Mixed Normal Conditional Heteroskedasticity

MARKUS HAAS
University of Munich

STEFAN MITTNIK
University of Munich, Center for Financial Studies, and Ifo Institute for Economic Research

MARC S. PAOLELLA
University of Zurich

ABSTRACT

Both unconditional mixed normal distributions and GARCH models with fat-tailed conditional distributions have been employed in the literature for modeling financial data. We consider a mixed normal distribution coupled with a GARCH-type structure (termed MN-GARCH) which allows for conditional variance in each of the components as well as dynamic feedback between the components. Special cases and relationships with previously proposed specifications are discussed and stationarity conditions are derived. For the empirically most relevant GARCH(1,1) case, the conditions for existence of arbitrary integer moments are given and analytic expressions of the unconditional skewness, kurtosis, and autocorrelations of the squared process are derived. Finally, employing daily return data on the NASDAQ index, we provide a detailed empirical analysis and compare both the in-sample fit and out-of-sample forecasting performance of the MN-GARCH as well as recently proposed Markov-switching models. We show that the MN-GARCH approach can generate a plausible disaggregation of the conditional variance process in which the components' volatility dynamics have a clearly distinct behavior, which is, for example, compatible with the well-known leverage effect.

KEYWORDS: GARCH, kurtosis, leverage effect, Markov-switching, skewness, stationarity, value-at-risk

We thank Eric Renault (the co-editor) and two anonymous referees for constructive comments. The research of M. Haas and S. Mittnik was supported by the Deutsche Forschungsgemeinschaft. Part of the research of M. Paolella was carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation. Address correspondence to Stefan Mittnik, Chair of Financial Econometrics, Institute of Statistics, University of Munich, D-80799 Munich, Germany, or e-mail: finmetrics@stat.uni-muenchen.de.

Journal of Financial Econometrics, Vol. 2, No. 2, pp. 211–250

DOI: 10.1093/jfinec/nbh009

Although generalized autoregressive conditional heteroskedastic (GARCH) models driven by normally distributed innovations and their numerous extensions can account for a substantial portion of both the volatility clustering and excess kurtosis found in financial return series, a GARCH-type model has yet to be constructed for which the filtered residuals consistently fail to exhibit clear-cut signs of nonnormality. On the contrary, it appears that the vast majority of GARCH-type models, when fit to returns over weekly and shorter horizons, imply quite heavy-tailed conditional innovation distributions. Moreover, there is a growing awareness of skewness in both unconditional and conditional return distributions [see, e.g., Kane (1977), Friend and Westerfield (1980), Rozelle and Fielitz (1980), Simkowitz and Beedles (1980), Mittnik and Rachev (1993), St. Pierre (1993), Franses and van Dijk (1996), Peiró (1999), and Harvey and Siddique (1999)]. A natural way of accommodating such stylized facts is to specify a GARCH-type structure driven by i.i.d. innovations from a fat-tailed and, possibly, asymmetric distribution. Moreover, building on work by Hansen (1994), the studies of Harvey and Siddique (1999), Paolella (1999), Brännäs and Nordman (2001), and Rockinger and Jondeau (2002) employ autoregressive-type structures to allow for time variation in the skewness and, in some cases, also kurtosis. Thus, while not as blatant as volatility clustering and heavy tails, time-varying skewness has emerged as another stylized fact of asset returns.

In this article, we investigate a model that incorporates the original assumption of normal innovations, yet can still adequately capture all three aforementioned stylized facts. Specifically, we let the conditional distribution be a mixture of normals [in short, mixed normal (MN)] and extend the usual GARCH structure by modeling the dynamics in volatility by a system of equations that permits feedback between the mixture components. With one component, the model reduces to the normal-GARCH model originally proposed in Bollerslev (1986). The excess kurtosis, which plagues normal-GARCH specifications, can be adequately modeled with only two components. In addition, with more than one component, time-varying skewness is accommodated, that is, it is inherent in the model without requiring explicit specification of a conditional skewness process. Moreover, the model can capture Black's (1976) leverage effect. These aspects will be demonstrated in the empirical example presented below.

The proposed model class has some characteristics similar to Markov-switching models, which have undoubtedly grown in importance (and complexity) since the seminal work of Hamilton (1988, 1989) [see, e.g., Hamilton and Susmel (1994), Gray (1996), and Dueker (1997)]. The relation between the mixed normal and the Markov-switching GARCH model will be explored in Section 1.5, and Section 3 provides an empirical comparison. An approach similar to ours has recently been suggested by Wong and Li (2001), who also argue against use of a latent Markov structure; see Section 1.4 below for some discussion of their work.

The remainder of this article is organized as follows. Section 1 reviews relevant properties of unconditional MN distributions, presents the MN-GARCH model and discusses various special cases. Section 2 details stationarity conditions and dynamic properties. Section 3 discusses various issues relevant to empirical modeling of asset returns and use of the proposed model class, including numerical estimation issues, methods for determining the required number of components in the mixture, goodness-of-fit assessment, diagnostic checking of the residuals, and issues involved in testing and comparing the quality of out-of-sample value-at-risk predictions. A detailed illustration of these techniques is provided by an analysis of the returns on the NASDAQ index, including a discussion and interpretation of the fitted model parameters, comparison with Markov-switching models, and a demonstration of the relevance of the implied time-varying skewness and kurtosis inherent in the MN-GARCH model. The section concludes with a short discussion of the results for the fully parameterized model and an extension to the mixed Student's *t*-GARCH model. Section 4 contains concluding remarks and ideas for further research. Various technical details are gathered in the appendix.

1 MIXED NORMAL MODELS

The MN distribution has a long and illustrious history in statistics. Its use for modeling heavy-tailed distributions apparently dates back to 1886, when the mathematician, astronomer, and economist Simon Newcomb used it in his astronomical studies. After the seminal work of Pearson in 1894 on the moments estimator for the univariate normal mixture with two components, maximum-likelihood (ML) estimation became very popular with the advent of the EM algorithm of Dempster, Laird, and Rubin (1977), while exact bayesian analysis of mixtures became feasible after the introduction of the Gibbs sampler of Geman and Geman (1984) into the statistical mainstream by Gelfand and Smith (1990). Further discussion on the history, modern inferential methods, and applications associated with mixtures of normals can be found in Titterington, Smith, and Makov (1985), McLachlan and Basford (1988), and McLachlan and Peel (2000).

1.1 Unconditional Mixed Normal Distribution

A random variable Y is said to have a univariate (finite) normal mixture distribution if its unconditional density is given by

$$f(y) = \sum_{j=1}^{k} \lambda_j \phi(y; \mu_j, \sigma_j^2),$$

where $\lambda_j > 0$, $j = 1, \ldots, k$, $\sum_{j=1}^{k} \lambda_j = 1$, are the mixing weights and

$$\phi(y; \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{ -\frac{1}{2} \left(\frac{y - \mu_j}{\sigma_j} \right)^2 \right\}, \quad j = 1, \ldots, k,$$

are the component densities. The normal mixture has finite moments of all orders, with expected value and variance given by

$$\mu = \mathrm{E}(Y) = \sum_{j=1}^{k} \lambda_j \mu_j, \quad \mathrm{var}(Y) = \sum_{j=1}^{k} \lambda_j (\sigma_j^2 + \mu_j^2) - \left(\sum_{j=1}^{k} \lambda_j \mu_j \right)^2. \tag{1}$$

Owing to its great flexibility, the MN model has been found useful for describing the unconditional distribution of asset returns [cf. Fama (1965), Kon (1984), Akgiray and Booth (1987), and Tucker and Pond (1988)]. Indeed, even a two-component mixture is rather capable of exhibiting the skewness and kurtosis typical of financial data [see Gridgeman (1970), who first proved leptokurtosis of scale mixtures, and McLachlan and Peel (2000)].

An advantage of the MN model not shared by other distributional assumptions is that it may lend itself to economic interpretation. Considering unconditional distributions, Kon (1984), for example, argues that a discrete mixture of normal distributions may account for a series of information flows, such as economy-wide, market-specific, and firm-specific, resulting in a three component mixture. There are also attempts to liaise the mixture approach and those recent models in financial economics trying to explain the stylized facts of financial time series by the interaction of heterogeneous groups of agents with the groups processing market information differently.[1] This has been undertaken by Vigfusson (1997), though the theoretical underpinnings of this view are still rudimentary.

The MN model can also be appropriate for samples where the components follow a repeating sequence in generating observations. Day-of-the-week effects, as mentioned by Fama (1965), could be a possible source of mixture distributions. By analyzing corresponding subsamples, however, Fama (1965) found that the Monday effect does not give rise to the observed departure from normality. However, the mixture may still be interpreted as representing trading days of different types: A component with relatively low variance, for example, could represent "business as usual" — typically associated with a large mixing weight — while components with high variances and smaller weights could correspond to times of high volatility caused by the arrival of substantive new information.

1.2 Conditionally Heteroskedastic MN Processes

Time series $\{\epsilon_t\}$ is generated by a k-component mixed normal GARCH(p, q) process, or, in short, MN-GARCH, if the conditional distribution of ϵ_t is a k-component MN with zero mean, that is,

$$\epsilon_t|\Psi_{t-1} \sim \mathrm{MN}(\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_k, \sigma^2_{1t}, \ldots, \sigma^2_{kt}), \tag{2}$$

where Ψ_t is the information set at time t; $\lambda_i \in (0, 1)$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} \lambda_i = 1$; and $\mu_k = -\sum_{i=1}^{k-1} (\lambda_i/\lambda_k)\mu_i$. Furthermore, the $k \times 1$ vector of component variances, denoted by $\sigma_t^{(2)}$, evolves according to

$$\sigma_t^{(2)} = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon^2_{t-i} + \sum_{j=1}^{p} \beta_j \sigma^{(2)}_{t-j}, \tag{3}$$

where $\sigma_t^{(2)} = [\sigma^2_{1t}, \sigma^2_{2t}, \ldots, \sigma^2_{kt}]'$; $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ik}]'$, $i = 0, \ldots, q$, are $k \times 1$ vectors; and $\beta_j, j = 1, \ldots, p$, are $k \times k$ matrices with typical element $\beta_{j,mn}$. The assumptions

[1] See Samanidou et al. (2004) for an overview of these models.

$\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 0, \ldots, q$, and $\beta_j \geq 0$, $j = 1, \ldots, p$,[2] correspond to the nonnegativity conditions of Bollerslev (1986) for the normal-GARCH model, although they may be unnecessarily strong [cf. Nelson and Cao (1992)]. They are, however, necessary for the diagonal MN-GARCH(1,1) model, a useful special case introduced and employed below.

Using lag-operator notation, $L^q y_t = y_{t-q}$, an MN-GARCH process can be written as

$$(I_k - \beta(L))\sigma_t^{(2)} = \alpha_0 + \alpha(L)\epsilon_t^2, \tag{4}$$

where $\beta(L) = \sum_{j=1}^{p} \beta_j L^j$; $\alpha(L) = \sum_{i=1}^{q} \alpha_i L^i$; and I_k is the identity matrix of dimension k.

As is common, a mean equation can also be introduced to incorporate exogenous variables and/or lagged values via an ARMA(u, v) structure. In particular, an ARMA-MN-GARCH model for variable r_t refers to a process with mean equation

$$r_t = a_0 + \sum_{i=1}^{u} a_i r_{t-i} + \epsilon_t + \sum_{j=1}^{v} b_j \epsilon_{t-j}, \tag{5}$$

with constant a_0, AR parameters $a_1, \ldots, a_u$, MA parameters $b_1, \ldots, b_v$, and with $\epsilon_t | \Psi_{t-1}$ given by Equations (2) and (4). Note that Equation (5) implies common mean dynamics in the mixture components. Mixture autoregressive models with different AR structures in each component have also been employed in the literature [e.g., Wong and Li (2000) and Lanne and Saikkonen (2003)]. Lanne and Saikkonen (2003) show that a mixture of autoregressions with two regimes improves forecasts of weekly U.S. three-month Treasury bill rates relative to standard AR models. Although more general (nonlinear) mean dynamics can be merged with the heteroskedastic structure of Equation (3) without any conceptual difficulty, we adopt the specification of Equation (5), as we are mainly interested in high-frequency return series. A common finding is that there is at least some degree of predictability in stock returns, but these (weak) dependencies are usually captured by low-order AR structures [Campbell, Lo, and MacKinlay (1997: chap. 2)]. Similar findings hold for exchange rates. Even though significant mean nonlinearities are occasionally found in in-sample studies of daily exchange-rate returns, such models very rarely outperform linear models out of sample [see, e.g., Boero and Marrocu (2002)] for a survey of recent experience with forecasting exchange rates and further evidence against nonlinear models for point prediction]. Given this empirical rationale, the advantage of focusing on dynamic mixture effects in the conditional variance — apart from the obvious advantage of avoiding parameter proliferation to which multiregime models are naturally prone — is that the proposed model is able to disentangle the mean and the variance dynamics, which leads to theoretical results that enhance the analysis of the volatility process (see Sections 2 and 3). In contrast, as follows from

[2] In the case of nonscalars, symbols $>$ and $\geq$ indicate element-wise inequality.

Equation (1), in mixture autoregressive models with GARCH errors, the conditional variance is affected by both the AR and the GARCH dynamics.[3]

1.3 Special Cases

1.3.1 Full and Diagonal MN-GARCH A particularly interesting special case for modeling asset returns arises by restricting matrix $\beta(L)$ in Equation (4) to be diagonal (subsequently referred to as a *diagonal* MN-GARCH process). In addition to allowing for a clear interpretation of the dynamics of the component variances (see the end of Section 1.5 for a discussion), we find — not only for the example reported below — that it tends to be preferred over the full model when employing standard model selection criteria.

1.3.2 Partial MN-GARCH Models where only subsets of the component variances are driven by GARCH processes may be appropriate for describing the volatility dynamics of a given series. For example, occasionally occurring jumps in the level of volatility may be captured by a component with a relatively large, but constant variance. Below we consider diagonal partial models, where a model denoted by MN(k, g), $g \leq k$, uses k component densities, g of which follow a GARCH process and $k - g$ components are restricted to be constant. If, for example, models with $g = 1$ fit the data well, then the unconditional properties of the normal mixture (skewness and kurtosis) account for most of the improvement relative to the standard GARCH model with conditional normality, and volatility clustering is adequately captured by introducing one GARCH component.

1.3.3 Symmetric MN$_s$-GARCH We also entertain models for which all the component means are restricted to be zero, that is, $\mu_1 = \mu_2 = \cdots \mu_k = 0$, which imposes a symmetric conditional error distribution. These are denoted by MN$_s(k, g)$-GARCH. Because both the conditional innovations and the GARCH structure are symmetric, the unconditional error distribution will also be symmetric.

1.4 Relationship with Other MN-GARCH Specifications

To the best of our knowledge, Vlaar and Palm (1993) and Palm and Vlaar (1997) first suggested the normal mixture in a GARCH context. Their model sets, for all t, $\sigma_{2t}^2 = \sigma_{1t}^2 + \delta^2$ [cf. the parameterization in Ball and Torous (1983)][4] and thus is nested in Equation (3). In our notation, it takes the form

$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{bmatrix} = \begin{bmatrix} \alpha_{01} \\ \alpha_{01} + \delta^2 \end{bmatrix} + \begin{bmatrix} \alpha_{11} \\ \alpha_{11} \end{bmatrix} \epsilon_{t-1}^2 + \begin{bmatrix} \beta_{11} & 0 \\ \beta_{11} & 0 \end{bmatrix} \begin{bmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{bmatrix},$$

which permits skewness by allowing the component means to differ from zero. Lin and Yeh (2000) employed this process to model Taiwanese stock market returns.

[3] Theoretical results on the mixture autoregressive model with GARCH errors are not available and their dynamic properties need to be evaluated by simulation, as in Lanne and Saikkonen (2003).

[4] Vlaar and Palm (1993: p.357) motivate this specification by arguing that "this procedure is preferred to that of independent variances, since it seems reasonable to assume that the same GARCH effect is present in all variances."

Bauwens, Bos, and van Dijk (1999) consider an MN-GARCH model with two components, in which the component variances are proportional to each other, that is, for all t, $\sigma_{2t}^2 = \tau\sigma_{1t}^2$, specializing Equation (3) to[5]

$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \tau\alpha_0 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \tau\alpha_1 \end{bmatrix} \epsilon_{t-1}^2 + \begin{bmatrix} \beta_{11} & 0 \\ \tau\beta_{11} & 0 \end{bmatrix} \begin{bmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{bmatrix}.$$

It may be argued that the proportionality property is less appealing, since it implies that the dynamic behavior of the vector $\sigma_t^{(2)}$ is restricted as its elements are forced to be linearly dependent. This feature also applies to the Palm and Vlaar specification. In fact, a "linear" MN-GARCH process that nests these two models can be defined by $\sigma_{2t}^2 = a + b\sigma_{1t}^2$, where $a, b \geq 0$, $a + b > 0$. In Section 3, an informal check of this "linearity assumption" will be performed by computing the correlations of the vector $\sigma_t^{(2)}$ for the example considered.

A mixture model with conditional heteroskedasticity that is not nested in the structure of Equations (2)–(5) has recently been proposed by Wong and Li (2001). Their approach is different as they couple the mixture of autoregressions of Wong and Li (2000) with conditional heteroskedasticity. Similar to the popular Markov-switching autoregressions in econometrics [Hamilton (1989)], their model allows for nonlinear mean dynamics. The variance dynamics in the model are restricted to ARCH effects (vis-à-vis GARCH), though (standard) GARCH(1, 1) models are known to provide better descriptions of market volatility than even high-order-ARCH specifications.[6] In view of this and the discussion at the end of Section 1.2, the approach adopted in Equations (2)–(5) seems preferable in the context of high-frequency stock and exchange-rate returns.

1.5 Relationship with the Markov-Switching GARCH Model

The model proposed here has some characteristics that are similar to those of a Markov-switching (MS) GARCH model. Given the latter's popularity in the literature, common properties and differences of the two approaches are discussed next. The MS-GARCH model will also be considered in the empirical application in Section 3. Most popular among the MS-GARCH models is the recombining approach suggested by Gray (1996). This standing derives from the ability of this model to overcome the problem of path dependence which plagued the MS-GARCH models of Cai (1994) and Hamilton and Susmel (1994) prior to Gray (1996).[7]

[5] Alternatively, because $\sigma_{2,t-1}^2 = \tau\sigma_{1,t-1}^2$, we could define $\beta = \beta_{11} I_2$.

[6] One reason for this restriction may be that theoretical properties of mixtures of autoregressions with generalized ARCH errors seem to be quite difficult to obtain.

[7] The path dependence refers to the fact that in these models the current variance depends on the entire history of regimes, rendering parameter estimation quite complicated. A consequence of this is that Gray's model is usually used in empirical applications, although the theoretical properties of this model are unknown, while Francq, Roussignol, and Zakoian (2001) established these for the MS-GARCH models considered in Cai (1994) and Hamilton and Susmel (1994). Variants of MS-GARCH models in response to path dependence were also developed by Dueker (1997) and Klaassen (2002).

As in the MN-GARCH model defined by Equations (2)–(5), the conditional distribution in Gray's model is a discrete mixture of k distributions, which are commonly taken to be normal. It is assumed that at each point of time one of the k components generates observation r_t, where the process that selects the actual component is a hidden Markov chain with k-dimensional state space. Let $p_{jt|t-1}$, $j=1,\ldots,k$, be the probability that r_t is generated by the jth regime, conditional on Ψ_{t-1}.[8] Then, from Equation (1), the conditional variance of r_t is

$$h_t =: \mathrm{var}(r_t|\Psi_{t-1}) = \sum_{j=1}^{k} p_{jt|t-1}(\sigma_{jt}^2 + \mu_{jt}^2) - \left(\sum_{j=1}^{k} p_{jt|t-1}\mu_{jt}\right)^2. \tag{6}$$

To overcome the path dependence, Gray (1996) lets the conditional *overall process* variance of Equation (6) drive the conditional *regime* variances, that is,

$$\sigma_{jt}^2 = \alpha_{0j} + \alpha_{1j}\epsilon_{t-1}^2 + \beta_j h_{t-1}, \;\; j = 1,\ldots,k, \tag{7}$$

where

$$\epsilon_t = r_t - \mathrm{E}(r_t|\Psi_{t-1}) = r_t - \sum_{j=1}^{k} p_{jt|t-1}\mu_{jt}. \tag{8}$$

In the empirical application below, we assume for the conditional mean of the jth component, μ_{jt},

$$\mu_{jt} = \mu_j + \sum_{i=1}^{v} a_i r_{t-i}, \quad j = 1,\ldots,k. \tag{9}$$

As in Equation (5), the autoregressive parameters a_i, $i=1,\ldots,v$, are common to all regimes, while the intercept in the mean equation is allowed to switch.

There are essentially two features in which Gray's MS-GARCH model of Equations (6)–(9) differs from the system of Equations (2)–(5). The most pronounced is the difference in the stochastic process selecting the components. Compared to the Markov process, we assume an i.i.d. multinomial distribution for the mixing process. We argue for the latter because it enables us to disentangle the mixture effects of the conditional distribution from the dynamic properties of the returns. In particular, allowing for skewness in MS-GARCH models through different component means is inevitably associated with autocorrelation in the raw returns [see Timmermann (2000) for a derivation of the autocorrelation function of Markov-switching models]. Thus the lack of autocorrelation in the raw returns paired with dependencies in higher moments, which is one of the most pronounced theoretical properties of GARCH, is lost with MS-GARCH. This feature is highly desirable, for example, in the context of (weak-form) market efficiency, stating that available return data cannot be used to improve the prediction of future ones [Fama (1970)].

The second difference refers to the specification of the GARCH Equation (7). Prima facie, Equation (7) appears reasonable, because, as in the standard

[8] Hamilton (1989) developed an algorithm to compute these probabilities from model parameters and past observations.

(single-component) GARCH model, it is the past return variance that affects today's variance in regime j through β_j. However, it is the basic intuition of GARCH models that shocks drive the variance, and consequently, the intention behind the generalization from ARCH to GARCH was that the latter can parsimoniously represent a high-order ARCH process [Bollerslev (1986), Bera and Higgins (1993)]. For example, in the standard, single-component GARCH(1,1) model, α_1 measures the magnitude of a shock's immediate impact on the next period's variance, while β_1 is a mere parameter of inertia, reflecting the impact of a current shock on future variances. Thus it is preferable to have σ^2_{jt-1} instead of h_{t-1} on the right-hand side of Equation (7). In this case, the parameters have a clear interpretation, namely, $\alpha_{1j}(1-\beta_j)^{-1}$ is the total impact of a unit shock to component j's future variances; α_{1j} measures the magnitude of a shock's immediate impact on the next period's σ^2_{jt}; and β_j reflects the memory in component j's variance in response to such a shock, which can be characterized by the mean lag $(1-\beta_j)^{-1}$ or the median lag $-\log 2/\log\beta_j$ in component j [Bollerslev (1986)]. That said, it is also clear why, as already indicated in Section 1.3.1, the diagonal specification of Equation (3) may be particularly relevant.

2 STATIONARITY AND PERSISTENCE

2.1 Weak Stationarity

2.1.1 The General Case Given the existence of the unconditional expectation $\mathrm{E}\sigma_t^{(2)}$, standard calculations using the law of iterated expectations show that

$$\mathrm{E}\sigma_t^{(2)} = [I-\beta(1)-\alpha(1)\lambda']^{-1}[\alpha_0+\alpha(1)c], \quad \text{with } c=\sum_{j=1}^k \lambda_j\mu_j^2 = \lambda'\mu^{(2)}. \tag{10}$$

As Equation (10) suggests and Appendix A shows, the necessary and sufficient condition for the existence of the unconditional variance is

$$\det[I-\beta(1)-\alpha(1)\lambda'] > 0. \tag{11}$$

Equation (11) assumes a simple form in the diagonal MN-GARCH case, which is discussed next.

2.1.2 The Diagonal Case For diagonal MN-GARCH processes, defining $\tilde{\beta}_j = 1-\sum_{i=1}^p \beta_{i,jj}$, Equation (A.4) implies

$$\begin{aligned}\det[I-\beta(1)-\alpha(1)\lambda'] &= \det[I-\beta(1)] - \lambda'[I-\beta(1)]^{+}\alpha(1)\\ &= \left[1-\sum_{j=1}^k \frac{\lambda_j}{\tilde{\beta}_j}\sum_{i=1}^q \alpha_{ij}\right]\prod_{j=1}^k \tilde{\beta}_j \\ &= \left[\sum_{j=1}^k \frac{\lambda_j}{\tilde{\beta}_j}\left(1-\sum_{i=1}^q \alpha_{ij} - \sum_{i=1}^p \beta_{i,jj}\right)\right]\prod_{j=1}^k \tilde{\beta}_j,\end{aligned} \tag{12}$$

where $[I-\beta(1)]^{+}$ denotes the adjoint matrix of $I-\beta(1)$. The last expression in Equation (12) implies that it is not necessary that the inequalities $1-\sum_{i=1}^{q}\alpha_{ij}-\sum_{i=1}^{p}\beta_{i,jj}>0$ have to hold for each $j\in\{1,\ldots,k\}$, but rather for their weighted sum with the jth weight being given by $\lambda_j/\tilde{\beta}_j$ and the weights not summing to one.[9] The mixing weight of each component is inflated by the component's contribution to the deterministic part of $\sigma_t^{(2)}$ in Equation (3).

Using Equation (12), Condition (11) can be written as

$$1-\sum_{j=1}^{k}\lambda_j\sum_{i=1}^{q}\alpha_{ij}\left(1-\sum_{i=1}^{p}\beta_{i,jj}\right)^{-1}>0,$$

which shows that it is a direct generalization of the well-known stationarity condition stated in Bollerslev (1986), which can be expressed as

$$1-\sum_{i=1}^{q}\alpha_i\left(1-\sum_{i=1}^{p}\beta_i\right)^{-1}>0.$$

Using Equation (A.2), the unconditional variance of a diagonal MN-GARCH process becomes

$$\begin{aligned}\mathrm{E}(\epsilon_t^2)&=\lambda'\mathrm{E}(\sigma_t^{(2)})+c=c\frac{\det[I-\beta+\alpha_0\lambda'/c]}{\det[I-\beta-\alpha_1\lambda']}\\&=\frac{c+\sum_j\lambda_j\alpha_{0j}/\tilde{\beta}_j}{1-\sum_j\lambda_j/\tilde{\beta}_j\sum_i\alpha_{ij}}=\frac{c+\sum_j\lambda_j\alpha_{0j}/\tilde{\beta}_j}{\sum_j\lambda_j/\tilde{\beta}_j(1-\sum_i\alpha_{ij}-\sum_i\beta_{i,jj})}.\end{aligned}$$

For $k=1$, this reduces to $\mathrm{E}(\epsilon_t^2)=\alpha_0/(1-\sum_i\alpha_i-\sum_i\beta_i)$, as in Bollerslev (1986).

According to Equation (12), the process can have finite variance even though some components are not covariance stationary, as long as the corresponding component weights are sufficiently small. This result is similar to the condition for strict stationarity given by Francq, Roussignol, and Zakoian (2001) for a regime-switching GARCH(1,1) model, for which they show that the condition derived by Nelson (1995) for the single-regime GARCH model need not hold in each regime, but for a weighted average of the GARCH parameters across regimes, with the weights being the stationary probabilities of the Markov chain.

The diagonal MN-GARCH(1,1) model is of special interest in practical applications. Thus the dynamic properties of this process, given by Equation (2) and

$$\sigma_t^{(2)}=\alpha_0+\alpha_1\epsilon_{t-1}^2+\beta\sigma_{t-1}^{(2)},\tag{13}$$

are treated in more detail in Appendix B. In discussing the GARCH(1,1) specification, we will simply write β for matrix β_1 and denote its (diagonal) elements by β_{jj},

[9] Clearly $\prod_j\tilde{\beta}_j>0$ must be assumed, since otherwise the deterministic part of difference Equation (3) would be explosive.

$j = 1, \ldots, k$. In particular, Appendix B presents the condition for the existence of as well as an expression for the unconditional fourth moment and hence kurtosis of the process, together with the autocorrelation function of the squared errors. In case of the existence of the fourth moment, the computation of the autocorrelation function of the squares may be of great interest. Given that GARCH models are mainly used to capture dependencies in second moments, investigators may wish to check how well the dynamic properties implied by an estimated model mimic the corresponding quantities estimated from the data at hand. Expressions for the unconditional third and fourth moments, provided they exist, are given in Equations (B.10) and (B.11). Although higher-order moments are rarely of interest in applications, Equation (B.13) also provides a condition for the existence of the $(2m)$th moment for $m \in \mathbb{N}$.

It is shown that the fourth moment exists if the largest eigenvalue of the $k^2 \times k^2$ matrix

$$C_{22} = 3(\alpha_1 \otimes \alpha_1)\mathrm{vec}[\mathrm{diag}(\lambda)]' + \beta \otimes (\alpha_1 \lambda') + (\alpha_1 \lambda') \otimes \beta + \beta \otimes \beta \tag{14}$$

is less than one. Again, this resembles the result of Bollerslev (1986) for the single-component GARCH(1,1) model, where $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ is required for the fourth moment to exist.[10]

Although the details are deferred to the appendix, we mention here that the evaluation of the autocorrelations of the squared MN-GARCH(1,1) process can be based on an ARMA(k,k) representation of ϵ_t^2 (derived in Appendix A), namely,

$$\det[I - (\beta + \alpha_1 \lambda')L]\epsilon_t^2 = c^\star + \det(I - \beta L)w_t, \tag{15}$$

where $c^\star$ is a constant and $w_t = \epsilon_t^2 - \mathrm{E}(\epsilon_t^2 | \Psi_{t-1})$ is weak white noise. For $\tau \geq k+1$, the autocorrelation at lag $\tau, r(\tau)$, may thus be obtained from the Yule-Walker equations corresponding to Equation (15), analogous to the single-component GARCH process [Bollerslev (1986)]. The first k autocorrelations can be obtained from a closed-form expression for $\mathrm{cov}(\epsilon_t^2, \epsilon_{t-\tau}^2)$, which uses the matrix $\mathrm{E}(\sigma_t^{(2)} \sigma_t^{(2)'})$ as input and is given in Equation (B.15). Equation (15) implies that $r(\tau)$ follows a linear difference equation for $\tau \geq k+1$, that is, it is dominated by an exponential decay at a rate equal to the largest eigenvalue of $C_{11} = \beta + \alpha_1 \lambda'$.

Next, we show that this quantity is a natural measure of the persistence in volatility for MN-GARCH processes.

---

10 The conditions for the existence of $\mathrm{E}(\epsilon_t^4)$ in the GARCH(p,q) model have only recently been established in Chen and An (1998) and Ling and McAleer (2002) [see also He and Teräsvirta (1999) and Karanasos (1999)]. The latter two articles also derive expressions for the autocorrelations of the squared GARCH(p,q) process. However, although this is not made explicit by Chen and An (1998) and Ling and McAleer (2002), the state space representation of the process used therein can easily be employed to obtain much simpler expressions for the autocorrelation function than the rather complicated formulas in He and Teräsvirta (1999) and Karanasos (1999).

2.2 Measuring Volatility Persistence

As shown in Appendix A, the largest eigenvalue, $\rho_{\max}$, of matrix

$$C_{11} = \begin{bmatrix} \beta_1 + \alpha_1\lambda' & \beta_2 + \alpha_2\lambda' & \cdots & \beta_{r-1} + \alpha_{r-1}\lambda' & \beta_r + \alpha_r\lambda' \\ I_k & 0_k & \cdots & 0_k & 0_k \\ 0_k & I_k & & 0_k & 0_k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_k & 0_k & \cdots & I_k & 0_k \end{bmatrix}, \tag{16}$$

with $r = \max\{p, q\}$ and 0_k denoting a $k \times k$ matrix of zeros, can serve as a measure of volatility persistence, since the impact of past variances declines geometrically at the rate $\rho_{\max}$. In the case of an MN-GARCH(1,1) model, this is the largest eigenvalue of the matrix $\beta(1) + \alpha(1)\lambda'$. Analogous to the expression for the single-component case, that is, a normal-GARCH(1,1) model [Bollerslev, Engle, and Nelson (1994)], the conditional expectation of future variances in this model is given by

$$\mathrm{E}\left[\sigma_{t+k}^{(2)} | \Psi_{t-1}\right] = \sigma^{(2)} + [\beta(1) + \alpha(1)\lambda']^k \left(\sigma_t^{(2)} - \sigma^{(2)}\right), \tag{17}$$

where, from Equation (10),

$$\sigma^{(2)} = \mathrm{E}\left(\sigma_t^{(2)}\right) = [I - \beta(1) - \alpha(1)\lambda']^{-1}(\alpha_0 + \alpha_1 c)$$

and $[\beta(1) + \alpha(1)\,\lambda']^k$ tends to zero geometrically at rate $\rho_{\max}$ as $k \to \infty$.

3 NASDAQ RETURNS AND MIXED NORMAL GARCH

We investigate daily returns on the NASDAQ index from its inception in February 1971 to June 2001, a sample of $T = 7681$ observations.[11] Continuously compounded percentage returns, $r_t = 100(\log P_t - \log P_{t-1})$, are considered, where P_t denotes the index level at time t. Figure 1 shows a plot of the return series. While the usual stylized fact of strong volatility clustering is apparent from Figure 1, it is not as obvious that the data are also negatively skewed. The usual measure for asymmetry involving the third moment of the data (let alone its asymptotically valid standard error under normality) may be meaningless, given that third and higher moments of financial data may not exist. Estimating, for example, an unconditional Student's t-distribution produces a degrees-of-freedom estimate of 2.4 with an approximate standard error of 0.08. One possible way to infer if asymmetry is statistically significant is to compare the fit of symmetric and asymmetric mixed normal models using standard likelihood-ratio tests. This is done in the next section.

[11] The data were obtained from the Internet site http://www.marketdata.nasdaq.com, maintained by the Economic Research Department of the National Association of Securities Dealers, Inc.

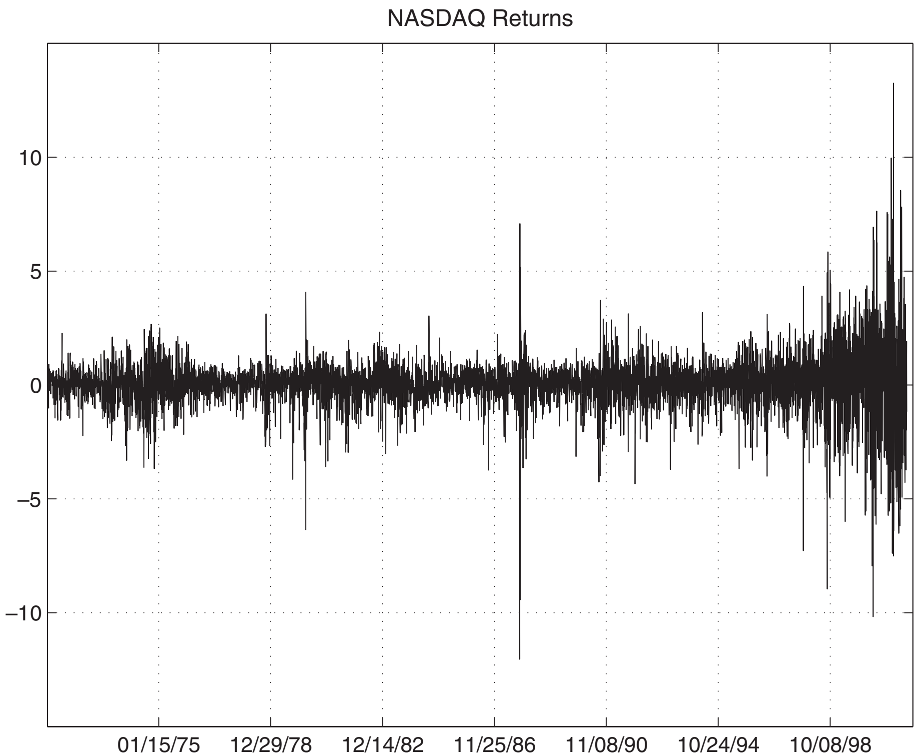

Figure 1 Percentage returns on the NASDAQ index.

Sample (partial) autocorrelation plots of the returns (not shown) and the Bayesian information criterion (BIC) suggest an AR(3) model for the mean equation, which will accompany all reported GARCH structures below.[12]

3.1 Estimation Issues

We estimate the ARMA-MN-GARCH model by conditional ML, conditioning, due to the ARMA structure of Equation (5), on the first u return observations and set the first v values of ϵ_t to zero. For the GARCH structure, we set the initial values of $\sigma_t^{(2)}$ and ϵ_t^2 equal to their unconditional expectations given in Equation (10).[13]

A question that naturally arises in the estimation of mixture distributions is identifiability. Although there is always nonidentifiability due to label switching, this can be ruled out by restricting the parameter space such that no duplication appears, for example, by imposing $\lambda_1 > \lambda_2 > \cdots > \lambda_k$. However, there may be a more serious problem if the class of density functions to be mixed is linearly

[12] All ARMA(r, s)-GARCH(1,1) models for combinations $r + s \leq 4$ were estimated, assuming conditionally normal innovations.

[13] The quasi-Newton maximization method available in Matlab (version 5.2, function fminu) with (automatically computed) numeric gradient and Hessian was used, and a convergence criterion of 10^{-4}.

dependent [Yakowitz and Spragins (1968)]. However, the class of finite mixtures of normals is identifiable, as was established by Teicher (1963) for the univariate setting studied herein; the generalization to multivariate mixtures was established in Yakowitz and Spragins (1968).

Bayesian inference via Markov chain Monte Carlo methods such as the Metropolis Hastings algorithm [see Chib and Greenberg (1995, 1996) and Bauwens, Lubrano, and Richard (1999), and the references therein] is theoretically possible.[14] However, given the large sample sizes typically used in high-frequency financial applications and the lack of strong prior information, conditional ML estimation should yield very similar results. Furthermore, obtaining the ML estimates is computationally easier, in terms of programming effort and run time, as well as the assessment of convergence.

3.2 Determining the Number of Mixture Components

The number of required component densities is generally unknown and needs to be determined empirically. Unfortunately, standard test theory breaks down in this context [see, e.g., McLachlan and Peel (2000:chap. 6)].

Standard model selection criteria, such as the AIC [Akaike (1973)] and the BIC [Schwarz (1978)], are widely used in the GARCH literature and may also be employed to compare models with different numbers of components. For a model with K parameters and log-likelihood, L, evaluated at the maximum likelihood estimator, $\mathrm{AIC} = -2L + 2K$ and $\mathrm{BIC} = -2L + K \log T$, with BIC being more conservative than AIC in that it favors more parsimonious models. Because these measures rely on the same conditions employed in the asymptotic theory of the likelihood ratio test, their small and large sample properties are, likewise, not known. However, the literature on mixtures provides some encouraging evidence in the context of unconditional models, suggesting that the BIC provides a reasonably good indication for the number of components [see McLachlan and Peel (2000: chap. 6), for a survey and further references]. According to Kass and Raftery (1995), a BIC difference of less than 2 corresponds to "not worth more than a bare mention," while differences between 2 and 6 imply positive evidence, differences between 6 and 10 give rise to strong evidence, and differences greater than 10 invoke very strong evidence. The results of Mittnik and Paolella (2000) suggest that, with respect to out-of-sample prediction, these measures are useful for choosing among GARCH-type models with competing distributional assumptions.

3.3 Goodness-of-Fit and Diagnostic Checking

In addition to the likelihood-based model selection via AIC and BIC, we examine the distributional properties of the residuals of the models both in and out of sample. The in-sample fit is evaluated over the whole sample period. The

[14] Chib (1995) discusses the use of marginal likelihoods obtained from MCMC output for determining the number of mixture components in a Bayesian context.

out-of-sample one-step-ahead forecasts are based on the most recent 3000 observations. Using the first 3000 observations for specification of the mean equation employed for out-of-sample forecasting, an AR(3) structure is chosen according to the BIC and using the same procedure as for the entire sample period (see footnote 12 for details). Each estimated parameter vector is used to forecast the return density of the next 20 days, that is, the next 4 weeks, before the models are reestimated using the most recent 3000 observations.

With the MN-GARCH model, it is not possible to directly evaluate the distributional properties of the estimated residuals $\hat{\epsilon}_t$ because, even if the model were correctly specified, standardized residuals would not be identically distributed. To circumvent this, we transform the residuals by computing the corresponding value of the conditional c.d.f., that is,

$$\hat{u}_t = \hat{F}(\hat{\epsilon}_t|\Psi_{t-1}), \quad t = 1, \ldots, T. \tag{18}$$

Under a correct specification, the transformed residuals, $\hat{u}_t$, are i.i.d. uniform [Rosenblatt (1952); see also Diebold, Gunther, and Tay (1998) and Kim, Shephard, and Chib (1998)]. To formally test for uniformity of the transformed values Equation (18)], we follow Palm and Vlaar (1997) and use the Pearson goodness-of-fit test statistic, given by

$$X^2 = \sum_{i=1}^{G} \frac{(n_i - n_i^*)^2}{n_i^*}, \tag{19}$$

where G is the number of (equally spaced) intervals (or bins) into which we group the transformed data, n_i is the number of observations in interval i, and n_i^* is the expected number of observations under the null hypothesis of uniformity. The choice of the number of bins is not evident, but given the large number of observations, we will use $G = 100$ below.

If Equation (19) is used to test a simple hypothesis, the statistic has an asymptotic χ^2 distribution with $G-1$ degrees of freedom under the null. However, if the hypothesis is composite, and Equation (19) is used to test the in-sample fit, that is, the same observations are used both to estimate the parameter vector and to test the goodness-of-fit, the degrees of freedom have to be reduced to compensate for the downward bias in χ^2, as the χ^2 values tend to be smaller when evaluated at the estimated rather than the true parameter values. As a consequence, the asymptotic distribution of Equation (19) is actually unknown, but is bounded between the $\chi^2(G-K-1)$ and $\chi^2(G-1)$ distributions, where K is the number of estimated parameters[15] [see Stuart, Ord, and Arnold (1999:chap. 25)]. To reflect the uncertainty about the true asymptotic distribution of χ^2, we will act as if it were $\chi^2(G-K-1)$ distributed, so that the test tends to favor models with fewer parameters when resulting in similar fit. To test the out-of-sample fit, the $\chi^2(G-1)$ distribution will be used.

[15] If the parameters are determined by minimizing Equation (19), the exact asymptotic distribution is $\chi^2(G-K-1)$ [Stuart, Ord, and Arnold (1999)].

A drawback of the above test is the degree of arbitrariness that is inherent in the choice of the number of classes, G.[16] In addition, one may wish to test whether the specified distribution captures some specific characteristics of the data such as (conditional) skewness and kurtosis.[17] This can be accomplished by a further transformation of the transformed residuals, $\hat{u}_t$, namely

$$z_t = \Phi^{-1}(\hat{u}_t), \tag{20}$$

where Φ is the standard normal c.d.f., such that the z_t's are i.i.d. $N(0, 1)$ distributed, provided the underlying model is correct. Berkowitz (2001) shows that inaccuracies in the specified density will be preserved in the transformed data.[18] Thus this transformation allows the use of moment-based normality tests for checking features such as correct specification of skewness and kurtosis. Palm and Vlaar (1997) and Berkowitz (2001) also made use of the filtered residuals of Equation (20) for testing the dynamic properties of the conditional distribution. We employ the Lagrange multiplier test of Engle (1982) to test whether the conditional volatility is successfully captured by the fitted models. The relevant test statistic,

$$\mathrm{LM}_{\mathrm{ARCH}} = TR^2, \tag{21}$$

is approximately $\chi^2(q)$ distributed, with R^2 denoting the coefficient of determination obtained for the regression $z_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i z_{t-i}^2 + u_t$.

With risk management applications in mind, we also evaluate the models' forecasting performance with respect to value-at-risk (VaR), a widely employed measure to analyze the downside risk of a given financial position [see, e.g., Alexander (2001a)]. The VaR for period t with shortfall probability ξ, denoted by $\mathrm{VaR}_t(\xi)$, associated with model $\mathcal{M}$ is defined by

$$\hat{F}_{t-1}^{\mathcal{M}}(\mathrm{VaR}_t(\xi)) = \xi, \tag{22}$$

where $\hat{F}_{t-1}^{\mathcal{M}}$ is the return distribution function at time t predicted with model $\mathcal{M}$ using information up to time $t-1$. For a correctly specified model, we expect $100 \times \xi\%$ of the observed return values not to exceed the respective VaR forecast. Thus we report the quantities

$$U_\xi = 100 \times \frac{x}{T} = 100 \times \hat{\xi}, \tag{23}$$

where T denotes the number of forecasts evaluated, x is the observed shortfall frequency, and $\hat{\xi} = x/T$ is the empirical shortfall probability. If $\hat{\xi}$ is less (higher) than ξ, then model $\mathcal{M}$ tends to overestimate (underestimate) the risk of the position.

---

[16] For example, the use of values between $G = 50$ and $G = 150$ gave rise to p-values less than 0.01 in 1%, 2%, 1%, and 5% of the cases for models MN(2, 2), MN(3, 2), MN(3, 3), and MN(4, 4), respectively (for the model notation, see Section 1.3). These numbers refer to tests for in-sample fit.

[17] As skewness and kurtosis of a mixture model are (complicated) functions of the model parameters, time variability of the component variances implies time varying skewness and kurtosis.

[18] Use of Equation (20) was also advocated by Palm and Vlaar (1997).

To formally test whether a model correctly estimates the risk inherent in a given financial position, that is, whether the empirical shortfall probability coincides with the specified shortfall probability, ξ, we use the likelihood ratio test statistic [see, e.g., Kupiec (1995)],

$$\mathrm{LR_{VaR}} = -2(\log[\xi^x(1-\xi)^{T-x}] - \log[\hat{\xi}^x(1-\hat{\xi})^{T-x}]) \overset{\text{asy}}{\sim} \chi^2(1). \qquad (24)$$

3.4 Estimated Models

In the following comparison, all models entertained share a common AR(3)-GARCH(1,1) specification, that is, following the notation in Section 1.2, $u=3$, $v=0$, and $p=q=1$. Within the MN-GARCH model class, for a given number of components, k, it turns out that the diagonal model is always preferred over the full model when using the BIC criterion. With respect to the AIC, only for $k=2$, is the full model preferred. For this reason, we restrict our attention to the diagonal models in the following analysis, but briefly discuss the characteristics of the full model for $k=2$ and $k=3$ at the end of this section.

We also fit MS-GARCH models, discussed in Section 1.5, with two, three, and four components. The notation MS(k, g) is used for the MS-GARCH model with k components, g of which exhibit GARCH dynamics and $k-g$ have constant variance.

Table 1 reports the likelihood-based goodness-of-fit measures for the fitted models and their rankings with respect to each of the criteria. Not surprisingly, the worst performer is the standard (one-component) normal-GARCH model. For each criterion, an MN(k, k)-GARCH model is favored, that is, models without suppression of any of the components' dynamics to a constant. The symmetric MN_s-GARCH models perform relatively poorly. Likelihood ratio tests may be used to test for significant differences in the component means, which may also be regarded as testing for skewness in the NASDAQ returns. In the case of two components, the test statistic is $-2\{L[\mathrm{MN}_s(2,2)]-L[\mathrm{MN}(2,2)]\} = 118.8$, which, with one degree of freedom, is clearly significant at any reasonable level. In the case of three components, the likelihood ratio value is 126.0, which is equally significant with two degrees of freedom.[19]

The performance of the MS-GARCH models, with the exception of model MS(4,4), is somewhat surprising, given that they allow for a greater flexibility of the mixing process. However, it should be kept in mind that, for comparisons between the MN- and MS-GARCH models, the likelihood criteria are even less informative than for those between members of the MN class, because the models are not nested due to their differently specified volatility processes (see the discussion in Section 1.5). Model MS(4, 4) exhibits both the highest likelihood and the smallest AIC value, but ranks considerably lower according to the BIC, given that it already involves 31 parameters.

[19] We also estimated a series of single-regime GARCH(1, 1) models with flexible conditional distributions such as skewed t and hyperbolic. None of these is favored against the best MN models, according to the criteria shown in Table 1. The results are reported in an earlier version of this article [Haas, Mittnik, and Paolella (2002)], and are available upon request.

Table 1 Likelihood-based goodness-of-fit.

		L		AIC		BIC	
Distributional model	K	Value	Rank	Value	Rank	Value	Rank
Normal	7	−9142.8	14	18299.5	14	18348.1	14
MN(2, 1)	10	−8962.7	13	17945.4	13	18014.8	13
MN(3, 1)	13	−8931.4	11	17888.7	12	17979.0	12
MN(3, 2)	15	−8857.5	6	17745.1	6	17849.2	2
MN(2, 2)	12	−8872.5	7	17768.9	7	17852.4	3
MN(3, 3)	17	−8845.5	4	17725.0	4	**17843.1**	**1**
MN(4, 4)	22	−8831.7	3	17707.5	2	17860.3	4
MN(5, 5)	27	−8828.1	2	17710.2	3	17897.7	7
MN_s(2, 2)	11	−8931.9	12	17885.7	11	17962.2	11
MN_s(3, 3)	15	−8908.5	10	17847.0	10	17951.2	10
MS(2, 2)	13	−8889.3	9	17804.6	8	17894.9	6
MS(3, 2)	19	−8884.2	8	17806.4	9	17938.4	9
MS(3, 3)	21	−8847.0	5	17735.9	5	17881.8	5
MS(4, 4)	31	**−8818.4**	**1**	**17698.7**	**1**	17914.0	8

The table shows three likelihood-based goodness-of-fit measures for models fitted to the NASDAQ return series. Each model includes an AR(3) specification for the mean. The left-most column specifies the type and order of the model: MN(k,g) is short for mixed normal(k, g)-GARCH(1,1), and indicates the MN-GARCH(1,1) process with k components, g of which follow a GARCH process and $k-g$ components being restricted to having constant variances. MN_s is similar but refers to the symmetric MN-GARCH model. MS(k, g) refers to the Markov switching model with k components, where g of them follow a GARCH process. The column labeled K reports the number of parameters of a model; L is the log-likelihood; $AIC=-2L+2K$; and $BIC=-2L+K\log T$. For each of the three criteria, the criterion value and the ranking of the models are shown. Boldface entries indicate the best model for the particular criterion.

We shall make an attempt to empirically compare the variance processes of the MN(k, k) and MS(k, k) models for the present example. Namely, the empirical correlation matrices of the component variance processes, $\{\sigma_{1t}^2, \ldots, \sigma_{kt}^2\}_{t=1}^T$, can be compared. The correlations between the component variances may also indicate how well the GARCH structure in the mixture components can be described by linear relations of the form $\sigma_{it}^2 = a + b\sigma_{jt}^2$, $i \neq j$, as in the models of Vlaar and Palm (1993) and Bauwens, Bos, and van Dijk (1999), discussed in Section 1.4. Table 2 reports the correlation matrices for $k=2,3,4$. From these matrices, it can be seen that the linear relation between the component variances is very strong in the two- and three-component MS models, while colinearity is less pronounced for the MN specifications. This may be a consequence of the fact that, in Gray's model, all regime variances have the same right-hand-side variables in the GARCH Equation (7). In the four-component MS model, the colinearity reduces somewhat, but it is still stronger than in the MN(4, 4) model. As the estimated MN(4, 4) model implies an existing fourth moment (although this conclusion is very problematic, see the discussion below), we can also compute the theoretical correlation matrix of the vector of component variances as implied by the parameter estimates, using the expressions for $\mathrm{E}(\sigma_t^{(2)})$ and $\mathrm{E}(\sigma_t^{(2)}\sigma_t^{(2)'})$ derived in Appendix B. Doing so, we

Table 2 Empirical correlation structures of component variances.

Model	Number of mixture components, k: 2	3	4
MN(k,k)	$\begin{bmatrix} 1 & 0.89 \\ & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.90 & 0.72 \\ & 1 & 0.91 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.82 & 0.99 & 0.88 \\ & 1 & 0.79 & 0.62 \\ & & 1 & 0.91 \\ & & & 1 \end{bmatrix}$
MS(k,k)	$\begin{bmatrix} 1 & 0.99 \\ & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.96 & 0.99 \\ & 1 & 0.96 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.79 & 0.99 & 0.97 \\ & 1 & 0.87 & 0.92 \\ & & 1 & 0.99 \\ & & & 1 \end{bmatrix}$

Shown are the empirical correlation matrices of the component variance processes $\{\sigma_{1t}^2, \ldots, \sigma_{kt}^2\}_{t=1}^T$ implied by the fitted models for the NASDAQ returns.

obtain

$$\operatorname{corr}(\sigma_t^{(2)}) = \begin{bmatrix} 1 & 0.79 & 0.99 & 0.89 \\ & 1 & 0.76 & 0.60 \\ & & 1 & 0.91 \\ & & & 1 \end{bmatrix}, \tag{25}$$

a fairly close match of the empirical correlations reported in Table 2. These results suggest that there is a need for a more general MN-GARCH structure than implied by the linear subclass discussed in Section 1.4.

In view of the results reported in Table 1, the MN$(3,3)$ and MN$(4,4)$, as well as the MN$(2,2)$ and MN$(3,2)$ models are retained for further consideration, and estimation results along with approximate standard errors[20] are shown in Table 3.[21] There, the components are ordered with respect to decreasing component means, μ_j, which also corresponds to an ordering with respect to increasing α_{1j} (with the necessary exception of the third component of model MN$(3,2)$), decreasing mixing weights, and a decreasing β_{jj}.[22] The results indicate a clear relationship between the component mean, μ_j, and the component dynamics determined by α_{1j} and β_{jj}. As μ_j drops, α_{1j} increases, reflecting an increasing responsiveness to (negative) shocks, while there tends to be more inertia in σ_{jt}^2 when shocks are positive, as is reflected by the growing values of β_{jj}.

Another striking result is that the volatility dynamics are stable in the sense that $\alpha_{1j} + \beta_{jj} < 1$ when $\mu_j \geq 0$ and unstable, that is, $\alpha_{1j} + \beta_{jj} > 1$, for $\mu_j < 0$.[23]

[20] Standard errors were obtained by numerically computing the Hessian matrix at the ML estimates. The delta method was used to approximate the standard errors of functions of estimated quantities, namely, $\alpha_{1i} + \beta_{ii}$, $i = 1, \ldots, 4$, as well as the weights and means of the last component of each of the models.

[21] For comparison purposes, results for the standard normal-GARCH model are also given.

[22] A sole exception are components 1 and 2 in the MN$(4,4)$ model.

[23] We use the term "stable" to refer to covariance stationarity. This does not imply that components are explosive [cf. Nelson (1995)].

Table 3 MN-GARCH parameter estimates for NASDAQ returns.

	Normal	MN(2, 2)	MN(3, 3)	MN(3, 2)	MN(4, 4)
α_{01}	0.014	0.002	0.000	0.001	0.003
	(0.0018)	(0.0008)	(0.0007)	(0.0008)	(0.0032)
α_{11}	0.117	0.051	0.022	0.038	0.067
	(0.0083)	(0.0066)	(0.0080)	(0.0068)	(0.0192)
β_{11}	0.869	0.920	0.956	0.934	0.855
	(0.0089)	(0.0090)	(0.0137)	(0.0101)	(0.0461)
$\alpha_{11}+\beta_{11}$	0.986	0.971	0.978	0.972	0.922
	(0.0032)	(0.0037)	(0.0063)	(0.0044)	(0.0345)
λ_1		0.820	0.541	0.724	0.373
	1	(0.0255)	(0.0879)	(0.0427)	(0.1182)
μ_1	0	0.091	0.164	0.119	0.200
		(0.0100)	(0.0233)	(0.0133)	(0.0367)
$E\sigma_1^2$	0.986	0.525	0.370	0.460	0.329
α_{02}	—	0.075	0.012	0.027	0.000
		(0.0235)	(0.0055)	(0.0122)	(0.0011)
α_{12}	—	0.512	0.197	0.379	0.015
		(0.0941)	(0.0425)	(0.0685)	(0.0054)
β_{22}	—	0.727	0.835	0.768	0.980
		(0.0457)	(0.0260)	(0.0357)	(0.0071)
$\alpha_{12}+\beta_{22}$	—	1.239	1.031	1.146	0.995
		(0.0588)	(0.0244)	(0.0413)	(0.0031)
λ_2	0	0.180	0.433	0.272	0.317
		(0.0255)	(0.0832)	(0.0431)	(0.0743)
μ_2	—	−0.415	−0.153	−0.281	0.035
		(0.0575)	(0.0548)	(0.0508)	(0.0700)
$E\sigma_2^2$	—	1.741	0.926	1.355	0.506
α_{03}	—	—	0.332	0.825	0.005
			(0.1913)	(0.6246)	(0.0086)
α_{13}	—	—	1.303		0.246
			(0.5179)	—	(0.0724)
β_{33}	—	—	0.567		0.824
			(0.1389)	—	(0.0393)
$\alpha_{13}+\beta_{33}$	—	—	1.870		1.070
			(0.4209)	0	(0.0416)
λ_3	0	0	0.026	0.004	0.289
			(0.0111)	(0.0028)	(0.0841)
μ_3	—	—	−0.865	−2.281	−0.232
			(0.2223)	(0.7251)	(0.0824)
$E\sigma_3^2$	—	—	2.936	0.825	0.974

continued

Table 3 (*continued*)

	Normal	MN(2, 2)	MN(3, 3)	MN(3, 2)	MN(4, 4)
α_{04}	—	—	—	—	0.373 (0.2158)
α_{14}	—	—	—	—	1.427 (0.6015)
β_{44}	—	—	—	—	0.546 (0.1524)
$\alpha_{14}+\beta_{44}$	—	—	—	—	1.973 (0.5002)
λ_4	0	0	0	0	0.021 (0.0089)
μ_4	—	—	—	—	−0.894 (0.2412)
$E\sigma_4^2$	—	—	—	—	2.941
$\rho_{\max}$	0.986	0.985	0.989	0.986	0.994
$\rho(C_{22})$	0.999	1.004	1.002	1.003	0.999

Standard errors of parameter estimates are given in parentheses. Column MN(k,g) indicates the MN-GARCH(1,1) with k components, g of which follow a GARCH process and $k-g$ components being restricted to having constant variances. $E\sigma_i^2$, $i=1,\ldots,4$, denotes the unconditional variance of component i, as computed from Equation (10). $\rho_{\max}$ is the measure of volatility persistence, that is, the largest eigenvalue of the matrix in Equation (16), and $\rho(C_{22})$ is the largest eigenvalue of the matrix in Equation (14). $\rho(C_{22})>1$ indicates that the unconditional fourth moment does not exist.

However, all estimated models themselves are stationary. Their respective volatility persistence measures, $\rho_{\max}$, reported in the bottom part of Table 3, are less than unity. This is due to the fact that the unstable components have sufficiently small mixing weights. In model MN(3, 3), the first component is rather similar to the first component in model MN(2, 2) and responds rather slowly to shocks. The second component, although just unstable ($\alpha_{12}+\beta_{22}=1.037$), is more similar to the normal-GARCH model and has an intermediate position. The third component, however, very strongly reacts to shocks, as is reflected by the large value of α_{13}; it is also characterized by a relatively high value for constant α_{03} and is highly unstable, with $\alpha_{13}+\beta_{33}=1.870$. Observe also that, in each model with two or more components, the higher the volatility (as measured by the estimate of $\alpha_{1i}+\beta_{ii}$ and the unconditional component variances $E\sigma_i^2$, $i=1,\ldots,4$), the lower $\hat{\mu}_i$; that is, negative means arise in conjunction with higher variance. This finding is compatible with the well-known leverage effect [Black (1976)], which refers to the tendency that negative returns coincide with high volatility [see, e.g., Bekaert and Wu (2000)].[24]

[24] A number of GARCH models exist that incorporate an asymmetric relation between risk and return, such as the EGARCH of Nelson (1991) and the model of Glosten, Jagannathan, and Runkle (1993); see also Bollerslev, Engle, and Nelson (1994).

The responsiveness to shocks of the individual components is illustrated in Figure 2. It shows (the square roots of) the variances in the normal-GARCH and those associated with the components of the MN(3, 3) model. The graphs clearly reveal the rather calm behavior of the first component and the more or less hectic movements of component 3, while σ_{2t} mimics the evolution of σ_t for the normal-GARCH model. The relatively large value of constant $\alpha_{03} = 0.332$ in component 3 is reflected in the high floor of σ_{3t} at roughly one.

Returning to Table 3, the first two components of the MN(3, 2) model resemble those of the MN(2, 2) model. With a component mean of $\mu_3 = -2.281$ and the rather small weight of $\lambda_3 = 0.004$, the third component captures the large negative shocks and amounts to a jump process that does not include any conditional volatility dynamics. Model MN(4, 4) is quite similar to model MN(3, 3), but with the stable component with positive mean being split into two positive stable components.

The last row of Table 3 reports the largest eigenvalue of matrix C_{22} defined in Equation (14). $\rho(C_{22}) > 1$ indicates that the unconditional fourth moment does not exist. The values are all very close to unity. They fall just below one for the single-component and the MN(4, 4) process, allowing us, at least in principle, to compute the autocorrelation function of the squared errors for these two models, using the results in Bollerslev (1988) and Appendix B. Figure 3 shows the theoretical autocorrelations from the fitted models along with their empirical counterpart computed from the AR(3) residuals. Similar to the findings in Ding, Granger, and Engle (1993) for returns on the S&P 500 index, for the NASDAQ, the autocorrelations of the squares are significantly positive over very long-lags and show some type of long-memory behavior. Also, as shown in Ding, Granger, and Engle for the S&P 500 returns, the initial theoretical autocorrelations of the standard GARCH(1,1) model shown in the top plot of Figure 3 are too large and the autocorrelations decrease too fast for high lags. The theoretical autocorrelation function of the MN(4, 4) model, shown in the bottom plot of Figure 3, fits its empirical counterpart much better over a long horizon, but also decreases somewhat too fast for high lags, although less dramatically. This follows from the fact (shown in Appendix B) that the autocorrelation function of MN-GARCH models, similar to those of standard GARCH models, are dominated by exponential decay.

The conditional skewness of the fitted MN(3, 3) model is plotted in the top panel of Figure 4.[25] With the increase in the conditional variance of the process toward the end of the dataset, the implied skewness moves toward zero. The center of Figure 4 shows the conditional density when the skewness reaches its most extreme value of -1.56. Its remarkable departure from symmetry and the wide range of implied skewness values (see top panel) emphasize the importance of time-varying skewness in this dataset. The bottom panel shows the implied kurtosis, which appears to have a "natural lower bound" of three, which can be explained by the leptokurtosis of a scale mixture (where the component means are equal) and the fact that $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$ are relatively close in value.

[25] Figure 4 shows the usual moment-based measures of skewness and kurtosis, as defined in the footnote of Table 4.

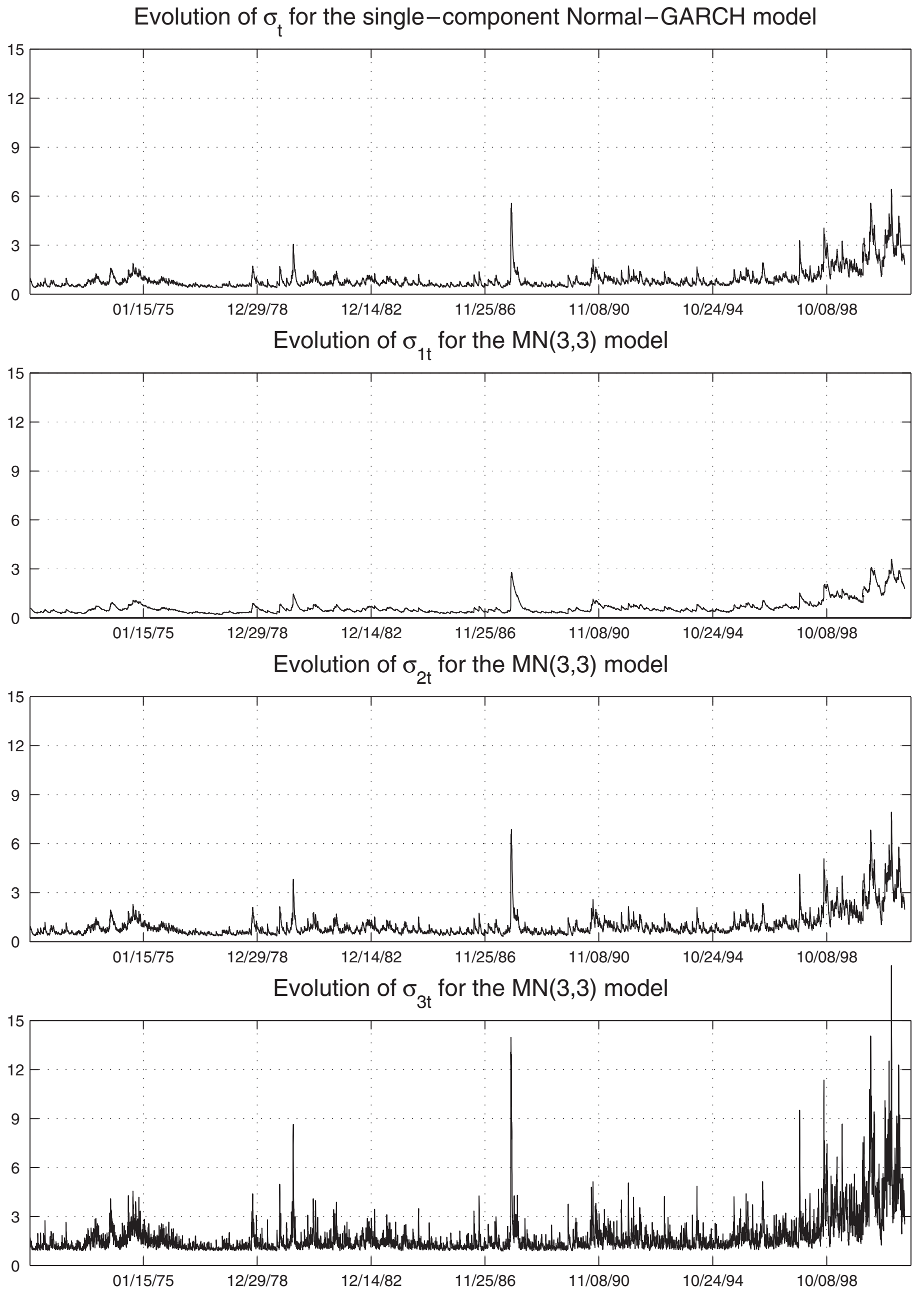

Figure 2 Volatility evolution for the single-component normal-GARCH and the MN(3,3)-GARCH models.

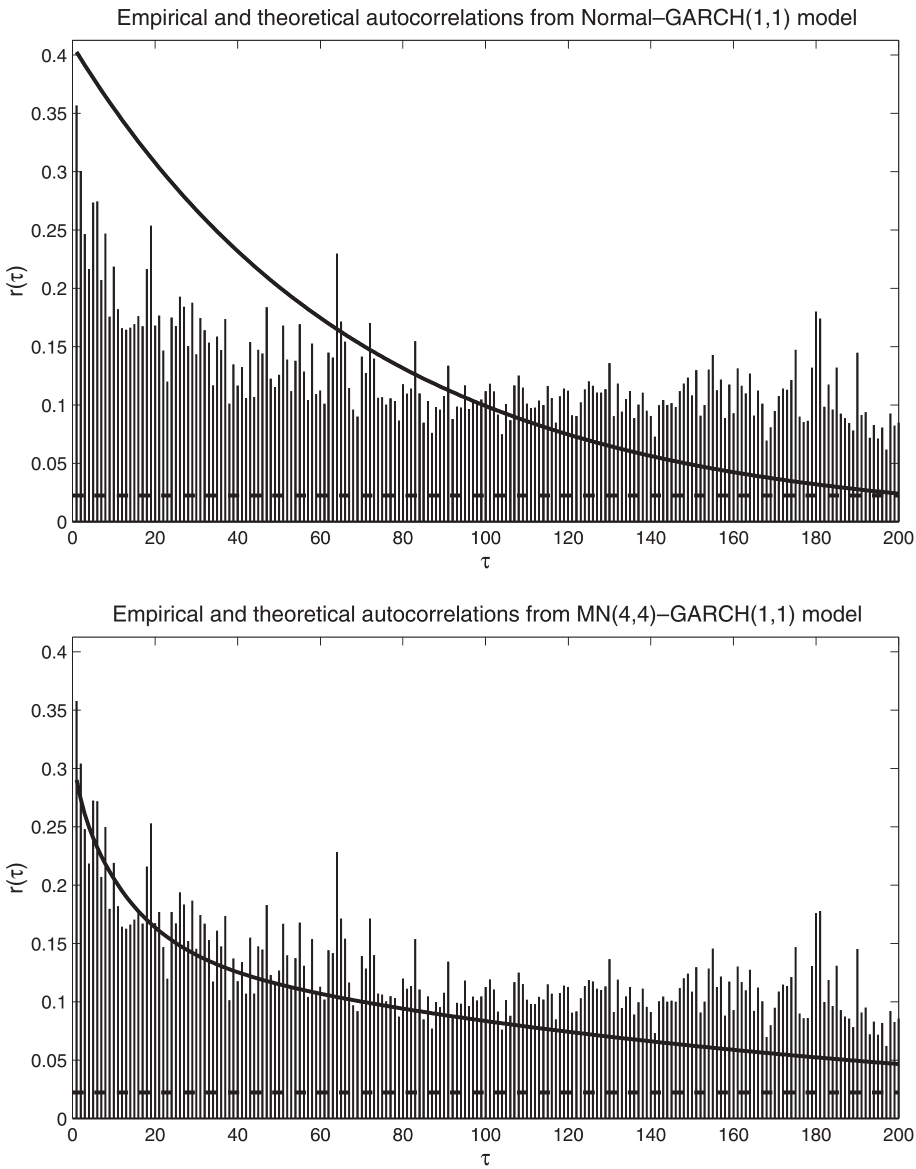

Figure 3 The top plot shows the empirical autocorrelations of the squared residuals from the AR(3) model for the NASDAQ returns, along with their theoretical counterpart computed from the fitted single-component GARCH(1,1) model. The latter are computed by $r(1) = \frac{\alpha_1(1-\alpha_1\beta_1-\beta_1^2)}{1-2\alpha_1\beta_1-\beta_1^2}$ and $r(\tau) = (\alpha_1 + \beta_1)r(\tau - 1)$, for $\tau > 1$ [see Bollerslev (1988)]. The bottom plot repeats this, but for the MN(4,4)-GARCH(1,1) model, for which $r(\tau)$ is derived in Appendix B. Dashed lines represent approximate 95% confidence bands.

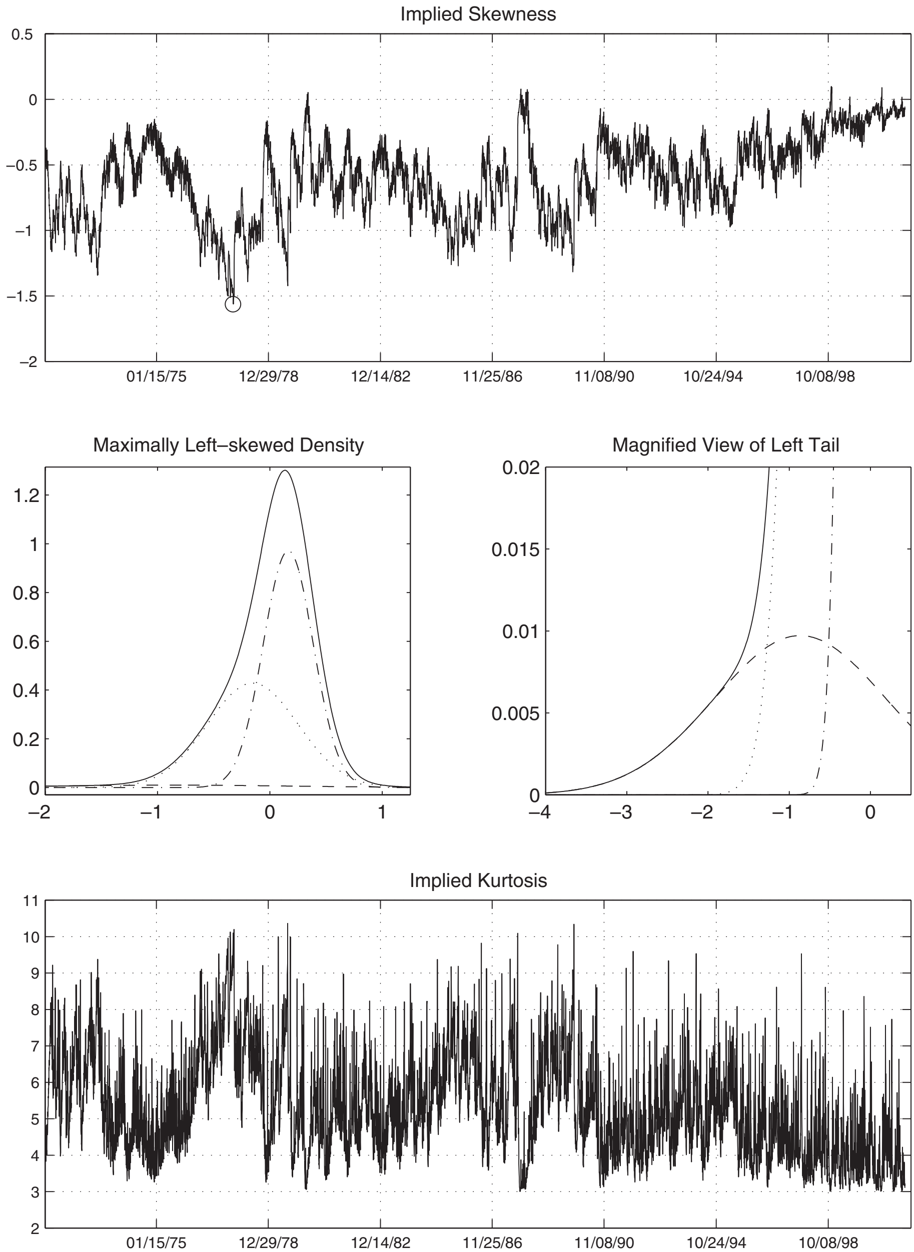

Figure 4 Top plot shows the implied skewness of fitted conditional densities for the NASDAQ data using the MN(3,3) model, with the inscribed circle indicating the maximal implied left skewness of −1.56, the density (solid line) of which is plotted in the center panel together with the weighted component densities (dashed, dotted, and dash-dotted lines); the right graph in the center panel is a magnification of the left tail. The bottom plot shows the implied kurtosis.

3.5 Diagnostic Checks and Forecasting Performance

This section reports the results of the in- and out-of-sample diagnostic tests discussed in Section 3.3. We do not include the MS(4, 4) model in the out-of-sample comparison, since the out-of-sample results for model MS(3, 3) show already clear-cut signs of overparameterization (see below). More importantly, perhaps, sensible estimation of the MS(4, 4) model over smaller samples turned out to be close to impracticable.

The results for the in-sample fit are shown in Table 4. All "skewed" models of the MN(k, g) class pass the Pearson goodness-of-fit test [Equation (19)], while the symmetric MN(k, g) and the MS-GARCH models do not. However, the two-component MN(2, 2) model fails to adequately capture the kurtosis, as does the MS(2, 2) specification. Also note that the two symmetric MN_s mixture models are able to accomodate the excess kurtosis from the residuals, but clearly fail to capture the skewness. This underscores once more the importance of skewness in the NASDAQ returns. It is striking that all MS-GARCH models do not pass the moment test for skewness, although they do in fact allow for skewness through different regime means. We suppose that this failure is a result of the linkage between the variance dynamics and the component means reflected in Equation (6). The fact that the component means determine both the conditional variances and the skewness permits the means of the MS models to properly capture the skewness in the data. With respect to conditional heteroskedasticity, all MN models — except the symmetric $MN_s(2, 2)$ model — pass the in-sample LM tests for ARCH. Among the MS models, only the MS(3, 3) and MS(4, 4) models pass the test. The fact that the single-component GARCH model solidly fails all tests conducted indicates that the different component variance processes capture important features of the data.

Summarizing the results in Table 4, it appears that the asymmetric diagonal MN(3, 3) and MN(4, 4) models provide an adequate description of the NASDAQ series over the entire sample period, which coincides with the recommendations of the AIC and BIC criteria reported in Table 1.

The forecasting performance of the models with respect to the tests discussed in Section 3.3 as well as VaR prediction are shown in Table 5. Out of sample, none of the models but the three-component MS model passes the Pearson test for uniformity of the residuals after the transformation in Equation (18). However, this model fails to appropriately forecast the conditional skewness and kurtosis and the conditional volatility. In addition, as reported in the bottom part of Table 5, it does not produce satisfactory forecasts of downside shortfall probabilities, which are of great interest in risk management. Altogether, the unsatisfactory performance of the MS(3, 3) model may be interpreted as reflecting overparameterization. This view is supported by the observation that MS(2, 2) and the MS(3, 2) specifications perform somewhat better in this forecasting exercise. Although there is still a certain amount of "excess skewness" in the "normalized" residuals [Equation (20)], these models adequately capture conditional kurtosis and volatility, and their VaR forecasts

Table 4 In-sample fit of AR(3)-GARCH(1,1) MN and MS models.

	Normal	$MN_s(2,2)$	$MN_s(3,3)$	MN(2, 2)	MN(3, 2)	MN(3, 3)	MN(4, 4)	MS(2, 2)	MS(3, 2)	MS(3, 3)	MS(4, 4)
χ^2	376.5***	206.3***	242.4***	94.39	96.39	93.24	82.09	119.4**	136.5***	99.88**	104.8***
Skewness	−0.672***	−0.298***	−0.290***	−0.088***	−0.054*	−0.046	−0.037	−0.167***	−0.159***	−0.159***	−0.149***
Kurtosis	2.521***	0.064	0.027	0.134**	0.094*	0.011	−0.012	0.206***	0.151***	0.005	−0.074
JB	2609.7***	115.1***	108.1***	15.6***	6.5**	2.7	1.8	49.2***	39.7***	32.2***	30.2***
ARCH(1)	14.93***	4.364**	1.458	1.788	1.927	1.164	0.714	8.642***	7.745***	1.350	0.060
ARCH(2)	15.05***	4.976*	1.471	1.907	1.936	1.191	0.714	8.629**	7.732**	1.454	1.340
ARCH(3)	15.05***	4.970	1.574	2.138	2.285	1.456	1.288	9.214**	8.275**	1.998	2.414
ARCH(4)	15.20***	5.040	1.581	2.167	2.425	1.488	1.368	9.648**	8.694*	2.010	2.479
ARCH(5)	15.20***	5.849	2.264	2.398	2.644	1.646	1.476	9.656*	8.712	2.732	4.012
ARCH(10)	20.78**	11.71	7.587	6.596	7.474	5.850	5.428	17.06*	16.28*	9.384	9.448

The upper part of the table reports test results on the distributional properties of the transformed residuals. χ^2 refers to the Pearson goodness-of-fit test of Equation (19) after the transformation of Equation (18), with $G = 100$ and $G - K - 1$ degrees of freedom, where K is the number of parameters of the respective models. The last three rows of the upper part are based on the transformation of Equation (20). "Skewness" denotes the coefficient of skewness $\gamma_1 = m_3/m_2^{3/2}$ and "Kurtosis" the coefficient of excess kurtosis $\gamma_2 - 3 = m_4/m_2^2 - 3$. Here

$$m_i = T^{-1}\sum_{t=1}^{T}(z_t - \tilde{z})^i, \quad i = 2,3,4, \quad \tilde{z} = T^{-1}\sum_{t=1}^{T} z_t.$$

Under normality, $T\gamma_1^2/6 \sim \chi^2(1)$ and $T(\gamma_2 - 3)^2/24 \sim \chi^2(1)$ asymptotically. JB is the value of the Jarque and Bera (1987) Lagrange multiplier test for normality, that is,

$$\mathrm{JB} = T\gamma_1^2/6 + T(\gamma_2 - 3)^2/24 \overset{\text{asy}}{\sim} \chi^2(2)$$

under normality [cf. Lütkepohl (1991: 152–156)]. The lower part reports the values of Engle's (1982) Lagrange multiplier test for conditional heteroskedasticity, given in Equation (21) and applied to the residuals after the transformation of Equation (20).

Asterisks *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5 Out-of-sample fit of AR(3)-GARCH(1,1) MN and MS models.

	Normal	$MN_s(2,2)$	$MN_s(3,3)$	MN(2, 2)	MN(3, 2)	MN(3, 3)	MN(4, 4)	MS(2, 2)	MS(3, 2)	MS(3, 3)
χ^2	204.9***	184.5***	178.3***	126.8**	118.2*	134.1**	124.2**	163.8***	155.2***	109.1
Skewness	−0.630***	−0.260***	−0.245***	−0.048	0.010	0.027	0.028	−0.065*	−0.066*	−0.084**
Kurtosis	2.704***	0.090	0.003	0.362***	0.205***	0.124*	0.150**	0.056	0.020	0.238***
JB	1735.0***	54.3***	46.8***	27.3***	8.2**	3.6	5.0*	3.9	3.4	16.6***
ARCH(1)	7.271***	1.950	1.646	5.527**	1.152	4.025**	0.024	1.631	1.429	4.560**
ARCH(2)	10.08***	8.623**	8.047**	8.331**	3.548	6.025**	1.485	2.394	2.164	7.709**
ARCH(3)	10.08**	8.829**	8.128**	8.350**	3.676	6.300*	2.178	3.024	2.654	8.707**
ARCH(4)	10.42**	8.880*	8.206*	8.660*	3.718	6.527	2.192	3.374	3.219	8.717*
ARCH(5)	10.51*	11.08**	10.22*	8.914	4.994	6.549	2.393	3.733	3.561	10.33*
ARCH(10)	14.09	15.32	14.32	13.74	7.582	12.48	5.100	8.156	7.741	13.28
$U_{0.001}$	0.855***	0.278***	0.214**	0.278***	0.214**	0.171	0.128	0.256***	0.235**	0.385***
$U_{0.005}$	1.496***	0.812***	0.833***	0.598	0.556	0.684*	0.598	0.556	0.534	0.855***
$U_{0.01}$	2.073***	1.688***	1.731***	1.047	1.068	1.218	1.090	0.897	0.919	1.282*
$U_{0.025}$	4.038***	4.209***	3.803***	2.756	2.778	2.863	2.799	2.671	2.735	3.120***
$U_{0.05}$	6.667***	7.521***	7.137***	5.876***	6.154***	5.855***	5.962***	6.218***	6.132***	6.432***
$U_{0.1}$	10.68	12.33***	12.48***	11.30***	11.13**	11.03**	11.20***	11.73***	11.54***	11.94***

For the upper and middle part, see the legend of Table 4 for explanations, but note the following modification: χ^2 refers to the Pearson goodness-of-fit test of Equation (19) after the transformation of Equation (18), with $G = 100$ and $G - 1$ degrees of freedom. The bottom part reports results from forecasting the value-at-risk (VaR). Reported are the values U_ξ defined in Equation (23), referring to the empirical percentage shortfall frequency corresponding to the theoretical shortfall probability ξ. Asterisks *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

are reasonably accurate for the practically important levels ranging from $\xi = 0.005$ to $\xi = 0.025$.

Turning to the MN models, we find that the symmetry restrictions are inadequate. The models belonging to the symmetric MN subclass capture the kurtosis, but as before, fair poorly with respect to skewness. Also, these symmetric models give no adequate description of the volatility process out of sample, as they do not pass the ARCH test reported in the middle part of Table 5. The skewed MN models all successfully predict the conditional skewness and provide accurate forecasts of downside risk, but there is some excess kurtosis in the residuals for all of them (although with differing degrees). A surprising result in Table 5 is perhaps the performance of model MN(4, 4). As this model with 22 parameters (vis-à-vis 21 parameters of the MS(3, 3) model) clearly raises suspicion of being highly overparameterized, a miserable out-of-sample performance is expected. However, both the volatility and the VaR forecasts of this specification are satisfactory. As in the in-sample case, model MN(2, 2) fails to adequately capture important features of the data (but accounts for the skewness), whereas models MN(3, 2), MN(3, 3), and MN(4, 4) now seem to be acceptable among the candidates of this class. Taken altogether, the MS-GARCH models MS(2,2) and MS(3,2) also perform well.

3.6 Empirical Results for the Nondiagonal Models

When relaxing the diagonality restriction on β, we obtain a triangular structure for $\hat{\beta}$, namely

$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{bmatrix} = \begin{bmatrix} 0.002 \\ 0.077 \end{bmatrix} + \begin{bmatrix} 0.051 \\ 0.538 \end{bmatrix} \epsilon_{t-1}^2 + \begin{bmatrix} 0.918 & 0.000 \\ 0.447 & 0.572 \end{bmatrix} \begin{bmatrix} \sigma_{1t-1}^2 \\ \sigma_{2t-1}^2 \end{bmatrix}, \tag{26}$$

with $\lambda = (0.806,0.194)'$ and $\mu = (0.095, -0.395)'$. The log likelihood, AIC, and BIC of the model are -8869.4, 17764.8, and 17855.1, respectively. Comparison with Table 1 shows that, while AIC prefers the full model, BIC prefers the diagonal specification.

Estimation of a full three-component model yields

$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \\ \sigma_{3t}^2 \end{bmatrix} = \begin{bmatrix} 0.005 \\ 0.000 \\ 0.007 \end{bmatrix} + \begin{bmatrix} 0.029 \\ 0.283 \\ 0.251 \end{bmatrix} \epsilon_{t-1}^2 + \begin{bmatrix} 0.221 & 0.202 & 0.010 \\ 1.572 & 0.235 & 0.001 \\ 0.000 & 0.000 & 0.968 \end{bmatrix} \begin{bmatrix} \sigma_{1t-1}^2 \\ \sigma_{2t-1}^2 \\ \sigma_{3t-1}^2 \end{bmatrix},$$

with $\lambda = (0.622, 0.372, 0.006)'$, $\mu = (0.145, -0.220, -1.375)'$, and $\rho_{\max} = 0.994$. The log-likelihood is -8842.4, which is a negligible improvement compared to the diagonal model with -8845.5.

3.7 Extension to Fat-Tailed Components: The Mixed-t GARCH

An obvious extension of the MN-GARCH model is to replace the normal distribution with a fatter-tailed alternative. This may improve the fit of MN models for a given number of components. This is, for example, the case for the

MN(2, 2)-GARCH(1, 1) model and potentially renders a third component unnecessary. In the case of flexible component densities, these are characterized by an additional shape parameter, which may or may not differ across the components. As the conditional variance of ϵ_t is affected by this shape parameter, we have

$$E[\epsilon_t^2|\Psi_{t-1}] = \sum_{j=1}^{k} \lambda_j \kappa_j \sigma_{jt}^2 + c,$$

where c is as in Equation (10) and κ_j is a function of the shape parameter of the jth component. For example, if the component densities are Student's t with ν_j degrees of freedom, $j = 1, \ldots, k$, then $\kappa_j = \nu_j/(\nu_j - 2)$. The stationarity conditions are easily extended using Equation (B.13) and the discussion in Appendix B.

Using a mixture of Student's t distributions for the NASDAQ, consider first the case where the degrees of freedom parameter, ν, is the same for all mixture components. The resulting model then generalizes that proposed by Neely (1999), who uses the Student's t with the Vlaar and Palm (1993) model. In the two-component case, we estimate $\hat{\nu} = 14.8$ (with standard error 3.4), indicating a relatively mild deviation from normality. With log-likelihood value -8862.4 and $K = 13$ parameters, this results in an AIC value of 17750.9 and a BIC value of 17841.2. Thus the AIC favors the MN(3, 3)-GARCH(1, 1) formulation, while the BIC is virtually indifferent. This indicates that the introduction of fatter-tailed component densities does not imply that we can neglect the dynamics associated with the additional component.

Allowing the degrees of freedom to differ across mixture components,[26] for the low-volatility component we have $\hat{\nu}_1 = 44.2$ (with "standard error" 47.9), while the high-volatility, nonstationary component yields $\hat{\nu}_2 = 8.63$ (with standard error 1.81). The values for log-likelihood, AIC, and BIC of this model with $K = 14$ parameters are -8857.9, 17743.8, and 17841.0, respectively. Regarding the comparison with the MN(3, 3) model, the discussion in the previous paragraph applies here as well.

For the three-component Student's t model (with identical degrees of freedom), our estimate is $\hat{\nu} = 109.8$ (with a — meaningless — standard error of 310), and the log-likelihood value is -8845.1, clearly indicating the adequacy of the mixture of normals for the data at hand.

4 CONCLUSION

We have investigated the properties and the usefulness of a class of conditionally heteroskedastic and heavy-tailed models for financial return series that retains the normality assumption via a mixed normal structure. The model gives rise to rich dynamics, including time-varying skewness and kurtosis, which are not captured by GARCH models driven with innovations governed by the usual

[26] A Markov model with switching in the degrees of freedom parameter was examined by Dueker (1997) for the returns on the S&P 500 index.

single-component asymmetric fat-tailed distributions. In an application based on the NASDAQ index returns, we have shown that the model offers a disaggregation of the conditional variance process that is amenable to economic interpretation, including the well-known leverage effect. In- and out-of-sample diagnostic tests show that the class of MN-GARCH models adequately captures relevant properties of return data.

There are several possible generalizations of the proposed model that deserve future investigation. First, allowing for time-varying mixture weights, as proposed in Vlaar and Palm (1993) and implemented, for example, in Beine and Laurent (2003), who model exchange rates with component weights depending on central bank interventions. In applications to stock market returns, it might be more promising to experiment with specifications relating current mixing weights to past returns. For example, a process that places more weight on the high-variance component tomorrow when today's return is negative may be well suited for modeling the leverage effect. Second, more general GARCH structures, such as those proposed by Ding, Granger, and Engle (1993), or long-memory effects as proposed by Baillie, Bollerslev, and Mikkelsen (1996), could be entertained in order to further improve the analysis of dependencies in the squared process. In addition, multivariate models for more realistic problems of risk management are desirable, where it may be particularly promising to couple the univariate specification developed here with factor structures as in Alexander (2001b).

APPENDIX A: DERIVATION OF STATIONARITY CONDITION (11)

By deriving a GARCH equation for the conditional variance of ϵ_t,

$$\mathrm{E}[\epsilon_t^2|\Psi_{t-1}] = \lambda'\sigma_t^{(2)} + c,$$

we show that the process is weakly stationary if the eigenvalues of matrix C_{11}, defined by Equation (16), are less than one in absolute value or, equivalently, if the roots of the characteristic equation

$$\det[I - \alpha(z)\lambda' - \beta(z)] = 0$$

are outside the unit circle. By use of the nonnegativity conditions for α_i and β_i, we show that this is equivalent to Equation (11).

Consider the MN-GARCH process of Equation (4). Using the fact that, for any invertible matrix C, $C^{-1} = C^+/\det C$, where C^+ denotes the adjoint matrix of C, Equation (4) can be written as

$$\det[I - \beta(L)]\sigma_t^{(2)} = [I - \beta(1)]^+\alpha_0 + [I - \beta(L)]^+\alpha(L)\epsilon_t^2. \tag{A.1}$$

Without loss of generality, it can be assumed that the roots of $\det[I - \beta(z)] = 0$ lie outside the unit circle, since otherwise the nonstochastic part of Equation (3) would be explosive.

The conditional variance of ϵ_t is given by a linear combination of the conditional component variances, that is,

$$\mathrm{E}[\epsilon_t^2|\Psi_{t-1}] = \sigma_t^2 = \lambda'\sigma_t^{(2)} + c,$$

The variance of the process $\{\epsilon_t\}$ thus follows the univariate GARCH equation

$$\det[I - \beta(L)]\sigma_t^2 = \lambda'[I - \beta(L)]^+\alpha(L)\epsilon_t^2 + c^*, \tag{A.2}$$

where $c^* = \lambda'[I - \beta(1)]^+ \alpha_0 + \det[I - \beta(1)]c = c \det[I - \beta(1) + \alpha_0\lambda'/c]$ is constant. The argument is completed by following the same lines as in Gouriéroux (1997: 37). Defining $w_t = \epsilon_t^2 - \sigma_t^2$ and replacing, in Equation (A.2), σ_t^2 by $\epsilon_t^2 - w_t$, we obtain the ARMA$(\max\{pk, p(k-1)+q\}, pk)$ representation

$$[\det[I - \beta(L)] - \lambda'[I - \beta(L)]^+\alpha(L)]\epsilon_t^2 = \det[I - \beta(L)]w_t + c^* \tag{A.3}$$

for the ϵ_t^2 process. Hence the sequence $\mathrm{E}(\epsilon_t^2)$ converges and the process $\{\epsilon_t\}$ is weakly stationary if the roots of the characteristic equation[27]

$$\det[I - \beta(z)] - \lambda'[I - \beta(z)]^+\alpha(z) = \det[I - \beta(z) - \alpha(z)\lambda'] = 0 \tag{A.4}$$

are larger than unity or, equivalently, the spectral radius, $\rho(\cdot)$, of the transition matrix of Equation (16) satisfies $\rho(C_{11}) < 1$.[28] If $\rho(C_{11}) < 1$, then Equation (11) holds and, by the nonnegativity of C_{11}, guarantees the required positivity in Equation (10).[29]

Next, assume that $\det[I - \alpha(1)\lambda' - \beta(1)] > 0$ and note that, by the Frobenius theorem [Gantmacher (1959: 66)], the largest root in magnitude of C_{11} is real and nonnegative, so it suffices to show that the determinant condition implies that there is no real root of C_{11} equal to or larger than one. Define, analogous to C_{11}, the matrix

$$B = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_{r-1} & \beta_r \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix},$$

where $r = \max\{p, q\}$. As was mentioned above, it can be assumed without loss of generality that the eigenvalues of B are inside the unit circle, that is, $\det[I - \beta(z)] = 0 \Rightarrow |z| > 1$. From Equation (A.4), the characteristic equation of

[27] Recall that a GARCH process is serially uncorrelated, hence the process is weakly stationary if the variance exists.

[28] The first equality in Equation (A.4) is a consequence of the Sherman-Morrison formula for determinants, stating that, for matrix A and vectors u and v, $\det(A + uv') = \det A + v'A^+u$ [see, e.g., Henderson and Searle (1981)].

[29] It is well known [e.g., Bowden (1972)] that $[I - \alpha(z)\lambda' - \beta(z)]^{-1}$ is the upper left block of matrix $(I - C_{11}z)^{-1} \geq 0$ for $z^{-1} > \rho(C_{11})$.

matrix C_{11} is

$$\begin{aligned}\det(zI - C_{11}) &= \det\left(z^r I - \sum_{i=1}^{p}\beta_i z^{r-i} - \sum_{i=1}^{q}\alpha_i\lambda' z^{r-i}\right)\\ &= \det(zI - B)\left[1 - \lambda'\left(z^r I - \sum_{i=1}^{p}\beta_i z^{r-i}\right)^{-1}\sum_{i=1}^{q}\alpha_i z^{r-i}\right]\\ &= \det(zI - B)\left[1 - \lambda'\left(I - \sum_{i=1}^{p}\beta_i z^{-i}\right)^{-1}\sum_{i=1}^{q}\alpha_i z^{-i}\right].\end{aligned}$$

From nonnegativity, $\sum_{i=1}^{q}\alpha_i z^{-i}$ monotonically decreases in z. $(I - \sum_{i=1}^{p}\beta_i z^{-i})^{-1}$ forms the first k rows and columns of $(I - Bz^{-1})^{-1} = \sum_{i=0}^{\infty} B^i z^{-i} \geq 0$ for $z > \rho(B)$. It decreases monotonically in z. Hence it follows that, if $\det(I - C_{11}) = \det[I - \alpha(1)\lambda' - \beta(1)] > 0$, then $\rho(C_{11}) < 1$.

APPENDIX B: DYNAMIC PROPERTIES OF THE DIAGONAL MN(1,1) MODEL

We derive the condition for the existence of the fourth moment of the diagonal MN(1, 1)-GARCH model and show how to compute the autocorrelation function of the squares of such a process.

The derivation of the condition for $\mathrm{E}(\epsilon_t^4)$ has some similarity with Bollerslev (1986), in that we simultaneously consider $\sigma_t^{(2)}$ and the matrix of cross-products, $\sigma_t^{(2)}\sigma_t^{(2)'}$. Straightforward calculation gives

$$\begin{aligned}\sigma_t^{(2)}\sigma_t^{(2)'} &= (\alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta\sigma_{t-1}^{(2)})(\alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta\sigma_{t-1}^{(2)})'\\ &= \alpha_0\alpha_0' + \alpha_0\alpha_1'\epsilon_{t-1}^2 + \alpha_1\alpha_0'\epsilon_{t-1}^2 + \alpha_0\sigma_{t-1}^{(2)'}\beta + \beta\sigma_{t-1}^{(2)}\alpha_0'\\ &\quad + \alpha_1\alpha_1'\epsilon_{t-1}^4 + \alpha_1\sigma_{t-1}^{(2)'}\epsilon_{t-1}^2\beta + \beta\sigma_{t-1}^{(2)}\epsilon_{t-1}^2\alpha_1' + \beta\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}\beta.\end{aligned} \tag{B.5}$$

Let $W_t = [\sigma_t^{(2)'}, \mathrm{vec}(\sigma_t^{(2)}\sigma_t^{(2)'})']'$. As in the proof of Theorem 2 in Bollerslev (1986), we begin by evaluating $\mathrm{E}[W_t|\Psi_{t-2}]$ and note that

$$\mathrm{E}[\sigma_t^{(2)}|\Psi_{t-2}] = \alpha_0 + \alpha_1\lambda'\mu^{(2)} + (\alpha_1\lambda' + \beta)\sigma_{t-1}^{(2)}. \tag{B.6}$$

For the terms in Equation (B.5), we have, using $\mathrm{vec}(xy') = y \otimes x$ and $\mathrm{vec}(ABC) = (C' \otimes A)\,\mathrm{vec}(B)$,

$$\begin{aligned}\mathrm{vec}(\alpha_0\alpha_0') &= \alpha_0 \otimes \alpha_0,\\ \mathrm{E}[\mathrm{vec}(\alpha_0\alpha_1'\epsilon_{t-1}^2)\,|\,\Psi_{t-2}] &= (\alpha_1 \otimes \alpha_0)\lambda'(\sigma_{t-1}^{(2)} + \mu^{(2)}) = ((\alpha_1\lambda') \otimes \alpha_0)\sigma_{t-1}^{(2)}\\ &\quad + (\alpha_1 \otimes \alpha_0)\lambda'\mu^{(2)},\\ \mathrm{E}[\mathrm{vec}(\alpha_1\alpha_0'\epsilon_{t-1}^2)\,|\,\Psi_{t-2}] &= (\alpha_0 \otimes \alpha_1)\lambda'(\sigma_{t-1}^{(2)} + \mu^{(2)}) = (\alpha_0 \otimes (\alpha_1\lambda'))\sigma_{t-1}^{(2)}\\ &\quad + (\alpha_0 \otimes \alpha_1)\lambda'\mu^{(2)},\end{aligned}$$

$$\begin{aligned}
\mathrm{vec}(\alpha_0\sigma_{t-1}^{(2)'}\beta) &= (\beta\otimes\alpha_0)\sigma_{t-1}^{(2)},\\
\mathrm{vec}(\beta\sigma_{t-1}^{(2)}\alpha_0') &= (\alpha_0\otimes\beta)\sigma_{t-1}^{(2)},\\
\mathrm{E}[\mathrm{vec}(\alpha_1\alpha_1'\epsilon_{t-1}^4)\,|\,\Psi_{t-2}] &= \mathrm{vec}(\alpha_1\alpha_1')\Big[3\lambda'\sigma_{t-1}^{(4)}+\lambda'\mu^{(4)}+6\lambda'(\mu^{(2)}\odot\sigma_{t-1}^{(2)})\Big]\\
&= \{3(\alpha_1\otimes\alpha_1)\mathrm{vec}[\mathrm{diag}(\lambda)]'\}\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'})\\
&\quad +6(\alpha_1\otimes\alpha_1)(\lambda\odot\mu^{(2)})'\sigma_{t-1}^{(2)}+(\alpha_1\otimes\alpha_1)\lambda'\mu^{(4)},\\
\mathrm{E}[\mathrm{vec}(\alpha_1\sigma_{t-1}^{(2)'}\epsilon_{t-1}^2\beta)\,|\,\Psi_{t-2}] &= \mathrm{vec}(\alpha_1\sigma_{t-1}^{(2)'}\lambda'(\sigma_{t-1}^{(2)}+\mu^{(2)})\beta)\\
&= (\beta\otimes(\alpha_1\lambda'))\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'})+(\beta\otimes\alpha_1)\lambda'\mu^{(2)}\sigma_{t-1}^{(2)},\\
\mathrm{E}[\mathrm{vec}(\beta\sigma_{t-1}^{(2)}\epsilon_{t-1}^2\alpha_1')\,|\,\Psi_{t-2}] &= \mathrm{vec}(\beta\sigma_{t-1}^{(2)}\lambda'(\sigma_{t-1}^{(2)}+\mu^{(2)})\alpha_1')\\
&= ((\alpha_1\lambda')\otimes\beta)\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'})+(\alpha_1\otimes\beta)\lambda'\mu^{(2)}\sigma_{t-1}^{(2)},\\
\mathrm{vec}(\beta\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}\beta) &= (\beta\otimes\beta)\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}).
\end{aligned}$$

Define

$$d=\begin{bmatrix}d_1\\ d_2\end{bmatrix}=\begin{bmatrix}\alpha_0+\alpha_1\lambda'\mu^{(2)}\\ \alpha_0\otimes\alpha_0+(\alpha_0\otimes\alpha_1+\alpha_1\otimes\alpha_0)\lambda'\mu^{(2)}+(\alpha_1\otimes\alpha_1)\lambda'\mu^{(4)}\end{bmatrix},$$

$$C=\begin{bmatrix}C_{11} & 0_{k\times k^2}\\ C_{21} & C_{22}\end{bmatrix},$$

where

$$C_{11}=\alpha_1\lambda'+\beta,$$

$$\begin{aligned}C_{21} &= (\alpha_1\lambda')\otimes\alpha_0+\alpha_0\otimes(\alpha_1\lambda')+\alpha_0\otimes\beta+\beta\otimes\alpha_0+(\beta\otimes\alpha_1)\lambda'\mu^{(2)}\\
&\quad+(\alpha_1\otimes\beta)\lambda'\mu^{(2)}+6(\alpha_1\otimes\alpha_1)(\lambda\odot\mu^{(2)})',\end{aligned}$$

$$C_{22}=3(\alpha_1\otimes\alpha_1)\mathrm{vec}[\mathrm{diag}(\lambda)]'+\beta\otimes(\alpha_1\lambda')+(\alpha_1\lambda')\otimes\beta+\beta\otimes\beta.$$

Thus we have

$$\mathrm{E}[W_t|\Psi_{t-2}]=d+CW_{t-1}. \tag{B.7}$$

Recursive substitution in Equation (B.7) yields

$$\mathrm{E}[W_t|\Psi_{t-\tau-1}]=\sum_{j=0}^{\tau-1}C^j d+C^\tau W_{t-\tau}. \tag{B.8}$$

Hence, if the process is covariance stationary, that is, $\rho(C_{11})<1$, then $\mathrm{E}(W_t)$ exists and

$$\mathrm{E}(W_t)=\lim_{\tau\to\infty}\mathrm{E}[W_t|\Psi_{t-\tau-1}]=(I-C)^{-1}d \tag{B.9}$$

if and only if $\rho(C_{22})<1$. An expression for $\mathrm{E}\,\mathrm{vec}(\sigma_t^{(2)}\sigma_t^{(2)'})$, given it exists, can be obtained from Equation (B.9) via

$$\mathrm{E}(W_t)=(I-C)^{-1}d=\begin{bmatrix}(I-C_{11})^{-1} & 0_{k\times k^2}\\ (I-C_{22})^{-1}C_{21}(I-C_{11})^{-1} & (I-C_{22})^{-1}\end{bmatrix}\begin{bmatrix}d_1\\ d_2\end{bmatrix}.$$

Finally, the fourth moment can be computed from

$$\mathrm{E}(\epsilon_t^4) = 3\,\mathrm{vec}[\mathrm{diag}(\lambda)]'\mathrm{vec}[\mathrm{E}(\sigma_t^{(2)}\sigma_t^{(2)'})] + 6\lambda'[\mu^{(2)} \odot \mathrm{E}(\sigma_t^{(2)})] + \lambda'\mu^{(4)}. \tag{B.10}$$

Also, if $\mathrm{E}(\epsilon_t^4)$ exists, the unconditional third moment, and hence the skewness, exists and is given by

$$\mathrm{E}(\epsilon_t^3) = 3\lambda'[\mu \odot \mathrm{E}(\sigma_t^{(2)})] + \lambda'\mu^{(3)}. \tag{B.11}$$

Finally, we state the condition for the $(2m)^{\text{th}}$ moment to exist, where $m \in \mathbb{N}$. Note that $\mathrm{E}(\epsilon_t^{2m})$ exists if $\mathrm{E}[(\sigma_t^{(2)})^{\otimes m}]$ exists, where $A^{\otimes m}$ denotes the expression $A \otimes A \otimes \cdots \otimes A$ with m product terms. Let $\{\eta_t\}$ be an i.i.d. sequence of standard normal random variables, and let e_j be the jth unit vector in $\mathbb{R}^k$. Also, for simplicity, but without loss of generality, assume that the mixture is symmetric, that is, $\mu_1 = \mu_2 = \cdots = \mu_k = 0$. Then, by conditioning on the mixing process,

$$\mathrm{E}[(\sigma_t^{(2)})^{\otimes m}|\Psi_{t-2}] = \sum_{j=1}^{k}\lambda_j \mathrm{E}\left\{\left[\alpha_0 + (\alpha_1 e_j'\eta_{t-1}^2 + \beta)\sigma_{t-1}^{(2)}\right]^{\otimes m}\right\}. \tag{B.12}$$

Expanding the Kronecker product in Equation (B.12), a simultaneous recursion, similar to Equation (B.7), can be established for vector $[\mathrm{E}(\sigma_t^{(2)})', \ldots, \mathrm{E}(\sigma_t^{(2)\otimes(m-1)})', \mathrm{E}(\sigma_t^{(2)\otimes m})']'$, where the transition matrix is lower block-diagonal and the coefficient matrix of $\mathrm{E}[(\sigma_t^{(2)})^{\otimes m}]$ is the $k^m \times k^m$ matrix

$$\sum_{j=1}^{k}\lambda_j \mathrm{E}\left[(\alpha_1 e_j'\eta_t^2 + \beta)^{\otimes m}\right], \tag{B.13}$$

with expectation taken with respect to the i.i.d. sequence η_t. Hence, by the same arguments as above, $\mathrm{E}(\epsilon_t^{2m})$ exists if the largest eigenvalue of the expression in Equation (B.13) is smaller than one. This generalizes the condition $\mathrm{E}[(\alpha_1\eta_t^2 + \beta_1)^m] < 1$ for the single-component GARCH(1,1) model [Bollerslev (1986)]. Clearly Equation (B.13) allows the computation of moment conditions for more flexible error distributions by assuming, for example, that η_t is t-distributed, as in Neely (1999). Expressions for all existing moments could be developed. However, we do not attempt this, since the first four moments are of most interest in financial applications.

For the autocorrelation function, we make use of the fact that

$$\mathrm{E}(\epsilon_{t-\tau}^2\epsilon_t^2) = \mathrm{E}[\epsilon_{t-\tau}^2\mathrm{E}(\epsilon_t^2|\Psi_{t-\tau})] = \mathrm{E}[\epsilon_{t-\tau}^2\lambda'\mathrm{E}(\sigma_t^{(2)} + \mu^{(2)}|\Psi_{t-\tau})], \tag{B.14}$$

and Equation (17), which implies

$$\begin{aligned}\mathrm{E}[\sigma_t^{(2)}|\Psi_{t-\tau}] &= \mathrm{E}(\sigma_t^{(2)}) + (\alpha_1\lambda' + \beta)^{\tau-1}\left(\sigma_{t-\tau+1}^{(2)} - \mathrm{E}(\sigma_t^{(2)})\right)\\ &= \mathrm{E}(\sigma_t^{(2)}) + (\alpha_1\lambda' + \beta)^{\tau-1}\left(\alpha_0 + \alpha_1\epsilon_{t-\tau}^2 + \beta\sigma_{t-\tau}^{(2)} - \mathrm{E}(\sigma_t^{(2)})\right).\end{aligned}$$

Substituting in Equation (B.14) and employing the definition $\mathrm{cov}(\epsilon_{t-\tau}^2, \epsilon_t^2) = \mathrm{E}(\epsilon_{t-\tau}^2\epsilon_t^2) - \mathrm{E}^2(\epsilon_t^2)$, we get

$$\begin{aligned}&\mathrm{cov}(\epsilon_{t-\tau}^2, \epsilon_t^2)\\ &\quad= \lambda'(\alpha_1\lambda' + \beta)^{\tau-1}\Big\{\alpha_0\mathrm{E}(\epsilon_t^2) + \alpha_1\mathrm{E}(\epsilon_t^4) + \beta\Big[\mathrm{E}(\sigma_t^{(2)}\sigma_t^{(2)'}) + \mathrm{E}(\sigma_t^{(2)})\mu^{(2)'}\Big]\lambda\\ &\qquad -\mathrm{E}(\sigma_t^{(2)})\mathrm{E}(\epsilon_t^2)\Big\}. \end{aligned} \tag{B.15}$$

Finally, the autocorrelation at lag τ is

$$r(\tau) = \frac{\mathrm{cov}(\epsilon_{t-\tau}^2, \epsilon_t^2)}{\mathrm{var}(\epsilon_t^2)} = \frac{\mathrm{E}(\epsilon_t^2\epsilon_{t-\tau}^2) - \mathrm{E}^2(\epsilon_t^2)}{\mathrm{E}(\epsilon_t^4) - \mathrm{E}^2(\epsilon_t^2)}. \tag{B.16}$$

Received September 23, 2002; revised December 19, 2003; accepted January 24, 2004

REFERENCES

Akaike, H. (1973). "Information Theory and an Extension of the Maximum Likelihood Principle." In B. Petrov and F. Csaki (eds.), *2nd International Symposium on Information Theory*. Budapest: Akademiai Kiado.

Akgiray, V., and G. G. Booth. (1987). "Compound Distribution Models of Stock Returns: An Empirical Comparison." *Journal of Financial Research 10*, 269–280.

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

Alexander, C. (2001b). "Orthogonal GARCH." In C. Alexander (ed.), *Mastering Risk*, vol. 2. London: FT Prentice Hall.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen. (1996). "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics 74*, 3–30.

Ball, C. A., and W. N. Torous. (1983). "A Simplified Jump Process for Common Stock Returns." *Journal of Financial and Quantitative Analysis 18*, 53–65.

Bauwens, L., C. S. Bos, and H. K. van Dijk. (1999a). "Adaptive Polar Sampling with an Application to Bayes Measure of Value-at-Risk." Tinbergen Institute Discussion Paper TI 99-082/4, Erasmus University.

Bauwens, L., M. Lubrano, and J.-F. Richard. (1999b). *Bayesian Inference in Dynamic Econometric Models* . New York: Oxford University Press.

Beine, M., and S. Laurent. (2003). "Central Bank Interventions and Jumps in Double Long Memory Models of Daily Exchange Rates." *Journal of Empirical Finance 10*, 641–660.

Bekaert, G., and G. Wu. (2000). "Asymmetric Volatility and Risk in Equity Markets." *Review of Financial Studies 13*, 1–42.

Bera, A. K., and M. L. Higgins. (1993). "ARCH Models: Properties, Estimation and Testing." *Journal of Economic Surveys 7*, 305–366.

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

Black, F. (1976). "Studies of Stock Market Volatility Changes." In American Statistical Association (ed.), *Proceedings of the American Statistical Association, Business and Economic Statistics Section*. Alexandria, VA: American Statistical Association.

Boero, G., and E. Marrocu. (2002). "The Performance of Non-Linear Exchange Rate Models: A Forecasting Comparison." *Journal of Forecasting 21*, 513–542.

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

Bollerslev, T. (1988). "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process." *Journal of Time Series Analysis 9*, 121–131.

Bollerslev, T., R. F. Engle, and D. B. Nelson. (1994). "ARCH Models." In R. F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: Elsevier Science B.V.

Bowden, R. (1972). "The Generalized Characteristic Equation of a Linear Dynamic System." *Econometrica 40*, 201–203.

Brännäs, K., and N. Nordman. (2001). "Conditional Skewness Modelling for Stock Returns." Umeå Economic Studies 562, Umeå University.

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

Campbell, J., A. W. Lo, and A. C. MacKinlay. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Chen, M., and H. Z. An. (1998). "A Note on the Stationarity and the Existence of Moments of the GARCH Model." *Statistica Sinica 8*, 505–510.

Chib, S. (1995). "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S., and E. Greenberg. (1995). "Understanding the Metropolis-Hastings Algorithm." *American Statistician 49*, 327–335.

Chib, S., and E. Greenberg. (1996). "Markov Chain Monte Carlo Simulation Methods in Econometrics." *Econometric Theory 12*, 409–431.

Dempster, A. P., N. M. Laird, and D. B. Rubin. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society B 39*, 1–38.

Diebold, F. X., T. A. Gunther, and A. S. Tay. (1998). "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review 39*, 863–883.

Ding, Z., C. W. J. Granger, and R. F. Engle. (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.

Engle, R. F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica 50*, 987–1007.

Fama, E. (1965). "The Behavior of Stock Market Prices." *Journal of Business 38*, 34–105.

Fama, E. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance 25*, 383–417.

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

Franses, P. H., and D. van Dijk. (1996). "Forecasting Stock Market Volatility Using (Non-Linear) GARCH Models." *Journal of Forecasting 15*, 229–235.

Friend, W. E., and R. Westerfield. (1980). "Co-skewness and Capital Asset Pricing." *Journal of Finance 35*, 897–914.

Gantmacher, F. (1959). *Matrix Theory*, vol. 2. New York: Chelsea Publishing.

Gelfand, A. E., and A. F. M. Smith. (1990). "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association 85*, 389–409.

Geman, S., and D. Geman. (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*, 721–741.

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

Gouriéroux, C. (1997). *ARCH Models and Financial Applications* . New York: Springer.

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

Gridgeman, N. T. (1970). "A Comparison of Two Methods of Analysis of Mixtures of Normal Distributions." *Technometrics 12*, 823–833.

Haas, M., S. Mittnik, and M. S. Paolella. (2002). "Mixed Normal Conditional Heteroskedasticity." CFS Working Paper 2002/10, Center for Financial Studies.

Hamilton, J. D. (1988). "Rational-Expectations Econometric Analysis of Changes in Regime. An Investigation of the Term Structure of Interest Rates." *Journal of Economic Dynamics and Control 12*, 385–423.

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

Hansen, B. E. (1994). "Autoregressive Conditional Density Estimation." *International Economic Review 35*, 705–730.

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

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

Henderson, H. V., and S. R. Searle. (1981). "On Deriving the Inverse of a Sum of Matrices." *SIAM Review 23*, 53–60.

Jarque, C. M., and A. K. Bera. (1987). "A Test for Normality of Observations and Regression Residuals." *International Statistical Review 55*, 163–172.

Kane, A. (1977). "Skewness Preference and Portfolio Choice." *Journal of Financial and Quantitative Analysis 17*, 15–25.

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

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

Kim, S., N. Shephard, and S. Chib. (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies 65*, 361–393.

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

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

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

Lanne, M., and P. Saikkonen. (2003). "Modeling the U.S. Short-Term Interest Rate by Mixture Autoregressive Processes." *Journal of Financial Econometrics 1*, 96–125.

Lin, B. H., and S. K. Yeh. (2000). "On the Distribution and Conditional Heteroskedasticity in Taiwan Stock Prices." *Journal of Multinational Financial Management 10*, 367–395.

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

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer.

McLachlan, G. J., and K. E. Basford. (1988). *Mixture Models: Inference and Application to Clustering*. New York: Marcel Dekker.

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

Mittnik, S., and M. S. Paolella. (2000). "Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates." *Journal of Forecasting 19*, 313–333.

Mittnik, S., and S. Rachev. (1993). "Modeling Asset Returns with Alternative Stable Models." *Econometric Reviews 12*, 261–330.

Neely, C. J. (1999). "Target Zones and Conditional Volatility: The Role of Realignments." *Journal of Empirical Finance 6*, 177–192.

Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica 59*, 347–370.

Nelson, D. B. (1995). "Stationarity and Persistence in the GARCH(1,1) Model." In R. F. Engle (ed.), *ARCH. Selected Readings*. New York: Oxford University Press.

Nelson, D. B., and C. Q. Cao. (1992). "Inequality Constraints in the Univariate GARCH Model." *Journal of Business and Economic Statistics 10*, 229–235.

Palm, F. C., and P. J. G. Vlaar. (1997). "Simple Diagnostic Procedures for Modeling Financial Time Series." *Allgemeines Statistisches Archiv 81*, 85–101.

Paolella, M. (1999). *Tail Estimation and Conditional Modeling of Heteroskedastic Time-Series*. Ph.D. dissertation, University of Kiel.

Peiró, A. (1999). "Skewness in Financial Returns." *Journal of Banking and Finance 6*, 847–862.

Rockinger, M., and E. Jondeau. (2002). "Entropy Densities with an Application to Autoregressive Conditional Skewness and Kurtosis." *Journal of Econometrics 106*, 119–142.

Rosenblatt, M. (1952). "Remarks on a Multivariate Transformation." *Annals of Mathematical Statistics 23*, 470–472.

Rozelle, J., and B. Fielitz. (1980). "Skewness in Common Stock Returns." *Financial Review 15*, 1–23.

Samanidou, E., E. Zschischang, D. Stauffer, and T. Lux. (2004). "Microscopic Models of Financial Markets." In F. Schweitzer (ed.), *Microscopic Models of Economic Dynamics*. Berlin: Springer.

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

Simkowitz, M., and W. Beedles. (1980). "Asymmetric Stable Distributed Security Returns." *Journal of the American Statistical Association 75*, 306–312.

St. Pierre, E. F. (1993). *The Importance of Skewness and Kurtosis in the Time-Series of Security Returns*. Ph.D. dissertation, Florida State University.

Stuart, A., J. K. Ord, and S. Arnold. (1999). *Kendall's Advanced Theory of Statistics*, vol. 2A. London: Edward Arnold.

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

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

Titterington, D. M., A. F. Smith, and U. E. Makov. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: John Wiley & Sons.

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

Vigfusson, R. (1997). "Switching Between Chartists and Fundamentalists: A Markov Regime-Switching Approach." *International Journal of Finance and Economics 2*, 291–305.

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

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

Wong, C. S., and W. K. Li. (2001). "On a Mixture Autoregressive Conditional Heteroskedastic Model." *Journal of the American Statistical Association 96*, 982–995.

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