The skew–normal Markov–switching GARCH process

Abstract

Markov–switching GARCH (MS–GARCH) models provide an attractive framework for modeling the distribution of daily asset returns. The model class can capture most of the characteristic features of these variables, while conditional normality within the regimes is preserved. This, along with CLT arguments, is deemed an attractive property. In the framework of these models, asymmetries are usually captured by allowing for different regime means. However, this introduces autocorrelation of raw returns, which may not be desirable. In this paper, skewness is introduced into the MS–GARCH model by taking the regime densities as belonging to the class of (centered) skew–normal densities. The appearance of this distribution can be explained by a pre-asymptotic behavior of daily returns, and can still be viewed as "generic". The dynamic properties of the process are derived. In an application to various European stock index returns the model shows a strong performance.

Keywords—Conditional volatility, GARCH, Markov–switching, Skew–normal distribution, Stock market

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 (2008, 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 $\{\epsilon_t\}$ be generated by

$$\epsilon_t = \sigma_{\Delta_t,t}\eta_t, \tag{1}$$

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 $S = \{1,\ldots,k\}$ and primitive (i.e., irreducible and aperiodic) transition matrix $\boldsymbol{P}$,

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

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

$$\sigma_{jt} = \omega_j + \alpha_j|\epsilon_{t-1}| + \beta_j\sigma_{j,t-1}, \tag{3}$$

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

Model (1)–(3), which will be referred to as MS(k)–GARCH, is an absolute value version of the MS–GARCH model introduced by Haas et al. (2004b) and further studied, for example, in Abramson and Cohen (2007), Alexander and Lazar (2008a), Ardia (2008a,b), and Liu (2006, 2007). Alternative formulations of MS–(G)ARCH processes exist (Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996; Dueker, 1997; and Klaassen, 2002; for an overview, see also the review articles of Hamilton, 2008; Lange and Rahbek, 2008; and Teräsvirta, 2008), but it may be argued that the one given by (1)–(3) has several advantages, see Haas et al. (2004b), Bauwens and Storti (2007), and Ardia (2008a,b) for discussion. In the absolute value GARCH specification (3), as originally proposed by Taylor (1986), the regime–specific volatility dynamics are specified in terms of the standard deviations and absolute shocks $|\epsilon_t|$ instead of the variances and squared shocks as in the classic GARCH setup of Bollerslev (1986). This is motivated by the observation of Taylor (1986) and Ding et al. (1993) that the time series dependencies tend to be stronger in the absolute than in the squared returns, which was termed the *Taylor effect* by Granger and Ding (1995), see also Ané (2006) and Haas (2007) for further discussion.

Besides incorporating the economically intuitive concept of regime–specific volatility dynamics and providing accurate forecast densities (e.g., Ardia, 2008b, Zhuang and Chan, 2004; Maheu, 2005; and Sajjad et al., 2008), 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 (3), 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, 2008). 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), and Komunjer (2007), 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 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) to

$$\epsilon_t = \mu_{\Delta_t} + \sigma_{\Delta_t,t}\eta_t. \tag{4}$$

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. Firstly, many return series exhibit significant skewness but no autocorrelation, and secondly, 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 by replacing 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 while 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 asset returns. At this point, it may be worthwhile to note in passing that recently, in a more standard situation, Chang et al. (2008) also successfully use the skew–normal to account for 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 relevant in our context, and in Section 2.2 the use of skew–normal densities in the context of MS–GARCH models is discussed.

2.1 The skew–normal (SN) distribution

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

$$f(z;\gamma) = 2\phi(z)\Phi(\gamma z), \quad \gamma \in \mathbb{R}, \tag{5}$$

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 two times $\Phi(\gamma z)$. If $\gamma < 0$, this weight 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 skewness of the SN. An example of an SN density is shown in the top panel of Figure 1, which, in the left graph, plots the *standardized*, i.e., mean zero and unit variance version of (5) with $\gamma = -1.5$ (which is of the order of magnitude of what we find for stock returns), along with the standard normal pdf (dashed line). The right plot is a magnification of the left tail. The plots show that, apart from the asymmetries, the shape of the SN is very similar to the Gaussian. In particular, although there is more probability mass in the left tail (and correspondingly less in the right tail), both tails are thin in the sense that they exhibit rapid decay. More formally, the tail behavior of the SN distribution can be characterized by using Azzalini's result that $|Z| \stackrel{d}{=} |X|$, where $X \sim \mathrm{N}(0,1)$, and $\stackrel{d}{=}$ denotes equality in distribution. Thus, we have (see, e.g., Feller, 1950, p. 131)

$$p(|Z| > z) = p(|X| > z) = 2\Phi(-z) \cong \sqrt{\frac{2}{\pi}}\frac{e^{-z^2/2}}{z} = \frac{2\phi(z)}{z} \quad \text{as } z \to \infty,$$

where $f(x) \cong g(x)$ as $x \to \infty$ means $\lim_{x\to\infty} f(x)/g(x) = 1$, i.e., the tail behavior of the SN is Gaussian. Nevertheless, the skewness captured by the SN distribution can have important consequences for applications in risk management, such as the calculation of Value–at–Risk measures. To illustrate this, the left plot of the center panel of Figure 1 shows, for various practically relevant shortfall levels, ξ, the quantiles of the (standardized) SN distribution as a function of γ, where $\gamma = 0$ corresponds to the normal distribution. Although these differences may appear relatively mild for the standardized SN, they can become rather important in periods of high volatility.

Moments can and will also be useful to describe the properties of return distributions. From the result cited above, it is clear that the even moments are identical to those of the Gaussian. Azzalini (1985) calculated the first two odd moments, and Henze (1986) and Martínez et al. (2008) provide general expressions. In Appendix C, these are briefly discussed, and a new one is derived. In particular, it is shown that

$$\mathrm{E}(Z^{2\ell+1}) = \sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^\ell}\sum_{i=0}^{\ell}\frac{(-1)^i\delta^{2i+1}}{i!(\ell-i)!(2i+1)} = \mathrm{sign}(\delta)\frac{2^{\ell+1/2}\ell!}{\sqrt{\pi}}\mathrm{B}^{\mathrm{inc}}(\delta^2; 1/2, \ell+1), \tag{6}$$

$\ell = 0, 1, \ldots$, where $\delta = \gamma/\sqrt{1+\gamma^2}$, and $\mathrm{B}^{\mathrm{inc}}$ is the incomplete beta function defined in (C.19) in Appendix C. Application of (6) gives the well–known expressions

$$\mathrm{E}(Z) = \sqrt{\frac{2}{\pi}}\delta, \quad \text{and} \quad \mathrm{E}(Z^3) = \sqrt{\frac{2}{\pi}}(3\delta - \delta^3), \tag{7}$$

and the moment–based coefficient of skewness, m_3, is given by

$$m_3 = \frac{\sqrt{2}(4-\pi)\delta^3}{(\pi - 2\delta^2)^{3/2}} \in (-0.995, 0.995), \tag{8}$$

which was calculated by Azzalini (1985). Thus, although limited, the range of feasible skewness coefficients is sufficient for most return distributions. The relationship between γ and m_3 is displayed in the right plot of the center panel of Figure 1. We also observe from (7) 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 (5) in Section (2.2) to define the skew–normal MS–GARCH process.

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

$$F(z;\gamma) = \Phi(z) - 2T(z,\gamma), \tag{9}$$

where

$$T(z,\gamma) = \frac{1}{2\pi}\int_0^\gamma \frac{\exp\left\{-\frac{z^2}{2}(1+x^2)\right\}}{1+x^2}dx \tag{10}$$

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

$$T(z,\gamma) = \frac{\arctan(\gamma)}{2\pi} - \frac{1}{2\pi}\sum_{i=0}^{\infty}(-1)^i \frac{1 - e_i(z^2/2)e^{-z^2/2}}{2i+1}\gamma^{2i+1}, \tag{11}$$

where $e_n(x) = \sum_{k=0}^{n} x^k/k!$ is the exponential sum function. For $|\gamma| > 1$, integration by parts shows that we can apply (9) and (11) in conjunction with

$$F(z;\gamma) = \begin{cases} 2\Phi(z)\Phi(\gamma z) - F(\gamma z;\gamma^{-1}) & \text{if} \quad \gamma \geq 0, \\ 2\Phi(z)\Phi(\gamma z) + 1 - F(\gamma z;\gamma^{-1}) & \text{if} \quad \gamma < 0. \end{cases} \tag{12}$$

According to our experience, (11) 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, abbreviated SN(γ)–MS(k)–GARCH, is obtained by replacing the N(0, 1) distribution for η_t in (1) by an SN(γ), i.e., the model is described by (1)–(3) along with

$$\eta_t \overset{iid}{\sim} \mathrm{SN}(\gamma), \tag{13}$$

where, as ϵ_t in (1) is interpreted as an unexpected random shock, we take the centered version of the density defined in (5). In principle, 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

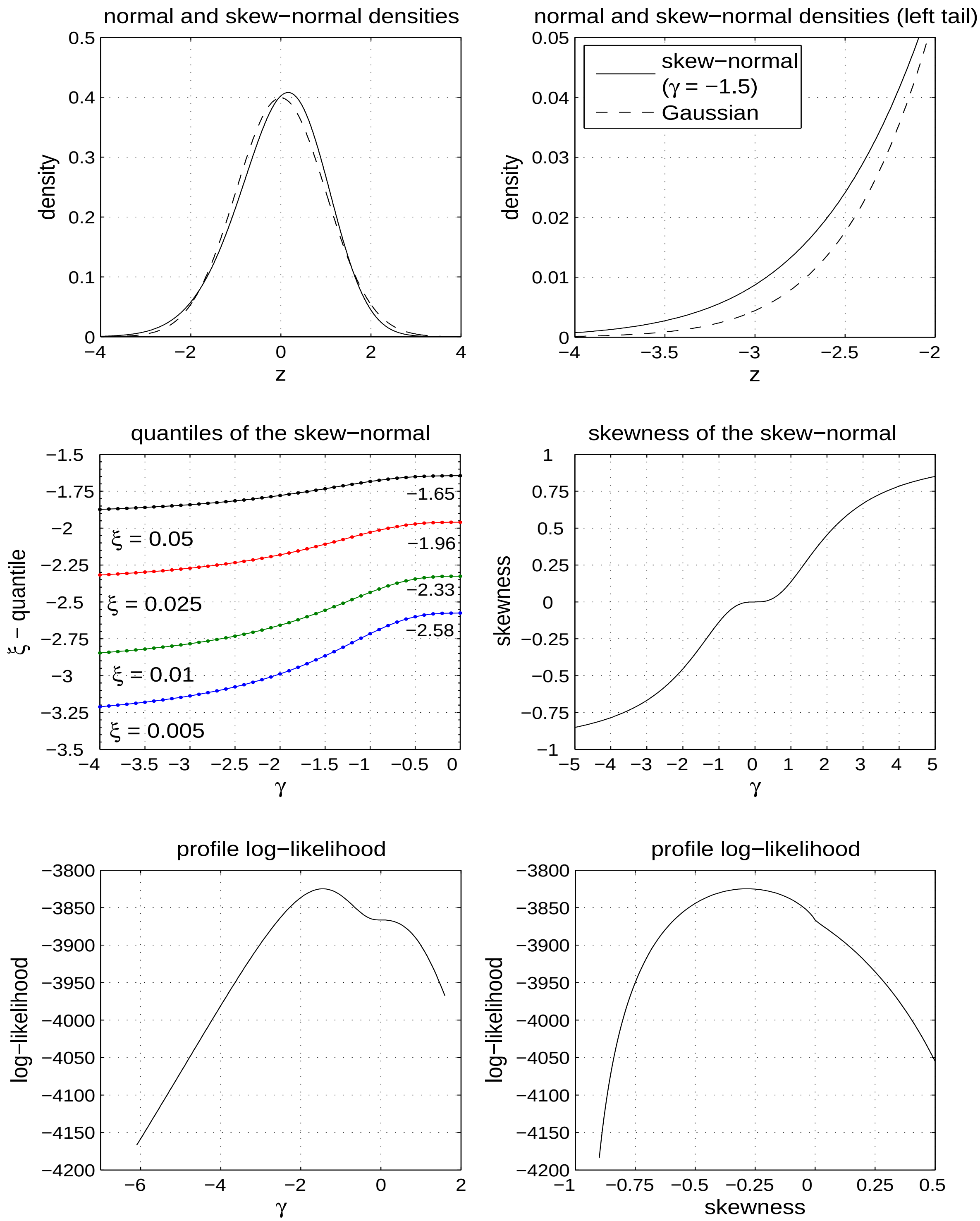

Figure 1: The top left plot compares the standardized skew–normal (SN) density with $\gamma = -1.5$ (solid line) with the standard normal density (dashed line). The right plot is a magnification of the left tail. The center panel shows, in its left and right graphs, respectively, various quantiles of the standardized SN as well as its moment–based skewness measure as a function of γ. For different parametrizations of the SN density, as discussed in Section 2.2.2, the bottom panel displays the profile log–likelihood for the skewness parameter when applying a single–regime SN–GARCH process to the DAX returns described at the beginning of Section 3.

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(γ)–MS(k)–GARCH model, skewness is *conditional* in the sense that it is a property of the innovations η_t in (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 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)–(3) with (13) is a rather natural generalization of Osborne's (1959) Gaussian model.

To some extent, an exception to what has been said above concerning the relation between skewness and mean dynamics is the normal mixture GARCH model discussed, among others, in Alexander and Lazar (2006, 2008b), Badescu et al. (2008), Bauwens et al. (2006), Bauwens and Rombouts (2007), Bentarzi and Hamdi (2008), Giannikis et al. (2007), Haas et al. (2004a), and Wong and Li (2001), where it is assumed that the transition matrix (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 $\boldsymbol{\pi}_\infty$, where use of the extension in (4), with $\mu_k = -\sum_{j=1}^{k}(\pi_{j,\infty}/\pi_{k,\infty})\mu_j$ to have $\mathrm{E}(\epsilon_t) = 0$, will not induce return autocorrelation, but allows great flexibility with respect to skewness (see, e.g., Haas et al., 2004a). 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.

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)–(3) and (13) 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. Having available explicit expressions for the autocorrelations of both the absolute as well as the squared process may be useful in assessing whether a given model is able to capture several empirical characteristics of the data, such as the Taylor effect referred to in Section 1. For the single–regime absolute value GARCH(1,1) model, He and Teräsvirta (1999) find that, with normal innovations, the Taylor effect may be present but only to a rather minor degree and for parameter constellations with very high values of the unconditional kurtosis. We shall see in Section 3.2 that the regime–switching models offer much more flexibility in this regard (also).

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 (5) 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 Figure 1, showing the profile log–likelihood for the skewness parameter γ in the context of a single–regime SN(γ)–GARCH(1,1) model, where $k = 1$ in (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 (8). As shown in the bottom right graph of Figure 1, 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 (8).

3 Empirical results

In this section, we investigate the volatility dynamics of daily returns of three European stock market indices, namely, the French CAC 40, the German DAX 30, and the British FTSE 100, from January 1990 to December 2007, a sample of $T = 4520$ observations for each series. 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 data were obtained from Datastream.

We first estimate the models over the (approximately) first ten years of data, i.e., the first 2500 observations. The remaining observations are retained for backtesting out–of–sample predictive densities and Value–at–Risk measures. First–order dynamics are negligible in the series, so that our model is just

$$r_t = \mu + \epsilon_t, \tag{14}$$

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

A major question to be investigated in 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 into 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 Mix–GARCH. 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. This restricted model will be included with both normal and skew–normal innovations, and in the first case it will be denoted by Mix_S in order to distinguish it from the asymmetric normal mixture GARCH process. Moreover, another restricted version of the multi–regime GARCH processes turns out to be empirically relevant, namely, a model where only the intercept in (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 mixture GARCH with switching intercept but common volatility dynamics will be denoted by $\text{Mix}_S^{\text{icept}}$. Note that, defining $\lambda = (1 - \beta_1)^{-1}(\omega_2 - \omega_1)$, this model can be written as $\sigma_{2t} = \sigma_{1t} + \lambda$, 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, 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

$$f(z;\nu,p,\theta)=\frac{\theta}{1+\theta^2}\frac{p}{\nu^{1/p}B(\nu,1/p)}\begin{cases}\left(1+\frac{(|z|\theta)^p}{\nu}\right)^{-(\nu+1/p)}, & \text{if} \quad z<0\\ \left(1+\frac{(z/\theta)^p}{\nu}\right)^{-(\nu+1/p)}, & \text{if} \quad z\geq 0,\end{cases} \tag{15}$$

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 (15) 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 (15), then $X = \sqrt{2}Z$ has a standard Student's t distribution with 2ν degrees of freedom. All models are estimated in the absolute value GARCH specification (3), see the discussion at the end of Section 1.

3.2 In–sample results

Before turning to the evaluation of out–of–sample predictive densities, we describe a few estimation results for the in–sample period, insofar as they reflect certain general properties of the multi–regime models studied herein. In particular, we discuss parameter estimates and the implied moment structure of (some of) the models under consideration. Parameter estimates for the CAC 40 are reported in Table 2; those for the other series are, at least with respect to the aspects highlighted in the following discussion, qualitatively similar and not reported.

The parameter estimates and the implied regime–specific unconditional variances, i.e., $\mathrm{E}(\epsilon_t^2|\Delta_t = j)$, $j = 1, 2$, show that all the two–component models identify a low–volatility regime (Regime 1) associated with a high unconditional probability ($\pi_{1,\infty}$) and a high–volatility regime which occurs less frequently, with the estimates of the unconditional bear market probability ranging from approximately 6.5 to 12%. However, regarding the regime–specific volatility *dynamics*, there is a striking difference between the models built on the assumption of independent regimes ("Mix") and those allowing them to be persistent ("Markov"). Namely, in the iid mixture GARCH processes, we observe that, compared to the low–volatility component, conditional volatility is "highly reactive to market shocks in the 'crash' regime, yet the effect dies out soon because the persistence parameters are all low" (Alexander and Lazar, 2008b),

Table 1: Overview of two–regime GARCH specifications

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

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

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

where the dynamics of the regime–specific variances are governed by (3), and the transition matrix (2) becomes

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

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 $\boldsymbol{\pi}_\infty = (\pi_{1,\infty}, \pi_{2,\infty})'$, 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 $\boldsymbol{P} = \boldsymbol{\pi}_\infty \mathbf{1}_2'$, where $\mathbf{1}_2 = (1, 1)'$, and the conditional distribution of Δ_t is equal to its stationary distribution, $\boldsymbol{\pi}_\infty$. The innovation η_t follows a (centered) skew–normal distribution with shape parameter γ (see (5)), which becomes Gaussian if $\gamma = 0$.

i.e., α_2 is considerably larger than α_1, whereas β_2 is smaller than β_1 (as also reported for mixture GARCH models in Alexander and Lazar, 2006; Badescu et al., 2008; Haas et al., 2004a, 2008; and Cheng et al., 2008). In the MS–GARCH specifications the dynamics of the regimes are much more similar, and the components differ mainly in their intercepts, representing different regime–specific volatility *levels*. This finding leads us to consider models with $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$ (see Table 1), which are more parsimonious but, in view of the results in Table 2, may still be sufficient to capture the main characteristics of the volatility process. In fact, it turns out that likelihood ratio tests (not reported) favor this restriction for the MS–GARCH models but reject for the iid mixture specifications. Is this a plausible result? First of all, note that it is not *a priori* surprising that implied within–regime volatility dynamics are different for the two specifications, because in case of persistent regimes there is an additional

Table 2: Parameter estimates for absolute value GARCH models (3) for the CAC 40 over the in–sample period (1990-1999)

	Normal				Skew–normal		
	Single	Mix$_S$	Mix	Markov	Single	Mix	Markov
ω_1	0.057 (0.014)	0.014 (0.008)	0.010 (0.007)	0.012 (0.004)	0.054 (0.015)	0.017 (0.009)	0.014 (0.005)
α_1	0.091 (0.014)	0.054 (0.013)	0.044 (0.015)	0.024 (0.005)	0.108 (0.016)	0.065 (0.014)	0.029 (0.007)
β_1	0.882 (0.020)	0.941 (0.015)	0.952 (0.017)	0.968 (0.007)	0.897 (0.017)	0.942 (0.013)	0.967 (0.007)
$\mathrm{E}(\epsilon_t^2 \vert \Delta_t = 1)$	1.631	1.279	n.a.	1.188	1.660	1.308	1.194
$\pi_{1,\infty}$	1	0.919 (0.051)	0.879 (0.068)	0.935	1	0.927 (0.048)	0.935
μ_1	0	0	0.057 (0.031)	0	0	0	0
ω_2	–	0.319 (0.186)	0.194 (0.108)	0.059 (0.093)	–	0.316 (0.191)	0.062 (0.098)
α_2	–	0.349 (0.147)	0.327 (0.114)	0.009 (0.014)	–	0.452 (0.199)	0.012 (0.017)
β_2	–	0.685 (0.111)	0.730 (0.077)	0.973 (0.038)	–	0.702 (0.101)	0.974 (0.036)
$\mathrm{E}(\epsilon_t^2 \vert \Delta_t = 2)$	–	4.445	n.a.	6.473	–	4.430	6.453
$\pi_{2,\infty}$	0	0.081 (0.051)	0.121 (0.068)	0.065	0	0.073 (0.048)	0.065
μ_2	–	0	−0.412 (0.207)	0	–	0	0
skewness	0	0	0	0	−0.179 (0.040)	−0.148 (0.057)	−0.073 (0.054)
γ	0	0	0	0	−1.134	−1.036	−0.764
$\boldsymbol{P}$	1	$\boldsymbol{\pi}_\infty \mathbf{1}_2'$	$\boldsymbol{\pi}_\infty \mathbf{1}_2'$	$\begin{pmatrix} \underset{(0.003)}{0.993} & \underset{(0.034)}{0.098} \\ \underset{(0.003)}{0.007} & \underset{(0.034)}{0.902} \end{pmatrix}$	1	$\boldsymbol{\pi}_\infty \mathbf{1}_2'$	$\begin{pmatrix} \underset{(0.003)}{0.993} & \underset{(0.034)}{0.094} \\ \underset{(0.003)}{0.007} & \underset{(0.034)}{0.906} \end{pmatrix}$
ψ	0	0	0	0.895 (0.035)	–	0	0.899 (0.035)
$\log L$	−3931.1	−3890.6	−3887.8	−3872.7	−3919.0	−3887.2	−3871.8

Approximate standard errors are given in parentheses. "Single" denotes a model with just one regime, i.e., $k = 1$ in (1)–(3). See Table 1 for the definitions of the different multi–regime models. "skewness" is the moment–based coefficient of skewness (8) of the *innovations*, η_t, as implied by the estimated shape parameter, γ, of the skew–normal density (5), and ψ is the measure for the persistence of the regimes, i.e., the largest subdominant eigenvalue of the transition matrix $\boldsymbol{P}$, given by $\psi = p_{11} + p_{22} - 1$, see the footnote of Table 1. $\log L$ denotes the value of the log–likelihood evaluated at the MLE.

source of time–varying volatility in the Markov–switching process. In fact, in all the multi–regime models (except the mixed normal with regime–specific means), the conditional overall standard deviation is $\sqrt{\pi_{1t|t-1}\kappa_2(\gamma)\sigma_{1t}^2 + \pi_{2t|t-1}\kappa_2(\gamma)\sigma_{2t}^2}$, where $\kappa_2(\gamma) = \mathrm{E}(\eta_t^2) = 1 - 2\delta^2/\pi$, and $\pi_{jt|t-1}$ is the time–t probability of regime j, $j = 1, 2$, given the information (return history) up to time $t-1$. In the iid mixture GARCH, these probabilities are constant, whereas, in the MS–GARCH with persistent regimes, a large shock at time $t-1$ will raise the probability of being in the high–volatility regime at time t, which contributes to the reactiveness of the overall conditional volatility to shocks and potentially makes the extreme shock responsiveness of the high–volatility component superfluous. The situation is depicted in Figure 2 for the evolution of conditional volatility in 1998. The top plot of Figure 2 shows the conditional standard deviations of the high–volatility regime both for the skew–normal iid mixture and MS–GARCH specifications. Compared to the highly volatile behavior of the former, the latter, due to the small $\alpha_2 = 0.012$, reacts rather sluggishly. However, the high–volatility regime's conditional probabilities are fixed in the iid model, whereas those of MS–GARCH, displayed in the center panel of Figure 2, substantially increase in the high–volatility periods (as the regimes are persistent, see Table 2). As a result, the overall volatilities of the two models, pictured in the plot thereunder, behave rather similar. That of MS–GARCH tends to be even somewhat higher in the turbulent periods, but then it reverts rather fast. Comparing the log–likelihood values, as reported in the last line of Table 2, we see that MS–GARCH is clearly preferred to the iid mixture, i.e., the hypothesis $p_{11} = 1 - p_{22}$ is rejected both with normal and skew–normal innovations. Another observation that deserves mention is that, according to the likelihood values, the skew–normal dominates the normal for both single–regime models and iid mixtures but not for MS–GARCH. However, this may not be particularly informative with respect to predictive performance, which we will turn to in Section 3.3.

Unconditional moments of the models are reported in Table 3, along with, in the first row, the corresponding sample analogues calculated from the demeaned return series. We note that the CAC 40 and the DAX 30 have a somewhat higher volatility than the FTSE 100, as they represent less diversified portfolios. Also, not surprisingly, all return series display considerable excess kurtosis. The skewness coefficients are negative for the CAC and the DAX and positive but rather small in value for the FTSE; however, all of them are clearly in the range that can be captured by the skew–normal, see (8). Comparing the moments implied by the models with those of the sample, we observe that, for all three series, the kurtosis implied by the single–regime models with both normal and skew–normal innovations does not match the empirical counterparts well. In contrast, the mixture models with innovations drawn from these distributions reproduce all the empirical moments more accurately, which may be taken

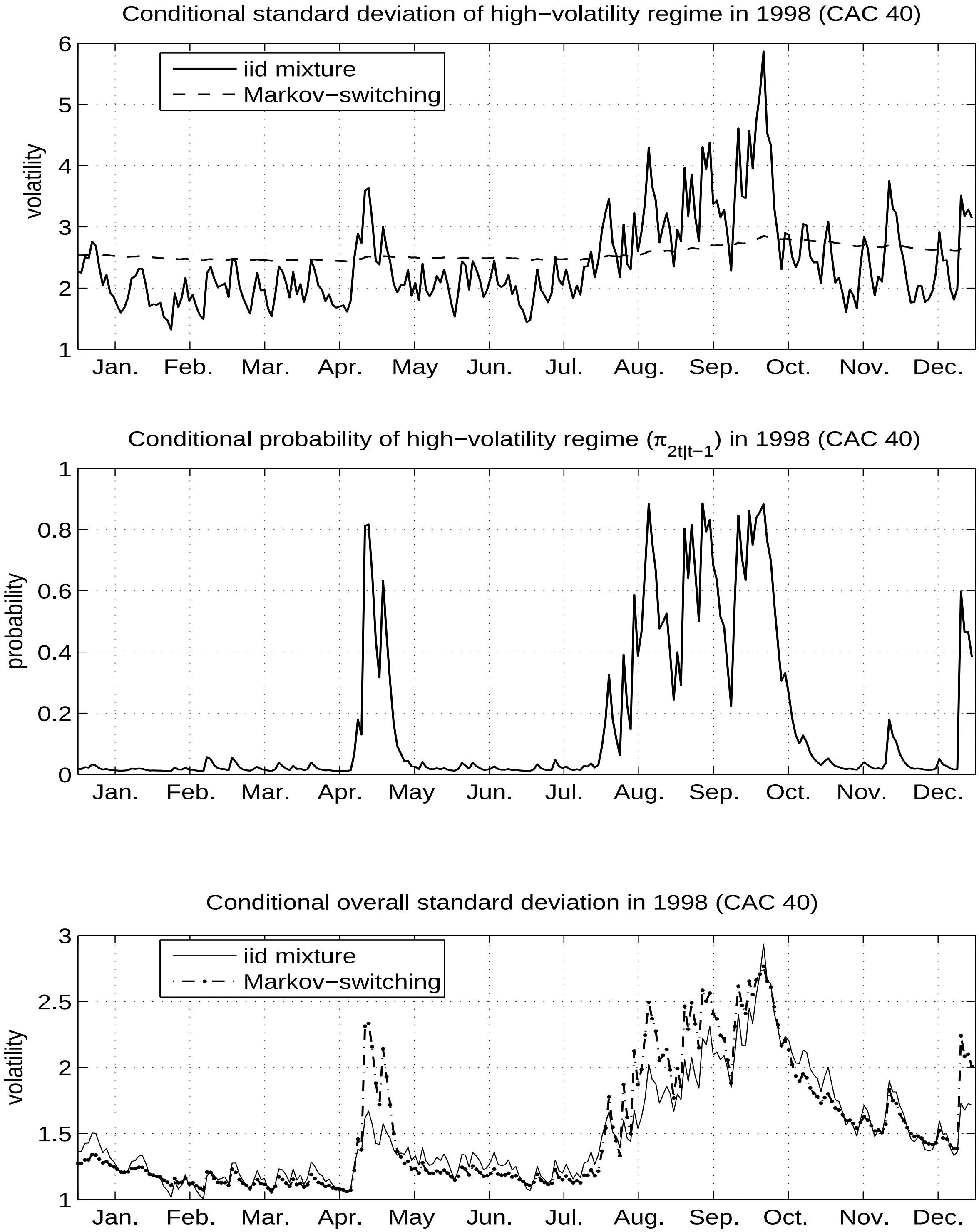

Figure 2: The figure compares the volatility processes of the iid mixture and the MS–GARCH models with skew–normal innovations for the CAC 40 during 1998. The top plot shows the evolution of the conditional standard deviation in the high–volatility regime, and the center panel exhibits this regime's conditional probability implied by the MS–GARCH specification. The bottom plot displays the overall conditional volatility for both models.

as a first, although informal, indication for what has been argued in favor of in Section 1, namely, that conditional within–regime fat–tailedness may be superfluous.

We shall also use the results of Appendix B.1 to illustrate the model's capability to capture the Taylor effect, as discussed in Section 1. We do so using the DAX 30 returns as an example since in–sample the effect is strongest for this series. For three fitted models, namely the single–regime GARCH(1,1) specifications with skew–normal and Student's t innovations and the skew–normal MS–GARCH process, Figure 3 displays the theoretical autocorrelation functions (ACFs) of both the absolute and the squared returns, along with their empirical counterparts estimated from the data. The upper half of Figure 3 shows that the Taylor effect, which is visible in the sample autocorrelations, can be modeled by means of the multi–regime process, whereas it is not present in the fitted single–regime skew–normal GARCH(1,1). The latter observation corresponds to the findings of He and Teräsvirta (1999) that Gaussian (absolute value) GARCH(1,1) processes produce the Taylor effect only for parameter constellations associated with "very strong leptokurtosis". For example, Figure 1 of He and Teräsvirta (1999) suggests that the kurtosis of such a process is typically greater than 10, whereas it is only 4.19 for the fitted skew–normal GARCH(1,1) model.

However, He and Teräsvirta (1999) only considered the model with Gaussian errors, with *unconditional* kurtosis controlled by the reaction parameter α in (3). In the single–regime models, more flexibility can be achieved by allowing the conditional distribution to have fat tails, as we show in the lower half of Figure 3, where the left panel reproduces the ACFs of the absolute values and the squares for the fitted t–GARCH(1,1) process, revealing that it incorporates a prominent Taylor effect. To further exemplify the role of conditional kurtosis, the bottom right plot presents the first–order autocorrelations as functions of the degrees of freedom, ν, of the conditional t density while keeping the parameters of the GARCH equation constant. It turns out that the effect shrinks and finally disappears as ν increases (and hence kurtosis decreases). See Gonçalves et al. (2008) for similar results in a pure ARCH(1) context.

Although the main purpose of Figure 3 is to illustrate that the Taylor effect can be captured by regime–switching models with light–tailed innovations (Gaussian or skew–normal), we also note that the theoretical autocorrelations implied by the fitted SN–MS(2)–GARCH process tend to be somewhat greater than their sample counterparts particularly for the squares.This is reconcilable with an observation made by He and Teräsvirta (1999) for Gaussian single–regime GARCH(1,1) processes, namely that sample autocorrelations are "heavily downward biased even for 100,000 observations." To assess the bias for our more complex model, we simulate the estimated SN–MS(2)–GARCH process for various sample sizes ranging from 2500 to 500,000, where SN random variables are generated using Henze's (1986) representation result mentioned

Table 3: Unconditional moments of estimated GARCH models

	CAC 40				DAX 30				FTSE 100			
	$E(\lvert\epsilon_t\rvert)$	Var.	Skew.	Kurt.	$E(\lvert\epsilon_t\rvert)$	Var.	Skew.	Kurt.	$E(\lvert\epsilon_t\rvert)$	Var.	Skew.	Kurt.
$\{\widehat{\epsilon}_t\}$	0.914	1.492	−0.133	5.47	0.902	1.583	−0.479	8.24	0.689	0.835	0.081	5.10
Normal												
Single	1.002	1.631	0	3.48	1.036	1.768	0	3.68	0.760	0.970	0	4.02
$\text{Mix}_S^{\text{icept}}$	0.963	1.593	0	4.76	0.894	1.530	0	7.47	0.684	0.815	0	4.71
Mix_S	0.941	1.535	0	4.95	0.888	1.497	0	7.13	0.676	0.803	0	5.10
$\text{Markov}^{\text{icept}}$	0.937	1.548	0	5.66	0.899	1.558	0	7.59	0.683	0.814	0	4.71
Markov	0.935	1.532	0	5.39	0.937	1.786	0	10.81	0.674	0.798	0	5.09
Skew–normal												
Single	1.004	1.660	−0.191	3.69	1.041	1.831	−0.313	4.19	0.759	0.969	−0.044	4.08
$\text{Mix}^{\text{icept}}$	0.960	1.581	−0.160	4.82	0.885	1.500	−0.305	7.48	0.684	0.814	−0.025	4.73
Mix	0.942	1.537	−0.175	5.07	0.879	1.476	−0.298	7.34	0.676	0.804	−0.045	5.17
$\text{Markov}^{\text{icept}}$	0.936	1.548	−0.094	5.68	0.880	1.490	−0.304	7.64	0.683	0.813	−0.020	4.73
Markov	0.935	1.534	−0.090	5.43	0.902	1.638	−0.335	10.60	0.674	0.800	−0.036	5.14
Student's t												
symmetric	0.916	1.471	0	4.93	0.834	1.370	0	9.37	0.664	0.773	0	4.91
skewed	0.934	1.509	−0.189	5.17	0.905	1.608	−0.387	21.54	0.704	0.862	−0.134	5.57

Shown are unconditional moments implied by the fitted GARCH models, as well as the corresponding sample quantities of the (demeaned) data over the in–sample period. These are reported in the row labeled $\{\widehat{\epsilon}_t\}$. "Var." is the variance, and "Skew." and "Kurt." denote the moment–based coefficient of skewness and kurtosis, respectively, i.e., the third and fourth standardized moments. For the notation of the two–regime models, see Table 1. The skewed Student's t distribution is defined by (15).

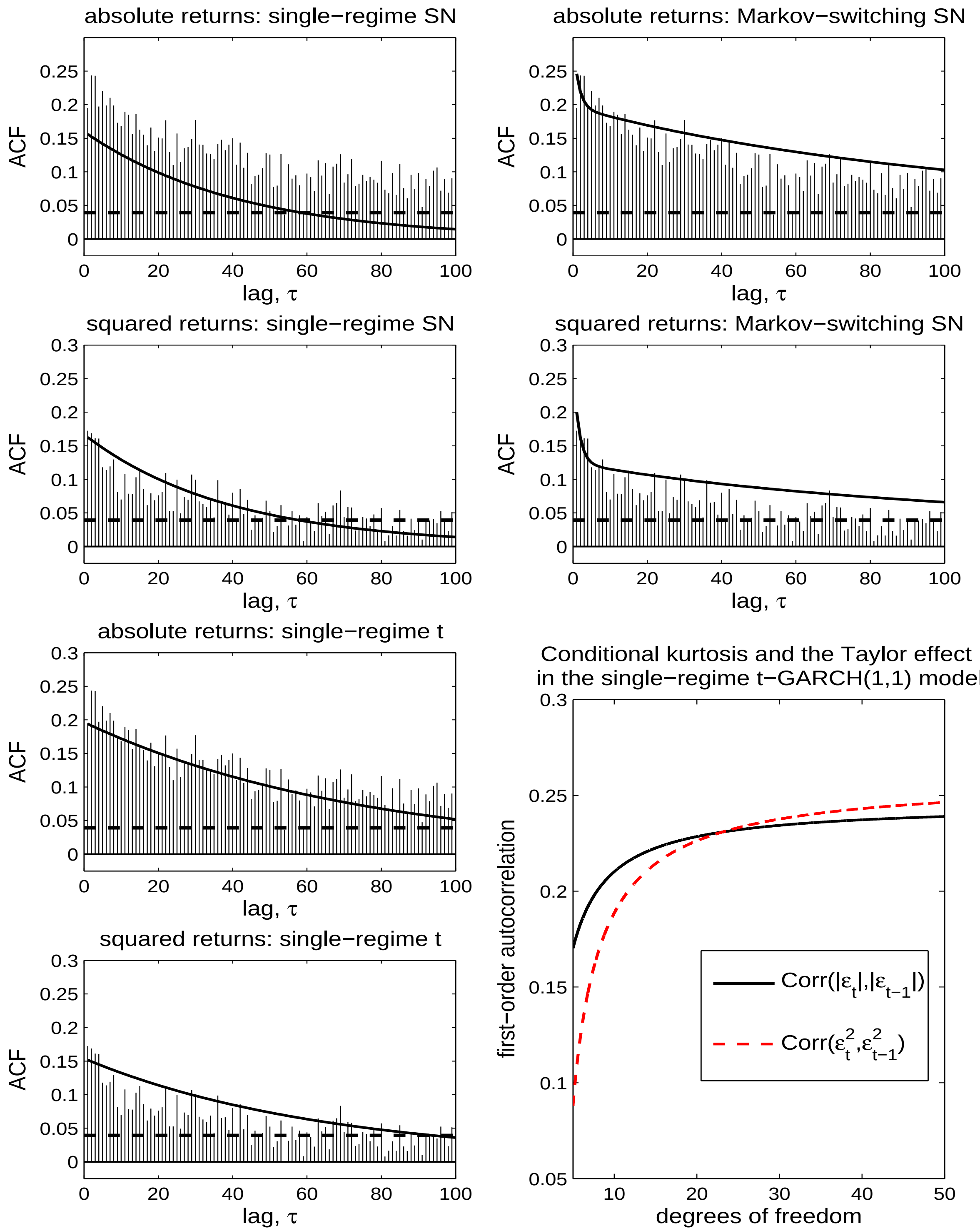

Figure 3: Plots in the upper half and in the left part of the lower half of the figure show empirical autocorrelations of absolute and squared returns of the in–sample DAX returns, along with their theoretical counterparts implied by several fitted GARCH models. The bottom right plot shows the first–order autocorrelations of absolute values and squares implied by a single–regime GARCH(1,1) process of the form $\sigma_t = \omega + (\alpha|\eta_{t-1}| + \beta)\sigma_{t-1}$, where η_t has a t distribution with unit variance, $\omega = 0.015$, $\alpha = 0.086$, and $\beta = 0.921$, as functions of the degrees of freedom, ν. The conditional kurtosis in this case is $3(\nu-2)/(\nu-4)$, i.e., kurtosis decreases with increasing ν. The estimate for the DAX returns is $\nu = 7.045$.

in Appendix C. For each sample size, 1000 realizations of the process are generated, and various moments are calculated for each replication. The averages of these are presented in Figure 4 along with the corresponding theoretical quantities. Indeed, for practically relevant sample sizes, we observe considerable downward bias in the autocorrelations as well as in the kurtosis measure, whereas the (upward) bias is less dramatic for the skewness, and no systematic bias is detected for the sample variance. For the autocorrelations, the bias at $T = 100,000$ appears to be smaller than that observed by He and Teräsvirta (1999) for the simple Gaussian GARCH(1,1), which is due to the fact that their processes are characterized by extreme unconditional kurtosis ranging from 13.8 to 291.

Quite evidently, comparisons such as those discussed in the preceding paragraphs cannot be used to assess a model's adequacy statistically. However, they may help in assessing whether several characteristics of interest implied by a model are "in the right ballpark" and, more generally, "how well [the models] are able to satisfy stylized facts in the observed series" (He and Teräsvirta, 1999), such as strong leptokurtosis and the Taylor effect. The results reported in Table 3 and Figure 3 suggest that regime-switching models with light–tailed innovations can adequately capture the excess kurtosis in financial returns, and they are also rather flexible with respect to the autocorrelation structure of the volatility process. More formal methods of out–of–sample model evaluation are considered in Section 3.3.

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 reestimate the models (approximately) every month (i.e., 20 trading days) employing a moving window of data, i.e., using the most recent 2,500 observations in the sample. In this manner, we obtain, for each model and series, 2020 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, $u_t = \widehat{F}(r_t|\Psi_{t-1})$, $t = 2501, \ldots, 4520$, where Ψ_t is the information up to time t, and $\widehat{F}(\cdot|\Psi_{t-1})$ is the conditional

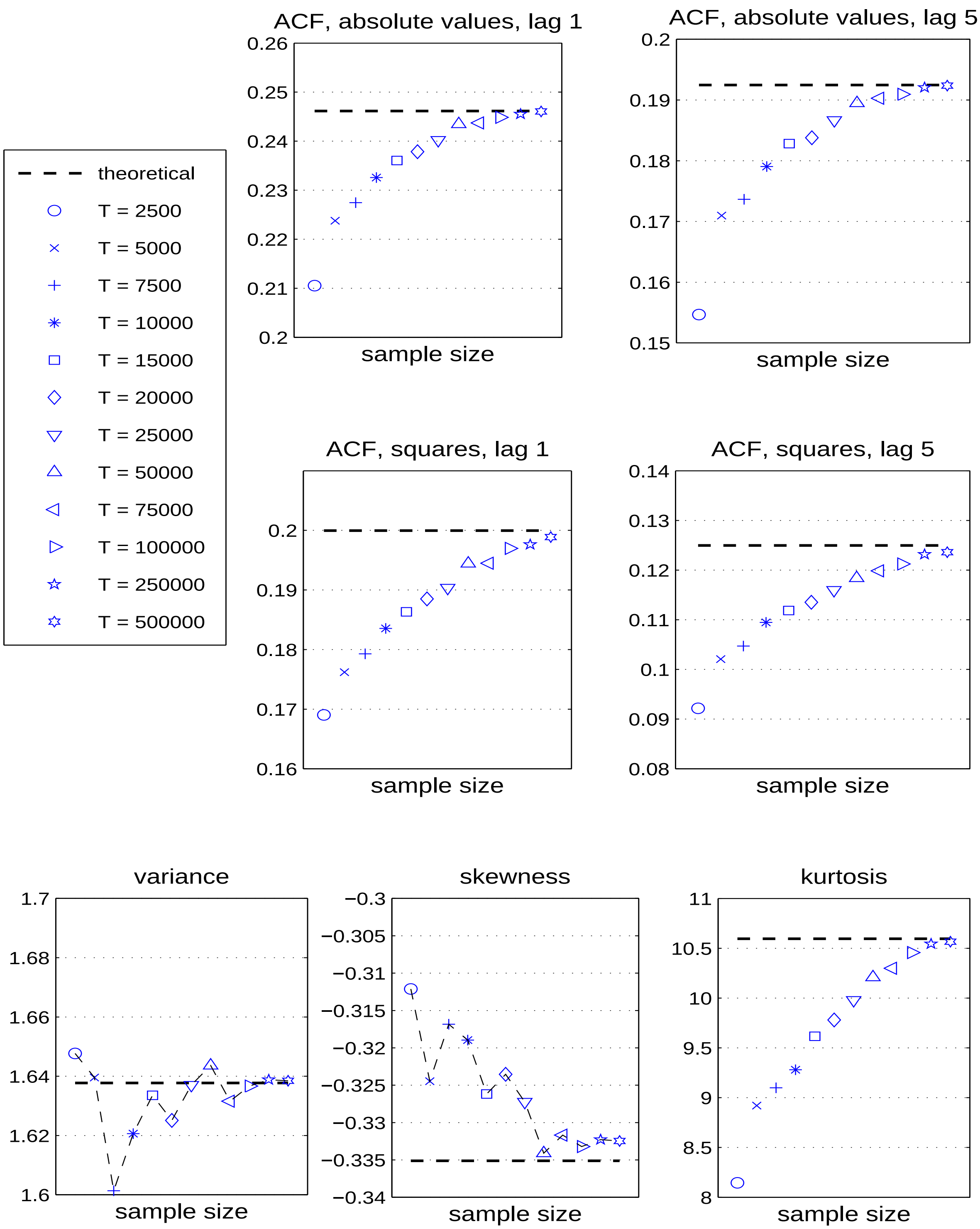

Figure 4: For the two–regime skew–normal MS–GARCH process fitted to the DAX returns, the figure shows various moments implied by the estimated parameter values (dashed lines) along with their expected finite–sample counterparts (symbols) for various sample sizes. The finite–sample expectations are calculated by averaging over 1000 simulation runs.

cdf of the return implied by the model under consideration. Subsequently, we apply a second transformation, namely,

$$\{z_t\} = \Phi^{-1}(\{u_t\}), \tag{16}$$

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 (16), and thus this transformation 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,

$$T\widehat{m}_3^2/6 \overset{asy}{\sim} \chi^2(1), \quad \text{and} \quad T(\widehat{m}_4 - 3)^2/24 \overset{asy}{\sim} \chi^2(1), \tag{17}$$

where $\widehat{m}_3$ and $\widehat{m}_4$ are the sample skewness and kurtosis of $\{z_t\}$, respectively, and $T = 2020$ is the number of forecasts evaluated. A joint test for correctly specified skewness and kurtosis is provided by the Jarque–Bera (JB) test,

$$JB = T\widehat{m}_3^2/6 + T(\widehat{m}_4 - 3)^2/24 \overset{asy}{\sim} \chi^2(2). \tag{18}$$

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 forecasts densities which can be accomplished by means of the skewed exponential power (SEP) distribution of Fernandez, Osiewalski, and Steel (1995), with density

$$f(z;\mu,\sigma,p,\theta) = \frac{\theta}{1+\theta^2}\frac{p}{\sigma 2^{1/p}\Gamma(1/p)}\begin{cases}\exp\left\{-\frac{1}{2}\left(\frac{|z-\mu|\theta}{\sigma}\right)^p\right\}, & \text{if} \quad z < \mu\\ \exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma\theta}\right)^p\right\}, & \text{if} \quad z \geq \mu,\end{cases} \tag{19}$$

where $\theta, p > 0$. 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 (16) 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 (19) and test the restrictions $\theta = 1$ and $p = 2$, respectively, with the corresponding LRTs again having an asymptotic $\chi^2(1)$ distribution. Finally, we 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 $p = 2$.

The second type 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 $\text{VaR}_t(\xi)$, is implicitly defined by $\widehat{F}(\text{VaR}_t(\xi)|\Psi_{t-1}) = \xi$. 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

$$I_t = \begin{cases} 1, & \text{if} \quad r_t < \mathrm{VaR}_t, \\ 0, & \text{if} \quad r_t \geq \mathrm{VaR}_t. \end{cases} \tag{20}$$

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)

$$\mathrm{LRT}_{\mathrm{VaR}} = -2\{x \log(\xi/\widehat{\xi}) + (T - x)\log[(1-\xi)/(1-\widehat{\xi})]\} \overset{asy}{\sim} \chi^2(1). \tag{21}$$

Equation (21) 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 $\xi = 0.0025, 0.005, 0.01, 0.025$ and 0.05, 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 reprted in Tables 4–9. Tables 4–6 provide the results for backtests of the entire predictive densities, and Tables 7-9 do the same for the VaR.

Concerning the results in Tables 4–6, we first note that there is no case where the residuals of any of the mixture models show significant positive excess kurtosis. The sole occasion where the hypothesis of a correctly specified kurtosis is rejected for a member of this family is the test based on the standardized fourth moment applied to the symmetric Gaussian mixture GARCH process (Mix_S) for the DAX 30 (Table 5). In this case the model suffers from platykurtosis, however. Thus, it appears that fat–tailed conditional within–regime distributions are indeed superfluous. In contrast, the results for the single–regime models based on light–tailed innovations show significant excess kurtosis for all indices, which is a well–known phenomenon.

Regarding the symmetry tests, we observe striking differences between the mixture models allowing for skewness and those enforcing symmetry of the conditional density. In particular,

Table 4: Evaluation of forecast densities: CAC 40

Model	Moment–based Skew.	Kurt.	JB	Berkowitz LRT μ	σ^2	θ	p	joint
Normal								
Single	-0.289^{***}	3.62^{***}	60.3^{***}	-0.052^{**}	0.997	0.882^{***}	1.77^{***}	30.7^{***}
$\text{Mix}_S^{\text{icept}}$	-0.179^{***}	3.03	10.9^{***}	-0.053^{**}	0.991	0.896^{***}	2.02	16.5^{***}
Mix_S	-0.195^{***}	3.02	12.8^{***}	-0.052^{**}	0.991	0.887^{***}	2.07	18.1^{***}
$\text{Mix}^{\text{icept}}$	-0.055	3.02	1.1	-0.048^{**}	0.989	0.970	2.02	5.7
Mix	-0.030	3.04	0.4	-0.044^{**}	0.989	0.988	2.00	4.3
$\text{Markov}^{\text{icept}}$	-0.166^{***}	3.05	9.5^{***}	-0.047^{**}	0.975	0.904^{***}	1.98	14.5^{***}
Markov	-0.172^{***}	3.07	10.3^{***}	-0.047^{**}	0.981	0.902^{***}	1.97	14.7^{***}
Skew–normal								
Single	-0.080	3.37^{***}	13.6^{***}	-0.049^{**}	0.991	0.974	1.81^{**}	10.8^{**}
$\text{Mix}^{\text{icept}}$	-0.054	3.02	1.0	-0.049^{**}	0.991	0.973	1.99	5.6
Mix	-0.064	2.97	1.5	-0.047^{**}	0.987	0.967	2.06	5.8
$\text{Markov}^{\text{icept}}$	-0.082	3.03	2.3	-0.046^{**}	0.974	0.954	1.99	7.1
Markov	-0.086	3.05	2.7	-0.046^{**}	0.979	0.953	1.97	7.0
Student's t								
symmetric	-0.182^{***}	2.97	11.2^{***}	-0.053^{**}	0.995	0.889^{***}	2.08	17.7^{***}
skewed	-0.055	2.99	1.0	-0.044^{**}	0.985	0.975	2.02	4.7

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)–(3). The abbreviations for the different multi–regime models are detailed in Table 1. All the tests are based on the residuals defined in (16), 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 (18). Significance of these is assessed on the basis of the distributional results in (17) and (18).

The Berkowitz LRT tests are based on the SEP distribution defined in (19); 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 (19) 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 (19) 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.

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

Table 5: Evaluation of forecast densities: DAX 30

	Moment–based			Berkowitz LRT				
Model	Skew.	Kurt.	JB	μ	σ^2	θ	p	joint
Normal								
Single	-0.252^{***}	3.33^{***}	30.5^{***}	-0.041^{*}	1.005	0.887^{***}	1.74^{***}	28.6^{***}
$\text{Mix}_S^{\text{icept}}$	-0.169^{***}	2.85	11.6^{***}	-0.056^{**}	1.008	0.895^{***}	2.06	18.1^{***}
Mix_S	-0.164^{***}	2.82^{*}	11.9^{***}	-0.054^{**}	1.010	0.895^{***}	2.08	17.8^{***}
$\text{Mix}^{\text{icept}}$	−0.018	2.84	2.3	-0.045^{**}	1.005	0.985	2.06	4.7
Mix	−0.009	2.82	2.6	-0.043^{*}	1.009	0.990	2.07	4.3
$\text{Markov}^{\text{icept}}$	-0.179^{***}	2.90	11.7^{***}	-0.053^{**}	1.009	0.893^{***}	1.99	18.1^{***}
Markov	-0.186^{***}	2.90	12.6^{***}	-0.053^{**}	1.004	0.888^{***}	1.99	19.3^{***}
Skew–normal								
Single	0.028	3.19^{*}	3.4	-0.037^{*}	1.008	1.010	1.78^{**}	8.3^{*}
$\text{Mix}^{\text{icept}}$	0.044	2.86	2.3	-0.048^{**}	1.012	1.025	2.00	5.3
Mix	0.028	2.84	2.5	-0.047^{**}	1.013	1.016	2.05	5.1
$\text{Markov}^{\text{icept}}$	0.015	2.89	1.1	-0.048^{**}	1.014	1.007	1.99	5.0
Markov	−0.001	2.89	1.0	-0.049^{**}	1.006	0.997	2.00	4.9
Student's t								
symmetric	-0.178^{***}	2.80^{*}	13.9^{***}	-0.053^{**}	1.005	0.884^{***}	2.16	20.5^{***}
skewed	−0.036	2.88	1.7	−0.029	1.000	0.985	1.97	2.0

Shown are the results of tests for correctly specified one–step–ahead predictive densities for the DAX 30. See the legend of Table 4 for further explanations.

Table 6: Evaluation of forecast densities: FTSE 100

Model	Moment–based Skew.	Kurt.	JB	Berkowitz LRT μ	σ^2	θ	p	joint
Normal								
Single	-0.346^{***}	3.53^{***}	64.2^{***}	-0.050^{**}	1.008	0.858^{***}	1.78^{**}	38.6^{***}
$\text{Mix}_S^{\text{icept}}$	-0.242^{***}	3.02	19.8^{***}	-0.053^{**}	1.006	0.868^{***}	1.98	24.8^{***}
Mix_S	-0.242^{***}	2.97	19.7^{***}	-0.052^{**}	1.003	0.864^{***}	2.02	24.7^{***}
$\text{Mix}^{\text{icept}}$	-0.133^{**}	3.01	5.9^{*}	-0.047^{**}	1.004	0.927^{**}	1.98	9.9^{**}
Mix	-0.133^{**}	3.09	6.7^{**}	-0.043^{*}	1.002	0.935^{**}	1.96	8.3^{*}
$\text{Markov}^{\text{icept}}$	-0.243^{***}	3.01	19.9^{***}	-0.050^{**}	0.984	0.865^{***}	2.01	24.5^{***}
Markov	-0.251^{***}	3.01	21.2^{***}	-0.051^{**}	0.994	0.860^{***}	2.02	25.6^{***}
Skew–normal								
Single	-0.150^{***}	3.34^{***}	17.5^{***}	-0.048^{**}	1.007	0.942^{*}	1.81^{**}	13.4^{***}
$\text{Mix}^{\text{icept}}$	-0.107^{**}	2.99	3.8	-0.049^{**}	1.007	0.945^{*}	1.98	8.1^{*}
Mix	-0.108^{**}	2.94	4.2	-0.048^{**}	1.003	0.942^{*}	2.01	8.0^{*}
$\text{Markov}^{\text{icept}}$	-0.144^{***}	3.01	7.0^{**}	-0.049^{**}	0.985	0.923^{**}	1.99	10.9^{**}
Markov	-0.143^{***}	2.99	6.9^{**}	-0.048^{**}	0.995	0.924^{**}	2.00	10.6^{**}
Student's t								
symmetric	-0.243^{***}	2.99	19.9^{***}	-0.052^{**}	1.004	0.860^{***}	2.06	25.3^{***}
skewed	-0.108^{**}	2.97	4.0	-0.037^{*}	0.997	0.945^{*}	2.01	5.8

Shown are the results of tests for correctly specified one–step–ahead predictive densities for the FTSE 100. See the legend of Table 4 for further explanations.

the symmetric mixtures are characterized by significant negative skewness for all indices. The asymmetric mixtures capture the skewness successfully for the CAC 40 and the DAX 30, but, although still superior to their symmetric companions, don't pass the symmetry tests in case of the FTSE 100. This is very likely due to the fact that the skewness of the FTSE 100 is not stable over time, as significant skewness is not there in the in–sample period. The corresponding statistics for the skewed t model (15) also indicate significant asymmetries in case of the FTSE 100, which points to the same direction.

Finally, whereas all the models describe the volatility in a satisfactory manner, the hypothesis of a correctly specified mean, i.e., $\mu = 0$, is rejected in most cases. This, however, simply reflects the well–known fact that asset returns are very hard to predict and therefore may not be seen as a serious problem for the models put to test in this paper.

The results of the VaR backtests reported in Tables 7–9 carry a similar message as Tables 4–6. The performance of the symmetric models is by far outperformed by the asymmetric specifications, which deliver rather accurate VaR measures, including, somewhat surprisingly, the single–regime skew–normal GARCH(1,1) model.

Summarizing, we conclude that the skew–normal MS–GARCH process helps to cure the weaknesses of the Gaussian MS–GARCH model without sacrifying its attractive properties as discussed in Section 1.

4 Conclusions

In this paper, we propose and motivate the use of Azzalini's (1985) skew–normal distribution in the framework of Markov–switching 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 shows that the model fits the data well and delivers reliable out–of–sample risk measures.

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 dimensions, optimization of the likelihood will become burdensome, however, and the development of Bayesian estimation routines may be called for.

Table 7: Evaluation of Value–at–Risk (VaR) measures: CAC 40

	Long positions					Short positions				
$100 \times \xi$	0.25	0.5	1	2.5	5	0.25	0.5	1	2.5	5
Normal										
Single	0.594***	0.842**	1.782***	3.564***	5.990**	0.198	0.347	0.644*	1.881*	3.713***
$\text{Mix}_S^{\text{icept}}$	0.347	0.495	1.485**	3.515***	6.139**	0.149	0.149***	0.545**	1.881*	3.713***
Mix_S	0.347	0.594	1.238	3.366**	6.287**	0.000***	0.149***	0.446***	1.683**	3.812**
$\text{Mix}^{\text{icept}}$	0.347	0.396	0.990	2.970	5.693	0.149	0.446	0.743	2.079	4.109*
Mix	0.297	0.396	0.842	2.525	5.495	0.198	0.495	0.842	2.376	4.257
$\text{Markov}^{\text{icept}}$	0.347	0.644	1.238	3.168*	6.188**	0.050**	0.198**	0.594**	1.733**	3.812**
Markov	0.396	0.693	1.238	3.218**	6.188**	0.050**	0.297	0.693	1.733**	3.713***
Skew–normal										
Single	0.347	0.594	1.188	2.871	5.347	0.248	0.495	0.990	2.178	4.010**
$\text{Mix}^{\text{icept}}$	0.297	0.495	1.139	3.069	5.495	0.149	0.347	0.693	2.030	4.059**
Mix	0.248	0.396	1.040	2.871	5.495	0.099	0.297	0.693	2.129	4.257
$\text{Markov}^{\text{icept}}$	0.297	0.446	1.040	3.020	5.693	0.050**	0.297	0.842	1.980	4.208*
Markov	0.347	0.446	0.990	3.119*	5.594	0.050**	0.396	0.842	2.030	4.208*
Student's t										
symmetric	0.347	0.545	1.485**	3.366**	6.139**	0.000***	0.149***	0.545**	1.733**	3.713***
skewed	0.347	0.446	0.990	2.574	5.446	0.149	0.347	0.644*	1.980	4.109*

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)–(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 (21).

Table 8: Evaluation of Value–at–Risk (VaR) measures: DAX 30

	Long positions					Short positions				
$100 \times \xi$	0.25	0.5	1	2.5	5	0.25	0.5	1	2.5	5
Normal										
Single	0.545**	0.792*	1.485**	3.465***	6.634***	0.099	0.248*	0.495**	1.683**	4.158*
$\text{Mix}_S^{\text{icept}}$	0.198	0.446	1.188	3.614***	6.980***	0.050**	0.099***	0.495**	1.436***	4.109*
Mix_S	0.149	0.446	1.188	3.663***	7.030***	0.050**	0.099***	0.396***	1.436***	4.158*
$\text{Mix}^{\text{icept}}$	0.099	0.248*	0.693	2.970	5.743	0.149	0.347	0.644*	2.228	4.950
Mix	0.099	0.248*	0.842	3.020	5.792	0.149	0.297	0.792	2.475	4.901
$\text{Markov}^{\text{icept}}$	0.198	0.792*	1.238	3.465***	6.881***	0.050**	0.149***	0.495**	1.782**	4.257
Markov	0.198	0.743	1.337	3.515***	6.832***	0.050**	0.149***	0.495**	1.782**	4.109*
Skew–normal										
Single	0.297	0.545	0.990	2.723	5.297	0.248	0.495	0.891	2.673	5.099
$\text{Mix}^{\text{icept}}$	0.099	0.297	0.990	2.525	5.743	0.149	0.396	0.743	2.624	5.297
Mix	0.099	0.347	0.941	2.574	5.743	0.050**	0.347	0.693	2.426	5.099
$\text{Markov}^{\text{icept}}$	0.198	0.297	0.990	2.871	5.990**	0.198	0.347	0.842	2.624	4.950
Markov	0.198	0.347	1.040	2.723	6.040**	0.099	0.347	0.693	2.426	4.901
Student's t										
symmetric	0.347	0.495	1.040	3.515***	6.881***	0.000***	0.149***	0.347***	1.337***	4.109*
skewed	0.198	0.495	0.842	2.574	5.941*	0.050**	0.198**	0.545**	2.574	5.050

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 7 for further explanations.

Table 9: Evaluation of Value–at–Risk (VaR) measures: FTSE 100

	Long positions					Short positions				
$100 \times \xi$	0.25	0.5	1	2.5	5	0.25	0.5	1	2.5	5
Normal										
Single	0.792***	0.990***	1.782***	3.762***	6.287**	0.099	0.149***	0.495**	1.634***	3.762***
$\text{Mix}_S^{\text{icept}}$	0.545**	0.891**	1.634***	3.960***	6.683***	0.050**	0.099***	0.297***	1.584***	3.762***
Mix_S	0.396	0.842**	1.584**	3.762***	6.832***	0.000***	0.099***	0.248***	1.436***	3.911**
$\text{Mix}^{\text{icept}}$	0.495*	0.693	1.337	3.515***	5.990**	0.099	0.149***	0.545**	1.931*	4.356
Mix	0.248	0.545	1.238	3.317**	5.891*	0.099	0.149***	0.495**	2.030	4.505
$\text{Markov}^{\text{icept}}$	0.446	0.792*	1.584**	3.515***	6.386***	0.050**	0.099***	0.396***	1.337***	3.564***
Markov	0.495*	0.792*	1.683***	3.663***	6.436***	0.050**	0.050***	0.347***	1.436***	3.812**
Skew–normal										
Single	0.495*	0.842**	1.238	3.465***	5.792	0.099	0.198**	0.644*	2.129	4.257
$\text{Mix}^{\text{icept}}$	0.446	0.644	1.238	3.515***	5.990**	0.099	0.099***	0.644*	2.079	4.505
Mix	0.297	0.644	1.287	3.515***	6.089**	0.099	0.149***	0.545**	2.129	4.554
$\text{Markov}^{\text{icept}}$	0.347	0.693	1.386*	3.317**	6.040**	0.099	0.099***	0.495**	1.881*	4.257
Markov	0.347	0.743	1.436*	3.465***	5.990**	0.050**	0.149***	0.495**	1.980	4.455
Student's t										
symmetric	0.495*	0.792*	1.485**	3.762***	6.535***	0.099	0.099***	0.198***	1.485***	3.713***
skewed	0.198	0.693	1.139	3.218**	6.040**	0.099	0.099***	0.495**	1.980	4.604

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 7 for further explanations.

Appendix

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

As the SN(γ)–MS(k)–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 (5).

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

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

where

$$
\begin{aligned}
E_\ell^+(\gamma) &= 2\int_m^\infty x^\ell \phi(x)\Phi(\gamma x)dx, \\
E_\ell^-(\gamma) &= 2\int_{-\infty}^m x^\ell \phi(x)\Phi(\gamma x)dx.
\end{aligned}
$$

We observe that

$$E_\ell^-(\gamma) = (-1)^\ell E_\ell^+(-\gamma),$$

so it is sufficient to calculate $E_\ell^+(\gamma)$. To this end, we note that, using $\phi'(x) = -x\phi(x)$,

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

where

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

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

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

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

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

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 $\mathrm{E}(|Z|^3) = 2\sqrt{2/\pi}$.

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

In this appendix, we state the conditions for the process (1)–(3) and (13) 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

$$\boldsymbol{X}_t = \begin{pmatrix} \sigma_{1t} \\ \sigma_{2t} \\ \vdots \\ \sigma_{kt} \\ |\epsilon_{t-1}| \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \\ 0 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 & 0 & \cdots & 0 & 0 \\ 0 & \beta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta_k & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{B.7}$$

and let $\boldsymbol{e}_j$ be the jth unit vector in $\mathbb{R}^{k+1}$, i.e., the jth column of $\boldsymbol{I}_{k+1}$, denoting the identity matrix of dimension $k+1$. Then the process can be written as a first–order stochastic vector difference equation,

$$\boldsymbol{X}_t = \boldsymbol{\omega} + \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1}, \tag{B.8}$$

where $\boldsymbol{C}_{\Delta_{t-1},t-1} = \boldsymbol{\alpha}\boldsymbol{e}'_{\Delta_{t-1}}|\eta_{t-1}| + \boldsymbol{\beta}$ 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

$$\mathbb{P}_f = \begin{pmatrix} p_{11}f(1) & \cdots & p_{k1}f(1) \\ \vdots & \cdots & \vdots \\ p_{1k}f(k) & \cdots & p_{kk}f(k) \end{pmatrix}. \tag{B.9}$$

Moreover, let

$$\boldsymbol{C}_{mm}(j) = \mathrm{E}(\boldsymbol{C}_{jt}^{\otimes m}), \quad j = 1, \ldots, k, \tag{B.10}$$

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

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

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.9) and (B.10), then the SN–MS–GARCH(k) process defined by (1)–(3) and (13) has a unique strictly stationary solution with finite mth moment.*

To check the conditions provided by Theorem 1, we have to calculate the quantities $\boldsymbol{C}_{mm}(j)$ defined in (B.10) for $m = 1, \ldots, 4$, which is straightforward, although tedious. For example,

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

where $\kappa_i(\gamma)$ is the ith absolute moment of a centered skew–normal random variable with skewness parameter γ. For i odd, these moments have been calculated in Appendix A, with $\kappa_1(\gamma)$ and $\kappa_3(\gamma)$ given in (A.5) and (A.6), respectively, while for i even they follow from the results in the literature, and those relevant for the moments up to order four a given by

$$\kappa_2(\gamma) = 1 - \frac{2}{\pi}\delta^2, \quad \text{and} \quad \kappa_4(\gamma) = 3 - \frac{12}{\pi}\delta^2 + \frac{8\pi - 12}{\pi^2}\delta^4.$$

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

Using Lemma B.1 and the fact that $\boldsymbol{X}_t$ belongs to the information set generated by $\Psi_{t-1} = \{\epsilon_s : s \leq t-1\}$, we have

$$\pi_{j,\infty}\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1} = j) = \pi_{j,\infty}\boldsymbol{\omega} + \sum_{i=1}^{k} p_{ij}\boldsymbol{C}_{11}(j)\pi_{i,\infty}\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-2} = i), \quad j = 1, \ldots, k,$$

so that

$$\boldsymbol{Q}_1 = \boldsymbol{\pi}_\infty \otimes \boldsymbol{\omega} + \mathbb{P}_{C_{11}}\boldsymbol{Q}_1,$$

where the $k(k+1) \times 1$ vector $\boldsymbol{Q}_1$ is defined as

$$\boldsymbol{Q}_1 = (\pi_{1,\infty}\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1} = 1)', \ldots, \pi_{k,\infty}\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1} = k)')'. \tag{B.11}$$

Then $\mathrm{E}(\boldsymbol{X}_t) = \sum_j \pi_{j,\infty}\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1} = j) = (\boldsymbol{1}_k' \otimes \boldsymbol{I}_{k+1})\boldsymbol{Q}_1$, where $\boldsymbol{1}_k$ is a k–dimensional column of ones, and $\mathrm{E}(|\epsilon_t|) = \boldsymbol{e}_{k+1}'\mathrm{E}(\boldsymbol{X}_t)$. Similarly, we obtain

$$\boldsymbol{Q}_2 = \boldsymbol{\pi}_\infty \otimes \boldsymbol{\omega} \otimes \boldsymbol{\omega} + \mathbb{P}_{C_{21}}\boldsymbol{Q}_1 + \mathbb{P}_{C_{22}}\boldsymbol{Q}_2,$$

where $\boldsymbol{Q}_2$ is the $k(k+1)^2 \times 1$ vector defined analogously to $\boldsymbol{Q}_1$ but with vector elements $\pi_{j,\infty}\mathrm{E}[\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')|\Delta_{t-1} = j]$ instead of $\pi_{j,\infty}\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1} = j)$, $j = 1, \ldots, k$, and $\mathbb{P}_{C_{21}}$ is obtained by applying definition (B.9) to

$$\boldsymbol{C}_{21}(j) = \boldsymbol{\omega} \otimes \boldsymbol{C}_{11}(j) + \boldsymbol{C}_{11}(j) \otimes \boldsymbol{\omega}, \quad j = 1, \ldots, k. \tag{B.12}$$

Expressions for $\mathrm{E}[\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')]$ and thus $\mathrm{E}(\epsilon_t^2)$ are readily deduced. In the same manner, by expanding $\boldsymbol{X}_t^{\otimes m} = (\boldsymbol{\omega} + \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1})^{\otimes m}$ in Kronecker powers of $\boldsymbol{X}_t$, we get $\boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$, defined by replacing $\pi_{j,\infty}\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1} = j)$ in $\boldsymbol{Q}_1$ with $\pi_{j,\infty}\mathrm{E}\{\mathrm{vec}[\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')\boldsymbol{X}_t']|\Delta_{t-1} = j\}$ and $\pi_{j,\infty}\mathrm{E}\{\mathrm{vec}[\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')']|\Delta_{t-1} = j\}$ in $\boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$, respectively, $j = 1, \ldots, k$, which in turn can be used to calculate $\mathrm{E}(|\epsilon_t|^3)$ and $\mathrm{E}(\epsilon_t^4)$. The details are lengthy and tedious, however, and are available from the author on request. A Matlab program for doing all the calculations can also be provided.

Once we have $\boldsymbol{Q}_3$, the unconditional *raw* (as opposed to absolute) third moment of the process is also obtained by use of Lemma B.1. We have

$$\begin{aligned}
\mathrm{E}(\epsilon_t^3) &= \sum_{j=1}^{k} \pi_{j,\infty}\mathrm{E}(\epsilon_t^3|\Delta_t = j) = \sum_{j=1}^{k} \pi_{j,\infty}\mathrm{E}(\eta_t^3\sigma_{jt}^3|\Delta_t = j) \\
&= \sqrt{\frac{2}{\pi}}\frac{4-\pi}{\pi}\delta^3 \sum_{j=1}^{k}\sum_{i=1}^{k} p_{ij}\pi_{i,\infty}\mathrm{E}(\sigma_{jt}^3|\Delta_{t-1} = i) \\
&= \sqrt{\frac{2}{\pi}}\frac{4-\pi}{\pi}\delta^3 \sum_{j=1}^{k}(\boldsymbol{p}_{\bullet j} \otimes \boldsymbol{e}_j' \otimes \boldsymbol{e}_j' \otimes \boldsymbol{e}_j')\boldsymbol{Q}_3, \qquad \text{(B.13)}
\end{aligned}$$

where $\boldsymbol{p}_{\bullet j}$ is the jth row of the transition matrix $\boldsymbol{P}$, $j = 1, \ldots, k$, and the unconditional skewness follows.

B.1 Autocorrelation structure

To derive the autocorrelation structure of both the absolute and the squared process, define

$$\begin{aligned}
\boldsymbol{Y}_t &= \begin{pmatrix} \boldsymbol{X}_t \\ \mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t') \end{pmatrix}, \quad \tilde{\boldsymbol{\omega}} = \begin{pmatrix} \boldsymbol{\omega} \\ \boldsymbol{\omega} \otimes \boldsymbol{\omega} \end{pmatrix}, \\
\tilde{\boldsymbol{C}}_{jt} &= \begin{pmatrix} \boldsymbol{C}_{jt} & \boldsymbol{0}_{(k+1)\times(k+1)^2} \\ \boldsymbol{\omega} \otimes \boldsymbol{C}_{jt} + \boldsymbol{C}_{jt} \otimes \boldsymbol{\omega} & \boldsymbol{C}_{jt} \otimes \boldsymbol{C}_{jt} \end{pmatrix}, \quad j = 1, \ldots, k,
\end{aligned}$$

so that $\boldsymbol{Y}_t = \tilde{\boldsymbol{\omega}} + \tilde{\boldsymbol{C}}_{\Delta_{t-1},t-1}\boldsymbol{Y}_{t-1}$. Multiplying by $\boldsymbol{Y}_{t-\tau}'$, $\tau \geq 1$, and applying Lemma B.1,

$$\begin{aligned}
\pi_{j,\infty}\mathrm{E}(\boldsymbol{Y}_t\boldsymbol{Y}_{t-\tau}'|\Delta_{t-1} = j) &= \tilde{\boldsymbol{\omega}}\sum_{i=1}^{k} p_{ij}^{(\tau)}\pi_{i,\infty}\mathrm{E}(\boldsymbol{Y}_t'|\Delta_{t-1} = i) \qquad \text{(B.14)} \\
&\quad + \sum_{i=1}^{k} p_{ij}\tilde{\boldsymbol{C}}(j)\pi_{i,\infty}\mathrm{E}(\boldsymbol{Y}_{t-1}\boldsymbol{Y}_{t-\tau}'|\Delta_{t-2} = i), \quad j = 1, \ldots, k,
\end{aligned}$$

where

$$\tilde{\boldsymbol{C}}(j) = \begin{pmatrix} \boldsymbol{C}_{11}(j) & \boldsymbol{0}_{(k+1)\times(k+1)^2} \\ \boldsymbol{C}_{21}(j) & \boldsymbol{C}_{22}(j) \end{pmatrix}, \quad j = 1, \ldots, k. \tag{B.15}$$

Now let $\boldsymbol{R}$ and $\boldsymbol{S}(\tau)$, $\tau = 0, 1, \ldots$, be the $k \times (k+1)(k+2)$ and $k(k+1)(k+2) \times (k+1)(k+2)$ matrices, respectively, where, in this order, the vectors $\pi_{j,\infty}\mathrm{E}(\boldsymbol{Y}_t'|\Delta_{t-1} = j)$ and the matrices $\pi_{j,\infty}\mathrm{E}(\boldsymbol{Y}_t\boldsymbol{Y}_{t-\tau}'|\Delta_{t-1} = j)$, $j = 1, \ldots, k$, are arranged one underneath the other. Then $\boldsymbol{R}$ and $\boldsymbol{S}(0)$ can be obtained straightforwardly by recombining the elements of $\boldsymbol{Q}_i$, $i = 1, \ldots, 4$, and from (B.14) we get the recursion

$$\boldsymbol{S}(\tau) = (\boldsymbol{P}^\tau \otimes \tilde{\boldsymbol{\omega}})\boldsymbol{R} + \mathbb{P}_{\tilde{C}}\boldsymbol{S}(\tau - 1), \quad \tau \geq 1, \tag{B.16}$$

where $\mathbb{P}_{\tilde{C}}$ is defined via (B.9) and (B.15). We finally calculate $\mathrm{E}(|\epsilon_t||\epsilon_{t-\tau}|) = \boldsymbol{e}_{k+1}'\mathrm{E}(\boldsymbol{X}_t\boldsymbol{X}_{t-\tau}')\boldsymbol{e}_{k+1}$, where $\mathrm{E}(\boldsymbol{X}_t\boldsymbol{X}_{t-\tau}')$ is the upper left $(k+1)\times(k+1)$ block of $\mathrm{E}(\boldsymbol{Y}_t\boldsymbol{Y}_{t-\tau}') = (\boldsymbol{1}_k'\otimes\boldsymbol{I}_{(k+1)(k+2)})\boldsymbol{S}(\tau)$, and $\mathrm{E}(\epsilon_t^2\epsilon_{t-\tau}^2)$ follows analogously. The autocorrelations are then given by

$$\begin{aligned} \varrho_1(\tau) &= \mathrm{Corr}(|\epsilon_t|, |\epsilon_{t-\tau}|) = \frac{\mathrm{Cov}(|\epsilon_t|, |\epsilon_{t-\tau}|)}{\mathrm{Var}(|\epsilon_t|)} = \frac{\mathrm{E}(|\epsilon_t||\epsilon_{t-\tau}|) - \mathrm{E}^2(|\epsilon_t|)}{\mathrm{E}(\epsilon_t^2) - \mathrm{E}^2(|\epsilon_t|)}, && \text{(B.17)} \\ \varrho_2(\tau) &= \mathrm{Corr}(\epsilon_t^2, \epsilon_{t-\tau}^2) = \frac{\mathrm{Cov}(\epsilon_t^2, \epsilon_{t-\tau}^2)}{\mathrm{Var}(\epsilon_t^2)} = \frac{\mathrm{E}(\epsilon_t^2\epsilon_{t-\tau}^2) - \mathrm{E}^2(\epsilon_t^2)}{\mathrm{E}(\epsilon_t^4) - \mathrm{E}^2(\epsilon_t^2)}. && \text{(B.18)} \end{aligned}$$

C Alternative expressions for the odd moments of the SN distribution

In this appendix, we derive equation (6) for the odd moments of an SN random variable and relate it to earlier expressions presented in the literature. Azzalini (1985) showed that the moment generating function (mgf) associated with (5) is given by $m(t) = 2e^{t^2/2}\Phi(\delta t)$, where $\delta = \gamma/\sqrt{1+\gamma^2}$, i.e.,

$$\begin{aligned} m(t) &= 2e^{t^2/2}\Phi(\delta t) = e^{t^2/2} + \sqrt{\frac{2}{\pi}}e^{t^2/2}\int_0^{\delta t}\sum_{i=0}^{\infty}\frac{(-1)^i x^{2i}}{i!2^i}dx \\ &= e^{t^2/2} + \sqrt{\frac{2}{\pi}}e^{t^2/2}\sum_{i=0}^{\infty}(-1)^i\frac{\delta^{2i+1}t^{2i+1}}{2^i i!(2i+1)} \\ &= e^{t^2/2} + \sqrt{\frac{2}{\pi}}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^i\delta^{2i+1}t^{2(i+j)+1}}{2^{i+j}i!j!(2i+1)} \\ &\overset{\ell=i+j}{=} e^{t^2/2} + \sqrt{\frac{2}{\pi}}\sum_{\ell=0}^{\infty}t^{2\ell+1}\left\{\sum_{i=0}^{\ell}\frac{(-1)^i\delta^{2i+1}}{2^\ell i!(\ell-i)!(2i+1)}\right\}, \end{aligned}$$

which shows that the mgf can be decomposed into a Gaussian part (accounting for the even moments), and a further term determining the odd moments. Thus,

$$\begin{aligned}
\mathrm{E}(Z^{2\ell+1}) &= \sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^\ell}\sum_{i=0}^{\ell}\frac{(-1)^i\delta^{2i+1}}{i!(\ell-i)!(2i+1)} \\
&= \sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^\ell\ell!}\sum_{i=0}^{\ell}\binom{\ell}{i}(-1)^i\frac{\delta^{2i+1}}{2i+1} = \sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^\ell\ell!}\sum_{i=0}^{\ell}\binom{\ell}{i}(-1)^i\int_0^\delta x^{2i}dx \\
&= \sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^\ell\ell!}\int_0^\delta(1-x^2)^\ell dx \overset{y=x^2}{=} \mathrm{sign}(\delta)\sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^{\ell+1}\ell!}\int_0^{\delta^2}y^{-1/2}(1-y)^\ell dy \\
&= \mathrm{sign}(\delta)\sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^{\ell+1}\ell!}\mathrm{B}(1/2,\ell+1)\mathrm{B}^{\mathrm{inc}}(\delta^2;1/2,\ell+1) \\
&= \mathrm{sign}(\delta)\frac{2^{\ell+1/2}\ell!}{\sqrt{\pi}}\mathrm{B}^{\mathrm{inc}}(\delta^2;1/2,\ell+1), \quad \ell=0,1,\ldots,
\end{aligned}$$

where $\mathrm{B}(a,b)=\int_0^1 y^{a-1}(1-y)^{b-1}dy$ denotes the beta function,

$$\mathrm{B}^{\mathrm{inc}}(x;a,b) = \frac{1}{\mathrm{B}(a,b)}\int_0^x y^{a-1}(1-y)^{b-1}dy \tag{C.19}$$

is the incomplete beta function, and the last line follows from $\mathrm{B}(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ and $\Gamma(\ell+3/2)=\sqrt{\pi}2^{-(\ell+1)}(2\ell+1)(2\ell-1)\cdots5\cdot3\cdot1$.

The first general expression for the odd moments of the SN distribution was given by Henze (1986), who showed that an SN random variable Z has a representation $Z=\delta|U|+\sqrt{1-\delta^2}V$, where U and V are independent standard normal variables. Then he used the binomial formula to obtain

$$\mathrm{E}(Z^{2\ell+1}) = \sqrt{\frac{2}{\pi}}\frac{(2\ell+1)!}{2^\ell}\frac{\gamma}{(1+\gamma^2)^{\ell+1/2}}\sum_{m=0}^{\ell}\frac{m!(2\gamma)^{2m}}{(2m+1)!(\ell-m)!}. \tag{C.20}$$

To see that (6) and (C.20) are identical, use $\delta=\gamma/\sqrt{1+\gamma^2}$ to express the sum in (6) in terms of γ, i.e.,

$$\begin{aligned}
\sum_{i=0}^{\ell}\frac{(-1)^i\delta^{2i+1}}{i!(\ell-i)!(2i+1)} &= \frac{\gamma}{(1+\gamma^2)^{\ell+1/2}}\sum_{i=0}^{\ell}\frac{(-1)^i\gamma^{2i}(1+\gamma^2)^{\ell-i}}{i!(\ell-i)!(2i+1)} \\
&= \frac{\gamma}{(1+\gamma^2)^{\ell+1/2}}\sum_{i=0}^{\ell}\sum_{j=0}^{\ell-i}\binom{\ell-i}{j}\frac{(-1)^i\gamma^{2(i+j)}}{i!(\ell-i)!(2i+1)} \\
&\overset{m=i+j}{=} \frac{\gamma}{(1+\gamma^2)^{\ell+1/2}}\sum_{m=0}^{\ell}\left\{\frac{1}{(\ell-m)!}\sum_{i=0}^{m}\frac{(-1)^i}{i!(m-i)!(2i+1)}\right\}\gamma^{2m}.
\end{aligned}$$

Comparing coefficients, equality of (6) and (C.20) follows from the combinatoric identity

$$\sum_{i=0}^{m}\binom{m}{i}\frac{(-1)^i}{2i+1} = \frac{(2^m m!)^2}{(2m+1)!} = \frac{2^{2m}}{\binom{2m}{m}(2m+1)}, \tag{C.21}$$

which is well–known and easy to prove (see, e.g., Paolella, 2006, p. 20).

A further method for obtaining a general expression for the odd moments was proposed by Martínez et al. (2008), who observed that integration by parts directly leads to the recursive formula

$$\mathrm{E}(Z^{2\ell+1}) = 2\ell\mathrm{E}(Z^{2\ell-1}) + \sqrt{\frac{2}{\pi}}\frac{\gamma}{(1+\gamma^2)^{\ell+1/2}}\frac{(2\ell)!}{2^\ell \ell!}.$$

This gives

$$\begin{aligned}\mathrm{E}(Z^{2\ell+1}) &= \sqrt{\frac{2}{\pi}}\sum_{i=0}^{\ell}\frac{2^\ell \ell!}{2^i i!}\frac{(2i)!}{2^i i!}\frac{\gamma}{(1+\gamma^2)^{i+1/2}} \\ &= \sqrt{\frac{2}{\pi}}\frac{\gamma}{(1+\gamma^2)^{\ell+1/2}}2^\ell \ell! \sum_{m=0}^{\ell}\left\{\sum_{i=0}^{\ell-m}\binom{\ell-i}{m}\binom{2i}{i}\frac{1}{2^{2i}}\right\}\gamma^{2m}, \qquad \text{(C.22)}\end{aligned}$$

which, by comparison with (C.20), gives rise to the combinatorial identity

$$\sum_{i=0}^{\ell-m}\binom{\ell-i}{m}\binom{2i}{i}2^{2(\ell-m-i)} = \frac{\binom{2\ell+1}{\ell}\binom{\ell+1}{m+1}}{\binom{2m+1}{m}},$$

which is presumably well–known, and otherwise can easily be verified by induction. We can also proceed the other way round, i.e., use $\gamma = \delta/\sqrt{1-\delta^2}$ to express (C.20) and (C.22) in terms of δ. Doing so, and comparing coefficients with those in (6), we obtain, respectively, the identities

$$\sum_{i=0}^{m}(-1)^i\frac{\binom{\ell-i}{m-i}\binom{\ell+1}{i+1}}{\binom{2i+1}{i+1}}2^{2i} = \frac{m+1}{2m+1}\binom{\ell+1}{m+1},$$

and

$$\sum_{i=m}^{\ell}\binom{2i}{i}\binom{i}{m}2^{2(\ell-i)} = \frac{m+1}{2m+1}\binom{2\ell+1}{\ell+1}\binom{\ell+1}{m+1}.$$

References

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

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

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

Alexander, C. and Lazar, E. (2008a). Markov switching GARCH diffusion. ICMA Centre Discussion Paper in Finance 2008-01, University of Reading.

Alexander, C. and Lazar, E. (2008b). Modelling Regime Specific Stock Volatility Behaviour. Working paper, ICMA Centre at the University of Reading.

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

Ardia, D. (2008a). Bayesian Estimation of a Markov–switching Threshold Asymmetric GARCH Model with Student-t Innovations. Forthcoming in: *Econometrics Journal*.

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

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

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

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

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

Bentarzi, M. and Hamdi, F. (2008). Mixture Periodic Autoregressive Conditional Heteroskedastic Models. *Computational Statistics and Data Analysis*, 53:1–16.

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

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

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.

Cheng, X., Yu, P. L. H., and Li, W. K. (2008). On a Dynamic Mixture GARCH Model. Forthcoming in: *Journal of Forecasting*.

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., 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.

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. 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 & Data Analysis*, 52:3027–3046.

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.

Gonçalves, E., Leite, J., and Mendes-Lopes, N. (2008). A Mathematical Approach to detect the Taylor Property in TARCH Processes. Forthcoming in: *Statistics and Probability Letters*.

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.

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

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

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. (2008). Asymmetric Multivariate Normal Mixture GARCH. Forthcoming in: *Computational Statistics and Data Analysis*.

Hamilton, J. D. (2008). Regime Switching Models. In Durlauf, S. N. and Blume, L. E., editors, *New Palgrave Dictionary of Economics, 2nd edition*, volume 7, pages 53–57. Palgrave Macmillan, Houndmills.

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.

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. (2008). An Introduction to Regime Switching Time Series Models. In Andersen, T. G., Davis, R. A., Kreiß, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*. Springer, Berlin.

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.

Marcucci, J. (2005). Forecasting Stock Market Volatility with Regime–Switching GARCH Models. *Studies in Nonlinear Dynamics & 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.

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.

Paolella, M. S. (2006). *Fundamental Probability*. John Wiley & Sons, Chichester.

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.

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.

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

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

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

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

Teräsvirta, T. (2008). An Introduction to Univariate GARCH Models. In Andersen, T. G., Davis, R. A., Kreiß, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*. Springer, Berlin.

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

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.

Zhuang, X.-F. and Chan, L.-W. (2004). Volatility Forecasts in Financial Time Series with HMM–GARCH Models. In Yang, Z. R., Everson, R., and Yin, H., editors, *Intelligent Data Engineering and Automated Learning - IDEAL 2004*, volume 3177 of *Lecture Notes in Computer Science*, pages 807–812. Springer, Berlin.