Multivariate Regime Switching GARCH with an Application to International Stock Markets

Abstract

We develop a multivariate generalization of the Markov–switching GARCH model introduced by Haas, Mittnik, and Paolella (2004b) and derive its fourth–moment structure. The dynamic properties of our model are established by employing a new GARCH(1,1) representation of a multivariate GARCH(p, q) process, which also simplifies the computation of higher moments and autocorrelations for a couple of more common uni– and multivariate GARCH(p, q) models. An application to international stock markets illustrates the relevance of accounting for volatility regimes from both a statistical and economic perspective, including out–of–sample portfolio selection and computation of Value–at–Risk.

JEL classification: C32; C51; G10; G11

Keywords—conditional volatility, Markov–switching, multivariate GARCH, portfolio selection, regime–dependent correlations, second–order dependence

1 Introduction

There is considerable evidence that the distribution of asset returns depends on an unobserved state (or *regime*) of the market (see, e.g., Turner, Startz, and Nelson, 1989; Hamilton and Susmel, 1994; Ramchand and Susmel, 1998; Perez–Quiros and Timmermann, 2001; Ang and Bekaert, 2002, 2004; and Guidolin and Timmermann, 2005a,b, 2006). In particular, researchers often identify a low– and a high–volatility regime, where the correlations between assets tend to be higher in the adverse state of the market. These findings have important implications for asset allocation and risk management purposes, because "it is in times of extreme market conditions that the benefits from diversification ... are most urgently needed" (Campbell, Koedijk, and Kofman, 2002). In addition, if the next period's regime is not known with certainty, investors will want to hedge against the possible occurrence of the high–volatility regime.

Markov–switching (MS) models, as introduced by Hamilton (1989), have been found to be useful for capturing regime–dependent return distributions. Even the most simple version of such an MS model, where the time variability of the parameters is governed solely by the unobserved regime variable, can generate rather flexible return distributions, including skewness, excess kurtosis, volatility clustering, and regime–dependent correlation structures (cf. Ryden, Teräsvirta, and Åsbrink, 1998; and Timmermann, 2000). However, for returns sampled at a daily or weekly frequency, it has been observed that the volatility dynamics are not adequately captured by the switching between constant regime–specific variances and covariances (Pagan and Schwert, 1990; Gray, 1996; Timmermann, 2000; Marcucci, 2005), i.e., a considerable part of the conditional heteroskedasticity is linked to within–regime ARCH–type dynamics rather than to the discrete regime process. This has motivated the introduction of the MS ARCH model in Cai (1994) and Hamilton and Susmel (1994), which was generalized to MS GARCH by Gray (1996) and Dueker (1997) and further elaborated by Klaassen (2002). A discussion of these models is provided in Haas, Mittnik, and Paolella (2004b). These authors also propose a new MS GARCH process and argue that their version can be viewed as the most natural specification of a multi–regime GARCH model. Their model has been further investigated in Liu (2006) and Abramson and Cohen (2007).

In this paper, we develop a multivariate generalization of the MS GARCH process introduced in Haas, Mittnik, and Paolella (2004b) and derive a number of its dynamic properties which are relevant for the analysis of the volatility dynamics. Although different versions of multivariate MS GARCH models have occasionally appeared in the literature (e.g., Ramchand and Susmel, 1998; and Ang and Chen, 2002), this is the first attempt to provide a general framework for the analysis of such models. Moreover, the main tool for deriving the dynamic properties of our model, an apparently newly introduced GARCH(1,1) representation of a multivariate GARCH(p, q) model, is also useful for computing the moment structure of more common uni– and multivariate GARCH(p, q) models, and its application leads to expressions which are often easier to implement than those previously obtained in the literature. This will be illustrated at the end of Section 2.3.2 by means of the standard univariate GARCH(p, q) model of Bollerslev (1986). We also note that, although in this paper we employ the vech representation of a multivariate GARCH model, the GARCH(1,1) representation put forward in Section 2.3 is easily adjusted to be applicable to multi–regime versions of Jeantheau's (1998) generalization of Bollerslev's (1990) constant conditional correlation model, the moment structure of which, in the single–regime case, has been considered by Ling and McAleer (2003) and He and Teräsvirta (2004). An MS model related to this family of processes has recently been proposed by Pelletier (2006).

The paper is organized as follows. In Section 2, we define the model, discuss special cases and estimation issues, and derive its dynamic properties. In Section 3, we present an application to international stock market indices, including an evaluation of out–of–sample fit in the context of portfolio selection and computation of Value–at–Risk. Section 4 draws conclusions and indicates areas for further research.

2 The Model and its Properties

In this Section, we define the multivariate Markov–switching GARCH process and derive its dynamic properties. The model studied in this paper represents a multi–regime version of the vech form of a multivariate GARCH(p, q) model, as introduced by Bollerslev, Engle, and Wooldridge (1988). As detailed in Section 2.2, this specification nests several more parsimonious parametrizations, but any other, nonnested variant of a multivariate GARCH process could as well be engaged. In particular, as mentioned in the Introduction, if the extended constant conditional correlation model of Jeantheau (1998) is adopted, all the theoretical results of Sections 2.3.1 and 2.3.2 go through with only minor modifications.

2.1 Definition of the Model

Let the M–dimensional time series $\{\epsilon_t\}$ satisfy

$$\epsilon_t = H_{\Delta_t,t}^{1/2} \xi_t, \tag{1}$$

where $\xi_t \overset{iid}{\sim} \mathrm{N}(0_{M\times 1}, I_M)$, I_n denotes the identity matrix of dimension n, and $\{\Delta_t\}$ is a Markov chain with finite state space $S = \{1, 2, \ldots, k\}$ and a primitive (i.e., irreducible and aperiodic) $k \times k$ transition matrix P,

$$P = \begin{bmatrix} p_{11} & \cdots & p_{k1} \\ \vdots & \cdots & \vdots \\ p_{1k} & \cdots & p_{kk} \end{bmatrix}, \tag{2}$$

where $p_{ij} = p(\Delta_t = j|\Delta_{t-1} = i)$, $i, j = 1, \ldots, k$. Moreover, it is assumed that $\{\xi_t\}$ and $\{\Delta_t\}$ are independent. We will denote by $\pi_t = [\pi_{1t}, \ldots, \pi_{kt}]'$ and $\pi_\infty = [\pi_{1,\infty}, \ldots, \pi_{k,\infty}]'$ the distribution at time t and the stationary distribution of the Markov chain, respectively.

Stack the $N := M(M+1)/2$ independent elements of the regime–dependent conditional covariance matrices, H_{jt}, and the "squared" ϵ_t (i.e., $\epsilon_t\epsilon_t'$) in $h_{jt} := \mathrm{vech}(H_{jt})$, $j = 1, \ldots, k$, and $\eta_t := \mathrm{vech}(\epsilon_t\epsilon_t')$, respectively. Then the regime–dependent covariance matrices evolve according to a multivariate GARCH(p, q) equation in vech form,

$$h_{jt} = A_{0j} + \sum_{i=1}^{q} A_{ij}\eta_{t-i} + \sum_{i=1}^{p} B_{ij}h_{j,t-i}, \quad j = 1, \ldots, k, \tag{3}$$

where A_{ij}, $i = 0, \ldots, q$, and B_{ij}, $i = 1, \ldots, p$, are parameter matrices of appropriate dimension, $j = 1, \ldots, k$. We will refer to the model defined by (1)–(3) as a multivariate Markov–switching GARCH$(p, q; k)$ process, or, in short, MMSG$(p, q; k)$.

To compactify the notation and facilitate the analysis of the model, let $h_t := [h_{1t}', \ldots, h_{kt}']'$, $A_0 = [A_{01}', \ldots, A_{0k}']'$, $A_i = [A_{i1}', \ldots, A_{ik}']'$, $i = 1, \ldots, q$, 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

$$h_t = A_0 + \sum_{i=1}^{q} A_i\eta_{t-i} + \sum_{i=1}^{p} B_ih_{t-i}. \tag{4}$$

In the univariate framework, it is argued in Haas, Mittnik, and Paolella (2004b) that the model (1)–(3) is the "most natural" extension of the GARCH approach to the multi–regime setting, and their reasoning directly carries over to the multivariate situation. Briefly, in the single–regime case, the most general conditional heteroskedastic specification is an ARCH(∞), i.e., $h_t = \nu + \Phi(L)\eta_t$, where $\Phi(L) = \sum_{i=1}^{\infty} \Phi_i L^i$, and L is the lag operator, $L^i y_t = y_{t-i}$. To make this applicable, one usually specifies $\Phi(L) = (I_N - B(L))^{-1}A(L)$, where $B(L) = \sum_{i=1}^{p} B_iL^i$ and $A(L) = \sum_{i=1}^{q} A_iL^i$ are lag polynomials of order p and q, respectively. This leads to a GARCH(p, q) process, i.e., $h_t = A_0 + A(L)\eta_t + B(L)h_t$, where $A_0 = (I_N - B(1))v$. Specification (1)–(3) is based on the same logic applied to each regime.

An interesting special case of model (1)–(3) arises when the transition matrix P in (2) has rank 1, i.e., $P = \pi_\infty \mathbf{1}_k'$, where $\mathbf{1}_k$ is a k–dimensional column of ones. This results in a multivariate normal mixture GARCH$(p, q; k)$ model, or, in short, MNMG$(p, q; k)$, with a constant vector of mixing weights given by π_∞. As discussed in Haas, Mittnik, and Paolella (2004b), we can then, in contrast to (1), allow for different (nonzero) regime means, and thus for conditional and unconditional skewness, without abandoning a central property of GARCH processes, namely, lack of serial correlation in connection with pronounced dependencies in power–transformed absolute returns, e.g., squared returns. If μ_j is the mean of component j's density, (1) generalizes to $\epsilon_t = \mu_{\Delta_t} + H_{\Delta_t,t}^{1/2}\xi_t$, and we impose $\mu_k = -\sum_{j=1}^{k-1}(\pi_{j,\infty}/\pi_{k,\infty})\mu_j$ in order to make sure that $\{\epsilon_t\}$ is a zero mean process. The normal mixture GARCH model was introduced in the univariate setting by Haas, Mittnik, and Paolella (2004a) and further considered, for example, by Alexander and Lazar (2006), Bauwens, Preminger, and Rombouts (2006), Haas, Mittnik, and Mizrach (2006), and Bauwens and Rombouts (2007). Multivariate extensions of the model are investigated in Bauwens, Hafner, and Rombouts (2006), and Haas, Mittnik, and Paolella (2006).

2.2 Estimation Issues

For estimation purposes, the general vech representation as given by (3) is not directly applicable, and parameter constraints are required in order to guarantee positive definiteness of all conditional covariance matrices. Such a parametrization is provided by the BEKK model of Engle and Kroner (1995) which specifies the covariance matrices as

$$H_{jt} = A_{0j}^{\star}A_{0j}^{\star\prime} + \sum_{\ell=1}^{L}\sum_{i=1}^{q} A_{ij,\ell}^{\star}\epsilon_{t-i}\epsilon_{t-i}' A_{ij,\ell}^{\star\prime} + \sum_{\ell=1}^{L}\sum_{i=1}^{p} B_{ij,\ell}^{\star} H_{j,t-i} B_{ij,\ell}^{\star\prime}, \quad j = 1, \ldots, k, \tag{5}$$

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 (5) has been estimated, the matrices A_{ij} and B_{ij} of the vech model (3) can be recovered via

$$A_{ij} = \sum_{\ell=1}^{L} D_M^{+}(A_{ij,\ell}^{\star} \otimes A_{ij,\ell}^{\star})D_M, \quad i = 1, \ldots, q, \quad j = 1, \ldots, k, \tag{6}$$

and similarly 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 addition, while, for $L = 1$ in (5), which is the standard choice in practice, the BEKK model already involves fewer parameters than the unrestricted vech form, further simplifications can be obtained by assuming that the $A_{ij}^{\star}$ and $B_{ij}^{\star}$ in (5) are diagonal matrices, and we will do so below in Section 3. As noted by Ding and Engle (2001), the diagonal BEKK model is equivalent to a restricted diagonal vech model, where, if $L = p = q = 1$ in (5), the conditional covariance matrices can be written as

$$H_{jt} = A^\star_{0j} A^{\star\prime}_{0j} + (a_{1j} a'_{1j}) \odot (\epsilon_{t-1} \epsilon'_{t-1}) + (b_{1j} b'_{1j}) \odot H_{j,t-1}, \quad j = 1, \ldots, k, \tag{7}$$

where $a_{1j} = A^\star_{1j} \mathbf{1}_M$ and $b_{1j} = B^\star_{1j} \mathbf{1}_M$ are $M \times 1$ vectors, $j = 1, \ldots, k$, and $\mathbf{1}_M$ is an M–dimensional column of ones. Representation (7) follows from the identity $D(A \odot B)E = (DAE) \odot B$ for conformable matrices A, B, D, and E, with D and E diagonal (cf. Horn and Johnson, 1991, Lemma 5.1.2), and the fact that $\mathbf{1}_M \mathbf{1}'_M$ is an $M \times M$ matrix of ones. Clearly specification (7) imposes some strong restrictions on the cross–dynamics. However, as noted by Bauwens, Laurent, and Rombouts (2006), although diagonal BEKK models are, due to these restrictions, not suitable if volatility transmission is the object under study, "they usually do a good job in representing the dynamics of variances and covariances". A recent application of the parametrization (7) in the context of dynamic correlations is provided by Cappiello, Engle, and Sheppard (2006).

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. In our application in Section 3, we will use the diagonal BEKK representation as given by (7).

2.3 Dynamic Properties

In this section, we derive conditions for the existence of and develop expressions for the unconditional overall and regime–specific covariance matrices, the unconditional fourth moment matrix (and hence kurtosis), and the dynamic correlation structure of the squares of an MMSG$(p, q; k)$ process. The moment conditions are investigated in Section 2.3.1 and summarized in Proposition 3, and the autocorrelation structure will be studied in Section 2.3.2. However, we first introduce some further notation.

We will have to calculate conditional expectations of the vector h_t, given in (4), based on different information sets. In general, the information at time t consists of the values of the process up to time t, $\Psi_t := \{\eta_s : s \leq t\}$, and hence h_{t+1}, and a probability distribution $\pi_t = [\pi_{1t}, \ldots, \pi_{kt}]'$ over S. In addition, the regime history up to time t will be denoted by $\underline{\Delta}_t := \{\Delta_s : s \leq t\}$.

Furthermore, we denote as $\rho(A)$ the spectral radius of a square matrix A, i.e.,

$$\rho(A) := \max\{|z| : z \text{ is an eigenvalue of } A\}, \tag{8}$$

and we use the notation ${}_m e_i$ to denote the ith unit vector in $\mathbb{R}^m$.

2.3.1 Moment Conditions

To derive the moment conditions for the model defined by (1)–(3), we write the model in GARCH(1,1) form. To this end, we define $X_t = [h_t', \ldots, h_{t-p+1}', \eta_{t-1}', \ldots, \eta_{t-r+1}']'$, where $r = \max\{q, 2\}$. Thus, X_t is of dimension $N(kp + r - 1) =: Nm$, where $m := (kp + r - 1)$. Furthermore, let

$$\tilde{A}_0 = \begin{bmatrix} A_0 \\ 0_{N(m-k)\times 1} \end{bmatrix}, \quad \tilde{A}_1 = \begin{bmatrix} A_1 \\ 0_{Nk(p-1)\times N} \\ I_N \\ 0_{N(r-2)\times N} \end{bmatrix}, \quad \tilde{B}_1 = \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ \tilde{B}_{21} & \tilde{B}_{22} \end{bmatrix}, \tag{9}$$

where

$$\begin{aligned} \tilde{B}_{11} &= \begin{bmatrix} B_1 & \cdots & B_{p-1} & B_p \\ & I_{Nk(p-1)} & & 0_{Nk(p-1)\times Nk} \end{bmatrix}, \quad \tilde{B}_{12} = \begin{bmatrix} A_2 & \cdots & A_r \\ & 0_{Nk(p-1)\times N(r-1)} & \end{bmatrix}, \\ \tilde{B}_{21} &= 0_{N(r-1)\times Nkp}, \quad \tilde{B}_{22} = \begin{bmatrix} 0_{N\times N(r-2)} & 0_{N\times N} \\ I_{N(r-2)} & 0_{N(r-2)\times N}, \end{bmatrix}, \end{aligned} \tag{10}$$

and $A_2 = 0_{Nk\times N}$ if $q = 1$. Therefore, we can write

$$X_t = \tilde{A}_0 + \tilde{A}_1 \eta_{t-1} + \tilde{B}_1 X_{t-1}. \tag{11}$$

We term (11) a "GARCH(1,1)" representation because X_t, just as h_t, is deterministic with respect to the information set at time $t-1$; thus, the formal structure of (11) resembles that of a GARCH(1,1), and methods similar to those developed for the basic GARCH(1,1) model can be employed to investigate the dynamic properties of equation (11).[1] We may also note that we could as well let $r = q$, so that, if $q = 1$, $X_t = [h_t', \ldots, h_{t-p+1}']'$. However, in the Markov–switching GARCH framework, it will turn out that inclusion of η_{t-1} in the state vector X_t greatly simplifies the computation of the unconditional moments of the process, although this blows up the state equation somewhat.

In the following analysis, we will make use of results of Francq and Zakoïan (2005) and Hafner (2003), which we state as Lemmas 1 and 2, respectively. To state Lemma 1, define the τ–step transition probabilities $p_{ij}^{(\tau)} := p(\Delta_t = j|\Delta_{t-\tau} = i)$, $i, j \in S$, as given by the elements

[1] Comte and Lieberman (2000) obtain a fourth moment condition for the standard multivariate GARCH(p, q) model using a state–space representation which they also term a "GARCH(1,1)" representation. It is, however, actually an ARMA(1,1) representation, as it is based on the innovations $u_t = \eta_t - h_t$ rather than the squared process η_t directly. ARMA representations, which are frequently adopted for the analysis of standard GARCH processes (e.g., Hafner, 2003; Zadrozny, 2005; Karanasos and Kim, 2006; and Haas, Mittnik, and Paolella, 2006), are not suitable for GARCH models subject to Markov switching.

of P^{τ}. Consider the matrix

$$\mathbb{P}_f^{(\tau)} = \begin{bmatrix} p_{11}^{(\tau)} f(1) & p_{21}^{(\tau)} f(1) & \cdots & p_{k1}^{(\tau)} f(1) \\ p_{12}^{(\tau)} f(2) & p_{22}^{(\tau)} f(2) & \cdots & p_{k2}^{(\tau)} f(2) \\ \vdots & \vdots & & \vdots \\ p_{1k}^{(\tau)} f(k) & p_{2k}^{(\tau)} f(k) & \cdots & p_{kk}^{(\tau)} f(k) \end{bmatrix} \tag{12}$$

for any function $f : S \mapsto \mathcal{M}_{n \times n'}(\mathbb{R})$, where $\mathcal{M}_{n \times n'}(\mathbb{R})$ denotes the space of real $n \times n'$ matrices, and positive integers τ, n, and n'. When $\tau = 1$, we drop the superscript and define $\mathbb{P}_f := \mathbb{P}_f^{(1)}$.

Lemma 1 *(Francq and Zakoïan, 2005, Lemma 1) Let $f : S \mapsto \mathcal{M}_{n \times n}(\mathbb{R})$, and $g : S \mapsto \mathcal{M}_{n \times n'}(\mathbb{R})$. Then, for $\tau > 0$, and $h > \tau$,*

$$E[f(\Delta_t) f(\Delta_{t-1}) \cdots f(\Delta_{t-\tau+1}) g(\Delta_{t-\tau}) | \Delta_{t-h}] = (\mathbf{1}_k' \otimes I_n) \mathbb{P}_f^{\tau} \mathbb{P}_g^{(h-\tau)} ({}_k e_{\Delta_{t-h}} \otimes I_{n'}),$$

where $\mathbf{1}_k$ is a k–dimensional column of ones, and ${}_k e_j$, $j \in S$, is the jth unit vector in $\mathbb{R}^k$.

If g does not depend on the prevailing regime, i.e., $g(1) = \cdots = g(k)$, we have

$$\begin{aligned} \mathrm{E}[f(\Delta_t) f(\Delta_{t-1}) \cdots f(\Delta_{t-\tau+1}) g | \Delta_{t-h}] &= (\mathbf{1}_k' \otimes I_n) \mathbb{P}_f^{\tau} (P^{h-\tau} \otimes g)({}_k e_{\Delta_{t-h}} \otimes I_{n'}) \\ &= (\mathbf{1}_k' \otimes I_n) \mathbb{P}_f^{\tau} [(P^{h-\tau}\ {}_k e_{\Delta_{t-h}}) \otimes g] \\ &= (\mathbf{1}_k' \otimes I_n) \mathbb{P}_f^{\tau} (\pi_{t-\tau} \otimes g). \end{aligned} \tag{13}$$

Lemma 2 *(Hafner, 2003, Theorem 1) For an M–dimensional normally distributed random vector x with zero mean and covariance matrix H,*[2] *we have*

$$vec\{E[vech(xx') vech(xx')']\} = G_M vec(hh'), \tag{14}$$

where $h = vech(H)$,

$$G_M = 2(L_M \otimes D_M^+)(I_M \otimes K_{MM} \otimes I_M)(D_M \otimes D_M) + I_{N^2}, \tag{15}$$

and $N := M(M+1)/2$ is the number of independent elements in H. Matrices L_M, D_M, D_M^+, and K_{MM} are defined in Appendix A.

Now we have, from (11),

$$\begin{aligned} \mathrm{E}(X_t | \Psi_{t-2}, \underline{\Delta}_{t-1}) &= \tilde{A}_0 + [\tilde{A}_1 ({}_m e'_{\Delta_{t-1}} \otimes I_N) + \tilde{B}_1] X_{t-1} \\ &= \tilde{A}_0 + [({}_m e'_{\Delta_{t-1}} \otimes \tilde{A}_1) + \tilde{B}_1] X_{t-1}, \end{aligned} \tag{16}$$

[2] Actually, Hafner (2003) considered the more general class of spherical distributions which includes the normal as a special case. Hafner's (2003) result for the normal distribution is based on earlier work of Magnus and Neudecker (1979).

where ${}_me_j$ is the jth unit vector in $\mathbb{R}^m$, and $m = kp + r - 1$.[3] Similarly,

$$\begin{aligned}
\mathrm{E}[\mathrm{vec}(X_tX_t')|\Psi_{t-2},\underline{\Delta}_{t-1}] &= \tilde{A}_0 \otimes \tilde{A}_0 + 2\tilde{N}_{Nm}[(\tilde{B}_1 \otimes \tilde{A}_0) + ({}_me'_{\Delta_{t-1}} \otimes \tilde{A}_1 \otimes \tilde{A}_0)]X_{t-1} \\
&\quad +(\tilde{A}_1 \otimes \tilde{A}_1)\mathrm{E}[\mathrm{vec}(\eta_{t-1}\eta_{t-1}')|\Psi_{t-2},\underline{\Delta}_{t-1}] \\
&\quad +2\tilde{N}_{Nm}\mathrm{E}[\mathrm{vec}(\tilde{A}_1\eta_{t-1}X_{t-1}'\tilde{B}_1{}')|\Psi_{t-2},\underline{\Delta}_{t-1}] \\
&\quad +(\tilde{B}_1 \otimes \tilde{B}_1)\mathrm{vec}(X_{t-1}X_{t-1}), \qquad (17)
\end{aligned}$$

where we used the identity (A.1) in Appendix A to compactify this expression. The expectations involved in (17) can be evaluated as

$$\begin{aligned}
\mathrm{E}[\mathrm{vec}(\tilde{A}_1\eta_{t-1}X_{t-1}'\tilde{B}_1{}')|\Psi_{t-2},\underline{\Delta}_{t-1}] &= \mathrm{vec}[\tilde{A}_1\mathrm{E}(\eta_{t-1}|\Psi_{t-2},\underline{\Delta}_{t-1})X_{t-1}'\tilde{B}_1{}'] \\
&= \mathrm{vec}[\tilde{A}_1({}_me'_{\Delta_{t-1}} \otimes I_N)X_{t-1}X_{t-1}'\tilde{B}_1{}'] \\
&= (\tilde{B}_1 \otimes_m e'_{\Delta_{t-1}} \otimes \tilde{A}_1)\mathrm{vec}(X_{t-1}X_{t-1}'),
\end{aligned}$$

and, applying Lemma 2,

$$\begin{aligned}
\mathrm{E}[\mathrm{vec}(\eta_{t-1}\eta_{t-1}')|\Psi_{t-2},\underline{\Delta}_{t-1}] &= G_M\mathrm{vec}[({}_me'_{\Delta_{t-1}} \otimes I_N)X_{t-1}X_{t-1}'({}_me_{\Delta_{t-1}} \otimes I_N)] \\
&= G_M({}_me'_{\Delta_{t-1}} \otimes I_N \otimes_m e'_{\Delta_{t-1}} \otimes I_N)\mathrm{vec}(X_{t-1}X_{t-1}').
\end{aligned}$$

We define

$$Y_t = \begin{bmatrix} X_t \\ \mathrm{vec}(X_tX_t') \end{bmatrix}, \quad d = \begin{bmatrix} \tilde{A}_0 \\ \tilde{A}_0 \otimes \tilde{A}_0 \end{bmatrix}, \quad C(j) = \begin{bmatrix} C_{11}(j) & 0_{Nm\times(Nm)^2} \\ C_{21}(j) & C_{22}(j) \end{bmatrix}, \quad j \in S,$$

where

$$\begin{aligned}
C_{11}(j) &= {}_me_j' \otimes \tilde{A}_1 + \tilde{B}_1, \qquad (18) \\
C_{21}(j) &= 2\tilde{N}_{Nm}({}_me_j' \otimes \tilde{A}_1 \otimes \tilde{A}_0 + \tilde{B}_1 \otimes \tilde{A}_0), \\
C_{22}(j) &= (\tilde{A}_1 \otimes \tilde{A}_1)G_M({}_me_j' \otimes I_N \otimes_m e_j' \otimes I_N) \qquad (19) \\
&\quad +2\tilde{N}_{Nm}(\tilde{B}_1 \otimes_m e_j' \otimes \tilde{A}_1) + \tilde{B}_1 \otimes \tilde{B}_1, \quad j = 1, \dots, k.
\end{aligned}$$

Using these definitions, we can state Proposition 3. As in 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 3 *The MMSG($p, q; k$) process defined by (1)–(3) is covariance stationary if and only if $\rho(\mathbb{P}_{C_{11}}) < 1$, where $\mathbb{P}_{C_{11}}$ is defined by (12) and (18). Moreover, the unconditional fourth moment matrix $E(\eta_t\eta_t')$ exists and does not depend on initial conditions if and only if, in addition, $\rho(\mathbb{P}_{C_{22}}) < 1$, where $\mathbb{P}_{C_{22}}$ is defined by (12) and (19). Expressions for the unconditional second and fourth moments are given in Equations (26) and (28), respectively.*

[3] Note that $\tilde{A}_1({}_me'_{\Delta_{t-1}} \otimes I_N) = (1 \otimes \tilde{A}_1)({}_me'_{\Delta_{t-1}} \otimes I_N) =_m e'_{\Delta_{t-1}} \otimes \tilde{A}_1$.

To derive this result, we note that we can write (16) and (17) as

$$\mathrm{E}(Y_t|\Psi_{t-2},\underline{\Delta}_{t-1}) = d + C(\Delta_{t-1})Y_{t-1}. \tag{20}$$

Iterating (20) gives

$$\mathrm{E}(Y_t|\Psi_{t-\tau-1},\underline{\Delta}_{t-1}) = \sum_{i=0}^{\tau-1}\left(\prod_{j=1}^{i} C(\Delta_{t-j})\right) d + \left(\prod_{i=1}^{\tau} C(\Delta_{t-i})\right) Y_{t-\tau},$$

where $\prod_{j=1}^{0} C(\Delta_{t-j}) := I_{Nm+(Nm)^2}$. Now we apply Lemma 1 and (13) to obtain

$$\begin{aligned}\mathrm{E}(Y_t|\Psi_{t-\tau-1},\pi_{t-\tau-1}) &= \sum_{i=0}^{\tau-1}(\mathbf{1}_k'\otimes I_{Nm+(Nm)^2})\mathbb{P}_C^i(\pi_{t-i-1}\otimes d) \\ &\quad +(\mathbf{1}_k'\otimes I_{Nm+(Nm)^2})\mathbb{P}_C^\tau(\pi_{t-\tau-1}\otimes Y_{t-\tau}).\end{aligned} \tag{21}$$

It will be convenient to write (21) in a slightly different form. To this end, let Q be the $k(Nm+N^2m^2)\times k(Nm+N^2m^2)$ permutation matrix such that

$$Q(\pi_{t-\tau-1}\otimes Y_{t-\tau}) = \begin{bmatrix} \pi_{t-\tau-1}\otimes X_{t-\tau} \\ \pi_{t-\tau-1}\otimes \mathrm{vec}(X_{t-\tau}X_{t-\tau}') \end{bmatrix},$$

so that (21) can be written as

$$\begin{aligned}\mathrm{E}(Y_t|\Psi_{t-\tau-1},\pi_{t-\tau-1}) &= \sum_{i=0}^{\tau-1}(\mathbf{1}_k'\otimes I_{Nm+(Nm)^2})Q'(Q\mathbb{P}_CQ')^iQ(\pi_{t-i-1}\otimes d) \\ &\quad +(\mathbf{1}_k'\otimes I_{Nm+(Nm)^2})Q'(Q\mathbb{P}_CQ')^\tau Q(\pi_{t-\tau-1}\otimes Y_{t-\tau}) \\ &= \sum_{i=0}^{\tau-1}\mathbb{I}\tilde{\mathbb{P}}_C^i M_{t-i-1} + \mathbb{I}\tilde{\mathbb{P}}_C^\tau \Pi,\end{aligned} \tag{22}$$

where

$$\begin{aligned}\mathbb{I} &= \begin{bmatrix} \mathbf{1}_k'\otimes I_{Nm} & 0_{Nm\times k(Nm)^2} \\ 0_{(Nm)^2\times kNm} & \mathbf{1}_k'\otimes I_{(Nm)^2}\end{bmatrix}, \quad \tilde{\mathbb{P}}_C = \begin{bmatrix} \mathbb{P}_{C_{11}} & 0_{kNm\times k(Nm)^2} \\ \mathbb{P}_{C_{21}} & \mathbb{P}_{C_{22}}\end{bmatrix}, \\ M_t &= \begin{bmatrix} M_{1t} \\ M_{2t}\end{bmatrix} = \begin{bmatrix} \pi_t\otimes\tilde{A}_0 \\ \pi_t\otimes\tilde{A}_0\otimes\tilde{A}_0\end{bmatrix}, \quad \Pi = \begin{bmatrix}\Pi_1 \\ \Pi_2\end{bmatrix} = \begin{bmatrix} \pi_{t-\tau-1}\otimes X_{t-\tau} \\ \pi_{t-\tau-1}\otimes\mathrm{vec}(X_{t-\tau}X_{t-\tau}')\end{bmatrix}.\end{aligned}$$

From the block–triangular structure of $\tilde{\mathbb{P}}_C$, we have

$$\mathrm{E}(X_t|\Psi_{t-\tau-1},\pi_{t-\tau-1}) = (\mathbf{1}_k'\otimes I_{Nm})\mathbb{P}_{C_{11}}^\tau\Pi_1 + (\mathbf{1}_k'\otimes I_{Nm})\sum_{i=0}^{\tau-1}\mathbb{P}_{C_{11}}^i M_{1,t-i-1}. \tag{23}$$

The first term on the right–hand side of (23) tends to zero as $\tau\to\infty$, provided that $\rho(\mathbb{P}_{C_{11}})<1$. We can write the second term on the right–hand side of (25) as

$$\begin{aligned}\sum_{i=0}^{\tau-1}\mathbb{P}_{C_{11}}^i(\pi_{t-i-1}\otimes\tilde{A}_0) &= \sum_{i=0}^{\tau-1}\mathbb{P}_{C_{11}}^i(\pi_\infty\otimes\tilde{A}_0) + \sum_{i=0}^{\tau-1}\mathbb{P}_{C_{11}}^i((\pi_{t-i-1}-\pi_\infty)\otimes\tilde{A}_0) \\ &= \sum_{i=0}^{\tau-1}\mathbb{P}_{C_{11}}^i M_{1,\infty} + \sum_{i=0}^{\tau-1}\mathbb{P}_{C_{11}}^i((P^{\tau-i}-P_\infty)\otimes I_{Nm})(\pi_{t-\tau-1}\otimes\tilde{A}_0),\end{aligned}$$

where $P_\infty := \lim_{\tau\to\infty} P^\tau = \pi_\infty \mathbf{1}_k'$. We have $P^{\tau-i} - P_\infty = (P - P_\infty)^{\tau-i}$, and, as P is irreducible and aperiodic (primitive), the matrix $P - P_\infty$ has all its roots strictly inside the unit circle (Moustakides, 1999).[4] Moreover, it is well–known that, if A is a square matrix, then for any $\varepsilon > 0$ there is a nonsingular matrix Q such that $||QAQ^{-1}||_2 \leq \rho(A) + \varepsilon$, where $\|\cdot\|_2$ denotes the spectral norm, and we can write $||A^i||_2 = ||Q^{-1}(QAQ^{-1})^i Q||_2 \leq ||Q^{-1}||_2 ||QAQ^{-1}||_2^i ||Q||_2$. We also observe that, for any τ, $||\pi_{t-\tau-1} \otimes \tilde{A}_0||_2 = \|\pi_{t-\tau-1}\|_2 \|\tilde{A}_0\|_2 \leq \|\tilde{A}_0\|_2$. Thus, if $\rho(\mathbb{P}_{C_{11}}) < 1$, we can find a ζ satisfying $\rho(\mathbb{P}_{C_{11}}) < \zeta < 1$, and an η satisfying $\rho(P - P_\infty) < \eta < 1$, and $\eta \neq \zeta$, such that, for an appropriately defined constant R, we can write

$$\begin{aligned} &\left\| \sum_{i=0}^{\tau-1} \mathbb{P}_{C_{11}}^i ((P^{\tau-i} - P_\infty) \otimes I_{Nm})(\pi_{t-\tau-1} \otimes \tilde{A}_0) \right\|_2 \\ &\leq \sum_{i=0}^{\tau-1} ||\mathbb{P}_{C_{11}}||_2^i \cdot ||((P - P_\infty) \otimes I_{Nm})||_2^{\tau-i} \cdot ||\pi_{t-\tau-1} \otimes \tilde{A}_0||_2 \\ &\leq R \sum_{i=0}^{\tau-1} \zeta^i \eta^{\tau-i} = \frac{R\eta}{\eta - \zeta}(\eta^\tau - \zeta^\tau) \overset{\tau\to\infty}{\longrightarrow} 0. \end{aligned} \tag{24}$$

Therefore, if $\rho(\mathbb{P}_{C_{11}}) < 1$,

$$\lim_{\tau\to\infty} \sum_{i=0}^{\tau-1} \mathbb{P}_{C_{11}}^i M_{1,t-i-1} = \sum_{i=0}^{\infty} \mathbb{P}_{C_{11}}^i M_{1,\infty} = (I_{kNm} - \mathbb{P}_{C_{11}})^{-1} M_{1,\infty},$$

which does not depend on the initial conditions. On the other hand, if $\rho(\mathbb{P}_{C_{11}}) \geq 1$, $\mathbb{P}_{C_{11}}^\tau$ will not tend to a zero matrix as $\tau \to \infty$, and the first term on the right–hand side of (23) will not converge to a finite limit that is independent of the initial conditions. Consequently, a necessary and sufficient condition for $\{\epsilon_t\}$ being covariance stationary is $\rho(\mathbb{P}_{C_{11}}) < 1$, and, in this case,

$$\mathrm{E}(X_t) = \lim_{\tau\to\infty} \mathrm{E}(X_t | \Psi_{t-\tau-1}, \pi_{t-\tau-1}) = (\mathbf{1}_k' \otimes I_{Nm})(I_{kNm} - \mathbb{P}_{C_{11}})^{-1} M_{1,\infty}. \tag{25}$$

The unconditional covariance matrix of ϵ_t can then be extracted from

$$\mathrm{E}(\eta_t) = ({}_m e_{kp+1}' \otimes I_N)\mathrm{E}(X_t). \tag{26}$$

By a similar analysis, it follows from (22) that $\mathrm{E}(X_t X_t')$ is finite and does not depend on the initial conditions if and only if, in addition, $\rho(\mathbb{P}_{C_{22}}) < 1$. In this case,

$$\mathrm{E}[\mathrm{vec}(X_t X_t')] = (\mathbf{1}_k' \otimes I_{(Nm)^2})(I_{k(Nm)^2} - \mathbb{P}_{C_{22}})^{-1}[M_{2,\infty} + \mathbb{P}_{C_{21}}(I_{kNm} - \mathbb{P}_{C_{11}})^{-1} M_{1,\infty}], \tag{27}$$

[4] Also not that, for square matrices A and B, $A^n \otimes B^n = (A \otimes B)^n$, and $\rho(A \otimes B) = \rho(A)\rho(B)$. Therefore, $(P^{\tau-i} - P_\infty) \otimes I_{Nm} = (P - P_\infty)^{\tau-i} \otimes I_{Nm} = ((P - P_\infty) \otimes I_{Nm})^{\tau-i}$, and $\rho((P - P_\infty) \otimes I_{Nm}) = \rho(P - P_\infty)$. Alternatively, we can use the result that $\|A \otimes B\|_2 = \|A\|_2 \|B\|_2$, and $\|I_{Nm}\|_2 = 1$ (Langville and Stewart, 2004).

and the unconditional fourth moment matrix of ϵ_t can be extracted from

$$\mathrm{E}(\eta_t\eta_t') = ({}_m e_{kp+1}' \otimes I_N)\mathrm{E}(X_t X_t')({}_m e_{kp+1} \otimes I_N). \tag{28}$$

Alternatively, in case of existence, the unconditional moments can be calculated using a more direct approach, which will be useful in Section 2.3.2 when computing the autocorrelation matrices of the squared process. This method is similar to the approach devised by Francq and Zakoïan (2001) for the moments of the Markov–switching ARMA model. We observe, from (11), that

$$\pi_{j,\infty}\mathrm{E}(X_t|\Delta_{t-1} = j) = \pi_{j,\infty}\tilde{A}_0 + \tilde{A}_1\pi_{j,\infty}\mathrm{E}(\eta_{t-1}|\Delta_{t-1} = j) + \tilde{B}_1\pi_{j,\infty}\mathrm{E}(X_{t-1}|\Delta_{t-1} = j),$$

where

$$\begin{aligned}\mathrm{E}(\eta_{t-1}|\Delta_{t-1} = j) &= \mathrm{E}[\mathrm{E}(\eta_{t-1}|\Psi_{t-2}, \Delta_{t-1} = j)|\Delta_{t-1} = j] \\ &= ({}_m e_j' \otimes I_N)\mathrm{E}(X_{t-1}|\Delta_{t-1} = j),\end{aligned}$$

and, using $\pi_{j,\infty}p(\Delta_{t-2} = i|\Delta_{t-1} = j) = \pi_{i,\infty}p(\Delta_{t-1} = j|\Delta_{t-2} = i) = \pi_{i,\infty}p_{ij}$,

$$\begin{aligned}\pi_{j,\infty}\mathrm{E}(X_{t-1}|\Delta_{t-1} = j) &= \sum_{i=1}^{k}\pi_{j,\infty}p(\Delta_{t-2} = i|\Delta_{t-1} = j)\mathrm{E}(X_{t-1}|\Delta_{t-1} = j \cap \Delta_{t-2} = i) \\ &= \sum_{i=1}^{k} p_{ij}\pi_{i,\infty}\mathrm{E}(X_{t-1}|\Delta_{t-2} = i),\end{aligned}$$

where the second equation uses that the expectation of X_{t-1} is independent of Δ_{t-1} once Δ_{t-2} is given (cf. Francq and Zakoïan, 2005, Lemma 3). Thus,

$$\begin{aligned}\pi_{j,\infty}\mathrm{E}(X_t|\Delta_{t-1} = j) &= \pi_{j,\infty}\tilde{A}_0 + \sum_{i=1}^{k} p_{ij}({}_m e_j' \otimes \tilde{A}_1 + \tilde{B}_1)\pi_{i,\infty}\mathrm{E}(X_{t-1}|\Delta_{t-2} = i), \\ j &= 1, \ldots, k.\end{aligned} \tag{29}$$

Define the $kNm \times 1$ vector $U = [\pi_{1,\infty}\mathrm{E}(X_t|\Delta_{t-1} = 1)', \ldots, \pi_{k,\infty}\mathrm{E}(X_t|\Delta_{t-1} = k)']'$. Equation (29) implies

$$U = \pi_\infty \otimes \tilde{A}_0 + \mathbb{P}_{C_{11}}U, \tag{30}$$

from which we recover (25). We can use Equation (30) to compute the unconditional covariance matrix of ϵ_t in regime j, $j = 1, \ldots, k$, as

$$\mathrm{E}(\eta_t|\Delta_t = j) = \pi_{j,\infty}^{-1}({}_{km}e_{pk+1+(j-1)m}' \otimes I_N)(I_{kNm} - \mathbb{P}_{C_{11}})^{-1}(\pi_\infty \otimes \tilde{A}_0), \tag{31}$$

Now define, similar to U, V to be the $k(Nm)^2 \times 1$ vector with elements $\pi_{j,\infty}\mathrm{E}[\mathrm{vec}(X_t X_t')|\Delta_{t-1} = j]$, $j = 1, \ldots, k$. An argument similar to the one leading to (30) shows that

$$V = \pi_\infty \otimes \tilde{A}_0 \otimes \tilde{A}_0 + \mathbb{P}_{C_{21}}U + \mathbb{P}_{C_{22}}V, \tag{32}$$

from which we recover (27).

It may be worth pointing out that the covariance stationarity condition $\rho(\mathbb{P}_{C_{11}}) < 1$ allows some regimes to be nonstationary, in the sense that the covariance stationarity condition for the single–component GARCH(p, q) process, i.e., $\det[z^n I_N - \sum_{i=1}^{n}(A_{ij} + B_{ij})z^{n-i}] \neq 0$ for $|z| \geq 1$, where $n = \max\{p, q\}$, $A_{ij} = 0_{N\times N}$ for $i > q$, and $B_{ij} = 0_{N\times N}$ for $i > p$, is not satisfied for some regimes.[5] Nevertheless, the overall process can still be stationary, as long as the persistencies (the "staying probabilities" p_{jj}) and unconditional probabilities of the corresponding regimes are sufficiently small. This parallels the situation in the univariate case (Francq, Roussignol, and Zakoïan, 2001; Wong and Li, 2001; Haas, Mittnik, and Paolella, 2004a,b; Alexander and Lazar, 2006; Liu, 2006; and Abramson and Cohen, 2007) and will be empirically illustrated in Section 3. Note that, for a given set of regime–specific GARCH parameters, $\rho(\mathbb{P}_{C_{11}})$ depends on both the unconditional regime probabilities as well as the persistence of the regimes. To illustrate, consider the simplest case of a univariate MSG(0,1;2) process, where, in obvious notation, $\sigma_{jt}^2 = \alpha_{0j} + \alpha_{1j}\epsilon_{t-1}^2$, $\alpha_{0j} > 0$, $\alpha_{1j} \geq 0$, $j = 1, 2$. The elements of the transition matrix (12) can be written in terms of the stationary probability of Regime 1, $\pi_{1,\infty} = 1 - \pi_{2,\infty}$, and the degree of regime–persistence, $\delta = p_{11} + p_{22} + 1$, i.e., $p_{11} = \pi_{1,\infty} + \delta(1 - \pi_{1,\infty})$, and $p_{22} = 1 - \pi_{1,\infty} + \delta\pi_{1,\infty}$. Using the notation $s := \pi_{1,\infty}\alpha_{11} + (1 - \pi_{1,\infty})\alpha_{12}$, and $s' := \pi_{1,\infty}\alpha_{12} + (1 - \pi_{1,\infty})\alpha_{11}$, we have $\rho(\mathbb{P}_{C_{11}}) = (s + \delta s' + \sqrt{(s + \delta s')^2 - 4\delta\alpha_{11}\alpha_{12}})/2$, so that

$$\left.\frac{d\rho(\mathbb{P}_{C_{11}})}{d\delta}\right|_{d\pi_{1,\infty}=0} = \frac{1}{2}\left(s' + \frac{(s + \delta s')s' - 2\alpha_{11}\alpha_{12}}{\sqrt{(s + \delta s')^2 - 4\delta\alpha_{11}\alpha_{12}}}\right). \tag{33}$$

The sign of (33) is not immediately obvious, because $(s + \delta s')s' - 2\alpha_{11}\alpha_{12}$ may be negative. However, straightforward calculations show that, if this is the case, positivity of (33) is equivalent to $ss' - \alpha_{11}\alpha_{12} = \pi_{1,\infty}(1 - \pi_{1,\infty})(\alpha_{11} - \alpha_{12})^2 > 0$, which, by the irreducibility assumption, i.e., $\pi_{1,\infty} \in (0, 1)$, holds as long as $\alpha_{11} \neq \alpha_{12}$. The intuition behind this result is that, the larger δ, relative to the (fixed) unconditional regime probabilities, the longer the chain tends to stay in the high–volatility regime, so that large shocks can accumulate. On the other hand, given the transition matrix, we can work out those combinations of parameters α_{11} and α_{12} giving rise to (covariance) stationary and nonstationary processes. Figure 1 shows the stationarity regions for three transition matrices, given by

$$P_1 = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \quad \text{and} \quad P_3 = \begin{bmatrix} 0.25 & 0.75 \\ 0.75 & 0.25 \end{bmatrix}, \tag{34}$$

respectively. In all three processes, both regimes have a stationary probability of 0.5, but $\delta > 0$, $\delta = 0$ (normal mixture GARCH), and $\delta < 0$ for P_1, P_2, and P_3, respectively. It is also

[5] For this condition, see Bollerslev and Engle (1993), and Engle and Kroner (1995).

easily confirmed that, as in Figure 1, the stationarity border is concave if $\delta > 0$, linear if $\delta = 0$, and convex if $\delta < 0$.

We finally note that, using the results of Balestra and Holly (1990), the analysis of the present section could be extended to even moments of any order. Such an extension was provided by Liu (2006) for the univariate model. However, the resulting expressions rapidly become unmanageable, as the number terms to be evaluated is explosive. Thus, in view of the fact that such higher moments are of minor interest in most financial applications, we do not attempt this.

2.3.2 Autocorrelation Function of the Squared Process

Now we turn to the computation of the sequence of autocorrelation matrices of the squared process. We have, for $\tau \geq 1$,

$$\begin{aligned}\pi_{j,\infty}\mathrm{E}(X_t X'_{t-\tau}|\Delta_{t-1}=j) &= \tilde{A}_0 \pi_{j,\infty}\mathrm{E}(X'_{t-\tau}|\Delta_{t-1}=j) + \tilde{A}_1 \pi_{j,\infty}\mathrm{E}(\eta_{t-1}X'_{t-\tau}|\Delta_{t-1}=j) \\ &\quad + \tilde{B}_1 \pi_{j,\infty}\mathrm{E}(X_{t-1}X'_{t-\tau}|\Delta_{t-1}=j),\end{aligned}$$

where

$$\begin{aligned}\pi_{j,\infty}\mathrm{E}(X'_{t-\tau}|\Delta_{t-1}=j) &= \sum_{i=1}^{k} \pi_{j,\infty} p(\Delta_{t-\tau-1}=i|\Delta_{t-1}=j)\mathrm{E}(X'_{t-\tau}|\Delta_{t-1}=j \cap \Delta_{t-\tau-1}=i) \\ &= \sum_{i=1}^{k} p_{ij}^{(\tau)} \pi_{i,\infty}\mathrm{E}(X'_{t-\tau}|\Delta_{t-\tau-1}=i),\end{aligned}$$

$$\begin{aligned}\pi_{j,\infty}\mathrm{E}(\eta_{t-1}X'_{t-\tau}|\Delta_{t-1}=j) &= \pi_{j,\infty}\mathrm{E}[\mathrm{E}(\eta_{t-1}|\Psi_{t-2},\Delta_{t-1}=j)X'_{t-\tau}|\Delta_{t-1}=j] \\ &= \pi_{j,\infty}({}_m e'_j \otimes I_N)\mathrm{E}(X_{t-1}X'_{t-\tau}|\Delta_{t-1}=j) \\ &= \sum_{i=1}^{k} p_{ij}({}_m e'_j \otimes I_N)\pi_{i,\infty}\mathrm{E}(X_{t-1}X'_{t-\tau}|\Delta_{t-2}=i),\end{aligned}$$

and

$$\pi_{j,\infty}\mathrm{E}(X_{t-1}X'_{t-\tau}|\Delta_{t-1}=j) = \sum_{i=1}^{k} p_{ij}\pi_{i,\infty}\mathrm{E}(X_{t-1}X'_{t-\tau}|\Delta_{t-2}=i).$$

Therefore,

$$\begin{aligned}\pi_{j,\infty}\mathrm{E}(X_t X'_{t-\tau}|\Delta_{t-1}=j) &= \tilde{A}_0 \sum_{i=1}^{k} p_{ij}^{(\tau)} \pi_{i,\infty} E(X'_{t-\tau}|\Delta_{t-\tau-1}=i) \qquad (35) \\ &\quad + \sum_{i=1}^{k} p_{ij}({}_m e'_j \otimes \tilde{A}_1 + \tilde{B}_1)\pi_{i,\infty}\mathrm{E}(X_{t-1}X'_{t-\tau}|\Delta_{t-2}=i), \\ j &= 1,\ldots,k.\end{aligned}$$

Now let $W(\tau)$ be the $kNm \times Nm$ matrix obtained by replacing $\pi_{j,\infty}\mathrm{E}[\mathrm{vec}(X_t X_t')|\Delta_{t-1} = j]$ with $\pi_{j,\infty}\mathrm{E}(X_t X_{t-\tau}'|\Delta_{t-1} = j)$ in V (Equation (32)), and let $\tilde{U}$ be the $k \times Nm$ matrix where $\pi_{j,\infty}\mathrm{E}(X_t|\Delta_{t-1} = j)$ is replaced with $\pi_{j,\infty}\mathrm{E}(X_t'|\Delta_{t-1} = j)$ in U (Equation (30)). Then we can write (35) as

$$W(\tau) = (P \otimes \tilde{A}_0)P^{\tau-1}\tilde{U} + \mathbb{P}_{C_{11}} W(\tau - 1), \quad \tau \geq 1, \tag{36}$$

$W(0)$ is obtained by reshaping V, defined in (32),

$$\mathrm{E}(X_t X_{t-\tau}') = (\mathbf{1}_k' \otimes I_{Nm})W(\tau),$$

and

$$\begin{aligned}\mathrm{E}(\eta_t \eta_{t-\tau}') &= ({}_m e_{kp+1}' \otimes I_N)\mathrm{E}(X_t X_{t-\tau}')({}_m e_{kp+1} \otimes I_N) \\ &= (\mathbf{1}_k' \otimes_m e_{kp+1}' \otimes I_N)W(\tau)({}_m e_{kp+1} \otimes I_N).\end{aligned} \tag{37}$$

The autocovariance function at lag τ, $\Gamma(\tau)$, is then given by

$$\Gamma(\tau) = \mathrm{E}(\eta_t \eta_{t-\tau}') - \mathrm{E}(\eta_t)\mathrm{E}(\eta_t'), \tag{38}$$

and the autocorrelation matrices, $R(\tau)$, can be calculated by

$$R(\tau) = D^{-1/2}\Gamma(\tau)D^{-1/2}, \tag{39}$$

where $D = I_N \odot \Gamma(0)$, and $\Gamma(0) = \mathrm{E}(\eta_t \eta_t') - \mathrm{E}(\eta_t)\mathrm{E}(\eta_t')$.

The solution of (36) is

$$\begin{aligned}W(\tau) &= \sum_{i=0}^{\tau-1} \mathbb{P}_{C_{11}}^i (P \otimes \tilde{A}_0)P^{\tau-1-i}\tilde{U} + \mathbb{P}_{C_{11}}^\tau W(0) \qquad (40) \\ &= \sum_{i=0}^{\tau-1} \mathbb{P}_{C_{11}}^i (P \otimes \tilde{A}_0)P_\infty \tilde{U} + \sum_{i=0}^{\tau-1} \mathbb{P}_{C_{11}}^i (I_k \otimes \tilde{A}_0)(P^{\tau-i} - P_\infty)\tilde{U} + \mathbb{P}_{C_{11}}^\tau W(0) \\ &= \sum_{i=0}^{\tau-1} \mathbb{P}_{C_{11}}^i (\pi_\infty \otimes \tilde{A}_0)\mathrm{E}(X_t') + \sum_{i=0}^{\tau-1} \mathbb{P}_{C_{11}}^i (I_k \otimes \tilde{A}_0)(P - P_\infty)^{\tau-i}\tilde{U} + \mathbb{P}_{C_{11}}^\tau W(0),\end{aligned}$$

where the last line follows from $(P \otimes \tilde{A}_0)P_\infty \tilde{U} = (P \otimes \tilde{A}_0)\pi_\infty \mathrm{E}(X_t') = (\pi_\infty \otimes \tilde{A}_0)\mathrm{E}(X_t')$. Therefore, from the analysis in (24), $\lim_{\tau\to\infty} W(\tau) = U\mathrm{E}(X_t')$, and $\lim_{\tau\to\infty} \mathrm{E}(X_t X_{t-\tau}') = (\mathbf{1}_k' \otimes I_{Nm}) \lim_{\tau\to\infty} W(\tau) = \mathrm{E}(X_t)\mathrm{E}(X_t')$, so that $\lim_{\tau\to\infty} \Gamma(\tau) = 0_{N\times N}$. Equation (40) can also be used to obtain a closed-from solution for $\Gamma(\tau)$ in the case of two regimes, i.e., $k = 2$, which is of particular relevance for the applications. In this case, $P^\tau = F_1 + \delta^\tau F_2$, where $-1 < \delta = p_{11} + p_{22} - 1 < 1$,

$$F_1 = P_\infty = \begin{bmatrix} \pi_{1,\infty} & \pi_{1,\infty} \\ \pi_{2,\infty} & \pi_{2,\infty} \end{bmatrix}, \quad F_2 = \begin{bmatrix} \pi_{2,\infty} & -\pi_{1,\infty} \\ -\pi_{2,\infty} & \pi_{1,\infty} \end{bmatrix},$$

$\pi_{1,\infty} = (1-p_{22})/(2-p_{11}-p_{22})$, and $\pi_{2,\infty} = 1-\pi_{1,\infty}$. If $\det(\delta I_{kNm} - \mathbb{P}_{C_{11}}) \neq 0$, then we get, after a few computations,

$$\begin{aligned} W(\tau) &= (I_{kNm} - \mathbb{P}_{C_{11}}^{\tau})(I_{kNm} - \mathbb{P}_{C_{11}})^{-1}(P \otimes \tilde{A}_0)F_1\tilde{U} \\ &\quad + (\delta^{\tau} I_{kNm} - \mathbb{P}_{C_{11}}^{\tau})(\delta I_{kNm} - \mathbb{P}_{C_{11}})^{-1}(P \otimes \tilde{A}_0)F_2\tilde{U} + \mathbb{P}_{C_{11}}^{\tau} W(0), \end{aligned}$$

and our final expression for the autocovariance function in the case of two regimes, provided that $\det(\delta I_{kNm} - \mathbb{P}_{C_{11}}) \neq 0$, is

$$\begin{aligned} \Gamma(\tau) &= \mathrm{E}(\eta_t \eta_{t-\tau}') - \mathrm{E}(\eta_t)\mathrm{E}(\eta_t') \\ &= (\mathbf{1}_k' \otimes_m e_{kp+1}' \otimes I_N) \left\{ \mathbb{P}_{C_{11}}^{\tau}[W(0) - U\mathrm{E}(X_t)'] \right. \\ &\quad \left. + (\delta^{\tau} I_{kNm} - \mathbb{P}_{C_{11}}^{\tau})(\delta I_{kNm} - \mathbb{P}_{C_{11}})^{-1}(P \otimes \tilde{A}_0)F_2\tilde{U} \right\} ({}_m e_{kp+1} \otimes I_N). \end{aligned} \tag{41}$$

This completes our discussion of the fourth moment structure of the MMSG$(p,q;k)$ process. Simpler expressions for the moment conditions and the moments, including the autocovariance function, can be obtained for the MNMG$(p,q;k)$ process discussed at the end of Section 2.1, which is a special case of the MMSG$(p,q;k)$ process. Such results appear in Bauwens, Hafner, and Rombouts (2006) for the MNMG$(1,1;k)$ model, where the fourth moment structure is investigated only for the symmetric case, i.e., with all component densities having zero mean. The more general asymmetric MNMG$(p,q;k)$ case is studied in Haas, Mittnik, and Paolella (2006) using a VARMA representation. While the extension to the asymmetric MNMG$(p,q;k)$ process is conceptually straightforward, it involves lengthy calculations, and, therefore, we refer to Haas, Mittnik, and Paolella (2006) for the relevant formulas.

We finally show how the GARCH(1,1) representation (11) can be used to obtain the fourth–moment structure of the classic univariate GARCH(p,q) model, giving rise to expressions which are much simpler and easier to implement than the ingenious but rather complicated formulas derived by He and Teräsvirta (1999) and Karanasos (1999). In this case, and assuming normally distributed innovations, the condition for the existence of $\mathrm{E}(\epsilon_t^4)$ is $\rho(C_{22}) < 1$, where

$$C_{22} = 3(\tilde{A}_1 e_1' \otimes \tilde{A}_1 e_1') + 2\tilde{N}_{p+r-1}(\tilde{B}_1 \otimes \tilde{A}_1 e_1') + \tilde{B}_1 \otimes \tilde{B}_1,$$

e_1' is the first unit vector in $\mathbb{R}^{p+r-1}$, and $\tilde{N}_n$ is defined in Appendix A. This requirement bears a striking resemblance to the classic condition $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ for the fourth moment of the univariate GARCH(1,1) model (under conditional normality) to exist (Bollerslev, 1986). This similarity provides a further rationale for referring to (11) as the GARCH(1,1) representation of a GARCH(p,q) process. For the autocorrelation structure, recursive substitution in $\mathrm{E}(X_t|\Psi_{t-2}) = \tilde{A}_0 + C_{11}X_{t-1}$, where $C_{11} = \tilde{A}_1 e_1' + \tilde{B}_1$, gives $\mathrm{E}(X_t|\Psi_{t-\tau-1}) =$

$\sum_{i=0}^{\tau-1} C_{11}^i \tilde{A}_0 + C_{11}^\tau X_{t-\tau} = \mathrm{E}(X_t) + C_{11}^\tau [X_{t-\tau} - \mathrm{E}(X_t)]$, where $\mathrm{E}(X_t) = (I_{p+r-1} - C_{11})^{-1}\tilde{A}_0$. Substituting in $\mathrm{E}(X_t X_{t-\tau}') = \mathrm{E}[\mathrm{E}(X_t|\Psi_{t-\tau-1})X_{t-\tau}']$ and subtracting $\mathrm{E}(X_t)\mathrm{E}(X_t')$ shows that an expression for the autocovariance function is

$$\mathrm{cov}(\epsilon_t^2, \epsilon_{t-\tau}^2) = e_{p+1}' C_{11}^\tau [\mathrm{E}(X_t X_t') - \mathrm{E}(X_t)\mathrm{E}(X_t')] e_{p+1},$$

where e_{p+1} is the $(p+1)$th unit vector in $\mathbb{R}^{p+r-1}$, and $\mathrm{E}(X_t X_t')$ is easily deduced from the development in Section 2.3.1. Using a slight modification of the procedure sketched in this paragraph, the moment structure of a variant of the asymmetric power GARCH(p,q) process of Ding, Granger, and Engle (1993), as considered in Ling and McAleer (2002), can also be derived, and, using Lemma 2, the extension to the multivariate GARCH(p,q) model in vech form is likewise straightforward.[6]

3 Application to International Stock Market Returns

We now provide an application of the model developed in Section 2 to international stock markets. We consider discrete[7] dollar–denominated weekly (Thursday to Thursday) percentage returns of the S&P500, FTSE, and DAX indices over the period from January 1984 to August 2005, a sample of $T = 1127$ observations.[8] We thus assume the perspective of an US–investor not hedging currency risk. We denote the return vector at time t by $r_t = [r_{1t}, r_{2t}, r_{3t}]'$, where r_{1t}, r_{2t}, and r_{3t} are the time–t returns of the S&P500, the FTSE, and the DAX, respectively.

A few descriptive statistics of the three series, along with the Jarque–Bera test for normality and Engle's (1982) Lagrange multiplier test for ARCH, are summarized in Table 1. As the latter test has been derived under conditional normality and may not be robust to "outliers", the values reported in the last three columns of Table 1 are calculated by excluding the return observation from October 15 to October 22, 1987.[9] All three series display considerable excess kurtosis, and the Jarque–Bera test strongly rejects normality in all cases. Likewise, the ARCH test rejects the null of no ARCH effects for all three indices.

Although not reported, graphical identification tools as well as Ljung–Box statistics do not

[6] Details are available from the authors on request.

[7] Due to limited liability, it is clear that (mixed) normality of discrete returns can only be an approximation to the return distribution. However, use of continuously compounded returns would complicate the derivation of optimal portfolios in Section 3.2.

[8] All data have been obtained from Datastream.

[9] While this does not affect qualitatively the results for the S&P500 and the DAX, the ARCH–LM test does not reject homoskedasticity of the FTSE returns in case this observation is not excluded from the sample.

suggest the presence of noteworthy autocorrelation. Thus, we model the return series as

$$r_t = \nu + \epsilon_t, \tag{42}$$

where ν is a constant mean vector, and $\{\epsilon_t\}$ follows a multivariate (diagonal BEKK) GARCH process.

We finally note that the stock return series display a considerable degree of comovement, with pairwise correlations ranging from approximately 0.525 to 0.6.

3.1 Estimation Results and Regime–Evidence

Four different versions of the general GARCH model developed in Section 2 with $p = q = 1$ are considered, assuming a diagonal BEKK structure as given by (5) and (7). Namely, this includes a single–component model, which corresponds to $k = 1$ in (1)–(3), and which is just the standard multivariate Normal–GARCH process, which we denote by MNG(1,1). Additionally, we estimate three two–component models ($k = 2$). The first of these is the MMSG$(1, 1; 2)$ process (1)–(3) without a priori restrictions on the transition matrix P, while the second and third are the symmetric and asymmetric multivariate normal mixture GARCH processes discussed in the last paragraph of Section 2.1, which we denote by MNMG$_s$(1,1;2) and MNMG(1,1;2), respectively.[10]

Table 2 reports likelihood–based goodness–of–fit measures for the models and their rankings with respect to these criteria, i.e., the value of the maximized log–likelihood function, and the BIC criterion of Schwarz (1978). To provide evidence for the presence of both regime–switching and GARCH effects in the data, Table 2 also reports the results for the corresponding models with constant (within–regime) covariance matrices, that is, with $A_1 = 0_{N\times N}$ $(0_{2N\times N})$ and $B_1 = 0_{N\times N}$ $(0_{2N\times 2N})$ in (4), i.e., for models MNG(0,0), MMSG(0,0;2), MNMG$_s$(0,0;2), and MNMG(0,0;2). In particular, it is well–known that even the basic Markov–switching model with constant within–regime parameters generates volatility clustering (Ryden, Teräsvirta, and Åsbrink, 1998; and Timmermann, 2000), and it may be the case that the conditional heteroskedasticity accommodated by the switching of regimes is sufficient to capture the second–order dependencies observed in the data, thus rendering the GARCH structures superfluous.[11] However, the results reported in Table 2 point in the opposite direction. A noteworthy implication of Table 2 is that *all* the (within–regime) homoskedastic models are inferior to *all*

[10] The models are estimated by conditional maximum likelihood with all the (component) covariance matrices being initialized by the sample covariance matrix.

[11] Positive evidence for this conjecture is presented in Ang and Bekaert (2002) for *monthly* returns.

the GARCH models; in particular, even the standard single–component MNG(1,1) specification dominates the basic two–component Markov–switching process MMSG(0,0;2) according to both log–likelihood and BIC, although it has less parameters.

Within the class of GARCH models, the single–regime MNG(1,1) ranks lowest according to the BIC, while the MMSG(1,1;2) ranks best.[12] While MMSG(1,1;2) dominates both the symmetric and asymmetric independent switching (normal mixture) models, the comparison between $\mathrm{MNMG}_s(1,1;2)$ and $\mathrm{MNMG}(1,1;2)$ produces less definite results. Although the former performs better according to the BIC, two times the difference in log–likelihood between the models is $2 \times (7131.7 - 7127.8) = 7.8$, so that a likelihood ratio test for symmetry with three degrees of freedom gives rise to a p–value of 0.050, which makes the discrimination between the models on the basis of their likelihood values somewhat vague.

Summarizing the evidence presented in Table 2, we conclude that both persistent regimes as well as regime–specific GARCH structures appear to be important features of the joint distribution of the international stock returns under study.

The parameter estimates for models MMSG(1,1;2), MNMG_s(1,1;2), and MNMG(1,1;2) are reported, in this order, in Tables 3–5, with the regimes being ordered with respect to a declining (stationary) regime probability, i.e., $\pi_{1,\infty} > \pi_{2,\infty}$. The equations driving the dynamics of the covariance matrices are reported in the form (7), which is the representation most amenable to interpretation.[13] In addition, we report the regime–specific measures for persistence in volatility, 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 (6), as well as, in the last row of the tables, the largest eigenvalues of the matrices $\mathbb{P}_{C_{11}}$ and $\mathbb{P}_{C_{22}}$ defined in Proposition 3, which provide information about the existence of the unconditional second and fourth moments, respectively. Furthermore, the impled unconditional overall and regime–specific covariance and correlation matrices are shown in Table 6, where, for purpose of comparison, the single–component MNG(1,1) model is also included.[14]

In discussing the parameter estimates reported in Tables 3–5, we first draw attention to a pattern common to all three specifications. All the mixture models identify two compo-

[12] With regard to the comparison between single–regime and multi–regime models, it may be worthwhile to mention that, in the literature on mixture models, there is some evidence that the BIC does a good job in discriminating between models with a different number of components (see McLachlan and Peel, 2000, Chap. 6, for a survey and references).

[13] Standard errors of functions of estimated quantities are obtained via the delta method.

[14] The term "unconditional correlation matrix" refers to the correlations calculated from the unconditional covariance matrices, which are given by (31). Due to the nonlinearity involved in the calculation of correlation coefficients, this is not identical to the unconditional expectation of the conditional correlation matrix, the expression for which is unknown.

nents with distinctly different covariance processes. More precisely, the first regime, i.e., the component with the larger (unconditional) regime probability, is stationary in the sense that $\rho(A_{1j} + B_{1j}) < 1$, and it can be characterized as the low–volatility regime. The latter statement can be inferred from Table 6, but this is also reflected in the fact that, in Tables 3–5, $A^{\star}_{01}A^{\star\prime}_{01} < A^{\star}_{02}A^{\star\prime}_{02}$ holds elementwise for all three models. The second regime is nonstationary in the sense that $\rho(A_{2j}+B_{2j}) > 1$, and it represents the high–volatility regime which occurs less frequently (approximately 10% of the weeks). However, all estimated models are stationary and have finite fourth moments, because, for all models, both $\rho(\mathbb{P}_{C_{11}})$ and $\rho(\mathbb{P}_{C_{22}})$ are below unity.[15] Furthermore, Table 6 shows that, in the multi–regime models, correlations are higher in turbulent markets, a phenomenon which has been extensively discussed in the literature on international portfolio diversification (see, e.g., Longin and Solnik, 1995; Ramchand and Susmel, 1998; Ang and Bekaert, 2002, 2004; Butler and Joaquin, 2002; Forbes and Rigobon, 2002; Guidolin and Timmermann, 2005b; Baur, 2006; and Cappiello, Engle, and Sheppard, 2006). Most of the differences are moderate, however, and may not be significant statistically (see Ang and Bekaert, 2002, for similar results). An informal comparison of Table 6 with columns 3–5 of Table 1 also shows that all models fit the unconditional covariance/correlation structure reasonably well.

Comparing the general Markov–switching process MMSG(1,1;2) in Table 3 with the independent switching models in Tables 4 and 5 reveals the reason for the superior fit of the former, as reported in Table 2. Namely, it appears that the regimes are persistent in the sense that $\delta = p_{11} + p_{22} - 1 > 0$, or, equivalently, $p_{jj} > \pi_{j,\infty}$, $j = 1, 2$, which cannot be captured by the normal mixture models. Although the persistence is relatively weak,[16] it is statistically significant, and it implies that, if we are in the low– (high–) variance regime currently, the probability of being in the low– (high–) variance regime in the following week will be larger than if the current regime were the high– (low–) variance regime. If regimes are persistent, it is clear that this persistence should be incorporated into the model, because this means that regimes are predictable, and such predictability can be exploited for asset allocation and risk management purposes.

With regard to the asymmetric MNMG model, as reported in Table 5, we note that the low–volatility regime is associated with positive mean returns, while the means of the high–volatility regime are negative, which is in line with the results reported in Ang and Bekaert

[15] The same is true for model MNG(1,1), where $\rho(\mathbb{P}_{C_{11}}) = 0.995$, and $\rho(\mathbb{P}_{C_{22}}) = 0.991$.

[16] For purpose of comparison, in model MMSG(0,0;2), i.e., the Markov–switching model with constant covariance matrices, we have $p_{11} = 0.958$ and $p_{22} = 0.927$, so that $\delta = 0.885$. Clearly part of the persistence captured by the GARCH effects in model MMSG(1,1;2) is accommodated by the persistence of the regimes in this case.

(2002) for monthly returns on international stock indices. This implies that the regimes can be characterized as bull and bear markets, respectively. However, and this also conforms to the findings of Ang and Bekaert (2002), given the relatively large standard errors of the regime–specific mean vectors, the economic significance of this classification is unclear. This is in accordance with the ambiguous results emerging from the likelihood–based comparison of models MNMG$_s$(1,1;2) and MNMG(1,1;2), as discussed earlier in this section. Consequently, restricting the means across regimes to be equal, as in the MMSG model, is not likely to be a serious constraint in the present application.

Finally, Figures 2 and 3 show the empirical autocorrelations of the squared residuals for the three series, along with their theoretical counterparts implied by the four estimated GARCH models. Note, however, that, just as the ARCH–LM test statistics in Table 1, the empirical quantities have been computed by excluding the return from October 15 to October 22, 1987. With the exception of model MMSG(1,1;2), the theoretical autocorrelations of the fitted models, when compared to their empirical counterparts, tend to be too low at the beginning. An inspection of Equation (41) for the theoretical autocovariance function of two–regime MMSG process shows that, compared to the case where $\delta = 0$, it offers a greater degree of flexibility due to an additional component which decays at rate δ. In the present situation, with $\delta = 0.479$, this component accounts for the fast decay of the autocorrelation function at the first lags observed in Figure 2, thus capturing the large low–order autocorrelations of the FTSE and the DAX.

3.2 Application to Portfolio Selection

3.2.1 Volatility Regimes and Portfolio Selection

In the MMSG$(p, q; k)$ process, the one–period–ahead distribution of the M–dimensional return vector at time t, r_t, is a k–component multivariate mixture of normals with vector of mixing weights $\pi_t = [\pi_{1t}, \ldots, \pi_{kt}]'$, i.e., its density is given by

$$f(r_t|\Psi_{t-1}) = \sum_{j=1}^{k} \pi_{jt}\phi(r_t; \mu_j, H_{jt}), \tag{43}$$

where $\phi(\cdot; \mu, H)$ denotes the normal density with mean μ and covariance matrix H, and μ_j and H_{jt}, $j = 1, \ldots, k$, are the component means and (conditional) component covariance matrices, respectively. In (43), we allow for regime–specific means in order to include the asymmetric MNMG$(p, q; k)$ process discussed at the end of Section 2.1. The mean and the covariance matrix of (43) can easily be deduced from the properties of the normal distribution and are given by

$$\mathrm{E}(r_t|\Psi_{t-1}) = \sum_{j=1}^{k} \pi_{jt}\mu_j, \tag{44}$$

and

$$\mathrm{cov}(r_t|\Psi_{t-1}) = \sum_{j=1}^{k} \pi_{jt}(H_{jt} + \mu_j\mu_j') - \left(\sum_{j=1}^{k} \pi_{jt}\mu_j\right)\left(\sum_{j=1}^{k} \pi_{jt}\mu_j\right)', \tag{45}$$

respectively. If r_t has a k–component multivariate normal mixture distribution as given by (43), then the return on a portfolio formed from these assets, $r_{p,t}$, i.e., $r_{p,t} = w'r_t$, where w is an $M \times 1$ vector of portfolio weights, has a k–component univariate normal mixture distribution, i.e., it has density

$$f(r_{p,t}|\Psi_{t-1}) = \sum_{j=1}^{k} \frac{\pi_{jt}}{\sqrt{2\pi}\tilde{\sigma}_{jt}} \exp\left\{-\frac{(r_{p,t} - \tilde{\mu}_j)^2}{2\tilde{\sigma}_{jt}^2}\right\}, \tag{46}$$

where $\tilde{\mu}_j = w'\mu_j$, and $\tilde{\sigma}_{jt} = \sqrt{w'H_{jt}w}$, $j = 1, \ldots, k$.

When applied to financial return data, it is usually found that the market regimes differ mainly in their variances and covariances, while the component means are rather close in value, and often their differences are not significant statistically. This reflects the observation that excess kurtosis is a much more pronounced (and ubiquitous) property of asset returns than skewness, and was, for the data under study, also reported in Section 3.1. Thus, in the following discussion, and in order to concentrate on the impact of volatility regimes, we shall assume that, in (43), $\mu_1 = \cdots = \mu_k =: \mu$, which implies a symmetric return distribution and is referred to as a *scale mixture*. This in turn implies that, in (46), for a given portfolio weight vector, w, $\tilde{\mu}_1 = \cdots = \tilde{\mu}_k = w'\mu =: \mu_p$, and, by (45), $\mathrm{var}(r_p) =: \sigma_p^2 = \sum_j \pi_j \tilde{\sigma}_j^2$, where, for simplicity of notation, we temporarily drop the time index of the variables.

Do investors dislike the uncertainty with respect to the next period's volatility regime which is reflected in (43), and what are the consequences for optimal portfolio choice? To provide a partial answer to this question, we compare an investment under (43) with the same investment under a single Gaussian distribution with the same mean and covariance matrix, as given by (44) and (45). For normally distributed wealth, W, we can define a $\mu - \sigma^2$ preference function, $V(\mu, \sigma^2) := \mathrm{E}[U(W)] = \int_{-\infty}^{\infty} U(W)\phi(W, \mu, \sigma^2)dW = \int_{-\infty}^{\infty} U(\sigma W + \mu)\phi(W)dW$, where U is the investor's expected utility function, and $\phi(W) := \phi(W; 0, 1)$ is the standard normal's density. Using the fact that $V(\mu, \sigma^2)$ satisfies the differential equation $\partial^2 V/\partial\mu^2 = 2(\partial V/\partial\sigma^2)$, Chipman (1973) inferred that $V_{\sigma^2\sigma^2} = \partial^2 V/(\partial\sigma^2)^2 = (\partial^4 V/\partial\mu^4)/4 = \mathrm{E}[U''''(\sigma W + \mu)]/4$. Therefore, if $U'''' < 0$, $V(\mu, \sigma^2)$ is concave in σ^2, and

$$\sum_{j=1}^{k} \pi_j V(\mu_p, \tilde{\sigma}_j^2) < V\left(\mu_p, \sum_{j=1}^{k} \pi_j \tilde{\sigma}_j^2\right) = V(\mu_p, \sigma_p^2), \tag{47}$$

so that the investor dislikes regime uncertainty, or, in decision–theoretic terms, the normal distribution fourth–order stochastically dominates any $\mu - \sigma^2$ equivalent scale mixture. Is $U'''' < 0$ reasonable economically? The answer is in the affirmative. A negative fourth derivative is known to be a necessary condition for decreasing absolute prudence, which is usually deemed plausible (Kimball, 1990; Gollier, 2001). If we take this for granted, then the interpretation of (47) is that investors prefer a certain state of the world, with a given variance, over the "veil of ignorance" with respect to the prevailing volatility regime, i.e., investors would like to rule out the possibility of the high–variance states of the world with their above–average volatility. When it comes to portfolio selection, we expect that, when confronted with mixed normally distributed asset returns, investors will, compared to a Gaussian distribution with the same mean and covariance matrix, allocate a larger fraction of wealth to those assets with relatively favorable diversification properties in the adverse states of the world, i.e., they want to hedge against the occurrence of the adverse states of the market. At the same time, reducing the variance in the high–volatility states means that the conditional mixture distribution becomes less fat–tailed, because for $|r_{p,t}| \to \infty$ the difference between any two mixture densities of the form (46) with $\tilde{\mu_1} = \cdots = \tilde{\mu}_k$ is dominated by the mixture component with the greatest variance.

To construct out–of–sample portfolios for the models under consideration, we first reestimate all of them using roughly the first ten years of data, i.e., the first 500 observations. The parameter vectors thus obtained are then used to predict the return density of the next four weeks and to derive optimal portfolios, where we restrict our analysis to the simplest case of one–period–ahead all–equity portfolios, as, for example, in Jondeau and Rockinger (2005,2006). Subsequently, the model parameters are updated (approximately) every month (i.e., four weeks) using the most recent information in the sample and employing an expanding window of data. In this manner, we obtain, for each model, and given our sample size $T = 1127$, 627 realized one–week–ahead out–of–sample portfolio returns.

To select portfolios, we assume that the expected utility function, U, with initial wealth fixed at 100,[17] can reasonably be approximated by

$$U(r_{p,t}) = -\exp\{-c r_{p,t}\}, \quad c > 0, \tag{48}$$

where c is the coefficient of constant absolute risk aversion (CARA), and we consider (48) for

[17] This is because percentage returns are used, i.e., $r_{it} = 100 \times (P_{it} - P_{i,t-1})/P_{i,t-1}$, $i = 1, 2, 3$, where P_{it} is the ith index level at time t (denominated in dollars).

different values of c.[18] Using (48), a (conditional) Gaussian investor will solve

$$\max_{w_t} w_t'\mu_t - \frac{c}{2} w_t' H_t w_t \quad \text{s.t.} \quad \mathbf{1}_3' w_t = 1 \text{ and } w_t \geq 0, \quad t = 501, \ldots, 1127, \tag{49}$$

where $w_t = [w_{US,t}, w_{UK,t}, w_{Ger,t}]'$ is the vector of portfolio weights, whereas, in view of (46), an investor assuming that returns follow a (conditional) normal mixture distribution will maximize

$$\mathrm{E}[U(r_{p,t})|\Psi_{t-1}] = -\sum_{j=1}^{k} \pi_{jt} \exp\left\{-cw_t'\mu_{jt} + \frac{c^2}{2} w_t' H_{jt} w_t\right\} \quad \text{s.t.} \quad \mathbf{1}_3' w_t = 1 \text{ and } w_t \geq 0, \tag{50}$$

$t = 501, \ldots, 1127$, where, in (49) and (50), the mean vectors depend on t, because the parameter estimates are updated every month, and the π_{jt}'s to be used in (50) are the one–step–ahead regime forecasts originating from Hamilton's (1989, 1994) filter algorithm.[19]

It must be stressed that the portfolio choice experiment conducted herein is necessarily of an illustrative nature. For example, in practice, the forecasts of the mean returns would not be based on a model as simple as (42) (see, for example, Ang and Bekaert, 2002; and Guidolin and Timmermann, 2005a,b, for a discussion of useful predictor variables), nor would they be based solely on statistical methods in all cases. Also, from a practical viewpoint, investigation of problems more general than the construction of single–period all–equity portfolios deserves attention. A more detailed study of portfolio selection under switching volatility regimes is beyond the scope of this paper, and, consequently, we will pay particular attention to the model's capability of providing accurate portfolio return predictive densities rather than to genuine portfolio performance measures.

3.2.2 An Illustrative Example

Before we investigate the distributional properties of the out–of–sample portfolio returns arising from the various GARCH models under consideration, we single out a characteristic example to illustrate the impact of volatility regimes on portfolio choice in line with our reasoning surrounding Equation (47).[20] Namely, on August 4, 2005, i.e., at the beginning of the last week

[18] CARA may be an undesirable property of an expected utility function, as it is often argued since Arrow (1971) that risk aversion is decreasing in wealth. In this case, we can use mixtures of CARA utility functions of the form $U(W) = -\sum_{j=1}^{n} a_j e^{-c_j W}$, where $a_j, c_j > 0$, $j = 1, \ldots, n$, and $c_i \neq c_j$ for $i \neq j$. By Theorem 5 of Pratt (1964), such functions exhibit strictly decreasing risk aversion, and they still admit a closed–form expression for expected utility under (mixed) normality, leading to numerically rather tractable optimization problems. We will not pursue this here, however.

[19] The functions `quadprog` and `fmincon` in Matlab 6.5 are used to carry out the optimizations in (49) and (50), respectively. To choose the starting values for the mixture investors, we evaluate (50) over a fine grid of portfolio weights and pick the weight vector which gives the highest expected utility.

[20] Note, however, that we abstain from accounting for estimation risk, i.e., we do not test for statistical significance of the differences in the optimal portfolio weights.

of our out–of–sample period, the (conditional) Gaussian investor, relying on model MNG(1,1), maximizes expected utility with respect to the predictive density $r_t|\Psi_{t-1} \sim \mathrm{N}(\mu_t^N, H_t)$, where $\mu_t^N = [0.22, 0.19, 0.28]'$, and

$$\begin{aligned} \mathrm{vech}(H_t) &= [\sigma_{US}^2, \sigma_{US,UK}, \sigma_{US,Ger}, \sigma_{UK}^2, \sigma_{UK,Ger}, \sigma_{Ger}^2]' \\ &= [2.15, 1.40, 1.90, 2.97, 2.77, 5.11]', \end{aligned} \tag{51}$$

whereas the predictive density of the Markov–switching (MS) investor, employing model MMSG(1,1;2), is given by $r_t|\Psi_{t-1} \sim \pi_{1t}\mathrm{N}(\mu_t^{MS}, H_{1t}) + (1-\pi_{1t})\mathrm{N}(\mu_t^{MS}, H_{2t})$, where $\pi_{1t} = 0.88$, $\mu_t^{MS} = [0.24, 0.21, 0.33]'$, and

$$\begin{aligned} \mathrm{vech}(H_{1t}) &= [1.73, 0.94, 1.22, 2.09, 1.80, 3.52]', \\ \mathrm{vech}(H_{2t}) &= [6.02, 5.56, 8.59, 10.6, 10.9, 20.8]', \end{aligned} \tag{52}$$

implying, by (45), an overall conditional covariance matrix of

$$\mathrm{vech}(\pi_{1t}H_{1t} + (1-\pi_{1t})H_{2t}) = [2.23, 1.48, 2.08, 3.10, 2.86, 5.54]', \tag{53}$$

which is similar to (51).

The optimal portfolios for the Gaussian and the MS investors under CARA utility, as shown in the left and right plot of Figure 4, respectively, display some considerable differences. While the differences for the lower degrees of risk aversion, c, are negligible, the presence of volatility regimes becomes more important as c increases. Namely, inspection of (52) reveals that, in model MMSG(1,1;2), the lower variance of US equity, relative to the UK market, as observed in (53), is mainly due to its considerably smaller variance in the high–volatility regime, while both variances are more similar in Regime 1. However, as c grows, the desire to hedge against the high–volatility regime becomes more and more important for the MS investor, and so, in accordance with the discussion in Section 3.2.1, her portfolio converges to the global minimum variance portfolio (GMVP) of Regime 2, given by $w_{GMVP}^{MS}(\Delta_2) = [0.917, 0.083, 0.000]'$. The Gaussian investor, ignoring the presence of market regimes, only cares about the overall portfolio variance, and thus overestimates the benefits from holding UK equity. As c increases, her optimal portfolio converges to the GMVP associated with (51), i.e., $w_{GMVP}^N = [0.676, 0.324, 0.000]'$.

Although the similarity of (51) and (53) suggests that the differences between the optimal portfolios of both investors can mainly be attributed to the presence of volatility regimes, rather than to the differences in the overall conditional covariance matrices implied by models MNG(1,1) and MMSG(1,1;2), we have also considered an MS investor characterized by a $\mu-\sigma^2$

preference function of the form $V(\mu_p, \sigma_p^2) = \mu_p - (c/2)\sigma_p^2$. This cannot be derived from an expected utility framework, but it helps in disentangling the two aforementioned sources of discrepancy between the optimal portfolio weights of the Gaussian and the MS investors under CARA utility. It turns out that the optimal portfolios for the MS investor with mean–variance preferences essentially reproduce those of the Gaussian investor, confirming that the differences between the left and right panels of Figure 4 are mainly due to the regime–uncertainty inherent in the Markov–switching framework.

The example presented in this Section is typical for our application insofar as, on average, and in line with the results of Ang and Bekaert (2002), US equity has relatively favorable distributional properties in the high–volatility regime, as compared to UK and German stocks. Thus, the MS and mixture investors with higher risk aversion tend to hold a larger fraction of wealth in US equity than their Gaussian counterparts. Whether this can explain part of the home bias of US investors (see, e.g., Lewis, 1999) is an interesting question but requires a broader framework than that employed in the current investigation.

3.2.3 Out–of–sample Portfolio Results

In this section, we consider the distributional properties of the out–of–sample portfolio returns originating from the single– and multi–regime GARCH models and an investor characterized by the utility function (48). To better appreciate the performance of the models, Table 7, which is similar to Table 1, reports descriptive statistics of the joint return distribution over the out–of–sample period. Comparison of Tables 1 and 7 reveals some noteworthy differences. In particular, compared to the entire sample, the mean returns are somewhat smaller in the out–of–sample period, as are, with the exception of the DAX, the magnitudes of the skewness and kurtosis coefficients. However, normality is still strongly rejected for all indices, and the ARCH–LM test detects highly significant heteroskedasticity. Moreover, the correlations between the stock markets under study have been somewhat higher during the last years of our sample.

For selected values of c in (48), ranging from 0.01 to 1, summary statistics for the respective sequences of portfolio returns, $r_{p,t}$, $t = 501\ldots, 1127$, are documented in Tables 8 and 9. Further increasing c did not result in any notable differences compared to $c = 1$. Two types of statistics are reported. The first set of summary statistics, which is reported in Table 8, directly refers to properties of the $r_{p,t}$–series. Next to the usual moment–based summary measures, we follow de Goeij and Marquering (2004) and calculate the average realized utility, i.e.,

$$\overline{U(r_{p,t})} = \frac{1}{627}\sum_{t=501}^{1127} U(r_{p,t}) = \frac{1}{627}\sum_{t=501}^{1127} -\exp\{-cr_{p,t}\}, \tag{54}$$

where, in Table 7, we scale (54) by e^{-3c} for convenience of reporting.

For the second type of summary statistics, as motivated by the discussion in the last paragraph of Section 3.2.1 and presented in Table 9, we employ the technique proposed by Berkowitz (2001) to assess the quality of the predictive portfolio return distributions implied by the respective models, which is of great interest for risk management purposes. That is, we calculate the sequence of "realized" portfolio return distribution functions, $u_t := \widehat{F}(r_{p,t}|\Psi_{t-1})$, $t = 501, \ldots, 1127$, where $\widehat{F}(\cdot|\Psi_{t-1})$ is the conditional cumulative distribution function (cdf) of the portfolio return implied by the model under consideration. Subsequently, we apply a second transformation, namely,

$$\{z_t\} = \Phi^{-1}(\{u_t\}), \tag{55}$$

where Φ is the standard normal cdf. The sequence $\{z_t\}$ is iid N(0,1) if the underlying model is correct, and Berkowitz (2001) shows that inaccuracies in the predictive density will be preserved in the transformed data. Thus this transformation allows the use of moment–based normality tests for checking features such as correct specification of skewness and kurtosis. In addition, we apply the ARCH–LM test to (55) in order to judge whether the volatility dynamics are successfully captured by the fitted models.

Finally, we evaluate each model's performance in measuring the portfolio Value–at–Risk (VaR), a widely employed tool in risk management (see, e.g., Christoffersen and Pelletier, 2004). Briefly, for a given model, the VaR at level α for period t, denoted by $\mathrm{VaR}_t(\alpha)$, is defined by $\widehat{F}(\mathrm{VaR}_t(\alpha)|\Psi_{t-1}) = \alpha$. A *violation* or *hit* is said to occur at time t if $r_{p,t} < \mathrm{VaR}_t(\alpha)$. The empirical shortfall probability is $\widehat{\alpha} = x/T$, where x is the empirical shortfall frequency, and T is the number of forecasts evaluated. From both the risk management and the regulatory perspective, the main interest is whether the model's actual shortfall probability is *greater* than α. Therefore, the check whether $\widehat{\alpha}$ is significantly larger than α is conducted using a one–sided binomial test, where the p–values are calculated by $p = \sum_{i=x}^{T}\binom{T}{i}\alpha^i(1-\alpha)^{T-i}$. In our application, we consider the VaR levels $\alpha = 0.005, 0.01$, and 0.05.

We begin our discussion with the first set of summary statistics, as reported in Table 8. In agreement with intuition, for all models, investors with a higher degree of risk aversion realize a lower mean return and a smaller variance than the less risk–averse investors. For each c, the Markov–switching (MS) investor achieves the highest mean return, and particularly so when compared to the Gaussian investor. Given the caveats indicated in the last paragraph of Section 3.2.1, this observation should be interpreted with utmost care. Nevertheless, it may very well be the case that, in the presence of fat–tailed distributions, use of a (scale) mixture instead of a single normal distribution can help in estimating mean returns more accurately (cf. Aitkin and Tunnicliffe–Wilson, 1980). Note, however, that the MS investors with the lowest risk aversion, i.e., $c = 0.01$ or 0.025, who realize a particularly high mean return, also have to pay a price in form of a relatively large variance. The higher moments for the other investors (with $c > 0.025$) are not exceedingly different across the four models, although the MS investors exhibit a somewhat lower (negative) skewness. Interestingly, the realized skewness of the MS investors is uniformly lower than that of those using the asymmetric normal mixture GARCH model, although the latter explicitly allows for possible asymmetries in the return distribution. Comparing the average realized utilities, as defined in (54), we observe that, for each c, the MS investor is better off then her corresponding Gaussian and normal mixture look–alikes, which points to the existence of economic gains from accounting for *persistent* volatility regimes. Note that neither the absolute differences nor the percentage improvements between the realized utilities can be interpreted, because expected utility functions are unique only up to affine transformations (cf. Takayama, 1994, p. 267).

Next, we consider the portfolio density forecasts of the models, as summarized in Table 9. While the series $\{z_t\}$ defined in (55) of model MNG(1,1) display highly significant skewness and excess kurtosis for each coefficient of risk aversion, those of the MMSG(1,1;2) process do not show any significant deviations from normality for $c \leq 0.1$. For the larger c–values, there is still no excess kurtosis, but the zero skewness hypothesis is rejected at either the 5% or 10% level. The asymmetric MNMG(1,1;2) is the only model for which the transformed series (55) consistently do not exhibit any significant nonnormalities, while the symmetric MNMG(1,1;2) variant fails to pass the skewness test for all c except the lowest.

With respect to the conditional heteroskedasticity, the ARCH–LM test detects significant ARCH effects in the $\{z_t\}$ of the MNG(1,1) model for all c–values, while model MMSG(1,1;2) passes the test for all degrees of risk aversion with the exception of $c = 0.025$, where the ARCH–LM(5) test rejects at the 10% level. For the normal mixture GARCH models, and in particular the symmetric version, the null of a correctly specified volatility process is rejected more often than for the MS model, but they still do better than the single–regime specification.

Finally, turning to the evaluation of the portfolio VaR measures, we first note the outstanding performance of the asymmetric MNMG(1,1;2) process. For each c–value, this model accurately measures the VaR at the practically most important level $\alpha = 0.01$, and does also well for the other two VaR levels. The symmetric MNMG(1,1;2) model is less adequate, while model MMSG(1,1;2) provides reasonably accurate VaR measures for the lower levels $\alpha = 0.005$ and 0.01, but not for $\alpha = 0.05$. Lastly, and not surprisingly, the single–regime MNG(1,1) model apparently fails to capture the portfolio risk at the lower VaR levels but, interestingly, not so for $\alpha = 0.05$.

In summary, the asymmetric MNMG(1,1;2) process and model MMSG(1,1;2) seem to be acceptable with respect to the predictive density evaluations in Table (9), with MNMG(1,1;2) delivering superior VaR measures. At first glance, it may seem difficult to reconcile the favorable performance of the predictive densities generated by the asymmetric normal mixture GARCH model with the results of the utility–based comparison of Table (8), but, in this regard, it is convenient to recall the point made by West, Edison, and Cho (1993), that "utility and statistical measures may be dramatically different". Clearly Table 9 suggests that, in applications to VaR, asymmetries may be important to assess the risk in the lower tail.

4 Conclusions

Several extensions and modifications of the analysis conducted in this paper are worth exploring. While the diagonal BEKK structure, which is used in the present paper to specify the dynamics of variances and covariances, is parsimonious enough to be applicable to a relatively large number of assets, different specifications will be preferable for very high–dimensional problems. For example, in the introduction, we already mentioned the class of constant correlation models. As noted by Pelletier (2006), an appealing feature of such models when enriched with a multi–regime structure is that, although the within–regime correlations are constant, the overall correlation matrix will be time–varying. If the return vector to be modeled is of very high dimension, further simplification can be achieved by resorting to common correlation–type models for the within–regime correlation matrices (e.g., Elton, Gruber, and Padberg, 1976; and Kwan, 2006) which, in a static framework, have been found to exhibit favorable performance in predicting asset return correlations (e.g., Eun and Resnick, 1984). Clearly such extensions require development of appropriate estimation techniques.

Appendix

A The Commutation, Elimination, and Duplication Matrices

To conveniently write down the unconditional moments of the multivariate regime–switching GARCH model developed herein, 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. 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 make 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,

$$2\tilde{N}_n\text{vec}(A) = \text{vec}(A + A'), \tag{A.1}$$

which follows directly from its definition.

References

Abramson, A. and Cohen, I. (2007). On the Stationarity of Markov–switching GARCH Processes. Forthcoming in: *Econometric Theory*, 23.

Aitkin, M. and Tunnicliffe-Wilson, G. (1980). Mixture Models, Outliers, and the EM Algorithm. *Technometrics*, 22:325–331.

Alexander, C. (2001). *Market Models. A Guide to Financial Data Analysis.* John Wiley & Sons, Chichester.

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

Ang, A. and Bekaert, G. (2002). International Asset Allocation With Regime Shifts. *Review of Financial Studies*, 15:1137–1187.

Ang, A. and Bekaert, G. (2004). How Regimes Affect Asset Allocation. *Financial Analysts Journal*, 60:86–99.

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

Arrow, K. J. (1971). *Essays in the Theory of Risk-Bearing.* Markham, Chicago.

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.

Baur, D. (2006). Multivariate Market Association and its Extremes. *Journal of International Financial Markets, Institutions & Money*, 16:355–369.

Bauwens, L., Hafner, C. M., and Rombouts, J. V. K. (2006a). Multivariate Mixed Normal Conditional Heteroskedasticity. Forthcoming in: *Computational Statistics and Data Analysis.*

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

Bauwens, L., Preminger, A., and Rombouts, J. V. K. (2006c). Regime–switching GARCH Models. CORE Discussion Paper 2006/11, Center for Operations Research and Econometrics, Université Catholique de Louvain.

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

Berkowitz, J. (2001). Testing Density Forecasts, with Applications to Risk Management. *Journal of Business and Economic Statistics*, 19:465–474.

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

Bollerslev, T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *Review of Economics and Statistics*, 73:498–505.

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

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A Capital Asset Pricing Model with Time-varying Covariances. *Journal of Political Economy*, 96:116–131.

Butler, K. C. and Joaquin, D. C. (2002). Are the Gains from International Portfolio Diversification Exaggerated? The Influence of Downside Risk in Bear Markets. *Journal of International Money and Finance*, 21:981–1011.

Cai, J. (1994). A Markov Model of Switching–Regime ARCH. *Journal of Business and Economic Statistics*, 12:309–316.

Campbell, R., Koedijk, K., and Kofman, P. (2002). Increased Correlation in Bear Markets: A Downside Risk Perspective. *Financial Analysts Journal*, 58:87–94.

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.

Chipman, J. S. (1973). The Ordering of Portfolios in Terms of Mean and Variance. *Review of Economic Studies*, 40:167–190.

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

Comte, F. and Lieberman, O. (2000). Second–Order Noncausality in Multivariate GARCH Processes. *Journal of Time Series Analysis*, 21:535–557.

de Goeij, P. and Marquering, W. (2004). Modeling the Conditional Covariance Between Stock and Bond Returns: A Multivariate GARCH Approach. *Journal of Financial Econometrics*, 2:531–564.

Ding, Z. and Engle, R. F. (2001). Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing. *Academia Economic Papers*, 29:157–184.

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.

Dueker, M. J. (1997). Markov Switching in GARCH Processes and Mean–Reverting Stock–Market Volatility. *Journal of Business and Economic Statistics*, 15:26–34.

Elton, E. J., Gruber, M. J., and Padberg, M. W. (1976). Simple Criteria for Optimal Portfolio Selection. *Journal of Finance*, 31:13411357.

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

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

Eun, C. S. and Resnick, B. G. (1984). Estimating the Correlation Structure of International Share Prices. *Journal of Finance*, 39:1311–1324.

Forbes, K. J. and Rigobon, R. (2002). No Contagion, Only Interdependence: Measuring Stock Market Comovements. *Journal of Finance*, 57:2223–2261.

Francq, C., Roussignol, M., and Zakoïan, J.-M. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22:197–220.

Francq, C. and Zakoïan, J.-M. (2001). Stationarity of Multivariate Markov–Switching ARMA Models. *Journal of Econometrics*, 102:339–364.

Francq, C. and Zakoïan, J.-M. (2005). The L^2–structures of Standard and Switching–regime GARCH Models. *Stochastic Processes and their Applications*, 115:1557–1582.

Gollier, C. (2001). *The Economics of Risk and Time.* MIT Press, Cambridge, MA.

Gray, S. F. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime–Switching Process. *Journal of Financial Economics*, 42:27–62.

Guidolin, M. and Timmermann, A. (2005a). Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns. *Economic Journal*, 115:111–143.

Guidolin, M. and Timmermann, A. (2005b). International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences. Working Paper 2005–034, Federal Reserve Bank of St. Louis.

Guidolin, M. and Timmermann, A. (2006). An Econometric Model of Nonlinear Dynamics in the Joint Distribution of Stock and Bond Returns. *Journal of Applied Econometrics*, 21:1–22.

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. (2004a). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2:211–250.

Haas, M., Mittnik, S., and Paolella, M. S. (2004b). A New Approach to Markov–Switching GARCH Models. *Journal of Financial Econometrics*, 2:493–530.

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.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357–384.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, New Jersey.

Hamilton, J. D. and Susmel, R. (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics*, 64:307–333.

He, C. and Teräsvirta, T. (1999). Fourth Moment Structure of the GARCH(p, q) Process. *Econometric Theory*, 15:824–846.

He, C. and Teräsvirta, T. (2004). An Extended Constant Correlation GARCH Model and its Fourth-Moment Structure. *Econometric Theory*, 20:904–926.

Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press, Cambridge.

Jeantheau, T. (1998). Strong Consistency of Estimators for Multivariate ARCH Models. *Econometric Theory*, 14:70–86.

Jondeau, E. and Rockinger, M. (2005). Conditional Asset Allocation under Non–Normality: How Costly is the Mean–Variance Criterion? Research Paper 132, FAME International Center for Financial Asset Management and Engineering, University of Geneva.

Jondeau, E. and Rockinger, M. (2006). Optimal Portfolio Allocation under Higher Moments. *European Financial Management*, 12:29–55.

Karanasos, M. (1999). The Second Moment and the Autocovariance Function of the Squared Errors of the GARCH Model. *Journal of Econometrics*, 90:63–76.

Karanasos, M. and Kim, J. (2006). A Re-Examination of the Asymmetric Power ARCH Model. *Journal of Empirical Finance*, 13:113–128.

Kimball, M. S. (1990). Precautionary Saving in the Small and in the Large. *Econometrica*, 58:53–73.

Klaassen, F. (2002). Improving GARCH Volatility Forecasts with Regime–Switching GARCH. *Empirical Economics*, 27:363–394.

Kwan, C. C. Y. (2006). Some Further Analytical Properties of the Constant Correlation Model for Portfolio Selection. *International Journal of Theoretical and Applied Finance*, 9:1071–1091.

Langville, A. N. and Stewart, W. J. (2004). The Kronecker Product and Stochastic Automata Networks. *Journal of Computational and Applied Mathematics*, 167:429–447.

Lewis, K. K. (1999). Trying to Explain the Home Bias in Equities and Consumption. *Journal of Economic Literature*, 37:571–608.

Ling, S. and McAleer, M. (2002). Necessary and Sufficient Moment Conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) Models. *Econometric Theory*, 18:722–729.

Ling, S. and McAleer, M. (2003). Asymptotic Theory for a Vector ARMA-GARCH Model. *Econometric Theory*, 19:280–310.

Liu, J.-C. (2006). Stationarity of a Markov–switching GARCH Model. *Journal of Financial Econometrics*, 4:573–593.

Longin, F. and Solnik, B. (1995). Is the Correlation in International Equity Returns Constant: 1960-1990? *Journal of International Money and Finance*, 14:3–26.

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.

Marcucci, J. (2005). Forecasting Stock Market Volatility with Regime–Switching GARCH Models. *Studies in Nonlinear Dynamics & Econometrics*, 9(4):Article 6.

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

Moustakides, G. V. (1999). Extension of Wald's First Lemma to Markov Processes. *Journal of Applied Probability*, 36:48–59.

Pagan, A. R. and Schwert, G. W. (1990). Alternative Models for Conditional Stock Volatility. *Journal of Econometrics*, 45:267–290.

Pelletier, D. (2006). Regime Switching for Dynamic Correlations. *Journal of Econometrics*, 131:445–473.

Perez-Quiros, G. and Timmermann, A. (2001). Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities. *Journal of Econometrics*, 103:259–306.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, 32:122–136.

Ramchand, L. and Susmel, R. (1998). Volatility and Cross Correlation Across Major Stock Markets. *Journal of Empirical Finance*, 5:397–416.

Ryden, T., Teräsvirta, T., and Åsbrink, S. (1998). Stylized Facts of Daily Return Series and the Hidden Markov Model. *Journal of Applied Econometrics*, 13:217–244.

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

Takayama, A. (1994). *Analytical Methods in Economics*. Harvester Wheatsheaf, New York.

Timmermann, A. (2000). Moments of Markov Switching Models. *Journal of Econometrics*, 96:75–111.

Turner, C. M., Startz, R., and Nelson, C. R. (1989). A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market. *Journal of Financial Economics*, 25:3–22.

West, K. D., Edison, H. J., and Cho, D. (1993). A Utility-based Comparison of some Models of Exchange Rate Volatility. *Journal of International Economics*, 35:23–45.

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

Zadrozny, P. A. (2005). Necessary and Sufficient Restrictions for Existence of a Unique Fourth Moment of a Univariate GARCH(P, Q) Process. CESifo Working Paper 1505, Ifo Institute for Economic Research.

Table 1: Distributional properties of international stock market returns.

	mean	covariance/ correlation matrix			skew	kurt	JB	ARCH–LM(q)		
		S&P500	FTSE	DAX				$q=1$	$q=5$	$q=10$
S&P500	0.200	4.540	0.543	0.524	–0.686	7.625	1093 (0.000)	16.70 (0.000)	55.44 (0.000)	76.85 (0.000)
FTSE	0.200	2.848	6.070	0.595	–0.589	8.156	1314 (0.000)	18.71 (0.000)	33.98 (0.000)	40.59 (0.000)
DAX	0.258	3.324	4.366	8.882	–0.241	5.119	221.8 (0.000)	27.73 (0.000)	69.66 (0.000)	80.91 (0.000)

p–values are given in parentheses. "skew" denotes the moment–based coefficient of skewness, $\gamma = m_3/m_2^{3/2}$, and "kurt" 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, $\text{JB} = T\gamma^2/6 + T(\kappa-3)^2/24 \overset{asy}{\sim} \chi^2(2)$ (see, e.g., Alexander, 2001, p. 286). ARCH–LM(q) refers to Engle's (1982) Lagrange multiplier test for ARCH effects, which is obtained by running the regression $\widehat{\epsilon}_t^2 = \alpha_0 + \sum_{i=1}^q \widehat{\epsilon}_{t-i}^2 + u_t$, where $\{\widehat{\epsilon}_t\}$ is the demeaned return series. Then, under the null of no ARCH, the quantity TR^2 is approximately distributed as $\chi^2(q)$, where T is the number of observations, and R^2 is the coefficient of determination obtained for the regression. As explained in the text, the ARCH tests reported in the table are calculated by excluding the return from October 15 to October 22, 1987.

Table 2: Likelihood–based goodness–of–fit.

Models with constant (within–regime) covariances				
	MNG(0,0)	MMSG(0,0;2)	MNMG$_s$(0,0;2)	MNMG(0,0;2)
K	9	17	16	19
Log–likelihood				
Value (Rank)	–7400.7 (8)	–7234.9 (5)	–7295.4 (7)	–7293.8 (6)
BIC				
Value (Rank)	14865 (8)	14589 (5)	14703 (6)	14721 (7)
GARCH models				
	MNG(1,1)	MMSG$(1,1;2)$	MNMG$_s(1,1;2)$	MNMG$(1,1;2)$
K	15	29	28	31
Log–likelihood				
Value (Rank)	–7203.6 (4)	**–7124.9** (1)	–7131.7 (3)	–7127.8 (2)
BIC				
Value (Rank)	14513 (4)	**14454** (1)	14460 (2)	14473 (3)

The table shows likelihood–based goodness–of–fit measures for models fitted to the international stock market return series. K refers to the number of parameters of a model, "Log–likelihood" is the value of the maximized log–likelihood function, and BIC is the Bayesian information criterion of Schwarz (1978), i.e., $\text{BIC} = -2 \times \text{Log–likelihood} + K \log T$, where T is the sample size. Smaller values of BIC are preferred. For both criteria, the criterion value and the ranking of the models are shown. Boldface entries indicate the best model for the particular criterion.

Table 3: Parameter estimates for the MMSG(1,1;2) model.

$$r_t = \nu + \epsilon_t, \text{ where } \nu = [\underset{(0.051)}{0.245}, \underset{(0.060)}{0.223}, \underset{(0.070)}{0.339}]', \ \epsilon_t \sim \text{MMSG}(1,1;2)$$

$$P = \begin{bmatrix} \underset{(0.029)}{0.934} & \underset{(0.139)}{0.455} \\ \underset{(0.029)}{0.066} & \underset{(0.139)}{0.545} \end{bmatrix}, \quad \pi_{1,\infty} = 0.873, \quad \pi_{2,\infty} = 0.127, \quad \delta = p_{11} + p_{22} - 1 = \underset{(0.147)}{0.479}$$

$$H_{1t} = \begin{bmatrix} \underset{(0.017)}{0.043} & \underset{(0.008)}{0.014} & \underset{(0.010)}{0.018} \\ & \underset{(0.016)}{0.023} & \underset{(0.012)}{0.023} \\ & & \underset{(0.028)}{0.057} \end{bmatrix} + \begin{bmatrix} \underset{(0.007)}{0.030} & \underset{(0.005)}{0.026} & \underset{(0.006)}{0.035} \\ & \underset{(0.007)}{0.023} & \underset{(0.007)}{0.031} \\ & & \underset{(0.009)}{0.041} \end{bmatrix} \odot (\epsilon_{t-1}\epsilon'_{t-1}) + \begin{bmatrix} \underset{(0.010)}{0.947} & \underset{(0.007)}{0.956} & \underset{(0.009)}{0.942} \\ & \underset{(0.009)}{0.965} & \underset{(0.009)}{0.951} \\ & & \underset{(0.011)}{0.937} \end{bmatrix} \odot H_{1,t-1}$$

$\rho(A_{11} + B_{11}) = 0.988$

$$H_{2t} = \begin{bmatrix} \underset{(0.063)}{0.055} & \underset{(0.111)}{0.111} & \underset{(0.098)}{0.099} \\ & \underset{(0.302)}{0.254} & \underset{(0.165)}{0.188} \\ & & \underset{(0.208)}{0.185} \end{bmatrix} + \begin{bmatrix} \underset{(0.074)}{0.064} & \underset{(0.032)}{0.057} & \underset{(0.036)}{0.048} \\ & \underset{(0.034)}{0.051} & \underset{(0.020)}{0.043} \\ & & \underset{(0.020)}{0.036} \end{bmatrix} \odot (\epsilon_{t-1}\epsilon'_{t-1}) + \begin{bmatrix} \underset{(0.034)}{0.972} & \underset{(0.023)}{0.969} & \underset{(0.023)}{0.977} \\ & \underset{(0.020)}{0.966} & \underset{(0.015)}{0.973} \\ & & \underset{(0.014)}{0.981} \end{bmatrix} \odot H_{2,t-1}$$

$\rho(A_{12} + B_{12}) = 1.036$

$\rho(\mathbb{P}_{C_{11}}) = 0.994, \quad \rho(\mathbb{P}_{C_{22}}) = 0.991$

Standard errors of parameter estimates are given in parentheses. The multivariate Markov–switching GARCH$(p, q; k)$ (MMSG$(p, q; k)$) model is defined in Equations (1)–(3), where the dynamics of the component covariance matrices are specified as in (7). Matrices $\mathbb{P}_{C_{11}}$ and $\mathbb{P}_{C_{22}}$ are defined in Lemma 1 and (18)–(19), and $\rho(A)$ denotes the largest eigenvalue in modulus of a square matrix A.

Table 4: Parameter estimates for the symmetric MNMG(1,1;2) model.

$$r_t = \nu + \epsilon_t, \text{ where } \nu = [\underset{(0.051)}{0.248}, \underset{(0.063)}{0.231}, \underset{(0.070)}{0.331}]', \ \epsilon_t \sim \text{MNMG}_s(1,1;2)$$

$$P = \pi_\infty \mathbf{1}_2', \quad \pi_{1,\infty} = \underset{(0.027)}{0.905}, \quad \pi_{2,\infty} = \underset{(0.027)}{0.095}$$

$$H_{1t} = \begin{bmatrix} \underset{(0.018)}{0.048} & \underset{(0.008)}{0.014} & \underset{(0.011)}{0.025} \\ & \underset{(0.014)}{0.023} & \underset{(0.014)}{0.028} \\ & & \underset{(0.031)}{0.070} \end{bmatrix} + \begin{bmatrix} \underset{(0.005)}{0.031} & \underset{(0.003)}{0.030} & \underset{(0.005)}{0.039} \\ & \underset{(0.004)}{0.030} & \underset{(0.005)}{0.038} \\ & & \underset{(0.008)}{0.048} \end{bmatrix} \odot (\epsilon_{t-1}\epsilon_{t-1}') + \begin{bmatrix} \underset{(0.009)}{0.947} & \underset{(0.006)}{0.952} & \underset{(0.008)}{0.938} \\ & \underset{(0.005)}{0.958} & \underset{(0.007)}{0.944} \\ & & \underset{(0.011)}{0.930} \end{bmatrix} \odot H_{1,t-1}$$

$$\rho(A_{11} + B_{11}) = 0.988$$

$$H_{2t} = \begin{bmatrix} \underset{(0.110)}{0.187} & \underset{(0.215)}{0.451} & \underset{(0.085)}{0.239} \\ & \underset{(0.492)}{1.091} & \underset{(0.155)}{0.546} \\ & & \underset{(0.313)}{0.436} \end{bmatrix} + \begin{bmatrix} \underset{(0.032)}{0.142} & \underset{(0.059)}{0.053} & \underset{(0.033)}{0.078} \\ & \underset{(0.042)}{0.020} & \underset{(0.038)}{0.029} \\ & & \underset{(0.030)}{0.043} \end{bmatrix} \odot (\epsilon_{t-1}\epsilon_{t-1}') + \begin{bmatrix} \underset{(0.002)}{0.939} & \underset{(0.010)}{0.936} & \underset{(0.008)}{0.954} \\ & \underset{(0.021)}{0.934} & \underset{(0.012)}{0.952} \\ & & \underset{(0.018)}{0.969} \end{bmatrix} \odot H_{2,t-1}$$

$$\rho(A_{12} + B_{12}) = 1.081$$

$$\rho(\mathbb{P}_{C_{11}}) = 0.987, \quad \rho(\mathbb{P}_{C_{22}}) = 0.977$$

Standard errors of parameter estimates are given in parentheses. The symmetric multivariate normal mixture GARCH$(p,q;k)$ (MNMG$_s(p,q;k)$) model is as defined in Equations (1)–(3), but with the transition matrix, P, restricted to have rank one, i.e., $P = \pi_\infty \mathbf{1}_2'$. The dynamics of the component covariance matrices are specified as in (7). Matrices $\mathbb{P}_{C_{11}}$ and $\mathbb{P}_{C_{22}}$ are defined in Lemma 1 and (18)–(19), and $\rho(A)$ denotes the largest eigenvalue in modulus of a square matrix A.

Table 5: Parameter estimates for the asymmetric MNMG(1,1;2) model.

$$r_t = \nu + \epsilon_t, \text{ where } \nu = [\underset{(0.054)}{0.207}, \underset{(0.065)}{0.190}, \underset{(0.075)}{0.276}]', \ \epsilon_t \sim \text{MNMG}(1,1;2)$$

$$P = \pi_\infty \mathbf{1}_2', \quad \pi_{1,\infty} = \underset{(0.030)}{0.897}, \quad \pi_{2,\infty} = \underset{(0.030)}{0.103}$$

$$\mu_1 = [\underset{(0.045)}{0.119}, \underset{(0.052)}{0.094}, \underset{(0.062)}{0.121}]', \quad \mu_2 = [\underset{(0.379)}{-1.032}, \underset{(0.452)}{-0.816}, \underset{(0.570)}{-1.054}]'$$

$$H_{1t} = \begin{bmatrix} \underset{(0.016)}{0.039} & \underset{(0.008)}{0.010} & \underset{(0.010)}{0.020} \\ & \underset{(0.015)}{0.021} & \underset{(0.013)}{0.025} \\ & & \underset{(0.029)}{0.064} \end{bmatrix} + \begin{bmatrix} \underset{(0.006)}{0.031} & \underset{(0.005)}{0.031} & \underset{(0.006)}{0.039} \\ & \underset{(0.006)}{0.032} & \underset{(0.006)}{0.040} \\ & & \underset{(0.008)}{0.049} \end{bmatrix} \odot (\epsilon_{t-1}\epsilon_{t-1}') + \begin{bmatrix} \underset{(0.009)}{0.948} & \underset{(0.006)}{0.952} & \underset{(0.008)}{0.939} \\ & \underset{(0.007)}{0.957} & \underset{(0.008)}{0.943} \\ & & \underset{(0.011)}{0.929} \end{bmatrix} \odot H_{1,t-1}$$

$$\rho(A_{11} + B_{11}) = 0.988$$

$$H_{2t} = \begin{bmatrix} \underset{(0.157)}{0.234} & \underset{(0.365)}{0.557} & \underset{(0.176)}{0.302} \\ & \underset{(1.055)}{1.349} & \underset{(0.451)}{0.654} \\ & & \underset{(0.464)}{0.578} \end{bmatrix} + \begin{bmatrix} \underset{(0.088)}{0.172} & \underset{(0.063)}{0.043} & \underset{(0.040)}{0.076} \\ & \underset{(0.030)}{0.011} & \underset{(0.030)}{0.019} \\ & & \underset{(0.026)}{0.034} \end{bmatrix} \odot (\epsilon_{t-1}\epsilon_{t-1}') + \begin{bmatrix} \underset{(0.040)}{0.917} & \underset{(0.041)}{0.916} & \underset{(0.027)}{0.939} \\ & \underset{(0.056)}{0.915} & \underset{(0.034)}{0.938} \\ & & \underset{(0.024)}{0.962} \end{bmatrix} \odot H_{2,t-1}$$

$$\rho(A_{12} + B_{12}) = 1.089$$

$$\rho(\mathbb{P}_{C_{11}}) = 0.986, \quad \rho(\mathbb{P}_{C_{22}}) = 0.977$$

Standard errors of parameter estimates are given in parentheses. The asymmetric multivariate normal mixture GARCH$(p, q; k)$ (MNMG$(p, q; k)$) model is as defined in Equations (1)–(3), but with two modifications. Firstly, the transition matrix, P, is restricted to have rank one, i.e., $P = \pi_\infty \mathbf{1}_2'$. Secondly, the regimes are characterized by different mean returns, i.e., in (1), we have $\epsilon_t = \mu_{\Delta_t} + H_{\Delta_t,t}^{1/2}\xi_t$, where $\pi_{1,\infty}\mu_1 + \pi_{2,\infty}\mu_2 = 0_{3\times1}$. The dynamics of the component covariance matrices are specified as in (7). Matrices $\mathbb{P}_{C_{11}}$ and $\mathbb{P}_{C_{22}}$ are defined in Lemma 1 and (18)–(19), and $\rho(A)$ denotes the largest eigenvalue in modulus of a square matrix A.

Table 6: Unconditional (regime–dependent) covariance matrices and implied correlations.

Model	$\mathrm{E}(\epsilon_t\epsilon_t')$ & implied corr.	$\mathrm{E}(\epsilon_t\epsilon_t' \mid \Delta_t = 1)$ & implied corr.	$\mathrm{E}(\epsilon_t\epsilon_t' \mid \Delta_t = 2)$ & implied corr.
MNG(1,1)	$\begin{bmatrix} 4.59 & 0.49 & 0.52 \\ 2.54 & 5.79 & 0.63 \\ 3.51 & 4.83 & 10.2 \end{bmatrix}$	–	–
MMSG(1,1;2)	$\begin{bmatrix} 4.78 & 0.55 & 0.55 \\ 3.08 & 6.52 & 0.63 \\ 4.05 & 5.38 & 11.2 \end{bmatrix}$	$\begin{bmatrix} 3.53 & 0.52 & 0.51 \\ 2.14 & 4.91 & 0.60 \\ 2.75 & 3.80 & 8.13 \end{bmatrix}$	$\begin{bmatrix} 13.3 & 0.62 & 0.62 \\ 9.55 & 17.6 & 0.68 \\ 12.9 & 16.2 & 32.2 \end{bmatrix}$
MNMG$_s$(1,1;2)	$\begin{bmatrix} 4.24 & 0.52 & 0.50 \\ 2.74 & 6.50 & 0.60 \\ 3.12 & 4.63 & 9.15 \end{bmatrix}$	$\begin{bmatrix} 3.33 & 0.49 & 0.48 \\ 2.04 & 5.23 & 0.59 \\ 2.34 & 3.64 & 7.27 \end{bmatrix}$	$\begin{bmatrix} 12.9 & 0.61 & 0.56 \\ 9.37 & 18.5 & 0.63 \\ 10.5 & 14.1 & 27.0 \end{bmatrix}$
MNMG(1,1;2)	$\begin{bmatrix} 4.29 & 0.51 & 0.50 \\ 2.71 & 6.51 & 0.60 \\ 3.10 & 4.58 & 8.91 \end{bmatrix}$	$\begin{bmatrix} 3.30 & 0.48 & 0.47 \\ 1.99 & 5.25 & 0.59 \\ 2.29 & 3.61 & 7.11 \end{bmatrix}$	$\begin{bmatrix} 11.7 & 0.57 & 0.54 \\ 8.03 & 16.7 & 0.61 \\ 8.89 & 12.0 & 23.3 \end{bmatrix}$

The table shows the unconditional overall and regime–dependent covariance matrices, as implied by the estimated multivariate GARCH models, along with the associated correlation structures in the upper triangular parts of the matrices. The return vector is $r_t = [r_{1t}, r_{2t}, r_{3t}]'$, where r_{1t}, r_{2t}, and r_{3t} are the time–t returns of the S&P500, the FTSE, and the DAX, respectively. For the mixture models, the unconditional overall covariance matrix, $\mathrm{E}(\epsilon_t\epsilon_t')$, is given by (26), and the regime–specific unconditional covariance matrices are given by (31). The relation between the overall and the regime–specific covariance matrices is $\mathrm{E}(\epsilon_t\epsilon_t') = \pi_{1,\infty}\mathrm{E}(\epsilon_t\epsilon_t' \mid \Delta_t = 1) + \pi_{2,\infty}\mathrm{E}(\epsilon_t\epsilon_t' \mid \Delta_t = 2)$ for MMSG(1,1;2) and MNMG$_s$(1,1;2), and $\mathrm{E}(\epsilon_t\epsilon_t') = \pi_{1,\infty}\mathrm{E}(\epsilon_t\epsilon_t' \mid \Delta_t = 1) + \pi_{2,\infty}\mathrm{E}(\epsilon_t\epsilon_t' \mid \Delta_t = 2) + \pi_{1,\infty}\mu_1\mu_1' + \pi_{2,\infty}\mu_2\mu_2'$ for MNMG(1,1;2), where μ_1 and μ_2 are given in Table 5.

Table 7: Distributional properties of stock market returns over the out–of–sample period.

	mean	covariance/ correlation matrix S&P500	FTSE	DAX	skew	kurt	JB	ARCH–LM(q) $q=1$	$q=5$	$q=10$
S&P500	0.186	4.701	0.647	0.659	–0.233	4.635	75.50	6.333	32.09	40.53
							(0.000)	(0.012)	(0.000)	(0.000)
FTSE	0.150	3.063	4.775	0.720	–0.117	4.127	34.64	24.77	30.29	43.36
							(0.000)	(0.000)	(0.000)	(0.000)
DAX	0.216	4.340	4.775	9.219	–0.228	5.496	168.2	17.19	42.52	51.15
							(0.000)	(0.000)	(0.000)	(0.000)

p–values are given in parentheses. See Table 1 for explanations.

Table 8: Properties of realized portfolio returns.

Risk aversion, c	0.01	0.025	0.05	0.1	0.5	1
Multivariate Normal GARCH(1,1) (MNG(1,1))						
mean($r_{p,t}$)	0.191	0.171	0.153	0.149	0.156	0.157
var($r_{p,t}$)	6.012	4.958	4.374	4.101	3.933	3.920
skew($r_{p,t}$)	–0.356	–0.383	–0.439	–0.374	–0.320	–0.313
kurt($r_{p,t}$)	5.057	5.043	5.292	5.145	5.138	5.137
$e^{-3c}\overline{U(r_{p,t})}$	–0.9689	–0.9252	–0.8589	–0.7455	–0.3938	–1.4596
Multivariate Markov–switching GARCH(1,1;2) (MMSG(1,1;2))						
mean($r_{p,t}$)	0.223	0.207	0.176	0.163	0.170	0.173
var($r_{p,t}$)	7.398	5.656	4.576	4.137	3.971	4.099
skew($r_{p,t}$)	–0.312	–0.275	–0.321	–0.309	–0.275	–0.253
kurt($r_{p,t}$)	4.803	4.566	4.796	4.966	5.020	5.085
$e^{-3c}\overline{U(r_{p,t})}$	–0.9686	–0.9246	–0.8581	–0.7445	–0.3853	–1.3006
Multivariate symmetric normal mixture GARCH(1,1;2) (MNMG$_s$(1,1;2))						
mean	0.209	0.192	0.165	0.160	0.162	0.158
variance	7.135	5.506	4.534	4.154	4.021	4.177
skewness	–0.433	–0.376	–0.415	–0.369	–0.311	–0.287
kurtosis	5.394	5.045	5.235	5.168	5.165	5.307
$e^{-3c}\overline{U(r_{p,t})}$	–0.9688	–0.9249	–0.8586	–0.7449	–0.3960	–1.5238
Multivariate asymmetric normal mixture GARCH(1,1;2) (MNMG(1,1;2))						
mean	0.198	0.179	0.162	0.158	0.159	0.154
variance	6.590	5.253	4.468	4.144	4.081	4.323
skewness	–0.431	–0.372	–0.412	–0.367	–0.311	–0.284
kurtosis	5.336	5.115	5.344	5.245	5.232	5.369
$e^{-3c}\overline{U(r_{p,t})}$	–0.9688	–0.9251	–0.8586	–0.7449	–0.4018	–1.7230

Shown are summary statistics for the out–of–sample portfolio returns, $r_{p,t}$, $t = 501, \ldots, 1127$. "skew($r_{p,t}$)" and "kurt($r_{p,t}$)" denote the moment–based coefficients of skewness and kurtosis, respectively, of the $r_{p,t}$–series; see the legend of Table 1 for the definition of these measures. $\overline{U(r_{p,t})}$ is the average realized utility, as given by (54).

Table 9: Evaluation of portfolio return predictive densities.

Risk aversion, c	0.01	0.025	0.05	0.1	0.5	1
Multivariate Normal GARCH(1,1) (MNG(1,1))						
skew(z_t)	-0.389^{***}	-0.375^{***}	-0.398^{***}	-0.370^{***}	-0.348^{***}	-0.345^{***}
kurt(z_t)	4.535^{***}	4.130^{***}	4.004^{***}	3.844^{***}	3.760^{***}	3.755^{***}
ARCH–LM(1)	3.303^{*}	4.474^{**}	5.822^{**}	6.481^{**}	7.693^{***}	7.926^{***}
ARCH–LM(5)	7.118	9.442^{*}	10.47^{*}	11.55^{**}	12.62^{**}	12.75^{**}
ARCH–LM(10)	10.64	13.54	14.80	16.39^{*}	18.20^{*}	18.43^{**}
VaR(0.005)	0.011^{**}	0.016^{***}	0.018^{***}	0.016^{***}	0.013^{**}	0.013^{**}
VaR(0.01)	0.014	0.018^{*}	0.024^{***}	0.026^{***}	0.024^{***}	0.024^{***}
VaR(0.05)	0.053	0.049	0.056	0.064^{*}	0.061	0.061
Multivariate Markov–switching GARCH(1,1;2) (MMSG(1,1;2))						
skew(z_t)	-0.102	-0.130	-0.150	-0.159	-0.188^{*}	-0.202^{**}
kurt(z_t)	2.949	2.828	2.731	2.687	2.761	2.847
ARCH–LM(1)	1.803	2.195	1.846	1.608	0.771	0.644
ARCH–LM(5)	7.814	10.05^{*}	7.624	6.594	4.469	4.557
ARCH–LM(10)	9.660	12.93	10.31	10.58	10.95	11.48
VaR(0.005)	0.011^{**}	0.010^{*}	0.010^{*}	0.006	0.006	0.010^{*}
VaR(0.01)	0.016	0.014	0.016	0.018^{*}	0.019^{**}	0.018^{*}
VaR(0.05)	0.070^{**}	0.070^{**}	0.067^{**}	0.078^{***}	0.072^{**}	0.072^{**}
Multivariate symmetric normal mixture GARCH(1,1;2) (MNMG$_s$(1,1;2))						
skew(z_t)	-0.149	-0.174^{*}	-0.187^{*}	-0.198^{**}	-0.228^{**}	-0.246^{**}
kurt(z_t)	3.057	2.925	2.833	2.796	2.881	2.999
ARCH–LM(1)	4.278^{**}	5.813^{**}	5.361^{**}	4.749^{**}	3.003^{*}	2.379
ARCH–LM(5)	8.864	11.81^{**}	10.96^{*}	10.07^{*}	7.242	7.072
ARCH–LM(10)	11.31	15.15	14.34	15.02	13.68	12.70
VaR(0.005)	0.011^{**}	0.013^{**}	0.011^{**}	0.011^{**}	0.010^{*}	0.011^{**}
VaR(0.01)	0.018^{*}	0.016	0.018^{*}	0.022^{***}	0.022^{***}	0.021^{**}
VaR(0.05)	0.065^{*}	0.069^{**}	0.065^{*}	0.073^{***}	0.070^{**}	0.070^{**}
Multivariate asymmetric normal mixture GARCH(1,1;2) (MNMG(1,1;2))						
skew(z_t)	-0.049	-0.049	-0.065	-0.072	-0.086	-0.095
kurt(z_t)	3.124	2.928	2.832	2.790	2.897	3.019
ARCH–LM(1)	3.540^{*}	4.391^{**}	3.483^{*}	3.247^{*}	2.281	2.431
ARCH–LM(5)	7.914	9.050	7.124	6.242	4.888	6.052
ARCH–LM(10)	11.52	13.89	11.51	11.68	10.64	10.55
VaR(0.005)	0.010^{*}	0.008	0.008	0.008	0.008	0.008
VaR(0.01)	0.013	0.014	0.013	0.013	0.013	0.014
VaR(0.05)	0.062^{*}	0.059	0.056	0.064^{*}	0.059	0.062^{*}

Shown are summary statistics for the transformed portfolio return series (55). "skew(z_t)" and "kurt(z_t)" denote the moment based coefficients of skewness, γ, and kurtosis, κ, as are defined in the legend of Table 1, of the $\{z_t\}$ defined in (55). Under normality, $T\gamma^2/6 \overset{asy}{\sim} \chi^2(1)$ and $T(\kappa-3)^2/24 \overset{asy}{\sim} \chi^2(1)$. ARCH–LM is the Lagrange multiplier test for ARCH effects applied to (55), the details of which are also provided in the legend of Table 1. "VaR(α)" refers to the Value–at–Risk (VaR) measures implied by the respective models. Reported are the empirical shortfall probabilities, x/T, observed for a nominal VaR level α, $\alpha = 0.005, 0.01, 0.05$, where x is the shortfall frequency, and T is the number of forecasts evaluated. Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

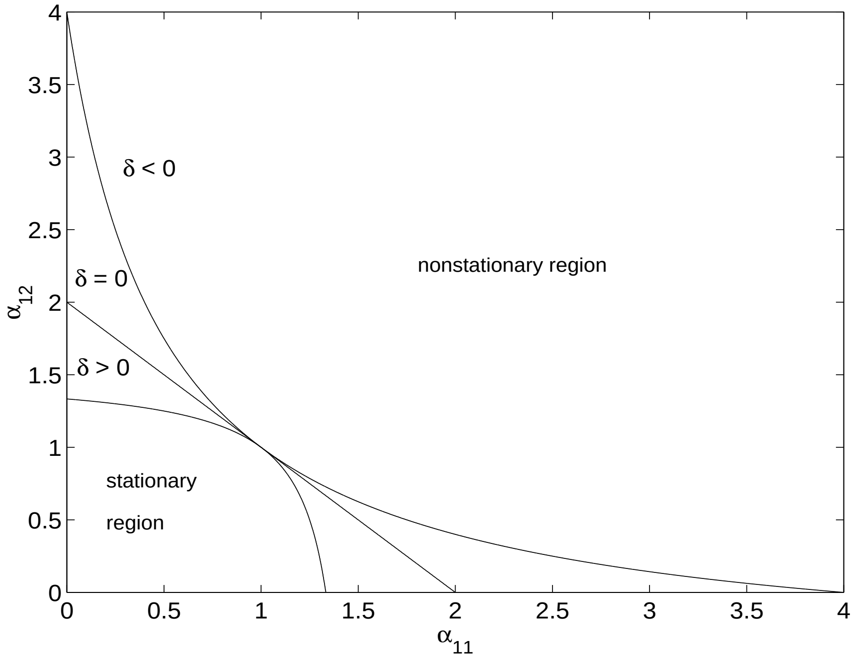

Figure 1: Stationarity regions for the univariate two–regime Markov–switching ARCH(1) processes characterized by the transition matrices given in (34).

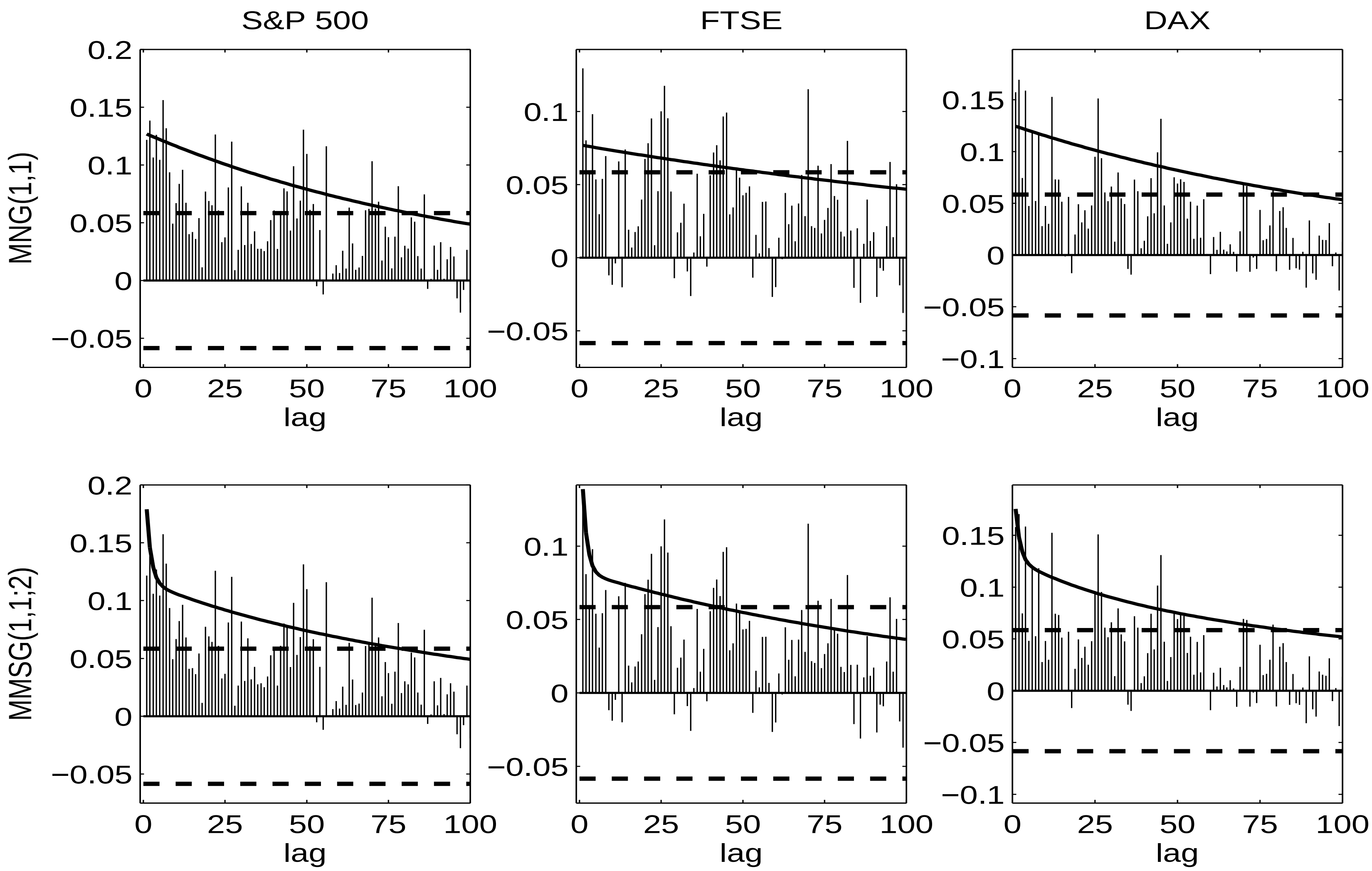

Figure 2: Shown are, from left to right, the sample autocorrelations of the squared returns of the S&P500, the FTSE, and the DAX, along with their theoretical counterparts as implied by, from top to bottom, the fitted MNG(1,1) and MMSG(1,1;2) models. The theoretical autocorrelation function is defined in (39). Dashed lines represent approximate 95% one–at–a–time confidence intervals.

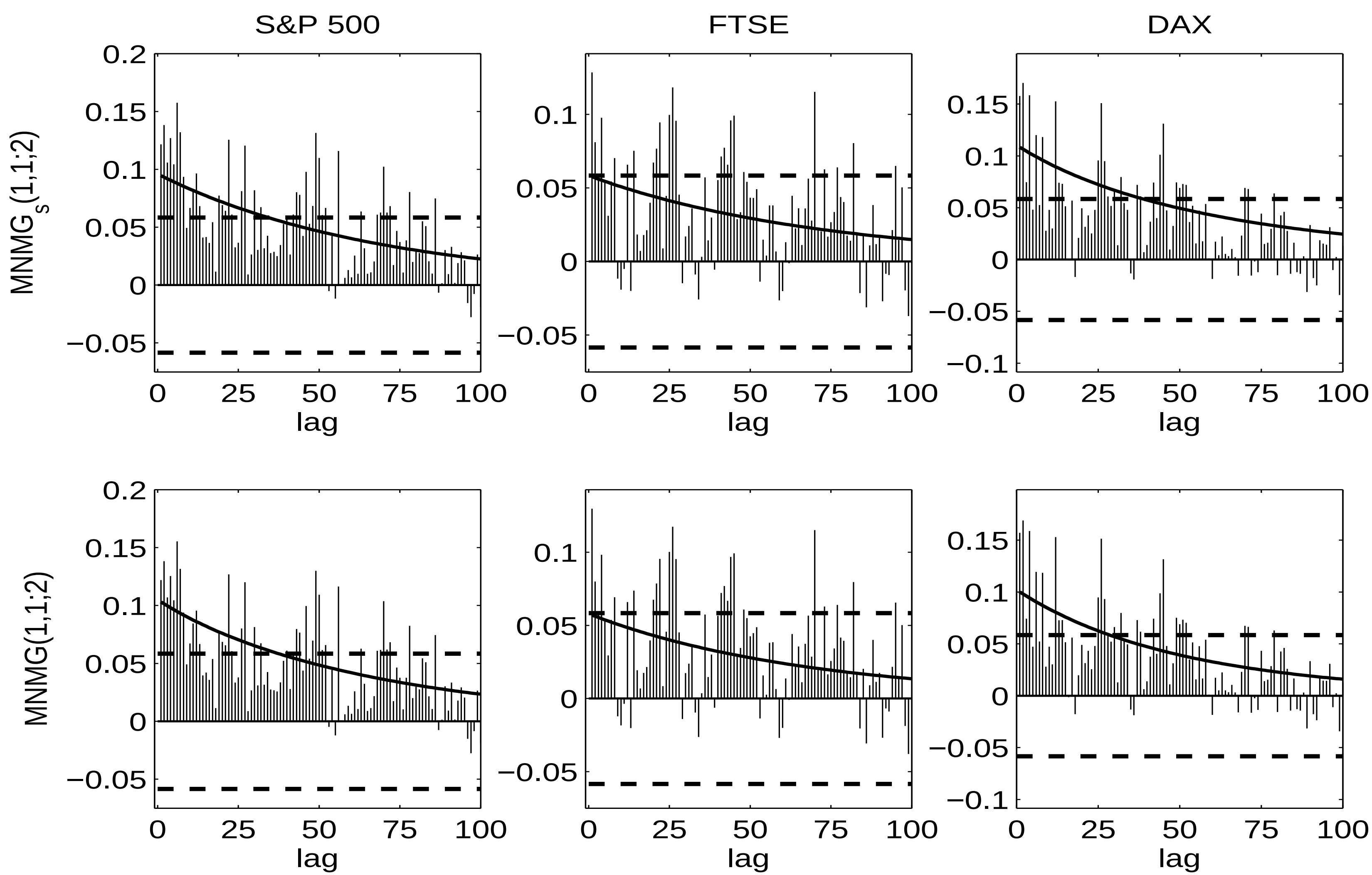

Figure 3: Shown are, from left to right, the sample autocorrelations of the squared returns of the S&P500, the FTSE, and the DAX, along with their theoretical counterparts as implied by, from top to bottom, the fitted MNMG$_s$(1,1) and MNMG(1,1;2) models. The theoretical autocorrelation function is defined in (39). Dashed lines represent approximate 95% one–at–a–time confidence intervals.

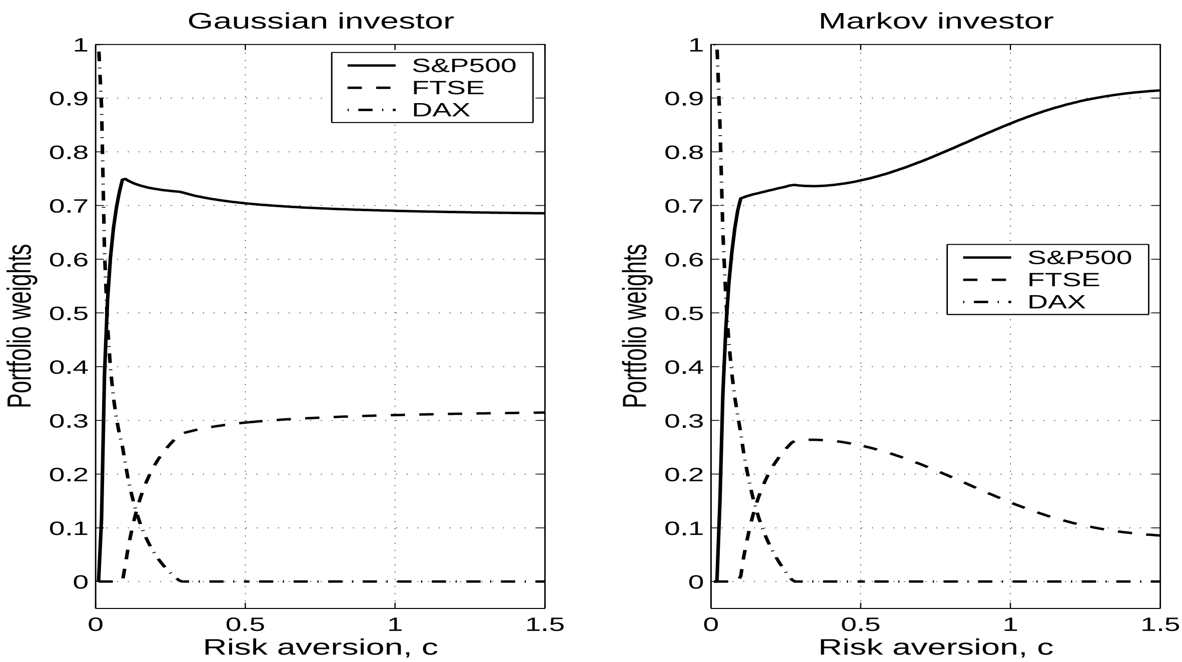

Figure 4: Shown are the optimal one-week-ahead portfolio weights of two groups of investors on August 4, 2005. The left plot shows, as a function of risk aversion, c, the optimal weights of investors employing a single–regime GARCH(1,1) process, i.e., model MNG(1,1). The right plot repeats this, but for investors using the Markov–switching GARCH model, i.e., model MMSG(1,1;2).