MODELING FINANCIAL TIME SERIES WITH THE SKEW SLASH DISTRIBUTION

CRISTINA G. DE LA FUENTE, PEDRO GALEANO AND MICHAEL P. WIPER*

Abstract

Financial returns often present moderate skewness and high kurtosis. As a consequence, it is natural to look for a model that is flexible enough to capture these characteristics. The proposal is to undertake inference for a generalized autoregressive conditional heteroskedastic (GARCH) model, where the innovations are assumed to follow a skew slash distribution. Both classical and Bayesian inference are carried out. Simulations and a real data example illustrate the performance of the proposed methodology.

Keywords: Financial returns; GARCH model; Kurtosis; Skew slash distribution; skewness.

Universidad Carlos III de Madrid, Departamento de Estadística. E-mail addresses: cgfuente@est-econ.uc3m.es (Cristina de la Fuente), pedro.galeano@uc3m.es (Pedro Galeano), and mwiper@est-econ.uc3m.es(Michael P. Wiper).

Modeling financial time series with the skew slash distribution

Cristina G. de la Fuente, Pedro Galeano & Michael P. Wiper

Departamento de Estadística

Universidad Carlos III de Madrid

Abstract

Financial returns often present moderate skewness and high kurtosis. As a consequence, it is natural to look for a model that is flexible enough to capture these characteristics. The proposal is to undertake inference for a generalized autoregressive conditional heteroskedastic (GARCH) model, where the innovations are assumed to follow a skew slash distribution. Both classical and Bayesian inference are carried out. Simulations and a real data example illustrate the performance of the proposed methodology.

Keywords: Financial returns; GARCH model; Kurtosis; Skew slash distribution; skewness.

1   Introduction

Financial returns are often difficult to model because they can present complicated relationships with previous observations. Returns often present a structure that may be reasonably explained with conditional heteroskedastic models, such as the generalized autoregressive conditional heteroskedastic (GARCH) model, proposed by Bollerslev (1986), the exponential generalized autoregressive conditional heteroskedastic (EGARCH) model of Nelson (1991), the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model by Glosten et al (1993) or the Threshold GARCH (TGARCH) model by Zakoian (1994).

The case where the stochastic component of these type of processes is assumed to be Gaussian has been widely explored, but this does not seem to be able to properly capture the essence of the returns because the residuals after fitting a conditional heteroskedastic model usually show

moderate skewness and high kurtosis. Deficient modeling has the drawback of leading to poor prediction, quantile estimation, and so on. Therefore, it is important to design models that better capture the particular features of financial data. Several alternatives have been proposed to the use of the Gaussian distribution, including the Student's t-distribution proposed by Bollerslev (1987), the generalized error distribution proposed by Nelson (1991) or a mixture of two zero mean Gaussian distributions proposed by Bai et al (2003). Nevertheless, it still has not been possible to show the existence of a distribution that adequately describes the behavior of financial returns in all situations.

In this paper, we enlarge the number of alternative distributions to the Gaussian one by proposing the skew slash distribution (Wang and Genton, 2006) to model the innovations in conditional heteroskedastic models. The skew slash distribution is based on a scaled mixture of skew normal distributions; this feature means that it can capture both moderate skewness and high kurtosis, exactly the features that have often been observed in financial data. This model has also been applied to describe data with similar characteristics in other settings, see e.g. Lachos et al (2009), where a Bayesian approach to inference for the skew slash distribution is developed.

Inference for financial time series models can be carried out using both classical, maximum likelihood, approaches or Bayesian methods. Here, we shall consider both methods. Firstly, maximum likelihood estimation is carried out using a direct constrained optimization algorithm. Secondly, we introduce a new approach to Bayesian inference for the skew slash distribution which is based on an alternative parameterization of this distribution to that of Lachos et al (2009) and enables an easy incorporation of the particular parameter restrictions inherent to the GARCH framework.

The rest of this paper is as follows. In Section 2, the skew slash distribution is introduced and explicit, closed formulae for the moments are derived. In Section 3, the GARCH(1,1) model with skew slash innovations is outlined and both classical and Bayesian approaches to inference for this model are developed. In Section 4, the performance of the proposed methodology is evaluated through a small simulation study and the analysis of the log return series of the Standard & Poor's index. Finally, the paper finishes with some conclusions and possible extensions of our approach.

2 The skew slash distribution

A random variable Z follows a standard skew normal distribution with parameter λ, and is denoted by $Z \sim \mathrm{SN}(\lambda)$, if its probability density function (pdf) is given by

$$f_Z(z) = 2\phi(z)\Phi(\lambda z), \tag{1}$$

for $z \in \mathbb{R}$, where $\lambda \in \mathbb{R}$ and $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the density and the cumulative distribution function (cdf) of the standard normal distribution, see O'Hagan and Leonard (1976) and Azzalini (1985). The parameter λ controls the skewness of the distribution. In particular, positive (negative) values of λ imply positive (negative) skewness; when $\lambda = 0$, the distribution reduces to a standard normal.

One useful property of the skew normal distribution is that it admits the following stochastic representation as shown in Henze (1985)

$$Z \equiv \delta|X_0| + \sqrt{1-\delta^2}X_1, \quad \text{where } \delta = \lambda/\sqrt{1+\lambda^2}, \tag{2}$$

where $\equiv$ signifies equivalence in distribution, and X_0, X_1 are independent standard normal random variables.

The skew slash distribution (Wang and Genton, 2006) generalizes the skew normal distribution as follows. If Z is a skew normal variable as in (1) and U is an independent, beta distributed random variable, $U \sim \mathrm{Be}(\nu, 1)$ where $\nu > 0$, then we will say that $W = U^{-1}Z$ follows a standard, skew slash distribution, $W \sim \mathrm{SSL}(\lambda, \nu)$.[1]

A four parameter skew slash distribution with location parameter η, scale parameter σ, skewness parameter λ and kurtosis parameter ν, can be defined in a straightforward generalization as

$$W = \eta + \sigma U^{-1}Z.$$

In this case, we shall write $W \sim \mathrm{SSL}(\eta, \sigma, \lambda, \nu)$.

---

[1]In fact, Wang and Genton define the skew slash variable using $U^{-1/2}$, but this definition is equivalent.

Consequently, the skew slash distribution is a scale-mixture of a variable with a skew normal distribution. The pdf of a skew slash variable is given by

$$f_W(w) = \int\limits_0^1 2\nu u^{\nu-1} \phi\left(w; \eta, u^{-2}\sigma^2\right) \Phi\left(\frac{\lambda u (w - \eta)}{\sigma^2}\right) du,$$

where $\phi\left(w; \eta, u^{-2}\sigma^2\right)$ denotes the pdf of a Gaussian distribution with mean η and variance $u^{-2}\sigma^2$.

Generalizing from (2), it is also easy to see that the skew slash distribution also admits a stochastic representation as

$$W \equiv \eta + \sigma U^{-1}\left(\delta |X_0| + \sqrt{1 - \delta^2} X_1\right). \tag{3}$$

This representation will be very useful in the context of Bayesian inference.

2.1 Moments

In order to derive the central moments of the skew slash distribution, two results are summarized in the following lemma, whose proofs, provided in the Appendix, are useful.

Lemma 1 *For $j \in \mathbb{N}$,*

1. *If $U \sim \mathrm{Be}\,(\nu, 1)$, then*
$$E\left[U^{-j}\right] = \frac{\nu}{\nu - j}, \quad \nu > j.$$

2. *If $Z \sim \mathrm{SN}\,(\lambda)$, then*
$$E\left[Z^j\right] = 2^{(j-2)/2} \pi^{-1} \left(\frac{1}{1 + \lambda^2}\right)^{j/2} \sum_{i=0}^{j} a_{ij} \lambda^i,$$

   where
$$a_{ij} = \binom{j}{i} \left(1 + (-1)^{j-i}\right) \Gamma\left(\frac{i+1}{2}\right) \Gamma\left(\frac{j-i+1}{2}\right),$$

   for $i = 0, 1, \ldots, j$. In particular, $a_{ij} = 0$ if $j - i$ is an odd number.

These results can be used to derive the expectation and central moments of the skew slash distribution, as summarized in the following Proposition.

4

Proposition 2 *If $W \sim \mathrm{SSL}\left(\eta, \sigma, \lambda, \nu\right)$, then*

1. *The expectation of W is given by*

$$E\left[W\right] = \eta + \sigma \left(\frac{2}{\pi}\right)^{1/2} \frac{\nu}{\nu - 1} \left(\frac{\lambda^2}{1 + \lambda^2}\right)^{1/2}, \quad \nu > 1.$$

2. *For $k \in \mathbb{N}$, the central moments of the skew slash distribution are given by*

$$m_k\left[W\right] = E\left[\left(W - E\left[W\right]\right)^k\right] = 2^{(k-2)/2} \sigma^k \left(\frac{1}{1 + \lambda^2}\right)^{k/2} \sum_{l=0}^{k} c_{lk} \lambda^l, \tag{4}$$

for $\nu > k$, where

$$c_{lk} = \sum_{m=0}^{l} b_{m,k-l+m,k}, \tag{5}$$

and

$$b_{m,k-l+m,k} =$$

$$(-1)^{l-m} \left[1 + (-1)^{k-l}\right] \pi^{-(l-m+2)/2} \frac{\nu}{\nu - (k - l + m)} \left(\frac{\nu}{\nu - 1}\right)^{l-m} \binom{k}{l} \binom{l}{m} \Gamma\left(\frac{m+1}{2}\right) \Gamma\left(\frac{k-l+1}{2}\right),$$

for fixed k, fixed $l \in \{0, 1, \ldots, k\}$ and $m \in \{0, 1, \ldots, l\}$, respectively. In particular, note that $c_{lk} = 0$ if $k - l$ is an odd number.

Given the general expression of the central moments of the skew slash distribution given in (4), it is straightforward to obtain the variance and the skewness and kurtosis coefficients of W, as summarized in the following corollary.

Corollary 3   1. *The variance of the skew slash distribution is given by*

$$V\left[W\right] = m_2\left[W\right] = \sigma^2 \frac{c_{02} + c_{22}\lambda^2}{1 + \lambda^2},$$

where

$$c_{02} = \frac{\nu}{\nu-2},$$

$$c_{22} = \frac{\nu}{\nu-2} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2,$$

(6)

respectively. In terms of δ

$$V[W] = \sigma^2\left(\frac{\nu}{\nu-2} - \left(\frac{\nu}{\nu-1}\right)^2\frac{2}{\pi}\delta^2\right) \quad \text{for } \nu > 2.$$

2. *The skewness coefficient of the skew slash distribution is given by*

$$S[W] = \frac{m_3[W]}{m_2[W]^{3/2}} = 2^{1/2}\frac{c_{13}\lambda + c_{33}\lambda^3}{(c_{02} + c_{22}\lambda^2)^{3/2}},$$

for $\nu > 3$, where c_{02} and c_{22} are given in (6) and

$$c_{13} = \frac{3}{\pi^{1/2}}\left(\frac{\nu}{\nu-3} - \frac{\nu}{\nu-2}\frac{\nu}{\nu-1}\right),$$

$$c_{33} = \frac{4}{\pi^{3/2}}\left(\frac{\nu}{\nu-1}\right)^3 - \frac{3}{\pi^{1/2}}\frac{\nu}{\nu-2}\frac{\nu}{\nu-1} + \frac{2}{\pi}\frac{\nu}{\nu-3},$$

respectively.

3. *The kurtosis coefficient of the skew slash distribution is given by*

$$K[W] = \frac{m_4[W]}{m_2[W]^2} = 2\frac{c_{04} + c_{24}\lambda^2 + c_{44}\lambda^4}{(c_{02} + c_{22}\lambda^2)^2},$$

for $\nu > 4$, where c_{02} and c_{22} are given in (6) and

$$c_{04} = \frac{3}{2}\frac{\nu}{\nu-4},$$

$$c_{24} = \frac{6}{\pi}\frac{\nu}{\nu-2}\left(\frac{\nu}{\nu-1}\right)^2 - \frac{12}{\pi}\frac{\nu}{\nu-3}\frac{\nu}{\nu-1} + 3\frac{\nu}{\nu-4},$$

$$c_{44} = -\frac{6}{\pi^2}\left(\frac{\nu}{\nu-1}\right)^4 + \frac{12}{\pi}\frac{\nu}{\nu-2}\left(\frac{\nu}{\nu-1}\right)^2 - \frac{8}{\pi}\frac{\nu}{\nu-3}\frac{\nu}{\nu-1} + \frac{3}{2}\frac{\nu}{\nu-4},$$
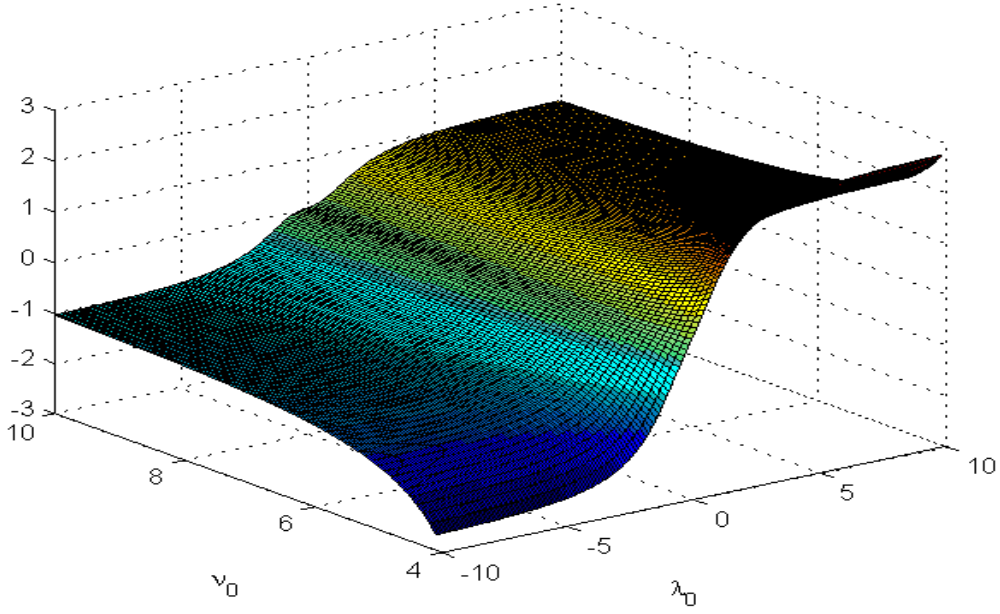
respectively.

6

Figure 1: Skewness coefficient of the SSL distribution as a function of λ_0 and ν_0

A direct consequence of these results is that the skew slash distribution is able to generate both skewness and high kurtosis. To illustrate this, Figures 1 and 2 show some values of the skewness and the kurtosis coefficients for values of λ and ν in the intervals $(-10, 10)$ and $(4, 10)$, respectively. Observe that, as with the skew normal distribution, the skewness coefficient is 0 for $\lambda = 0$, and positive (negative) for positive (negative) values of λ. Note also that ν has a small effect on skewness. On the other hand,, the kurtosis coefficient gets larger as ν gets smaller. Note also that λ has a small effect on the kurtosis.

3   The GARCH model with skew slash innovations

Although the proposed approach can be adopted in any conditional heteroskedastic model, for simplicity and because it is one of the most popular models for estimating the dynamics of financial
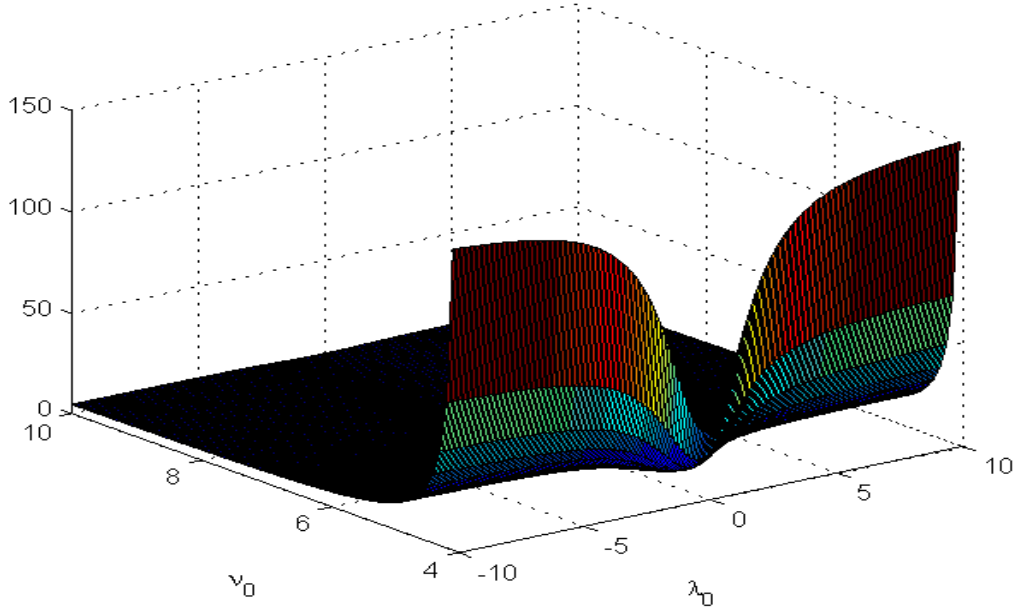
Figure 2: Kurtosis coefficient of the SSL distribution as a function of λ_0 and ν_0

returns, for illustration we use the GARCH(1,1) defined as follows

$$
\begin{aligned}
y_t &= \mu + \sqrt{h_t}\varepsilon_t; \\
h_t &= \omega + \alpha \left(y_{t-1} - \mu\right)^2 + \beta h_{t-1},
\end{aligned}
\tag{7}
$$

for $t = 1, \ldots, T$, where h_t is the conditional variance of y_t, given $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$, the information set available until time $t-1$, and the innovations, ε_t, are independent and identically distributed random variables such that $E\left[\varepsilon_t\right] = 0$ and $V\left[\varepsilon_t\right] = 1$, for $t = 1, \ldots, T$. It is typically assumed that h_0 and y_0 are known constants, $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$, to ensure positivity of h_t and $\alpha + \beta < 1$, to ensure covariance stationarity.

Here, it is proposed to model the innovations, ε_t, as skew slash distributed random variables. This seems to be a desirable approach because financial returns usually present moderate skewness and high kurtosis in the innovations when modeled through a GARCH framework and, as it has already been shown, the skew slash distribution is able to help capture this kind of characteristics

8

due to its nature. It is important to note that usual stationarity conditions of the GARCH process are directly applicable using the results in Carrasco and Chen (2002). In particular, from Proposition 12 in that paper, if $\nu > 2$, then, $E[h_t] < \infty$ and $E\left[y_t^2\right] < \infty$. Moreover, if h_0 is a constant, then $\{(y_t, \varepsilon_t)\}$ is strictly stationary and β-mixing with exponential decay.

We shall consider two slightly different but equivalent formulations of this basic model to facilitate inference in the classical and Bayesian contexts, respectively, which incorporate the moment restrictions on the innovation mean and variance. Firstly, we describe a classical approach.

3.1 Classical inference

Here we shall assume directly that $\varepsilon_t \sim \mathrm{SSL}\,(\eta, \sigma, \lambda, \nu)$ with $E\left[\varepsilon_t\right] = 0$ and $V\left[\varepsilon_t\right] = 1$. Therefore, from Proposition 2

$$E\left[\varepsilon_t\right] = 0 = \eta + \sigma \left(\frac{2}{\pi}\right)^{1/2} \frac{\nu}{\nu - 1} \left(\frac{\lambda^2}{1 + \lambda^2}\right)^{1/2}$$

so that

$$\eta = -\sigma \left(\frac{2}{\pi}\right)^{1/2} \frac{\nu}{\nu - 1} \left(\frac{\lambda^2}{1 + \lambda^2}\right)^{1/2}. \tag{8}$$

Also, from Corollary 3

$$V\left[\varepsilon_t\right] = 1 = \sigma^2 \frac{c_{02} + c_{22}\lambda^2}{1 + \lambda^2}$$

so that

$$\sigma^2 = \frac{1 + \lambda^2}{c_{02} + c_{22}\lambda^2} \tag{9}$$

where c_{02} and c_{22} are given in (6).

Assume that a series of returns $\mathbf{y} = (y_1, \ldots, y_T)$ is observed. Under the GARCH model with skew slash innovations, the likelihood function of a GARCH(1,1) model is given by

$$
\begin{aligned}
f\left(\mathbf{y}|\theta\right) &= f\left(y_T|\mathcal{F}_{T-1}\right) f\left(y_{T-1}|\mathcal{F}_{T-2}\right) \cdots f\left(y_1|\mathcal{F}_0\right) \\
&= \prod_{t=1}^{T} \left\{ h_t^{-1/2} \int_0^1 2\nu_0 u^{\nu-1} \phi\left(\frac{y_t - \mu}{\sqrt{h_t}}; \eta, u^{-2}\sigma^2\right) \Phi\left(\frac{\lambda u\left(y_t - \mu - \eta\sqrt{h_t}\right)}{\sigma^2\sqrt{h_t}}\right) du \right\},
\end{aligned}
$$

where $\theta = (\mu, \omega, \alpha, \beta, \lambda, \nu)'$ is the vector of parameters of the model, $f\left(y_1, \ldots, y_m|\theta\right)$ is the joint

pdf of $y_1, \ldots, y_m$, and η and σ^2 are given in (8) and (9), respectively.

The ML estimator is obtained by maximizing the log conditional likelihood function

$$\mathcal{L}\left(\theta | y_1, \ldots, y_T\right) = \sum_{t=1}^{T} \ell_t\left(\theta\right),$$

where

$$\ell_t\left(\theta\right) = -\frac{1}{2} \log h_t + \log \left\{ \int_0^1 2\nu u^{\nu-1} \phi\left(\frac{y_t - \mu}{\sqrt{h_t}}; \eta, u^{-2}\sigma^2\right) \Phi\left(\frac{\lambda u\left(y_t - \mu - \eta\sqrt{h_t}\right)}{\sigma^2\sqrt{h_t}}\right) du \right\}$$

The maximization of the log conditional likelihood is a highly nonlinear problem but can be carried out by standard numerical algorithms. By the stationary and mixing properties of the processes y_t and h_t, previously mentioned, it is reasonable to apply usual large sample results of ML estimation and to assume that the ML estimator of θ, denoted by $\widehat{\theta}$, is asymptotically Gaussian distributed with mean θ and covariance matrix $-E\left[\partial^2 \mathcal{L}\left(\theta | y_1, \ldots, y_T\right) / \partial\theta\partial\theta'\right]^{-1}$. Then, approximated standard errors of the parameters can be obtained by taking the square roots of the elements of $\partial^2 \mathcal{L}\left(\widehat{\theta} | y_1, \ldots, y_T\right) / \partial\theta\partial\theta'$.

3.2 Bayesian inference

In order to consider a Bayesian approach to the skew slash GARCH model, it is convenient from a practical viewpoint to define $\tilde{h}_t = h_t/\omega$, and (7) becomes

$$\begin{aligned} y_t &= \mu + \sqrt{\tilde{h}_t}\tilde{\varepsilon}_t; \\ \tilde{h}_t &= 1 + \tilde{\alpha}\left(y_{t-1} - \mu\right)^2 + \beta\tilde{h}_{t-1}, \end{aligned} \tag{10}$$

so that $E[\tilde{\varepsilon}_t] = 0$, $V[\tilde{\varepsilon}_t] = \omega$ and $0 < \tilde{\alpha} = \alpha/\omega < 1/\omega$. Assume that $\tilde{\varepsilon}_t \sim \text{SSL}(\eta, \sigma, \lambda, \nu)$, where, in order to comply with the mean restriction, we set $E[\tilde{\varepsilon}_t] = 0$ so that

$$\delta\sigma = -\sqrt{\frac{\pi}{2}} \frac{\nu - 1}{\nu} \eta.$$

10

Then, incorporating this condition into the stochastic representation of the skew slash distribution in (3) we have

$$\tilde{\varepsilon}_t \equiv \eta \left(1 - \sqrt{\frac{\pi}{2}} \frac{\nu - 1}{\nu} U_t^{-1} \left[|X_{0t}| + \frac{1}{\lambda} X_{1t} \right] \right),$$

where $U_t | \nu \sim \mathrm{Be}(\nu, 1)$ and X_{0t}, X_{1t} are standard normal random variables. Then, defining $R_t = U_t^{-1} |X_{0t}|$, and $\rho = \left(\sqrt{\frac{2}{\pi}} \frac{\nu}{\nu - 1} \frac{\lambda}{\eta} \right)^2$, we have

$$\tilde{\varepsilon}_t \equiv \eta \left(1 - \sqrt{\frac{\pi}{2}} \frac{\nu - 1}{\nu} R_t \right) + \frac{1}{U_t \sqrt{\rho}} Z_{1t}$$

Given the above formulation, we have that:

$$y_t | \mu, \tilde{\alpha}, \beta, \eta, \rho, \nu, x_t, u_t, y_{t-1}, \tilde{h}_{t-1} \sim \mathrm{N} \left(\mu + \sqrt{\tilde{h}_t} \eta \left(1 - \sqrt{\frac{\pi}{2}} \frac{\nu - 1}{\nu} x_t \right), \frac{\tilde{h}_t}{u_t^2 \rho} \right).$$

In order to undertake Bayesian inference, it is first necessary to define prior distributions for the model parameters $\mu, \tilde{\alpha}, \beta, \eta, \nu$, and ρ. One of the advantages of the Bayesian approach in this context is that, for many of these parameters, real prior information in the form of expert knowledge, or based on economic theory, will be available and this information can be incorporated into the analysis. Thus, for example, economic theory suggests that the drift parameter, μ, in the GARCH model should be very close to zero. Therefore, a reasonable prior distribution that incorporates this knowledge is a normal distribution centered at 0. Secondly, analysts with experience in GARCH models will often be able to provide good prior estimates of the volatility parameters, α and β. Thirdly, the parameter ν represents the number of finite moments of the error distribution and analysts will often be able to give good estimates of this parameter. In other cases, where the parameter does not have such a clear interpretation, more non-informative prior distributions can be used to represent prior uncertainty. Here, we shall assume the following prior distributions:

$$\mu \sim \mathrm{N}\left(0, \frac{1}{c_m}\right) \quad \text{where } c_m >> 1$$

$$f(\tilde{\alpha}, \beta | \omega) \propto \omega \frac{\Gamma(c)}{\Gamma(cp_a)\Gamma(cp_b)\Gamma(c(1-p_a-p_b))}(\tilde{\alpha}\omega)^{cp_a-1}\beta^{cp_b-1}(1-\tilde{\alpha}\omega-\beta)^{c(1-p_a-p_b)}$$

$$\eta \sim \mathrm{N}\left(m_e, \frac{1}{c_e}\right)$$

$$\nu \sim \mathrm{Ga}\left(a_n, b_n\right) \quad \text{truncated onto } \nu > 2$$

$$\rho \sim \mathrm{Ga}\left(\frac{a_r}{2}, \frac{b_r}{2}\right) \quad \text{where } a_r, b_r > 0 \text{ are small.}$$

In the above, $\omega = \omega(\eta, \nu, \rho) = V[\tilde{\varepsilon}_t | \eta, \nu, \rho]$ as derived in Corollary 3. Note that the prior distribution for $\tilde{\alpha}, \beta$ is derived by assuming a Dirichlet prior distribution for $(\alpha, \beta, 1 - \alpha - \beta)$.

Given this prior structure, exact inference is impossible. However, expressions for the posterior conditional distributions can be derived. Thus, we have:

$$f(\mu | \mathbf{y}, \tilde{\alpha}, \beta, \eta, \rho, \nu, \mathbf{r}, \mathbf{u}) \propto$$

$$e^{-\frac{c_m\mu^2}{2}} \prod_{t=1}^{T} \frac{1}{\sqrt{\tilde{h}_t(\mu)}} \exp\left(-\frac{u_t^2\rho}{2\tilde{h}_t(\mu)}\left(y_t - \mu - \sqrt{\tilde{h}_t(\mu)}\eta\left(1 - \sqrt{\frac{\pi}{2}}\frac{\nu-1}{\nu}r_t\right)\right)^2\right),$$

where we have expressed the volatility $\tilde{h}_t$ as a function of μ to make the dependence clear.

$$f(\tilde{\alpha}, \beta | \mathbf{y}, \mu, \eta, \rho, \nu, \mathbf{r}, \mathbf{u}) \propto \tilde{\alpha}^{c_d p_a - 1}\beta^{c_d p_b - 1}(1 - \tilde{\alpha}\omega - \beta)^{c_d(1-p_a-p_b)-1} \times$$

$$\prod_{t=1}^{T} \frac{1}{\sqrt{\tilde{h}_t(\tilde{\alpha}, \beta)}} \exp\left(-\frac{u_t^2\rho}{2\tilde{h}_t(\tilde{\alpha}, \beta)}\left(y_t - \mu - \sqrt{\tilde{h}_t(\tilde{\alpha}, \beta)}\eta\left(1 - \sqrt{\frac{\pi}{2}}\frac{\nu-1}{\nu}r_t\right)\right)^2\right).$$

where we have written $\tilde{h}_t$ as a function of $(\tilde{\alpha}, \beta)$ to make the dependence clear. For η,

$$f(\eta | \mathbf{y}, \mu, \tilde{\alpha}, \beta, \rho, \nu, \mathbf{r}, \mathbf{u}) \propto \omega(\eta)\omega(\eta)^{c_d p_a - 1}(1 - \tilde{\alpha}\omega(\eta) - \beta)^{c_d(1-p_a-p_b)-1}\phi\left(\sqrt{c_e^\star}(\eta - m_e^\star)\right)$$

where $\phi(\cdot)$ is a normal density and

$$c_e^\star = c_e + \sum_{t=1}^{T} u_t^2 \rho \left(1 - \sqrt{\frac{\pi}{2} \frac{\nu - 1}{\nu}} r_t \right)^2$$

$$c_e^\star m_e^\star = c_e m_e + \sum_{t=1}^{T} \frac{u_t^2 \rho}{\sqrt{\tilde{h}_t}} \left(1 - \sqrt{\frac{\pi}{2} \frac{\nu - 1}{\nu}} r_t \right) (y_t - \mu)$$

where we have made the dependence of ω on η clear. For ν,

$$f(\nu|\mathbf{y}, \mu, \tilde{\alpha}, \beta, \rho, \eta, \mathbf{r}, \mathbf{u}) \propto \omega(\nu)^{c_d p_a} (1 - \tilde{\alpha}\omega(\nu) - \beta)^{c_d(1 - p_a - p_b) - 1} \times$$

$$\nu^{a_n + T - 1} e^{-\nu\left(b_n - \sum_{t=1}^{T} \log u_t\right)} \times$$

$$\exp\left(-\sum_{t=1}^{T} \frac{u_t^2 \rho}{\tilde{h}_t} \left(y_t - \mu - \sqrt{\tilde{h}_t} \eta \left(1 - \sqrt{\frac{\pi}{2} \frac{\nu - 1}{\nu}} r_t \right) \right)^2 \right)$$

where we have made the dependence of ω on ν clear. For ρ,

$$f(\rho|\mathbf{y}, \mu, \tilde{\alpha}, \beta, \eta, \nu, \mathbf{r}, \mathbf{u}) \propto \omega(\rho)\omega(\rho)^{c_d p_a - 1}(1 - \alpha\omega(\rho) - \beta)^{c_d(1 - p_a - p_b) - 1} g\left(\rho \left| \frac{a_g^\star}{2}, \frac{b_g^\star}{2} \right. \right)$$

where $g(\cdot|\cdot, \cdot)$ is a gamma density and

$$a_g^\star = a_g + T$$

$$b_g^\star = b_g + \sum_{t=1}^{T} \frac{u_t^2}{\tilde{h}_t} \left(y_t - \mu - \sqrt{\tilde{h}_t} \eta \left(1 - \sqrt{\frac{\pi}{2} \frac{\nu - 1}{\nu}} x_t \right) \right)^2,$$

where again, we have made the dependence of ω on ρ obvious. Finally, for $t = 1, \ldots, T$, we have:

$$f(r_t|\mathbf{y}, \mu, \tilde{\alpha}, \beta, \eta, \rho, \nu, u_t) \propto \phi\left(r_t \left| m_t^\star, \frac{1}{c_t^\star} \right. \right) \mathbb{I}_{\mathbb{R} \cup \{0\}} (r_t),$$

13

where

$$c_t^\star = u_t^2 \left(1 + \frac{\pi}{2}\eta^2\rho \left(\frac{\nu-1}{\nu} \right)^2 \right)$$

$$c_t^\star m_t^\star = \frac{u_t^2\eta\rho}{\sqrt{\tilde{h}_t}}\sqrt{\frac{\pi}{2}}\frac{\nu-1}{\nu}\left(\mu + \sqrt{\tilde{h}_t}\eta - y_t \right),$$

and

$$u_t^2|\mathbf{y},\mu,\alpha,\beta,\eta,\rho,\nu,r_t \sim \mathrm{Ga}\left(\frac{\nu+2}{2}, \left(r_t^2 + \frac{\rho}{\tilde{h}_t}\left(y_t - \mu - \sqrt{\tilde{h}_t}\eta\left(1 - \sqrt{\frac{\pi}{2}}\frac{\nu-1}{\nu}r_t \right) \right)^2 \right) \right)\mathbb{I}_{(0,1)}(u_t).$$

A Metropolis Hastings within Gibbs algorithm can then be designed to sample the joint posterior parameter distribution by successively sampling the above distributions. In the cases of r_t and u_t, sampling is straightforward. In the cases of η and ρ, sampling can be carried out using a Metropolis Hastings algorithm based on truncated normal and gamma candidate distributions, respectively. In the case of ν, we consider a griddy Gibbs sampler based on the reparameterisation $\tau = (\nu-1)/\nu$. For μ, we use a Metropolis sampler based on a normal distribution centred at the current value. Finally, in the case of $\tilde{\alpha}, \beta$ we consider a sampler based on a combination of an independence sampler centred on the maximum likelihood estimates and a Metropolis Hastings sampler centred on the current values.

4   Examples

In this section, we illustrate our approach with a small simulation and the analysis of real data from the Standard & Poor's 500 Index. In both cases, the Bayesian prior parameters are specified as $c_m = 10$, $c = 10$, $p_a = 0.1$, $p_b = 0.8$, $m_e = 0$, $c_e = 1$, $a_n = b_n = 1$ and $a_r = b_r = 1$. The sampler is run for 5000 iterations to burn in and 10000 iterations in equilibrium.

4.1   Simulation results

We focus first on the maximum likelihood method. Then, we generate three sets of 2500 time series from a SSL GARCH model with GARCH parameters $\mu = 0$, $\omega = 0.01$, $\alpha = 0.1$ and $\beta = 0.85$ and

14

with skew slash parameters $\eta = 0.652$, $\sigma^2 = 0.855$, $\lambda = -1$ and $\nu = 5$. Note that $\eta = 0.652$ and $\sigma^2 = 0.855$ are set to verify the two moment conditions. Each set corresponds to the sample sizes $T = 1000$, $T = 2000$ and $T = 3000$, respectively.

Table 1 shows the mean and standard error of the model parameters over the 2500 time series estimated through maximum likelihood. Observe that the parameter estimation results are apparently very good, even given the small sample size. Also, as expected, the larger the sample size, the smaller the standard errors and the better the mean estimates.

Table 1: Mean and Standard deviation for the ML estimators, with T simulated observations.

	$T = 1000$	$T = 2000$	$T = 3000$
μ	-0.00136	-0.00058	-0.00085
	(0.01237)	(0.00861)	(0.00699)
ω	0.01161	0.01068	0.01025
	(0.00505)	(0.00286)	(0.00237)
α_1	0.10115	0.10114	0.09981
	(0.02610)	(0.01906)	(0.01542)
β_1	0.84053	0.84519	0.84854
	(0.04301)	(0.02736)	(0.02190)
λ	-1.02685	-1.01937	-1.02641
	(0.27714)	(0.19226)	(0.15933)
ν	5.43052	5.22406	5.16790
	(1.72854)	(0.93940)	(0.56823)

Unfortunately, the Bayesian procedure is time consuming and a comparative exercise as the one performed with maximum likelihood is not possible. However, we consider one time series generated with sample size $T = 3000$ to compare the estimated volatilies using both the maximum likelihood and the Bayesian procedure. Figure 3 shows the true (black line), classically estimated (red line) and Bayesian posterior mean (blue line) volatility estimates. It can be seen that the volatility predictions for both models are almost indistinguishable from the true volatilies in this case.
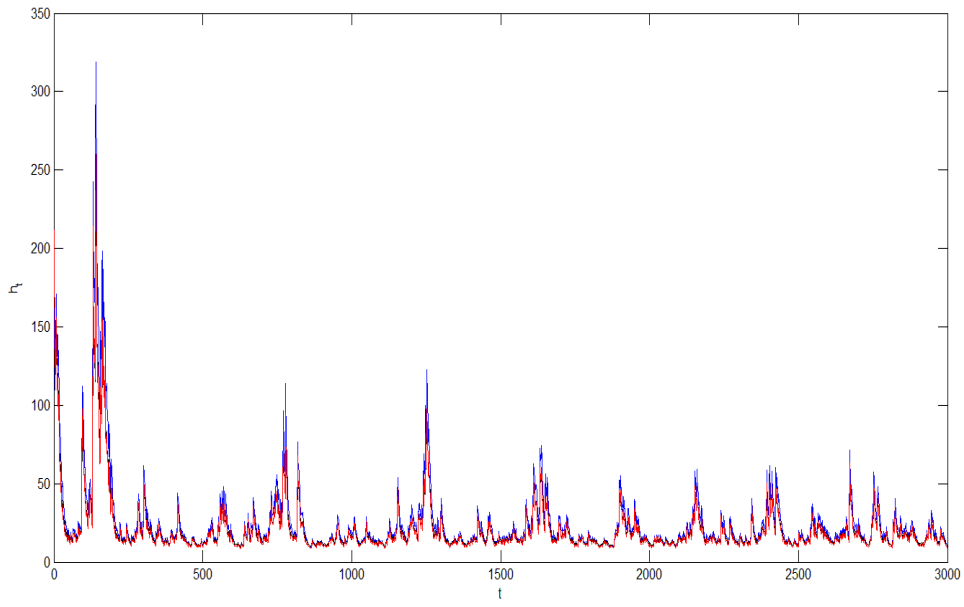
Figure 3: True volatilities (black line) and classical (red line) and Bayesian (blue line) volatility estimates

4.2 Real data example

As an illustration of the usefulness of the proposed approach based on the skew slash distribution, this section analyzes the Standard & Poor's 500 Index (S&P 500). The S&P 500 index is a free-float capitalization-weighted index of the prices of 500 of the main companies in leading industries of the U.S. economy. Figure 4 shows the time plot of the simple returns of the daily closing prices of the index (in percentages) for the period from January 1983 until December 1997, leading to 3035 index returns. Observe that the returns appear to vary more in the first part of the series, which is the period that includes the one-day crash of "Black Monday" corresponding to October, 19, 1987. The Black Monday fall was the largest one-day percentage decline in stock market history. The sample mean, standard deviation, skewness and kurtosis of the return series are 0.0555, 0.9890, -3.2210 and 70.8934, respectively. Observe that the return series is left-skewed and the kurtosis is very large, indicating that the return distribution has higher peaks and heavier tails than a normal distribution with the same variance. Thus, it seems that the proposed skew slash specification can be adequate to address these issues.

Figure 5 shows the estimated innovation densities under both the classical (red line) and Bayesian (blue line) approaches. The densities are quite similar and both exhibit some negative skewness although the Bayesian predictive density is heavier tailed. This can also be seen by looking at a kernel density estimate of the Bayesian posterior density of ν which is concentrated on values of ν at 4 or below, see Figure 6, which suggests that there is some evidence that the fourth and higher moments of the innovation distribution do not exist. Indeed, the ML estimate of ν is given by 3.3486, which is very close to the posterior mean of ν in the Bayesian approach.

Finally, the fitted volatilities are shown in Figure 7. It can be seen that both approaches capture the same features, but that in this case, the fitted volatilities estimated via the classical approach are slightly higher.
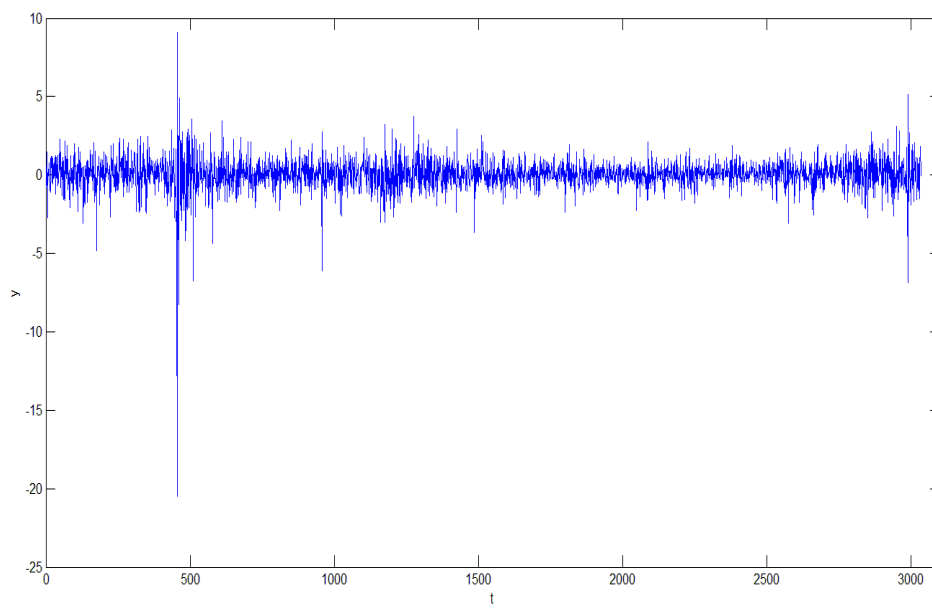
17

Figure 4: Simple returns of the Standard & Poor's 500 Index in the period from January 1983 until December, 1997.
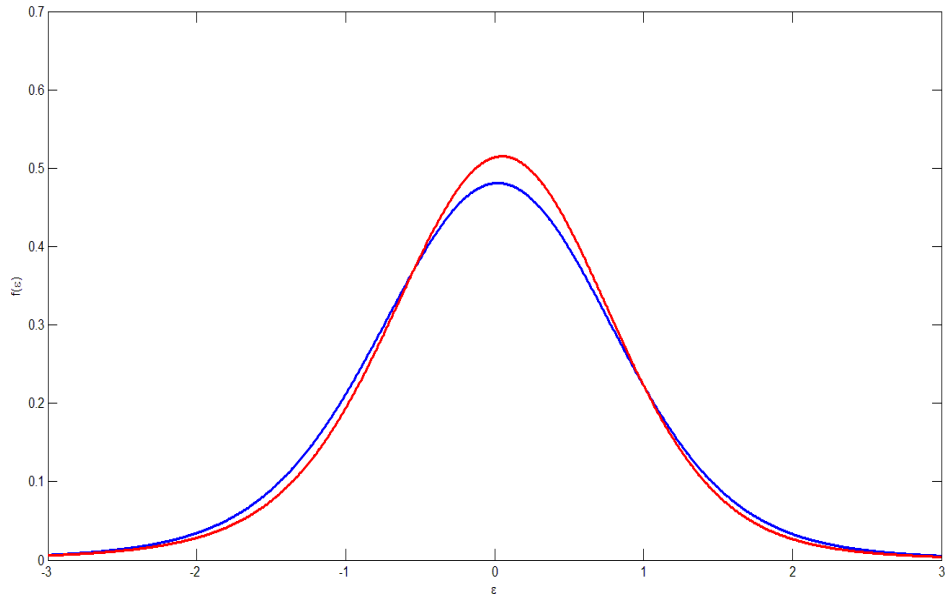
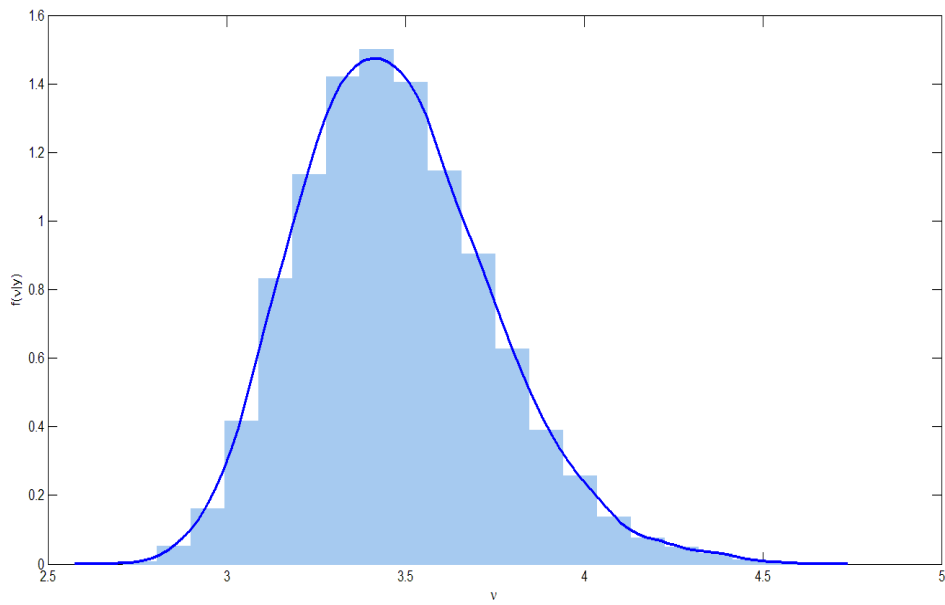Figure 5: Estimated innovation density: red = classical, blue = Bayesian

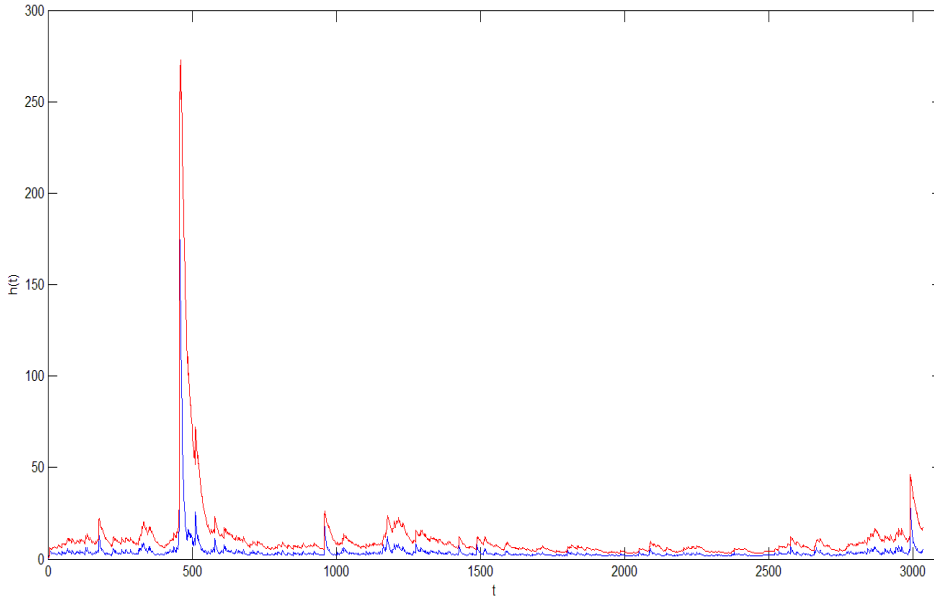Figure 6: Bayesian posterior density estimate for ν

19

Figure 7: Estimated volatilities: red = classical, blue = Bayesian

5   Conclusions and extensions

In this paper, we have introduced a new approach to modeling the innovations in a GARCH(1,1) model using a skew slash distribution, which allows us to capture the asymmetry and high kurtosis that are often observed in financial time series. We have shown how both classical and Bayesian inferential techniques can be used for model fitting and have illustrated our methods with both simulated and real time series data.

A number of extensions are possible. Firstly, although we have introduced our approach here with respect to the GARCH(1,1) model, it would also be interesting to extend both our approaches to classical and Bayesian inference to the more general GARCH(p,q) models. Obviously, this would extend the complexity of the optimization algorithm used within the classical algorithm and alternative approaches such as the EM algorithm might be considered in this context. A recent relevant article that uses the EM algorithm in the context of skew slash and other models is Lachos et al (2010).

Secondly, it is possible to extend our approach to multivariate financial time series where a multivariate skew slash distribution can be considered for the error model. In this line, Barbosa Cabral et al (2012) develop an approach to EM based inference for multivariate skew slash distributions.

Finally, it would be useful to apply this model for value at risk and conditional value at risk estimation in the financial time series context. Work on all these problems is currently in progress.

References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.

- Bai, X., Russell, J.R. and Tiao, G.C., (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, **114**, 349–360.

- Barbosa Cabral, C.R., Lachos, V.H. and Prates, M.O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics & Data Analysis*, **56**, 126—142.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.

- Bollerslev, T. (1987). A conditional heteroscedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, **69**, 542–547.

- Glosten L.R., Jagannathan, R. and Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, **48**, 1779–1801.

- Henze, N. (1985). A probabilistic representation of the 'skew-normal' distribution. *Scandinavian Journal of Statistics*, **13**, 271–275.

- Lachos, V.H., Dey, D.K. and Cancho, V.G. (2009) Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. *Journal of Statistical Planning*

and Inference, **139**, 4098—4110.

- Lachos, V.H., Ghosh, P. and Arrellano Valle, R.B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**, 303–322.

- Nelson, D.B. (1991). Conditional heteroscedasticity in asset pricing: a new approach. *Econometrica*, **59**, 347–370.

- O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, **63**, 201–203.

- Wang and Genton (2006). The multivariate skew slash distribution. *Journal of Statistical Planning and Inference*, **136**, 209—220.

- Zakoian, J.M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, **18**, 931—955.

Appendix

Proof of Lemma 1.

1. The equality holds from

$$E\left[U^{-j}\right] = \int_0^1 u^{-j}\nu u^{\nu-1}du = \nu \int_0^1 u^{-j-\nu+1}du = \frac{\nu}{\nu-j}, \nu > j.$$

2. First, from the stochastic representation of Z in (2)

$$E\left[Z^j\right] = \left(\frac{1}{1+\lambda^2}\right)^{j/2} E\left[(\lambda\,|X_0| + X_1)^j\right] = \left(\frac{1}{1+\lambda^2}\right)^{j/2} \sum_{i=0}^{j}\binom{j}{i}\lambda^i E\left[|X_0|^i\right] E\left[X_1^{j-i}\right]$$

22

Now, the moments of the standard normal distribution are given by

$$E\left[X_1^{j-i}\right] = \left(1 + (-1)^{j-i}\right) 2^{(j-i-2)/2}\pi^{-1/2}\Gamma\left(\frac{j-i+1}{2}\right),$$

where $\Gamma\left(\cdot\right)$ is the Gamma function, while the moments of the half normal distribution are given by

$$E\left[|X_0|^i\right] = 2^{i/2}\pi^{-1/2}\Gamma\left(\frac{i+1}{2}\right).$$

Therefore,

$$E\left[|X_0|^i\right] E\left[X_1^{j-i}\right] = \left(1 + (-1)^{j-i}\right) 2^{(j-2)/2}\pi^{-1}\Gamma\left(\frac{i+1}{2}\right)\Gamma\left(\frac{j-i+1}{2}\right)$$

that leads to

$$E\left[Z^j\right] = 2^{(j-2)/2}\pi^{-1}\left(\frac{1}{1+\lambda^2}\right)^{j/2}\sum_{i=0}^{j} a_{ij}\lambda^i,$$

where

$$a_{ij} = \binom{j}{i}\left(1 + (-1)^{j-i}\right)\Gamma\left(\frac{i+1}{2}\right)\Gamma\left(\frac{j-i+1}{2}\right),$$

for $i = 0, 1, \ldots, j$.

∎

Proof of Proposition 2.

1. From Lemma 1, we have that

$$E\left[U^{-1}\right] = \frac{\nu}{\nu-1}, \ \nu > 1$$

$$E\left[Z\right] = \left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\lambda^2}{1+\lambda^2}\right)^{1/2}.$$

Consequently,

$$E\left[W\right] = E\left[\eta + \sigma U^{-1}Z\right] = \eta + \sigma E\left[U^{-1}\right] E\left[Z\right] = \eta + \sigma\left(\frac{2}{\pi}\right)^{1/2}\frac{\nu}{\nu-1}\left(\frac{\lambda^2}{1+\lambda^2}\right)^{1/2}, \text{ for } \nu > 1.$$

23

2. First, $W - E[W]$ can be written as

$$W - E[W] = \sigma \left(U^{-1}Z - E[U^{-1}]E[Z] \right),$$

leading to

$$(W - E[W])^k = \sigma^k \left(U^{-1}Z - E[U^{-1}]E[Z] \right)^k = \sigma^k \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} U^{-j} Z^j E[U^{-1}]^{k-j} E[Z]^{k-j}$$

Now, taking expectations in the previous equation,

$$m_k(W) = \sigma^k E \left[\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} U^{-j} Z^j E[U^{-1}]^{k-j} E[Z]^{k-j} \right] =$$

$$= \sigma^k \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} E[U^{-j}] E[Z^j] E[U^{-1}]^{k-j} E[Z]^{k-j} =$$

$$= 2^{(k-2)/2} \sigma^k \left(\frac{1}{1+\lambda^2} \right)^{k/2} \sum_{j=0}^{k} \sum_{i=0}^{j} b_{ijk} \lambda^{k-j+i},$$

for $\nu > \frac{k}{2}$, where

$$b_{ijk} = (-1)^{k-j} \binom{k}{j} \frac{\nu}{\nu - j} \left(\frac{\nu}{\nu - 1} \right)^{k-j} \pi^{-(k-j+2)/2} a_{ij},$$

with k fixed, $j \in \{0, 1, \ldots, k\}$, and $i \in \{0, 1, \ldots, j\}$, respectively.

Finally, the two sums can be reduced as follows.

$$m_k(W) = 2^{(k-2)/2} \sigma^k \left(\frac{1}{1+\lambda^2} \right)^{k/2} \sum_{l=0}^{k} c_{lk} \lambda^l,$$

where

$$c_{lk} = \sum_{m=0}^{l} b_{m,k-l+m,k},$$

24

and

$$b_{m,k-l+m,k} =$$

$$(-1)^{l-m} \left[1 + (-1)^{k-l} \right] \pi^{-(l-m+2)/2} \frac{\nu}{\nu - (k-l+m)} \left(\frac{\nu}{\nu - 1} \right)^{l-m} \binom{k}{l} \binom{l}{m} \Gamma \left(\frac{m+1}{2} \right) \Gamma \left(\frac{k-l+1}{2} \right),$$

for fixed k, fixed $l \in \{0, 1, \ldots, k\}$, and $m \in \{0, 1, \ldots, l\}$, respectively.

∎

Proof of Corollary.   The three equations lead from (4) for $k = 2$, 3 and 4, respectively.   ∎