No. 2002/10

Mixed Normal Conditional Heteroskedasticity

Markus Haas, Stefan Mittnik and Marc S. Paolella

Center for Financial Studies
an der Johann Wolfgang Goethe-Universität ▪ Taunusanlage 6 ▪ D-60329 Frankfurt am Main
Tel: (+49)069/242941-0 ▪ Fax: (+49)069/242941-77 ▪ E-Mail: ifk@ifk-cfs.de ▪ Internet: http://www.ifk-cfs.de

CFS Working Paper No. 2002/10

Mixed Normal Conditional Heteroskedasticity

Markus Haas, Stefan Mittnik and Marc S. Paolella*

September 2002

Abstract:
Both unconditional mixed-normal distributions and GARCH models with fat-tailed conditional distributions have been employed for modeling financial return data. We consider a mixed-normal distribution coupled with a GARCH-type structure 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. An empirical application to NASDAQ-index data indicates the appropriateness of the model class and illustrates that the approach can generate a plausible disaggregation of the conditional variance process, in which the components' volatility dynamics have a clearly distinct behavior that is, for example, compatible with the well-known leverage effect.

JEL Classification: C22, C51, G10

Keywords: Finance, GARCH, Kurtosis, Skewness, Stationarity.

* Markus Haas and Marc S. Paolella, Institute of Statistics and Econometrics, University of Kiel, Olshausenstr. 40, 24118 Kiel, Germany
Stefan Mittnik, Institute of Statistics and Econometrics, University of Kiel, Olshausenstr. 40, 24118 Kiel, Germany and Center for Financial Studies, Taunusanlage 6, 60329 Frankfurt am Main, Germany

1 Introduction

Although Generalized Autoregressive Conditional Heteroskedastic Normal (GARCH) models 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.[1] A natural way of accommodating such stylized facts is to specify a GARCH–type structure driven by iid innovations from a fat–tailed and, possibly, asymmetric distribution. A sizeable and growing number of candidate densities exist, a number of which are considered in the application below. Moreover, building on work by Hansen (1994), the studies of Paolella (1999), Harvey and Siddique (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 paper, we investigate a model which 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, MN) and extend the usual GARCH structure by modeling the dynamics in volatility by a system of equations which 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 induced, i.e., it is inherent in the model without requiring explicit specification of a conditional skewness process. Moreover, the model can capture the leverage effect. These aspects will be demonstrated in the empirical example below.

The MN formulation also allows for a plausible interpretation of two or more heterogeneous groups of market participants. For example, "bullish" and "bearish" investors could behave differently. Shleifer and Summers (1990) distinguish between "arbitrageurs" or "ratio-

[1] See, for example, Kane (1977), Friend and Westerfield (1980), Rozelle and Fielitz (1980), Simkowitz and Beedles (1980), St. Pierre (1993), Mittnik and Rachev (1993), Franses and van Dijk (1996), Peiró (1999), and Harvey and Siddique (1999).

nal speculators" and "noise traders" who react differently to arriving news. Moreover, noise traders may fall into different subgroups, as in Lux (1997), who distinguishes between optimistic and pessimistic naive traders[2] and shows that the interaction between fundamentalists and chartists may be a source of time–varying second moments. The proposed model class is not only appealing in that it can allow for heterogeneous agents, it also provides, in the empirical application considered, a superior fit compared to competing models. Thus, apart from a viewpoint of financial theory, it will be of interest to practitioners such as risk managers.

The proposed model indeed has some similar characteristics to Markov switching models, which have undoubtedly grown in importance (and complexity) since the seminal work of Hamilton (1988, 1989) (see, for example, Hamilton and Susmel, 1994; Gray, 1996; and Dueker, 1997). However, compared to the aforementioned GARCH-type models with fat–tailed innovation distributions, these models have not been shown to be advantageous with respect to either estimation or, more importantly, out–of–sample forecasting (Pagan and Schwert, 1990; Hamilton and Susmel, 1994; Dacco and Satchell, 1999). Nor is the notion of a constant set of unique and recurring regimes any more plausible in a financial context than the decomposition considered herein. An approach similar to ours has recently been explored by Wong and Li (2001), who also argue against use of a latent Markov structure; see Section 5 below for some discussion of their work.

The remainder of this paper is as follows. Section 2 reviews relevant properties of unconditional MN distributions, presents the MN-GARCH model and discusses various special cases. Section 3 details stationarity conditions. An empirical application is presented in Section 4. Section 5 provides concluding remarks.

2 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 (Newcomb, 1980). After the seminal work of Pearson in 1894 on the moments estimator for the univariate normal mixture with two components, maximum likelihood (ML) estimation has become very popular with the advent of the EM algorithm of Dempster *et al.* (1977), while exact Bayesian analysis of mixtures has become feasible after the introduction of the Gibbs sampler of Geman and Geman (1984) into the statistical mainstream by Gelfand and Smith (1990) and Gelfand *et al.*

[2] In Lux (1997), noise traders also differ with respect to their trading–strategies: Some try to find out "the mood of the market", others follow various "technical" trading rules.

(1990). Further historical aspects, modern inferential methods, and discussion of applications associated with mixtures of normals are given in Titterington *et al.* (1985), McLachlan and Basford (1988), and McLachlan and Peel (2000).

2.1 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

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

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

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

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

$$\mu = \mathrm{E}(Y) = \sum_{j=1}^{k} \lambda_j \mu_j, \qquad m_2 = \mathrm{Var}(Y) = \sum_{j=1}^{k} \lambda_j \sigma_j^2 + \sum_{j=1}^{k} \lambda_j \left(\mu_j - \mathrm{E}(Y)\right)^2. \tag{1}$$

Owing to its great flexibility (see, for example, the various density plots in Marron and Wand, 1992), the MN has also 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. To demonstrate the skewness property, let Y be a k–component mixed normal random variable with mean $\mu = \sum_{j=1}^{k} \lambda_j \mu_j$. Since for constant a, $\int (y-a)^n \phi(y; \mu, \sigma^2) dy = \int y^n \phi(y; \mu - a, \sigma^2) dy$, we have

$$\begin{aligned} m_3 = \mathrm{E}\left[(Y-\mu)^3\right] &= \sum_{j=1}^{k} \lambda_j \left(\mu_j - \mu\right)^3 + 3\sum_{j=1}^{k} \lambda_j \sigma_j^2 \left(\mu_j - \mu\right) \qquad (2) \\ &= \sum_{j=1}^{k} \lambda_j \left(\mu_j - \mu\right)^3 + 3\sum_{j=1}^{k} \sum_{i<j} \lambda_j \lambda_i (\sigma_j^2 - \sigma_i^2)(\mu_j - \mu_i), \end{aligned}$$

which shows that common component means, i.e., $\mu_1 = \cdots = \mu_k = \mu$, imply symmetry. For $k = 2$, the above expression becomes

$$\begin{aligned} m_3 &= \frac{\lambda_1}{\lambda_2}\left(1 - \frac{\lambda_1}{\lambda_2}\right)\left(\mu_1 - \mu\right)^3 + 3\lambda_1\left(\mu_1 - \mu\right)\left(\sigma_1^2 - \sigma_2^2\right) \\ &= \lambda_1 \lambda_2 \left[\left(\lambda_2 - \lambda_1\right)\left(\mu_1 - \mu_2\right)^3 + 3\left(\mu_1 - \mu_2\right)\left(\sigma_1^2 - \sigma_2^2\right)\right]. \end{aligned}$$

If $\mu_1 \neq \mu_2$, then it is necessary and sufficient that $\lambda_1 \neq \lambda_2$ and/or $\sigma_1^2 \neq \sigma_2^2$ for Y to be asymmetric.[3]

With regard to kurtosis, let Y be a k-component mixed normal random variable but with $\mu_1 = \ldots = \mu_k = \mu$, so that $\mathrm{E}(Y) = \sum_j \lambda_j \mu_j = \mu$. Then, from Jensen's inequality, $\sum_j \lambda_j \sigma_j^4 = \sum_j \lambda_j (\sigma_j^2)^2 > (\sum_j \lambda_j \sigma_j^2)^2$, so that[4]

$$\gamma_2 = \frac{m_4}{m_2^2} = \frac{\mathrm{E}\left[(Y-\mu)^4\right]}{\left[\mathrm{E}\left[(Y-\mu)^2\right]\right]^2} = 3\frac{\sum_j \lambda_j \sigma_j^4}{\left(\sum_j \lambda_j \sigma_j^2\right)^2} > 3. \tag{3}$$

An advantage of the MN model not shared by other distributional assumptions is that it lends itself to economic interpretation in several ways. A mixture of two or more normals could arise from different groups of actors, with one group acting, for example, more volatile than the other or, possibly, processing market information differently. Considering unconditional distributions, Kon (1984), for example, argues that returns on individual stocks may be drawn from a noninformation distribution, a firm-specific distribution and a market-wide information distribution, i.e., a three component mixture.

The MN model can also be appropriate for samples where the components follow a repeating sequence in generating observations. As an example, day-of-the-week effects, as mentioned by Fama (1965), are a possible source of mixture distributions. More specifically, political and economic news arrivals occur continuously, and, if they are assimilated continuously by investors, the variance of the distribution of price changes between two points in time would be proportional to the actual number of days elapsed (as in the Monday-effect). 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.

[3]Necessity is rather obvious. Sufficiency follows from the fact that symmetry implies $m_3 = 0$. If $\mu_1 = \mu \pm \lambda_2\sqrt{3\left(\sigma_1^2 - \sigma_2^2\right)/\left(\lambda_1 - \lambda_2\right)} =: \mu \pm \lambda_2\tau$, then $m_3 = 0$ but the density is not symmetric. Symmetry, i.e., $f(\mu + y) = f(\mu - y)$, would imply that, for all y,

$$\lambda_1\phi\left(y; \pm\lambda_2\tau, \sigma_1^2\right) + \lambda_2\phi\left(y; \mp\lambda_1\tau, \sigma_2^2\right) = \lambda_1\phi\left(y; \mp\lambda_2\tau, \sigma_1^2\right) + \lambda_2\phi\left(y; \pm\lambda_1\tau, \sigma_2^2\right),$$

which does not hold for any $\tau \neq 0$ because the class of finite normal mixtures is identifiable (Teicher, 1963). That the density can only be symmetric about its mean is clear; see, e.g., Dudewicz and Mishra (1988, pp. 216-217).

[4]If, however, the means are far enough apart (so that the density is not highly peaked around its center), the kurtosis can actually be less than three.

2.2 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, i.e.,

$$\epsilon_t|\Psi_{t-1} \sim \text{MN}\left(\lambda_1,\ldots,\lambda_k,\mu_1,\ldots,\mu_k,\sigma_{1t}^2,\ldots,\sigma_{kt}^2\right), \tag{4}$$

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

$$\sigma_t^{(2)} = \alpha_0 + \sum_{i=1}^{q}\alpha_i\epsilon_{t-1}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^{(2)}, \tag{5}$$

where $\sigma_t^{(2)} = \left[\sigma_{1t}^2,\sigma_{2t}^2,\ldots,\sigma_{kt}^2\right]^T$; $\alpha_i = [\alpha_{i1},\alpha_{i2},\ldots,\alpha_{ik}]^T$, $i=0,\ldots,q$; and $\beta_j, j=1,\ldots,p$, are $k\times k$ matrices with typical element $\beta_{j,mn}$. Restrictions $\alpha_0 > 0$, $\alpha_i \geq 0, i=0,\ldots,q$, and $\beta_j \geq 0, j=1,\ldots,p$, are assumed.[5] They correspond to the non-negativity conditions of Bollerslev (1986) for the Normal-GARCH model, although they may be unnecessarily strong (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

$$\left(I_k - \beta(L)\right)\sigma_t^{(2)} = \alpha_0 + \alpha(L)\,\epsilon_t^2, \tag{6}$$

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

$$r_t = a_0 + \sum_{i=1}^{u}a_i r_{t-i} + \epsilon_t + \sum_{j=1}^{v}b_j\epsilon_{t-j}, \tag{7}$$

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 (4) and (6).

2.3 Special Cases

2.3.1 Diagonal MN-GARCH

A particularly interesting special case for modeling asset returns arises by restricting matrix $\beta(L)$ in (6) to be diagonal (subsequently referred to as a *diagonal* MN-GARCH process). In

[5] In case of vectors and matrices, symbol $\geq$ indicates elementwise inequality.

addition to allowing for a clear interpretation of the dynamics of the component variances, we find—not only for the example reported below—that it tends to be preferred over the full model when employing various model–selection criteria.

2.3.2 Partial MN-GARCH

With the interpretation of different groups of actors in mind, it is conceivable that the market is driven by a mixture in which some components exhibit constant variance. Such components could be associated with informed traders, whereas the dynamic components could be due to noise traders, possibly overreacting to news. 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.

2.3.3 Symmetric MN$_s$–GARCH

We also entertain models for which all the component means are restricted to be zero, i.e., $\mu_1 = \mu_2 = \cdots \mu_k = 0$, 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.

2.4 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. The model they proposed is restricted such that, for all t, $\sigma_{2t}^2 = \sigma_{1t}^2 + \delta^2$ (cf. the parameterization in Ball and Torous, 1983)[6] and can be nested in (5). 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.

[6] 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 *et al.* (1999a) consider a MN-GARCH model with two components, in which the component variances are proportional to each other, i.e., for all t, $\sigma_{2t}^2 = \tau\sigma_{1t}^2$, specializing (5) to[7]

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

It may be argued that the proportionality property is less appealing, since it implies that both components exhibit essentially the same dynamic behavior and does not allow for two (or more) differently acting groups of market participants having, for example, different speeds of adjustment. This feature also applies to the Palm and Vlaar specification.

Another special MN-GARCH model has been proposed in Lin and Yeh (2000). Their model is also characterized by imposing the same dynamics on each component variance, i.e., only the constants α_{0j}, $j = 1, \ldots, k$, in the GARCH equations are component–specific, while the coefficients of lagged squared error terms and variances are the same in each equation. For $k = 2$, this amounts to restricting (5) to

$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{bmatrix} = \begin{bmatrix} \alpha_{01} \\ \alpha_{02} \end{bmatrix} + \begin{bmatrix} \alpha_{11} \\ \alpha_{11} \end{bmatrix} \epsilon_{t-1}^2 + \begin{bmatrix} \beta_{11} & 0 \\ 0 & \beta_{11} \end{bmatrix} \begin{bmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{bmatrix}.$$

Finally, it should be noted that MN-GARCH processes are related to the t–GARCH model (Bollerslev, 1987) in that the t distribution can be represented as an infinite gamma–mixture of normals.

3 Stationarity and Persistence

3.1 Weak Stationarity

3.1.1 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

$$\mathrm{E}\,\sigma_t^{(2)} = \left[I - \beta(1) - \alpha(1)\lambda^T\right]^{-1}\left[\alpha_0 + \alpha(1)c\right], \tag{8}$$

where (see Appendix A for derivation)

$$c = \sum_{j=1}^{k} \lambda_j \mu_j^2 = \frac{1}{\lambda_k}\left[\sum_{j=1}^{k-1} \lambda_j \mu_j^2 - \sum\sum_{j<r<k} \lambda_j \lambda_r (\mu_j - \mu_r)^2\right].$$

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

As relationship (8) suggests and Appendix B shows, the necessary and sufficient condition for the existence of the unconditional variance is

$$\det\left[I-\beta(1)-\alpha(1)\lambda^T\right]>0. \tag{9}$$

An interpretation of (9) is provided in Appendix B. Condition (9) assumes a simple form in the special diagonal MN-GARCH case, which is discussed next.

3.1.2 The Diagonal Case

For diagonal MN-GARCH processes, defining $\tilde{\beta}_j = 1-\sum_{i=1}^{p}\beta_{i,jj}$, we have, from (B.7), that

$$\begin{aligned}\det\left[I-\beta(1)-\alpha(1)\lambda^T\right] &= \det\left[I-\beta(1)\right]-\sum_{j=1}^{k}\lambda_j\det B(1)_j \\ &= \left[1-\sum_{j=1}^{k}\frac{\lambda_j}{\tilde{\beta}_j}\sum_{i=1}^{q}\alpha_{ij}\right]\prod_{j=1}^{k}\tilde{\beta}_j \\ &= \left[\sum_{j=1}^{k}\frac{\lambda_j}{\tilde{\beta}_j}\left(1-\sum_{i=1}^{q}\alpha_{ij}-\sum_{i=1}^{p}\beta_{i,jj}\right)\right]\prod_{j=1}^{k}\tilde{\beta}_j,\end{aligned} \tag{10}$$

where matrix $B(1)_j$ is defined in (B.2). This last expression 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 all $j\in\{1,\ldots,k\}$, but rather for their weighted sum with the j^{th} weight being given by $\lambda_j/\tilde{\beta}_j$ and the weights not summing to one.[8] The mixing weight of each component is inflated by the component's contribution to the deterministic part of $\sigma_t^{(2)}$ in (5). This condition is stronger than just $\sum_{j=1}^{k}\lambda_j\left(1-\sum_{i=1}^{q}\alpha_{ij}-\sum_{i=1}^{p}\beta_{i,jj}\right)>0$ due to the feedback between the components.

By writing requirement (10) as

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

we see 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 (B.5), the unconditional variance of a diagonal MN-GARCH process becomes

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

[8] Clearly, $\prod_j\tilde{\beta}_j>0$ must be assumed, since otherwise the deterministic part of difference equation (5) would be explosive.

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

According to (10), the process can have finite variance even though some components are not covariance stationary, as long as the corresponding weights are sufficiently small. This result is similar to the condition for strict stationarity given by Francq et al. (2001) for a regime–switching GARCH(1,1) model. They show that, in this model, the condition derived by Nelson (1991) for the single–regime GARCH model need not hold in each regime but for a weighted average of the GARCH–parameters in each regime, where the weights are the stationary probabilities of the Markov chain.

3.2 Measuring Volatility Persistence

As is demonstrated in Appendix B, the largest eigenvalue, $\rho_{\max}$, of matrix

$$\Phi = \begin{bmatrix} \beta_1 + \alpha_1 \lambda^T & \beta_2 + \alpha_2 \lambda^T & \cdots & \beta_{r-1} + \alpha_{r-1}\lambda^T & \beta_r + \alpha_r \lambda^T \\ I_k & 0_k & \cdots & 0_k & 0_k \\ 0_k & I_k & & 0_k & 0_k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_k & 0_k & \cdots & I_k & 0_k \end{bmatrix}, \qquad (11)$$

with $r = \max\{p, q\}$ and 0_k denotes 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 $\beta(1) + \alpha(1)\lambda^T$. Analogous to the expression for the single component case, i.e., a Normal–GARCH(1,1) model (Bollerslev *et al.*, 1994), the conditional expectation of future variances in this model is given by

$$\mathrm{E}\left[\sigma_{t+k}^{(2)} | \Psi_{t-1}\right] = \sigma^{(2)} + \left(\beta(1) + \alpha(1)\lambda^T\right)^k \left(\sigma_t^{(2)} - \sigma^{(2)}\right),$$

where, from (8),

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

and $\left[\beta(1) + \alpha(1)\lambda^T\right]^k$ tends to zero geometrically with rate $\rho_{\max}$.

4 Conditional Heteroskedasticity of NASDAQ Returns

We investigate the daily returns on the NASDAQ index from its inception in February 1971 to June 2001, a sample of $T = 7,681$ observations.[9] Continuously compounded percentage

[9] 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.

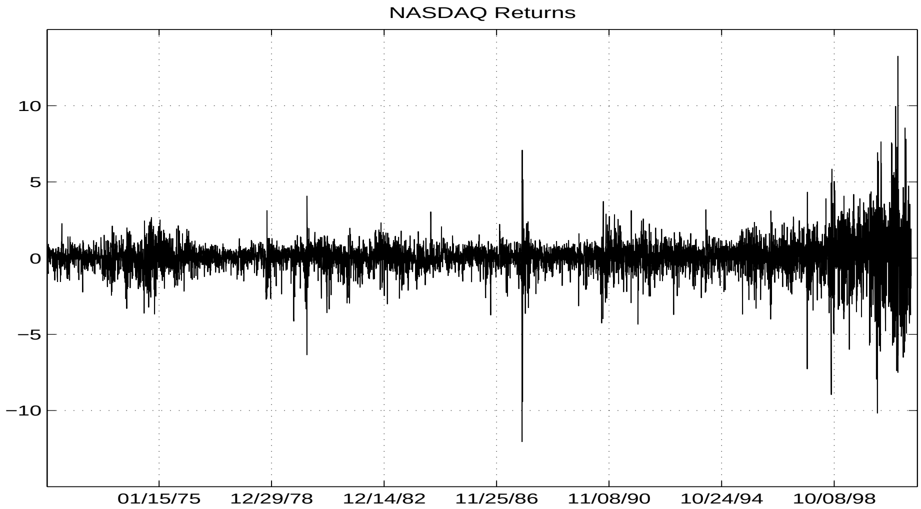

Figure 1: Percentage Returns on NASDAQ Index.

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) is virtually meaningless to report, given that 3rd and higher moments of financial data may not exist. In this case, estimating an unconditional Student's t distribution resulted in 2.4 degrees of freedom (and approximate standard error 0.08). One possible way to infer if asymmetry is statistically significant is to use a flexible parametric density which allows for asymmetry and fit it both restricted and unrestricted, from which a likelihood ratio test for asymmetry can be constructed. This was done using the noncentral t distribution, as suggested for use in a financial modeling context by Harvey and Siddique (1999); the asymmetric generalized t distribution in Mittnik and Paolella (2000); and the stable Paretian distribution (see, e.g., Mittnik, Paolella and Rachev, 2000). The likelihood ratio values were 69, 73 and 66, respectively, which are clearly tremendously significant at any conventional testing level.

Sample autocorrelation plots of Normal-GARCH(1,1) residuals (not shown) suggest a low order AR model for the mean equation. The Bayesian Information Criterion (BIC) favors an AR(3), which will accompany all GARCH structures estimated below.[10]

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

4.1 Estimation Issues

We estimate the ARMA–MN–GARCH model by conditional ML, conditioning, due to the ARMA structure (7), on the first u return observations and set the first v values of ϵ_t to zero and, for the GARCH structure, set the initial values of $\sigma_t^{(2)}$ and ϵ_t^2 equal to their unconditional expectations given in (8).[11] Because it is not clear what the "typical" parameter values would be for the GARCH structure with $k \geq 2$ components when applied to financial return series, we simply set the starting values to $\lambda_i = 1/k$, $\alpha_{01} = \ldots = \alpha_{0k} = 0.05$, $\alpha_{11} = \ldots = \alpha_{1k} = 0.1$, $\beta_{1,11} = \ldots = \beta_{1,kk} = 0.8$ and the off–diagonal elements of β matrices to zero. For several real data sets including the one used below (as well as many simulated series), these proved adequate, with convergence occurring usually within 20 to 50 iterations. Use of other, even very unrealistic, starting values led in virtually all cases to the same estimates.

Bayesian inference via Markov Chain Monte Carlo methods such as the Metropolis Hastings algorithm (see Chib and Greenberg, 1996, and Bauwens *et al.*, 1999b, and the references therein) is theoretically possible, but for the large sample sizes typically available in 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, both in terms of programming effort as well as in run time and assessment of convergence. For the diagonal model discussed below, an EM algorithm could also be constructed, but would offer little, if any gain, given the slow convergence of the method, and because each M–step would itself require numerical optimization.

4.2 Determining the Number of Mixture Components

For mixture models in general, the number of required component densities is unknown and needs to be empirically determined. Unfortunately, standard test theory breaks down in this context; see, for example, Wolfe (1971), Everitt and Hand (1981), Aitkin *et al.* (1981), Hartigan (1985), Ghosh and Sen (1985) McLachlan and Peel (2000, Ch. 6), and Chen *et al.* (2001). These authors perform and refer to simulation studies suggesting that the asymptotic distribution of the usual likelihood ratio test statistic mimics a χ^2 distribution with degrees of freedom larger than the number of fixed parameters under the null. This draws into question the results of Kon (1984), in which standard theory is used to provide evidence that some stocks are best modeled by a mixture of four components. Similar criticism applies to Kim and Kon (1994), although the values of the likelihood ratio statistics reported there (ranging

[11] The quasi-Newton maximization method available in Matlab (version 5.2, function `fminu`) was used, with (automatically computed) numeric gradient and Hessian, and a convergence criterion of 0.0001.

from 423 to 1854) are high enough to keep their conclusions valid under more appropriate methods of model selection.

Standard model selection criteria such as the AIC (Akaike, 1973) and the BIC (Schwarz, 1978) are widely used in the GARCH literature and can be used to compare models with differing numbers of components. For a model with K parameters and log–likelihood, L, evaluated at the maximum likelihood estimator, AIC $= -2L + 2K$ and 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, in particular, Dasgupta and Raftery, 1998; Fraley and Raftery, 1998; Leroux, 1992; Roeder and Wassermann, 1997; and McLachlan and Peel, 2000, Ch. 6). According to Kass and Raftery (1995), a BIC difference of less than two corresponds to "not worth more than a bare mention", while differences between two and six imply positive evidence, differences between six and ten give rise to strong evidence, and differences greater than ten invoke very strong evidence. The results of Mittnik and Paolella (2000) suggest that, with respect to out–of–sample prediction, these measures are indeed useful for choosing among GARCH–type models with competing distributional assumptions.

4.3 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. 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 cdf, that is,

$$\hat{u}_t = \hat{F}\left(\hat{\epsilon}_t | \Psi_{t-1}\right), \quad t = 1, \ldots, T. \tag{12}$$

Under a correct specification, the transformed residuals, $\hat{u}_t$, are iid uniform (Rosenblatt, 1952; see also Diebold et al., 1998). Thus, an inspection of the quantile fit can be based on the T $\hat{u}_t$–values. Below, we report for selected ξ–values the percentage of $\hat{u}_t$–values, denoted by U_ξ, for which $\hat{u}_t \leq \xi$; i.e.,

$$U_\xi = 100 \times T^{-1} \sum_{t=1}^{T} \mathcal{I}_{[0,\xi]}\left(\hat{u}_t\right), \tag{13}$$

where $\mathcal{I}$ denotes the indicator function. For a correctly specified model, we expect $U_\xi \approx 100\times\xi$. Also, a histogram of the $\hat{u}_t$'s provides a valuable visual check of the goodness of fit. To formally test for uniformity of the transformed values (12), we use the Pearson goodness–of–fit test, as was suggested by Palm and Vlaar (1997). The test statistic is given by

$$X^2 = \sum_{i=1}^{g} \frac{(n_i - n_i^*)^2}{n_i^*}, \tag{14}$$

where g is the number of (equally spaced) subintervals over the $[0,1]$–interval; 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. Below, we will report the results for $g = 100$.

If (14) 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, the X^2–values tend to be smaller when evaluated at the estimated rather than the true parameter values. As a consequence, the asymptotic distribution of (14) 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[12] (see Stuart et al., 1999, Ch. 25). To reflect the uncertainty about the true asymptotic distribution of X^2, we will act as if it were $\chi^2(g-K-1)$–distributed, so that the test tends to favor models with less parameters resulting in similar fit.

A drawback of the above test is the degree of arbitrariness that is inherent in the choice of the number of classes, g.[13] In addition, one may wish to test whether the specified distribution captures some specific characteristics of the data such as (conditional) skewness and kurtosis.[14] This can be accomplished by the further transformation

$$z_t = \Phi^{-1}(\hat{u}_t), \tag{15}$$

where Φ is the standard normal cdf, such that the z_t's are iid N(0,1) distributed, if the underlying model is correct. Berkowitz (2001) shows that inaccuracies in the specified density will be preserved in the transformed data.[15] Thus, this transformation allows the use of normal probability plots or moment–based normality tests for checking features such as correct specification of skewness and kurtosis.

[12] If the parameters are determined by minimizing (14), the exact asymptotic distribution is $\chi^2(g-K-1)$.

[13] For example, the use of values between $g = 50$ and $g = 150$ gave rise to p–values below 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 2.3).

[14] 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.

[15] Use of values (15) was also advocated by Palm and Vlaar (1997).

4.4 Competing Models

In the following comparison, all models entertained share a common AR(3)–GARCH(1,1) specification, i.e., following the notation in Section 2.2, $u = 3$, $v = 0$ and $p = q = 1$. Within the MN-GARCH model class, for a given number of components, k, it turns out that the diagonal model was always preferred over the full model when using the BIC criterion. With respect to the AIC, only for $k = 2$ was the full model preferred. For this reason, we restrict our attention to the diagonal models in the following analysis. We briefly discuss the characteristics of the full model for $k = 2$ and $k = 3$ at the end of this section.

In addition to several MN-GARCH specifications, we also fit the AR(3)–GARCH(1,1) model assuming a variety of conditional innovation distributions. To save space, we do not reproduce the density specifications here and refer the reader to the corresponding citations provided. Along with the Student's t (Bollerslev, 1987), two asymmetric generalizations are used, namely the non–central t distribution (Harvey and Siddique, 1999) and the so–called t_3 distribution used in Mittnik and Paolella (2000). Further candidates include the hyperbolic (Eberlein and Keller, 1995; Küchler *et al.*, 1999; Paolella, 1999), the generalized logistic (or EGB2) distribution (Paolella, 1997, Wang *et al.*, 2001) and the asymmetric two–sided Weibull (Mittnik *et al.*, 1998), abbreviated ADW.

Table 1 reports the likelihood–based goodness–of–fit measures for the fitted models and the rankings of the models with respect to each of the criteria. Not surprisingly, the worst performer is the standard Normal–GARCH model. For each criterion, the best model is among the MN–GARCH class. Furthermore, each of the chosen models is of the form MN(k, k)–GARCH, i.e., without suppression of any of the components' dynamics to a constant. When ranking according to the log–likelihood and the AIC, the top 5 models all belong to the MN–GARCH class, whereas, according to the BIC, 4 of the 5, including the top three, belong to that class. All symmetric MN$_s$–GARCH models perform relatively poorly. This is not surprising, given the pronounced negative skewness of the unconditional distribution.

In view of these results, the models MN$(3, 3)$ and MN$(4, 4)$, as well as MN$(2, 2)$ and MN$(3, 2)$, are retained for further consideration. The estimated parameter values of interest along with their approximate standard errors[16] are shown in Table 2. (Due to the GARCH(1,1) specification we simply write β for matrix β_1 and denote the typical element of β by β_{ij}.) For comparison purposes, results for the standard Normal–GARCH model are also given.

In Table 2, the components are ordered with respect to decreasing component means μ_j,

[16]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.

Table 1: Likelihood–based goodness of fit[a]

Distributional Model	K	L Value	L Rank	AIC Value	AIC Rank	BIC Value	BIC Rank
Normal	7	-9142.8	16	18299.5	16	18348.1	16
MN(2,1)	10	-8962.7	15	17945.4	15	18014.8	15
MN(3,1)	13	-8931.4	12	17888.7	14	17979.0	14
MN(3,2)	15	-8857.5	4	17745.1	4	17849.2	2
MN(2,2)	12	-8872.5	5	17768.9	5	17852.4	3
MN(3,3)	17	-8845.5	3	17725.0	3	**17843.1**	**1**
MN(4,4)	22	-8831.7	2	**17707.5**	**1**	17860.3	5
MN(5,5)	27	**-8828.1**	**1**	17710.2	2	17897.7	9
$\mathrm{MN}_s(2,2)$	11	-8931.9	13	17885.7	13	17962.2	13
$\mathrm{MN}_s(3,3)$	15	-8908.5	10	17847.0	10	17951.2	12
Student's t	8	-8932.7	14	17881.3	12	17936.9	11
non–central t	9	-8908.2	9	17834.4	9	17896.9	8
t_3	10	-8884.3	6	17788.6	6	17858.1	4
hyperbolic	9	-8904.7	8	17827.4	8	17889.9	7
EGB2	9	-8895.0	7	17808.1	7	17870.6	6
ADW	9	-8927.5	11	17873.1	11	17935.6	10

[a]The leftmost column refers to the conditional distribution used with an AR(3)–GARCH(1,1) model specification fitted for the NASDAQ returns. The column labeled K refers to the number of parameters for the respective models; 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.

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} (with the exception of components 1 and 2 in model MN(4,4)). 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, the increasing α_{1j} reflects an increasing responsiveness to (negative) shocks, while there is more inertia in σ_{jt}^2 when shocks tend to be positive, as is reflected by the increasing 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 in the sense that $\alpha_{1j} + \beta_{jj} > 1$ for $\mu_j < 0$. However, all estimated models themselves are stationary, as can be seen from the respective volatility persistence measures, $\rho_{\max}$, reported in the last row of Table 2. 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 ($\alpha_{12} + \beta_{22} = 1.032$), is more similar to the Normal–GARCH model and has an intermediate position. The third component, however, tends to heavily "overreact" to shocks, as reflected by the large value of α_{13}; it is also characterized by a remarkably 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 $\mathrm{E}\sigma_i^2$, $i = 1, \ldots, 4$), the lower is $\widehat{\mu}_i$, i.e., 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 for high volatility to coincide with negative returns (see, for example, Bekaert and Wu, 2000).[17]

The different responsiveness of the components to shocks is illustrated in Figure 2, which shows the square roots of the variance in the Normal–GARCH and of the component variances in model MN(3,3). The graphs clearly reveal the calm and rather hectic behavior of components 1 and 3, respectively, while σ_{2t} mimics the evolution of σ_t for the Normal–GARCH model. The relatively large constant $\alpha_{03} = 0.332$ in component 3 is reflected in the floor of σ_{3t} at roughly 1.

Returning to Table 2, the first two components of the MN(3,2) 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 which 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.

Table 3 provides quantile values (13), skewness, kurtosis and the Jarque–Bera Lagrange multiplier test for normality for the "normalized" residuals, $\hat{v}_t = \Phi^{-1}(\hat{u}_t)$, $t = 1, \ldots, 7678$, of the four candidate MN and the symmetric MN_s models. The corresponding histograms, with one-at-a-time (i.e., not simultaneous) 95% confidence intervals, and normal probability plots are displayed in Figure 4. The graph for MN(4, 4) mimics that for MN(3, 3) and is not shown.

The quantiles U_ξ of the asymmetric MN models match the target values ξ rather well—both in the left and right tails—as can be seen from Table 3 and Figure 4. Note, however, that the left–tail fit of the MN(2, 2) model is not as good as that for the models with $k > 2$; this coincides with the preferences of the AIC and BIC criteria for higher parameterized MN models and is especially evident from the skewness and kurtosis statistics reported in Table 3.

[17] A number of GARCH models exist that incorporate an asymmetric relation between risk and return, e.g., the EGARCH of Nelson (1991) and the model of Glosten et al. (1993); see also Bollerslev et al. (1994).

Table 2: MN-GARCH Parameter Estimates for NASDAQ Returns[a]

	Normal	MN(2, 2)	MN(3, 3)	MN(3, 2)	MN(4, 4)
α_{01}	0.014 (0.0018)	0.002 (0.0008)	0.000 (0.0007)	0.001 (0.0008)	0.003 (0.0032)
α_{11}	0.117 (0.0083)	0.051 (0.0066)	0.022 (0.0080)	0.038 (0.0068)	0.067 (0.0192)
β_{11}	0.869 (0.0089)	0.920 (0.0090)	0.956 (0.0137)	0.934 (0.0101)	0.855 (0.0461)
$\alpha_{11} + \beta_{11}$	0.986 (0.0032)	0.971 (0.0037)	0.978 (0.0063)	0.972 (0.0044)	0.922 (0.0345)
λ_1	1	0.820 (0.0255)	0.541 (0.0879)	0.724 (0.0427)	0.373 (0.1182)
μ_1	0	0.091 (0.0100)	0.164 (0.0233)	0.119 (0.0133)	0.200 (0.0367)
$\mathrm{E}\sigma_1^2$	0.986	0.525	0.370	0.460	0.329
α_{02}	–	0.075 (0.0235)	0.012 (0.0055)	0.027 (0.0122)	0.000 (0.0011)
α_{12}	–	0.512 (0.0941)	0.197 (0.0425)	0.379 (0.0685)	0.015 (0.0054)
β_{22}	–	0.727 (0.0457)	0.835 (0.0260)	0.768 (0.0357)	0.980 (0.0071)
$\alpha_{12} + \beta_{22}$	–	1.239 (0.0588)	1.031 (0.0244)	1.146 (0.0413)	0.995 (0.0031)
λ_2	0	0.180 (0.0255)	0.433 (0.0832)	0.272 (0.0431)	0.317 (0.0743)
μ_2	–	−0.415 (0.0575)	−0.153 (0.0548)	−0.281 (0.0508)	0.035 (0.0700)
$\mathrm{E}\sigma_2^2$	–	1.741	0.926	1.355	0.506
α_{03}	–	–	0.332 (0.1913)	0.825 (0.6246)	0.005 (0.0086)
α_{13}	–	–	1.303 (0.5179)	–	0.246 (0.0724)
β_{33}	–	–	0.567 (0.1389)	–	0.824 (0.0393)
$\alpha_{13} + \beta_{33}$	–	–	1.870 (0.4209)	0	1.070 (0.0416)
λ_3	0	0	0.026 (0.0111)	0.004 (0.0028)	0.289 (0.0841)
μ_3	–	–	−0.865 (0.2223)	−2.281 (0.7251)	−0.232 (0.0824)
$\mathrm{E}\sigma_3^2$	–	–	2.936	0.825	0.974
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)
$\mathrm{E}\sigma_4^2$	–	–	–	–	2.941
$\rho_{\max}$	0.986	0.985	0.989	0.986	0.994

[a]Standard errors 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. $\mathrm{E}\sigma_i^2$, $i = 1, \ldots, 4$, denotes the unconditional variance of component i, as computed from (8), and $\rho_{\max}$ is the measure of volatility persistence, that is, the largest eigenvalue of matrix (11).

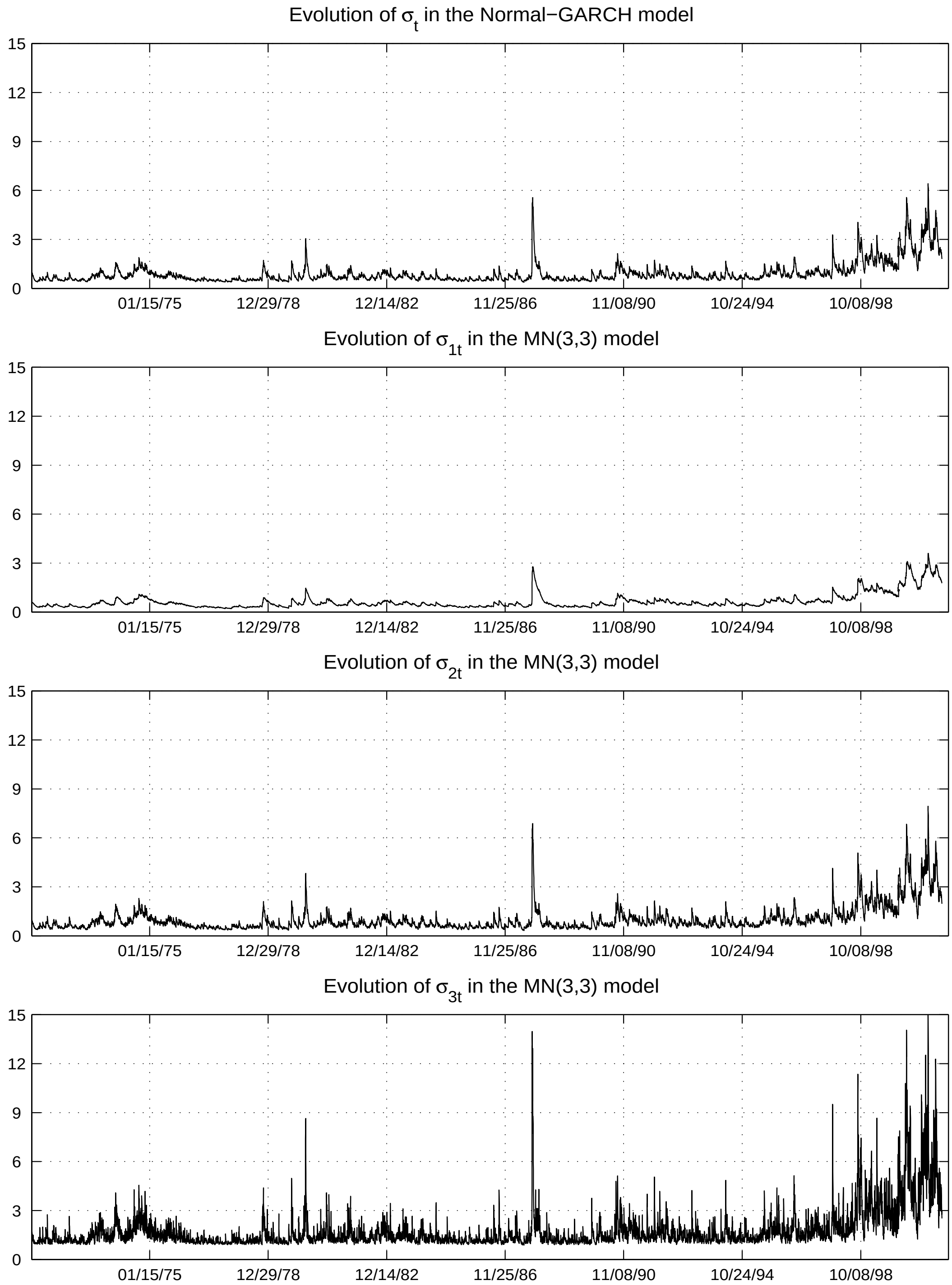

Figure 2: Volatility evolution for the Normal–GARCH and the MN(3,3)–GARCH models

While there is no significant skewness and excess kurtosis in the "normalized" residuals of the MN(3,3) and MN(4,4) models, the model MN(2,2) fails to adequately capture these properties. However, all three models as well as the MN(3,2) pass the Pearson goodness–of–fit test at the 10% level. Note that the symmetric mixture models $\mathrm{MN}_s(2,2)$ and $\mathrm{MN}_s(3,3)$ are able to accommodate the excess kurtosis from the residuals, but clearly fail to capture the skewness. Taken altogether, it appears that the asymmetric diagonal–MN$(3,3)$ and diagonal–MN$(4,4)$ models provide an adequate description of the NASDAQ series.

Using (1) and (2), the conditional skewness $m_3/m_2^{3/2}$ of the fitted MN(3,3) model is shown in the top plot of Figure 3. Because of the increase in the conditional variance of the process towards the end of the data set, the implied skewness moves towards zero. The middle plots in the figure show the conditional density when the skewness reached its most extreme value of -1.56. Its remarkable deviation from symmetry and the wide range of implied skewness values in the top plot emphasize the importance of time-varying skewness in this data set. The bottom plot shows the implied kurtosis, which appears to have a "natural lower bound" of 3, which is explainable from (3) and the fact that $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$ are relatively close in value.

4.5 Empirical Results for the Non–diagonal Models

For the two–component full model, we obtain a triangular structure for β. The estimated model is of the form

$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{bmatrix} = \begin{bmatrix} 0.002 \\ 0.077 \end{bmatrix} + \begin{bmatrix} 0.051 \\ 0.538 \end{bmatrix} \epsilon_{t-1}^2 + \begin{bmatrix} 0.918 & 0.000 \\ 0.447 & 0.572 \end{bmatrix} \begin{bmatrix} \sigma_{1t-1}^2 \\ \sigma_{2t-1}^2 \end{bmatrix}, \tag{16}$$

with $\lambda = (0.806, 0.194)^T$ and $\mu = (0.095, -0.395)^T$. The log likelihood, AIC and BIC of the model are –8869.4, 17764.8 and 17855.1, respectively. Thus, while AIC would prefer the full model, BIC prefers the diagonal specification.

Model (16) gives rise to a dynamic behavior similar to that of the diagonal model discussed in Section 4.4. For example, the volatility–persistence value is $\rho_{\max} = 0.987$, which is close to the value 0.985 for the diagonal MN$(2,2)$ in Table 2. Expression (B.4) shows how the components of the full model respond to innovations. The component–specific volatility persistence measures, given by $\alpha_{1j} + \beta_{jj}$ in the diagonal case, are now computed as the Frobenius roots of matrices $\beta_1(1) + \alpha_1(1)e_j^T$, $j = 1, 2$. (Here, e_j denotes the j^{th} $k \times 1$ unit vector.) These are 0.969 and 1.192, respectively, and are similar to those implied for the MN$(2,2)$ model, namely, $\alpha_{11} + \beta_{11} = 0.971$ and $\alpha_{12} + \beta_{22} = 1.239$ (from line 4 in Table 2). Also, both processes give rise to quite similar unconditional variances.

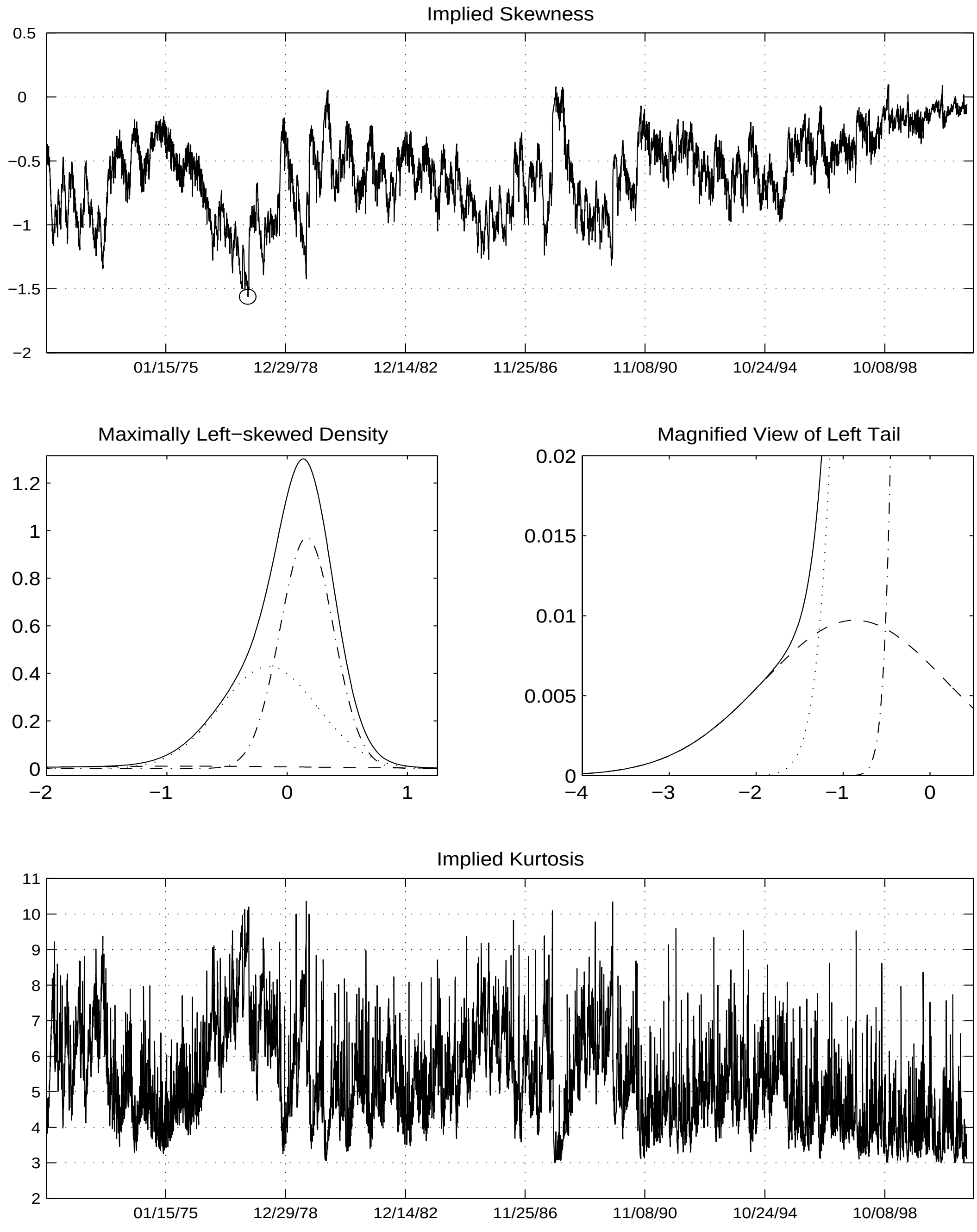

Figure 3: Top plot shows the implied skewness of fitted conditional densities for the NASDAQ data using the MN(3,3) model, with the inscribed circle indicating the maximal implied left skewness of -1.56, the density (solid line) of which is plotted in the middle panel together with the weighted component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

The three–component full model is

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

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

4.6 Extension to Fat–Tailed Components: The Mixed–t–GARCH

An extension of the MN-GARCH model which very naturally suggests itself is to replace the normal distribution with a fatter-tailed alternative. This would, in the NASDAQ case, help to accomodate the lack of fit of the MN(2,2)-GARCH(1,1) and potentially render unnecessary the MN(3,3) model, also resulting in a more parsimonious model. In this case, the component densities 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

$$\mathrm{E}\left[\epsilon^2_t | \Psi_{t-1}\right] = \sum_{j=1}^{k} \lambda_j \kappa_j \sigma^2_{jt} + c$$

where c is as in (8) and κ_j is a function of the shape parameter of the j^{th} component. For example, if the component densities are Student's t with ν_j degrees of freedom, $j = 1, \ldots, k$, then $\kappa_j = \nu_j/(\nu_j - 2)$. Straightforward calculations show that the stationarity condition (9) is easily extended, with the vector of mixing weights λ being replaced by $\lambda \odot \kappa$, where $\odot$ denotes element–by–element multiplication (the Hadamard product) of conformable matrices and $\kappa = [\kappa_1, \ldots, \kappa_k]^T$.

Using a mixture of Student's t distributions for the NASDAQ, first consider the same degrees of freedom parameter, ν, for each mixture component, jointly estimated with the remaining parameters. The resulting model then generalizes that proposed by Neely (1999) who used the Student's t with the Vlaar and Palm (1993) model. In the two-component case, $\hat{\nu} = 14.8$ (with standard error 3.4), indicating a relatively mild deviation from normality. With log likelihood value -8862.4 and 13 parameters, this resulted in a AIC value of 17750.9 and a BIC value of 17841.2, i.e., the AIC favors the MN(3,3)-GARCH(1,1) formulation, while the BIC is virtually indifferent. Thus, the introduction of the fatter-tailed density cannot replace the added dynamics which are allowed for in the MN(3,3) case.

Allowing the degrees of freedom to differ in this model, the low–volatility component has $\hat{\nu}_1 = 44.2$ (with "standard error" 47.9), and the high volatility unstationary component has

$\hat{\nu}_2 = 8.63$ (with standard error 1.81). The log likelihood of this model with 14 parameters is -8857.9, implying AIC $= 17743.8$ and BIC $= 17841.0$. Compared to the MN(3,3) model, the discussion of the former paragraph still applies.

Using the Student's t in the three component model (with equal degrees of freedom) resulted in $\hat{\nu} = 109.8$ (with the meaningless standard error of 310) and log likelihood value -8845.1, clearly demonstrating the adequacy of the normal.

It must be emphasized that these results are based on a single data set; the "mixed-t-GARCH" might indeed be useful in other contexts. However, its use ventures into the ad-hoc realm which we have endeavored to avoid.

Table 3: In–sample Fit of AR(3)–GARCH(1,1) MN Models[a]

	Normal	$\text{MN}_\text{s}(2,2)$	$\text{MN}_\text{s}(3,3)$	$\text{MN}(2,2)$	$\text{MN}(3,2)$	$\text{MN}(3,3)$	$\text{MN}(4,4)$
$U_{0.1}$	0.703	0.234	0.182	0.195	0.143	0.130	0.130
$U_{0.5}$	1.420	0.755	0.821	0.495	0.612	0.560	0.560
U_1	1.954	1.511	1.719	0.886	1.042	1.094	1.055
$U_{2.5}$	3.647	3.829	3.985	2.370	2.527	2.761	2.748
U_5	5.978	6.851	6.981	5.522	5.548	5.731	5.640
U_{10}	9.820	11.64	11.66	10.78	10.68	10.51	10.63
U_{90}	93.01	91.63	91.61	90.49	90.41	90.28	90.10
U_{95}	96.69	96.20	96.26	95.29	95.26	95.08	95.21
$U_{97.5}$	98.41	98.50	98.42	97.67	97.64	97.60	97.51
U_{99}	99.30	99.65	99.57	99.05	99.04	99.01	99.05
$U_{99.5}$	99.69	99.88	99.83	99.58	99.51	99.49	99.49
$U_{99.9}$	99.90	99.99	99.99	99.91	99.88	99.91	99.92
X^2	0.000	0.000	0.000	0.276	0.168	0.186	0.325
Skewness	-0.672^{***}	-0.298^{***}	-0.290^{***}	-0.088^{***}	-0.054^{*}	-0.046	-0.037
Kurtosis	2.521^{***}	0.064	0.027	0.134^{**}	0.094^{*}	0.011	-0.012
JB	2609.7^{***}	115.1^{***}	108.1^{***}	15.6^{***}	6.5^{**}	2.7	1.8

[a]The upper part of the table reports the empirical quantiles filtered by the fitted models, with U_ξ denoting the ξ%–quantile. The lower part reports test results on the distributional properties of the transformed residuals. X^2 refers to the p–value of the Pearson goodness–of–fit test (14) after transformation (12), 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 are based on transformation (15). "Skewness" denotes the coefficient of skewness $\gamma_1 = m_3/m_2^{3/2}$ and "Kurtosis" the coefficient of excess kurtosis $\gamma_2 - 3 = m_4/m_2^2 - 3$. 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–Bera (1987) Lagrange multiplier test for normality, i.e., JB $= T\gamma_1^2/6 + T(\gamma_2-3)^2/24$ (cf. Lütkepohl, 1991, pp. 152–156). Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

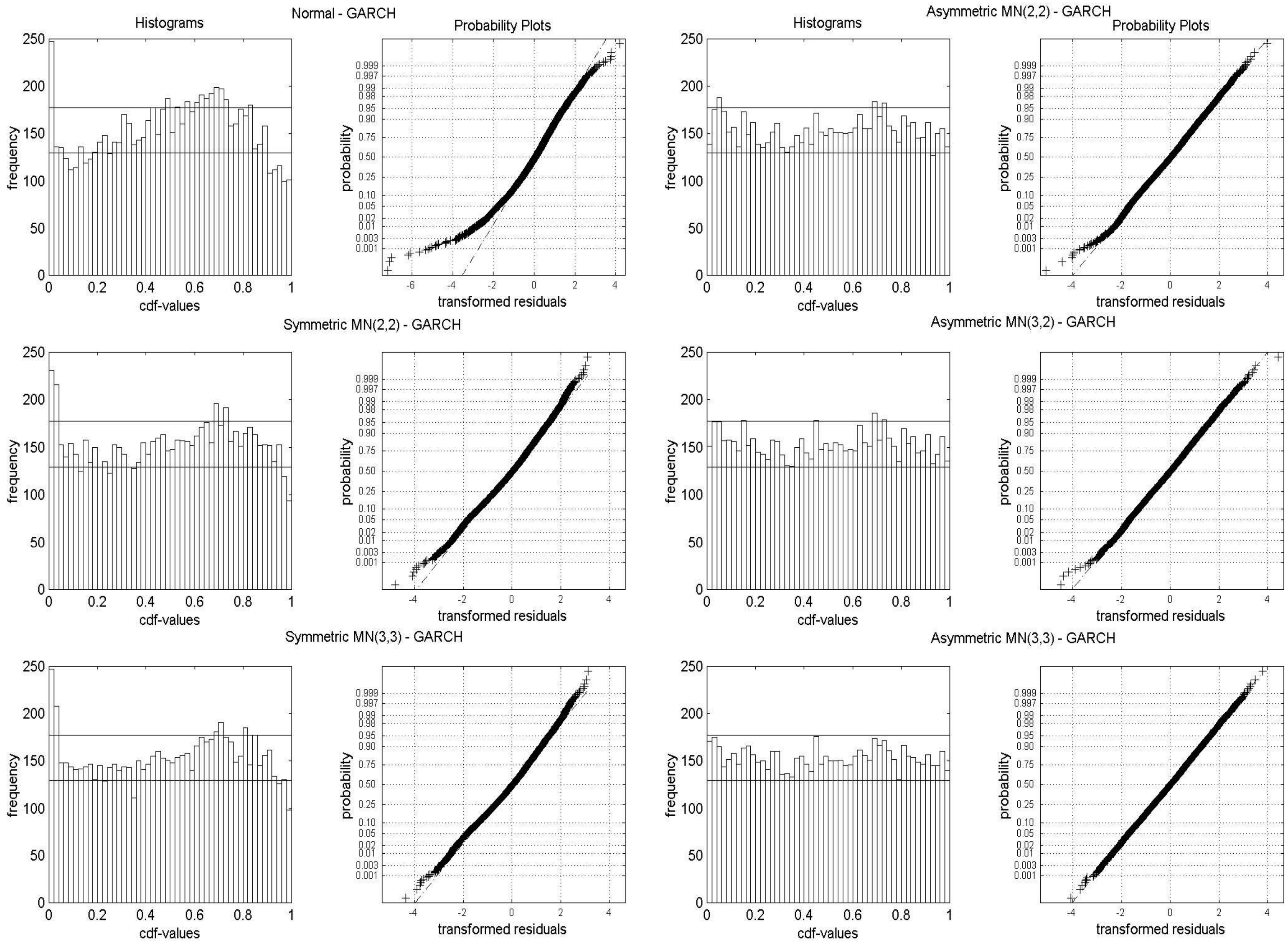

Figure 4: Diagnostics of the transformed residuals of AR(3)–MN–GARCH(1,1) models. Shown are histograms of the cdf–values $\hat{u}_t$ along with 95% confidence intervals and normal probability plots of cdf–values transformed by the inverse standard normal cdf.

5 Conclusions

We have investigated the properties and the usefulness of a class of conditionally heteroskedastic models for financial return series which re-employs the normality assumption via a mixed normal structure. The model gives rise to rich dynamics including time–varying skewness and kurtosis, which is otherwise not encountered in GARCH models driven by innovations from the "usual" asymmetric fat-tailed distributions. When applied to the returns on the NASDAQ index, the model class fairs extremely well compared to commonly used competing distributional specifications. Moreover, it offers a disaggregation of the conditional variance process which is amenable to economic interpretation, including the well–known leverage effect.

There are several possible generalizations of the proposed model which might be worth future investigation. First, allowing for time–varying mixture weights, as proposed in Vlaar and Palm (1993) and implemented, for example, in Beine and Laurent (1999) to model exchange rates, with the weights depending on central bank interventions. Second, more general, asymmetric GARCH structures, such as those proposed by Ding *et al.* (1993); and Sentana (1995), could be entertained. Third, the use of a weighted likelihood function, as employed in Mittnik and Paolella (2000) for achieving better out-of-sample forecasting performance, might also prove useful in this context. Finally, models with more general dynamics in the mean equation might be advantageous for modeling certain nonlinear time series. To this end, Wong and Li (2001) proposed a mixture autoregressive ARCH($k; p_1, \ldots, p_k; q_1, \ldots, q_k$) model, which allows for rather general mean dynamics. It is defined by

$$F\left(y_t|\Psi_{t-1}\right) = \sum_{j=1}^{k} \lambda_j \Phi\left(\frac{\epsilon_{j,t}}{\sigma_{j,t}}\right), \tag{17}$$

where $\epsilon_{j,t} = y_t - \gamma_{j,0} - \sum_{i=1}^{p_j} \gamma_{j,i} y_{t-i}$ and $\sigma_{j,t}^2 = \alpha_{j,0} + \sum_{i=1}^{q_j} \alpha_{j,i} \epsilon_{j,t-i}^2$. In modeling asset returns, however, the benefits of additional efforts in modeling the mean dynamics tend to be negligible, so that specification (5) may be preferable in the present context. Furthermore, in contrast to (17), the model structure adapted here allows a clear separation between the dynamics in the mean and in volatility, and, moreover, it leads to tractable stationarity conditions with insightful interpretations.[18]

[18] In (17), the dynamics in the means also account for conditional heteroskedasticity. The interaction between AR– and ARCH–dynamics leads to rather complicated stationarity conditions, especially for autoregressive orders exceeding one.

Appendix

A Computation of Constant c in (8)

For $Y \sim \mathrm{MN}\left(\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2\right)$, we have

$$\mathrm{Var}\left[Y\right] = \sum_{j=1}^{k} \lambda_j \sigma_j^2 + \sum_{j=1}^{k} \lambda_j \mu_j^2 - \left(\sum_{j=1}^{k} \lambda_j \mu_j\right)^2 = \sum_{j=1}^{k} \lambda_j \sigma_j^2 + c.$$

Hence, with $\mathrm{E}\left[Y\right] = \sum_{j=1}^{k} \lambda_j \mu_j = 0 \Leftrightarrow \mu_k = -\sum_{j=1}^{k-1} \frac{\lambda_j}{\lambda_k} \mu_j$,

$$\begin{aligned}
c &= \sum_{j=1}^{k} \lambda_j \mu_j^2 = \sum_{j=1}^{k-1} \lambda_j \mu_j^2 + \lambda_k \left(\sum_{j=1}^{k-1} \frac{\lambda_j}{\lambda_k} \mu_j\right)^2 \\
&= \frac{1}{\lambda_k}\left(\sum_{j=1}^{k-1} \lambda_j \mu_j^2 \left(1 - \sum_{j=1}^{k-1} \lambda_j\right) + \sum_{j=1}^{k-1}\sum_{r=1}^{k-1} \lambda_j \lambda_r \mu_j \mu_r\right) \\
&= \frac{1}{\lambda_k}\left(\sum_{j=1}^{k-1} \lambda_j \mu_j^2 - \sum_{j=1}^{k-1} \lambda_j \mu_j^2 \sum_{r=1}^{k-1} \lambda_r + \sum_{j=1}^{k-1}\sum_{r=1}^{k-1} \lambda_j \lambda_r \mu_j \mu_r\right) \\
&= \frac{1}{\lambda_k}\left(\sum_{j=1}^{k-1} \lambda_j \mu_j^2 - \sum_{j=1}^{k-1} \lambda_j \mu_j^2 \sum_{r \neq j} \lambda_r + 2 \sum\sum_{j<r<k} \lambda_j \lambda_r \mu_j \mu_r\right) \\
&= \frac{1}{\lambda_k}\left(\sum_{j=1}^{k-1} \lambda_j \mu_j^2 - \sum\sum_{j<r<k} \lambda_j \lambda_r \left(\mu_j - \mu_r\right)^2\right).
\end{aligned}$$

For $k = 2$, the expression for constant c reduces to $c = \frac{\lambda_1}{1-\lambda_1} \mu_1^2$.

B Derivation of Stationarity Condition (9)

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

$$\mathrm{E}\left(\epsilon_t^2 | \Psi_{t-1}\right) = \lambda^T \sigma_t^{(2)} + c,$$

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

$$\det\left[I - \alpha\left(z\right) \lambda^T - \beta\left(z\right)\right] = 0$$

are outside the unit circle. By use of the non-negativity conditions for the α_i and β_i, this is equivalent to condition (9).

Consider the MN–GARCH process (6). Using the fact that, for any invertible matrix C, $C^{-1} = C^{+} / \det C$, where C^{+} denotes the adjoint matrix of C, (6) can be written as

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

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 non-stochastic part of difference equation (5) would be explosive.

The construction of the adjoint matrix implies that

$$[I - \beta(L)]^{+} \alpha(L) = \begin{bmatrix} \det B(L)_1 \\ \det B(L)_2 \\ \vdots \\ \det B(L)_k \end{bmatrix}, \tag{B.2}$$

where $B(L)_j$ is matrix $I - \beta(L)$ with the j^{th} column being replaced by $\alpha(L)$. Thus, (B.1) gives rise to k univariate equations of the form

$$\det[I - \beta(L)]\, \sigma_{tj}^2 = \left(\det B(L)_j\right) \epsilon_t^2 + \det A(1)_j, \quad j = 1, \ldots, k, \tag{B.3}$$

where $A(1)_j$ is matrix $I - \beta(1)$ with the j^{th} column being replaced by α_0.

Note that (B.3) can not be interpreted as a GARCH equation for σ_{jt}^2, because σ_{jt}^2 is not the conditional variance of ϵ_t. If such an interpretation were correct, weak stationarity would require the roots of

$$\det[I - \beta(z)] - \det B(z)_j = \det\left[I - \beta(z) - \alpha(z)\, e_j^T\right] = 0, \quad j = 1, \ldots, k, \tag{B.4}$$

to be outside the unit circle, where e_j is the j^{th} unit vector in $\mathbb{R}^k$.[19] The conditional variance of ϵ_t is given by a linear combination of the conditional component variances, i.e.,

$$\mathrm{E}\left[\epsilon_t^2 | \Psi_{t-1}\right] = \sigma_t^2 = \sum_{j=1}^{k} \lambda_j \sigma_{jt}^2 + c.$$

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

$$\det[I - \beta(L)]\, \sigma_t^2 = \left(\sum_{j=1}^{k} \lambda_j \det B(L)_j\right) \epsilon_t^2 + c^*, \tag{B.5}$$

where $c^* = \sum_j \lambda_j A(1)_j + \det(I - \beta(1))\, c = c \det\left[I - \beta(1) + \alpha_0 \lambda^T / c\right]$ is constant. The argument is completed by following the same lines as in Gourieroux (1997, p. 37). Defining $w_t = \epsilon_t^2 - \sigma_t^2$ and replacing, in (B.5), σ_t^2 by $\epsilon_t^2 - w_t$, we obtain an ARMA($\max\{pk, p(k-1) + q\}, pk$) representation for the ϵ_t^2 process,

$$\left[\det\left(I - \beta(L)\right) - \sum_{j=1}^{k} \lambda_j \det B(L)_j\right] \epsilon_t^2 = \det[I - \beta(L)]\, w_t + c^*. \tag{B.6}$$

[19]The first equality in (B.4) follows directly from the linearity of $\det(\cdot)$ in columns.

Hence, the sequence $\mathrm{E}\left(\epsilon_t^2\right)$ converges and the process $\left\{\epsilon_t^2\right\}$ is weakly stationary if the roots of the characteristic equation[20]

$$\begin{aligned}
\det\left[I-\beta(z)\right]-\sum_{j=1}^{k}\lambda_j \det B(z)_j &= \det\left[I-\beta(z)-\alpha(z)\lambda^T\right] \qquad (B.7)\\
&= \sum_{j=1}^{k}\lambda_j \det\left[I-\beta(z)-\alpha(z)e_j^T\right]\\
&= 0
\end{aligned}$$

are larger than unity or, equivalently, the spectral radius, $\rho(\cdot)$, of the transition matrix (11) satisfies $\rho(\Phi)<1$.[21] If $\rho(\Phi)<1$, then (9) holds, and, by the non-negativity of Φ, guarantees the required positivity in (8).[22]

Next, assume that $\det\left[I-\alpha(1)\lambda^T-\beta(1)\right]>0$ and note that, by the Frobenius Theorem (Gantmacher, 1959, p. 66), the largest root in magnitude of Φ is real and non–negative, so it suffices to show that the determinant condition implies that there is no real root of Φ equal to or larger than one. Define, analogous to Φ, 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, i.e., $\det\left[I-\beta(z)\right]=0\Rightarrow |z|>1$. From (B.6), the characteristic equation of matrix Φ is

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

From non-negativity, $\sum_{i=1}^{q}\alpha_i z^{-i}$ monotonically decreases in z. $\left(I-\sum_{i=1}^{p}\beta_i z^{-i}\right)^{-1}$ forms the first k rows and columns of $\left(I-Bz^{-1}\right)^{-1}=\sum_{i=0}^{\infty}B^i z^{-i}\geq 0$ for $z>\rho(B)$. It decreases

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

[21]The first equality can be obtained by repeated use of the linearity of det($\cdot$) in columns. It is, however, a direct consequence of the Sherman–Morrison formula for determinants, stating that, for matrix A and vectors u and v, $\det\left(A+uv^T\right)=\det A+v^TA^+u$ (see, e.g., Henderson and Searle, 1981).

[22]It is well–known (see, e.g., Bowden, 1972), that $\left[I-\alpha(z)\lambda^T-\beta(z)\right]^{-1}$ is the upper left block of matrix $(I-\Phi z)^{-1}\geq 0$ for $z^{-1}>\rho(\Phi)$.

monotonically in z. Hence, it follows that, if $\det(I - \Phi) = \det\left[I - \alpha(1)\lambda^T - \beta(1)\right] > 0$, then $\rho(\Phi) < 1$.

To appreciate the stationarity condition of MN–GARCH processes, note that, combining (B.6) with (B.3) shows that stability does not require the condition $\det\left(I - \beta(1) - \alpha(1) e_j^T\right) > 0$ to hold for each of the k relationships in (B.3). It suffices that positivity holds for the weighted average with weights λ_j, $j = 1, \ldots, k$.

References

Aitkin, M., Anderson, D., and Hinde, J. (1981). Statistical Modelling of Data on Teaching Styles (with discussion). *Journal of the Royal Statistical Society A*, 144:419–461.

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

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

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

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

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

Beine, M. and Laurent, S. (1999). Central Bank Interventions and Jumps in Double Long Memory Models of Daily Exchange Models. Paper presented at the annual Helenau Workshop, Humboldt University, Berlin.

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

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 *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

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

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model For Speculative Prices and Rates of Returns. *The Review of Economics and Statistics*, 69:542–547.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). ARCH Models. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, Vol. 4, Chapter 49. Elsevier Science B.V., Amsterdam, The Netherlands.

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

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

Chen, H., Cheng, J., and Kalbfleisch, J. D. (2001). A Modified Likelihood Ratio Test for Homogeneity in Finite Mixture Models. *Journal of the Royal Statistical Society*, 63:19–29.

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

Dacco, R. and Satchell, S. (1999). Why do Regime–Switching Models Forecast so Badly? *Journal of Forecasting*, 18:1–16.

Dasgupta, A. and Raftery, A. E. (1998). Detecting Features in Spatial Point Processes With Clutter via Model–Based Clustering. *Journal of the American Statistical Association*, 93:294–302.

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

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

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

Dudewicz, E. J. and Mishra, S. N. (1988). *Modern Mathematical Statistics*. John Wiley & Sons, New York.

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

Eberlein, E. and Keller, U. (1995). Hyperbolic Distributions in Finance. *Bernoulli*, 1:281–299.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions*. Chapman and Hall, London.

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

Feng, Z. D. and McCulloch, C. E. (1996). Using Bootstrap Likelihood Ratios in Finite Mixture Models. *Journal of the Royal Statistical Society*, 58:609–617.

Fraley, C. and Raftery, A. E. (1998). How Many Clusters? Which Clustering Method? Answers via Model–Based Cluster Analysis. *The Computer Journal*, 41:578–588.

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

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

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

Gantmacher, F. (1959). *Matrix Theory, Vol. 2*. Chelsea Publishing Company, New York.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, 85:973–985.

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

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

Ghosh, J. H. and Sen, P. K. (1985). On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results. In Cam, L. M. L. and Olshen, R. A., editors, *Proceedings of the Berkeley Conference in Honor of Jery Neyman and Jack Kiefer*, Vol. 2, 789–806. Wadsworth, Monterey.

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

Gouriéroux, C. (1997). *ARCH Models and Fiancial Applications*. Springer–Verlag, New York.

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

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 Susmel, R. (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.

Hartigan, J. A. (1985). A Failure of Likelihood Asymptotics for Normal Mixtures. In Cam, L. M. L. and Olshen, R. A., editors, *Proceedings of the Berkeley Conference in Honor of Jery Neyman and Jack Kiefer*, Vol. 2, 807–810. Wadsworth, Monterey.

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

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

Jarque, C. M. and Bera, A. K. (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.

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

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

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

Kuechler, U., Neumann, K., Soerensen, M., and Streller, A. (1999). Stock Returns and Hyperbolic Distributions. *Mathematical and Computer Modelling*, 29:1–15.

Leroux, B. G. (1992). Consistent Estimation of a Mixing Distribution. *The Annals of Statistics*, 20:1350–1360.

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

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer–Verlag, Heildelberg.

Lux, T. (1997). Time Variation of Second Moments from a Noise Trader/Infection Model. *Journal of Economic Dynamics and Control*, 22:1–38.

Marron, J. S. and Wand, M. P. (1992). Exact Mean Integrated Squared Error. *Annals of Statistics*, 20:712–736.

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

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

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

Mittnik, S., Paolella, M. S., and Rachev, S. T. (1998). Unconditional and Conditional Distributional Models for the Nikkei Index. *Asia-Pacific Financial Markets*, 5:99–128.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (2000). Diagnosing and Treating the Fat Tails in Financial Returns Data. *Journal of Empirical Finance*, 7:389–416.

Mittnik, S. and Rachev, S. (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 Engle, R. F., editor, *ARCH. Selected Readings*, 176–192. Oxford University Press, New York.

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

Newcomb, S. (1980). A Generalized Theory of the Combination of Observations so as to Obtain the Best Result. In Stigler, M., editor, *American Contributions to Mathematical Statistics in the Nineteenth Century*, Vol. 2, 343–366. Arno Press, New York.

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

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

Paolella, M. (1999). *Tail Estimation and Conditional Modeling of Heteroscedastic Time-Series.* Ph.D. Thesis, Institute of Statistics and Econometrics, University of Kiel. Pro Business, Berlin.

Paolella, M. S. (1997). Generalized GARCH Models with Nonnormal Distributions. Working Paper 111, Institute of Statistics and Econometrics, University of Kiel.

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

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

Roeder, K. and Wasserman, L. (1997). Practical Bayesian Density Estimation Using Mixtures of Normals. *Journal of the American Statistical Association*, 92:894–902.

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

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

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

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

Shleifer, A. and Summers, L. H. (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspectives*, 4:19–33.

Simkowitz, M. and Beedles, W. (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*. PhD thesis, Florida State University College of Business.

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

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

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

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

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

Wang, K.-L., Fawson, C., Barrett, C. B., and McDonald, J. B. (2001). A Flexible Parametric GARCH Model with an Application to Exchange Rates. *Journal of Applied Econometrics*, 16:521–536.

Wolfe, J. H. (1971). A Monte Carlo Study of Sampling Distribution of the Likelihood Ratio for Mixtures of Multinormal Distributions. Technical Bulletin 72–2, U.S. Naval Personnel and Training Research Laboratory, San Diego.

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

CFS Working Paper Series:

No.	Author(s)	Title
2001/12	**Daniel Gross**	Trade Flows and the International Business Cycle
2002/01	**Stefan Feinendegen** **Daniel Schmidt** **Mark Wahrenburg**	Die Vertragsbeziehung zwischen Investoren und Venture Capital-Fonds: Eine empirische Untersuchung des europäischen Venture Capital-Marktes
2002/02	**Issam Hallak**	Price Discrimination on Syndicated Loans and the Number of Lenders: Empirical Evidence from the Sovereign Debt Syndication
2002/03	**Raimond Maurer** **Christian Schlag**	Money-Back Guarantees in Individual Pension Accounts: Evidence from the German Pension Reform
2002/04	**Holger Claessen** **Stefan Mittnik**	Forecasting Stock Market Volatility and the Informational Efficiency of the DAX-index Options Market
2002/05	**Bernd Kaltenhäuser**	Return and Volatility Spillovers to Industry Returns: Does EMU Play a Role?
2002/06	**Erik Theissen**	Internalisierung und Marktqualität: Was bringt Xetra Best?
2002/07	**Werner Neus** **Uwe Walz**	Exit Timing of Venture Capitalists in the Course of an Initial Public Offering
2002/08	**Andreas Bascha** **Uwe Walz**	Financing Practices in the German Venture Capital Industry An Empirical Assessment
2002/09	**Marco Hoeberichts**	The Credibility of Central Bank Announcements
2002/10	**Markus Haas,** **Stefan Mittnik** **Marc S. Paolella**	Mixed Normal Conditional Heteroskedasticity

Copies of working papers are available at the Center for Financial Studies or can be downloaded (http://www.ifk-cfs.de).