I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0125

MODELLING FINANCIAL TIME SERIES USING GARCH-TYPE MODELS WITH A SKEWED STUDENT DISTRIBUTION FOR THE INNOVATIONS

P. LAMBERT and S. LAURENT

http://www.stat.ucl.ac.be

Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations

Philippe Lambert* and Sébastien Laurent†

Abstract

Fernández and Steel (1998) propose a 4-parameter skewed Student distribution where the parameters specifying the location, the dispersion, the asymmetry and the tail thickness have a meaningful interpretation. We first reparametrize their density as a function of the conditional mean and of the conditional variance and derive its cumulative density function and quantile function. We also proceed to a Monte Carlo simulation to assess its practical applicability in a MLE estimation procedure in the GARCH framework. Finally, this general tool is illustrated by the analysis of the NASDAQ on the period 1985-1996. Using both in- and out-of-sample density forecast tests, we validate the choice of this density and reject the normal and Student densities.

*Université Catholique de Louvain, Institut de Statistique, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: lambert@stat.ucl.ac.be

†Corresponding author: Université de Liège, Département d'Economie, Sart-Tilman, B31, 4000 Liège, Belgium and Department of Quantitative Economics, Maastricht University, Netherlands. E-mail: S.Laurent@ulg.ac.be or S.Laurent@ke.unimaas.nl

1 Introduction

During the past decade, the statistical analysis of financial time series has focused on the conditional second moment as most financial asset returns exhibit temporal bursts of volatility. ARCH-type models (Engle, 1982) and its various extensions, e.g. GARCH (Bollerslev, 1986), TARCH (Zakoian, 1994), APARCH (Ding, Granger, and Engle, 1993), EGARCH (Nelson, 1991) and FI-GARCH (Baillie, Bollerslev, and Mikkelsen, 1996) among others, are commonly used to describe the conditional variance while an ARMA structure is often considered for the conditional mean. Even if the choice of appropriate statistical models for the first two moments is an important issue, few attention has been devoted to the specification of the conditional distribution. These sophisticated linear models for the conditional mean and for the conditional variance often rely on simplistic assumptions on the stochastic structure (normality). Indeed, it is widely accepted that financial returns, on a weekly, daily or intraday basis, are fat-tailed and even skewed. However, if we are only interested in the first two conditional moments, the normality assumption may be justified by the fact that the quasi maximum likelihood estimator is consistent assuming that the conditional mean and the conditional variance are specified correctly (see Weiss, 1986 and Bollerslev and Wooldridge, 1982). This estimator is, however, inefficient with the degree of inefficiency increasing with the degree of departure from normality (Engle and González-Rivera, 1991). Searching for a more suitable distribution may thus be of primary importance to gain in efficiency. Moreover, from a practical point of view, the issue of skewness (asymmetry) and kurtosis (fat-tails) is important in many respects for financial applications. Peiró (1999) emphasizes the relevance of the modelling of higher-order features in asset pricing models[1], portfolio selection[2] and option pricing theories.[3] Modelling skewness and kurtosis has an impact on all conditional quantiles. Therefore, not surprisingly, they were found to be crucial in Value-at-Risk applications (see Giot, 2000 and Giot and Laurent, 2001 for instance). In practice, it is frequent to reject normality in favor of a fat-tailed density. However, switching from a normal density to (say) a Student density may be hazardous as it does not imply that this last assumption holds. And if it does not, it is very likely that the estimates won't be consistent (see Newey and Steigerwald, 1997).[4] As a consequence, we have to be very cautious with the choice of the density and check its appropriateness. To accommodate the excess of (unconditional-) skewness and/or kurtosis, GARCH models have been first combined with Student-, mixture of normal- or Student- distributed errors. In general, it has been found that these densities cannot capture all the skewness and leptokurtosis (Beine and Laurent, 1999; Bollerslev, 1987; Jorion, 1988; Neely, 1999; Vlaar and Palm, 1993, among others), although they seem adequate in some rare cases. Liu and Brorsen (1995) and Lambert and Laurent (2001) consider the asymmetric stable density in combination with a GARCH model. A major drawback of the stable density is that, except when the tail parameter $\alpha = 2$ (i.e. normality), the variance does not exist, a fact usually not supported empirically. Lee and Tse (1991), Knight, Satchell, and Tran (1995) and Harvey and Siddique (1999)[5] propose alternative skewed fat-tailed densities, with respectively the Graham-Charlier Expansion, the Double-Gamma distribution and the non-central t. However, as pointed out by Bond (2000) in a recent survey on asymmetric conditional density functions, estimation of these densities in a GARCH framework often proved troublesome and highly sensitive to initial values. McDonald (1984, 1991) introduced the exponential generalized beta distribution of the second kind (EGB2), a flexible distribution that is able to accommodate not only thick tails but also asymmetry. The usefulness of this density has been proved recently by Wang, Fawson, Barrett, and McDonald (2000) in a GARCH framework. These authors show that a more flexible density than the normal and the Student is required in the modelling of six daily nominal exchange rate returns vis-a-vis the US dollar. However, goodness of fit tests clearly reject the EGB2 for all the currencies that they consider, even if it seems that

[1] Asset pricing models are indeed incomplete unless the full conditional model is specified.

[2] Chunhachinda, Dandapani, Hamid, and Prakash (1997) find that the incorporation of skewness into the investor's portfolio decision causes a major change in the construction of the optimal portfolio'.

[3] Corrado and Su (1996, 1997) show that when skewness and kurtosis adjustment terms are added to the Black and Scholes formula, improved accuracy is obtained for pricing options.

[4] Note the the consistency is preserved when assuming (even wrongly) normality for the conditional density.

[5] This list is by no means exhaustive.

it outperforms the normal and the Student. Hansen (1994) is the first to propose a skewed Student distribution in which conditional higher moments may vary over time. Its density nests the symmetric Student when the asymmetry coefficient (λ) equals 0, with $-1 < \lambda < 1$. This density is quite easy to implement and do not imply serious problems of convergence. However, Hansen (1994) does not discuss the relation between λ and higher moments. More recently, Fernández and Steel (1998) developed a more general tool (based on the method of inverse scaling of the probability density function on the left and the right of the mode) which has the advantage that all the parameters have a clear interpretation (see below). Moreover, contrary to Hansen (1994), Fernández and Steel (1998) discuss the relation between the asymmetry coefficient and the first three moments. In order to keep in the ARCH tradition, we first re-express Fernández and Steel's (1998) density as a function of the mean and of the variance and derive its cumulative density function and quantile function. We also proceed to a Monte Carlo simulation to assess its practical applicability in a MLE estimation procedure in the GARCH framework. Finally, we show the usefulness of this method by the analysis of the NASDAQ on the period 1985-1996. Using both in- and out-of-sample density forecast tests, we show that this density seems to be adequate in describing this database compared to the normal and the Student distributions.

2 Several candidates for the density of the innovation process

Let us consider a univariate time series y_t with $t = 1, \dots, T$. If ψ_{t-1} is the information set (i.e. all the information available) at time $t-1$, we can define its functional form as:

$$y_t = E[y_t|\psi_{t-1}] + \varepsilon_t \tag{1}$$

where ε_t is the disturbance term (or unpredictable part). Without loss of generality, we can define an Autoregressive Conditional Heteroscedastic (ARCH) process, $\{\varepsilon_t\}$ by:

$$\begin{aligned} \varepsilon_t &= z_t\sigma_t & (2)\\ z_t &\sim i.i.d.\ (0,1) & (3)\\ \sigma_t &= h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1; \vartheta) & (4) \end{aligned}$$

where z_t is an independently and identically distributed ($i.i.d.$) process with $E(z_t) = 0$, $Var(z_t) = 1$ and ϑ a parameter vector. By definition, ε_t is serially uncorrelated with mean zero, and its conditional variance equals σ_t^2. To estimate this kind of model by maximum likelihood, one has to make an additional assumption on the innovation process by choosing a density function for z_t.

2.1 The normal distribution

A common choice for the distribution of z_t is the normal $N(0,1)$. If θ is the vector of unknown parameters in the model, the log-likelihood function is:

$$\begin{aligned} L_N(\theta) &= \sum_{t=1}^{T} \left[\log g(\varepsilon_t\sigma_t^{-1}) - \log\sigma_t\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T} \left[\ln(2\pi) + \ln\left(\sigma_t^2\right) + z_t^2\right], \qquad (5) \end{aligned}$$

where $g(.)$ is the gaussian probability density function (pdf). This normality assumption is to a certain extent justified by the fact that consistent estimates are found for the first two conditional moments (provided that they are correctly specified), even when normality does not hold.

2.2 The Student distribution

As reported by Palm (1996), Pagan (1996) and Bollerslev, Chou, and Kroner (1992), the use of the Student-t distribution is widespread in the literature. In particular, Bollerslev (1987), Hsieh (1989), Baillie and Bollerslev (1989) and Palm and Vlaar (1997) among others show that this distribution better captures the observed kurtosis. As a reminder, if z_t is distributed as a Student with mean 0, variance 1 and degree of freedom υ (or $t(0,1,\upsilon)$), the log-likelihood function becomes:

$$\begin{aligned} L_{St}(\theta) &= \ln\left[\Gamma\left(\frac{\upsilon+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{\upsilon}{2}\right)\right] - 0.5\ln\left[\pi\left(\upsilon-2\right)\right] \\ &- 0.5\sum_{t=1}^{T}\left[\ln\sigma_t^2 + (1+\upsilon)\ln\left(1+\frac{z_t^2}{\upsilon-2}\right)\right] \end{aligned} \tag{6}$$

where $\upsilon \in I\!N \setminus [0,2]$. Compared to the normal distribution, the Student-t implies the estimation of the additional parameter υ standing for the number of degrees of freedom. The thickness of the tails is decreasing with υ. The constraints on the tail parameter can be relaxed (after reparametrization) by allowing υ to take values in $(0,2]$. In these cases, the variance is infinite and σ_t^2, which is not the variance anymore, remains a dispersion parameter.

2.3 Skewed densities

More recently, Fernández and Steel (1998) proposed an extension of the Student distribution by adding a skewness parameter. Their procedure allows the introduction of skewness in any continuous unimodal and symmetric (about 0) distribution $g(.)$ by changing the scale at each side of the mode. To understand how to build this new family of densities, it is fruitful to express it in terms of a mixture of two truncated densities.

2.3.1 Construction

Let u be an *i.i.d.* continuous random variable with a symmetric unimodal density function $g(.)$ with mean 0 and variance 1,

$$u \sim i.i.d.\ g(0,1) \tag{7}$$

and x, a Bernoulli process, with probability of success $\frac{\xi^2}{1+\xi^2}$. Let us consider the following mixture:

$$\epsilon = x\xi\left|u\right| - (1-x)\frac{1}{\xi}\left|u\right| \tag{8}$$

One can show that the unconditional density $f(\epsilon|\xi)$ of ϵ is:

$$\begin{aligned} f(\epsilon|\xi) &= Prob\,(x=1)\,g(\epsilon|\xi,x=1) + Prob\,(x=0)\,g(\epsilon|\xi,x=0) \\ &= \frac{2}{\xi+\frac{1}{\xi}}\left[g\left(\frac{\epsilon}{\xi}\right)I_{[0,\infty)}\left(\epsilon\right) + g\left(\epsilon\xi\right)I_{(-\infty,0)}\left(\epsilon\right)\right] \end{aligned} \tag{9}$$

Thus, $f(\epsilon|\xi)$ is a unimodal density with the same mode as $g(\epsilon)$ and a skewness parameter $\xi > 0$ such that the ratio of probability masses above and below the mode is:

$$\frac{\Pr(\epsilon \geq 0|\xi)}{\Pr(\epsilon < 0|\xi)} = \xi^2 \tag{10}$$

Note that the density $f(\epsilon|1/\xi)$ is the symmetric of $f(\epsilon|\xi)$ with respect to the mode. Therefore, working with $\xi' = \log(\xi)$ might be preferable to indicate the sign of the skewness. If we set $y_t = \mu_t + \epsilon_t\sigma_t$, where ϵ_t has a skewed Student distribution (as obtained by considering a Student distribution with mean 0 and variance 1 for u in Equation (7)), then we obtain a distribution for y_t where all these parameters have a clear interpretation:

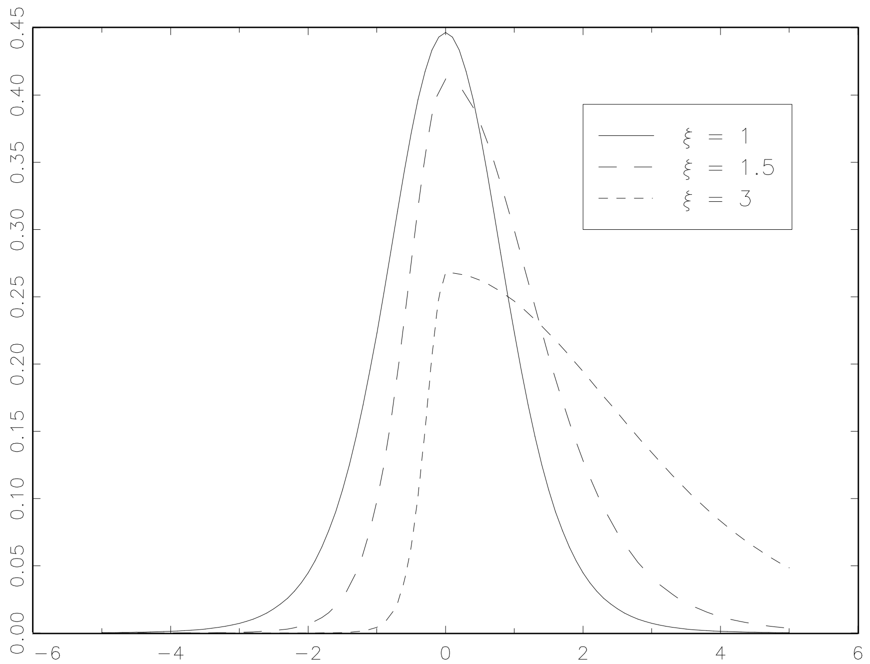

Figure 1: Skewed Student densities with $\upsilon = 8$ and $\xi = 1$, 1.5 and 3.

- μ_t, as the conditional mode, models the location,
- $\sigma_t^2 > 0$ (which is not the conditional variance anymore) models the dispersion,
- $\xi > 0$ models the skewness,
- $\upsilon > 0$ models the tail thickness.

Note that the four important aspects of the distribution can thus be specified. This density has been used successfully by Lambert and Laurent (2001) and Von Rohr and Hoeschele (1999). The skewed normal distribution, directly obtained by taking for $g(u)$ in Equation (7), is a limiting case ($\upsilon \to \infty$) of the skewed Student with the same tail properties as the traditional normal.

2.3.2 Moments

Fernández and Steel (1998) show that if the r^{th} ($r \in \Re$) order moment of $g(.)$ exists, the associated skewed distribution in Equation (9) also has a finite r^{th} moment. In particular,

$$E\left(\epsilon^r|\xi\right) = M_r \frac{\xi^{r+1} + \frac{(-1)^r}{\xi^{r+1}}}{\xi + \frac{1}{\xi}} \tag{11}$$

where

$$M_r = \int_0^\infty 2s^r g\left(s\right) ds, \tag{12}$$

and M_r is the r^{th} order moment of $g(.)$ truncated to the positive real values. Provided that these quantities are finite, we can easily obtain:

$$\begin{aligned} E\left(\epsilon|\xi\right) &= M_1\left(\xi - \frac{1}{\xi}\right) \\ V\left(\epsilon|\xi\right) &= E\left(\epsilon^2|\xi\right) - E\left(\epsilon|\xi\right)^2 \end{aligned} \tag{13}$$

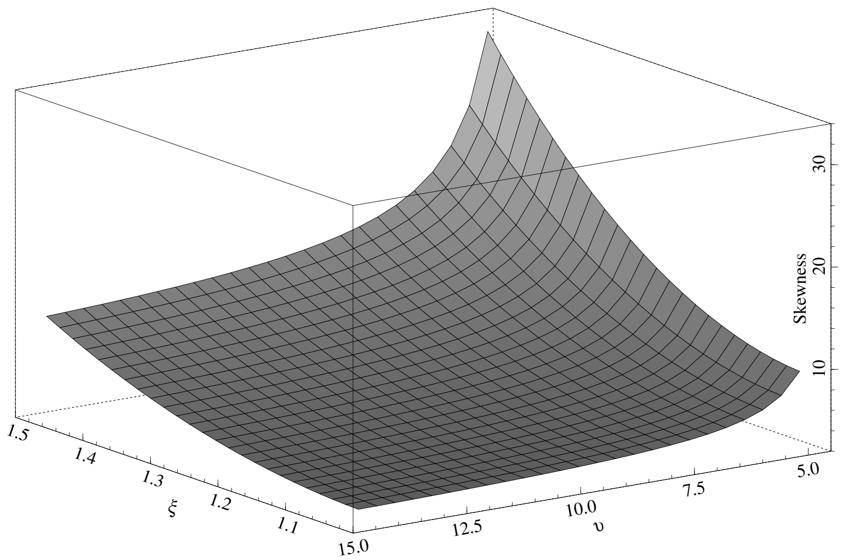

Figure 2: Skewness implied by several combinations of $1 \leq \xi \leq 1.5$ and $3.5 \leq \upsilon \leq 15$

$$= \left(M_2 - M_1^2\right)\left(\xi^2 + \frac{1}{\xi^2}\right) + 2M_1^2 - M_2 \tag{14}$$

$$Sk\left(\epsilon|\xi\right) = \frac{E\left(\epsilon^3|\xi\right) - 3E\left(\epsilon|\xi\right)E\left(\epsilon^2|\xi\right) + 2E\left(\epsilon|\xi\right)^3}{Var\left(\epsilon|\xi\right)^{\frac{3}{2}}}$$

$$= \frac{\left(\xi - \frac{1}{\xi}\right)\left(M_3 + 2M_1^3 - 3M_1M_2\right)\left(\xi^2 + \frac{1}{\xi^2}\right) + 3M_1M_2 - 4M_1^3}{Var\left(\epsilon|\xi\right)^{\frac{3}{2}}} \tag{15}$$

$$Ku\left(\epsilon|\xi\right) = \frac{E\left(\epsilon^4|\xi\right) - 4E\left(\epsilon|\xi\right)E\left(\epsilon^3|\xi\right) + 6E\left(\epsilon^2|\xi\right)E\left(\epsilon|\xi\right)^2 - 3E\left(\epsilon|\xi\right)^4}{Var\left(\epsilon|\xi\right)^2} \tag{16}$$

where $E\left(.|\xi\right)$, $V\left(.|\xi\right)$, $Sk\left(.|\xi\right)$ and $Ku\left(.|\xi\right)$ are respectively the mean, variance, skewness and kurtosis, given ξ.[6] As shown in Equations (15) and (16), both ξ and υ define skewness and kurtosis. Figures 2 and 3 investigate the relation between these two parameters and the skewness (with $\upsilon > 3$ to ensure the existence of the skewness). For simplicity, we do not tackle the case $0 < \xi \leq 1$ and only report the graphs when $\xi \geq 1$. It is clear from these two figures that the dominating feature of skewness is the ξ parameter.[7] From figure 3 we can see that skewness may be very high when υ approaches 3. Figures 4 and 5 traces the kurtosis surface for several combinations of $\xi > 1$ and $\upsilon > 4$ (to insure the existence of the kurtosis). The dominating feature of kurtosis is obviously the υ parameter, even if the higher the asymmetry parameter, the higher the kurtosis. Consequently, even if both ξ and υ determine skewness and kurtosis, Figures 2-5 show that skewness (resp. kurtosis) is mainly governed by ξ (resp. υ).

2.3.3 Standardized skewed Student density

One drawback of this parameterization of the skewed Student density is that μ_t and σ_t^2 are not the conditional mean and the conditional variance but the conditional mode and some measure of conditional dispersion. In order to keep in the ARCH tradition, it is important to express the density in terms of the mean and of the variance, and, thus, to reparameterize Equation (9). In

[6] Even if a closed form of the kurtosis is theoretically available, it is not tractable.

[7] This property applies also for Hansen's skewed Student density. See Jondeau and Rockinger (2000).

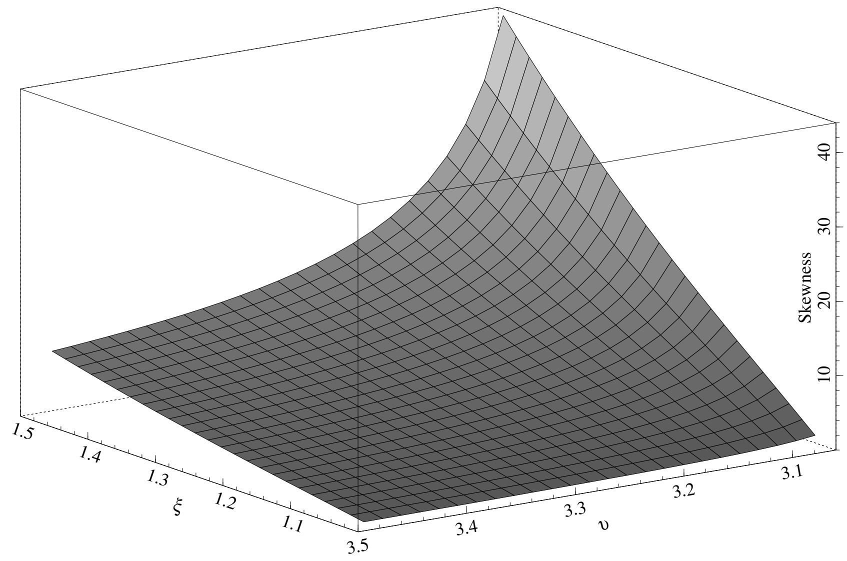

Figure 3: Skewness implied by several combinations of $1 \leq \xi \leq 1.5$ and $3.05 \leq \upsilon \leq 3.5$

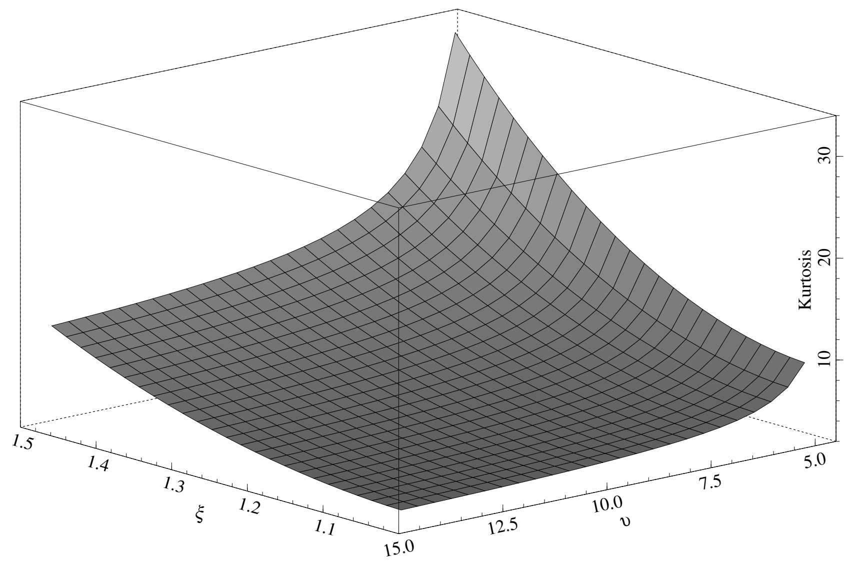

Figure 4: Kurtosis implied by several combinations of $1 \leq \xi \leq 1.5$ and $4.5 \leq \upsilon \leq 15$

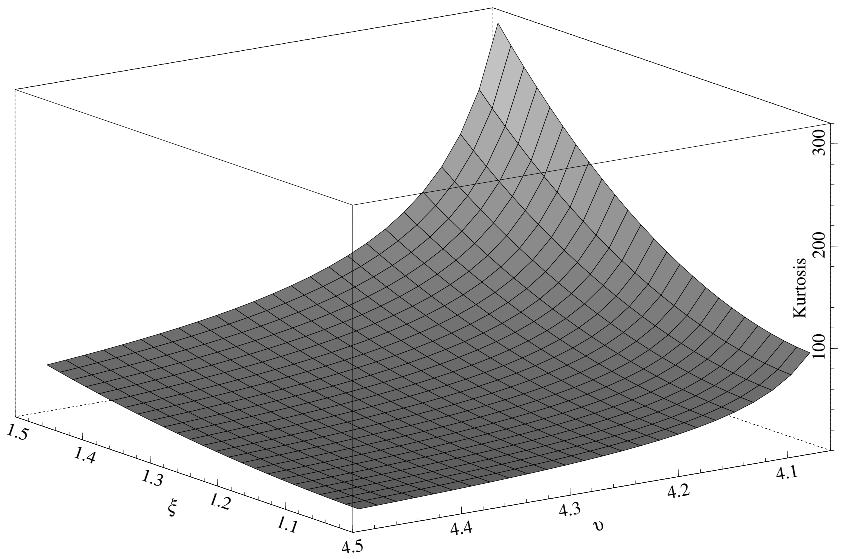

Figure 5: Kurtosis implied by several combinations of $1 \leq \xi \leq 1.5$ and $4.05 \leq \upsilon \leq 4.5$

such a way it will be possible to take z_t in Equation 3 a skewed Student distribution with zero mean and unit variance. More specifically, assume that ϵ has a Student distribution with density $g(.)$ and degree of freedom υ. Then, the r^{th} moment of ϵ truncated to the positive real values is:

$$M_{r|\upsilon} = \frac{\Gamma\left(\frac{\upsilon-r}{2}\right)\Gamma\left(\frac{1+r}{2}\right)(\upsilon-2)^{\frac{1+r}{2}}}{\sqrt{\pi(\upsilon-2)}\Gamma\left(\frac{\upsilon}{2}\right)} \tag{17}$$

Using Equation (13) and (14), and provided that $\upsilon > 2$, it follows that:

$$E(\epsilon|\xi,\upsilon) = \frac{\Gamma\left(\frac{\upsilon-1}{2}\right)\sqrt{\upsilon-2}}{\sqrt{\pi}\Gamma\left(\frac{\upsilon}{2}\right)}\left(\xi - \frac{1}{\xi}\right) \equiv m \tag{18}$$

and

$$V(\epsilon|\xi,\upsilon) = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2 \equiv s^2 \tag{19}$$

Now consider the standardized random variable

$$z_t = \frac{\epsilon_t - m}{s} \tag{20}$$

It has mean 0, variance 1 and density:

$$f(z|\xi,\upsilon) = \frac{2}{\xi + \frac{1}{\xi}} s\left\{g\left[\xi(sz+m)|\upsilon\right]I_{(-\infty,0)}(z+m/s) + g\left[(sz+m)/\xi|\upsilon\right]I_{[0,\infty)}(z+m/s)\right\} \tag{21}$$

For a standardized skewed Student, the log-likelihood is:

$$\begin{aligned} L_{SkSt}(\theta) &= \ln\left[\Gamma\left(\frac{\upsilon+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{\upsilon}{2}\right)\right] - 0.5\ln\left[\pi(\upsilon-2)\right] + \ln\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + \ln(s) \\ &- 0.5\sum_{t=1}^{T}\left[\ln\sigma_t^2 + (1+\upsilon)\ln\left(1 + \frac{sz_t+m}{\upsilon-2}\xi^{-I_t}\right)\right] \end{aligned} \tag{22}$$

where $I_t = \begin{cases} 1 \text{ if } z_t \geq -\frac{m}{s} \\ -1 \text{ if } z_t < -\frac{m}{s} \end{cases}$. Figure 6 displays several standardized skewed Student densities with $\upsilon = 5, 30, +\infty$ and $\xi = 1, 1.3, 1.5$ and 2.

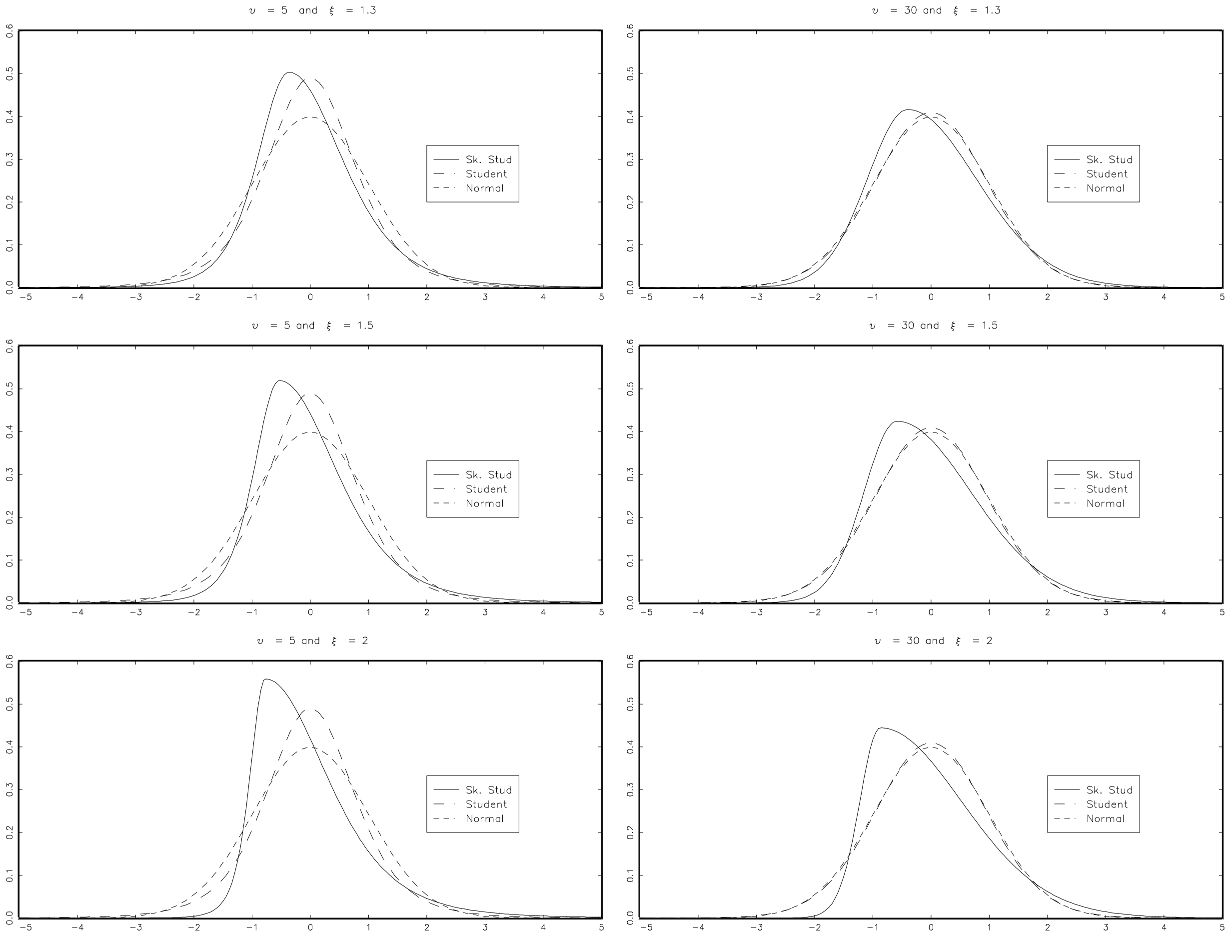

Figure 6: Normal, Student and skewed Student densities.

2.3.4 Distribution and quantile functions of a skewed distribution

Using the same notation and hypotheses as in the previous section, we can relate the cumulative distribution function (cdf) F and the quantile function F^{-1} corresponding to a standardized skewed density $f(z|\xi)$ to the cdf G and quantile function G^{-1} of the original symmetric density. We have

$$F(z|\xi) = \begin{cases} \frac{2}{1+\xi^2}\frac{1}{s}\, G(\xi(sz+m)) & \text{if} \quad z < -\frac{m}{s} \\ 1 - \frac{2}{1+\xi^{-2}}\frac{1}{s}\, G(-\xi^{-1}(sz+m)) & \text{if} \quad z \geq -\frac{m}{s} \end{cases} \tag{23}$$

for the cdf and

$$F^{-1}(p|\xi) = \begin{cases} \frac{\frac{1}{\xi}G^{-1}\left(\frac{p}{2}(1+\xi^2)\right)-m}{s} & \text{if} \quad p < \frac{1}{1+\xi^2} \\ \frac{-\xi G^{-1}\left(\frac{1-p}{2}(1+\xi^{-2})\right)-m}{s} & \text{if} \quad p \geq \frac{1}{1+\xi^2} \end{cases} \tag{24}$$

for the quantile function.

3 GARCH model with skewed distribution for the innovations

Before analyzing real data, and in order to assess the practical applicability of the MLE procedure of the skewed Student distribution, we first present the results of a simulation study. It is not our intention to provide a comprehensive Monte Carlo study of the MLE. The reliability of the inference concerning the model parameters will not be examined. Our results, however, will provide some preliminary evidence with respect to the finite sample properties of the MLE when a skewed Student distribution for the innovations is coupled with a GARCH model. More specifically, consider a GARCH(1,1) model:

$$\begin{aligned} y_t &= \mu + \varepsilon_t & (25) \\ \varepsilon_t &= \sqrt{h_t} z_t & (26) \\ h_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2 & (27) \end{aligned}$$

with $z_t \sim N(0,1)$ or $z_t \sim t(0,1,\upsilon)$, or $z_t \sim st(0,1,\xi,\upsilon)$ with $t = 1,...,3000$. Consider for the parameters typical values found in the empirical literature relative to daily financial returns: $\mu = 0, \alpha_0 = 0.1, \alpha_1 = 0.1, \beta_1 = 0.8$. To avoid start-up problem, the first 3000 realizations (out of 6000) were discarded for each replication, leaving the announced 3000 data. To gauge the accuracy of MLEs, Tables 1, 2 and 3 report estimation results of the above mentioned models, under three different error assumptions, the true Data Generating Process (DGP) being (25-27) respectively with $z_t \sim st(0,1,\exp(0.3), 8.0)$, $z_t \sim \chi^2(3)$ and $z_t \sim \Gamma(1,2)$, where χ^2 and $\Gamma(.)$ denote respectively the Chi-square and Gamma distributions.[8] The last two lines report the generated and estimated skewness and kurtosis. ¿From Table 1, it is clear that the maximum likelihood method for the GARCH model, under the right density (i.e. the skewed student, see column 5), works reasonably well with the considered sample size. Table 1 also illustrates the well known result of Weiss (1986) and Bollerslev and Wooldridge (1992) that (if the mean and the variance are specified correctly) the quasi- (or pseudo-) maximum likelihood estimator under the normal and Student assumptions are respectively consistent (but inefficient) and inconsistent. Moreover, the Monte Carlo results suggest that the MLE of the skewed Student (with a GARCH model governing the conditional variance) are only slightly biased and are more efficient than QMLE. Tables 2 and 3 show that the skewed Student does a good job in modelling the first and second moments when the errors are Chi-square or Gamma distributed (who are both skewed and kurtosed) leading to very small biases in the mean and variance parameters (as the normal density), compared to the usual Student density.

[8] The Chi-square and Gamma distributions have been standardized in order to have mean 0 and variance 1.

	DGP	normal	Student	skewed Student
μ	0.0	-0.0008	-0.0531	0.0009
		[0.0133]	[0.0136]	[0.0131]
α_0	0.1	0.1067	0.1022	0.1031
		[0.0237]	[0.0192]	[0.0181]
α_1	0.1	0.1013	0.0974	0.1008
		[0.0169]	[0.0136]	[0.0134]
β_1	0.8	0.7914	0.7994	0.7963
		[0.0347]	[0.0280]	[0.0264]
$\log(\xi)$	0.3	-	-	0.3008
				[0.0209]
υ	8.0	-	7.1649	8.2000
			[0.7096]	[0.8794]
Skewness	0.76	0	0	0.75
Kurtosis	5.13	3	5.44	5.03

Table 1: Skewed Student errors. Data Generating process: $y_t = \varepsilon_t = \sqrt{h_t} z_t$, $h_t = 0.1 + 0.1\varepsilon_{t-1}^2 + 0.8h_{t-1}^2$ and $z_t \sim st(0, 1, \exp(0.3), 8.0)$. Note: Robust Standard deviations of the estimated parameters are reported in brackets. The last two lines report the generated and estimated skewness and kurtosis.

	DGP	normal	Student	skewed Student
μ	0.0	0.0001	-0.1779	-0.0021
		[0.0133]	[0.0161]	[0.0122]
α_0	0.1	0.1049	0.1035	0.1023
		[0.0250]	[0.0182]	[0.0067]
α_1	0.1	0.1027	0.0968	0.1018
		[0.0197]	[0.0134]	[0.0055]
β_1	0.8	0.7919	0.8095	0.7990
		[0.0376]	[0.0247]	[0.0079]
$\log(\xi)$	-	-	-	1.8591
				[0.1344]
υ	-	-	3.8108	7.1504
			[0.2400]	[0.6885]
Skewness	1.63	0	0	1.75
Kurtosis	7	3	11.40	9.02

Table 2: Chi-square errors. Data Generating process: $y_t = \varepsilon_t = \sqrt{h_t} z_t$, $h_t = 0.1 + 0.1\varepsilon_{t-1}^2 + 0.8h_{t-1}^2$ and $z_t \sim \chi^2(3)$. Note: see Table 1.

	DGP	normal	Student	skewed Student
μ	0.0	-0.0009	-0.1407	-0.0056
		[0.0133]	[0.0160]	[0.0128]
α_0	0.1	0.1048	0.1006	0.1004
		[0.0246]	[0.0181]	[0.0092]
α_1	0.1	0.1011	0.0938	0.1003
		[0.0186]	[0.0130]	[0.0072]
β_1	0.8	0.7940	0.8080	0.7997
		[0.0365]	[0.0258]	[0.0129]
$\log(\xi)$	-	-	-	1.2207
				[0.0589]
υ	-	-	4.7417	11.1242
			[0.3487]	[1.6085]
Skewness	1.41	0	0	1.33
Kurtosis	6	3	7.44	5.74

Table 3: Gamma errors. Data Generating process: $y_t = \varepsilon_t = \sqrt{h_t} z_t$, $h_t = 0.1 + 0.1\varepsilon_{t-1}^2 + 0.8h_{t-1}^2$ and $z_t \sim \Gamma(1,2)$. Note: see Table 1.

4 Density forecasts

To compare the adequacy of the different distributions, we employ in- and out-of-sample density forecasts proposed by Diebold, Gunther, and Tay (1998) (henceforward DGT).[9] The idea of density forecasts is quite simple. Let $f_i(y_i|\Omega_i)_{i=1}^m$ be a sequence of m one-step-ahead density forecasts produced by a given model, where Ω_i is the conditioning information set, and $p_i(y_i|\Omega_i)_{i=1}^m$ the sequence of densities defining the Data Generating Process y_i (which is never observed). The main question to answer is then: is the density suitable ? If the answer is yes, it means that:

$$H_0 : f_i(y_i|\Omega_i)_{i=1}^m = p_i(y_i|\Omega_i)_{i=1}^m \tag{28}$$

DGT use the fact that under (28), the probability integral transform $\zeta_i = \int_{-\infty}^{y_i} f_i(t)dt$ is *i.i.d.* $U(0,1)$, i.e. independent and identically distributed uniform. To check H_0, they propose to use goodness-of-fit test and independence test for *i.i.d.* $U(0,1)$. The *i.i.d.*-ness property of ζ_i can be evaluated by plotting the correlograms of $\left(\zeta - \overline{\zeta}\right)^j$, for $j = 1,2,3,4,...$, to detect potential dependence in the conditional mean, variance, skewness, kurtosis, etc. Departure from uniformity can also be evaluated by plotting an histogram of ζ_i. According to Bauwens, Giot, Grammig, and Veredas (2000), *a humped shape of the ζ-histogram would indicate that the issued forecasts are too narrow and that the tails of the true density are not accounted for. On the other hand, a U-shape of the histogram would suggest that the model issues forecasts that either under- or overestimate too frequently.*[10] To illustrate the usefulness of this testing procedure, Figures 7 to 9 plot the *ζ-histograms* (with 40 cells) of 5000 one-step-ahead forecasts based on the same GDP as in the previous section. In Figure 7, $t(0,1,8)$ errors are generated while normality is assumed for the innovations when computing the MLEs. In Figures 8 and 9 skewed Student $st(0,1,exp(0.3),8)$

[9]For more details about density forecasts and applications in finance, see the special issue of *Journal of Forecasting* (Timmermann, 2000)

[10]Confidence intervals for the ζ-histogram can be obtained by using the properties of the histogram under the null hypothesis of uniformity. We would like to thank Luc Bauwens and his coauthors for providing us a procedure to compute them.

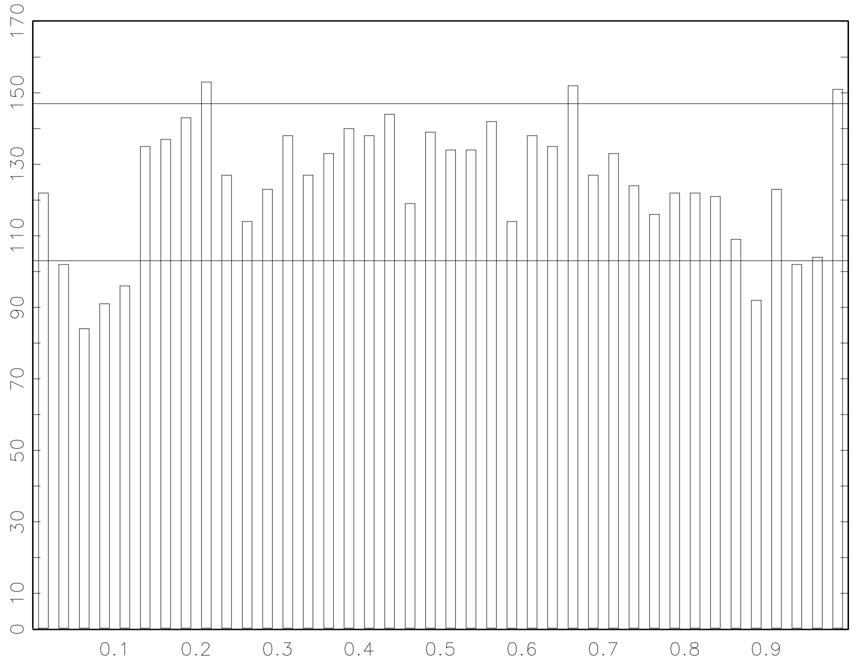

Figure 7: ζ-histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $t(0,1,8)$ errors. The MLEs were computed assuming normality for the innovations.

errors are generated while respectively symmetric and asymmetric Student errors are assumed. Figures 7 and 8 clearly suggest that the assumption made on the error term is not appropriate. Moreover, Figure 8 show that an *inverted S* shape of the histogram would indicate that the errors are skewed, i.e. the true density is probably not symmetric. However, from Figure 9, it is clear that the probability integral transform is Uniformly distributed. To check the uniformity of ξ, we rely on the Pearson goodness-of-fit test that compares the empirical distribution with the theoretical one. In order to carry out this testing procedure, one first needs to classify the residuals in cells according to their magnitude.[11] Then, for a given number of cells denoted κ, one computes the following test statistic:

$$P(\kappa) = \sum_{i=1}^{\kappa} \frac{(n_i - En_i)^2}{En_i} \tag{29}$$

where n_i is the number of observations in group i and En_i, the expected number of observations (based on the estimated MLE). For *i.i.d.* observations, under the null of a correct distribution, $P(\kappa)$ has an asymptotic $\chi^2(\kappa-1)$ distribution. A practical problem with the Chi-square test (29) is that the statistic does not have a χ^2 distribution if some of the parameters are estimated. Chernoff and Lehmann (1954) derived the asymptotic distribution for the test statistic if the parameters are estimated by maximum likelihood based on ungrouped data. Based on this result, Vlaar and Palm (1993) show that the asymptotic distribution of $P(\kappa)$ is bounded between a $\chi^2(\kappa-1)$ and a $\chi^2(\kappa-k-1)$ where k is the number of estimated parameters.

5 Application

The analyzed database consists of 3000 observations of the NASDAQ from January 1985 until December 1996. As pointed out by El Babsiri and Zakoian (2001), although asymmetric GARCH

[11] See Vlaar and Palm (1993) for more details.

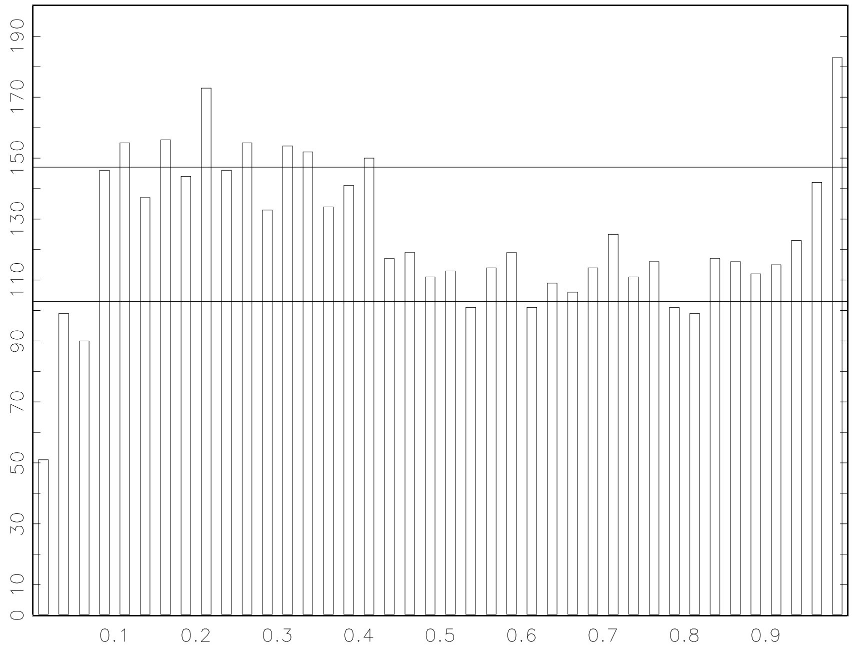

Figure 8: ζ-histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $st(0, 1, \exp(0.3), 8)$ errors. The MLEs were computed assuming the innovations are Student-t distributed.

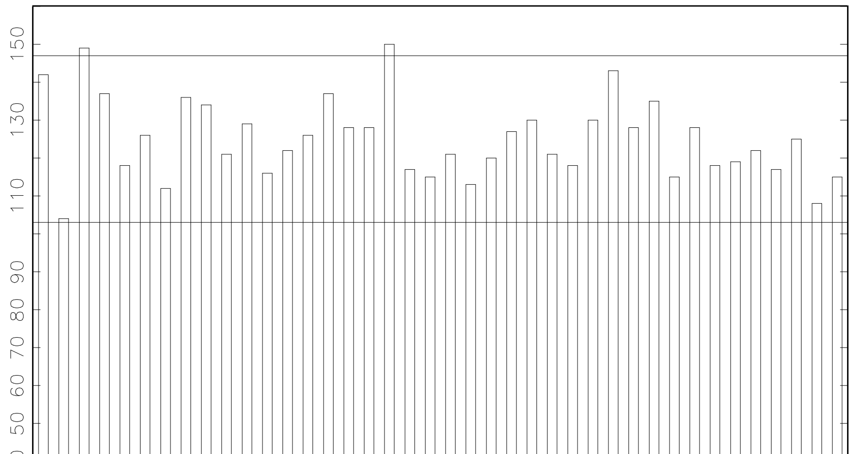

Figure 9: ζ-histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $st(0, 1, \exp(0.3), 8)$ errors. The MLEs were computed assuming the innovations are skewed Student-t distributed

models allow positive and negative changes to have different impacts on future volatilities, the two components of the innovation have - up to a constant - the same volatilities, while it is desirable to allow an asymmetric confidence interval around the prediction value. Here, we propose to analyze that long time series by assuming normal, Student and skewed Student distributions for the innovations. Dynamics will be introduced in the conditional mean and the conditional variance with an AR(1)-APARCH(1,1) specification:

$$y_t = \mu + \rho(y_{t-1} - \mu) + \varepsilon_t \tag{30}$$

$$\varepsilon_t = \sigma_t z_t \tag{31}$$

$$\sigma_t^\delta = \alpha_0 + \alpha_1 \left(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1}\right)^\delta + \beta_1 \sigma_{t-1}^\delta \tag{32}$$

where $\mu, \rho, \alpha_0, \alpha_1, \beta_1, \gamma$ and δ are parameters to be estimated. The APARCH is probably one of the most promising ARCH-type model. Indeed, it nests at least seven GARCH models (see Ding, Granger, and Engle, 1993). δ ($\delta > 0$) plays the role of a Box-Cox transformation of σ_t, while γ ($-1 < \gamma < 1$), reflects the so-called leverage effect[12] (see Black, 1976; French, Schwert, and Stambaugh, 1987; Pagan and Schwert, 1990). Following Ding, Granger, and Engle (1993), if it exists, a stationary solution of (32) is given by:

$$E\left(\sigma_t^\delta\right) = \frac{\alpha_0}{1 - \alpha_1 E\left(|z| - \gamma z\right)^\delta - \beta_1} \tag{33}$$

which depends on the density of z. Such a solution exists if $\alpha_1 E\left(|z| - \gamma z\right)^\delta + \beta_1 < 1$.[13] Notice that if we set $\gamma = 0$, $\delta = 2$ and that z_t has zero mean and unit variance, we recover the stationarity condition of the GARCH(1,1) model ($\alpha_1 + \beta_1 < 1$). Ding, Granger, and Engle (1993) derived the expression $E\left(|z| - \gamma z\right)^\delta$ in the gaussian case. Paolella (1997) gives such expressions for various non standardized densities. It can be shown that for the standardized skewed Student:[14]

$$E\left(|z| - \gamma z\right)^\delta = \left\{\xi^{-(1+\delta)}\left(1+\gamma\right)^\delta + \xi^{1+\delta}\left(1-\gamma\right)^\delta\right\} \frac{\Gamma\left(\frac{\delta+1}{2}\right)\Gamma\left(\frac{\upsilon-\delta}{2}\right)(\upsilon-2)^{\frac{1+\delta}{2}}}{\left(\xi+\frac{1}{\xi}\right)\sqrt{(\upsilon-2)\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)}$$

Table 4 hereafter presents the quasi-maximum likelihood estimation results of the AR(1)-APARCH(1,1) under the normal, Student and skewed Student assumptions for the first 2000 observations while Table 5 reports some statistics of interest. These results have been obtained using G@RCH 2.0 (Laurent and Peters, 2001), an Ox 2.20 package (see Doornik, 1999) with a friendly interface dedicated to the estimation of ARCH-type models.[15] Let us comment Tables 4 and 5:

1. First, the extra flexibility of the APARCH specification is required. Both the asymmetry coefficient (γ) and the power (δ) estimates suggest that a usual GARCH model is not appropriate to model the NASDAQ. This is also confirmed by likelihood ratio tests (not reported here to save space).

2. Second, likelihood ratio tests (not reported) and the AIC criterion clearly suggest that the skewed Student outperforms the two other candidates. The distribution of the NASDAQ is highly kurtosed and left skewed. To illustrate the difference between the normal, the Student and the skewed Student, Figure 10 plots the fitted densities of the innovations, namely a $st(0, 1, \exp(-0.179), 6.039)$ (solid line), a $t(0, 1, 5.57)$ (dashed line) and a $N(0, 1)$ (short dashes). The asymmetry coefficient equals -0.179, which means the skewed Student density allocates nearly 59% of the mass to the left side of the mode. It seems moreover that the asymmetry feature of the APARCH model (characterizing the conditional variance) and the "skewness" coefficient of the unconditional density are both necessary to explain the overall asymmetry of the series.

[12] A positive (resp. negative) value of γ means that past negative (resp. positive) shocks have a deeper impact on current conditional volatility than past positive shocks.

[13] $\alpha_1 E\left(|z| - \gamma z\right)^\delta + \beta_1$ may be viewed as a measure of volatility persistence.

[14] Notice that setting $\xi = 1$ leads to the stationarity condition of the symmetric Student density (with unit variance).

[15] The G@RCH web site is: http://www.egss.ulg.ac.be/garch/

	normal	Student	skewed Student
μ	0.0554	0.1010	0.0626
	(0.0206)	(0.0195)	(0.0197)
ρ	0.2946	0.2861	0.2787
	(0.0288)	(0.0298)	(0.0239)
α_0	0.0533	0.0374	0.0371
	(0.0231)	(0.0134)	(0.0132)
α_1	0.1641	0.1451	0.1503
	(0.0377)	(0.0255)	(0.0261)
γ	0.3768	0.2787	0.2419
	(0.1124)	(0.1064)	(0.0953)
β_1	0.8024	0.8258	0.8229
	(0.0537)	(0.0368)	(0.0360)
δ	1.0544	1.2914	1.3322
	(0.3794)	(0.2929)	(0.2967)
$\log(\xi)$	0	0	-0.1789
			(0.0319)
υ	∞	5.5706	6.0388
		(0.7490)	(0.8182)
Log-Lik	-2090.6	-2008.7	-1993.1

Table 4: Estimation results. Asymptotic heteroskedasticity-consistent standard errors are given in parenthesis. *Log-Lik* refers to the log-likelihood value at maximum.

3. The stationary condition of the APARCH model is satisfied for the three distributions, as $\alpha_1 E\left(|z| - \gamma z\right)^{\delta} + \beta_1 < 1$ (at the MLEs).
4. The AR(1)-APARCH(1,1) seems to be adequate in describing the dynamic of the first two moments of the NASDAQ, for the period of interest. Indeed, the Box-Pierce statistics[16] $Q(20)$ and $Q^2(20)$ are all non significant at any reasonable level.
5. The relevance of the skewed Student distribution is also confirmed by the Pearson goodness-of-fit statistic, $P(50)$ and $P(60)$. While the normal and the Student distributions are rejected indubitably (the p-values equal about 0), the skewed Student density seems to be supported by the data (both by the non adjusted and adjusted tests with 50 and 60 cells).

Finally, to assess the relevance of the skewed Student density, we perform some out-of-sample forecasts. Table 6 gives the goodness-of-fit tests (density forecasts test) on the one-day-ahead forecasts of the AR(1)-APARCH(1,1). This test has been conducted on the last 1000 observations (about 4 years), using the estimated parameters reported in Table 4. From Table 6, it is obvious that both the normal and the Student pdf's are not adequate for density forecast purposes. On the other hand, the skewed Student passes this test (with less evidence for the adjusted version with 50 cells).

[16] The level of significance of the Box-Pierce statistics on the standardized residuals have to be adjusted by the number of ARMA parameters while the Box-Pierce statistics on the squared standardized residuals have to be adjusted by the number of GARCH parameters (see Bollerslev and Mikkelsen, 1996).

	normal	Student	skewed Student
$\alpha_1 E(\|z\| - \gamma)^\delta + \beta_1$	0.934	0.940	0.967
Q(20)	25.009	23.926	26.435
$Q^2(20)$	10.185	11.625	10.804
P(50)	122.120	78.176	50.648
	(0.000)	(0.005)	(0.408)
	[0.000]	[0.000]	[0.120]
P(60)	110.828	85.183	52.210
	(0.000)	(0.014)	(0.722)
	[0.001]	[0.001]	[0.388]
AIC	2.099	2.019	2.004

Table 5: Statistics of interest. Q(20) and $Q^2(20)$ are respectively the Box-Pierce statistics at lag 20 of the standardized and squared standardized residuals. P(50) and P(60) are the Pearson Goodness-of-fit statistics with 50 and 60 cells. P-values of the non adjusted and adjusted test are given respectively in parenthesis and in brackets.

	normal	Student	skewed Student
P(50)	122.120	78.100	57.600
	(0.006)	(0.005)	(0.187)
	[0.000]	[0.001]	[0.035]
P(60)	117.915	88.160	56.120
	(0.027)	(0.008)	(0.582)
	[0.001]	[0.005]	[0.256]

Table 6: Density forecast test - Pearson Goodness-of-fit test. P-values of the non adjusted and adjusted test are given respectively in parentheses and in brackets.

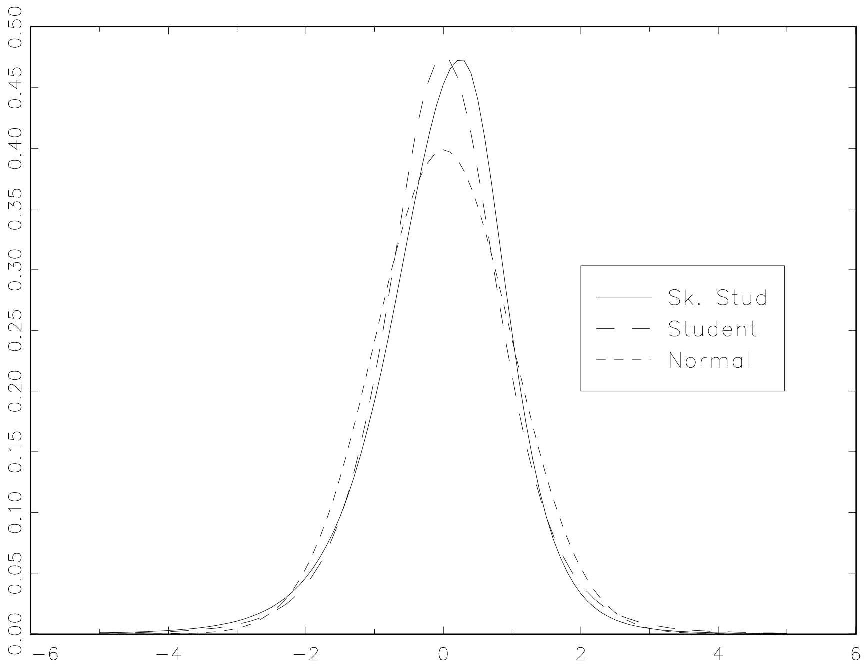

Figure 10: $st(0, 1, \exp(-0.179), 6.039)$ (solid line), $t(0, 1, 5.57)$ (dashed line) and $N(0, 1)$ (short dashes).

6 Conclusion

In this paper we first parametrized the skewed Student density proposed by Fernández and Steel (1998) in terms of the mean and of the variance parameters. This density is very promising in many respects. First, we have shown its practical applicability in a MLE estimation procedure in the GARCH framework using a Monte Carlo simulation. The very small biases found in the parameter estimates of a GARCH(1,1) model when (wrongly) assuming a skewed Student distribution (see Section 3) makes it a worthwhile alternative to pseudo-likelihood methods as in addition it allows for skewness and kurtosis. Moreover, this density, based on a mixture of two truncated symmetric densities, is easy to implement. Indeed, its pdf, cdf and inverse cdf are based on the corresponding functions of its symmetric versions (which are available in most statistical packages). We have shown the practical advantages of the skewed Student distribution by analyzing the NASDAQ on a 12 year period (on a daily basis). Pearson goodness-of-fit tests reject the normal and the Student, but not the skewed Student. Finally, the adequacy of this specification has been reinforced by in- and out-of-sample one-step-ahead density forecast tests. The skewed Student density appears to be a promising specification to accommodate both the high kurtosis and the skewness inherent to most asset returns.

Acknowledgment

The authors would like to thank G. Haesbroeck, B. Lejeune, J.K. Lindsey, J. Rombouts, J-P. Urbain and the participants of the econometric seminar of the CORE and the University of Maastricht for helpful comments.

References

Baillie, R., and T. Bollerslev (1989): "The Message in Daily Exchange Rates: A Conditional-Variance Tale," *Journal of Business and Economic Statistics*, 7, 297–305.

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2000): "A Comparison of Financial Duration Models Via Density Forecasts," CORE DP 2060.

Beine, M., and S. Laurent (1999): "Central Bank Interventions and Jumps in Double Long Memory Models of Daily Exchange Rates," Mimeo, University of Liège.

Black, F. (1976): "Studies of Stock Market Volatility Changes," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

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

——— (1987): "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542–547.

Bollerslev, T., R. Chou, and K. Kroner (1992): "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.

Bollerslev, T., and H. O. Mikkelsen (1996): "Modeling and Pricing Long-Memory in Stock Market Volatility," *Journal of Econometrics*, 73, 151–184.

Bollerslev, T., and J. Wooldridge (1992): "Quasi-maximum likelihood estimation inference in dynamic models with time-varying covariances," *Econometric Theory*, 11, 143–172.

Bond, S. (2000): "A Review of Asymmetric Conditional Density Functions in Autoregressive Conditional Heteroscedasticity Models," mimeo, Duke University, Durham.

Chernoff, H., and E. Lehmann (1954): "The Maximum Likelihood Estimate in Chi-Square Tests for Goodness of Fit," *Annals of Mathematical Statistics*, 25, 579–586.

Chunhachinda, P., K. Dandapani, S. Hamid, and A. Prakash (1997): "Portfolio selection and skewness: evidence from international stock markets," *Journal of Banking and Finance*, 21, 143–167.

Corrado, C., and T. Su (1996): "Skewness and kurtosis in S&P 500 index returns implied by option prices," *Journal of Financial Research*, 19, 175–192.

——— (1997): "Implied volatility skews and stock return skewness and kurtosis implied by stock option prices," *European Journal of Finance*, 3, 73–85.

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

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

Doornik, J. A. (1999): *An Object Oriented Matrix Programming Language*. Timberlake Consultant Ltd., third edn.

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

ENGLE, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

ENGLE, R., AND G. GONZÁLEZ-RIVERA (1991): "Semiparametric ARCH Model," *Journal of Business and Economic Statistics*, 9, 345–360.

FERNÁNDEZ, C., AND M. STEEL (1998): "On Bayesian modelling of fat tails and skewness," *Journal of the American Statistical Association*, 93, 359–371.

FRENCH, K., G. SCHWERT, AND R. STAMBAUGH (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

GIOT, P. (2000): "Intraday Value-at-Risk," CORE DP 2045, Maastricht University METEOR RM/00/030.

GIOT, P., AND S. LAURENT (2001): "Valut-at-Risk for Long and Short Positions," CORE DP 2001-22, Maastricht University METEOR RM/01/005.

HANSEN, B. (1994): "Autoregressive conditional density estimation," *International Economic Review*, 35, 705–730.

HARVEY, C., AND A. SIDDIQUE (1999): "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34, 465–487.

HSIEH, D. (1989): "Modeling Heteroskedasticity in Daily Foreign Exchange Rates," *Journal of Business and Economic Statistics*, 7, 307–317.

JONDEAU, E., AND M. ROCKINGER (2000): "Conditional Volatility, Skewness, and Kurtosis: Existence and Persistence," Mimeo, HEC-School of Management, Jouyen-Josas, Fance.

JORION, P. (1981): "On jump processes in the foreign exchange and stock markets," *The Review of Financial Studies*, 68, 165–176.

KNIGHT, J., S. SATCHELL, AND K. TRAN (1995): "Statistical Modelling of Asymmetric Risk in Asset Returns," *Applied Mathematical Finance*, 2, 155–172.

LAMBERT, P., AND S. LAURENT (2001): "Modeling skewness dynamics in series of financial data using skewed location-scale distributions," Discussion Paper 0119, Institut de Statistique, Louvain-la-Neuve.

LAURENT, S., AND J.-P. PETERS (2001): "G@RCH 2.0 : An Ox Package for Estimating and Forecasting Various ARCH Models," Mimeo, Université de Liège.

LEE, Y., AND T. TSE (1991): "Term Structure of Interest Rates in the Singapore Asian Dollar Market," *Journal of Applied Econometrics*, 6, 143–152.

LIU, S.-M., AND B. BRORSEN (1995): "Maximum Likelihood Estimation of a GARCH - STABLE Model," *Journal of Applied Econometrics*, 2, 273–285.

MCDONALD, J. (1984): "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 53, 647–663.

——— (1991): "Parametric Models for Partially Adaptive Estimation with Skewed and Leptokurtic Residuals," *Economics Letters*, 37, 272–278.

NEELY, C. (1999): "Target zones and conditional volatility: the role of realignments," *Journal of Empirical Finance*, 6, 177–192.

NELSON, D. (1991): "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica*, 59, 349–370.

NEWEY, W., AND D. STEIGERWALD (1997): "Asymptotic Biais for Quasi Maximum Likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, 3, 587–599.

PAGAN, A. (1996): "The Econometrics of Financial Markets," *Journal of Empirical Finance*, 3, 15–102.

PAGAN, A., AND G. SCHWERT (1990): "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267–290.

PALM, F. (1996): "GARCH Models of Volatility," *in Maddala, G.S., Rao, C.R., Handbook of Statistics*, pp. 209–240.

PALM, F., AND P. VLAAR (1997): "Simple Diagnostics Procedures for Modelling Financial Time Series," *Allgemeine Statistisches Archiv*, 81, 85–101.

PAOLELLA, M. S. (1997): "Using Flexible GARCH Models with Asymmetric Distributions," Working paper, Institute of Statistics and Econometrics Christian Albrechts University at Kiel.

PEIRÓ, A. (1999): "Skewness in financial returns," *Journal of Banking and Finance*, 23, 847–862.

TIMMERMANN, A. (2000): "Density Forecasting in Economics and Finance," *Journal of Forecasting*, 19, 120–123.

VLAAR, P., AND F. PALM (1993): "The message in weekly exchange rates in the european monetary system: mean reversion, conditional heteroskedasticity and jumps," *Journal of Business and Economic Statistics*, 11, 351–360.

VON ROHR, P., AND I. HOESCHELE (1999): "Robust Bayesian Analysis Using Skewed Student-T Distributions," 50th Annual Meeting of EAAP, August 22-26, Zurich, Switzerland.

WANG, K.-L., C. FAWSON, C. BARRETT, AND J. MCDONALD (2000): "A Flexible Parametric GARCH Model with an Appliction to Exchange Rates," *Forthcoming in Journal of Applied Econometrics.*

WEISS, A. (1986): "Asymptotic theory for arch models: estimation and testing," *Econometric Theory*, 2, 107–131.

ZAKOIAN, J.-M. (1994): "Threshold Heteroskedasticity Models," *Journal of Economic Dynamics and Control*, 15, 931–955.