A new class of multivariate skew densities, with application to GARCH models

Luc Bauwens[1] and Sébastien Laurent[2]

March 15, 2002

Abstract

We propose a practical and flexible solution to introduce skewness in multivariate symmetrical distributions. Applying this procedure to the multivariate Student density leads to a "multivariate skew-Student" density, for which each marginal has a different asymmetry coefficient. Similarly, when applied to the product of independent univariate Student densities, it provides a "multivariate skew density with independent Student components" for which each marginal has a different asymmetry coefficient and number of degrees of freedom. Combined with a multivariate GARCH model, this new family of distributions (that generalizes the work of Fernández and Steel, 1998) is potentially useful for modelling stock returns, which are known to be conditionally heteroskedastic, fat-tailed, and often skew. In an application to the daily returns of the CAC40, NASDAQ, NIKKEI and the SMI, it is found that this density suits well the data and clearly outperforms its symmetric competitors.

Keywords: Multivariate skew density, Multivariate Student density, Multivariate GARCH models.
JEL classification: C13, C32, C52.

---

[1]CORE and Department of Economics, Université catholique de Louvain.

[2]Department of Economics, Université de Liège, CORE, Université catholique de Louvain, and Department of Quantitative Economics, Maastricht University.

Correspondence to one of the authors at: CORE, Voie du Roman Pays, 34, B-1348 Louvain-La-Neuve, Belgium. Email: bauwens@core.ucl.ac.be, or S.Laurent@ulg.ac.be

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. We would like to thank Christian Gouriéroux, Gaëlle Le Fol, Bernard Lejeune, Franz Palm, Jeroen Rombouts and Jean-Pierre Urbain for the helpful comments and suggestions. The scientific responsibility is assumed by the authors.

1 Introduction

Many time series of asset returns can be characterized as serially dependent. This is revealed by the presence of positive autocorrelation in the squared returns, and sometimes to a much smaller extent by autocorrelation in the returns. The increased importance played by risk and uncertainty considerations in modern economic theory, has necessitated the development of new econometric time series techniques that allow for the modelling of time varying means, variances and covariances. Given the apparent lack of any structural dynamic economic theory explaining the variation in the second moment, econometricians have thus extended traditional time series tools such as Autoregressive Moving Average (ARMA) models (Box and Jenkins, 1970) for the mean to essentially equivalent models for the variance. Indeed, the dynamics observed in the dispersion is clearly the dominating feature in the data. The most widespread modelling approach to capture these properties is to specify a dynamic model for the conditional mean and the conditional variance, such as an ARMA-GARCH model or one of its various extensions (see the seminal paper of Engle, 1982).

Although there is a huge literature on univariate ARCH models, much less papers are concerned with their multivariate extensions. For this reason, Geweke and Amisano (2001) argue that "while univariate models are a first step, there is an urgent need to move on to multivariate modelling of the time-varying distribution of asset returns". Indeed, financial volatilities move together over time across assets and markets. Recognizing this commonality through a multivariate modelling framework can lead to obvious gains in efficiency and to more relevant financial decision making than working with separate univariate models.

Among the most widespread multivariate GARCH models, we find the Constant Conditional Correlations model (CCC) of Bollerslev (1990), the Vech of Kraft and Engle (1982) and Bollerslev, Engle, and Wooldridge (1988), the BEKK of Engle and Kroner (1995), the Factor GARCH of Ng, Engle, and Rothschild (1992), the General Dynamic Covariance (GDC) model of Kroner and Ng (1998), the Dynamic Conditional Correlations (DCC) model of Engle (2001) and the Time-Varying Correlation (TVC) model of Tse and Tsui (1998).[1]

The estimation of these models is commonly done by maximizing a Gaussian likelihood function. Even if it is unrealistic in practice, the normality assumption may be justified by the fact that the Gaussian QML estimator is consistent provided the conditional mean and the conditional variance are specified correctly. In this respect, Jeantheau (1998) has proved the strong convergence of the QML estimator of multivariate GARCH models, extending previous results of Lee and Hansen (1994) and Lumsdaine (1996).

[1]Alternatively, Harvey, Ruiz, and Shephard (1994) propose a multivariate stochastic variance model, which has been extended in various ways. Even if this kind of model is also attractive, we only focus our attention to multivariate GARCH models.

However, another well established stylized fact of financial returns, at least when they are sampled at high frequencies, is that they exhibit fat-tails, which corresponds to a kurtosis coefficient larger than three. For instance, Hong (1988) rejected the conditional normality claiming abnormally high kurtosis in the daily New York Stock Exchange stock returns. While the high kurtosis of the returns is a well-established fact, the situation is much more obscure with regard to the symmetry of the distribution. Many authors do not observe anything special on this point, but other researchers (for instance Simkowitz and Beedles, 1980; Kon, 1984 and So, 1987) have drawn the attention to the asymmetry of the distribution. French, Schwert, and Stambaugh (1987) found also conditional skewness significantly different from 0 in the standardized residuals when an ARCH-type model was fitted to the daily SP500 returns.

As far as financial applications are concerned, and in order to gain statistical efficiency, it is of primary importance to base modelling and inference on a more suitable distribution than the multivariate normal. On the first hand, Engle and González-Rivera (1991) show in a univariate framework that the Gaussian QML estimator of a GARCH model is inefficient, with the degree of inefficiency increasing with the degree of departure from normality. On the other hand, Peiró (1999) emphasizes the relevance of modelling of higher-order features for asset pricing models[2], portfolio selection[3] and option pricing theories[4], while Giot and Laurent (2001b) and Mittnik and Paolella (2000) show that for asset returns that are are skew and fat-tailed, it is crucial to account for these features in order to obtain accurate Value-at-Risk forecasts.

The challenge to econometricians is to design multivariate distributions that are both easy to use for inference and compatible with the skewness and kurtosis properties of financial returns. Otherwise it is very likely that the estimators will not be consistent (see Newey and Steigerwald, 1997 in a univariate framework). To the best of our knowledge, asymmetric and fat-tailed k-variate distributions with support on the full Euclidian space of dimension k are uncommon.

The main contribution of this paper is to propose a practical and flexible method to introduce skewness in multivariate symmetric distributions. Applying this procedure to the multivariate Student density leads to a "multivariate skew-Student" density, in which each marginal has a specific asymmetry coefficient. Similarly, when applied to the product of independent univariate Student densities, it provides a "multivariate skew density with independent Student components" for which each marginal has a specific asymmetry coefficient and number of degrees of freedom. Combined with a multivariate GARCH model, this new family of distributions is potentially useful for modelling stock returns. In an application to the daily returns of the CAC40, NASDAQ,

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

[3]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.

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

NIKKEI and the SMI, it is found that this density suits well the data and clearly outperforms its symmetric competitors.

The paper is organized as follows. In Section 2, we briefly review the univariate skew-Student density proposed by Fernández and Steel (1998) and extended by Lambert and Laurent (2001). In Section 3, we describe the new family of multivariate skew densities, and in Section 4 we use it in a multivariate GARCH model of daily returns. Finally, we offer our conclusions and ideas for further developments in Section 5.

2 Univariate case

A series of financial returns y_t $(t = 1, \ldots, T)$, known to be typically conditionally heteroscedastic, is typically modelled as follows:

$$y_t = \mu_t + \varepsilon_t \quad (1)$$

$$\varepsilon_t = \sigma_t z_t \quad (2)$$

$$\mu_t = c(\mu|\Omega_{t-1}) \quad (3)$$

$$\sigma_t = h(\mu, \eta|\Omega_{t-1}), \quad (4)$$

where $c(.|\Omega_{t-1})$ and $h(.|\Omega_{t-1})$ are functions of Ω_{t-1} (the information set at time $t-1$), depending on unknown vectors of parameters μ and η, and z_t is an independently and identically distributed (*i.i.d.*) process, independent of Ω_{t-1}, with $E(z_t) = 0$ and $Var(z_t) = 1$. Assuming their existence, μ_t is the conditional mean of y_t and σ_t^2 is its conditional variance.

2.1 Skew-Student densities

To accommodate the excess of (unconditional) kurtosis, GARCH models have been first combined with Student distributed errors by Bollerslev (1987). Indeed, although a GARCH model generates fat-tails in the unconditional distribution, when combined with a Gaussian conditional density, it does not fully account for the excess kurtosis present in many return series. The Student density is now very popular in the literature due to its simplicity and because it often outperforms the Gaussian density. However, the main drawback of this density is that it is symmetrical while financial time series can be skewed. To create asymmetric unconditional densities, GARCH models have been extended to include a leverage effect. For instance, the Threshold ARCH (TARCH) model of Zakoian (1994) allows past negative (resp. positive) shocks to have a deeper impact on current conditional volatility than past positive (resp. negative) shocks (see among others Black, 1976; French, Schwert, and Stambaugh, 1987; Pagan and Schwert, 1990). Combined with a Student distribution for the errors, this model is in general flexible enough to mimic the observe kurtosis of many stock returns but often fails in replicating the asymmetry feature of these series (even if it can explain a small part of it). To account for both the excess skewness and kurtosis, mixtures of normal or Student densities can be used in combination with a GARCH model. In general, it has been found that these densities cannot capture all the skewness and leptokurtosis (Ball and Roma, 1993; Beine and Laurent, 1999; Jorion, 1988; Neely, 1999; Vlaar and Palm, 1993), although they seem adequate in some rare cases. Liu and Brorsen (1995) and Lambert and Laurent (2000) use the asymmetric stable density. A major drawback of the stable density is that, except when the tail parameter is equal to two (corresponding to normality), the variance does not exist, a fact that is neither usually supported empirically nor theoretically desirable. Lee and Tse (1991), Knight, Satchell, and Tran (1995), and Harvey and Siddique (1999) propose alternative skew fat-tailed densities, respectively the Gram-Charlier expansion, the double-gamma distribution, and the non-central Student. However, as pointed out by Bond (2000) in a recent survey on asymmetric conditional density functions, the estimation of these densities in a GARCH framework often proves 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 both thick tails and asymmetry. The usefulness of this density has been illustrated recently by Wang, Fawson, Barrett, and McDonald (2001) in the 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. However, goodness-of-fit tests clearly reject the EGB2 distribution for all the currencies that they consider, even if it seems to outperform the normal and the Student densities. Alternatively, Brannas and Nordman (2001) propose to use a log-generalized gamma distribution or a Pearson IV distribution with three parameters to model NYSE-returns on a daily basis.

Hansen (1994) was the first to propose a skew-Student distribution for modelling financial time series. His density nests the symmetric Student when the asymmetry parameter is equal to 0. Estimation with this density does not raise serious problems of convergence. More generally, Fernández and Steel (1998) propose a method to introduce skewness in any continuous unimodal and symmetric (about 0) univariate distribution $g(.)$, by changing its scale at each side of the mode. Applying this procedure to the Student distribution leads to another skew-Student density, that may be assumed for the innovations of an ARCH model. In order to stay in the ARCH tradition, Lambert and Laurent (2001) have modified this density in order to standardize it (i.e. to make it zero mean and unit variance). Otherwise, it would be difficult to separate the fluctuations in the mean and the variance from the fluctuations in the shape of the conditional density (see Hansen, 1994).

Following Lambert and Laurent (2001), the random variable z_t is said to be $SKST(0, 1, \xi, \upsilon)$, i.e. distributed as standardized skew-Student with parameters $\upsilon > 2$ (the number of degrees of freedom) and $\xi > 0$ (a parameter related to the skewness, see below), if its density is given by

$$f(z_t|\xi,\upsilon) = \begin{cases} \frac{2}{\xi+\frac{1}{\xi}} s\, g\left[\xi\left(sz_t+m\right)|\upsilon\right] & \text{if} \quad z_t < -m/s \\ \frac{2}{\xi+\frac{1}{\xi}} s\, g\left[\left(sz_t+m\right)/\xi|\upsilon\right] & \text{if} \quad z_t \geq -m/s, \end{cases} \tag{5}$$

where $g(.|\upsilon)$ is a symmetric (zero mean and unit variance) Student density with υ (> 2) degrees of freedom,[5], denoted $x \sim ST(0,1,\upsilon)$, and defined by

$$g(x|\upsilon) = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\pi(\upsilon-2)}\,\Gamma\left(\frac{\upsilon}{2}\right)}\left[1+\frac{x^2}{\upsilon-2}\right]^{-(\upsilon+1)/2}, \tag{6}$$

and $\Gamma(.)$ is Euler's gamma function.

In (5), the constants $m = m(\xi,\upsilon)$ and $s = \sqrt{s^2(\xi,\upsilon)}$ are respectively the mean and the standard deviation of the non-standardized skew-Student density $SKST(m,s^2,\xi,\upsilon)$ of Fernández and Steel (1998), and are defined as follows:

$$m(\xi,\upsilon) = \frac{\Gamma\left(\frac{\upsilon-1}{2}\right)\sqrt{\upsilon-2}}{\sqrt{\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)}\left(\xi-\frac{1}{\xi}\right), \tag{7}$$

and

$$s^2(\xi,\upsilon) = \left(\xi^2+\frac{1}{\xi^2}-1\right)-m^2. \tag{8}$$

It can be shown that in (5), ξ^2 is equal to the ratio of probability masses above and below the mode, which makes the use of this density very attractive because ξ^2 can be interpreted as a skewness measure. Notice also that the density $f(z_t|1/\xi,\upsilon)$ is the "mirror" of $f(z_t|\xi,\upsilon)$ with respect to the (zero) mean, i.e. $f(z_t|1/\xi,\upsilon) = f(-z_t|\xi,\upsilon)$. Therefore, as remarked by Lambert and Laurent (2000), the sign of $\log\xi$ indicates the direction of the skewness: the third moment is positive (negative), and the density is skew to the right (left), if $\log\xi > 0$ (< 0).

The main advantages of this density are its ease of implementation, that its parameters have a clear interpretation, and that it performs well on financial datasets (see Paolella 1997, Lambert and Laurent 2001, Giot and Laurent, 2001a, and Giot and Laurent 2001b). Moreover, Lambert and Laurent (2001) show how to obtain the cumulative distribution function (cdf) and the quantile function of a standardized skew density from the cdf and quantile function of the corresponding symmetric density.

Efficient estimation of the model defined by Eq. (1-4) under the assumption that $z_t \sim i.i.d.$ $SKST(0,1,\xi,\upsilon)$ is performed by maximizing the log-likelihood function $L_T(\theta) = \sum_{t=1}^{T} l_t(\theta)$ where

[5] We choose on purpose to restrict the number of degrees of freedom to be larger than 2, since we want to construct a distribution with zero mean and unit variance. Fernández and Steel (1998) and Lambert and Laurent (2000) have considered the case when υ can be smaller than two. In this case the conditional distribution is parameterized in terms of its mode and dispersion.

$\theta = (\mu', \eta', \xi, \upsilon)'$ denotes the vector of parameters, and

$$
\begin{aligned}
l_t(\theta) \quad = \quad & \log\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + \log\Gamma\left(\frac{\upsilon+1}{2}\right) - 0.5\log\left[\pi(\upsilon-2)\right] - \log\Gamma\left(\frac{\upsilon}{2}\right) \\
& + \log\frac{s}{\sigma_t} - 0.5\,(1+\upsilon)\log\left[1+\frac{(sz_t+m)^2\xi^{-2I_t}}{\upsilon-2}\right] \qquad (9)
\end{aligned}
$$

with $z_t = (y_t - \mu_t)/\sigma_t$, and

$$
I_t \quad = \quad \begin{cases} 1 \text{ if } z_t \geq -\frac{m}{s} \\ -1 \text{ if } z_t < -\frac{m}{s}. \end{cases}
$$

In Eq. (9), μ_t, σ_t, m and s are functions of the parameters defined by Eq. (3), (4), (7), and (8), respectively. The estimation of a highly non-linear model like Eq. (1-4) relies on numerical techniques to approximate the derivatives of the likelihood function with respect to the parameter vector. To avoid numerical inefficiencies and highly speed-up estimation, Laurent (2001) provides numerically reliable analytical expressions for the score vector of Eq. (9).

Recently, Jones and Faddy (2000) have designed another skew-t distribution. Like the $SKST$ $(0, 1, \xi, \upsilon)$ density, it has two parameters (assuming zero location and unit scale parameters), say a and b. If $a = b$, the distribution is the usual symmetrical Student one, as defined above by Eq. (6), with $\upsilon = 2b$ (assuming $b > 1$). If $a - b > 0$ (< 0), the density is skew to the right (left): hence $a - b$ is a skewness parameter that, however, does not have an interpretation as clearcut as ξ^2 (the ratio of probability masses above and below the mode). A property of this skew-t density is that its long tail is thicker than its short tail (if $a > b$, the left tail behaves like $x^{-(2a+1)}$ at minus infinity, the long tail like $x^{-(2b+1)}$ at plus infinity). On the contrary the $SKST$ density has the same thickness of tails at plus and minus infinity, where it behaves like $x^{-(\upsilon+1)}$. While it may be of interest to have a different tail behavior at the two extremities, for financial applications it is not obvious that the thicker tail should be necessarily the long one. Jones and Faddy (2000) also provide the moments and the cdf of their skew-t density. Which of the two densities is to be preferred for modelling skew returns in a univariate GARCH model is an open question that is beyond the objective of this paper.[6]

2.2 Empirical illustration

In this illustration, we consider four stock market indexes: the French CAC40, US NASDAQ, Japanese NIKKEI and Swiss SMI from January 1991 to December 1998 (1816 daily observations; source: Datastream). The daily return is defined as $y_t = 100 \times (\log p_t - \log p_{t-1})$ where p_t is the stock index value of day t.

[6] In a multivariate GARCH model, the issue is clearly settled in favor of the multivariate generalization of the $SKST$ density that is proposed in Section 3.

We use the model defined by Eq. (1-4) with the following conditional mean and variance equations:

$$\mu_t = \mu + \phi(y_{t-1} - \mu) \quad (10)$$

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2, \quad (11)$$

where μ, ϕ, ω, β, and α are parameters to be estimated. An autoregressive (AR) model of order one is chosen for the conditional mean to allow for possible autocorrelation in the daily returns, while a GARCH(1,1) specification -see Bollerslev (1986)- is chosen for the conditional variance to account for volatility clustering in a simple way. More sophisticated ARCH models could easily be used, but this is not the focus of the paper.

To account for possible skewness and fat tails, we estimated the AR(1)-GARCH(1,1) model assuming a skew-Student density for the innovations. In order to assess the practical relevance of this density, we compare the estimation results with two other assumptions regarding the innovations density: the normal (obtained when υ tends to infinity and $\xi = 1$), and the symmetric Student (obtained by setting $\xi = 1$). Results concerning the CAC40 and the NASDAQ are gathered in Table 1 and those concerning the NIKKEI and the SMI are reported in Table 2. Several comments are in order:

- The AR(1)-GARCH(1,1) specification seems to be adequate for capturing the dynamics of the four series. Indeed, looking at the Box-Pierce statistics with 20 lags on the standardized residuals (Q_{20}) and the squared standardized residuals (Q_{20}^2), one cannot reject the assumption of lack of autocorrelation in the innovation process and its square (except perhaps for the CAC40 where the standardized residuals are still slightly serially correlated);

- The estimated number of degrees of freedom υ is about 6 for the NASDAQ, NIKKEI and SMI and about 9 for the CAC40, which indicates that the returns are fat-tailed. Moreover, the differences between the likelihood of the normal and the Student densities are so big that there is little doubt that the latter should be preferred to the former (despite the fact that the LR test is presumably non-standard);

- The estimated skewness parameter $\log \xi$ is negative and different from 0 at conventional levels of significance for the NASDAQ and the SMI, while it is not different from 0 for the CAC40 and the NIKKEI. The distribution of returns of the NASDAQ and the SMI is therefore characterized by negative skewness, while the other series appear to be symmetrically distributed over the period under consideration. Notice however that since the skew-Student density has the symmetric Student density as a limiting case, it is also adequate for the CAC40 and the NIKKEI (resulting perhaps in a small loss of efficiency);

Table 1: ML Estimation Results of AR-GARCH models for the CAC40 and the NASDAQ

	Normal	Student	skew-Student
	CAC40; NASDAQ	CAC40 ; NASDAQ	CAC40 ; NASDAQ
μ	0.051 ; 0.111	0.057 ; 0.137	0.053 ; 0.099
	[0.029] ; [0.027]	[0.027] ; [0.023]	[0.027] ; [0.023]
ϕ	0.052 ; 0.177	0.044 ; 0.171	0.044 ; 0.152
	[0.026] ; [0.026]	[0.023] ; [0.024]	[0.023] ; [0.024]
ω	0.094 ; 0.092	0.043 ; 0.055	0.042 ; 0.053
	[0.074] ; [0.036]	[0.028] ; [0.026]	[0.027] ; [0.025]
β	0.860 ; 0.766	0.915 ; 0.827	0.915 ; 0.826
	[0.076] ; [0.063]	[0.037] ; [0.052]	[0.037] ; [0.052]
α	0.078 ; 0.153	0.056 ; 0.124	0.056 ; 0.128
	[0.034] ; [0.043]	[0.022] ; [0.035]	[0.022] ; [0.035]
$\log \xi$	0 ; 0	0 ; 0	-0.014 ; -0.158
			[0.031] ; [0.034]
υ	∞ ; ∞	8.657 ; 5.685	8.714 ; 5.938
		[1.918] ; [0.753]	[1.933] ; [0.817]
Q_{20}	27.511 ; 17.830	30.652 ; 17.925	27.289 ; 19.526
Q^2_{20}	8.682 ; 7.815	11.302 ; 10.720	10.990 ; 10.983
P_{20}	30.608 ; 62.344	10.531 ; 37.815	17.782 ; 12.338
	(0.044) ; (0.000)	(0.938) ; (0.006)	(0.537) ; (0.870)
SIC	3.229 ; 2.802	3.197 ; 2.728	3.202 ; 2.720
Log-Lik	-2911.6 ; -2524.6	-2879.9 ; -2453.7	-2879.8 ; -2442.9

Each column reports the ML estimates of the model defined by Eq. (1)-(2)-(10)-(11), with robust standard errors underneath in brackets. The column headed "Normal" corresponds to $z_t \sim N(0,1)$, "Student" to $z_t \sim ST(0,1,\upsilon)$ as in (6), "Skew-Student" to $z_t \sim SKST(0,1,\xi,\upsilon)$ as in (5), and in all cases z_t is an *i.i.d.* process. Q_{20} is the Box-Pierce statistic of order 20 on the standardized residuals, Q^2_{20} is the same for their squares, P_{20} is the Pearson goodness-of-fit statistic (using 20 cells) with the associated p-value underneath in parentheses (see footnote 7). *SIC* is the Schwarz information criterion (divided by the sample size), and *Log-Lik* is the log-likelihood value at the maximum. The sample size is equal to 1816.

Table 2: ML Estimation Results of AR-GARCH models for the NIKKEI and the SMI

	Normal	Student	skew-Student
	NIKKEI; SMI	NIKKEI ; SMI	NIKKEI ; SMI
μ	-0.004 ; 0.110	-0.021 ; 0.123	-0.031 ; 0.102
	[0.030] ; [0.023]	[0.026] ; [0.020]	[0.028] ; [0.021]
ϕ	-0.014 ; 0.070	-0.026 ; 0.039	-0.028 ; 0.028
	[0.025] ; [0.026]	[0.023] ; [0.025]	[0.023] ; [0.026]
ω	0.061 ; 0.147	0.040 ; 0.063	0.039 ; 0.058
	[0.036] ; [0.066]	[0.016] ; [0.024]	[0.015] ; [0.022]
β	0.894 ; 0.731	0.902 ; 0.810	0.903 ; 0.817
	[0.035] ; [0.058]	[0.018] ; [0.049]	[0.018] ; [0.047]
α	0.080 ; 0.141	0.082 ; 0.136	0.082 ; 0.134
	[0.024] ; [0.028]	[0.016] ; [0.035]	[0.016] ; [0.033]
$\log \xi$	0 ; 0	0 ; 0	-0.035 ; -0.101
			[0.034] ; [0.034]
υ	∞ ; ∞	5.950 ; 6.273	5.895 ; 6.364
		[0.845] ; [1.085]	[0.825] ; [1.096]
Q_{20}	14.198 ; 12.271	14.29 ; 11.212	14.372 ; 11.836
Q_{20}^2	5.876 ; 1.377	6.617 ; 2.433	6.662 ; 2.454
P_{20}	47.225 ; 51.567	13.837 ; 26.818	15.843 ; 16.570
	(0.000) ; (0.000)	(0.793) ; (0.108)	(0.667) ; (0.618)
SIC	3.562 ; 2.894	3.494 ; 2.788	3.498 ; 2.787
Log-Lik	-3214.1 ; -2607.9	-3148.9 ; -2507.6	-3148.4 ; -2503.2

Note: see Table 1.

- Using the Schwarz information criterion to discriminate between the three densities, one should select the skew-Student for the NASDAQ and the SMI and the Student for the others;

- Finally and more importantly, the relevance of the skew-Student distribution is also confirmed by the Pearson goodness-of-fit statistics.[7] This test is in fact equivalent to an in-sample density forecast test, as proposed recently by Diebold, Gunther, and Tay (1998). While the normal and the Student distributions are clearly rejected for the NASDAQ (the p-values being very small), the skew-Student density seems to be supported (p-value = 0.87). Similarly, one can see that the skew-Student density is appropriate for modelling the SMI. Unsurprisingly, the normal density is rejected for the CAC40 and the NIKKEI while the Student and the skew-Student are not rejected at conventional levels of significance.

This example illustrates the potential usefulness of the skew-Student distribution in a univariate volatility model. The skewness parameters of the four series are different, and although the numbers of degrees of freedom are almost identical for the NASDAQ, NIKKEI and SMI, the innovations of the CAC40 seem to have less kurtosis. For modelling jointly the four series, it could therefore be useful to have a multivariate density that would allow for different skewness and perhaps different tail properties on each series.

3 Multivariate case

Consider a time series vector y_t, with k elements, $y_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$. A multivariate dynamic regression model with time-varying means, variances and covariances for the components of y_t generally takes the form:

$$y_t = \mu_t + \Sigma_t^{1/2} z_t \quad (12)$$

$$\mu_t = C(\mu|\Omega_{t-1}) \quad (13)$$

$$\Sigma_t = \Sigma(\mu, \eta|\Omega_{t-1}) \quad (14)$$

where $z_t \in \Re^k$ is an *i.i.d.* random vector independent of Ω_{t-1} with zero mean and identity variance matrix and $C(.|\Omega_{t-1})$ and $\Sigma(.|\Omega_{t-1})$ are functions of Ω_{t-1}. It follows that $E(y_t|\mu, \Omega_{t-1}) = \mu_t$ and $Var(y_t|\mu, \eta, \Omega_{t-1}) = \Sigma_t^{1/2}(\Sigma_t^{1/2})' = \Sigma_t$, i.e. μ_t is the conditional mean vector (of dimension $k \times 1$) and Σ_t the conditional variance matrix (of dimension $k \times k$).

[7] For a given number of cells denoted g, the Pearson goodness-of-fit statistics is $P(g) = \sum_{i=1}^{g} \frac{(n_i - En_i)^2}{En_i}$, where n_i is the number of observations in cell i and En_i is the expected number of observations (based on the ML estimates). For *i.i.d.* observations, Palm and Vlaar (1997) show that under the null of a correct distribution the asymptotic distribution of $P(g)$ is bounded between a $\chi^2(g-1)$ and a $\chi^2(g-k-1)$ where k is the number of estimated parameters. Since our conclusions hold for both critical values, we report the significance levels relative to the first one.

Under the assumption of correct specification of the conditional mean and conditional variance matrix, the efficient estimation of the above model is obtained by the ML method, assuming z_t to be *i.i.d.* with a correctly specified distribution that may depend upon a few unknown parameters. When the distribution of z_t is assumed to be the standard normal, the ML estimator obtained from the corresponding likelihood function is consistent even if the normality assumption is incorrect (see Bollerslev and Wooldridge, 1992). This well-known Gaussian QML procedure has the advantage of robustness with respect to the distributional assumption of the model. The QML estimator relying on a normal distribution is, however, inefficient, with the degree of inefficiency increasing with the degree of departure from normality (see Engle and González-Rivera, 1991 in a univariate framework).

3.1 Multivariate symmetrical densities

Like in the univariate case, a natural candidate, apart from the normal density, is the multivariate Student density with at least two degrees of freedom υ (in order to ensure the existence of second moments). It may be defined as:

$$g(z_t|\upsilon) = \frac{\Gamma\left(\frac{\upsilon+k}{2}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)\left[\pi(\upsilon-2)\right]^{\frac{k}{2}}}\left[1+\frac{z_t'z_t}{\upsilon-2}\right]^{-\frac{k+\upsilon}{2}}, \tag{15}$$

where $\Gamma(.)$ is the Gamma function. This density is denoted $ST(0, I_k, \upsilon)$.

The density function of y_t, easily derived from the density of z_t by using the transformation in Eq. (12), is given by:

$$f(y_t \mid \mu, \eta, \upsilon, \Omega_{t-1}) = \frac{\Gamma(\frac{\upsilon+k}{2})}{\Gamma(\frac{\upsilon}{2})[\pi(\upsilon-2)]^{\frac{k}{2}}} \mid \Sigma_t \mid^{-\frac{1}{2}} \left[1+\frac{(y_t-\mu_t)'\Sigma_t^{-1}(y_t-\mu_t)}{\upsilon-2}\right]^{-\frac{k+\upsilon}{2}}. \tag{16}$$

While genuine ML methods provide more efficient estimators than the Gaussian QML when the assumption made on the innovation process holds, it has the main disadvantage that unlike the Gaussian QML, it does not provide a consistent estimator when this assumption does not hold.

Consequently, recalling the findings of the previous section, there is a need for skew densities in the multivariate case. Such densities can be defined by introducing skewness in symmetric densities by means of new parameters, such that the symmetric density results as a particular case. In Section 3.2, we propose a simple and intuitive method to introduce skewness into a multivariate "symmetric" unimodal density (with zero mean and unit variance). Before that, we define the notion of symmetry that we rely on.

In the univariate case, the symmetry property corresponds to $g(x) = g(-x)$ assuming $g(x)$ is a unimodal probability density function and $E(x) = 0$. In the multivariate case, we use the following definition of symmetry of a standardized density $g(x)$:

Definition 1 (M-symmetry): *The unimodal density* $g(x)$ *defined on* $\Re^k$, *such that* $E(x) = 0$, *and* $Var(x) = I_k$, *is symmetrical if and only if for any* x, $g(x) = g(Qx)$, *for all diagonal matrices* Q *whose diagonal elements are equal to +1 or to -1. If* x *is a random vector with a density satisfying this definition, we write*

$$x \sim M\text{-}Sym(0, I_k, g). \tag{17}$$

In the bivariate case, this definition means that

$$g(x_1, x_2) = g(-x_1, x_2) = g(x_1, -x_2) = g(-x_1, -x_2), \tag{18}$$

and in the trivariate case

$$\begin{aligned} g(x_1, x_2, x_3) &= g(-x_1, x_2, x_3) = g(-x_1, -x_2, x_3) = g(-x_1, -x_2, -x_3) \\ &= g(x_1, -x_2, x_3) = g(x_1, -x_2, -x_3) = g(x_1, x_2, -x_3) = g(-x_1, x_2, -x_3). \end{aligned} \tag{19}$$

Spherically symmetric (SS) densities, defined by the property that the density depends on x through $x'x$ only, i.e.

$$g(x) \propto k(x'x), \tag{20}$$

for an appropriate integrable positive function $k(.)$, are M-symmetric. The most well known examples of SS-densities[8] are the standard normal density and the standard Student density $ST(0, I_k, \upsilon)$. However, there exist other distributions that have the desired property while not being spherically symmetric. A large class is defined by

$$g(x) = \prod_{i=1}^{k} g_i(x_i), \tag{21}$$

where $g_i(.)$, $\forall i$, is a univariate symmetric density (unimodal, with mean 0 and unit variance). If $g_i(.)$ ($\forall i$) is standard normal, there is no difference between (21) and (20) with $g(.) = N(0, I_k)$. Nevertheless, if $g_i(.) = ST(0, 1, \upsilon)$ ($\forall i$) and $g(.) = ST(0, I_k, \upsilon)$, there is a difference between (21) and (20) since the elements of (20) are not mutually independent whereas those of (21) are. Notice however that both multivariate densities have the same univariate marginal densities.

3.2 Multivariate skew densities

3.2.1 Literature review

Jones (2000) has generalized the univariate skew-*t* density of Jones and Faddy (2000), briefly described at the end of Section 2.1, to the multivariate case. His multivariate skew-*t* density is such that each marginal is a univariate skew-*t* as defined by Jones and Faddy (2000). However,

[8] Johnson (1987), chapter 6, provides graphical illustrations of several bivariate SS-densities.

his multivariate density has necessarily positive covariances, and is therefore useless for a model such as defined by Eq. (12), where it is essential that $\text{Var}(z_t) = I_k$.

Mauleón and Perote (1999) use the bivariate Edgeworth-Sargan density for z_t in a bivariate constant correlation GARCH model, where each conditional variance is specified like in a univariate GARCH(1,1) model. The Edgeworth-Sargan density has as leading term a bivariate standard normal density, to which are added terms that create the non-normality (these terms involve Hermite polynomials in each of the marginal densities of the leading term). However, they use only a symmetrical version of their density, because they choose not to include odd-order terms in the expansion (such terms would induce asymmetry). Actually they include four even-order terms in the expansion on each element of z_t, under the motivation that these terms induce fatter tails than for the leading normal density. This appears to us to be a costly way, in term of the number of parameters, to introduce the possibility of having fat tails. A multivariate Student density requires just one extra parameter, with the drawback of constraining the same thickness of tails on each element of z_t, but this is easily extended by taking a product of independent Student densities in the spirit of Eq. (21) (the last solution would require 2 parameters instead of 8 in the bivariate case). Moreover, Mauleón and Perote (1999) report some difficulties in obtaining the convergence of the numerical maximization of the log-likelihood function based on their Edgeworth-Sargan density. At least for the time being, this does not seem to be a fruitful approach.

Another recent paper, by Branco and Dey (2000), introduces a general class of multivariate skew-elliptical distributions, and is therefore related to our work.[9] Their work generalizes to the full class of elliptically contoured (EC) densities earlier results by Azzalini and Capitanio (1996), who have defined a multivariate skew-normal distribution. Any EC-density is obtained by linear transformation of a SS-density: if z (of dimension $k \times 1$) is SS-distributed with density $g(z)$, μ is a vector of location parameters, and Ω is a $k \times k$ positive-definite symmetric scale matrix, then $x = \mu + \Omega^{1/2} z$ is elliptically contoured, which is denoted $x \sim EC(\mu, \Omega; g)$ (where g denotes the density of x). To obtain a skew version of an EC-density, Branco and Dey (2000) start from $x^* \sim EC(\mu^*, \Omega^*; g^*)$, where $x^* = (x_0, x')'$ is a vector of $k+1$ elements. They partition μ^* and Ω^* as x^*, i.e.

$$\mu^* = \begin{pmatrix} 0 \\ \mu \end{pmatrix}, \quad \Omega^* = \begin{pmatrix} 1 & \delta' \\ \delta & \Omega \end{pmatrix}, \tag{22}$$

where μ and δ are $k \times 1$ vectors, and Ω is a $k \times k$ matrix. Then they define the distribution of x conditional on $x_0 > 0$ to be the skew-elliptical distribution based on the density $g^*(.)$, with parameters μ (location, or mean if it exists), Ω (scale matrix, or variance matrix if it exists), and

[9] Sahu, Dey, and Branco (2001) use the skew-elliptical density in Bayesian regression analysis, by assuming the error terms to have this kind of distribution, rather than a symmetrical distribution.

δ (a vector of skewness parameters), i.e. $x \sim SKE(\mu, \Omega, \delta; g)$. They show that the density of this random vector (call it z) is given by

$$f(z) = 2g(z)\, G_*[\lambda'(z - \mu)], \tag{23}$$

where $g(.)$ is the marginal density of x derived from the density of x^* (by properties of EC-distributions, it has the same functional form as g^*), $G_*(.)$ is the (univariate) cdf of an $EC(0, 1; g_*)$, with g_* appropriately defined (essentially from the conditional density of x_0 given x), and

$$\lambda = \frac{\delta'\Omega^{-1}}{(1 - \delta'\Omega^{-1}\delta)^{1/2}}. \tag{24}$$

It is therefore clear that the parameters δ (a set of covariances) create the skewness. If they are all equal to 0, $G_*[\lambda'(z - \mu)] = G_*(0) = 1/2$, by symmetry of $EC(0, 1; g_*)$, and the density (23) becomes symmetrical. However, there is a constraint linking these skewness parameters, namely that $\delta'\Omega^{-1}\delta$ must be smaller than unity, see Eq. (24). This is a constraint that is likely to complicate inference. In the context of GARCH models with standardized innovations, Ω is an identity matrix (and $\mu = 0$), hence δ is a vector of correlation coefficients, and the constraint is that the sum of squared correlations is less than one. To what extent this constraint limits the degree of skewness is not known.[10] Another drawback of this approach is that if one wants to introduce some dynamics in the skewness parameters, the constraint would be different for each observation, which would complicate the estimation dramatically. We conclude on this class of skew densities by saying that it seems an interesting, though seemingly more difficult to implement, alternative to the class of skew densities that we propose below, and that more work is needed to compare the different classes of skew densities

To accommodate both the skewness and kurtosis of six weekly rates of the European Monetary System (EMS) expressed in terms of the Deutsche mark, Vlaar and Palm (1993) propose to use a (Bernoulli) mixture of two multivariate normal densities (coupled with an MA(1)-GARCH(1,1) model with constant correlations, see Bollerslev, 1990). The size and the variance of the jumps are allowed to differ across currencies. However, to render the estimation feasible, they assume identical jump probability for all the series arguing that a stochastic shock leading to a jump is likely to simultaneously affect all of the currencies in the system. Even if this assumption is realistic for currencies that belong to the EMS, it is unrealistic for stock returns, for instance. Moreover, even if this density is expressed in such a way that $E(z_t) = 0$, the covariance matrix of z_t is not an identity matrix in their specification. Another drawback of this density is that the parameters that govern the skewness and kurtosis have not a clear interpretation because for each margin the jump probability, the size and the variance of the jumps explain at the same time the

[10] If $k = 1$, the constraint is not limitative.

variance, skewness and kurtosis in an highly non-linear way (see Vlaar and Palm, 1993 for more details). To conclude about this density, it suffers from a problem of non-identification of several parameters when the mixture is not relevant (for instance when the jump probability equals 0 or 1), which makes the testing procedures non-standard.

Finally, we cannot refrain from mentioning a class of multivariate densities that could be of interest: the so-called poly-t densities that contain the multivariate Student density as a particular case. Poly-t densities arise as posterior densities in Bayesian inference, see Drèze (1978), and can be heavily skew, have fat tails and even be multimodal. However, more work is required to discover how the skewness of these densities depends on their parameters (see Richard and Tompa, 1980 for results on moments of poly-t densities).

3.2.2 New skew densities

We generalize to the multivariate case the method proposed by Fernández and Steel (1998) to construct a skew density from a symmetrical one. Let us consider the k-dimensional random vector z^* defined by:

$$z^* = \lambda(\tau)\,|x|, \tag{25}$$

where

$$|x| = (|x_1|, \ldots, |x_k|)', \tag{26}$$

and

$$x \sim M\text{-}Sym(0, I_k, g). \tag{27}$$

Moreover, $\lambda(\tau)$ is a $k \times k$ diagonal matrix defined by:

$$\lambda(\tau) = \tau\xi - (I_k - \tau)\,\xi^{-1}, \tag{28}$$

where

$$\begin{aligned}
\tau &= diag(\tau_1, \ldots, \tau_k), \text{ with } \tau_i \in \{0, 1\}, \\
\tau_i &\sim Ber\left(\frac{\xi_i^2}{1+\xi_i^2}\right), \text{ with } \xi_i > 0, \\
\xi &= (\xi_1, \ldots, \xi_k). \\
\Xi &= diag(\xi).
\end{aligned}$$

$Ber\left(\frac{\xi_i^2}{1+\xi_i^2}\right)$ denotes a Bernoulli distribution with probability of success $\frac{\xi_i^2}{1+\xi_i^2}$. It is also assumed that the elements of τ are mutually independent.

For ease of exposition, we give the details of the derivation of the density of z^* in the bivariate case, before giving the general formula.

Bivariate case

We can write the density of z^* as a discrete mixture with respect to the distribution of τ:

$$\begin{aligned}
f(z^*|\xi) &= \Pr(\tau_1=1,\tau_2=1)f(z^*|\xi,\tau_1=1,\tau_2=1) \\
&+ \Pr(\tau_1=1,\tau_2=0)f(z^*|\xi,\tau_1=1,\tau_2=0) \\
&+ \Pr(\tau_1=0,\tau_2=1)f(z^*|\xi,\tau_1=0,\tau_2=1) \\
&+ \Pr(\tau_1=0,\tau_2=0)f(z^*|\xi,\tau_1=0,\tau_2=1). \qquad (29)
\end{aligned}$$

By dividing the range of all possible values of $z^* \in \Re^2$ into the four quadrants, we can write the right hand side of Eq. (29) in terms of the original M-symmetric density $g(.)$:

$$\begin{aligned}
f(z^*|\xi) &= 2^2\Pr(\tau_1=1,\tau_2=1)\,|\lambda(1,1)|^{-1}\,g[\lambda(1,1)^{-1}z^*]\,I_{(z_1^*\geq 0;z_2^*\geq 0)} \\
&+ 2^2\Pr(\tau_1=1,\tau_2=0)\,|\lambda(1,0)|^{-1}g[\lambda(1,0)^{-1}z^*]\,I_{(z_1^*\geq 0;z_2^*<0)} \\
&+ 2^2\Pr(\tau_1=0,\tau_2=1)\,|\lambda(0,1)|^{-1}g[\lambda(0,1)^{-1}z^*)\,I_{(z_1^*<0;z_2^*\geq 0)} \\
&+ 2^2\Pr(\tau_1=0,\tau_2=0)\,|\lambda(0,0)|^{-1}g[\lambda(0,0)^{-1}z^*)\,I_{(z_1^*<0;z_2^*<0)}, \qquad (30)
\end{aligned}$$

where e.g. $\lambda(1,1)$ stands for $\lambda(\tau_1=1,\tau_2=1)$ and for instance $I_{(z_1^*\geq 0;z_2^*\geq 0)}=1$ when $z_1^*\geq 0$ and $z_2^*\geq 0$, 0 otherwise. After some algebraic manipulations of (30) using (28) and the assumption of independence of τ_1 and τ_2, we obtain:

$$\begin{aligned}
f(z^*|\xi) &= 2^2\frac{\xi_1}{1+\xi_1^2}\frac{\xi_2}{1+\xi_2^2}\Big\{g[\lambda(1,1)^{-1}z^*]\,I_{(z_1^*\geq 0;z_2^*\geq 0)} \\
&+ g[\lambda(1,0)^{-1}z^*]\,I_{(z_1^*\geq 0;z_2^*<0)}+g[\lambda(0,1)^{-1}z^*]\,I_{(z_1^*<0;z_2^*\geq 0)} \\
&+ g[\lambda(0,0)^{-1}z^*]\,I_{(z_1^*<0;z_2^*<0)}\Big\}, \qquad (31)
\end{aligned}$$

and finally,

$$f(z^*|\xi)=2^2\frac{\xi_1}{1+\xi_1^2}\frac{\xi_2}{1+\xi_2^2}\,g(\kappa^*), \qquad (32)$$

where

$$\begin{aligned}
\kappa^* &= (\kappa_1^*,\kappa_2^*)' \qquad (33)\\
\kappa_i &= z_i^*\xi_i^{-I_i}\ (i=1,2) \qquad (34)\\
I_i &= \begin{cases} 1 \text{ if } z_i^*\geq 0 \\ -1 \text{ if } z_i^*<0. \end{cases}
\end{aligned}$$

Applying this procedure to the bivariate Student distribution given by Eq. (15) with $k=2$ and x instead of z_t, i.e. $x\sim ST(0,I_2,\upsilon)$, yields a "bivariate skew-Student" density, in which both marginals have different asymmetry parameters, ξ_1 and ξ_2.

Multivariate case

It is straightforward to show that for any dimension k,

$$f(z^*|\xi) = 2^k \left(\prod_{i=1}^{k} \frac{\xi_i}{1+\xi_i^2} \right) g(\kappa^*), \tag{35}$$

where κ^* is given in Eq. (33)-(34) for the bivariate case and is easily extended to the multivariate case. Recall that for each margin z_i^*, ξ_i has a clear interpretation since ξ_i^2 is equal to the ratio of probability masses above and below the mode. Remark also that when $k = 1$, one recovers the family of skew densities proposed by Fernández and Steel (1998).

Moments

A convenient property of this new family of skew densities is that the marginal moments are obtained by the same method and actually correspond to the same formulas as in the univariate case. The r-th order moment of $f(z^*|\xi)$ exists if the r-th order moment of $g(.)$ exists. In particular,

$$\mathrm{E}\left(z_i^{*r}|\xi\right) = M_{i,r} \frac{\xi_i^{r+1} + \frac{(-1)^r}{\xi_i^{r+1}}}{\xi_i + \frac{1}{\xi_i}} \tag{36}$$

where

$$M_{i,r} = \int_0^\infty 2u^r g_i(u) du, \tag{37}$$

and $g_i(.)$ is the marginal of x_i extracted from $g(x)$, while $M_{i,r}$ is the r-th order moment of $g_i(.)$ truncated to the positive real values. Provided that these quantities are finite, we obtain:

$$\mathrm{E}(z_i^*|\xi_i) = M_{i,1}\left(\xi_i - \frac{1}{\xi_i}\right) = m_i \tag{38}$$

$$\mathrm{Var}(z_i^*|\xi_i) = \left(M_{i,2} - M_{i,1}^2\right)\left(\xi_i^2 + \frac{1}{\xi_i^2}\right) + 2M_{i,1}^2 - M_{i,2} = s_i^2 \tag{39}$$

$$\mathrm{Sk}(z_i^*|\xi_i) = \frac{\left(\xi_i - \frac{1}{\xi_i}\right)\left(M_{i,3} + 2M_{i,1}^3 - 3M_{i,1}M_{i,2}\right)\left(\xi_i^2 + \frac{1}{\xi_i^2}\right) + 3M_{i,1}M_{i,2} - 4M_{i,1}^3}{\mathrm{Var}\left(z_i^*|\xi_i\right)^{\frac{3}{2}}} \tag{40}$$

$$\mathrm{Ku}(z_i^*|\xi_i) = \frac{\mathrm{E}\left(z_i^{*4}|\xi_i\right) - 4\mathrm{E}\left(z_i^*|\xi_i\right)\mathrm{E}\left(z_i^{*3}|\xi_i\right) + 6\mathrm{E}\left(z_i^{*2}|\xi_i\right)\mathrm{E}\left(z_i^*|\xi_i\right)^2 - 3\mathrm{E}\left(z_i^*|\xi_i\right)^4}{\mathrm{Var}\left(z_i^*|\xi_i\right)^2} \tag{41}$$

where Sk(.) and Ku(.) denote the skewness and kurtosis coefficients, respectively.[11]

Finally, it is obvious that the elements of z^* are uncorrelated (since those of x are uncorrelated by assumption), so that it is easy to transform z^* so as to have any specified covariance matrix.

Standardized skew densities

The main drawback of the skew density defined by Eq. (35) is that it is not centered on 0 and the covariance matrix is a function of ξ (and of υ if $g(.)$ is a multivariate Student density). As in the univariate case, one can solve this problem by standardizing z^*.

[11] An explicit expression of the kurtosis in terms of the $M_{i,r}$ and ξ_i is too cumbersome.

Let us consider the following random vector:

$$z = (z^* - m)./s \tag{42}$$

where $m = (m_1, \ldots, m_k)$ and $s = (s_1, \ldots, s_k)$ are the vectors of unconditional means and standard deviations of z^*, and ./ means element by element division. The above transformation amounts to standardize each component of z^*.

Note that if $g(.)$ is the multivariate Student density as described in Eq. (15), its marginal $g_i(.|\upsilon)$ is a univariate standardized Student and following Lambert and Laurent (2001),

$$m_i = \frac{\Gamma\left(\frac{\upsilon-1}{2}\right)\sqrt{\upsilon-2}}{\sqrt{\pi}\Gamma\left(\frac{\upsilon}{2}\right)}\left(\xi_i - \frac{1}{\xi_i}\right) \tag{43}$$

and

$$s_i^2 = \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1\right) - m_i^2. \tag{44}$$

Definition 2 *If (i) z is defined by Eq. (42-44), and (ii)* z^* *has a density given by Eq. (35), where* $g(x)$ *is the Student density given by Eq. (15), then z is said to be distributed as (multivariate) standardized skew-Student with asymmetry parameters* $\xi = (\xi_1, \ldots, \xi_k)$, *and number of degrees of freedom* $\upsilon(> 2)$. *This is denoted* $z \sim SKST(0, I_k, \xi, \upsilon)$. *The density of z is given by:*

$$f(z|\xi,\upsilon) = \left(\frac{2}{\sqrt{\pi}}\right)^k \left(\prod_{i=1}^k \frac{\xi_i s_i}{1+\xi_i^2}\right) \frac{\Gamma(\frac{\upsilon+k}{2})}{\Gamma(\frac{\upsilon}{2})(\upsilon-2)^{\frac{k}{2}}} \left(1 + \frac{\kappa'\kappa}{\upsilon-2}\right)^{-\frac{k+\upsilon}{2}}. \tag{45}$$

where

$$\kappa = (\kappa_1, \ldots, \kappa_k)' \tag{46}$$

$$\kappa_i = (s_i z_i + m_i)\,\xi_i^{-I_i} \tag{47}$$

$$I_i = \begin{cases} 1 \text{ if } z_i \geq -\frac{m_i}{s_i} \\ -1 \text{ if } z_i < -\frac{m_i}{s_i}. \end{cases}$$

By construction, $\mathrm{E}(z) = 0$ and $\mathrm{Var}(z) = I_k$. If $\xi = I_k$, the $SKST(0, I_k, \xi, \upsilon)$ density becomes the $ST(0, I_k, \upsilon)$ one, i.e. the symmetric Student density.

Assuming that y_t is specified as in Eq. (12) and $z_t \sim SKST(0, I_k, \xi, \upsilon)$, the density of y_t is straightforwardly obtained (see how Eq. (16) is obtained from Eq. (15)).

To illustrate, Figure 1 shows a graph of the $SKST(0, I_2, \xi, 6)$ density with $\xi_1 = 1$, $\xi_2 = 1.3$, and the Panel A of Figure 2 shows its contours.

The first graph is oriented to show the asymmetry to the right along the z_2-axis, while the density is symmetric in the direction of the first coordinate (z_1). The contours show more clearly the skewness properties of the density in the direction of z_2, and its symmetry in the direction of z_1. One also clearly sees that the mode is not centered in zero (unlike in the non-standardized version).

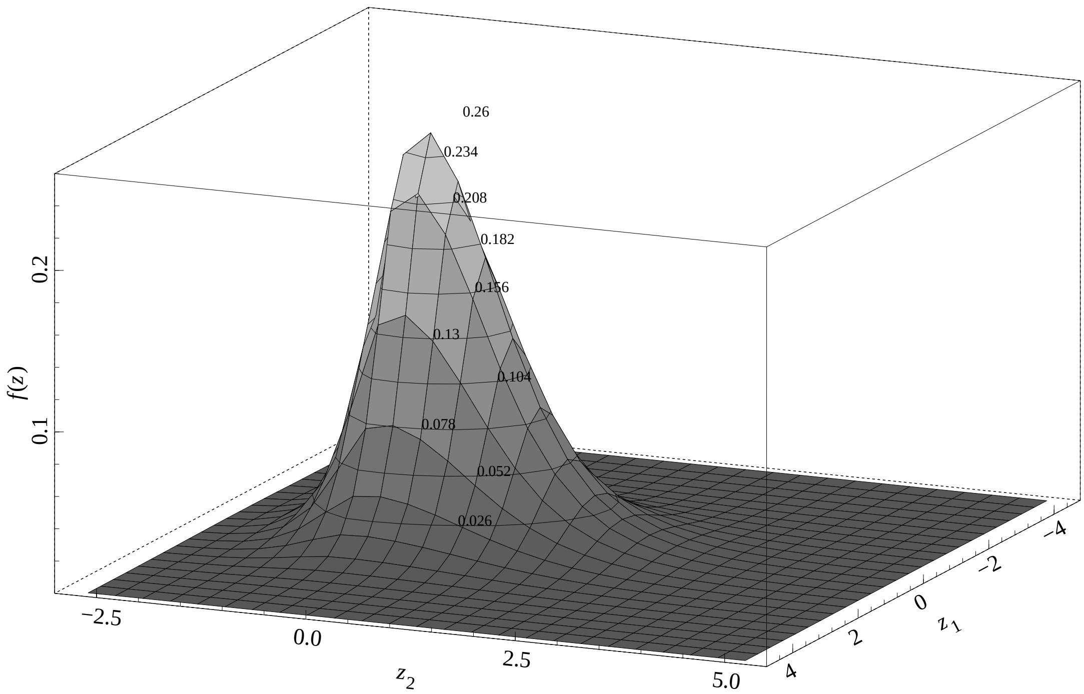

Figure 1: Graph of the $SKST\,(0, I_2, (1, 1.3), 6)$ density

3.3 Simulation

In order to assess the practical applicability of the ML method to the estimation of the skew-Student distribution, we present the results of a small simulation study. It is not our intention to provide a comprehensive Monte Carlo study. Our results, however, provide some evidence on the properties of the ML estimator when a multivariate standardized skew-Student distribution is assumed for the innovations. Consider the bivariate case with $y_t = (y_{1,t}, y_{2,t})$. The data generating process is given by Eq. (12), with $\mu_t = \mu = (0, 0)'$, $\Sigma_t = \Sigma$ a correlation matrix with off-diagonal element equal to -0.2, $z_t \sim SKST(0, I_2, \xi, \upsilon)$, where $(\log \xi_1, \log \xi_2) = (0.2, -0.2)$ and $\upsilon = 8$. This configuration implies that the innovations are skew (with skewness amounting to 0.53 and -0.53 respectively for z_1 and z_2) and have fat-tails (the kurtosis equals 4.80 for both). The sample size is set to 20,000. Table 3 reports the DGP as well as the estimation results under three assumptions for the innovations: normal, Student and (standardized) skew-Student densities.

From Table 3, it is clear that the ML method, under the correct density (i.e. the skew-Student, see column 5), works reasonably well in the sense that the estimates are very close to the "true" values. Table 3 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 Gaussian QML estimator is consistent (but inefficient). Moreover, this table also confirms the result of Newey and Steigerwald

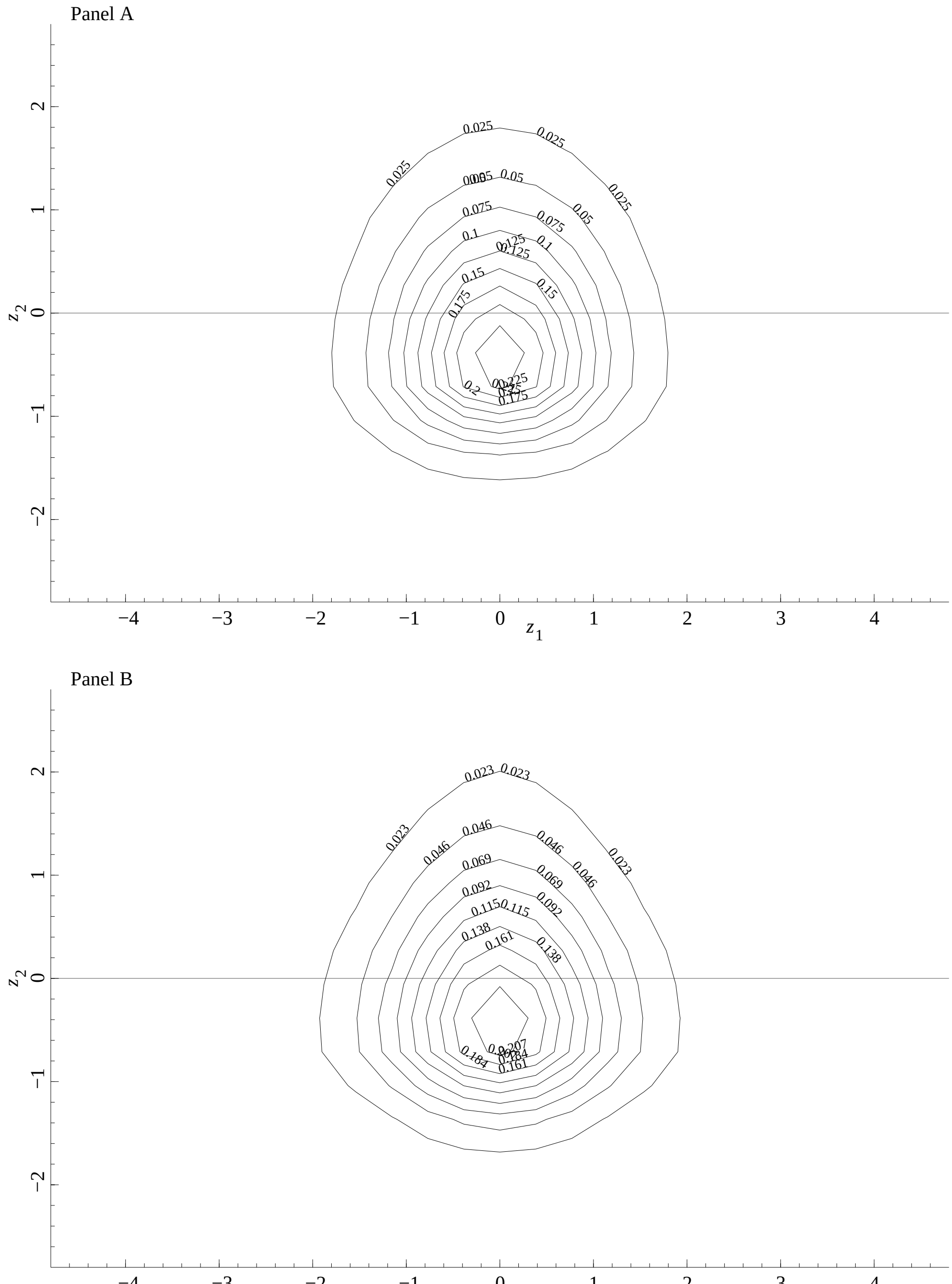

Figure 2: Panel A refers to the contours of the bivariate $SKST\,(0, I_2, (1, 1.3), 6)$ density illustrated in Figure 1. Panel B refers to the contours of a $SKST\text{-}IC(0, I_k, (1, 1.3), (6, 6))$ (see Section 3.4)

Table 3: QML Estimation Results of Simple skew-Student DGP

	DGP	Normal	Student	skew-Student
μ_1	0.0	-0.001	-0.037	-0.000
		[0.007]	[0.007]	[0.007]
μ_2	0.0	0.004	0.046	0.003
		[0.007]	[0.007]	[0.007]
σ_1^2	1.0	0.985	0.982	0.992
		[0.014]	[0.012]	[0.012]
σ_2^2	1.0	0.994	0.989	0.998
		[0.013]	[0.012]	[0.012]
ρ	0.2	-0.226	-0.219	-0.213
		[0.008]	[0.007]	[0.007]
$\log \xi_1$	0.2	-	-	0.184
				[0.010]
$\log \xi_2$	-0.2	-	-	-0.194
				[0.010]
υ	8.0	-	7.903	8.316
			[0.284]	[0.306]
Q_{20} and $Q_{20}^2(\hat{z}_1)$	-	14.791 ; 18.893	14.763 ; 17.884	14.743; 17.275
Q_{20} and $Q_{20}^2(\hat{z}_2)$	-	21.942 ; 14.492	21.852 ; 13.826	21.773; 9.936
$P_{40}(\hat{z}_1)$	-	475.768 (0.000)	316.240 (0.000)	34.504 (0.675)
$P_{40}(\hat{z}_2)$	-	585.384 (0.000)	355.068 (0.000)	30.496 (0.833)

DGP: $y_t = \mu + \Sigma^{1/2} z_t$, $t = 1, \ldots, 20000$, with $\mu = (\mu_1, \mu_2)'$, $z_t \sim SKST(0, I_2, \xi, \upsilon)$ as in (45), with $\xi = (\xi_1, \xi_2)$; σ_i^2 is the variance of y_i $(i = 1, 2)$, and ρ is the correlation coefficient between y_1 and y_2. The last four columns report the ML estimates (with the robust standard errors underneath in brackets) of the parameters of the model corresponding to the DGP with different assumptions on the distribution of z_t. The column headed "Normal" corresponds to $z_t \sim N(0, I_2)$, "Student" to $z_t \sim ST(0, I_2, \upsilon)$ as in (15), "skew-Student" to $z_t \sim SKST(0, I_2, [\xi_1, \xi_2], \upsilon)$. $Q_{20}(\hat{z}_i)$ and $Q_{20}(\hat{z}_i^2)$ are the Box-Pierce statistics of order 20 on the innovations $\hat{z}_i$ and their squares. $P_{40}(\hat{z}_i)$ is the Pearson goodness-of-fit statistic (using 40 cells) with the associated p-value beside (see footnote 7). $\hat{z}$ is given by $\hat{\Sigma}^{-1/2}(y_t - \hat{\mu})$, where $\hat{\Sigma}$ and $\hat{\mu}$ are obtained by replacing the parameters by their estimates in the corresponding formulas and $\hat{\Sigma}^{-1/2}$ is obtained from the spectral decomposition of $\hat{\Sigma}$.

(1997) that the QML estimator with a Student pseudo-likelihood is inconsistent when innovations are skew. One can see that μ is rather strongly biased under the Student density, whereas the other parameters seem less affected in this experiment. To check the model adequacy, we use the same diagnostic tools (on each innovation separately)[12] as in the empirical illustration of Section 2.2. These statistics suggest that the normal and Student densities are not appropriate, while the skew-Student is. Notice that rejecting that the margins are not correctly specified is sufficient to reject the assumption that the whole density is not appropriate. However, the converse is obviously not true. Indeed, accepting that the margins are well specified is necessary to accept that the whole density is appropriate, but it is not sufficient.

3.4 Multivariate skew densities with independent components

An obvious variation with respect to the previous class of multivariate skew densities is obtained by starting from the product of k independent $ST(0, 1, \upsilon_i)$ and applying to it the transformation defined by Eq. (25)-(26)-(28).

Definition 3 *If (i) z is defined by Eq. (42-44), where υ is simply replaced by υ_i, and (ii) z^* has a density given by Eq. (21), where $g_i(x)$ is the Student density given by Eq. (6), then z is said to be distributed as a (multivariate) skew density with independent Student components, with asymmetry parameters $\xi = (\xi_1, \ldots, \xi_k)$, and degrees of freedom $\upsilon = (\upsilon_1, \ldots, \upsilon_k)$ (with $\upsilon_i > 2$). This is denoted $z \sim SKST\text{-}IC(0, I_k, \xi, \upsilon)$. The density of z is given by:*

$$f(z|\xi, \upsilon) = \left(\frac{2}{\sqrt{\pi}}\right)^k \left[\prod_{i=1}^{k} \frac{\xi_i s_i}{1 + \xi_i^2} \frac{\Gamma(\frac{\upsilon_i+1}{2})}{\Gamma(\frac{\upsilon_i}{2})\sqrt{\upsilon_i - 2}} \left(1 + \frac{\kappa_i^2}{\upsilon_i - 2}\right)^{-\frac{1+\upsilon_i}{2}}\right], \tag{48}$$

where κ_i is defined in Eq. (47).

Note that Eq. (48) is obtained equivalently by taking the product of k independent $SKST$ $(0, 1, \xi_i, \upsilon_i)$. The main advantage of (48) with respect to (45) is that it enables a different tail behavior for each marginal, at the cost of introducing $k - 1$ additional parameters. However, nothing prevents to constrain several degrees of freedom parameters to be equal. If all the degrees of freedom parameters υ_i are equal to the degrees of freedom υ of (45), the densities (48) and (45) have exactly the same marginal moments. The fact that the components of (45) are not independent implies that its cross-moments of order 4 or higher are functions of a common single parameter υ and are thus less flexible than those of (48).

To illustrate, Panel B of Figure 2 shows the contours of the bivariate skew density with independent Student components whose parameters are $\xi_1 = 1$, $\xi_2 = 1.3$, $\upsilon_1 = \upsilon_2 = 6$. One can notice

[12] Multivariate tests of adequacy of a distribution are more appropriate tools but are usually difficult to implement. This is the reason why we use simple diagnostic tools, which should at least help to detect a major misspecification.

the difference with respect to the contours of the Panel A of the same figure, which corresponds to the skew-Student with non-independent margins. In Panel B, the contours look like less "elliptic" than in Figure Panel A (see also the graphs in Johnson, 1987, Chapter 6, for the symmetric versions of these densities).

4 Empirical application

In this section, we model jointly the four series already used in the univariate application. The specification used to model the first two conditional moments is the time-varying correlation GARCH model (TVC-GARCH) proposed by Tse and Tsui (1998), with first-order ARMA dynamics in the conditional variances and the conditional correlation, and an AR(1) equation for each conditional mean. This AR(1)-TVC(1,1)-GARCH(1,1) model is defined as follows:

$$y_t = \mu_t + \Sigma_t^{1/2} z_t \tag{49}$$

$$\mu_t = (\mu_{1,t}, \ldots, \mu_{4,t})', \quad z_t = (z_{1,t}, \ldots, z_{4,t})' \tag{50}$$

$$\mu_{i,t} = \mu_i + \phi_i(y_{i,t-1} - \mu_i) \quad (i = 1, \ldots, 4) \tag{51}$$

$$\Sigma_t = D_t \Gamma_t D_t \tag{52}$$

$$D_t = diag(\sigma_{1,t}, \ldots, \sigma_{4,t}) \tag{53}$$

$$\sigma_{i,t}^2 = \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i \varepsilon_{i,t-1}^2 \quad (i = 1, \ldots, 4) \tag{54}$$

$$\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{4,t})' = y_t - \mu_t \tag{55}$$

$$\Gamma_t = (1 - \theta_1 - \theta_2)\Gamma + \theta_1 \Gamma_{t-1} + \theta_2 \Psi_{t-1} \tag{56}$$

$$\Gamma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{23} & \rho_{34} & 1 \end{pmatrix} \tag{57}$$

$$\Psi_{t-1} = B_{t-1}^{-1} E_{t-1} E_{t-1}' B_{t-1}^{-1} \tag{58}$$

$$B_{t-1}^{-1} = diag\left(\sum_{h=1}^{m} \epsilon_{1,t-h}^2, \ldots, \sum_{h=1}^{m} \epsilon_{4,t-h}^2 \right)^{1/2} \tag{59}$$

$$E_{t-1} = (\epsilon_{t-1}, \ldots, \epsilon_{t-m}) \tag{60}$$

$$\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{4,t})' = D_t^{-1} \varepsilon_t \tag{61}$$

where μ_i, ϕ_i, ω_i, β_i, α_i $(i = 1, \ldots, 4)$, ρ_{ij} $(1 \leq i < j \leq 4)$, and θ_1, θ_2 are parameters to be estimated.[13] Ψ_{t-1} is thus the sample correlation matrix of $\{\epsilon_{t-1}, \ldots, \epsilon_{t-m}\}$. Since $\Psi_{t-1} = 1$ if $m = 1$, we must take $m \geq 4$ to have a non-trivial correlation. In this application, we set $m = 4$. Note that the TVC-MGARCH model nests the constant correlation GARCH model of Bollerslev

[13]The parameters θ_1 and θ_2 are assumed to be nonnegative with the additional constraint that $\theta_1 + \theta_2 < 1$.

(1990). Therefore, we can test $\theta_1 = \theta_2 = 0$ to check wether the constant correlation assumption is appropriate.

The estimation results of this model are gathered in Tables 4 and 5. A QML estimation procedure has been done with four different likelihoods: normal and Student in Table 4, skew-Student and skew density with independent Student components in Table 5.

The results are in line with those obtained in the univariate case. The AR(1)-TVC(1,1)-MGARCH(1,1) specification seems adequate in describing the dynamics of the series, witness the small values of the Box-Pierce statistics of order 20 on the residuals and their squares, $Q_{20}(\hat{z}_i)$ and $Q_{20}(\hat{z}_i^2)$ respectively. The residual vector $\hat{z}_t = (\hat{z}_{i,t}, \ldots, \hat{z}_{4,t})$ is defined as:

$$\hat{z}_t = \hat{\Sigma}_t^{-1/2}(y_t - \hat{\mu}_t), \tag{62}$$

where $\hat{\Sigma}_t$ and $\hat{\mu}_t$ are obtained by replacing the parameters by their estimates in the model formulas. $\hat{\Sigma}_t^{-1/2}$ has been obtained from the spectral decomposition of $\hat{\Sigma}_t$ (alternatively, a Cholesky factorization can be used).

A time-varying and very persistent correlation between the series is strongly supported if one looks at the estimates of θ_1 and θ_2 and the corresponding standard errors. On the first hand this justifies the use of a time-varying correlation specification and on the other hand the use of a multivariate model (comparing the sum of the univariate log-likelihoods with the corresponding multivariate likelihood, one can see that the multivariate approach increases the likelihood by more than 600 in all cases). Note that to facilitate the reading of the results concerning the unconditional correlation parameters (the matrix Γ), they are reported as in a 4 by 4 matrix. The upper triangle part of the matrix gives the estimated parameters while the lower triangle matrix (below the diagonal of ones) gives the associated standard errors. For instance, the estimated unconditional correlation between the CAC40 and the NIKKEI ($\hat{\rho}_{13}$) obtained with a Gaussian QML equals 0.374, with standard error 0.111.

It is clear from the estimation results reported in Table 4 that, apart from the dynamics in the first two conditional moments, the dominating feature of the four series is their fat-tail property. Indeed, the Student density increases the log-likelihood value by about 230 for only one additional parameter. Note that when comparing the standard errors related to the unconditional correlation parameters one can see that they are slightly reduced when switching from a Gaussian a a Student density. The normality assumption is also clearly rejected by the Pearson goodness-of-fit statistics (with very small p-values).[14] As in the univariate case, the Student density is clearly rejected for the NASDAQ (the p-value of the Pearson goodness-of-fit statistics being equal to 0.001).

This is confirmed by the results concerning the skew-Student density (see Table 5). First,

[14]The normality assumption is less questioned for the CAC40. This is in line with the result obtained in the univariate analysis.

Table 4: ML Estimation Results of AR-TVC-GARCH model: normal and Student distributions

	Normal				Student			
	CAC40	NASDAQ	NIKKEI	SMI	CAC40	NASDAQ	NIKKEI	SMI
μ_i	0.089	0.130	0.014	0.128	0.087	0.139	0.003	0.136
	[0.028]	[0.025]	[0.031]	[0.025]	[0.026]	[0.022]	[0.028]	[0.021]
ϕ_i	0.014	0.092	0.024	0.085	0.017	0.103	0.012	0.064
	[0.022]	[0.026]	[0.025]	[0.023]	[0.020]	[0.024]	[0.023]	[0.021]
ω_i	0.053	0.087	0.052	0.103	0.049	0.045	0.037	0.043
	[0.030]	[0.033]	[0.027]	[0.069]	[0.024]	[0.029]	[0.014]	[0.022]
β_i	0.922	0.782	0.906	0.822	0.928	0.866	0.909	0.885
	[0.032]	[0.058]	[0.025]	[0.089]	[0.025]	[0.062]	[0.015]	[0.044]
α_i	0.042	0.142	0.073	0.083	0.039	0.090	0.077	0.070
	[0.013]	[0.040]	[0.017]	[0.035]	[0.011]	[0.036]	[0.013]	[0.023]
ρ_{ij}								
CAC40	1	0.383	0.374	0.749	1	0.286	0.234	0.663
NASDAQ	[0.103]	1	0.219	0.397	[0.038]	1	0.122	0.287
NIKKEI	[0.111]	[0.088]	1	0.383	[0.038]	[0.037]	1	0.247
SMI	[0.069]	[0.087]	[0.117]	1	[0.027]	[0.038]	[0.039]	1
θ_1	0.992				0.964			
	[0.005]				[0.033]			
θ_2	0.004				0.013			
	[0.002]				[0.007]			
$\log \xi_i$	0				0			
υ	∞				7.664			
					[0.680]			
$Q_{20}(\hat{z}_i)$	19.445	21.551	19.394	18.068	24.742	16.953	13.601	7.585
$Q_{20}(\hat{z}_i^2)$	19.445	21.551	19.394	18.068	24.244	11.814	8.133	4.197
$P_{20}(\hat{z}_i)$	26.708	79.909	45.661	52.074	10.663	43.809	17.319	16.548
	(0.111)	(0.000)	(0.000)	(0.000)	(0.934)	(0.001)	(0.568)	(0.620)
SIC	11.726				11.478			
Log-Lik	-10544.3				-10315.2			

Each column reports the ML estimates of the model defined by Eq. (49)-(61), with robust standard errors underneath in brackets. The column headed "Normal" corresponds to $z_t \sim N(0, I_4)$ and "Student" to $z_t \sim ST(0, I_4, \upsilon)$ as in (15). In both cases z_t is an *i.i.d.* process. $Q_{20}(\hat{z}_i)$ is the Box-Pierce statistic of order 20 on the standardized residuals $\hat{z}_i$, $Q_{20}(\hat{z}_i^2)$ is the same for their squares, $P_{20}(\hat{z}_i)$ is the Pearson goodness-of-fit statistic (using 20 cells) with the associated unadjusted p-value beside. *SIC* is the Schwarz information criterion (divided by the sample size $T = 1816$), and *Log-Lik* is the log-likelihood value at the maximum.

Table 5: ML Estimation Results of AR-TVC-GARCH model: skew-Student and skew-Student with IC distributions

	Skew-Student				IC-Skew-Student			
	CAC40	NASDAQ	NIKKEI	SMI	CAC40	NASDAQ	NIKKEI	SMI
μ_i	0.085	0.103	-0.002	0.119	0.079	0.111	-0.014	0.116
	[0.027]	[0.023]	[0.029]	[0.022]	[0.028]	[0.023]	[0.029]	[0.023]
ϕ_i	0.015	0.081	0.011	0.058	0.011	0.075	0.005	0.060
	[0.020]	[0.024]	[0.023]	[0.022]	[0.021]	[0.024]	[0.023]	[0.022]
ω_i	0.049	0.043	0.036	0.043	0.050	0.050	0.036	0.053
	[0.024]	[0.027]	[0.014]	[0.022]	[0.029]	[0.024]	[0.014]	[0.028]
β_i	0.928	0.863	0.908	0.884	0.923	0.841	0.908	0.860
	[0.025]	[0.057]	[0.014]	[0.043]	[0.032]	[0.050]	[0.016]	[0.054]
α_i	0.039	0.095	0.077	0.071	0.043	0.114	0.080	0.087
	[0.011]	[0.034]	[0.013]	[0.023]	[0.014]	[0.032]	[0.014]	[0.030]
ρ_{ij}								
CAC40	1	0.288	0.234	0.661	1	0.311	0.272	0.679
NASDAQ	[0.037]	1	0.118	0.286	[0.049]	1	0.145	0.314
NIKKEI	[0.038]	[0.037]	1	0.245	[0.050]	[0.044]	1	0.280
SMI	[0.027]	[0.037]	[0.039]	1	[0.038]	[0.047]	[0.051]	1
θ_1	0.961				0.973			
	[0.037]				[0.032]			
θ_2	0.013				0.010			
	[0.007]				[0.007]			
$\log \xi_i$	0.035	-0.186	-0.013	-0.085	0.025	-0.172	-0.016	-0.076
	[0.034]	[0.037]	[0.036]	[0.036]	[0.034]	[0.037]	[0.036]	[0.037]
υ/υ_i	7.757				10.339	6.159	6.266	6.479
	[0.696]				[2.172]	[0.834]	[0.906]	[1.095]
$Q_{20}(\hat{z}_i)$	24.825	20.409	13.552	7.657	25.182	21.561	12.874	7.437
$Q_{20}(\hat{z}_i^2)$	24.415	11.005	8.138	4.170	23.810	9.820	8.432	4.211
$P_{20}(\hat{z}_i)$	11.435	18.730	16.989	18.906	10.708	17.121	22.741	14.829
	(0.908)	(0.474)	(0.590)	(0.462)	(0.934)	(0.581)	(0.248)	(0.733)
SIC	11.473				11.515			
Log-Lik	-10296.1				-10322.4			

Each column reports the ML estimates of the model defined by Eq. (49)-(61). The column headed "Skew-Student" corresponds to $z_t \sim SKST(0, I_4, \xi, \upsilon)$ as in (45), and "IC-Skew-Student" to $z_t \sim$ Eq. (48) (with $k = 4$). In both cases $\{z_t\}$ is an *i.i.d.* process. $Q_{20}(\hat{z}_i)$ is the Box-Pierce statistic of order 20 on the standardized residuals $\hat{z}_i$, $Q_{20}(\hat{z}_i^2)$ is the same for their squares, $P_{20}(\hat{z}_i)$ is the Pearson goodness-of-fit statistic (using 20 cells) with the associated unadjusted p-value beside. *SIC* is the Schwarz information criterion (divided by the sample size $T = 1816$), and *Log-Lik* is the log-likelihood value at the maximum.

comparing the log-likelihood values and the information criterion values suggests that this density outperforms the symmetric Student (the log-likelihood is increased by about 19 for 4 additional parameters). Second, the Pearson goodness-of-fit statistics suggest that the skew-Student is adequate in capturing the skewness of the NASDAQ and in general that all the marginals are well described by our model specification.

The last part of Table 5 gives the results for the skew density with independent Student components (see Section 3.4). Recall that unlike the skew-Student, this density has different degrees of freedom. The results suggest that the υ_i are about 6 for the last three series (the NASDAQ, NIKKEI and SMI) and are not statistically different. Even if the number of degrees of freedom of the CAC40 is higher (about 10) the precision of this estimator is even worse and one can hardly distinguish it from the other. Note that one cannot use a LR test to discriminate between the skew-Student and the skew-Student with independent components since the models are not nested. Finally, looking at the Pearson goodness-of-fit statistics one cannot reject the assumption that this last density is also adequate for modelling the excess skewness and kurtosis observed on the four marginals.

To assess the irrelevance of the normal density and the adequacy of the skew-Student density, Figures 3 and 4 plot the histogram of the probability integral transform $\hat{\zeta}_i = \int_{-\infty}^{\hat{z}_i} f_i(t)dt$ with the 95% confidence bands.

Under weak conditions (see Diebold, Gunther, and Tay, 1998), the adequacy of a density implies that the sequence of ζ_i is independent and identically uniformly distributed on the unit interval. Departure from uniformity is directly observable in the Gaussian case for the NASDAQ, NIKKEI and SMI. On the other hand, one cannot reject the assumption that the probability integral transforms of the skew-Student density are uniformly distributed.[15]

5 Conclusion

It is broadly accepted that high-frequency financial time series are heteroscedastic, fat-tailed and volatilities are related over time across assets and markets. To accommodate these stylized facts in a parametric framework a natural approach would be to rely on a multivariate GARCH or SV specification coupled with a Student density.

However, most asset returns are also skewed, which invalidates the choice of this density (it would lead to inconsistent estimates). To overcome this problem, we propose a practical and flexible method to introduce skewness in a wide class of multivariate symmetric distributions. By introducing a vector of skewness parameters, the new distributions bring additional flexibility for

[15]Confidence intervals for the ζ_i-histogram can be obtained by using the properties of the histogram under the null hypothesis of uniformity.

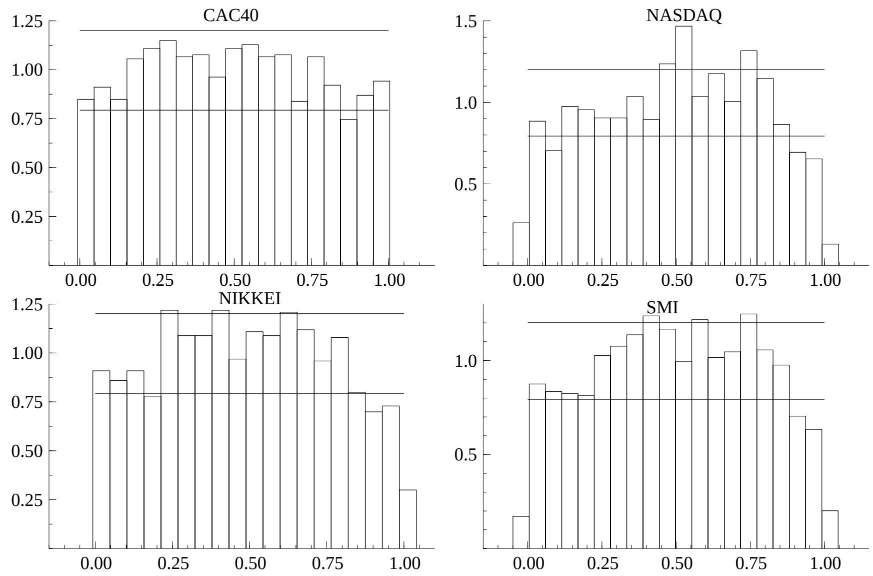

Figure 3: Histogram of the Probability Integral Transform of the CAC40, NASDAQ, NIKKEI and SMI innovations with a normal likelihood (with 20 cells).

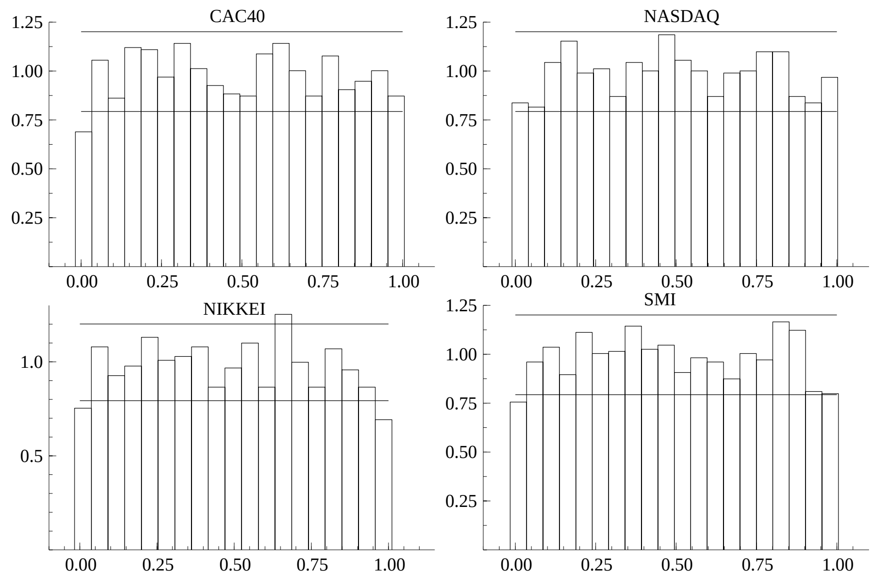

Figure 4: Histogram of the Probability Integral Transform of the CAC40, NASDAQ, NIKKEI and SMI innovations with a skew-Student likelihood (with 20 cells).

modelling time series of asset returns with multivariate volatility models. Applying the procedure to the multivariate Student density leads to a "multivariate skewed Student" density, in which each marginal has a different asymmetry coefficient. An easy variant provides a multivariate skewed density that can have different tail properties on each coordinate. These densities are found to outperform their symmetric competitors (the multivariate normal and Student) for modelling four daily stock market indexes, and therefore are of great potential interest for the empirical modelling of several asset returns together.

Additional empirical studies based on these flexible distributions should be carried out to explore deeply the skewness and kurtosis properties of asset returns, including the co-skewness and co-kurtosis aspects in a multivariate framework (see Hafner, 2001).

Another potential area of application of the new densities is in Bayesian inference, for the design of simulators for Monte-Carlo integration of posterior densities that are characterized by different skewness and tail properties in different directions of the parameter space. In this respect, some of the densities we have proposed are related to the split-Student importance function proposed by Geweke (1989). This is obviously a different research topic, that we leave for further work.

Finally, a natural extension of this paper would be to generalize the GARCH specification to higher moments. Indeed, in a univariate framework Hansen (1994), introduces dynamics through the 3rd and 4th order moments by conditioning the asymmetry and fat-tail parameters on past errors and their square. In the same spirit, Harvey and Siddique (1999) and Lambert and Laurent (2000) provide alternative specifications to introduce dynamics in higher order moments.

To conclude, this new family of multivariate skewed densities and in particular the multivariate skewed Student density seems to be a promising specification to accommodate both the high kurtosis and the skewness inherent in most asset returns.

References

Azzalini, A., and A. Capitanio (1996): "The Multivariate Skew-Normal Distribution," *Biometrika*, 83, 715–726.

Ball, C., and A. Roma (1993): "A Jump Diffusion Model for the European Monetary System," *Journal of International Money and finance*, 12, 475–492.

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.

——— (1990): "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model," *Review of Economics and Statistics*, 72, 498–505.

Bollerslev, T., R. Engle, and J. Wooldridge (1988): "A Capital Asset Pricing Mode1 with Time Varying Covariances," *Journal of Political Economy*, 96, 116–131.

Bollerslev, T., and J. Wooldridge (1992): "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances," *Econometric Reviews*, 11, 143–172.

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

Box, G., and G. Jenkins (1970): *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.

Branco, M., and D. Dey (2000): "A class of Multivariate Skew-Elliptical Distributions," *Forthcoming* in Journal of Multivariate Analysis.

Brannas, K., and N. Nordman (2001): "Conditional Skewness Modelling for Stock Returns," Umea Economic Studies 562.

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 (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.

Drèze, J. (1978): "Bayesian Regression Analysis using poly-t Densities," *Journal of Econometrics*, 6, 329–354.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

——— (2001): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," Mimeo, UCSD.

Engle, R., and G. González-Rivera (1991): "Semiparametric ARCH Model," *Journal of Business and Economic Statistics*, 9, 345–360.

Engle, R., and F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

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.

Geweke, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Geweke, J., and G. Amisano (2001): "Compound Markov Mixture Models with Application in Finance," Mimeo, University of Iowa.

Giot, P., and S. Laurent (2001a): "Modelling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models," Maastricht University METEOR RM/01/026.

——— (2001b): "Quantifying Market Risk for Long and Short Traders," Forthcoming in European Investment Review.

Hafner, C. (2001): "Fourth Moment of Multivariate GARCH Processes," CORE DP 2001-39.

Hansen, B. (1994): "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

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

Hong, C. (1988): "Options, Volatilities and the Hedge Strategy," Unpublished Ph.D. diss., University of San Diego, Dept. of Economics.

Jeantheau, T. (1998): "Strong Consistency of Estimators for Multivariate ARCH models," *Econometric Theory*, 14, 70–86.

Johnson, M. (1987): *Multivariate Statistical Simulation*. Wiley.

Jones, M. (2000): "Multivariate T and Beta Distributions Associated with the Multivariate F Distribution," *Forthcoming* in Metrika.

Jones, M., and M. Faddy (2000): "A Skew Extension of the t Distribution, with Applications," mimeo, Department of Statistics, Open University, Walton Hall, UK.

Jorion, P. (1988): "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.

Kon, S. (1982): "Models of Stock Returns, a Comparison," *Journal of Finance*, 39, 147–165.

Kraft, D., and R. Engle (1982): "Autoregressive Conditional Heteroskedasticity in Multiple Time Series," unpublished manuscript, Department of Economics, UCSD.

Kroner, F., and V. Ng (1998): "Modelling Asymmetric Comovements of Asset Returns," *The Review of Financial Studies*, 11, 817–844.

Lambert, P., and S. Laurent (2000): "Modelling Skewness Dynamics in Series of Financial Data," Discussion Paper, Institut de Statistique, Louvain-la-Neuve.

——— (2001): "Modelling Financial Time Series Using GARCH-Type Models and a Skewed Student Density," Mimeo, Université de Liège.

Laurent, S. (2001): "Analytical Derivates of the APARCH Model under a Skewed Student Assumption for the Innovation Process," Mimeo, Université de Liège.

Lee, S., and B. Hansen (1994): "Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and IGARCH Models," *Econometric Theory*, 10, 29–52.

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.

Lumsdaine, R. (1996): "Asymptotic Properties of the Quasi Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models," *Econometrica*, 64, 575–596.

Mauleón, I., and J. Perote (1999): "Estimation of Multivariate Densities with Financial Data: the Performance of the Multivariate Edgeworth-Sargan Density," Proceedings of the 12th Australian Finance and Banking Conference, Sidney.

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.

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

Neely, C. (1999): "Target Zones and Conditional Volatility: the Role of Realignments," *Journal of Empirical Finance*, 6, 177–192.

Newey, W., and D. Steigerwald (1997): "Asymptotic Bias for Quasi Maximum Likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, 3, 587–599.

Ng, V., R. Engle, and M. Rothschild (1992): "A Multi-Dynamic Factor Model for Stock Returns," *Journal of Econometrics*, 52, 245–265.

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

Palm, F., and P. Vlaar (1997): "Simple Diagnostics Procedures for Modelling Financial Time Series," *Allgemeines 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.

Richard, J.-F., and H. Tompa (1980): "On the Evaluation of poly-t Density Functions," *Journal of Econometrics*, 12, 335–351.

Sahu, S., D. Dey, and D. Branco (2001): "A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models," mimeo, Department of Statistics, Univerisity of Sao Paulo.

Simkowitz, M., and W. Beedles (1980): "Asymmetric Stable Distributed Security Returns," *Journal of the American Statistical Association*, 75, 306–312.

So, J. (1987): "The Distribution of Foreign Exchange Price Changes: Trading Day Effects and Risk measurement - A Comment," *Journal of Finance*, 42, 181–188.

Tse, Y., and A. Tsui (1998): "A Multivariate GARCH Model with Time-Varying Correlations," Mimeo, Department of Economics, National University of Singapore.

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.

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

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.