Studies in Nonlinear Dynamics & Econometrics

Volume 14, *Issue* 4 2010 *Article* 1

Skew-Normal Mixture and Markov-Switching GARCH Processes

Markus Haas*

*University of Munich, haas@stat.uni-muenchen.de

Skew-Normal Mixture and Markov-Switching GARCH Processes*

Markus Haas

Abstract

This paper introduces skew-normal (SN) mixture and Markov-switching (MS) GARCH processes for capturing the skewness in the distribution of stock returns. The model class is motivated by the fact that the common way of incorporating asymmetries into Gaussian MS GARCH models, i.e., regime-dependent means, leads to autocorrelated raw returns, which may not be desirable. The appearance of the SN distribution can be explained by a pre-asymptotic behavior of daily stock returns, and can still be viewed as "generic." The dynamic properties of the process are derived, and its in- and out-of-sample performance is compared with that of several competing models in an application to three major European stock markets over a period covering the recent financial turmoil. It turns out that parsimoniously parameterized SN mixture GARCH processes perform best overall. In particular, they outperform both a skewed *t* GARCH specification as well as normal mixture GARCH models with skewness generated via nonzero component means.

*I am grateful for constructive comments and suggestions from two anonymous referees, which led to significant improvements of the paper. I also thank Marc Paolella, Christian Pigorsch, seminar participants at the University of Bonn, and participants of the 2^{nd} Workshop on Computational and Financial Econometrics in Neuchatel (CFE'08), the annual meeting of the Verein für Socialpolitik 2009 in Magdeburg, and the annual meeting of the German Statistical Association 2009 in Wuppertal.

1 Introduction

The Markov–switching GARCH (MS GARCH) model provides an attractive framework for modeling the distribution of daily asset returns. To quote the recent textbook of Alexander (2008a, p. 163), it "tells us a lot about the volatility dynamics of equity markets and allows one to characterize its behaviour in [...] different market regimes", e.g., bull and bear markets, where the level of volatility is generally higher in the bearish regime, and the volatility dynamics may also be different. To set the stage for the subsequent discussion, let the time series $\{\varepsilon_t\}$ be generated by

$$\varepsilon_t = \sigma_{\Delta_t,t}\eta_t, \tag{1.1}$$

where $\{\eta_t\}$ is an iid sequence of standard normal variables, abbreviated $\eta_t \overset{iid}{\sim} \mathrm{N}(0,1)$; $\{\Delta_t\}$ is a Markov chain with finite state space $S = \{1,\dots,k\}$ and primitive (i.e., irreducible and aperiodic) transition matrix P,

$$P = \begin{pmatrix} p_{11} & \cdots & p_{k1} \\ \vdots & \dots & \vdots \\ p_{1k} & \cdots & p_{kk} \end{pmatrix}, \tag{1.2}$$

where $p_{ij} = p(\Delta_t = j|\Delta_{t-1} = i)$, $i, j = 1,\dots,k$; and the conditional standard deviation in regime j, σ_{jt}, is driven by a GARCH–type equation of the form

$$\sigma_{jt} = \omega_j + \alpha_j(|\varepsilon_{t-1}| - \lambda\varepsilon_{t-1}) + \beta_j\sigma_{j,t-1}, \tag{1.3}$$

where $\omega_j > 0$, $\alpha_j, \beta_j \geq 0$, $j = 1,\dots,k$, $-1 \leq \lambda \leq 1$. In addition, $\{\eta_t\}$ and $\{\Delta_t\}$ are assumed to be independent. The stationary distribution of the Markov chain will be denoted by $\pi_\infty = (\pi_{1,\infty},\dots,\pi_{k,\infty})'$.

Model (1.1)–(1.3) is an asymmetric and absolute value version of the MS GARCH model introduced by Haas et al. (2004b) and further studied, for example, in Liu (2006, 2007), Abramson and Cohen (2007), Alexander and Lazar (2008), and Ardia (2008, 2009). Alternative formulations of MS (G)ARCH processes exist (Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996; Dueker, 1997; Francq and Zakoïan, 2001; Klaassen, 2002; Bauwens et al., 2010; for an overview, see also the review articles of Hamilton, 2008; Lange and Rahbek, 2009; and Teräsvirta, 2009), but it may be argued that the one given by (1.1)–(1.3) has several advantages; for discussion, see Haas et al. (2004b), Ardia (2008, p. 110), and Lee (2009). Parameter λ in (1.3) allows volatility to react differently to positive and negative shocks of the same magnitude. Usually $\lambda > 0$, and this asymmetry in volatility, also referred to as *leverage effect*, is a robust feature of stock returns (e.g., Awartani and Corradi, 2005), and its incorporation via λ in (1.3) follows Ding et al. (1993).[1] Moreover, in the absolute value GARCH specification (1.3), as originally proposed by Taylor (1986), the regime–specific volatility dynamics are specified in terms of the standard deviations and absolute shocks $|\varepsilon_t|$ instead of the variances and squared shocks ε_t^2 as in the classic GARCH setup of Bollerslev (1986). Although the latter version is still the standard specification in the literature, it may be argued that models based on absolute returns are preferable as they are more robust to "outliers" (e.g., Ederington and Guan, 2005). A priori, both variants might appear somewhat arbitrary, since, as pointed out by Ding et al. (1993), "there is no obvious reason why one should assume the conditional variance is a linear function of lagged squared returns (residuals) as in Bollerslev's GARCH, or the conditional standard deviation a linear function of lagged absolute returns (residuals) as in Taylor/Schwert model." Consequently, Ding et al. (1993) proposed to introduce an additional power parameter d, thus generalizing (1.3) to the asymmetric power GARCH (APGARCH) process,

$$\sigma_{jt}^d = \omega_j + \alpha_j(|\varepsilon_{t-1}| - \lambda\varepsilon_{t-1})^d + \beta_j\sigma_{j,t-1}^d, \quad d > 0. \tag{1.4}$$

However, when the APGARCH model is applied to stock returns, it often turns out that the estimated d is not significantly different from unity, provided the volatility dynamics and the innovations' distribution are specified appropriately. To illustrate, Ding et al. (1993) estimated a (single–regime) Gaussian APGARCH model for the S&P 500 and found an optimal d significantly different from both one and two; moreover, the log–likelihood was considerably higher for $d = 2$ than for $d = 1$.[2] However, this appears to be due to the (inadequate) use of Gaussian innovations, since applications of APGARCH models to stock market data (including the S&P 500) employing more realistic specifications for the conditional density tend to find that d is close to and not significantly different from unity.[3] A possible explanation

[1]Clearly the degree of volatility asymmetry could likewise be made regime–dependent, but we will not entertain this possibility. Anticipating several results of Section 3, it appears that, due to the scarcity of observations from regimes occurring with low probability, models with regime–independent volatility dynamics may have several advantages over specifications without cross–component restrictions.

[2]Taylor (1986, p. 93), using Gaussian innovations, also observed for a number of stocks that the model based on squared returns "generally has a far higher maximum likelihood than the alternative specification based on absolute returns."

[3]For example, fitting an APGARCH process with Student's t innovations to 10 national stock indices and the MSCI world index, Brooks et al. (2000) find that the power parameter is significantly different from unity for only one series, whereas it is significantly different from two in nine cases. Similarly, Giot and Laurent (2003) apply a skewed Student's t APGARCH to three national stock markets and three individual stocks. For all but one of the stocks, the estimated d is rather close to (and statistically indistinguishable from) unity, leading the authors to conclude that "instead of modelling the conditional variance (GARCH), it is more relevant to model the conditional standard

for part of this observation might be that the unconditional kurtosis of a GARCH process depends on both the innovations' kurtosis as well as the variation of the conditional variance (cf. Bollerslev et al., 1994). Therefore, a model with a leptokurtic innovation density might require a less erratic behavior of the conditional volatility to generate a certain degree of unconditional kurtosis. We will report similar results for our data and models in Section 3; thus, use of the absolute value GARCH process appears reasonable, although it is clear that the abovementioned pattern is not an iron law and in all likelihood there are many series where it cannot be detected.[4]

Besides incorporating the economically intuitive concept of regime–specific volatility dynamics and providing accurate forecast densities (e.g., Ardia, 2009; Kaufmann and Scheicher, 2006; Marcucci, 2005; Sajjad et al., 2008; and Chen et al., 2009), regime–switching GARCH models possess a further attractive property. Namely, and in contrast to single–regime GARCH models, there is no need for specifying a fat–tailed distribution for η_t in (1.1), since in most applications filtered residuals from multi–regime models with Gaussian innovations do not display significant excess kurtosis (e.g., Alexander and Lazar, 2006; Haas et al., 2004a,b, 2009). This is due to the fact that there are already two sources of leptokurtosis in regime–switching GARCH models (i.e., GARCH and mixture effects), so that *conditional* normality *within* the regimes can be preserved, which, besides computational convenience, is often deemed appealing for theoretical reasons. To illustrate why, consider the problem of modeling log returns, r_t, defined by $r_t = 100 \times \log(P_t/P_{t-1})$, where P_t is the asset price at time t. The classical, "generic" assumption for the distribution of r_t has been the Gaussian, the rationale of which was clearly set out in Osborne (1959): Due to the central limit theorem and as the daily log return is the sum of a rather large number of intraday returns, he argued that "under fairly general conditions [...] we can expect that the distribution function of $[r_t]$ will be normal".

However, in addition to fat–tailedness, empirical return distributions are often characterized by significant asymmetries, although these are less ubiquitous and pronounced than excess kurtosis. See, e.g., Harvey and Siddique (1999), Jondeau and Rockinger (2003), Komunjer (2007), and Grigoletto and Lisi (2009) for evidence of skewness, whereas a more skeptical view is expressed in the contributions of Peiró (1999, 2002, 2004), who argues that "though there could exist some specific asymmetries that are relatively weak, asymmetry or skewness is not a stylized

---

deviation." More recently, Nieto and Ruiz (2008) and Lejeune (2009) obtain basically the same results for the S&P 500 (both) and the NASDAQ (Lejeune).

[4] An example is Ané (2006), who finds for the Japanese TOPIX index that the difference between the models with $d = 1$ and $d = 2$ is negligible.

fact of stock [...] returns" (Peiró, 2004). In the context of regime–switching models, asymmetries are usually captured by allowing for regime–specific means, i.e., by extending (1.1) to

$$\varepsilon_t = \mu_{\Delta_t} + \sigma_{\Delta_t,t}\eta_t. \tag{1.5}$$

However, as long as the Markov chain is persistent, this introduces autocorrelation of raw returns (cf. Poskitt and Chung, 1996; and Timmermann, 2000), which may not be desirable for two reasons. First, many return series exhibit significant skewness but no autocorrelation, and second, even if significant autocorrelations were present, such a specification would make it impossible to disentangle the asymmetries from the first–order dynamics of the returns. Thus, a more promising approach to introduce skewness into the MS GARCH model is to replace the normal distribution with a distribution which is rather close to the normal with respect to kurtosis and tail behavior but allows for some asymmetries. In the context of the Osborne–model referred to above, the appearance of such a distribution can be interpreted as a pre–asymptotic behavior of asset returns at a daily frequency, where, as skewness generally vanishes more slowly than kurtosis, the latter has been wiped out during the intraday summation process, whereas a moderate degree of the former is still present. In this paper, we follow this strategy by taking the regime densities as belonging to the class of skew–normal distributions introduced by Azzalini (1985). This distribution, while exhibiting Gaussian tail behavior, allows for skewness of a degree sufficient for most index return series. At this point, it may be worthwhile to note in passing that recently, in a more standard situation (i.e., approximating binomial probabilities), Chang et al. (2008) also successfully use the skew–normal to account for the pre–asymptotic behavior of sums of random variables.

The paper is organized as follows. Section 2 presents the model and a discussion of its properties. Section 3 reports the results of an application to three European stock indices, and Section 4 concludes and identifies issues for further research. Various technical details are gathered in a set of appendices.

2 The skew-normal Markov–switching GARCH process

In this section, we define the skew–normal MS GARCH process. Section 2.1 introduces the skew–normal distribution and details its properties, and Section 2.2 discusses its use in the context of MS GARCH models.

2.1 The skew–normal (SN) distribution

The density (pdf) of the skew–normal (SN) distribution is given by

$$f(z;\gamma) = 2\phi(z)\Phi(\gamma z), \quad \gamma \in \mathbb{R}, \tag{2.1}$$

where $\phi(z) = (2\pi)^{-1/2}\exp(-z^2/2)$ and $\Phi(z) = \int_{-\infty}^{z}\phi(\xi)d\xi$ are the standard normal pdf and distribution function (cdf), respectively. Thus, the pdf of the SN can be interpreted as a normal pdf times a weight factor given by $2\Phi(\gamma z)$. If $\gamma < 0$, this factor will be larger for negative values of z and vice versa, so that a negative (positive) γ gives rise to a negatively (positively) skewed density, and γ controls the degree of skewness of the SN. On the other hand, the tails of the SN are Gaussian, as is reflected in Azzalini's (1986) result that $|Z| \overset{d}{=} |X|$, where $X \sim \mathrm{N}(0,1)$.

Examples of SN densities are presented in Figure 1. For purpose of comparison with the standard normal, which is also shown, the SN densities have been standardized to have zero mean and unit variance. The case $\gamma = -1.5$ is representative for what we find for the European stock market indices in Section 3.2. It is characterized by only mild deviations from the Gaussian. As γ increases in magnitude, the asymmetry becomes stronger, and the density with $\gamma = -5$ is already indicative of the convergence to the (centered and scaled) half–normal distribution on $\mathbb{R}^-$ which occurs for $\gamma \to -\infty$. The corresponding figures for positively skewed SN variables are easily discerned from the fact that, if $Z \sim \mathrm{SN}(\gamma)$, then $-Z \sim \mathrm{SN}(-\gamma)$.

Moments can also be useful to describe the properties of return distributions. From the result cited at the end of the penultimate paragraph, it is clear that the even moments are identical to those of the Gaussian. Azzalini (1985) calculated the odd moments up to the third, and Henze (1986) and Martínez et al. (2008) provide general expressions. In particular, Henze (1986) shows that an SN random variable Z has a representation $Z = \delta|U| + \sqrt{1-\delta^2}V$, where $\delta = \gamma/\sqrt{1+\gamma^2}$, and U and V are independent standard normal variables; thus

$$\mathrm{E}(Z^{2\ell+1}) = \sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^\ell}\frac{\gamma}{(1+\gamma^2)^{\ell+1/2}}\sum_{m=0}^{\ell}\frac{m!(2\gamma)^{2m}}{(2m+1)!(\ell-m)!}, \quad \ell = 0,1,\ldots. \tag{2.2}$$

Application of (2.2) gives the well–known expressions

$$\mathrm{E}(Z) = \sqrt{\frac{2}{\pi}}\delta, \quad \text{and} \quad \mathrm{E}(Z^3) = \sqrt{\frac{2}{\pi}}(3\delta - \delta^3), \tag{2.3}$$

which shows that Z has a nonzero mean for $\gamma \neq 0$. As we interpret the error term as representing unexpected news in our context, we will use the centered version of (2.1) in Section 2.2 to define the skew–normal MS GARCH process.

The moment–based coefficient of skewness, m_3, is given by

$$m_3 = \frac{\sqrt{2}(4-\pi)\delta^3}{(\pi - 2\delta^2)^{3/2}} \in (-0.995, 0.995), \tag{2.4}$$

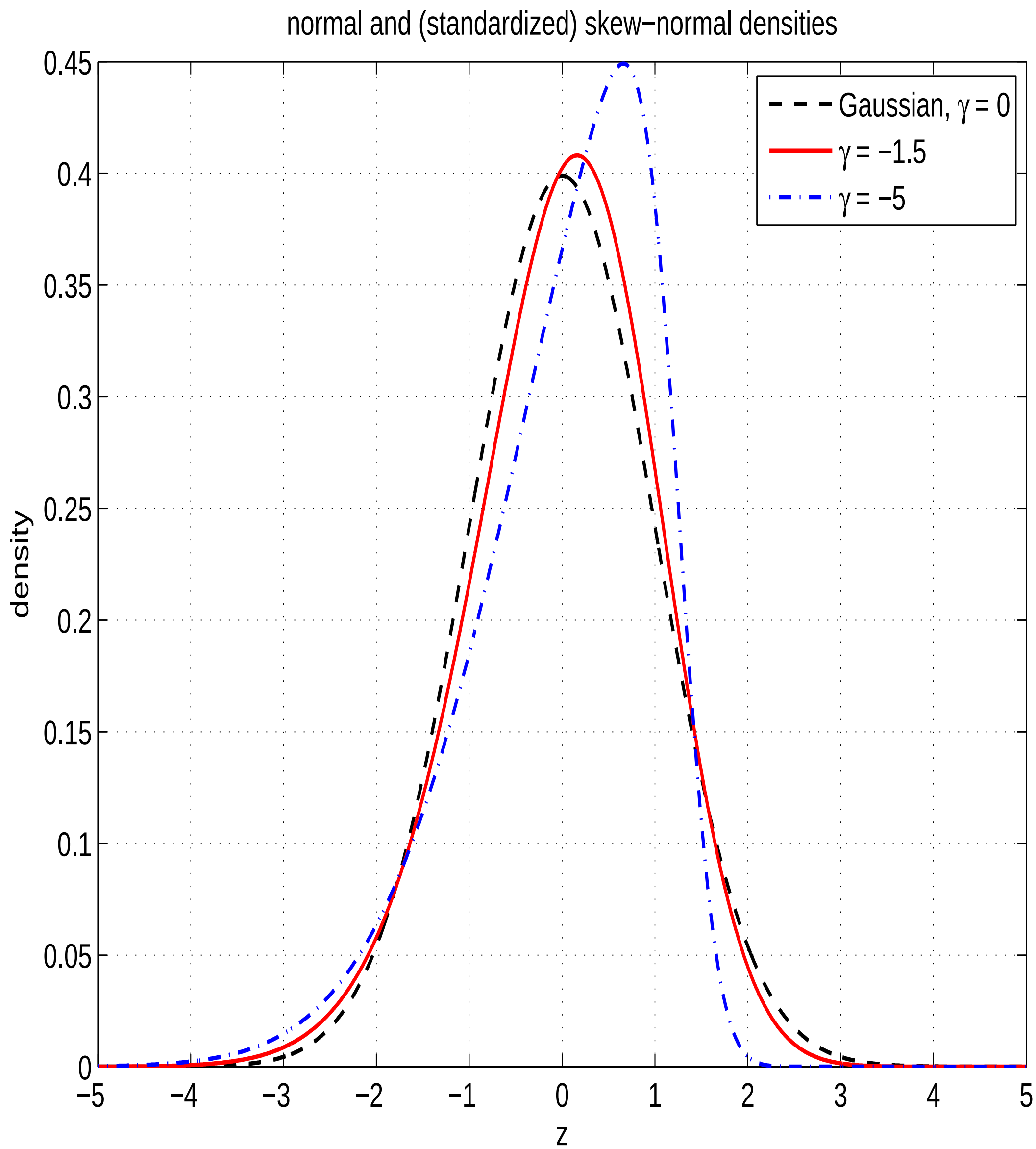

Figure 1: The Figure shows the standard normal density (dashed line) along with the skew–normal densities (2.1) with $\gamma = -1.5$ (solid line) and $\gamma = -5$ (dash–dotted line). For purpose of comparison with the standard normal, the skew–normal densities have been standardized to have zero mean and unit variance.

which was calculated by Azzalini (1985). Thus, although limited, the range of feasible skewness coefficients is sufficient for most index return distributions.[5] Azzalini (1985) also calculated the coefficient of kurtosis, m_4, which is given by

$$m_4 = 3 + \frac{8(\pi-3)\delta^4}{(\pi-2\delta^2)^2}, \tag{2.5}$$

and for $\gamma \to \pm\infty$ approaches its maximum value of 3.869. However, in the current context, the significance of this quantity appears rather limited. For example, in financial applications, the most emphasized aspect of excess kurtosis is fat–tailedness, whereas, as observed in Figure 1, for the SN it appears to mainly reflect a certain degree of peakedness when compared to the Gaussian. The asymmetry already implies that, for $\gamma < 0$, the SN has more mass in the left tail than the Gaussian, whereas there is actually less weight in the right tail. Thus, there is little additional information to extract from m_4 in (2.5).[6]

For applications in finance, such as Value–at–Risk calculations, availability of a convenient expression for the distribution function (cdf) is desirable. Azzalini (1985) has shown that the cdf of the SN is

$$F(z;\gamma) = \Phi(z) - 2T(z,\gamma), \tag{2.6}$$

where

$$T(z,\gamma) = \frac{1}{2\pi}\int_0^{\gamma} \frac{\exp\left\{-\frac{z^2}{2}(1+x^2)\right\}}{1+x^2}dx \tag{2.7}$$

is Owen's (1956) T–function, which can be integrated numerically. Alternatively, for $|\gamma| < 1$, we can use (Owen, 1956)

$$T(z,\gamma) = \frac{\arctan(\gamma)}{2\pi} - \frac{1}{2\pi}\sum_{i=0}^{\infty}(-1)^i \frac{1-e_i(z^2/2)e^{-z^2/2}}{2i+1}\gamma^{2i+1}, \tag{2.8}$$

where $e_n(x) = \sum_{k=0}^{n} x^k/k!$ is the exponential sum function. For $|\gamma| > 1$,[7] integration

[5] We also note that (2.4) represents the range of *conditional* skewness; the *unconditional* skewness of the SN Markov–switching GARCH process can be calculated using (B.7) and is larger in magnitude than the conditional skewness (since $\mathrm{E}(\sigma_t^3) > \mathrm{E}^{3/2}(\sigma_t^2)$), albeit (usually) by a small amount (e.g., compare the conditional skewness measures reported in Tables 5–7 with the corresponding unconditional counterparts in Table 8).

[6] Moreover, our estimates of γ in Section 3.2 will often be around -1 and never below -1.5, which corresponds to rather moderate kurtosis values of 3.06 ($\gamma = -1$) and 3.18 ($\gamma = -1.5$), the latter being approximately the kurtosis of a Student's t random variable with 37 degrees of freedom.

[7] The case $|\gamma| = 1$ can be handled by Azzalini's (1985) observations that $F(z;1) = \Phi^2(z)$ and

by parts shows that we can apply (2.6) and (2.8) in conjunction with

$$F(z;\gamma) = \begin{cases} 2\Phi(z)\Phi(\gamma z) - F(\gamma z;\gamma^{-1}) & \text{if} \quad \gamma \geq 0 \\ 2\Phi(z)\Phi(\gamma z) + 1 - F(\gamma z;\gamma^{-1}) & \text{if} \quad \gamma < 0. \end{cases} \tag{2.9}$$

According to our experience, (2.8) converges fast for any reasonable values of γ and z, and, as the coefficients of the series are alternating in sign and monotonically decreasing in absolute value, the approximation error can be controlled by means of Leibniz' criterion.

2.2 The skew–normal MS GARCH process

The skew–normal MS GARCH process with k regimes and skewness parameter γ is obtained by replacing the $\mathrm{N}(0,1)$ distribution for η_t in (1.1) by an SN(γ), i.e., the model is described by (1.1)–(1.3) along with

$$\eta_t \overset{iid}{\sim} \mathrm{SN}(\gamma), \tag{2.10}$$

where, as ε_t in (1.1) is interpreted as an unexpected random shock, we take the centered version of the density defined in (2.1). In principle, similar to the approach in Rombouts and Bouaddi (2009), the skewness parameter γ of the SN could be made regime–dependent. However, we refrain from doing so since we found that, when regime–specific shape parameters are allowed for, those of the regimes with low stationary probabilities are estimated with large standard errors (see Lin et al., 2007, for similar results in the context of unconditional mixtures of SNs), and in no case did the added flexibility lead to a significant improvement in–sample. More importantly, perhaps, when using rolling windows to reestimate the model and to calculate out–of–sample forecast densities, it turns out that the shape parameter of the low–probability regime tends to be highly unstable, and, if anything, the extra flexibility worsens the out–of–sample forecast quality of the model.

In the SN MS GARCH model, skewness is *conditional* in the sense that it is a property of the innovations η_t in (1.1) rather than a result of the dynamic structure of the model. Recently, He et al. (2008) argued that such an approach may be less attractive because it "requires giving up a standard assumption in econometric work, namely, that noise sent through a parametric filter to generate the output has a symmetric distribution around zero." However, these authors also show that, with symmetric innovations, substantive nonlinear mean dynamics are required to interact with the conditional heteroskedasticity in order to produce a realistic degree of skewness of the unconditional distribution. This seems to be an even less

$F(z;-\gamma) = 1 - F(-z;\gamma)$.

attractive alternative even if one were to agree with the concerns of He et al. (2008) concerning conditional skewness. However, we don't believe that these concerns are well–grounded, and that specification (1.1)–(1.3) with (2.10) is a rather natural generalization of Osborne's (1959) Gaussian model.[8]

To some extent, an exception to what has been said above concerning the relation between skewness and conditional mean dynamics is the normal mixture GARCH model discussed, among others, in Alexander and Lazar (2006, 2009), Badescu et al. (2008), Bauwens et al. (2007), Bauwens and Rombouts (2007), Bertholon et al. (2009), Giannikis et al. (2008), Haas et al. (2004a), Rombouts and Bouaddi (2009), Rombouts and Stentoft (2009), and Wong and Li (2001), where it is assumed that the transition matrix (1.2) has unit rank, so that there is no persistence in the Markov chain. In this case, the conditional return distribution is an iid mixture with weights π_∞, where use of the extension in (1.5), with $\mu_k = -\sum_{j=1}^{k-1}(\pi_{j,\infty}/\pi_{k,\infty})\mu_j$ to have $\mathrm{E}(\varepsilon_t) = 0$, will not induce return autocorrelation, but allows great flexibility with respect to skewness (see, e.g., Geweke and Amisano, 2009). This model is difficult to classify according to the He et al. (2008) criterion, since, although driven by symmetric innovations, the conditional mixture distribution at each point of time is asymmetric. It may also be interesting to consider skew–normal mixture GARCH processes with regime–dependent means. We then have two sources of skewness and might be able to assess their relative importance.

2.2.1 Properties of the model: Stationarity and moment structure

Conditions for stationarity and the finiteness of moments of the model defined by (1.1)–(1.3) and (2.10) follow from the results of Liu (2007) summarized in Theorem 1 in Appendix B. In this appendix, we also derive the unconditional variance, skewness, and kurtosis of the process, as well as the autocorrelation structure of its absolute and the squared values. For these derivations, the odd absolute moments of a centered SN random variable are required and calculated in Appendix A. As pointed out by He and Teräsvirta (1999), having available expressions for certain moments may be useful in assessing whether a given model is able "to satisfy stylized facts in the observed series. These observational facts include leptokurtosis and nonlinear dependence, of which the latter shows as positive autocorrelations of squared and absolute–valued observations." However, due to sampling error in particular of the higher moments, it is clear that simple comparisons between empirical and theoretical moments are a rather informal way of assessing a model's adequacy;

[8]An alternative approach to skewness based on the idea that negative and positive returns are driven by different volatility dynamics has recently been proposed by Pelagatti (2009), see also El Babsiri and Zakoïan (2001).

nevertheless, they may still be useful for spotting whether certain characteristics of interest implied by a model are in the right ballpark, or, as phrased by Rydén et al. (1998), to see if a model "does [...] capture the general tendency reasonably well."

2.2.2 Estimation issues

For numerically calculating maximum likelihood estimates, as well as for certain theoretical considerations, the "direct" parametrization of the SN density function given by (2.1) is not well suited due to an irregular shape of the corresponding likelihood function. This phenomenon is discussed in Azzalini (1985), Azzalini and Capitanio (1999), Chiogna (2005), and Arellano–Valle and Azzalini (2008), and is illustrated in the left plot of Figure 2, showing the profile log–likelihood for the skewness parameter γ in the context of a single–regime SN GARCH model, where $k = 1$ in (1.1)–(1.3), when applied to the DAX returns described at the beginning of Section 3 below. As a general feature, we observe that the profile log–likelihood exhibits an inflection point at $\gamma = 0$ and is far from a quadratic shape overall, which, besides being associated with theoretical complications, can seriously affect the numerical optimization process. As demonstrated by Azzalini and Capitanio (1999), Chiogna (2005), and Arellano–Valle and Azzalini (2008), the situation can be resolved by reparameterizing the density in terms of the coefficient of skewness (2.4). As shown in the right plot of Figure 2, this removes the inflection point at $\gamma = 0$ and also makes the profile log–likelihood much closer resembling a quadratic shape. In general, as noted by Azzalini and Capitanio (1999), "the more regular shape of the log–likelihood leads to faster convergence of the numerical maximization procedures when computing the MLE." Once this has been found, the parameter γ of the "direct" parametrization can be recovered via

$$\delta = \operatorname{sign}(m_3)\sqrt{\frac{m_3^{2/3}\pi}{2m_3^{2/3} + 2^{1/3}(4-\pi)^{2/3}}}, \quad \text{and} \quad \gamma = \frac{\delta}{\sqrt{1-\delta^2}},$$

where m_3 is the estimated skewness (2.4).

3 Empirical results

In this section, we investigate the volatility dynamics of daily returns of three European (total return) stock market indices, namely, the French CAC 40, the German DAX 30, and the British FTSE 100, from January 1990 to October 2009, a sample

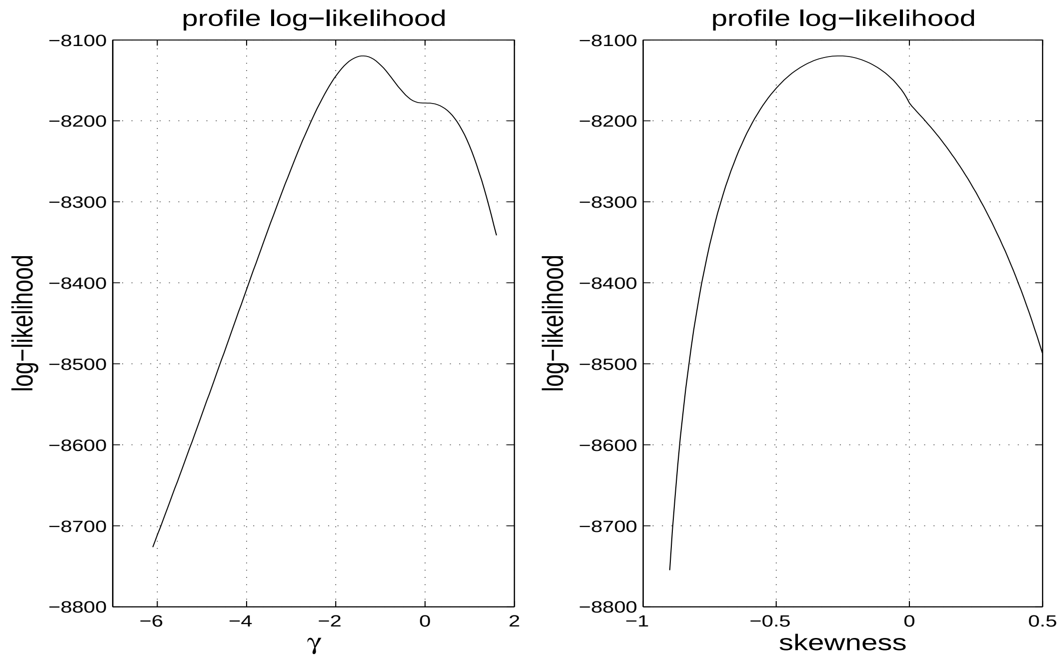

Figure 2: For the single–regime (symmetric, i.e., $\lambda = 0$ in (1.3)) SN–GARCH process applied to the DAX returns (described at the beginning of Section 3), the left plot displays the profile log–likelihood for the skewness parameter γ of the "direct" parametrization (2.1). The right plot repeats this, but for the reparametrization based on the skewness, see the discussion in Section 2.2.2.

of $T = 4980$ observations for each series.[9] Percentage log returns, r_t, are used, i.e., $r_t = 100 \times \log(I_t/I_{t-1})$, where I_t is the index level at time t. The index levels and returns are shown in Figure 3. Particularly for the CAC 40 and the FTSE 100, we observe that the sharp decline of the markets during the recent financial crisis was accompanied by a burst of (at least in our sample) unprecedented volatility, which can be expected to pose a serious challenge for all volatility models.

The first row of Table 8 (the row labeled $\{\widehat{\varepsilon}_t\}$) shows several moments of the three series. We note that the CAC 40 and the DAX 30 have a somewhat higher volatility (standard deviation) than the FTSE 100, as they represent less diversified portfolios. Also, not surprisingly, all return series display considerable excess kurtosis. In view of the latter, it is difficult to assess the significance of the skewness coefficients, which are all negative. The magnitude of the CAC 40's skewness may appear rather small; however, as pointed out by Kim and White (2004), the sample skewness measure is rather vulnerable to single large observations. For in-

[9]The data were obtained from Datastream.

stance, if we remove the single largest (three largest) [five largest] negative and positive observations from each return series, we get a skewness coefficient of -0.064 (-0.115) $[-0.156]$, -0.150 (-0.214) $[-0.221]$, and -0.121 (-0.147) $[-0.194]$ for the CAC 40, DAX 30, and FTSE 100, respectively, i.e., a considerable increase in all cases. Alternative, more robust measures of skewness have been proposed (cf. Kim and White, 2004); however, we shall concentrate on evaluating the importance of distributional asymmetries for out–of–sample predictive densities.

First–order dynamics are negligible in the series, so that our model is just

$$r_t = \mu + \varepsilon_t, \tag{3.1}$$

where μ is a constant mean parameter, and ε_t is described by one of the GARCH processes listed in the next section.

3.1 Models

The question that initially motivated this paper is whether the skew–normal MS GARCH is capable of curing the main defect of the Gaussian MS GARCH, i.e., its inability to reproduce the skewness observed in many financial return series. Consequently, both of these models will be included in our empirical study. A further model of interest is the iid mixture GARCH process discussed at the end of Section 2.2, as this offers another way of generating skewness without undesirable return autocorrelation. We denote this class of models by MixGARCH. For purpose of comparison, we also consider this model in the symmetric version, where all the component means are zero, so that it is a restricted version of MS GARCH. Both versions, i.e., with and without zero component means, will be considered both with normal and skew–normal innovations, and in the first case, i.e., with conditional normality, the zero–mean model will be denoted by Mix_S in order to distinguish it from the asymmetric Gaussian MixGARCH process. Moreover, another restricted version of the multi–regime GARCH processes turns out to be empirically relevant, namely, a model where only the intercept ω_j in (1.3) is allowed to switch. The volatility level is then regime–dependent, but its dynamics are not. These models will be labeled by the superscript $^{\text{icept}}$; for example, the symmetric iid normal mixture GARCH with switching intercept but common volatility dynamics will be denoted by $\text{Mix}_S^{\text{icept}}$. Note that, defining $\bar{\omega} = (1-\beta_1)^{-1}(\omega_2 - \omega_1)$, this model can be written as $\sigma_{2t} = \sigma_{1t} + \bar{\omega}$, a specification first considered by Vlaar and Palm (1992, 1993).

Models with $k = 1$ and $k = 2$ regimes will be evaluated, as models with more than two regimes will be plagued by substantial estimation error due to their large number of parameters (see the discussion and simulations in Alexander and Lazar,

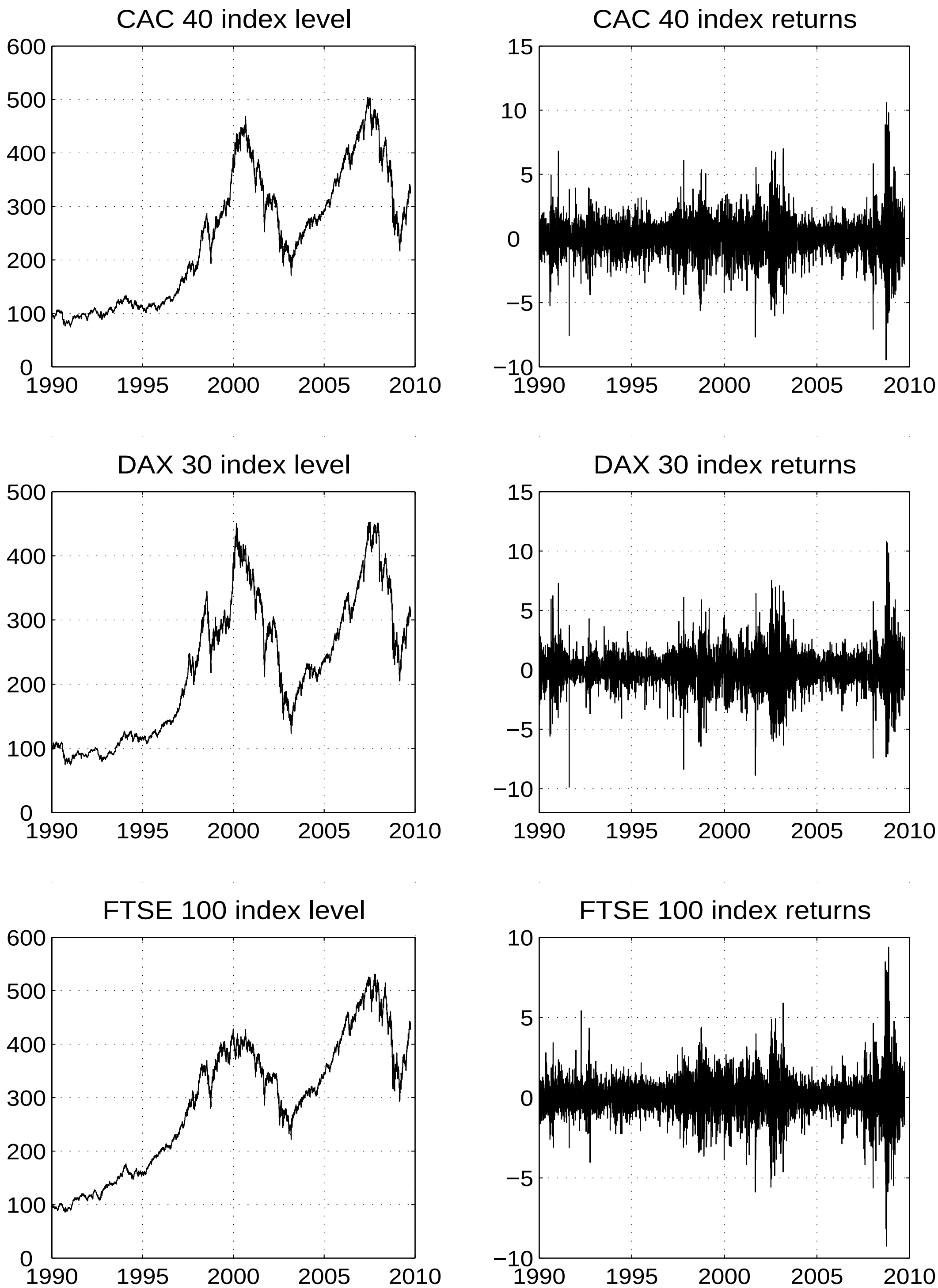

Figure 3: Index levels (left panel) and returns (right panel) of the CAC 40 (top), DAX 30 (center), and FTSE 100 (bottom), from January 1990 to October 2009.

Table 1: Overview of two–regime GARCH specifications

Model	Parameter restriction(s) Markov	GARCH	conditional density
Normal			
$\text{Mix}_S^{\text{icept}}$	$p_{11} = 1 - p_{22}$	$\alpha_1 = \alpha_2$; $\beta_1 = \beta_2$	$\mu_1 = \mu_2 = 0$; $\gamma = 0$
Mix_S	$p_{11} = 1 - p_{22}$	–	$\mu_1 = \mu_2 = 0$; $\gamma = 0$
$\text{Mix}^{\text{icept}}$	$p_{11} = 1 - p_{22}$	$\alpha_1 = \alpha_2$; $\beta_1 = \beta_2$	$\gamma = 0$; $\mu_2 = -\pi_{1,\infty}\mu_1/\pi_{2,\infty}$
Mix	$p_{11} = 1 - p_{22}$	–	$\gamma = 0$; $\mu_2 = -\pi_{1,\infty}\mu_1/\pi_{2,\infty}$
$\text{Markov}^{\text{icept}}$	–	$\alpha_1 = \alpha_2$; $\beta_1 = \beta_2$	$\mu_1 = \mu_2 = 0$; $\gamma = 0$
Markov	–	–	$\mu_1 = \mu_2 = 0$; $\gamma = 0$
Skew–normal			
$\text{Mix}^{\text{icept}}$	$p_{11} = 1 - p_{22}$	$\alpha_1 = \alpha_2$; $\beta_1 = \beta_2$	$\mu_1 = \mu_2 = 0$
Mix	$p_{11} = 1 - p_{22}$	–	$\mu_1 = \mu_2 = 0$
$\text{Markov}^{\text{icept}}$	–	$\alpha_1 = \alpha_2$; $\beta_1 = \beta_2$	$\mu_1 = \mu_2 = 0$
Markov	–	–	$\mu_1 = \mu_2 = 0$
Skew–normal with $\mu_1 \neq \mu_2$			
$\text{Mix}^{\text{icept}}$	$p_{11} = 1 - p_{22}$	$\alpha_1 = \alpha_2$; $\beta_1 = \beta_2$	$\mu_2 = -\pi_{1,\infty}\mu_1/\pi_{2,\infty}$
Mix	$p_{11} = 1 - p_{22}$	–	$\mu_2 = -\pi_{1,\infty}\mu_1/\pi_{2,\infty}$

The table presents the different multi–regime models included in the empirical comparison. Models with two regimes are considered, with the most general specification being given by

$$\varepsilon_t = \mu_{\Delta_t} + \sigma_{\Delta_t,t}\eta_t, \quad \Delta_t \in \{1,2\},$$

where the dynamics of the regime–specific variances are governed by (1.3), and the transition matrix (1.2) becomes

$$P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{pmatrix},$$

where $p_{ij} = p(\Delta_t = j|\Delta_{t-1} = i)$, $i, j = 1,2$, and the unconditional distribution of the Markov chain $\{\Delta_t\}$ is $\pi_\infty = (\pi_{1,\infty}, \pi_{2,\infty})'$, where $\pi_{1,\infty} = (1-p_{22})/(2-p_{11}-p_{22})$ and $\pi_{2,\infty} = 1-\pi_{1,\infty}$. If $\psi := p_{11}+p_{22}-1 = 0$, then $P = \pi_\infty 1_2'$, where $1_2 = (1,1)'$, and the conditional distribution of Δ_t is equal to its stationary distribution, π_∞. The innovation η_t follows a (centered) skew–normal distribution with shape parameter γ (see (2.1)), which becomes Gaussian if $\gamma = 0$. The restriction $\mu_2 = -\pi_{1,\infty}\mu_1/\pi_{2,\infty}$ is imposed in the mixture models in order to have a shock ε_t with (conditional and unconditional) zero mean.

2006). For later reference, an overview of the different multi–regime models is provided in Table 1.

Finally, to put the regime–switching models into perspective, we add to the list of competitors two rather popular and successful models which may serve as a benchmark, i.e., the single–regime GARCH(1,1) process with innovations following both a symmetric and an asymmetric t distribution, where for the skewed t we use the version proposed by Mittnik and Paolella (2000), which has density

$$f(z;\nu,p,\theta)=\frac{\theta}{1+\theta^2}\frac{p}{\nu^{1/p}B(\nu,1/p)}\begin{cases}\left(1+\frac{(|z|\theta)^p}{\nu}\right)^{-(\nu+1/p)} & \text{if} \quad z<0\\ \left(1+\frac{(z/\theta)^p}{\nu}\right)^{-(\nu+1/p)} & \text{if} \quad z\geq 0,\end{cases} \tag{3.2}$$

where $\nu,p,\theta>0$, and $B(\cdot,\cdot)$ is the beta function. Relevant for the subsequent discussion are the moments and the cdf of (3.2) which can be found in Kuester et al. (2006; note that there is a typo in their Equation (5) for the pdf). The mth moment is finite if $m<p\nu$, so that both p and ν interact in determining the tail behavior. Moreover, if $\theta=1$ and $p=2$ in (3.2), then $X=\sqrt{2}Z$ has a standard Student's t distribution with 2ν degrees of freedom.

3.2 In–sample results

Before turning to the evaluation of out–of–sample predictive densities, we describe a few estimation results for the entire sample period.

To assess the in–sample fit of the models, the values of the maximized log–likelihood along with the Bayesian information criterion (BIC) are reported in Tables 2, 3, and 4 for the CAC 40, DAX 30, and FTSE 100, respectively, where, for purpose of comparison, these tables also report the results for models with symmetric volatility dynamics, where $\lambda=0$ in (1.3). Parameter estimates for selected models are, with the same order of the market indices, displayed in Tables 5–7, where the regimes have been ordered with respect to a declining (stationary) regime probability, i.e., $\pi_{1,\infty}>\pi_{2,\infty}$. In the following, several facets of the results summarized in Tables 2–4 and 5–7 will be discussed in turn.

Importance of asymmetric volatility

We first note that, according to the BIC, and in accordance with the results in Rombouts and Bouaddi (2009), for all models and all three indices, the specifications with asymmetric volatility dynamics, where $\lambda\neq 0$ in (1.3), fit considerably better than their symmetric counterparts. The importance of including the leverage effect is also reflected in the magnitude and significance of the estimates of the asymmetry parameter λ reported in Tables 5–7. In view of the overwhelming superiority of the asymmetric GARCH models, unless otherwise stated, the subsequent discussion will focus exclusively on the members of this group, as reported in the right part of Tables 2–4.

Importance of skewness and its specification

The results suggest that (negative) skewness is a salient feature of all three return series. In Tables 5–7, for all models based on the SN distribution, the skewness pa-

Table 2: Likelihood–based goodness–of–fit: CAC 40

	Models with symmetric volatility				Models with asymmetric volatility			
			BIC				BIC	
Model	K	$\log L$	Value	Rank	K	$\log L$	Value	Rank
Normal								
Single	4	−8088.0	16210	32	5	−8028.3	16099	23
$\text{Mix}_S^{\text{icept}}$	6	−8025.8	16103	25	7	−7965.1	15990	7
Mix_S	8	−8020.6	16109	30	9	−7964.1	16005	13
$\text{Mix}^{\text{icept}}$	7	−8021.8	16103	27	8	−7960.2	15988	3
Mix	9	−8013.3	16103	28	10	−7959.5	16004	11
$\text{Markov}^{\text{icept}}$	7	−8001.2	16062	16	8	−7960.4	15989	4
Markov	9	−7999.5	16076	19	10	−7960.3	16006	14
Skew–normal								
Single	5	−8059.3	16161	31	6	−7999.2	16050	15
$\text{Mix}^{\text{icept}}$	7	−8017.9	16095	22	8	−7957.5	15983	2
Mix	9	−8011.9	16100	24	10	−7956.6	15998	10
$\text{Markov}^{\text{icept}}$	8	−7997.4	16063	17	9	−7951.6	15980	1
Markov	10	−7995.9	16077	20	11	−7951.3	15996	9
Skew–normal with $\mu_1 \neq \mu_2$								
$\text{Mix}^{\text{icept}}$	8	−8017.3	16103	26	9	−7956.4	15989	5
Mix	10	−8010.7	16107	29	11	−7955.5	16005	12
Student's t								
symmetric	5	−8025.5	16094	21	6	−7969.2	15990	6
skewed	7	−8007.5	16075	18	8	−7962.5	15993	8

Reported are likelihood–based goodness–of–fit measures for various GARCH models fitted to the CAC 40 returns. The left–most column specifies the type of model; the abbreviations for the different multi–regime models are detailed in Table 1. Models with symmetric volatility response impose $\lambda = 0$ in (1.3). K denotes the number of parameters of a model, $\log L$ is the value of the maximized log–likelihood, and BIC is the Bayesian information criterion, i.e., $\text{BIC} = -2 \times \log L + K \log T$, where T is the sample size.

rameter is significantly different from zero and negative. In the asymmetric mixture models based on the Gaussian distribution, the component with the lower probability (Regime 2) is associated with a higher volatility and a negative mean, which also gives rise to a left–skewed density. Comparing these skewed Gaussian mixture GARCH processes with their symmetric counterparts, i.e., comparing $\text{Mix}_S^{\text{icept}}$ with $\text{Mix}^{\text{icept}}$, and Mix_S with Mix, the BIC prefers the skewed versions for the CAC 40 and the DAX 30, whereas the symmetric specification is preferred for the FTSE

Table 3: Likelihood–based goodness–of–fit: DAX 30

	Models with symmetric volatility				Models with asymmetric volatility			
			BIC				BIC	
Model	K	$\log L$	Value	Rank	K	$\log L$	Value	Rank
Normal								
Single	4	−8178.0	16390	32	5	−8115.0	16273	30
$\text{Mix}_S^{\text{icept}}$	6	−8029.4	16110	24	7	−7980.6	16021	8
Mix_S	8	−8027.1	16122	28	9	−7979.3	16035	13
$\text{Mix}^{\text{icept}}$	7	−8024.7	16109	22	8	−7974.6	16017	6
Mix	9	−8022.4	16121	27	10	−7973.4	16032	11
$\text{Markov}^{\text{icept}}$	7	−8019.0	16098	19	8	−7977.3	16023	9
Markov	9	−8018.6	16114	26	10	−7976.8	16039	14
Skew–normal								
Single	5	−8119.7	16282	31	6	−8062.3	16176	29
$\text{Mix}^{\text{icept}}$	7	−8014.4	16088	17	8	−7963.3	15995	1
Mix	9	−8012.8	16102	20	10	−7962.7	16011	4
$\text{Markov}^{\text{icept}}$	8	−8008.9	16086	16	9	−7962.2	16001	2
Markov	10	−8008.6	16102	21	11	−7961.8	16017	5
Skew–normal with $\mu_1 \neq \mu_2$								
$\text{Mix}^{\text{icept}}$	8	−8014.0	16096	18	9	−7962.6	16002	3
Mix	10	−8012.3	16110	23	11	−7962.1	16018	7
Student's t								
symmetric	5	−8034.0	16111	25	6	−7990.9	16033	12
skewed	7	−8012.7	16085	15	8	−7981.0	16030	10

Reported are likelihood–based goodness–of–fit measures for various GARCH models fitted to the DAX 30 returns. See the footnote of Table 2 for further explanations.

100. However, for comparing nested models with a fixed number of components, a likelihood ratio test (LRT) may be more appropriate than the rather conservative BIC, and this gives, for both of the abovementioned comparisons, a p–value below 0.02 for the FTSE. Similarly, regarding the Student's t models, the BIC prefers the symmetric version for the CAC 40 and the FTSE 100; however, an LRT of the hypothesis $\theta = 1$ and $p = 2$ in (3.2) rejects with a p-value below 0.005, for each index.

As expected, the (symmetric and skewed) Student's t processes outperform both the normal and the SN single–regime specifications, but they are, for all three series,

Table 4: Likelihood–based goodness–of–fit: FTSE 100

	Models with symmetric volatility				Models with asymmetric volatility			
			BIC				BIC	
Model	K	$\log L$	Value	Rank	K	$\log L$	Value	Rank
Normal								
Single	4	−6796.8	13628	32	5	−6739.0	13520	16
$\text{Mix}_S^{\text{icept}}$	6	−6768.7	13588	25	7	−6708.4	13476	2
Mix_S	8	−6762.5	13593	29	9	−6707.2	13491	10
$\text{Mix}^{\text{icept}}$	7	−6764.5	13589	26	8	−6705.6	13479	5
Mix	9	−6754.8	13586	24	10	−6704.1	13493	11
$\text{Markov}^{\text{icept}}$	7	−6752.6	13565	18	8	−6708.2	13484	8
Markov	9	−6746.9	13570	20	10	−6706.8	13499	14
Skew–normal								
Single	5	−6782.5	13607	31	6	−6726.9	13505	15
$\text{Mix}^{\text{icept}}$	7	−6762.2	13584	23	8	−6702.7	13473	1
Mix	9	−6756.0	13589	27	10	−6701.6	13488	9
$\text{Markov}^{\text{icept}}$	8	−6748.8	13566	19	9	−6702.3	13481	6
Markov	10	−6744.1	13573	21	11	−6701.0	13496	12
Skew–normal with $\mu_1 \neq \mu_2$								
$\text{Mix}^{\text{icept}}$	8	−6761.7	13591	28	9	−6702.4	13481	7
Mix	10	−6754.3	13594	30	11	−6701.3	13496	13
Student's t								
symmetric	5	−6766.9	13576	22	6	−6713.3	13478	3
skewed	7	−6749.7	13559	17	8	−6705.5	13479	4

Reported are likelihood–based goodness–of–fit measures for various GARCH models fitted to the FTSE 100 returns. See the footnote of Table 2 for further explanations.

dominated in terms of BIC by (at least) the best representative of both the SN and the Gaussian mixture families. This leads us to the comparison of mixture models based on the normal with those based on the SN distribution. In view of the evidence for skewness reported in the previous paragraph, it is not surprising that the SN MS GARCH processes dominate their symmetric normal counterparts. Thus, comparisons between the asymmetric Gaussian and the SN MixGARCH specifications will perhaps be more interesting. Three combinations are possible and accounted for in Tables 2–4 and 5–7: First, skewness may be generated by Gaussian innovations with component–specific means (i.e., in (1.5), $\mu_1 \neq \mu_2$), and second, by SN innovations with zero means. A third alternative which nests both of these is SN innovations with $\mu_1 \neq \mu_2$. Comparing the first two alternatives, models employing the SN achieve a higher likelihood and, as they have the same number of parameters, a lower BIC for all indices. The two models are not nested and LRTs cannot be conducted straightforwardly. However, both can be tested against the third, more general alternative, and it turns out that the Gaussian model is rejected,[10] whereas the restricted SN model is not. Thus, once skewness is introduced via the SN density, the contribution of nonzero component means appears negligible, whereas allowing for asymmetric innovations adds significant improvement to the asymmetric Gaussian mixture GARCH. This is also reflected in the parameter estimates in Tables 5–7. In the SN mixture models with $\mu_1 \neq \mu_2$, as displayed in the two rightmost columns, the magnitude of the estimated means is considerably reduced as compared to the Gaussian models, whereas the estimated skewness parameter γ remains almost unchanged when moving from the restricted SN models in columns 5 and 6 of Tables 5–7 to those with $\mu_1 \neq \mu_2$.

Summarizing, it appears that the skewness in the three index return series is better described by a mixture of SN distributions rather than by a skewed Student's *t* density or a mixture of Gaussians with skewness introduced via different component means.

Regime–specific volatility dynamics and the mixing process

According to the BIC, the best model belongs to the class of SN mixture GARCH processes, for all three series. Moreover, parsimoniously parameterized models with restricted volatility dynamics and mixing processes tend to be preferred:

First, in all cases, the preferred model is characterized by regime–independent volatility dynamics, where $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. Note that this result does not depend on the rather conservative BIC, since LRTs lead to the same conclusion. This may be somewhat surprising at first sight, since the estimates of most unrestricted mixture models reported in Tables 5–7 exhibit a striking pattern which has similarly been documented in several earlier papers, namely, that in the high–volatility component, which occurs less frequently, conditional volatility is "highly reactive to market shocks in the crash regime, yet because the persistence parameters are [...] low, [...] the effect of a shock soon dies out" (Alexander and Lazar, 2009), i.e., α_2 is considerably larger than α_1, whereas β_2 is smaller than β_1 (see also Alexander and Lazar, 2006; Badescu et al., 2008; Haas et al., 2004a, 2009; Cheng et al., 2009; and Bauwens and Storti, 2009). However, given the rather small mixing weights of and thus the scarcity of observations drawn from the high–volatility components in

[10] p–values are smaller than 0.01 for the CAC 40 and the DAX 30 and smaller than 0.02 for the

most of these models, the parameters of the associated volatility processes are subject to large estimation error, rendering the differences in the components' volatility dynamics insignificant.

Second, for the DAX 30 and the FTSE 100, the preferred model is characterized by an iid rather than a persistent Markovian mixing process. Whereas, in Tables 5–7, the transition matrix P of the CAC 40 (Table 5) exhibits a high degree of persistence, there is less memory in the chain estimated for the DAX 30 (Table 6), and almost no memory in that of the FTSE 100. In the latter two models, the subdominant eigenvalue of P, given by $\psi = p_{11} + p_{22} - 1$, is statistically insignificant due to a rather high standard error of the second regime's staying probability p_{22}.

Rather parsimonious mixture models are thus preferred, and we note that, for example, the SN MixGARCH process with regime–independent volatility dynamics has the same number of parameters as the skewed t GARCH process given by (1.3) with $k = 1$ and (3.2).

The power parameter d in (1.4)

Next, we consider the issue of the optimal power parameter d in the general APGARCH specification (1.4). To figure out the impact of both the conditional density and the specification of the volatility dynamics on the optimal d, we estimate the APGARCH process (1.4) for $d \in \{0.75, 0.8, \ldots, 2\}$ under both the single–regime SN GARCH process and the SN Mix$^{\text{icept}}$ specification, which performs best overall according to the BIC. Both models are considered with symmetric as well as asymmetric volatility, i.e., $\lambda = 0$ and $\lambda \neq 0$ in (1.3), respectively. Figure 4 shows, for the four model variants and the three return series, the value of the maximized log–likelihood as a function of d. For the single–regime SN PGARCH with $\lambda = 0$, shown in the top panel of Figure 4, $d = 2$ turns out to be superior to $d = 1$. In fact, an LRT rejects $d = 1$ for all series at the 1% level, whereas $d = 2$ is rejected only for the DAX 30 at 5%. The situation is dramatically changed by introducing asymmetric volatility dynamics, as shown in next panel. The optimal d moves towards unity for all three series, and $d = 2$ is now rejected at the 1% level for all indices and is inferior to $d = 1$, which, however, is still rejected at the 10% and 5% levels for the CAC 40 and the DAX 30, respectively. Employing a conditional mixture distribution with symmetric volatility response (third panel of Figure 4), the hypothesis $d = 1$ is, just as for the single–regime model, rejected at the 1% level for all indices, but is superior to $d = 2$ for the DAX 30. Again, introduction of asymmetry into the volatility equation has a considerable effect on the optimal d, since $d = 2$ is rejected

FTSE 100.

Table 5: Parameter estimates for absolute value GARCH models with asymmetric volatility dynamics for the CAC 40

	Normal		Skew–normal					Skew–normal with $\mu_1 \neq \mu_2$	
	Mix$^{\text{icept}}$	Mix	Single	Mix$^{\text{icept}}$	Mix	Markov$^{\text{icept}}$	Markov	Mix$^{\text{icept}}$	Mix
ω_1	0.017 (0.003)	0.017 (0.003)	0.027 (0.004)	0.022 (0.004)	0.021 (0.004)	0.020 (0.004)	0.020 (0.004)	0.021 (0.004)	0.021 (0.004)
α_1	0.066 (0.006)	0.065 (0.006)	0.083 (0.008)	0.081 (0.008)	0.079 (0.008)	0.057 (0.006)	0.055 (0.007)	0.079 (0.008)	0.078 (0.008)
β_1	0.932 (0.006)	0.934 (0.006)	0.933 (0.006)	0.932 (0.006)	0.933 (0.006)	0.948 (0.005)	0.950 (0.005)	0.932 (0.006)	0.933 (0.006)
$\pi_{1,\infty}$	0.973 (0.014)	0.971 (0.014)	1	0.985 (0.010)	0.981 (0.011)	0.911	0.907	0.983 (0.010)	0.980 (0.011)
μ_1	0.027 (0.011)	0.027 (0.011)	0	0	0	0	0	0.012 (0.009)	0.012 (0.009)
ω_2	0.095 (0.021)	0.250 (0.174)	–	0.152 (0.044)	0.468 (0.380)	0.066 (0.011)	0.091 (0.070)	0.139 (0.039)	0.349 (0.287)
α_2	0.066 (0.006)	0.171 (0.099)	–	0.081 (0.008)	0.262 (0.171)	0.057 (0.006)	0.072 (0.024)	0.079 (0.008)	0.231 (0.145)
β_2	0.932 (0.006)	0.822 (0.106)	–	0.932 (0.006)	0.779 (0.148)	0.948 (0.005)	0.932 (0.041)	0.932 (0.006)	0.819 (0.122)
$\pi_{2,\infty}$	0.027 (0.014)	0.029 (0.014)	0	0.015 (0.010)	0.017 (0.011)	0.089	0.093	0.017 (0.010)	0.020 (0.011)
μ_2	−0.982 (0.399)	−0.902 (0.366)	–	0	0	0	0	−0.673 (0.474)	−0.579 (0.430)
λ	0.683 (0.076)	0.668 (0.077)	0.653 (0.075)	0.674 (0.074)	0.664 (0.075)	1 (–)	1 (–)	0.674 (0.074)	0.663 (0.075)
skewness	0	0	−0.206 (0.030)	−0.152 (0.038)	−0.152 (0.039)	−0.177 (0.040)	−0.178 (0.039)	−0.125 (0.043)	−0.127 (0.043)
γ	0	0	−1.217	−1.048	−1.049	−1.128	−1.131	−0.960	−0.968
P	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$	1	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$	$\begin{pmatrix} \underset{(0.003)}{0.992} & \underset{(0.047)}{0.081} \\ \underset{(0.003)}{0.008} & \underset{(0.047)}{0.919} \end{pmatrix}$	$\begin{pmatrix} \underset{(0.004)}{0.992} & \underset{(0.055)}{0.082} \\ \underset{(0.004)}{0.008} & \underset{(0.055)}{0.918} \end{pmatrix}$	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$
ψ	0	0	0	0	0	0.911 (0.093)	0.910 (0.144)	0	0

Approximate standard errors are given in parentheses. For models Markov$^{\text{icept}}$ and Markov, where the asymmetry parameter is on the boundary (i.e., $\gamma = 1$), we reestimated the model with γ fixed to unity, so that standard errors are not reported. "Single" denotes a model with just one regime, i.e., $k = 1$ in (1.1)–(1.3). See Table 1 for the definitions of the different multi–regime models. "skewness" is the moment–based coefficient of skewness (2.4) of the *innovations*, η_t, as implied by the estimated shape parameter, γ, of the skew–normal density (2.1), and ψ is the measure of the persistence of the regimes, i.e., the largest subdominant eigenvalue of the transition matrix P, given by $\psi = p_{11} + p_{22} - 1$, see the footnote of Table 1.

Table 6: Parameter estimates for absolute value GARCH models with asymmetric volatility dynamics for the DAX 30

	Normal		Skew–normal					Skew–normal with $\mu_1 \neq \mu_2$	
	Mix$^{\text{icept}}$	Mix	Single	Mix$^{\text{icept}}$	Mix	Markov$^{\text{icept}}$	Markov	Mix$^{\text{icept}}$	Mix
ω_1	0.015 (0.003)	0.014 (0.003)	0.036 (0.005)	0.023 (0.004)	0.022 (0.004)	0.020 (0.004)	0.019 (0.004)	0.022 (0.004)	0.021 (0.004)
α_1	0.075 (0.007)	0.073 (0.007)	0.101 (0.009)	0.101 (0.009)	0.100 (0.009)	0.091 (0.011)	0.092 (0.013)	0.100 (0.009)	0.098 (0.010)
β_1	0.925 (0.007)	0.928 (0.007)	0.920 (0.007)	0.922 (0.007)	0.923 (0.007)	0.929 (0.008)	0.929 (0.009)	0.922 (0.007)	0.924 (0.007)
$\pi_{1,\infty}$	0.970 (0.010)	0.965 (0.011)	1	0.983 (0.007)	0.981 (0.007)	0.978	0.977	0.981 (0.007)	0.980 (0.008)
μ_1	0.031 (0.010)	0.033 (0.011)	0	0	0	0	0	0.009 (0.008)	0.009 (0.009)
ω_2	0.123 (0.022)	0.402 (0.273)	–	0.209 (0.045)	0.667 (0.633)	0.172 (0.038)	0.411 (0.489)	0.202 (0.045)	0.551 (0.499)
α_2	0.075 (0.007)	0.160 (0.087)	–	0.101 (0.009)	0.213 (0.180)	0.091 (0.011)	0.118 (0.128)	0.100 (0.009)	0.198 (0.155)
β_2	0.925 (0.007)	0.775 (0.134)	–	0.922 (0.007)	0.769 (0.196)	0.929 (0.008)	0.852 (0.166)	0.922 (0.007)	0.799 (0.163)
$\pi_{2,\infty}$	0.030 (0.010)	0.035 (0.011)	0	0.017 (0.007)	0.019 (0.007)	0.022	0.023	0.019 (0.007)	0.020 (0.008)
μ_2	−0.981 (0.306)	−0.910 (0.288)	–	0	0	0	0	−0.489 (0.404)	−0.440 (0.381)
λ	0.564 (0.064)	0.566 (0.065)	0.555 (0.060)	0.556 (0.064)	0.558 (0.064)	0.589 (0.072)	0.583 (0.076)	0.559 (0.064)	0.561 (0.064)
skewness	0	0	−0.260 (0.028)	−0.231 (0.038)	−0.227 (0.039)	−0.217 (0.040)	−0.217 (0.041)	−0.215 (0.041)	−0.212 (0.042)
γ	0	0	−1.379	−1.292	−1.282	−1.251	−1.249	−1.245	−1.236
P	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$	1	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$	$\begin{pmatrix} \underset{(0.006)}{0.987} & \underset{(0.268)}{0.547} \\ \underset{(0.006)}{0.013} & \underset{(0.268)}{0.453} \end{pmatrix}$	$\begin{pmatrix} \underset{(0.011)}{0.985} & \underset{(0.500)}{0.625} \\ \underset{(0.011)}{0.015} & \underset{(0.500)}{0.375} \end{pmatrix}$	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$
ψ	0	0	0	0	0	0.440 (0.273)	0.360 (0.510)	0	0

Approximate standard errors are given in parentheses. See the legend of Table 5 for further explanations.

Table 7: Parameter estimates for absolute value GARCH models with asymmetric volatility dynamics for the FTSE 100

	Normal		Skew–normal					Skew–normal with $\mu_1 \neq \mu_2$	
	Mix$^{\text{icept}}$	Mix	Single	Mix$^{\text{icept}}$	Mix	Markov$^{\text{icept}}$	Markov	Mix$^{\text{icept}}$	Mix
ω_1	0.010 (0.002)	0.011 (0.002)	0.015 (0.003)	0.013 (0.002)	0.013 (0.003)	0.013 (0.002)	0.013 (0.003)	0.013 (0.002)	0.013 (0.003)
α_1	0.063 (0.006)	0.060 (0.006)	0.079 (0.008)	0.076 (0.008)	0.074 (0.008)	0.074 (0.008)	0.072 (0.008)	0.075 (0.008)	0.073 (0.008)
β_1	0.938 (0.006)	0.939 (0.006)	0.937 (0.006)	0.938 (0.006)	0.939 (0.006)	0.939 (0.006)	0.940 (0.006)	0.938 (0.006)	0.939 (0.006)
$\pi_{1,\infty}$	0.965 (0.018)	0.957 (0.023)	1	0.969 (0.017)	0.966 (0.019)	0.969	0.964	0.969 (0.017)	0.966 (0.019)
μ_1	0.016 (0.008)	0.019 (0.009)	0	0	0	0	0	0.005 (0.008)	0.006 (0.008)
ω_2	0.057 (0.014)	0.035 (0.046)	–	0.074 (0.019)	0.050 (0.097)	0.074 (0.019)	0.030 (0.036)	0.074 (0.019)	0.064 (0.133)
α_2	0.063 (0.006)	0.103 (0.065)	–	0.076 (0.008)	0.120 (0.111)	0.074 (0.008)	0.097 (0.052)	0.075 (0.008)	0.136 (0.140)
β_2	0.938 (0.006)	0.934 (0.053)	–	0.938 (0.006)	0.935 (0.080)	0.939 (0.006)	0.952 (0.031)	0.938 (0.006)	0.923 (0.107)
$\pi_{2,\infty}$	0.035 (0.018)	0.043 (0.023)	0	0.031 (0.017)	0.031 (0.019)	0.031	0.036	0.031 (0.017)	0.034 (0.019)
μ_2	−0.440 (0.229)	−0.408 (0.205)	–	0	0	0	0	−0.158 (0.241)	−0.174 (0.232)
λ	0.708 (0.085)	0.701 (0.089)	0.653 (0.082)	0.715 (0.087)	0.712 (0.092)	0.737 (0.093)	0.745 (0.098)	0.711 (0.086)	0.703 (0.092)
skewness	0	0	−0.147 (0.031)	−0.138 (0.041)	−0.141 (0.042)	−0.141 (0.041)	−0.144 (0.042)	−0.122 (0.047)	−0.121 (0.050)
γ	0	0	−1.032	−1.003	−1.013	−1.012	−1.024	−0.948	−0.944
P	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$	1	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$	$\begin{pmatrix} \underset{(0.015)}{0.975} & \underset{(0.219)}{0.784} \\ \underset{(0.015)}{0.025} & \underset{(0.219)}{0.216} \end{pmatrix}$	$\begin{pmatrix} \underset{(0.015)}{0.972} & \underset{(0.226)}{0.739} \\ \underset{(0.015)}{0.028} & \underset{(0.226)}{0.261} \end{pmatrix}$	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_2'$
ψ	0	0	0	0	0	0.190 (0.225)	0.234 (0.231)	0	0

Approximate standard errors are given in parentheses. See the legend of Table 5 for further explanations.

anew at the 1% level in all cases, whereas $d = 1$ is never rejected at conventional levels. The bottom row of Table 4 shows that, in accordance with the references referred to in Footnote 3, the log–likelihood as a function of d attains its maximum in the vicinity of 1 for all three indices.

Unconditional moments

Unconditional moments of the models are reported in Table 8, along with, in the first row, the corresponding sample analogues calculated from the demeaned return series. In addition to standard deviation, skewness, and kurtosis, Table 8 also reports the mean/standard deviation ratio of the absolute (demeaned) returns. The latter quantity is motivated by the observation of Granger and Ding (1995) that often, for the absolute errors, mean $\approx$ standard deviation (for further evidence, see also Rydén et al., 1998; Granger et al., 2000; and Bulla and Bulla, 2006), and Table 8 shows that this is indeed the case for the three stock return series investigated in this paper.

From the results in Table 8, all models perform reasonably well in matching the unconditional volatility, whereas there are more remarkable differences for the mean/standard deviation ratio (henceforth abbreviated m/s ratio) and the higher–order moments. In particular, the Gaussian single–regime model fails to reproduce the m/s ratio and the kurtosis of the CAC 40 and the DAX 30; it matches these quantities surprisingly well for the FTSE 100, however. Basically the same comment applies to the single–regime SN, which exhibits a somewhat closer fit for all three series, though. Overall, the moments implied by the mixture GARCH models display a much greater similarity to their empirical counterparts, and the models based on the SN distribution again do better than the Gaussian mixtures. Note, however, that Table 8 contains results only for the symmetric normal mixtures, since expressions for the moments of their asymmetric relatives are not available. The skewed Student's t GARCH model matches the empirical m/s ratios particularly well. Its theoretical kurtosis measures are rather large, but, as indicated in Section 2.2.1, an unreasonably large value of the fourth moment just reflects the fact that (the estimated) $\rho(C_{44})$ in (B.4) is close to unity and the unconditional fourth moment may actually not be finite, or at least its sample analogue may have an extremely fat–tailed distribution.

Along with the corresponding empirical autocorrelations, Figure 5 shows the theoretical autocorrelation functions (ACFs) of both the absolute and the squared returns, as implied by both the BIC–preferred SN mixture models[11] (left panel) and the skewed t GARCH process (right panel). These autocorrelations are defined in (B.8) and denoted by $\rho_1(\tau)$ and $\rho_2(\tau)$, respectively. All three series display a form

[11] That is, shown are the ACFs implied by model SN Markov$^{\text{icept}}$ for the CAC 40 and those of

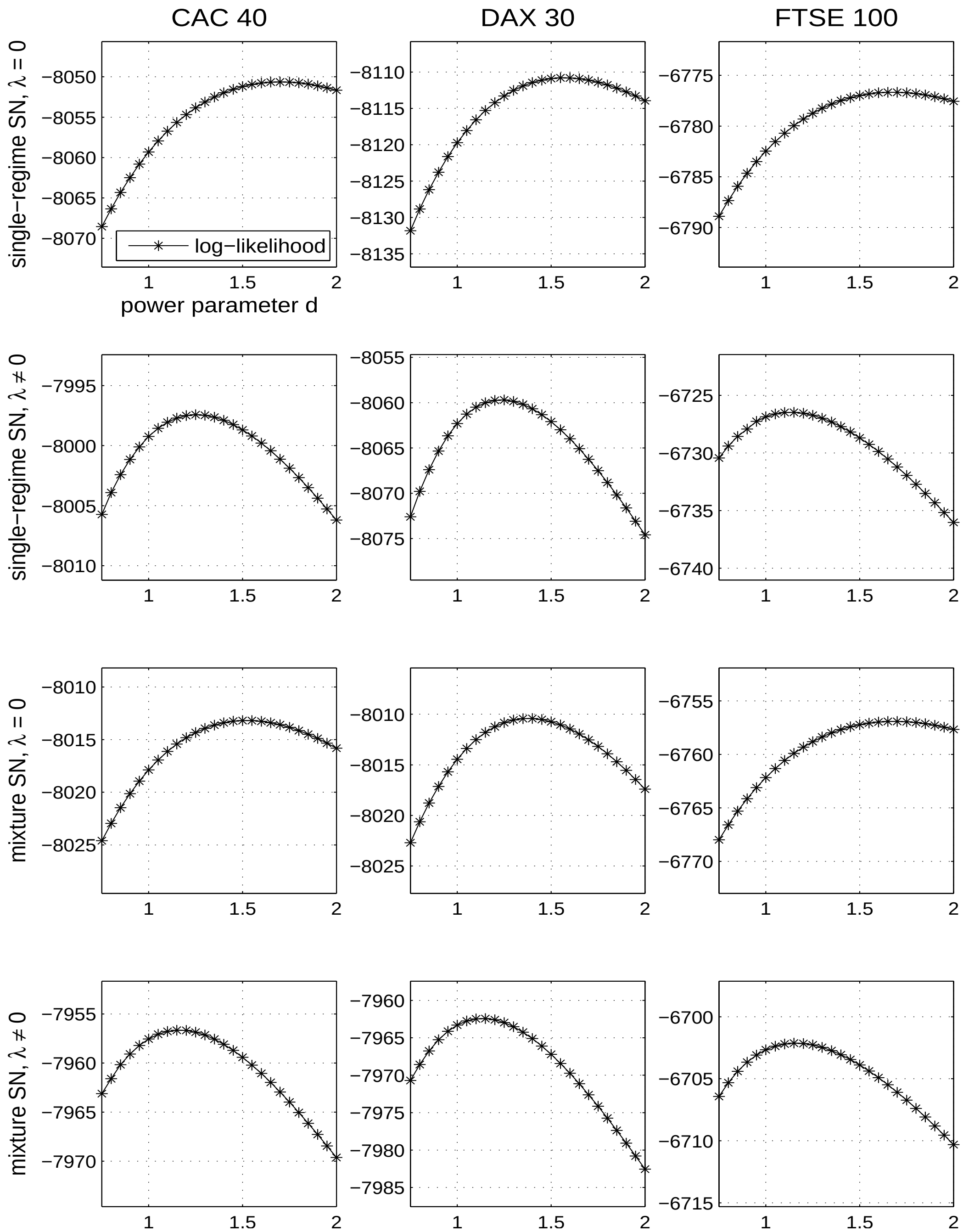

Figure 4: For the APGARCH specification (1.3), the figure shows the value of the maximized log–likelihood as a function of the power parameter d both for the single–regime SN GARCH ($k = 1$) and for the SN Mix$^{\text{icept}}$ GARCH model (cf. Table 1). Both models are considered with symmetric as well as asymmetric volatility, i.e., $\lambda = 0$ and $\lambda \neq 0$ in (1.3), respectively.

of the "Taylor effect" in that the ACF of the squares decays to zero faster than that of the absolute values; except for the first lag, it also starts considerably higher for the DAX 30 (cf. Ding et al., 1993; Ding and Granger, 1996; and Granger, 2005). Figure 5 shows that there is a pronounced Taylor effect in the ACFs implied by the skewed t GARCH model and in particular so for the DAX 30, whereas there are only minor differences between the ACFs of the absolute and the squared values implied by the SN mixture processes.[12] Thus, the skewed t GARCH process tracks the autocorrelations of the absolute values better for the DAX 30. The SN mixture GARCH, on the other hand, tends to track the ACF of the squares better, and it exhibits a particularly close fit to the first 50 lags of $\rho_1(\tau)$ for the CAC 40. In summary, both models overall provide a satisfactory fit to the temporal properties of the returns as measured by the ACF of squared and absolute valued observations. More formal methods of out-of–sample model evaluation are considered next.

3.3 Out–of–sample results

Now we turn to the evaluation of the one–step–ahead out–of–sample forecast densities of the models. We first reestimate the models over the (approximately) first ten years of data, i.e., the first 2500 observations, and then update the parameters (approximately) every month (i.e., 20 trading days) employing a moving window of data, i.e., using the most recent 2500 observations in the sample. In this manner, we obtain, for each model and series, 2480 one–step–ahead out–of–sample forecast densities, which will be evaluated by means of the actual observations. The techniques applied for this purpose are explained in Section 3.3.1, and Section 3.3.2 presents the empirical results.

3.3.1 Diagnostic checking

To assess the adequacy of the models, we evaluate both their entire forecast densities as well as their implied Value–at–Risk (VaR) measures. For the first type of evaluation procedure, we employ the transformation due to Rosenblatt (1952), see also Smith (1985). To this end, we calculate the sequence of "realized" return distribution functions, $u_t = \widehat{F}(r_t|\Psi_{t-1})$, $t = 2501,\ldots,4980$, where Ψ_t is the information

model SN Mix$^{\text{icept}}$ for the DAX 30 and FTSE 100.

[12] He and Teräsvirta (1999) considered the Gaussian (absolute value) GARCH(1,1) process and find that this produces the Taylor effect only for parameter constellations associated with "very strong leptokurtosis", but Figure 1 of He and Teräsvirta (1999) suggests that even then the effect appears to be quantitatively negligible. GARCH processes with leptokurtic innovations, however, are more likely to generate a significant Taylor effect (cf. Gonçalves et al., 2009; and Haas, 2009; for stochastic volatility models, similar conclusions regarding the role of conditional kurtosis are

Table 8: Unconditional moments of estimated GARCH models

	CAC 40				DAX 30				FTSE 100			
	Std.	$\frac{\mathrm{E}(\|\varepsilon_t\|)}{\mathrm{Std}(\|\varepsilon_t\|)}$	Skew.	Kurt.	Var.	$\frac{\mathrm{E}(\|\varepsilon_t\|)}{\mathrm{Std}(\|\varepsilon_t\|)}$	Skew.	Kurt.	Std.	$\frac{\mathrm{E}(\|\varepsilon_t\|)}{\mathrm{Std}(\|\varepsilon_t\|)}$	Skew.	Kurt.
$\{\widehat{\varepsilon}_t\}$	1.41	1.02	−0.039	7.81	1.49	0.98	−0.129	8.07	1.15	0.98	−0.114	9.63
Normal												
Single	1.42	1.16		4.89	1.48	1.17	0	4.82	1.15	1.09	0	6.82
$\mathrm{Mix}_S^{\mathrm{icept}}$	1.37	1.10	0	6.25	1.39	1.05	0	7.41	1.05	1.07	0	7.05
Mix_S	1.36	1.10	0	6.33	1.38	1.06	0	7.06	1.05	1.06	0	7.97
$\mathrm{Markov}^{\mathrm{icept}}$	1.36	1.09	0	6.15	1.39	1.05	0	7.41	1.05	1.07	0	6.93
Markov	1.36	1.09	0	6.08	1.38	1.06	0	6.87	1.05	1.06	0	7.99
Skew–normal												
Single	1.45	1.12	−0.256	5.87	1.52	1.12	−0.325	6.04	1.14	1.08	−0.197	7.48
$\mathrm{Mix}^{\mathrm{icept}}$	1.38	1.09	−0.199	6.64	1.42	1.03	−0.327	8.48	1.06	1.06	−0.188	7.54
Mix	1.38	1.09	−0.200	6.72	1.41	1.04	−0.318	8.13	1.06	1.05	−0.198	8.38
$\mathrm{Markov}^{\mathrm{icept}}$	1.40	1.07	−0.234	6.65	1.41	1.03	−0.306	8.18	1.06	1.06	−0.191	7.35
Markov	1.40	1.07	−0.236	6.77	1.40	1.04	−0.299	7.72	1.06	1.05	−0.204	8.59
Student's t												
symmetric	1.32	1.06	0	7.38	1.32	1.01	0	10.1	1.03	1.05	0	7.97
skewed	1.48	1.02	−0.250	11.1	1.59	0.94	−0.477	24.2	1.19	0.99	−0.290	15.4

Shown are unconditional moments implied by estimated GARCH models, as well as the corresponding sample quantities of the (demeaned) return series. The latter are reported in the row labeled $\{\widehat{\varepsilon}_t\}$. "Std." is the standard deviation, and "Skew." and "Kurt." denote the moment–based coefficients of skewness and kurtosis of ε_t, respectively, i.e., the third and fourth standardized moments. The mean and standard deviation of $|\varepsilon_t|$ are denoted by $\mathrm{E}(|\varepsilon_t|)$ and $\mathrm{Std}(|\varepsilon_t|)$, respectively. For the notation of the two–regime models, see Table 1. The skewed Student's t distribution is defined by (3.2).

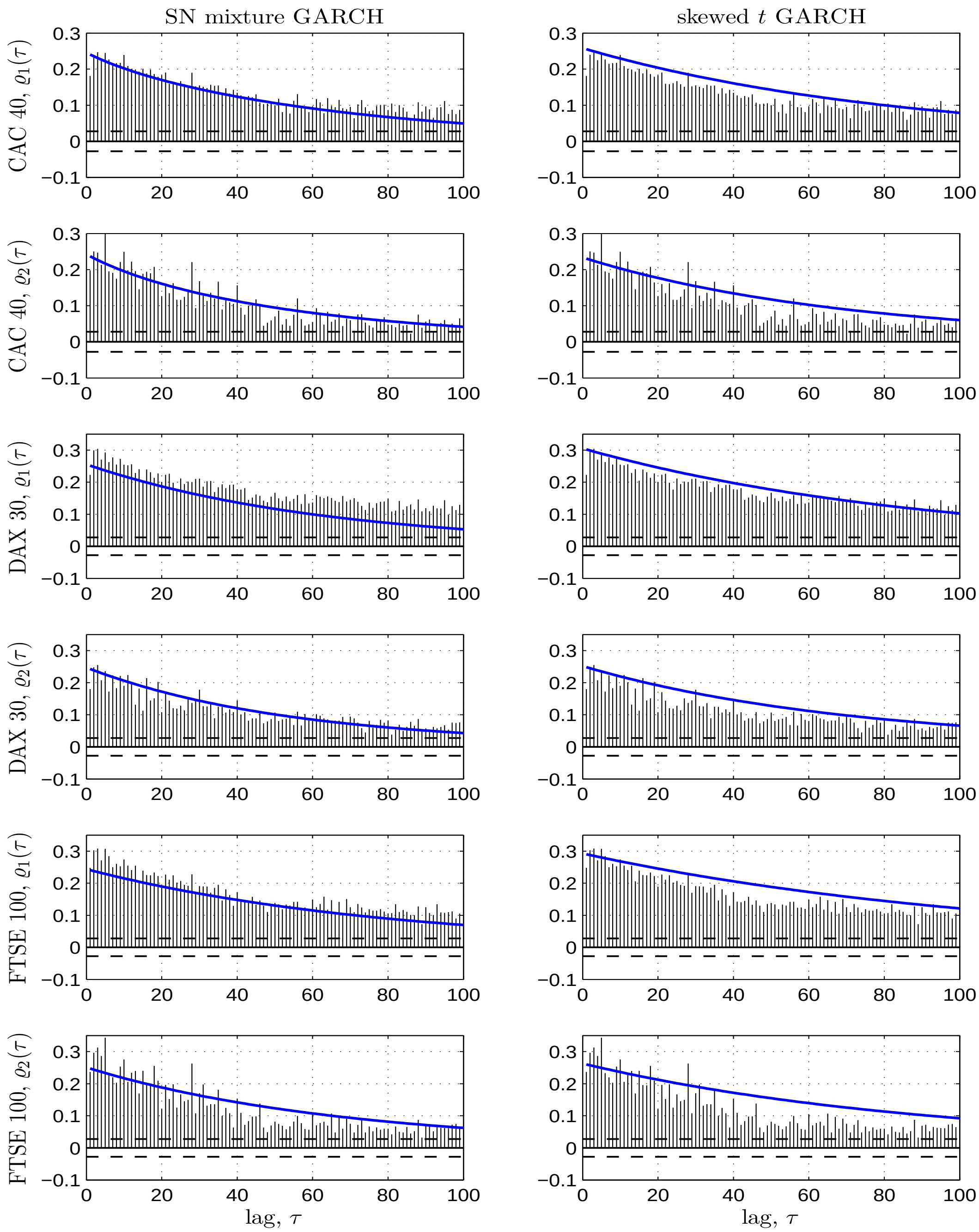

Figure 5: The left panel shows the empirical autocorrelations of absolute and squared (demeaned) returns ($\rho_1(\tau)$ and $\rho_2(\tau)$, respectively), along with their theoretical counterparts implied by the best SN mixture GARCH process according to BIC. The right panel repeats this for the skewed t GARCH process.

up to time t, and $\widehat{F}(\cdot|\Psi_{t-1})$ is the conditional cdf of the return implied by the model under consideration. Subsequently, we apply a second transformation, namely,

$$\{z_t\} = \Phi^{-1}(\{u_t\}), \tag{3.3}$$

where Φ^{-1} is the inverse of the standard normal cdf. The sequence $\{z_t\}$ is iid N(0,1) if the underlying model is correct, and Berkowitz (2001) shows that inaccuracies in the predictive density will be preserved in the transformed data (3.3). This transformation therefore allows the use of moment–based normality tests for checking features such as correct specification of skewness and kurtosis, using the result that, under normality,

$$T\widehat{m}_3^2/6 \overset{asy}{\sim} \chi^2(1), \quad \text{and} \quad T(\widehat{m}_4 - 3)^2/24 \overset{asy}{\sim} \chi^2(1), \tag{3.4}$$

where $\widehat{m}_3$ and $\widehat{m}_4$ are the sample skewness and kurtosis of $\{z_t\}$, respectively, and $T = 2480$ is the number of forecasts evaluated. A joint test for correctly specified skewness and kurtosis is provided by the Jarque–Bera (JB) test,

$$JB = T\widehat{m}_3^2/6 + T(\widehat{m}_4 - 3)^2/24 \overset{asy}{\sim} \chi^2(2). \tag{3.5}$$

In a second class of tests, we use the likelihood ratio approach devised by Berkowitz (2001) to construct tests for the first four moments of the forecast densities, which can be accomplished by means of the skewed exponential power (SEP) distribution of Fernandez, Osiewalski, and Steel (1995), with density

$$f(z;\mu,\sigma,p,\theta) = \frac{\theta}{1+\theta^2}\frac{p}{\sigma 2^{1/p}\Gamma(1/p)}\begin{cases}\exp\left\{-\frac{1}{2}\left(\frac{|z-\mu|\theta}{\sigma}\right)^p\right\} & \text{if} \quad z < \mu \\ \exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma\theta}\right)^p\right\} & \text{if} \quad z \geq \mu,\end{cases} \tag{3.6}$$

where $\sigma, \theta, p > 0$. This distribution nests the normal for $\theta = 1$ and $p = 2$. For $\theta < 1 (\theta > 1)$, the density is skewed to the left (right), and is fat–tailed for $p < 2$. To test for correct specification of the first two moments, we fix θ and p at their Gaussian values (i.e., as in Berkowitz, 2001, we use the normal likelihood), and then test (3.3) for $\mu = 0$ and $\sigma^2 = 1$, respectively. Both likelihood ratio tests (LRTs) are approximately $\chi^2(1)$ distributed. To test for asymmetries and excess kurtosis, we use the full model (3.6) and test the restrictions $\theta = 1$ and $p = 2$, respectively, with the corresponding LRTs again having an asymptotic $\chi^2(1)$ distribution. We also report the results of a joint test for zero mean, unit variance and absence of skewness and excess kurtosis by testing the hypothesis that $\mu = 0$, $\sigma = 1$, $\theta = 1$, and $p = 2$.

---

arrived at by Veiga, 2009).

Finally, we consider the Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) distances, given by

$$\mathrm{KS} = \max_{z_t} |\widehat{F}(z_t) - \Phi(z_t)|, \text{ and } \quad \mathrm{AD} = \max_{z_t} \frac{|\widehat{F}(z_t) - \Phi(z_t)|}{\sqrt{\Phi(z_t)(1-\Phi(z_t))}}, \tag{3.7}$$

where $\widehat{F}$ is the empirical distribution function of the z_t in (3.3), and which emphasize the fit of the hypothesized distribution in the center and the tails, respectively. For further discussion and applications of these measures in finance, see, e.g., Mittnik and Paolella (2000), Malevergne and Sornette (2003), and Alexander (2008b, p. 128). To assess the significance of KS, we use its asymptotic distribution, whereas that of AD is evaluated by simulation.

The second type of evaluation is based on the Value–at–Risk (VaR) measure relevant for risk management, see, e.g., Christoffersen and Pelletier (2004). For a given model, the VaR at level ξ for period t, denoted by $\mathrm{VaR}_t(\xi)$, is implicitly defined by $\widehat{F}(\mathrm{VaR}_t(\xi)|\Psi_{t-1}) = \xi$. A *violation* or *hit* is said to occur at time t if $r_t < \mathrm{VaR}_t(\xi)$. To test the models' suitability for calculating accurate ex–ante VaR measures, we define the binary sequence

$$I_t = \begin{cases} 1, & \text{if} \quad r_t < \mathrm{VaR}_t \\ 0, & \text{if} \quad r_t \geq \mathrm{VaR}_t. \end{cases} \tag{3.8}$$

Then the empirical shortfall probability is $\widehat{\xi} = x/T$, where $x = \sum_{t=1}^{T} I_t$ is the number of observed violations. To assess whether the empirical shortfall probability, $\widehat{\xi}$, is statistically indistinguishable from the nominal shortfall probability, ξ, we use the likelihood ratio test (cf. Kupiec, 1995)

$$\mathrm{LRT}_{\mathrm{VaR}} = -2\{x\log(\xi/\widehat{\xi}) + (T-x)\log[(1-\xi)/(1-\widehat{\xi})]\} \overset{asy}{\sim} \chi^2(1). \tag{3.9}$$

Equation (3.9) represents a test for correct *unconditional* coverage of a VaR model, but it neglects possible dependencies in the series of hits. More elaborate tests accounting for potential *violation clustering* have been developed (e.g., Christoffersen and Pelletier, 2004), but testing unconditional coverage is effective for our objective to figure out whether the skew–normal's shape is sufficiently flexible to remove the deficiencies of the Gaussian MS GARCH process. We consider the nominal VaR levels $\xi = 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05$, and 0.1, and calculate the one–step–ahead out–of–sample VaR measures both for long and short positions of the respective indices.

3.3.2 Empirical results

The results of the tests described in Section 3.3.1 are reported in Tables 9–14. Tables 9, 10, and 11 provide the results for the entire predictive densities for the CAC 40, DAX 30, and FTSE 100, respectively, and Tables 12, 13, and 14 do the same for the VaR backtesting results. We note that, in this section, the SN MixGARCH processes are only considered in the version with zero component means, since letting the means become regime–dependent has been found to add little value. In Tables 12–14, for each VaR level, ξ, we show the empirical *percentage* shortfall probability $100 \times \widehat{\xi}$ of the respective models, as well as, as a summary measure of performance, the mean absolute percentage error (MAPE) of a model, given by

$$\text{MAPE} = \frac{1}{7} \sum_{i=1}^{7} \left| \frac{\widehat{\xi}_i - \xi_i}{\xi_i} \right|, \tag{3.10}$$

where $\{\xi_1, \ldots, \xi_7\} = \{0.001, \ldots, 0.1\}$, and which we calculate separately for long and short positions.

The results in Tables 9–14 are in accordance with the in–sample results in Section 3.2 in that the SN mixture models appear to exhibit the best fit overall. Obviously, as in Section 3.2, models based on symmetric conditional densities are inferior to their skewed counterparts both according to the predictive density tests in Tables 9–11 as well as the VaR backtests in Tables 12–14. Thus, we compare the SN mixture GARCH processes with the other skewed contestants, i.e., the single–regime SN and skewed t models as well as the asymmetric Gaussian mixture GARCH specifications.

The single–regime SN GARCH process considerably improves the fit as compared to the Gaussian GARCH. In particular, it produces residuals with no significant skewness for the CAC 40 and DAX 30, but it still suffers from excess kurtosis in each case. Its failure in the tails is also reflected in the rather large (and significant) value of the AD statistic defined in (3.7). Perhaps somewhat surprisingly, however, its VaR measures are rather accurate according to the Kupiec test (3.9), which shows that considerable improvements can be achieved solely by accounting for skewness (in contrast to both skewness and kurtosis).

In contrast to the single–regime SN GARCH process, the skewed Student's t model matches the kurtosis very well for all three series, but, although it improves upon its symmetric relative, it still fails to successfully capture the skewness. This is also reflected in the VaR backtesting results in Tables 12–14. Whereas the accuracy of the VaR measures for the left tail (long positions) implied by the skewed t is comparable with that of the VaRs of the SN mixture models for the CAC 40 and the DAX 30, the VaRs implied by the mixture models are more precise in the right tail in the sense that the skewed t model tends to overestimate the risk in the short positions.[13] This means that the negative skewness in the out–of–sample residuals

[13]For the (symmetric) t GARCH model, similar results have been reported in So and Yu (2006).

(3.3) of the skewed Student's t model can be explained by it overestimating the thickness of the right tail. The SN mixture models, on the other hand, are rather accurate in both tails, at least for the CAC 40 and DAX 30. All models' performance is somewhat worse for the FTSE 100, but the relative superiority of the SN mixtures for this series becomes clear in Table 11. They filter a much greater part of the skewness out of the data than the other models, so that only the members of this class pass (i) the Jarque–Bera test (3.5) and (ii) all the likelihood ratio tests based on the SEP distribution (3.6). We note, however, that the skewed t GARCH model is the only one that passes both the KS and AD distances tests (3.7) for all three series. The second best according to this criterion is the SN $\text{Mix}_S^{\text{icept}}$ GARCH, which only fails to pass the KS test for the DAX 30.

The comparison between SN and asymmetric Gaussian mixture GARCH processes is also in favor of the former. Similar to the skewed t specification, the Gaussian models are less capable of capturing the skewness in the CAC 40 and FTSE 100 returns, and their VaR measures likewise tend to be inferior to those of the SN mixtures, according to both the Kupiec test (3.9) and the MAPE (3.10).

Among the SN mixture models, we observe that, analogous to the in–sample results, adding complexity to the basic SN $\text{Mix}^{\text{icept}}$ specification via either regime–specific volatility dynamics or a more flexible mixing process does not improve the results. Although all the models of this class produce results of roughly similar quality, the basic model may even be viewed as exhibiting the most robust performance. For example, as compared to their MixGARCH counterparts, the SN Markov–switching GARCH models are slightly less efficient in capturing the kurtosis of the CAC 40 returns (c.f. Table 9), and SN $\text{Mix}^{\text{icept}}$ is the only mixture model that passes the tail–centered AD test for all series.

4 Conclusions

In this paper, we propose and motivate the use of Azzalini's (1985) skew–normal distribution in the framework of independent and Markovian mixture GARCH processes. The dynamic properties of the new process are derived, including the unconditional variance, skewness, kurtosis, and the autocorrelation structure of absolute and squared returns. An application to three major European stock market indices over a period covering the recent financial crisis shows that the approach fits the data well and delivers reliable out–of–sample risk measures.

The initial motivation for introducing this class of models was that distributional asymmetries are difficult to incorporate into the Gaussian MS GARCH model without affecting the autocorrelation properties of the raw returns in an undesirable manner. It turns out that the skew–normal MS GARCH process indeed helps to

Table 9: Forecast density evaluation: CAC 40

	Moment–based			Berkowitz LRT					Distances	
Model	Skew.	Kurt.	*JB*	μ	σ^2	θ	p	joint	KS	AD
Normal										
Single	−0.283***	3.60***	70.9***	−0.031	0.997	0.881***	1.76***	33.9***	0.018	0.800***
$\text{Mix}_S^{\text{icept}}$	−0.186***	3.04	14.4***	−0.037*	0.988	0.891***	1.99	18.6***	0.017	0.096*
Mix_S	−0.223***	3.18*	23.8***	−0.039*	0.994	0.882***	1.95	22.4***	0.016	0.281**
$\text{Mix}^{\text{icept}}$	−0.093*	3.02	3.6	−0.032	0.986	0.942**	2.00	6.8	0.014	0.096*
Mix	−0.096*	3.11	5.2*	−0.032	0.986	0.948*	1.95	6.3	0.015	0.168**
$\text{Markov}^{\text{icept}}$	−0.240***	3.27***	31.6***	−0.030	1.015	0.880***	1.89	24.4***	0.016	0.326***
Markov	−0.239***	3.28***	31.9***	−0.030	1.027	0.879***	1.87	25.5***	0.017	0.367***
Skew–normal										
Single	−0.068	3.40***	18.3***	−0.029	0.988	0.973	1.80**	9.7**	0.019	0.276**
$\text{Mix}^{\text{icept}}$	−0.049	3.03	1.1	−0.032	0.985	0.971	1.98	3.9	0.016	0.064
Mix	−0.065	3.10	2.7	−0.032	0.994	0.965	1.97	4.1	0.016	0.154**
$\text{Markov}^{\text{icept}}$	−0.059	3.18*	4.8*	−0.027	1.005	0.971	1.91	4.1	0.015	0.190**
Markov	−0.045	3.19*	4.6	−0.026	1.034	0.977	1.90	5.3	0.014	0.143**
Student's t										
symmetric	−0.193***	2.98	15.4***	−0.036*	0.988	0.883***	2.07	20.2***	0.015	0.110*
skewed	−0.117**	2.99	5.6*	−0.026	0.987	0.933**	2.02	7.3	0.014	0.081

Shown are the results of tests for correctly specified one–step–ahead predictive densities for the CAC 40. "Single" denotes a model with just one regime, i.e., $k = 1$ in (1.1)–(1.3). The abbreviations for the different multi–regime models are detailed in Table 1. All the tests are based on the residuals defined in (3.3), which are iid standard normal for a correctly specified model. "Skew." and "Kurt." are the sample skewness and kurtosis of these residuals, respectively, and *JB* is the Jarque–Bera test statistic defined in (3.5). Significance of these is assessed on the basis of the distributional results in (3.4) and (3.5).

The Berkowitz LRT tests are based on the SEP distribution defined in (3.6); the procedure is as follows: Tests for the first two moments are based on the Gaussian likelihood, i.e., we fix $\theta = 1$ and $p = 2$ in (3.6) and then test the hypotheses $\mu = 0$ and $\sigma = 1$, respectively, with both tests being approximately $\chi^2(1)$ distributed. Tests for $\theta = 1$ (symmetry) and $p = 2$ ("mesokurtosis") use the full model (3.6) and are also approximately $\chi^2(1)$. For all these tests, we report the estimate of the parameter of interest. Finally, "joint" refers to the joint test of $\mu = 1$, $\sigma = 1$, $\theta = 1$, and $p = 2$, which is approximately $\chi^2(4)$. Here we report the value of the test statistic. See Section 3.3.1 for further details. KD and AD are the Kolmogorov–Smirnov and Anderson–Darling distances, respectively, as defined in (3.7). Asterisks *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 10: Forecast density evaluation: DAX 30

	Moment–based			Berkowitz LRT					Distances	
Model	Skew.	Kurt.	JB	μ	σ^2	θ	p	joint	KS	AD
Normal										
Single	-0.275^{***}	3.50^{***}	56.6^{***}	-0.026	1.021	0.883^{***}	1.70^{***}	38.9^{***}	0.023	1.25^{***}
$\text{Mix}_S^{\text{icept}}$	-0.186^{***}	2.93	14.8^{***}	-0.040^{**}	1.017	0.890^{***}	1.97	21.0^{***}	0.024	0.096^{*}
Mix_S	-0.207^{***}	3.05	18.0^{***}	-0.041^{**}	1.024	0.886^{***}	1.94	23.2^{***}	0.024	0.685^{***}
$\text{Mix}^{\text{icept}}$	-0.073	2.96	2.4	-0.033	1.012	0.956	2.00	5.2	0.020	0.206^{**}
Mix	-0.069	2.92	2.6	-0.031	1.008	0.956	2.02	4.9	0.021	0.145^{**}
$\text{Markov}^{\text{icept}}$	-0.208^{***}	3.01	18.0^{***}	-0.037^{*}	1.024	0.885^{***}	1.92	23.6^{***}	0.024	0.117^{*}
Markov	-0.232^{***}	3.16^{*}	25.0^{***}	-0.038^{*}	1.030	0.881^{***}	1.86	28.1^{***}	0.022	0.633^{***}
Skew–normal										
Single	-0.012	3.29^{***}	9.0^{**}	-0.022	1.017	0.990	1.75^{***}	10.3^{**}	0.024	0.270^{**}
$\text{Mix}^{\text{icept}}$	0.034	2.92	1.1	-0.033	1.017	1.017	1.96	3.6	0.028^{**}	0.057
Mix	0.016	3.01	0.1	-0.034^{*}	1.019	1.009	1.93	4.1	0.026^{*}	0.140^{**}
$\text{Markov}^{\text{icept}}$	0.022	3.00	0.2	-0.028	1.021	1.013	1.92	3.5	0.025^{*}	0.069
Markov	0.022	3.08	0.9	-0.024	1.053^{*}	1.012	1.87	7.2	0.021	0.115^{*}
Student's t										
symmetric	-0.196^{***}	2.88	17.3^{***}	-0.039^{*}	1.011	0.879^{***}	2.09	23.2^{***}	0.024	0.117^{*}
skewed	-0.085^{*}	2.92	3.6	-0.017	1.012	0.953^{*}	1.94	4.3	0.018	0.075

Shown are the results of tests for correctly specified one–step–ahead predictive densities for the DAX 30. See the legend of Table 9 for further explanations.

Table 11: Forecast density evaluation: FTSE 100

	Moment–based			Berkowitz LRT					Distances	
Model	Skew.	Kurt.	JB	μ	σ^2	θ	p	joint	KS	AD
Normal										
Single	−0.353***	3.55***	83.3***	−0.025	1.015	0.853***	1.81**	42.5***	0.016	0.892***
$\text{Mix}_S^{\text{icept}}$	−0.255***	3.02	26.8***	−0.029	1.016	0.858***	2.01	28.4***	0.017	0.104*
Mix_S	−0.267***	3.04	29.6***	−0.029	1.011	0.852***	2.01	30.3***	0.017	0.106*
$\text{Mix}^{\text{icept}}$	−0.183***	2.98	13.9***	−0.024	1.012	0.893***	2.02	15.8***	0.015	0.092
Mix	−0.184***	2.97	14.1***	−0.024	1.011	0.892***	2.03	16.0***	0.015	0.090
$\text{Markov}^{\text{icept}}$	−0.257***	3.02	27.2***	−0.029	1.016	0.857***	2.01	28.7***	0.017	0.105*
Markov	−0.263***	3.03	28.8***	−0.029	1.011	0.853***	2.02	29.7***	0.017	0.105*
Skew–normal										
Single	−0.118**	3.35***	18.2***	−0.023	1.012	0.957	1.84**	8.5*	0.015	0.594***
$\text{Mix}^{\text{icept}}$	−0.087*	2.98	3.1	−0.025	1.015	0.954	2.01	4.2	0.012	0.069
Mix	−0.088*	2.96	3.3	−0.024	1.009	0.952	2.03	4.3	0.012	0.070
$\text{Markov}^{\text{icept}}$	−0.086*	2.97	3.2	−0.024	1.014	0.955	2.01	4.1	0.012	0.069
Markov	−0.081*	2.97	2.8	−0.024	1.009	0.955	2.03	3.9	0.012	0.065
Student's t										
symmetric	−0.261***	3.00	28.1***	−0.029	1.010	0.847***	2.10	30.6***	0.016	0.122*
skewed	−0.166***	3.00	11.4***	−0.014	1.010	0.914***	1.97	10.3**	0.014	0.079

Shown are the results of tests for correctly specified one–step–ahead predictive densities for the FTSE 100. See the legend of Table 9 for further explanations.

Table 12: Evaluation of Value–at–Risk (VaR) measures: CAC 40

Long positions								
$100 \times \xi$	0.1	0.25	0.5	1	2.5	5	10	MAPE
Normal								
Single	0.28**	0.56***	0.73	1.33	3.39***	6.01**	10.9	0.645
$\text{Mix}_S^{\text{icept}}$	0.20	0.32	0.56	1.25	3.31**	6.29***	11.2*	0.340
Mix_S	0.20	0.36	0.56	1.21	3.35**	6.25***	11.1*	0.358
$\text{Mix}^{\text{icept}}$	0.16	0.24	0.48	0.85	2.82	5.73	11.0*	0.172
Mix	0.16	0.28	0.44	0.89	2.94	5.73	10.8	0.196
$\text{Markov}^{\text{icept}}$	0.20	0.40	0.81**	1.41*	3.31**	6.49***	11.1*	0.484
Markov	0.24*	0.48**	0.81**	1.49**	3.47***	6.53***	11.5**	0.614
Skew–normal								
Single	0.24*	0.36	0.52	0.97	2.70	5.44	10.4	0.309
$\text{Mix}^{\text{icept}}$	0.20	0.28	0.48	0.85	2.66	5.56	10.9	0.228
Mix	0.16	0.24	0.48	0.93	2.66	5.65	10.9	0.147
$\text{Markov}^{\text{icept}}$	0.16	0.36	0.56	1.01	2.78	5.77*	10.8	0.221
Markov	0.20	0.40	0.60	1.01	2.94	6.05**	11.1*	0.335
Student's t								
symmetric	0.24*	0.32	0.52	1.01	3.19**	6.13**	11.2*	0.340
skewed	0.20	0.28	0.48	0.97	2.98	5.81*	10.9	0.236
Short positions								
$100 \times \xi$	0.1	0.25	0.5	1	2.5	5	10	MAE
Normal								
Single	0.04	0.08**	0.24**	0.73	1.90**	3.79***	8.39***	0.387
$\text{Mix}_S^{\text{icept}}$	0.04	0.04***	0.20**	0.52***	1.90**	3.71***	8.63**	0.449
Mix_S	0.04	0.04***	0.20**	0.60**	1.81**	3.63***	8.67**	0.444
$\text{Mix}^{\text{icept}}$	0.04	0.12	0.24**	0.69*	2.10	4.07**	9.03	0.341
Mix	0.04	0.12	0.28*	0.85	2.26	3.87***	9.15	0.301
$\text{Markov}^{\text{icept}}$	0.04	0.04***	0.20**	0.73	1.94*	3.87***	8.99*	0.408
Markov	0.04	0.04***	0.20**	0.73	1.94*	4.03**	8.91*	0.405
Skew–normal								
Single	0.04	0.20	0.44	1.05	2.46	4.23*	8.91*	0.176
$\text{Mix}^{\text{icept}}$	0.04	0.16	0.24**	0.89	2.38	4.40	9.19	0.262
Mix	0.04	0.16	0.28*	0.85	2.26	4.35	9.31	0.262
$\text{Markov}^{\text{icept}}$	0.04	0.12	0.44	1.01	2.42	4.60	9.03	0.206
Markov	0.04	0.24	0.52	1.09	2.66	4.80	9.44	0.132
Student's t								
symmetric	0.04	0.04***	0.20**	0.48***	1.61***	3.63***	8.71**	0.472
skewed	0.04	0.04***	0.24**	0.85	2.06	4.15**	9.35	0.359

Shown are the results of the tests for the correct coverage of out–of–sample Value–at–Risk measures for the CAC 40. Shown are the results of tests for correctly specified one–step–ahead predictive densities for the CAC 40. "Single" denotes a model with just one regime, i.e., $k = 1$ in (1.1)–(1.3). The abbreviations for the different multi–regime models are detailed in Table 1. Reported are the percentage empirical shortfall probabilities $100 \times \widehat{\xi} = x/T$ observed for a nominal VaR level ξ, where x is the empirical shortfall frequency, and T is the number of forecasts evaluated. Asterisks *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively, as obtained from the likelihood ratio test (3.9). MAPE is the mean absolute percentage error, as defined in (3.10).

Table 13: Evaluation of Value–at–Risk (VaR) measures: DAX 30

Long positions								
$100 \times \xi$	0.1	0.25	0.5	1	2.5	5	10	MAPE
Normal								
Single	0.40***	0.65***	0.81**	1.37*	3.51***	6.49***	11.4**	0.919
$\text{Mix}_S^{\text{icept}}$	0.20	0.28	0.69	1.17	3.51***	6.94***	12.3***	0.386
Mix_S	0.24*	0.28	0.60	1.17	3.59***	6.98***	12.3***	0.427
$\text{Mix}^{\text{icept}}$	0.16	0.20	0.28*	0.73	2.70	5.97**	12.0***	0.284
Mix	0.20	0.20	0.28*	0.85	2.82	5.93**	11.9***	0.329
$\text{Markov}^{\text{icept}}$	0.20	0.48**	0.81**	1.33	3.55***	6.94***	12.1***	0.559
Markov	0.20	0.48**	0.89**	1.37*	3.71***	6.85***	12.1***	0.594
Skew–normal								
Single	0.16	0.40	0.65	0.97	2.58	5.56	11.0	0.256
$\text{Mix}^{\text{icept}}$	0.08	0.24	0.52	0.81	2.34	5.81*	11.3**	0.118
Mix	0.16	0.20	0.44	0.85	2.38	5.77*	11.3**	0.200
$\text{Markov}^{\text{icept}}$	0.12	0.32	0.48	0.89	2.42	5.73	11.4**	0.138
Markov	0.20	0.44*	0.48	0.93	2.50	5.93**	11.6***	0.320
Student's t								
symmetric	0.12	0.40	0.69	1.01	3.19**	6.77***	12.1***	0.291
skewed	0.08	0.36	0.52	0.89	2.70	6.29***	11.5**	0.184
Short positions								
$100 \times \xi$	0.1	0.25	0.5	1	2.5	5	10	MAE
Normal								
Single	0.08	0.20	0.28*	0.60**	1.90**	4.48	9.27	0.234
$\text{Mix}_S^{\text{icept}}$	0.00**	0.12	0.20**	0.44***	1.77**	4.48	9.40	0.446
Mix_S	0.04	0.16	0.20**	0.44***	1.77**	4.48	9.44	0.365
$\text{Mix}^{\text{icept}}$	0.04	0.20	0.28*	0.56**	2.14	4.96	9.84	0.262
Mix	0.04	0.16	0.24**	0.52***	2.14	5.08	9.96	0.301
$\text{Markov}^{\text{icept}}$	0.00**	0.16	0.24**	0.48***	1.81**	4.56	9.48	0.400
Markov	0.04	0.16	0.32	0.48***	1.90**	4.48	9.19	0.321
Skew–normal								
Single	0.20	0.28	0.48	0.97	2.66	5.08	9.96	0.185
$\text{Mix}^{\text{icept}}$	0.08	0.20	0.36	0.81	2.74	5.36	10.3	0.150
Mix	0.12	0.16	0.36	0.85	2.78	5.32	10.2	0.169
$\text{Markov}^{\text{icept}}$	0.12	0.24	0.28*	1.01	2.98	5.28	10.4	0.139
Markov	0.16	0.24	0.44	1.25	3.19**	5.65	10.6	0.210
Student's t								
symmetric	0.00**	0.04***	0.20**	0.36***	1.53***	4.23*	9.52	0.523
skewed	0.00**	0.16	0.28*	0.56**	2.30	5.12	10.4	0.338

Shown are the results of the tests for the correct coverage of out–of–sample Value–at–Risk measures for the DAX 30. See the legend of Table 12 for further explanations.

Table 14: Evaluation of Value–at–Risk (VaR) measures: FTSE 100

Long positions $100\times\xi$	0.1	0.25	0.5	1	2.5	5	10	MAPE
Normal								
Single	0.32***	0.69***	1.13***	1.65***	3.87***	6.25***	10.4	0.960
$\text{Mix}_S^{\text{icept}}$	0.28**	0.40	0.81**	1.69***	4.07***	6.45***	11.1*	0.682
Mix_S	0.24*	0.52**	0.85**	1.53**	4.11***	6.41***	11.1*	0.683
$\text{Mix}^{\text{icept}}$	0.16	0.36	0.60	1.37*	3.71***	6.25***	10.8	0.351
Mix	0.24*	0.28	0.56	1.33	3.71***	6.29***	10.9	0.406
$\text{Markov}^{\text{icept}}$	0.32***	0.48**	0.81**	1.69***	3.99***	6.45***	11.1*	0.781
Markov	0.24*	0.52**	0.85**	1.45**	3.99***	6.37***	11.1*	0.662
Skew–normal								
Single	0.24*	0.40	0.65	1.33	2.94	5.77*	10.2	0.429
$\text{Mix}^{\text{icept}}$	0.16	0.36	0.56	1.25	3.23**	5.93**	10.6	0.283
Mix	0.16	0.32	0.52	1.17	3.27**	5.85*	10.6	0.237
$\text{Markov}^{\text{icept}}$	0.16	0.36	0.56	1.21	3.19**	5.89**	10.7	0.274
Markov	0.16	0.28	0.52	1.17	3.15**	5.85*	10.6	0.206
Student's *t*								
symmetric	0.24*	0.32	0.73	1.49**	3.79***	6.25***	11.1*	0.504
skewed	0.24*	0.28	0.69	1.25	3.55***	6.21***	10.6	0.412
Short positions $100\times\xi$	**0.1**	**0.25**	**0.5**	**1**	**2.5**	**5**	**10**	**MAE**
Normal								
Single	0.04	0.12	0.16***	0.40***	1.37***	3.99**	9.03	0.448
$\text{Mix}_S^{\text{icept}}$	0.04	0.08**	0.12***	0.44***	1.41***	4.07**	9.31	0.468
Mix_S	0.04	0.04***	0.12***	0.40***	1.29***	4.03**	9.27	0.506
$\text{Mix}^{\text{icept}}$	0.08	0.12	0.20**	0.48***	1.53***	4.44	9.68	0.336
Mix	0.08	0.08**	0.16***	0.48***	1.53***	4.48	9.72	0.369
$\text{Markov}^{\text{icept}}$	0.04	0.08**	0.12***	0.44***	1.41***	4.11**	9.31	0.467
Markov	0.04	0.04***	0.08***	0.40***	1.33***	4.11**	9.35	0.512
Skew–normal								
Single	0.12	0.16	0.36	0.81	2.14	4.96	9.72	0.173
$\text{Mix}^{\text{icept}}$	0.04	0.12	0.24**	0.65*	2.14	5.00	9.96	0.305
Mix	0.04	0.16	0.24**	0.65*	1.94*	5.00	10.0	0.293
$\text{Markov}^{\text{icept}}$	0.04	0.12	0.28*	0.65*	2.14	5.00	9.96	0.293
Markov	0.04	0.16	0.24**	0.69*	1.94*	5.00	10.0	0.287
Student's *t*								
symmetric	0.04	0.08**	0.12***	0.32***	1.25***	4.03**	9.48	0.494
skewed	0.04	0.12	0.16***	0.56**	1.98*	4.92	9.68	0.355

Shown are the results of the tests for the correct coverage of out–of–sample Value–at–Risk measures for the FTSE 100. See the legend of Table 12 for further explanations.

cure the weaknesses of the Gaussian MS GARCH model; however, we also find that parsimoniously parameterized specifications with regime–independent volatility dynamics (i.e., a switching intercept only) and an independent mixing process tend to be preferred by the data. The in– and out–of–sample comparisons also showed that the skew–normal mixtures tend to be superior to both a single–regime skewed t GARCH model as well as to the normal mixture GARCH specification with regime–specific means.

Future work will consider the use of the skew–normal family in multivariate MS–GARCH models. When such models are applied to time series of higher dimension, optimization of the likelihood will become burdensome, however, and the development of Bayesian estimation routines may be called for.

Appendix

A Odd absolute moments of a centered skew–normal random variable

As the SN MS GARCH process defined in Section 2.2 is built on centered SN innovations, the absolute moments of such variables are used for deriving the moment structure of the process in Appendix B. In this appendix we therefore calculate the odd absolute moments of a centered skew–normal random variable Z, i.e., $Z = X - \sqrt{2/\pi}\delta$, where $\delta = \gamma/\sqrt{1+\gamma^2}$ and X has the density defined in (2.1).

Let $i \in \mathbb{N}$ be odd. Then we have, with $m = \sqrt{2/\pi}\delta$,

$$\begin{aligned} \mathrm{E}(|Z|^i) &= -\mathrm{E}(Z^i) + 4\int_0^\infty z^i \phi(z+m)\Phi(\gamma(z+m))dz \\ &= -\mathrm{E}(Z^i) + 4\int_m^\infty (x-m)^i \phi(x)\Phi(\gamma x)dx \\ &= -\mathrm{E}(Z^i) + 2\sum_{\ell=0}^{i} \binom{i}{\ell} (-m)^{i-\ell} E_\ell(\gamma), \end{aligned} \tag{A.1}$$

where

$$E_\ell(\gamma) = 2\int_m^\infty x^\ell \phi(x)\Phi(\gamma x)dx.$$

Using $\phi'(x) = -x\phi(x)$, we observe that

$$\begin{aligned}
E_\ell(\gamma) &= 2\int_m^\infty x\phi(x)x^{\ell-1}\Phi(\gamma x)dx \\
&= -2\phi(x)x^{\ell-1}\Phi(\gamma x)\Big|_m^\infty + 2\int_m^\infty [(\ell-1)x^{\ell-2}\Phi(\gamma x) + x^{\ell-1}\gamma\phi(\gamma x)]\phi(x)dx \\
&= m^{\ell-1}f(m;\gamma) + (\ell-1)E^+_{\ell-2}(\gamma) + M_{\ell-1}(\gamma), \qquad \text{(A.2)}
\end{aligned}$$

where

$$\begin{aligned}
M_\ell(\gamma) &= 2\gamma\int_m^\infty x^\ell\phi(x)\phi(\gamma x)dx = \sqrt{\frac{2}{\pi}}\frac{\gamma}{\sqrt{2\pi}}\int_m^\infty x^\ell \exp\left\{-\frac{x^2}{2}(1+\gamma^2)\right\}dx \\
&= \sqrt{\frac{2}{\pi}}\frac{\gamma}{(1+\gamma^2)^{(\ell+1)/2}}\int_{\tilde{m}}^\infty x^\ell\phi(x)dx \\
&= \sqrt{\frac{2}{\pi}}\frac{\gamma}{(1+\gamma^2)^{(\ell+1)/2}}\tilde{M}_\ell(\gamma), \qquad \text{(A.3)}
\end{aligned}$$

with $\tilde{m} = \sqrt{2/\pi}\gamma$, and

$$\begin{aligned}
\tilde{M}_\ell(\gamma) &= \int_{\tilde{m}}^\infty x^\ell\phi(x)dx \\
&= -x^{\ell-1}\phi(x)\Big|_{\tilde{m}}^\infty + (\ell-1)\int_{\tilde{m}}^\infty x^{\ell-2}\phi(x)dx \\
&= \tilde{m}^{\ell-1}\phi(\tilde{m}) + (\ell-1)\tilde{M}_{\ell-2}(\gamma), \quad \ell \geq 2. \qquad \text{(A.4)}
\end{aligned}$$

The starting values in (A.4) are

$$\tilde{M}_0(\gamma) = \int_{\tilde{m}}^\infty \phi(x)dx = \Phi(-\tilde{m}),$$

so that $M_0(\gamma) = \sqrt{2/\pi}\gamma/\sqrt{1+\gamma^2}\tilde{M}_0(\gamma) = m\Phi(-\tilde{m})$, and

$$\tilde{M}_1(\gamma) = \int_{\tilde{m}}^\infty x\phi(x)dx = \phi(\tilde{m}),$$

so that $M_1(\gamma) = m/\sqrt{1+\gamma^2}\phi(\tilde{m})$, and it then follows, for example,

$$M_2(\gamma) = \sqrt{\frac{2}{\pi}}\frac{\gamma}{(1+\gamma^2)^{3/2}}\tilde{M}_2(\gamma) = \frac{m}{1+\gamma^2}[\tilde{m}\phi(\tilde{m}) + \Phi(-\tilde{m})].$$

The starting values in (A.2) are

$$\begin{aligned}
E_0(\gamma) &= 1 - F(m;\gamma), \\
E_1(\gamma) &= f(m;\gamma) + M_0(\gamma) = f(m;\gamma) + m\Phi(-\tilde{m}),
\end{aligned}$$

and then

$$\begin{aligned} E_2(\gamma) &= mf(m;\gamma)+1-F(m;\gamma)+\frac{m}{\sqrt{1+\gamma^2}}\phi(\tilde{m}), \\ E_3(\gamma) &= m^2 f(m;\gamma)+2E_1(\gamma)+\frac{m}{1+\gamma^2}[\tilde{m}\phi(\tilde{m})+\Phi(-\tilde{m})]. \end{aligned}$$

For the moments relevant in our context, we finally get, after a few calculations,

$$\mathrm{E}(|Z|) = 2\left\{f(m;\gamma)+m[F(m;\gamma)-\Phi(\tilde{m})]\right\}, \tag{A.5}$$

and

$$\begin{aligned} \mathrm{E}(|Z|^3) &= 2(m^2+2)f(m;\gamma)+(3m+m^3)(2F(m;\gamma)-1) \\ &\quad +\left(2m+\frac{m}{1+\gamma^2}+3m^2\right)[\Phi(-\tilde{m})-\Phi(\tilde{m})] \qquad \text{(A.6)} \\ &\quad -\frac{4m^2}{\sqrt{1+\gamma^2}}\phi(\tilde{m}). \qquad \text{(A.7)} \end{aligned}$$

In case of a Gaussian distribution, where $\gamma=0$, we have, of course, $\mathrm{E}(|Z|) = 2f(m;\gamma) = 2\phi(0) = \sqrt{2/\pi}$, and $\mathrm{E}(|Z|^3) = 2\sqrt{2/\pi}$. The absolute moments for i even easily follow from those of the uncentered skew–normal distribution, and up to the fourth order are given by

$$\mathrm{E}(Z^2) = 1-\frac{2}{\pi}\delta^2, \quad \text{and} \quad \mathrm{E}(Z^4) = 3-\frac{12}{\pi}\delta^2+\frac{8\pi-12}{\pi^2}\delta^4.$$

We also need expressions of the form

$$\kappa_i(\gamma) := \mathrm{E}[(|Z|-\lambda Z)^i] = (1-\lambda)^i\int_0^\infty z^i f(z;\gamma)dz+(1+\lambda)^i\int_{-\infty}^0 |z|^i f(z;\gamma)dz. \tag{A.8}$$

For i odd, we have

$$\begin{aligned} \kappa_i(\gamma) &= (1-\lambda)^i\int_0^\infty z^i f(z;\gamma)dz+(1+\lambda)^i\int_{-\infty}^0 |z|^i f(z;\gamma)dz \\ &= \frac{(1-\lambda)^i+(1+\lambda)^i}{2}\mathrm{E}(|Z|^i)+\frac{(1-\lambda)^i-(1+\lambda)^i}{2} \\ &\quad \times\left\{\int_0^\infty z^i f(z;\gamma)dz-\int_{-\infty}^0 |z|^i f(z;\gamma)dz\right\} \\ &= \frac{(1-\lambda)^i+(1+\lambda)^i}{2}\mathrm{E}(|Z|^i)+\frac{(1-\lambda)^i-(1+\lambda)^i}{2}\mathrm{E}(Z^i), \end{aligned}$$

and for i even,

$$\kappa_i(\gamma) = (1+\lambda)^i\mathrm{E}(|Z|^i)+[(1-\lambda)^i-(1+\lambda)^i]\sum_{\ell=0}^{i}\binom{i}{\ell}(-m)^{i-\ell}E_\ell(\gamma).$$

B Moment structure of the skew–normal Markov–switching GARCH process

In this appendix, we state the conditions for the process (1.1)–(1.3) and (2.10) to have a stationary solution with finite moments and provide explicit expressions for these moments in case of existence. To this end, we define the matrices

$$X_t = \begin{pmatrix} \sigma_{1t} \\ \sigma_{2t} \\ \vdots \\ \sigma_{kt} \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_k \end{pmatrix}, \tag{B.1}$$

and let e_j be the jth unit vector in $\mathbb{R}^k$, i.e., the jth column of I_k, denoting the identity matrix of dimension k. Then the process can be written as a first–order stochastic vector difference equation,

$$X_t = \omega + C_{\Delta_{t-1},t-1} X_{t-1}, \tag{B.2}$$

where $C_{\Delta_{t-1},t-1} = \alpha e'_{\Delta_{t-1}}(|\eta_{t-1}| - \lambda \eta_{t-1}) + \beta$ is the stochastic coefficient matrix.

To state the condition for the process to have a stationary solution with a finite fourth moment (and hence kurtosis), we adopt the following notation due to Francq and Zakoïan (2002, 2005). For any function $f : S \mapsto M_{n\times n'}(\mathbb{R})$, where $M_{n\times n'}(\mathbb{R})$ is the space of real $n \times n'$ matrices, and $S = \{1, \ldots, k\}$ is the state space of $\{\Delta_t\}$, set

$$\mathbb{P}_f = \begin{pmatrix} p_{11} f(1) & \cdots & p_{k1} f(k) \\ \vdots & \cdots & \vdots \\ p_{1k} f(1) & \cdots & p_{kk} f(k) \end{pmatrix}. \tag{B.3}$$

Moreover, let

$$C_{mm}(j) = \mathrm{E}(C_{jt}^{\otimes m}), \quad j = 1, \ldots, k, \tag{B.4}$$

where $A^{\otimes m}$ denotes the mth Kronecker power of matrix A, i.e., the expression $A \otimes A \otimes \cdots \otimes A$ with m product terms, and let $\rho(A)$ denote the spectral radius of a square matrix A, i.e.,

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

Then the following is an application of results in Liu (2007).

Theorem 1 *(Liu, 2007, Theorem 3.2) If $\rho(\mathbb{P}_{C_{mm}}) < 1$, where $\mathbb{P}_{C_{mm}}$ is defined via (B.3) and (B.4), then the SN–MS GARCH(k) process defined by (1.1)–(1.3) and (2.10) has a unique strictly stationary solution with finite mth moment.*

Liu (2007, Theorem 2.2) also provides a weaker necessary and sufficient condition for strict stationarity which is, however, less practicable. Usually we expect that at least a few moments of financial return distributions are finite, so that the condition in Theorem 1 will hold for some $m \in \mathbb{N}$, and there is no need to check the less manageable condition in Theorem 2.2 of Liu (2009).[14]

To check the conditions provided by Theorem 1, we have to calculate the quantities $C_{mm}(j)$ defined in (B.4) for $m = 1, \ldots, 4$, which is straightforward, although tedious. For example,

$$\begin{aligned} C_{11}(j) &= \alpha e_j' \kappa_1(\gamma) + \beta, \\ C_{22}(j) &= (\alpha \otimes \alpha)(e_j' \otimes e_j')\kappa_2(\gamma) + (\alpha e_j' \otimes \beta + \beta \otimes \alpha e_j')\kappa_1(\gamma) + \beta \otimes \beta, \\ j &= 1, \ldots, k. \end{aligned}$$

where $\kappa_i(\gamma)$ is defined in (A.8).

If the eigenvalue condition of Theorem 1 holds for $m = 4$, we can explicitly calculate the fourth–moment structure of the process, where we employ the following lemma of Francq and Zakoïan (2005, 2008).

Lemma 1 *(Francq and Zakoïan, 2005, Lemma 3) For $\ell \geq 1$, if the variable $Y_{t-\ell}$ belongs to the information set generated by $\Psi_{t-\ell} := \{\varepsilon_s : s \leq t - \ell\}$, then*

$$\pi_{j,\infty} E(Y_{t-\ell} | \Delta_t = j) = \sum_{i=1}^{k} \pi_{i,\infty} p_{ij}^{(\ell)} E(Y_{t-\ell} | \Delta_{t-\ell} = i),$$

where the $p_{ij}^{(\ell)} := p(\Delta_t = j | \Delta_{t-\ell} = i)$, $i, j = 1, \ldots, k$, denote the ℓ–step transition probabilities, as given by the elements of P^ℓ

Using Lemma 1, we have

$$\pi_{j,\infty} \mathrm{E}(X_t | \Delta_t = j) = \pi_{j,\infty} \omega + \sum_{i=1}^{k} p_{ij} C_{11}(i) \pi_{i,\infty} \mathrm{E}(X_{t-1} | \Delta_{t-1} = i), \quad j = 1, \ldots, k,$$

so that

$$Q_1 = \pi_\infty \otimes \omega + \mathbb{P}_{C_{11}} Q_1,$$

where the $k^2 \times 1$ vector Q_1 is defined as

$$Q_1 = (\pi_{1,\infty} \mathrm{E}(X_t | \Delta_t = 1)', \ldots, \pi_{k,\infty} \mathrm{E}(X_t | \Delta_t = k)')'. \tag{B.5}$$

[14] In related work, Liu (2006, 2009) investigates the relation between strict and weak stationarity

Similarly, we obtain

$$Q_2 = \pi_\infty \otimes \omega \otimes \omega + \mathbb{P}_{C_{21}} Q_1 + \mathbb{P}_{C_{22}} Q_2,$$

where Q_2 is the $k^3 \times 1$ vector similar to Q_1 but with vector elements $\pi_{j\,\infty}\mathrm{E}[\mathrm{vec}(X_t X_t')|\Delta_t = j]$ instead of $\pi_{j,\infty}\mathrm{E}(X_t|\Delta_t = j)$, $j = 1,\ldots,k$, and $\mathbb{P}_{C_{21}}$ is obtained by applying definition (B.3) to

$$C_{21}(j) = \omega \otimes C_{11}(j) + C_{11}(j) \otimes \omega, \quad j = 1,\ldots,k. \tag{B.6}$$

In the same manner, by expanding $X_t^{\otimes m} = (\omega + C_{\Delta_{t-1},t-1}X_{t-1})^{\otimes m}$ in Kronecker powers of X_t, we get Q_3 and Q_4, defined by substituting $\pi_{j,\infty}\mathrm{E}\{\mathrm{vec}[\mathrm{vec}(X_t X_t')X_t']|\Delta_t = j\}$ and $\pi_{j,\infty}\mathrm{E}\{\mathrm{vec}[\mathrm{vec}(X_t X_t')\mathrm{vec}(X_t X_t')']|\Delta_t = j\}$ for $\pi_{j,\infty}\mathrm{E}(X_t|\Delta_t = j)$ in Q_1, respectively; the details are lengthy and tedious, however, and are available from the author on request. The moments of interest are then given by

$$\begin{aligned}
\mathrm{E}(|\varepsilon_t|^m) &= \sum_{j=1}^{k} \pi_{j,\infty}\mathrm{E}(|\varepsilon_t|^m|\Delta_t = j) = \mathrm{E}(|\eta_t|^m)\sum_{j=1}^{k}\pi_{j,\infty}\mathrm{E}(\sigma_{jt}^m|\Delta_t = j) \\
&= \mathrm{E}(|\eta_t|^m)\left(\sum_{j=1}^{k} e_j^{\otimes(m+1)}\right)' Q_m, \quad m = 1,\ldots,4,
\end{aligned}$$

and the unconditional *raw* (as opposed to absolute) third moment of the process is

$$\mathrm{E}(\varepsilon_t^3) = \mathrm{E}(\eta_t^3)\left(\sum_{j=1}^{k} e_j^{\otimes 4}\right)' Q_3 = \sqrt{\frac{2}{\pi}}\frac{4-\pi}{\pi}\delta^3\left(\sum_{j=1}^{k} e_j^{\otimes 4}\right)' Q_3, \tag{B.7}$$

and the unconditional skewness follows.

B.1 Autocorrelation structure

To derive the autocorrelation structure of the absolute process, we observe that

$$\begin{aligned}
\mathrm{E}(|\varepsilon_t||\varepsilon_{t-\tau}|) &= \mathrm{E}(|\eta_t|)\mathrm{E}(\sigma_{\Delta_t,t}\sigma_{\Delta_{t-\tau},t-\tau}|\eta_{t-\tau}|) \\
&= \mathrm{E}(|\eta_t|)\sum_{i=1}^{k}\sum_{j=1}^{k} p(\Delta_{t-\tau} = i \cap \Delta_t = j) \\
&\quad \times \mathrm{E}\left(\sigma_{jt}\sigma_{i,t-\tau}|\eta_{t-\tau}||\Delta_{t-\tau} = i \cap \Delta_t = j\right) \\
&= \mathrm{E}(|\eta_t|)\sum_{i=1}^{k}\sum_{j=1}^{k} \pi_{i,\infty}p_{ij}^{(\tau)}\mathrm{E}\left(\sigma_{jt}\sigma_{i,t-\tau}|\eta_{t-\tau}||\Delta_{t-\tau} = i \cap \Delta_t = j\right).
\end{aligned}$$

of (different versions of) MS GARCH processes and shows that their integrated versions are strictly

To calculate $\mathrm{E}\left(\sigma_{jt}\sigma_{i,t-\tau}|\eta_{t-\tau}||\Delta_{t-\tau}=i\cap\Delta_t=j\right)$, $i,j=1,\ldots,k$, $\tau\geq 2$, we evaluate

$$\begin{aligned}
S_{ij}(\tau) &:= \pi_{i,\infty}p_{ij}^{(\tau)}\mathrm{E}(X_tX'_{t-\tau}|\eta_{t-\tau}||\Delta_{t-\tau}=i\cap\Delta_t=j)\\
&= \pi_{i,\infty}p_{ij}^{(\tau)}\mathrm{E}(|\eta_t|)\omega\mathrm{E}(X'_{t-\tau}|\Delta_{t-\tau}=i)\\
&\quad +\pi_{i,\infty}p_{ij}^{(\tau)}\mathrm{E}(C_{\Delta_{t-1},t-1}X_{t-1}X'_{t-\tau}|\eta_{t-\tau}||\Delta_{t-\tau}=i\cap\Delta_t=j)\\
&= \pi_{i,\infty}p_{ij}^{(\tau)}\mathrm{E}(|\eta_t|)\omega\mathrm{E}(X'_{t-\tau}|\Delta_{t-\tau}=i)\\
&\quad +\pi_{i,\infty}p_{ij}^{(\tau)}\sum_{\ell=1}^{k}p(\Delta_{t-1}=\ell|\Delta_{t-\tau}=i\cap\Delta_t=j)\\
&\qquad \times\mathrm{E}(C_{\Delta_{t-1},t-1}X_{t-1}X'_{t-\tau}|\eta_{t-\tau}||\Delta_{t-\tau}=i\cap\Delta_{t-1}=\ell\cap\Delta_t=j)\\
&= \pi_{i,\infty}p_{ij}^{(\tau)}\mathrm{E}(|\eta_t|)\omega\mathrm{E}(X'_{t-\tau}|\Delta_{t-\tau}=i)\\
&\quad +\sum_{\ell=1}^{k}\pi_{i,\infty}p_{i\ell}^{(\tau-1)}p_{\ell j}C_{11}(\ell)\mathrm{E}(X_{t-1}X'_{t-\tau}|\eta_{t-\tau}||\Delta_{t-\tau}=i\cap\Delta_{t-1}=\ell)\\
&= \pi_{i,\infty}p_{ij}^{(\tau)}\mathrm{E}(|\eta_t|)\omega\mathrm{E}(X'_{t-\tau}|\Delta_{t-\tau}=i)+\sum_{\ell=1}^{k}p_{\ell j}C_{11}(\ell)S_{i\ell}(\tau-1), \quad i,j=1,\ldots,k.
\end{aligned}$$

Thus,

$$S(\tau)=\mathrm{E}(|\eta_t|)(P^{\tau}\otimes\omega)\tilde{Q}_1+\mathbb{P}_{C_{11}}S(\tau-1), \quad \tau\geq 2,$$

where

$$S(\tau)=\begin{pmatrix} S_{11}(\tau) & \cdots & S_{k1}(\tau)\\ \vdots & \cdots & \vdots\\ S_{1k}(\tau) & \cdots & S_{kk}(\tau)\end{pmatrix}, \quad \tilde{Q}_1=\begin{pmatrix}\pi_{1,\infty}\mathrm{E}(X'_t|\Delta_t=1) & \cdots & 0_{1\times k}\\ \vdots & \ddots & \vdots\\ 0_{1\times k} & \cdots & \pi_{k,\infty}\mathrm{E}(X'_t|\Delta_t=k)\end{pmatrix}.$$

For $\tau=1$, we find

$$S(1)=\mathrm{E}(|\eta_t|)(P\otimes\omega)\tilde{Q}_1+\mathbb{P}_{\tilde{C}_{11}}\tilde{Q}_2,$$

where $\mathbb{P}_{\tilde{C}_{11}}$ is defined by (B.3) along with $\tilde{C}_{11}(j)=\mathrm{E}(C_{jt}|\eta_t|)$, $j=1,\ldots,k$, and $\tilde{Q}_2$ is similar to $\tilde{Q}_1$ with $\pi_{j,\infty}\mathrm{E}(X'_t|\Delta_t=j)$ replaced with $\pi_{j,\infty}\mathrm{E}(X_tX'_t|\Delta_t=j)$, $j=1,\ldots,k$. Finally,

$$\mathrm{E}(|\varepsilon_t||\varepsilon_{t-\tau}|)=\mathrm{E}(|\eta_t|)\sum_{i=1}^{k}\sum_{j=1}^{k}e'_jS_{ij}(\tau)e_i=\mathrm{E}(|\eta_t|)\mathrm{vec}(I_k)'S(\tau)\mathrm{vec}(I_k), \quad \tau\geq 1.$$

---

stationary with infinite variance, which parallels Nelson's (1990) result for the IGARCH(1,1).

To find the autocorrelations of the squared process, we can define

$$\begin{aligned} Y_t &= \begin{pmatrix} X_t \\ \mathrm{vec}(X_t X_t') \end{pmatrix}, \quad \breve{\omega} = \begin{pmatrix} \omega \\ \omega \otimes \omega \end{pmatrix}, \\ \breve{C}_{jt} &= \begin{pmatrix} C_{jt} & 0_{k \times k^2} \\ \omega \otimes C_{jt} + C_{jt} \otimes \omega & C_{jt} \otimes C_{jt} \end{pmatrix}, \quad j = 1, \ldots, k, \end{aligned}$$

so that $Y_t = \breve{\omega} + \breve{C}_{\Delta_{t-1},t-1} Y_{t-1}$. Then we can apply the same reasoning as above to calculate

$$\mathrm{E}(\varepsilon_t^2 \varepsilon_{t-\tau}^2) = \mathrm{E}(\eta_t^2)\mathrm{E}(\sigma_{\Delta_t,t}^2 \sigma_{\Delta_{t-\tau},t-\tau}^2 \eta_{t-\tau}^2)$$

via evaluation of $\pi_{i,\infty} p_{ij}^{(\tau)} \mathrm{E}(Y_t Y_{t-\tau}' \eta_{t-\tau}^2 | \Delta_{t-\tau} = i \cap \Delta_t = j)$, $i, j = 1, \ldots, k$. The autocorrelations are given by

$$\begin{aligned} \rho_1(\tau) &= \mathrm{Corr}(|\varepsilon_t|, |\varepsilon_{t-\tau}|) = \frac{\mathrm{Cov}(|\varepsilon_t|, |\varepsilon_{t-\tau}|)}{\mathrm{Var}(|\varepsilon_t|)} = \frac{\mathrm{E}(|\varepsilon_t||\varepsilon_{t-\tau}|) - \mathrm{E}^2(|\varepsilon_t|)}{\mathrm{E}(\varepsilon_t^2) - \mathrm{E}^2(|\varepsilon_t|)}, \\ \rho_2(\tau) &= \mathrm{Corr}(\varepsilon_t^2, \varepsilon_{t-\tau}^2) = \frac{\mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-\tau}^2)}{\mathrm{Var}(\varepsilon_t^2)} = \frac{\mathrm{E}(\varepsilon_t^2 \varepsilon_{t-\tau}^2) - \mathrm{E}^2(\varepsilon_t^2)}{\mathrm{E}(\varepsilon_t^4) - \mathrm{E}^2(\varepsilon_t^2)}. \end{aligned} \tag{B.8}$$

References

Abramson, A. and Cohen, I. (2007): “On the Stationarity of Markov–Switching GARCH Processes,” *Econometric Theory*, 23, 485–500.

Alexander, C. (2008a): *Practical Financial Econometrics*. Chichester: John Wiley & Sons.

Alexander, C. (2008b): *Quantitative Methods in Finance*. Chichester: John Wiley & Sons.

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

Alexander, C. and Lazar, E. (2008): “Markov Switching GARCH Diffusion,” ICMA Centre Discussion Paper in Finance 2008-01, University of Reading.

Alexander, C. and Lazar, E. (2009): “Modelling Regime–Specific Stock Price Volatility,” *Oxford Bulletin of Economics and Statistics*, 71, 761–797.

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

Ardia, D. (2008): *Financial Risk Management with Bayesian Estimation of GARCH Models*. Berlin: Springer.

Ardia, D. (2009): "Bayesian Estimation of a Markov-Switching Threshold Asymmetric GARCH Model with Student-*t* Innovations," *Econometrics Journal*, 12, 105–126.

Arellano-Valle, R. B. and Azzalini, A. (2008): "The Centred Parametrization for the Multivariate Skew-Normal Distribution," *Journal of Multivariate Analysis*, 99, 1362–1382.

Awartani, B. M. A. and Corradi, V. (2005): "Predicting the Volatility of the S&P-500 Stock Index via GARCH Models: The Role of Asymmetries," *International Journal of Forecasting*, 21, 167–183.

Azzalini, A. (1985): "A Class of Distributions which Includes the Normal Ones," *Scandinavian Journal of Statistics*, 12, 171–178.

Azzalini, A. (1986): "Further Results on a Class of Distributions which Includes the Normal Ones," *Statistica*, 46, 199–208.

Azzalini, A. and Capitanio, A. (1999): "Statistical Applications of the Multivariate Skew Normal Distribution," *Journal of the Royal Statistical Society*, 61, 579–602.

Badescu, A., Kulperger, R., and Lazar, E. (2008): "Option Valuation with Normal Mixture GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 12(2), Article 5.

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

Bauwens, L., Preminger, A., and Rombouts, J. V. K. (2010): "Theory and Inference for a Markov Switching GARCH Model," *Econometrics Journal*, 13, 218–244.

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

Bauwens, L. and Storti, G. (2009): "A Component GARCH Model with Time Varying Weights," *Studies in Nonlinear Dynamics and Econometrics*, 13(2), Article 3.

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

Bertholon, H., Monfort, A., and Pegoraro, F. (2009): "Pricing and Inference with Mixtures of Conditionally Normal Processes," Working paper, Centre de Recherche en Économie et Statistique (CREST), Laboratoire der Finance-Assurance.

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

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

Brooks, R. D., Faff, R. W., McKenzie, M. D., and Mitchell, H. (2000): "A Multi-Country Study of Power ARCH Models and National Stock Market Returns," *Journal of International Money and Finance*, 19, 377–397.

Bulla, J. and Bulla, I. (2006): "Stylized Facts of Financial Time Series and Hidden Semi-Markov Models," *Computational Statistics and Data Analysis*, 51, 2192–2209.

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

Chang, C.-H., Lin, J.-J., Pal, N., and Chiang, M.-C. (2008): "A Note on Improved Approximation of the Binomial Distribution by the Skew-Normal Distribution," *American Statistician*, 62, 167–170.

Chen, C. W. S., So, M. P. K., and Lin, E. M. H. (2009): "Volatility Forecasting with Double Markov Switching GARCH Models," *Journal of Forecasting*, 28, 681–697.

Cheng, X., Yu, P. L. H., and Li, W. K. (2009): "On a Dynamic Mixture GARCH Model," *Journal of Forecasting*, 28, 247–265.

Chiogna, M. (2005): "A Note on the Asymptotic Distribution of the Maximum Likelihood Estimator for the Scalar Skew-Normal Distribution," *Statistical Methods and Applications*, 14, 331–341.

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

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

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

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

Ederington, L. H. and Guan, W. (2005): "Forecasting Volatility," *Journal of Futures Markets*, 25, 465–490.

El Babsiri, M. and Zakoïan, J.-M. (2001): "Contemporaneous Asymmetry in GARCH Processes," *Journal of Econometrics*, 101, 257–294.

Fernandez, C., Osiewalski, J., and Steel, M. F. (1995): "Modelling and Inference With V–Spherical Distributions," *Journal of the American Statistical Association*, 90, 1331–1340.

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

Francq, C. and Zakoïan, J.-M. (2002): "Autocovariance Structure of Powers of Switching–Regime ARMA Processes," *ESAIM: Probability and Statistics*, 6, 259–270.

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

Francq, C. and Zakoïan, J.-M. (2008): "Deriving the Autocovariances of Powers of Markov–Switching GARCH Models, with Applications to Statistical Inference," *Computational Statistics and Data Analysis*, 52, 3027–3046.

Geweke, J. and Amisano, G. (2009): "Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns," Forthcoming in: *Journal of Applied Econometrics*.

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

Giot, P. and Laurent, S. (2003): "Value–at–Risk for Long and Short Trading Positions," *Journal of Applied Econometrics*, 18, 641–664.

Gonçalves, E., Leite, J., and Mendes-Lopes, N. (2009): “A Mathematical Approach to Detect the Taylor Property in TARCH Processes,” *Statistics and Probability Letters*, 79, 602–610.

Granger, C. W. J. (2005): “The Past and Future of Empirical Finance: Some Personal Comments,” *Journal of Econometrics*, 129, 35–40.

Granger, C. W. J. and Ding, Z. (1995): “Some Properties of Absolute Return, An Alternative Measure of Risk,” *Annales D’économie et de Statistique*, 40, 67–91.

Granger, C. W. J., Spear, S., and Ding, Z. (2000): “Stylized Facts on the Temporal and Distributional Properties of Absolute Returns: An Update,” in Chan, W., Li, W. K., and Tong, H., eds, *Statistics and Finance: An Interface*. London: Imperial College Press.

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

Grigoletto, M. and Lisi, F. (2009): “Looking for Skewness in Financial Time Series,” *Econometrics Journal*, 12, 310–323.

Haas, M. (2009): “Persistence in Volatility, Conditional Kurtosis, and the Taylor Property in Absolute Value GARCH Processes,” *Statistics and Probability Letters*, 79, 1674–1683.

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

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

Haas, M., Mittnik, S., and Paolella, M. S. (2009): “Asymmetric Multivariate Normal Mixture GARCH,” *Computational Statistics and Data Analysis*, 53, 2129–2154.

Hamilton, J. D. (2008): “Regime Switching Models,” in Durlauf, S. N. and Blume, L. E., eds, *New Palgrave Dictionary of Economics, 2nd edition*, vol. 7. Houndmills: Palgrave Macmillan.

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

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

He, C., Silvennoinen, A., and Teräsvirta, T. (2008): "Parameterizing Unconditional Skewness in Models for Financial Time Series," *Journal of Financial Econometrics*, 6, 208–230.

He, C. and Teräsvirta, T. (1999): "Properties of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 92, 173–192.

Henze, N. (1986): "A Probabilistic Representation of the 'Skew-normal' Distribution," *Scandinavian Journal of Statistics*, 13, 271–275.

Jondeau, E. and Rockinger, M. (2003): "Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements," *Journal of Economic Dynamics and Control*, 27, 1699–1737.

Kaufmann, S. and Scheicher, M. (2006): "A Switching ARCH Model for the German DAX Index," *Studies in Nonlinear Dynamics and Econometrics*, 10(4), Article 3.

Kim, T.-H. and White, H. (2004): "On More Robust Estimation of Skewness and Kurtosis," *Finance Research Letters*, 1, 56–73.

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

Komunjer, I. (2007): "Asymmetric Power Distribution: Theory and Applications to Risk Measurement," *Journal of Applied Econometrics*, 22, 891–921.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006): "Value–at–Risk Prediction: A Comparison of Alternative Strategies," *Journal of Financial Econometrics*, 4, 53–89.

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

Lange, T. and Rahbek, A. (2009): "An Introduction to Regime Switching Time Series Models," in Andersen, T. G., Davis, R. A., Kreiß, J.-P., and Mikosch, T., eds, *Handbook of Financial Time Series*. Berlin: Springer.

Lee, H.-T. (2009): "A Copula–Based Regime–Switching GARCH Model for Optimal Futures Hedging," *Journal of Futures Markets*, 29, 946–972.

Lejeune, B. (2009): "A Diagnostic M-Test for Distributional Specification of Parametric Conditional Heteroscedasticity Models for Financial Data," *Journal of Empirical Finance*, 16, 507–523.

Lin, T. I., Lee, J. C., and Yen, S. Y. (2007): "Finite Mixture Modelling using the Skew Normal Distribution," *Statistics Sinica*, 17, 909–927.

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

Liu, J.-C. (2007): "Stationarity for a Markov–Switching Box–Cox Transformed Threshold GARCH Process," *Statistics and Probability Letters*, 77, 1428–1438.

Liu, J.-C. (2009): "Integrated Markov–Switching GARCH Processes," *Econometric Theory*, 25, 1277–1288.

Malevergne, Y. and Sornette, D. (2003): "Testing the Gaussian Copula Hypothesis for Financial Assets Dependences," *Quantitative Finance*, 3, 231–250.

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

Martínez, E. H., Varela, H., Gómez, H. W., and Bolfarine, H. (2008): "A Note on the Likelihood and Moments of the Skew-normal Distribution," *SORT*, 32, 57–66.

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.

Nelson, D. B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

Nieto, M. R. and Ruiz, E. (2008): "Measuring Financial Risk: Comparison of Alternative Procedures to Estimate VaR and ES," Statistics and Econometrics Working Paper 08-73, Department of Economics, Universidad Carlos III de Madrid.

Osborne, M. F. M. (1959): "Brownian Motion in the Stock Market," *Operations Research*, 7, 145–173.

Owen, D. B. (1956): "Tables for Computing Bivariate Normal Probabilities," *Annals of Mathematical Statistics*, 27, 1075–1090.

Peiró, A. (1999): "Skewness in Financial Returns," *Journal of Banking and Finance*, 23, 847–862.

Peiró, A. (2002): "Skewness in Individual Stocks at Different Investment Horizons," *Quantitative Finance*, 2, 139–146.

Peiró, A. (2004): "Asymmetries and Tails in Stock Index Retruns: Are Their Distributions Really Asymmetric?" *Quantitative Finance*, 4, 37–44.

Pelagatti, M. M. (2009): "Modelling Good and Bad Volatility," *Studies in Nonlinear Dynamics and Econometrics*, 13(1), Article 2.

Poskitt, D. S. and Chung, S.-H. (1996): "Markov Chain Models, Time Series Analysis and Extreme Value Theory," *Advances in Applied Probability*, 28, 405–425.

Rombouts, J. V. K. and Stentoft, L. (2009): "Bayesian Option Pricing Using Mixed Normal Heteroskedasticity Models," CREATES Research Papers 2009-07, School of Economics and Management, University of Aarhus.

Rombouts, J. V. K. and Bouaddi, M. (2009): "Mixed Exponential Power Asymmetric Conditional Heteroskedasticity," *Studies in Nonlinear Dynamics and Econometrics*, 13(3), Article 3.

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

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

Sajjad, R., Coakley, J., and Nankervis, J. C. (2008): "Markov–Switching GARCH Modelling of Value–at–Risk," *Studies in Nonlinear Dynamics and Econometrics*, 12(3), Article 7.

Smith, J. Q. (1985): "Diagnostic Checks of Non-Standard Time Series Models," *Journal of Forecasting*, 4, 283–291.

So, M. K. P. and Yu, P . L. H. (2006): "Empirical Analysis of GARCH Models in Value-at-Risk Analysis," *Journal of International Financial Markets, Institutions and Money*, 16, 180–197.

Taylor, S. J. (1986): *Modelling Financial Time Series*. Chichester: John Wiley & Sons.

Teräsvirta, T. (2009): "An Introduction to Univariate GARCH Models," in Andersen, T. G., Davis, R. A., Kreiß, J.-P., and Mikosch, T., eds, *Handbook of Financial Time Series*. Berlin: Springer.

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

Veiga, H. (2009): "Financial Stylized Facts and the Taylor-Effect in Stochastic Volatility Models," *Economics Bulletin*, 29, 265–276.

Vlaar, P. J. G. and Palm, F. C. (1992): "Mean Reversion, Conditional Heteroskedasticity and Jumps in the EMS," *Tinbergen Institute Research Bulletin*, 4, 221–230.

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.

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