchlists: Evidence from Rating Changes
2008/01	**Sumit Agarwal** **Chunlin Liu** **Nicholas Souleles**	The Reaction of Consumer Spending and Debt Tax Rebates – Evidence from Consumer Credit Data
2007/34	**Todd Sinai** **Nicholas Souleles**	Net Worth and Housing Equity in Retirement
2007/33	**Annamaria Lusardi** **Olivia S. Mitchell**	Financial Literacy and Retirement Planning: New Evidence from the Rand American Life Panel
2007/32	**Giuseppe Bertola** **Stefan Hochguertel**	Household Debt and Credit: Economic Issues and Data Problems
2007/31	**Giuseppe Bertola**	Finance and Welfare States in Globalizing Markets

Copies of working papers can be downloaded at http://www.ifk-cfs.de