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

0119

MODELING SKEWNESS DYNAMICS IN SERIES OF FINANCIAL DATA USING SKEWED LOCATION-SCALE DISTRIBUTIONS

P. LAMBERT and S. LAURENT

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

Modeling skewness dynamics in series of financial data using skewed location-scale distributions

Philippe Lambert* and Sébastien Laurent†

June 14, 2001

Abstract

We show how the ARMA-Power GARCH model for the conditional mean and variance can be adapted to analyze times series data showing asymmetry. Dynamics is introduced in the location and the dispersion parameters of the skewed Student and of the skewed stable distributions using the same type of structure found in the conditional mean and in the conditional variance in the ARMA-APARCH model.

We also propose a general dynamic model for skewness as measured by the odds ratio of having the next observation greater than the conditional mode.

This general tool is illustrated by the analysis of the DEM-USD exchange rate over the 1980-1996 period.

Keywords: ARMA, Power GARCH, (time-varying) skewness, kurtosis, stable distribution, skewed Student.

*Correspondence to Philippe Lambert, 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. Phone: +32-10-47-28-01. Fax: +32-10-47-30-32

†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

Traditional regression tools have shown their limitation in the modeling of continuous data, particularly in finance. Assuming that only the mean response could be changing with covariates while the variance remains constant often revealed to be an unrealistic assumption in practice. This fact is particularly obvious in series of financial data where clusters of volatility can be detected in a direct plot of the response. The modelling of the mean can then often be limited to simple first order autoregressive structure while the dynamics observed in the dispersion is clearly the dominating feature in the data.

Econometricians have thus extended traditional time series tools such as ARMA (Box and Jenkins 1970) and ARFIMA models (Hosking 1981) for the mean to essentially equivalent models for the variance. ARCH (Engle 1982), GARCH (Bollerslev 1986), Asymmetric Power ARCH (APARCH) (Ding, Granger, and Engle 1993) and Fractionally Integrated GARCH (FIGARCH) models (Baillie, Bollerslev, and Mikkelsen 1996), among others, are now commonly used to describe and forecast changes in volatility of financial time series.

Weiss (1986) and Bollerslev and Wooldridge (1992) show that the quasi- (or pseudo-) maximum likelihood estimator is consistent assuming that the conditional mean and the conditional variance are specified correctly. This estimator is, however, inefficient with the degree of inefficiency increasing with the degree of departure from normality (Engle and González-Rivera 1991).

Even if the specification of the conditional moments is an important issue, few attention has been devoted to the search of an appropriate distribution. In general, these sophisticated linear models for the mean and the variance usually rely on simplistic assumptions on the stochastic structure (normality).

Hansen (1994) argues that the reason why most applications have ignored higher-order features of the conditional distribution may be because only the mean and variance generate significant excitement. But this lack of excitement does not imply that higher-order features should be completely ignored.

Indeed, the issue of skewness (asymmetry) and kurtosis (fat-tails) in empirical finance is important in many respects. Giot (2000) stresses the role of fat-tail distributions in parametric Value-at-Risk applications. On the other hand, 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]

The empirical literature on high frequency financial time series has shown that the normality assumption is in general unrealistic. Moreover, GARCH models assuming Student-, mixture of normal- or Student- distributed errors cannot capture all the skewness and leptokurtosis (Beine and Laurent, 1999; Jorion, 1988; Neely, 1999; Vlaar and Palm, 1993), although they can do a good job in some rare cases.

To overcome this problem, Liu and Brorsen (1995) and Bidarkota and McCulloch (1998) tried to combine GARCH processes and stable densities. A major drawback of the stable distribution is that the variance of the innovation process does not exist. See McCulloch (1996) for a survey of the financial applications of stable distributions.

Hansen (1994) builds a skewed Student by mixing two symmetric Student densities. Paolella (1997) and Mittnik and Paolella (2000) demonstrate the simultaneous need for both APARCH specification and non-normal innovation distributions for modelling daily exchange rate returns (Bulgarian and East Asian currencies against the US dollar). Jondeau and Rockinger (2000) propose to use a conditional generalized-t distribution for the residuals (in a GARCH framework) and allow for time-varying conditional skewness and kurtosis.

In this paper, we consider possibly skewed and heavy-tailed distributions for the response. Traditional times series models are reformulated as dynamic models for the mean and the variance.

[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.

The common structure of ARMA and GARCH models are then clearly visible in the conditional mean and the conditional variance. These structures are adapted to model the location and the dispersion parameters of four parameter distributions. The distributions that we shall consider are the skewed Student and the skewed stable. The skewness and the tails properties will first be considered time invariant. Dynamics will then be introduced to allow skewness to change over time in a totally different way than Hansen (1994) who conditions the 3rd order moment on past residuals and their square. Our proposal has the major advantage that the conditional skewness parameter has a clear interpretation as the odds ratio of having the next observation above the mode. This allows sensible choices for the quantities on which one conditions to predict future skewness.

The paper is organized as follows. Section 2 briefly introduces the investigated distributions. Section 3 presents the modelling framework of the location and dispersion equations while Section 4 describes the maximum likelihood estimation procedure. A general dynamic model for skewness is proposed in Section 5. An empirical application on daily exchange rate returns compares the performance of the different distributions in Section 6. We conclude the paper by a discussion in Section 7.

2 Alternative distributions

2.1 The t-distribution

The Student t-distribution is a well known alternative to the normal distribution. As a reminder, a random variable Y is distributed as a Student$(\mu, \sigma^2, \upsilon)$ if the density can be written

$$g_\upsilon(y|\mu,\sigma^2) \quad = \quad \frac{\Gamma(\frac{\upsilon+1}{2})}{\sqrt{\pi(\upsilon-2)}\;\Gamma(\frac{\upsilon}{2})} \; \frac{1}{\left\{1+\frac{1}{\upsilon-2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^{\frac{\upsilon+1}{2}}} \; \frac{1}{\sigma} \tag{1}$$

where μ, $\sigma^2 > 0$ and $\upsilon \in I\!N \setminus [0,2]$ are respectively the mean, the variance and the degrees of freedom. As υ tends to infinity, the Student tends in distribution to the normal. The t-distribution can be extended by allowing the degree of freedom to take real values in $(2,\infty[$. The thickness of the tails is decreasing with υ. An alternative parametrisation allows υ to be less than 2. In these cases, the variance is infinite and σ^2, which is not the variance anymore, remains a dispersion parameter.

2.2 The skewed t-distribution

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 γ) distribution $g(y;\gamma)$ by changing the scale at each side of the mode. More specifically,

$$f(y|\gamma,\xi) \quad = \quad \frac{2}{\xi+\frac{1}{\xi}}\left\{g(\xi(y-\gamma);0)I_{(-\infty,\gamma)}(y) + g((y-\gamma)/\xi;0)I_{[\gamma,\infty)}(y)\right\} \tag{2}$$

is a unimodal density with the same mode as $g(y;\gamma)$ and a skewness parameter $\xi > 0$ such that the ratio of probability masses above and below the mode is

$$\frac{\Pr(Y \geq \gamma|\xi)}{\Pr(Y < \gamma|\xi)} \quad = \quad \xi^2 \tag{3}$$

Note that the density $f(y|\gamma, 1/\xi)$ is the symmetric of $f(y|\gamma,\xi)$ with respect to the mode. Therefore, working with $\xi' = \log(\xi)$ might be preferable to indicate the sign of the skewness.

If we take for $g(y;\gamma)$ the Student density $g_\upsilon(y|\mu=\gamma,\sigma^2)$ in Equation (1), we obtain the four parameter skewed Student distribution in Fernández and Steel (1998). These parameters all have a clear interpretation:

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

Four important aspects of the distribution can thus be independently specified.

The skewed normal distribution directly obtained by applying Equation (2) to the symmetric normal density with mean γ and variance σ^2 is a limiting case ($\upsilon \to \infty$) of the skewed Student with the same tail properties than the traditional normal.

2.3 The asymmetric stable distribution

Stable distributions are by definition the only possible limiting distributions for the sum of independent, identically distributed random variables (see for example, Samorodnitsky and Taqqu, 1994; Adler, Feldman, and Taqqu, 1998; Lambert and Lindsey, 1999). Therefore these limiting distributions generalize the normal that arises in the central limit theorems when the variance of the summed random variables is finite.

Unfortunately, these four parameter distributions are only known through their characteristic function $\phi(t)$. A possible parameterization of the latter is given by

$$\begin{aligned} \log \phi(t) &= i\gamma t - |t|^\alpha \delta^\alpha \ \exp\left[-i\beta \tfrac{\pi}{2} \eta_\alpha \ \mathrm{sign}(t)\right] \\ \eta_\alpha &= \min(\alpha, 2-\alpha) = 1 - |1-\alpha| \end{aligned} \tag{4}$$

where γ, $\delta > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$ are location, dispersion, skewness and tail parameters. The cases $\alpha = 1$, $\alpha = 2$ and $(\beta, \alpha) = (1, 0.5)$ respectively correspond to the Cauchy, the normal and Lévy distributions. For other parameter values, the density $g_\alpha(y|\gamma, \delta, \beta)$ can only be obtained using numerical approximations (see Part VII in Adler, Feldman, and Taqqu, 1998 ; Lambert and Lindsey, 1999 ; Hoffmann-Jørgensen , 1994, I, pp. 406–411). The density is unimodal Yamazato (1978) and bell-shaped Gawronski (1984). It is symmetric when $\beta = 0$ and right (left) skewed when β is negative (positive). Stable distributions ($\alpha \neq 2$) are said heavy tailed because the distribution and survivor functions show a power decay

$$\begin{aligned} G_\alpha(x) &\propto (-x)^{-\alpha} \ \text{ as } \ x \to -\infty \\ \overline{G}_\alpha(x) &= 1 - G_\alpha(x) \propto x^{-\alpha} \ \text{ as } \ x \to +\infty \end{aligned} \tag{5}$$

as in the Pareto case. This can be contrasted to the normal distribution ($\alpha = 2$) where an exponential decay is observed

$$\overline{G}_{\alpha=2}(x) \propto \frac{\exp(-x^2/2)}{x}$$

One important aspect of stable distributions is that the four parameters can be specified in an independent manner (the space parameter is simply the cartesian product of the four parameter spaces). This is a remarkable result for statisticians used to the exponential family where the specification of the mean often affects in an important manner other aspects of the shape of the considered distribution.

However, the range of possible skewness introduced by β in the stable distribution highly depends on the value of the tail parameter α as illustrated in Figure 1. If we measure skewness by the log of the ratio ξ^2 of probability masses above and below the mode as given in Equation (3), we see that the asymmetry introduced by non-zero values of β becomes negligible as α approaches 2. It presents at least two drawbacks:

- the analyst might think, for example, that the data present severe skewness because the estimated value of β is close to -1 or 1. However, the last figure shows that he might be completely wrong if the fitted stable distribution is characterized by an α parameter close to 2. A reparameterization of the stable distribution is thus desirable to reflect reliably the skewness introduced by the skewness parameter. This parameter would obviously be a function of β and α.
- the analyzed dataset might present (severe) skewness and have heavy tails (cf. Equation (5)) with an α close to 2. The stable family of distributions cannot adequately describe such a phenomenon.

2.4 The skewed stable distribution

The skewed stable distribution is an alternative to the traditional (asymmetric) stable distribution to model heavy tailed processes. The two drawbacks mentioned above about the stable disappear in this new four parameter family of distributions. Moreover, each of the four parameters involved describe in an unambiguous way four fundamental aspects of the underlying unimodal densities.

If $g_\alpha(y|\gamma,\delta,\beta=0)$ denotes the density of the symmetric unimodal stable distribution (with a mode at γ) obtained by inverting the characteristic function in Equation (4), the density of the skewed stable distribution is defined as

$$f_\alpha(y|\gamma,\delta,\xi) = \frac{2}{\xi+\frac{1}{\xi}}\left\{g_\alpha(\xi(y-\gamma)|0,\delta,\beta=0)I_{(-\infty,\gamma)}(y)+g_\alpha((y-\gamma)/\xi|0,\delta,\beta=0)I_{[\gamma,+\infty)}(y)\right\} \tag{6}$$

following the general method described by Fernández and Steel (1998). The four parameters in that distribution now have an unambiguous interpretation:

- γ is the mode,
- δ is the dispersion parameter,
- ξ^2 is the skewness parameter giving the ratio of probability masses above and below the mode,
- α is the tail parameter such that (when $\alpha \neq 2$)

$$\begin{aligned} F_\alpha(y|\gamma,\delta,\xi) &\propto (-y)^{-\alpha} \text{ as } y\to-\infty \\ \overline{F}_\alpha(y|\gamma,\delta,\xi) &\propto y^{-\alpha} \text{ as } y\to+\infty \end{aligned}$$

The location parameter γ, as the mode, is easier to interpret than the location measure considered in the usual stable distribution.

The amount of skewness corresponding to a given value of ξ is the same whatever the value of the tail parameter α. Moreover, skewness of any size can now be introduced in the distribution through the parameter ξ.

Finally, for large values of $|y|$, the stable and the skewed stable distributions show the same Paretian behavior in the tails.

2.5 Skewed location-scale distributions

Consider the following class of densities

$$f(y|\gamma,\delta,\xi) = \frac{2}{\xi+\frac{1}{\xi}}\left\{g(\xi(y-\gamma)|0,\delta)I_{(-\infty,\gamma)}(y)+g((y-\gamma)/\xi|0,\delta)I_{[\gamma,+\infty)}(y)\right\} \tag{7}$$

with

$$g(y|\gamma,\delta) = \frac{1}{\delta}\,h\left(\frac{y-\gamma}{\delta}\right)$$

where $h(\cdot)$ is a unimodal and symmetric density about 0 with existing first and second derivatives at 0. Thus, $f(y|\gamma,\delta,\xi)$ is the skewed version of a location-scale density $g(y|\gamma,\delta)$ unimodal and symmetric about γ with

- mode γ,
- skewness parameter such that ξ^2 is the ratio of probabilities above and below the mode,
- a dispersion parameter δ^2 measuring the inverse of the curvature of the density at the mode. Indeed, we have

$$\begin{cases} \left(-\frac{\partial^2 \log f(y|\gamma,\delta,\xi)}{\partial_L y^2}\right)^{-1}_{|y=\gamma} = \frac{h(0)}{-h''(0)}\ \frac{1}{\xi^2}\delta^2 \\ \left(-\frac{\partial^2 \log f(y|\gamma,\delta,\xi)}{\partial_R y^2}\right)^{-1}_{|y=\gamma} = \frac{h(0)}{-h''(0)}\ \xi^2\delta^2 \end{cases} \tag{8}$$

  as inverse curvature measure on, respectively, the left and the right of the mode γ (with ∂/∂_L and ∂/∂_R denoting respectively left and right derivatives).

We shall write

$$Y \sim \mathrm{SLS}(h,\gamma,\delta,\xi)$$

to denote that Y has a skewed location-scale (SLS) distribution with generating density $h(\cdot)$.

Special cases

1. Normal distribution: we have

$$h(x) = \frac{1}{\sqrt{2\pi}}\exp(-0.5\ x^2)$$

   Note that, as

$$\frac{h(0)}{-h''(0)} = 1\ ,$$

   Equation (8) just yields δ^2 for the inverse quadrature measure.

2. Skewed Student distribution: we have seen that

$$h(x) = h_\upsilon(x) = \frac{\Gamma(\frac{\upsilon+1}{2})}{\sqrt{\pi(\upsilon-2)}\ \Gamma(\frac{\upsilon}{2})}\ \frac{1}{\left(1+\frac{1}{\upsilon-2}\ x^2\right)^{\frac{\upsilon+1}{2}}}$$

   Note that, as

$$\frac{h(0)}{-h''(0)} = \frac{\upsilon-2}{\upsilon+1}\ ,$$

   Equation (8) just yields $\frac{\upsilon-2}{\upsilon+1}\delta^2$ for the inverse quadrature measure when $\xi^2=1$.

3. Skewed stable distribution: the generating density is

$$h(x) = h_\alpha(x) = g_\alpha(x|\gamma=0,\delta=1,\beta=0)$$

   obtained by inverting the characteristic function in Equation (4). Equation (8) just yields $\frac{h_\alpha(0)}{-h''_\alpha(0)}\delta^2$ for the inverse quadrature measure when $\xi^2=1$.

2.6 Distribution and quantile functions of a skewed distribution

Assume that $f(y|\gamma,\delta,\xi)$ is a density skewed using the technique proposed by Fernández and Steel (1998) in Equation (2) applied on a continuous and symmetric location-scale density $g(y|\gamma,\delta)$ with location and dispersion parameters γ and δ. We can relate the distribution function (cdf) F and the quantile function F^{-1} to the starting cdf G and quantile function G^{-1}. We have

$$F(y|\gamma,\delta,\xi) = \begin{cases} \frac{2}{1+\xi^2}\, G(\xi(y-\gamma)|0,\delta) & \text{if} \quad y < \gamma \\ 1 - \frac{2}{1+\xi^{-2}}\, G(-\xi^{-1}(y-\gamma)|0,\delta) & \text{if} \quad y \geq \gamma \end{cases}$$

for the cdf and

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

for the quantile function.

3 Reformulation of time series models

3.1 The ARMA(p_1,q_1) model

Efficient tools have been developed to model long sequences of data or time series. The ARMA model of Box and Jenkins (1970) is available in most generalist statistical packages to model the serial dependence arising in sequences of data.

More specifically, if $\{y_1,\ldots,y_n\}$ is a sequence of observations indexed by the (discrete) time t and L the lag operator, then an ARMA(p_1,q_1) model is given by

$$Y_t = \phi_{11}(Y_{t-1}-\psi_{1t-1}) + \ldots + \phi_{1p_1}(Y_{t-p_1}-\psi_{1t-p_1}) + \epsilon_t + \theta_{11}\epsilon_{t-1} + \ldots + \theta_{1q_1}\epsilon_{t-q_1} \tag{9}$$

or equivalently by

$$\begin{gathered} \Phi_1(L)(Y_t-\psi_{1t}) = \Theta_1(L)\epsilon_t \\ \Phi_1(L) = 1 - \textstyle\sum_{i=1}^{p_1} \phi_{1i}L^i \ ; \ \Theta_1(L) = 1 + \sum_{j=1}^{q_1} \theta_{1j}L^j \end{gathered} \tag{10}$$

where

- the roots of the polynomials $\Phi_1(L)$ and $\Theta_1(L)$ are supposed to lie outside the unit circle,
- the $\{\epsilon_t\}$ is $i.i.d.(0,\sigma^2)$.
- ψ_{1t} is the unconditional mean of Y_t which could change with (possibly) time varying (exogenous) covariates $\{\boldsymbol{x}_t\}$ if desired.

The "epsilon" notation in Equation (10) is interesting because it allows to derive the properties of the Y_t's (such as its unconditional moments) in an easy way. Moreover, expressing Y_t (or ϵ_t) as a function of its previous states and of past values of the ϵ's (Y_t's) can easily be done by inverting the $\Phi_1(L)$ ($\Theta_1(L)$) polynomials in Equation (10).

However, the ϵ notation hides a complex structure in the underlying conditional first moment where the hypotheses and the dynamics introduced by the ARMA model can be better understood. After some straightforward manipulations of Equation (10), we get, for the conditional mean μ_t,

$$\mu_t = \psi_{1t} + \sum_{i=1}^{p_1} \phi_{1i}(y_{t-i}-\psi_{1,t-i}) + \sum_{j=1}^{q_1} \theta_{1j}(y_{t-j}-\mu_{t-j}) \tag{11}$$

with

$$e_t = y_t - \mu_t$$

as observed value for ϵ_t.

The conditional mean is thus divided into three contributions:

1. a covariate part ψ_{1t} as in traditional regression model which is also the unconditional mean of Y_t,
2. the AR(p_1) part, that corrects the regression model using its past deficiencies as measured by the previous p_1 residuals, i.e. the differences between past observations and their respective unconditional mean,
3. the MA(q_1) part that corrects the regression and AR(p_1) models with their q_1 past errors.

3.2 The GARCH(p_2, q_2) model

The above ARMA model assumes that the conditional variance of Y_t, which is also the variance of ϵ_t, is constant and equal to σ^2. As the mean of ϵ_t is zero, this is simply $E(\epsilon_t^2)$.

The generalized autoregressive conditionally heteroskedastic model , i.e. GARCH(p,q), defined by Bollerslev (1986), allows the conditional variance of ϵ_t to change over time as specified by:

$$\sigma_t^2 = \mathrm{E}(\epsilon_t^2|\mathcal{F}_{t-1}) = \omega + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{12}$$

with $\omega > 0, \alpha_i \geq 0, \beta_j \geq 0$, where $e_t^2 = (y_t - \mu_t)^2$ is (say) the squared residual from the ARMA model for the mean.

We propose the following alternative definition

$$\sigma_t^2 = \psi_{2t} + \sum_{i=1}^{p_2} \phi_{2i}(e_{t-i}^2 - \psi_{2,t-i}) + \sum_{j=1}^{q_2} \theta_{2j}(e_{t-j}^2 - \sigma_{t-j}^2) \tag{13}$$

for the conditional variance, where ψ_{2t} is the unconditional variance which could be changing with (possibly) time varying covariates. It is closely related to the expression that Chung (1999) considered in the FIGARCH framework. Note that there is a one to one relation between (12) and (13). Indeed,

$$\omega = \psi_{2t}(1 - \sum_{i=1}^{p_2} \phi_{2i} L),\ \alpha_i = \phi_{2i} + \theta_{2j},\ \beta_j = -\phi_{2i},\ p = q_2,\ q = \max[p_2, q_2],$$

$$\phi_{2i} = 0 \text{ for } i > p_2 \quad \text{and} \quad \theta_{2j} = 0 \text{ for } j > q_2$$

Translating the positivity constraints of (12) for (13) leads to

$$\psi_{2t} > 0,\ \phi_{2i} + \theta_{2j} \geq 0,\ \phi_{2i} \leq 0$$

Equation (13) shares the same structure as the conditional mean in the ARMA model of Section 3.1. Using the same arguments, one can show that $\{\epsilon_t^2 - \psi_{2t}^2\}$ is first-order stationary when the roots of

$$\Phi_2(L) = 1 - \sum_{i=1}^{p_2} \phi_{2i} L^i$$

all lie outside the unit circle. As Equation (13) can be rewritten as

$$(\sigma_t^2 - \psi_{2t}) = \sum_{i=1}^{p_2} \phi_{2i}(e_{t-i}^2 - \psi_{2,t-i}) + \sum_{j=1}^{q_2} \theta_{2j}[(e_{t-j}^2 - \psi_{2,t-j}) - (\sigma_{t-j}^2 - \psi_{2,t-j})],$$

the mean m of $(\epsilon_t^2 - \psi_{2t})$ at stationarity must check

$$m = \sum_{i=1}^{p_2} \phi_{2i} m + \sum_{j=1}^{q_2} \theta_{2j}[m - m]$$

Thus, using the stationarity condition, it requires m to be zero and hence

$$\mathrm{E}(\epsilon_t^2) = \psi_{2t}$$

In particular, when

$$\psi_{2t} = \psi_2 \quad \forall\ t,$$

the variance of ϵ_t at stationarity is ψ_2. This is the main motivation for the above proposed alternative definition of the GARCH.

In the rest of the paper, we shall speak about an ARMA-GARCH model when the location and the dispersion parameters of the considered response conditional distribution can be expressed as in Equations (11) and (13).

3.3 The APARCH(p_2, q_2) model

The GARCH(p,q) model has been extended in various ways. One of the most interesting developments is the asymmetric power (G)ARCH or APARCH(p,q) model (Ding, Granger, and Engle 1993) which allows to take account of both the conditional asymmetry and (possible) long memory property empirically described in stock market volatility. This long memory is introduced by a GARCH like model for σ_t^ζ instead of the conditional variance σ_t^2:

$$\sigma_t^\zeta = \omega + \sum_{i=1}^{q} \alpha_i k(e_{t-i})^\zeta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\zeta$$

where

$$\begin{aligned}
&\sigma_t^\zeta = \mathrm{E}(\epsilon_t^\zeta | \mathcal{F}_{t-1}) \quad ; \quad e_t = y_t - \mu_t \\
&k(e_{t-i}) = |e_{t-i}| - \tau_i e_{t-i} \\
&\omega > 0, \alpha_i \geq 0, \beta_j \geq 0, \zeta > 0, -1 < \tau_i < 1
\end{aligned}$$

This specification has been motivated by a stylized fact detected by Taylor (1986) who first observed that the absolute returns ($|y_t|$) in financial time series are positively autocorrelated, even at long lags. Ding, Granger, and Engle (1993) found that that the closer ζ to 1, the larger the memory of the process.[4] The extra set of τ_i parameters allows a different effect of a positive and a negative shock on volatility. More details can be found in the above mentioned reference. The properties of the APARCH model has been studied recently by He and Terasvirta (1999a, 1999b). It is interesting to know that the APARCH model includes at least six other proposed alternatives to the GARCH model as special cases.

As in the previous section and to keep the same structure as Equations (11) and (13), we propose an alternative specification of the APARCH(p_2, q_2):

$$\sigma_t^\zeta = \psi_{2t} + \sum_{i=1}^{p_2} \phi_{2i} \left\{ k(e_{t-i})^\zeta - \psi_{2,t-i} \right\} + \sum_{j=1}^{q_2} \theta_{2j} \left\{ k(e_{t-j})^\zeta - \sigma_{t-j}^\zeta \right\} \tag{14}$$

For equity returns and interest rates, it is particularly unlikely that positive and negative shocks have the same impact on volatility (see Black, 1976; Glosten, Jagannathan, and Runkle, 1993; Nelson, 1991; Engle, Ng, and Rothschild, 1990 among others). In general, with exchange rate returns (like in Section 6), assuming $\tau_i = 0 \ \ \forall i$ is empirically supported.[5]

[4] Tse (1998) extended the APARCH by including a pure long memory feature (FIAPARCH). Even if such an extension is straightforward, few is known about the properties of such a model. This is the reason why we do not tackle this specification in this paper.

[5] Results, not reported here, highly support this assumption.

4 Time series models and skewed distributions

The objective is now to adapt ARMA-GARCH and ARMA-APARCH models to describe dynamics in skewed location-scale distributions. In the skewed Student (say) distribution, when the skewness parameter ξ differs from one, μ and σ^2 are not the mean and the variance anymore, but the mode and a dispersion parameter. Therefore, dynamics cannot be properly introduced in the skewed Student distribution by specifying dynamics in μ and σ^2 using Equations (11) and (13).

The same type of problems arise with the skewed stable distribution as the mean is possibly non finite (when $\alpha \leq 1$) and the variance infinite (except when $\alpha = 2$).

4.1 Parameter estimation

When the chosen distribution for the innovation is the traditional normal, estimating the parameters appearing in an ARMA-GARCH structure is straightforward. One way to tackle the problem is to assume that the responses

$$(Y_t|\mathcal{F}_{t-1}) \sim N(\mu_t, \sigma_t^2)$$

are conditionally independent with the first two conditional moments given by Equations (11) and (13). The parameters in the ARMA-GARCH model can then be estimated by maximizing (apart from initial conditions) the corresponding likelihood

$$L = \prod_{t=1}^{n} f(y_t|\mu_t, \sigma_t^2)$$

where $f(\cdot)$ is the normal density.

The technique is similar when the Student distribution is considered. Using the same equations for the first two conditional moments, we get as likelihood

$$L = \prod_{t=1}^{n} f_v(y_t|\mu_t, \sigma_t^2) \tag{15}$$

where the density $f_v(\cdot)$ is given by Equation (1).

4.2 Time series models with skewed location-scale densities

We assume that the distribution of the response Y_t conditional on the history F_{t-1} of the process is such that

$$(Y_t|\mathcal{F}_{t-1}) \sim \text{SLS}(h, \gamma_t, \delta_t, \xi_t)$$

with conditional parameters γ_t, δ_t and ξ_t.

4.2.1 ARMA model

Let us assume that both the dispersion and the skewness parameters are constant, i.e.

$$\delta_t = \delta \ ; \ \xi_t = \xi$$

We would like to propose a specification of the same type as the ARMA model of Equation (11) relating the conditional mode (instead of the possibly non existing mean) to the history of the process. In Equation (11), we see that past observed residuals

$$e_j = y_j - \mu_j \ (j < t)$$

are used to measure prediction errors before time t. The choice of this error measure is motivated by

$$\text{E}(\epsilon_{t-j} = Y_{t-j} - \mu_{t-j}|\mathcal{F}_{t-j-1}) = 0$$

provided that our conditional distribution (which is the predicting distribution) is indeed governing the observed process. Measuring a prediction error by comparing the observed y_j to the (predictive) conditional mean μ_j is of course arbitrary. It would be equally reasonable to predict Y_j using the conditional mode γ_j of the conditional distribution. This is precisely what we propose here.

Therefore, we propose an ARMA model for the conditional mode γ_t

$$\gamma_t = \psi_{1t} + \sum_{i=1}^{p_1} \phi_{1i}(y_{t-i} - \psi_{1,t-i}) + \sum_{j=1}^{q_1} \theta_{1j}(y_{t-j} - \gamma_{t-j}) \tag{16}$$

with the usual restrictions on the roots of the corresponding $\Phi_1(L)$ and $\Theta_1(L)$ polynomials. This is equivalent to assuming that

$$\Phi_1(L)(Y_t - \psi_{1t}) = \Theta_1(L)\epsilon_t \tag{17}$$

where $\{\epsilon_t\}$ is i.i.d. with $\mathrm{SLS}(h, 0, \delta, \xi)$ distribution.

Note that starting by assuming that Y_t is governed by Equation (17) is equally arbitrary to assuming directly that

$$(Y_t|\mathcal{F}_{t-1}) \sim \mathrm{SLS}(h, \gamma_t, \delta, \xi)$$

with γ_t defined by Equation (16).

Consider for simplicity an ARMA(p_1, q_1) for the conditional mode. We could say that the process $\{Y_t - \psi_{1t}\}$ is asymptotically mode stationary when the sequence

$$\mathrm{Mode}(Y_t - \psi_{1t})$$

converges to a m that does not depend on t. If such a m exists, it is the solution of

$$m = \sum_{i=1}^{p_1} \phi_{1i} m + \sum_{j=1}^{q_1} \theta_{1j}[m - m]$$

as obtained by applying the Mode$(\cdot)$ operator to the ARMA(p_1, q_1) equation

$$(\gamma_t - \psi_{1t}) = \sum_{i=1}^{p_1} \phi_{1i}(y_{t-i} - \psi_{1,t-i}) + \sum_{j=1}^{q_1} \theta_{1j}[(y_{t-j} - \psi_{1,t-j}) - (\gamma_{t-j} - \psi_{1,t-j})]$$

As we have

$$\sum_{i=1}^{p_1} \phi_{1i} < 1, \text{ [6]}$$

from the unit root condition on $\Phi_1(L)$, it requires m to be zero. Hence

$$\mathrm{Mode}(Y_t) = \psi_{1t}$$

We could reverse the above motivating argument for defining the residual from the mode and wonder to what function $x(y_j)$ we should confront γ_j to measure (using a single number) the adequacy of the location of the conditional distribution. The "ideal" choice for γ_j could be defined to be the value x that maximizes the conditional probability to measure y_j for the process, which is equivalent to finding the value of x that maximizes the density $f(y_j|x, \delta_j, \xi_j)$ in Equation (7). This yields

$$x(y_j) = y_j$$

Therefore, our prediction error at time j could be defined as

$$x(y_j) - \gamma_j = y_j - \gamma_j$$

i.e. the difference between the "ideal" γ_j and its actually considered value.

Note that Equation (16) can be used in particular with the skewed Student and the skewed stable distributions to define the conditional mode.

[6] This is the only required condition to have asymptotical mode stationarity.

4.2.2 APARCH model

We would like to define the conditional dispersion parameter δ_t (measuring the inverse curvature of the conditional density at the mode) using a formula similar to the definition of the APARCH conditional variance of Equation (14), assuming $\tau_i = 0 \;\; \forall i$ for simplicity.

If we substitute δ_j to σ_j in that equation, by what function $d(y_j)$ should we replace the absolute (conditional) residual $|e_j| = |y_j - \mu_j|$?

Using the same argumentation as in the previous section, we want to know to what function $d(y_j)$ we should confront δ_j to measure (using a single number) the adequacy of the dispersion of the conditional distribution. The "ideal" choice for δ_j could be defined to be the value d that maximizes the conditional probability to measure y_j for the process, which is equivalent to finding the value of d that maximizes the density $f(y_j|\gamma_j, d, \xi_j)$ in Equation (7). Solving

$$\frac{\partial f(y_j|\gamma_j, d, \xi_j)}{\partial d} = 0$$

for d, we get

$$d_j = d(y_j) = \begin{cases} \frac{|y_j - \gamma_j|}{z_0 \xi_j^{-1}} & \text{if} \;\; y_j < \gamma_j \\ \frac{|y_j - \gamma_j|}{z_0 \xi_j} & \text{if} \;\; y_j \geq \gamma_j \end{cases}$$

where $z_0 = z_0(h)$ is the solution of

$$-\frac{h'(z)}{h(z)} = \frac{1}{z}$$

This suggests the following APARCH model for the conditional dispersion parameter δ_t

$$\delta_t^\zeta = \psi_{2t} + \sum_{i=1}^{p_2} \phi_{2i} \left\{ |d_{t-i}|^\zeta - \psi_{2,t-i} \right\} + \sum_{j=1}^{q_2} \theta_{2j} \left\{ |d_{t-j}|^\zeta - \delta_{t-j}^\zeta \right\} \tag{18}$$

Special cases

1. Normal ARMA-APARCH: we have

$$h(z) = \frac{1}{\sqrt{2\pi}} \exp(-0.5 \; z^2)$$

Thus, z_0 is the solution of

$$-\frac{h'(z)}{h(z)} = \frac{1}{z} \iff z = \frac{1}{z}$$

and

$$|d_t|^\zeta = |y_j - \gamma_j|^\zeta.$$

Notice that if $\zeta = 2$, as in the usual GARCH specification, we recover the squared residual.

2. Skewed Student ARMA-APARCH: we have seen that

$$h(z) = h_\upsilon(z) = \frac{\Gamma(\frac{\upsilon+1}{2})}{\sqrt{\pi(\upsilon-2)}\;\Gamma(\frac{\upsilon}{2})} \; \frac{1}{\left(1 + \frac{1}{\upsilon-2} \; z^2\right)^{\frac{\upsilon+1}{2}}}$$

Thus, z_0 is the solution of

$$-\frac{h'(z)}{h(z)} = \frac{1}{z} \iff \frac{(\upsilon+1)z}{(\upsilon-2)+z^2} = \frac{1}{z} \iff z^2 = \frac{\upsilon-2}{\upsilon}$$

Hence, we have

$$|d_t|^\zeta = \begin{cases} \left(\frac{\upsilon}{\upsilon-2} \frac{|y_j-\gamma_j|}{\xi_j^{-1}}\right)^\zeta & \text{if} \;\; y_j < \gamma_j \\ \left(\frac{\upsilon}{\upsilon-2} \frac{|y_j-\gamma_j|}{\xi_j}\right)^\zeta & \text{if} \;\; y_j \geq \gamma_j \end{cases}$$

Note that when we have symmetry ($\xi_j = 1$), we only recover the ζ^{th} power absolute residual when the degrees of freedom $\upsilon \to +\infty$, i.e. in the normal case.

3. Skewed stable ARMA-APARCH: consider for $h(z)$ the density of the symmetric stable distribution with $\gamma = 0, \delta = 1, \beta = 0$ and arbitrary α. The main problem with the stable distribution is the non availability of an analytic form for its density, except in special cases. Hence, except for particular values of (β, α), we must solve numerically

$$-\frac{h'(z)}{h(z)} = \frac{1}{z}$$

for z to determine $z_0 = z_0(h)$. The value of z_0 for varying values of α is given on Figure 2, completing the definition of the ζ^{th} power absolute residual

$$|d_j|^\zeta = \begin{cases} \left(\frac{|y_j - \gamma_j|}{z_0 \xi_j^{-1}}\right)^\zeta & \text{if} \quad y_j < \gamma_j \\ \left(\frac{|y_j - \gamma_j|}{z_0 \xi_j}\right)^\zeta & \text{if} \quad y_j \geq \gamma_j \end{cases}$$

It can be used with Equation (18) to define the dispersion of the response conditional distribution. Further imposing ζ to be less than α is required to have finite expectation for $|d_j|^\zeta$.

Note that when $\alpha = 2$ and $\xi = 1$, we recover, for the skewed stable, the normal distribution with variance $\sigma^2 = 2\delta^2$. Thus, the obtained $z_0^2 = 2$ (see Figure 2) is consistent with the first special case.

5 Skewness dynamics in skewed distributions

The models presented up till now assume that the skewness parameter remains constant over time. This hypothesis might not be reasonable in some situations and alternative models allowing this characteristic of the response distribution to change with time should be available. Hansen (1994) proposes to condition the skewness (and the kurtosis) on past residuals and squared residuals. The same idea has been used by Campbell and Siddique (1999)[7] and Jondeau and Rockinger (2000)[8]. Here, we show how dynamics can be introduced in the skewness parameter (defined as the odds ratio of having an observation above the mode) using time varying covariates and ARMA like models.

5.1 Basic dynamic model

The starting point is Equation (3) for an arbitrary conditional mode γ_t. Denoting by π_t the probability to observe for the response at time t a larger value than the conditional mode γ_t, we have

$$\log(\xi_t^2) \quad = \quad \text{logit}(\pi_t)$$

If the skewness is assumed constant, then a reasonable estimator for the (logarithm of the) conditional skewness parameter is:[9]

$$\log(\hat{\xi}_t^2) \quad = \quad \text{logit}\left(\frac{n_{t-1}^{>}}{n_{t-1}^{>} + n_{t-1}^{<}}\right) = \log\left(\frac{n_{t-1}^{>}}{n_{t-1}^{<}}\right)$$

[7] Campbell and Siddique (1999) use a noncentral-t distribution, scaled to have a unit variance and condition the third moment on past residuals to the power one and three. These authors also investigate the existence of coskewness using a bivariate GARCH on daily and monthly stock indexes.

[8] Jondeau and Rockinger (2000) express skewness and kurtosis of Hansen's GARCH model as a function of the underlying parameters. The cost of such a flexibility is that for a dataset of about 7,000 observations, they have to impose not less that 20,000 restrictions. This difficult estimation problem is solved using a recent sophisticated sequential quadratic optimization algorithm.

[9] Working with $\log(\hat{\xi}_t^2)$ instead of $\hat{\xi}_t^2$ avoids to worry about the positiveness of $\hat{\xi}_t^2$.

where ξ_t^2 is the conditional odds to have observation at time t above the mode, $n_{t-1}^>$ and $n_{t-1}^<$ denote respectively the number of times an observation has been observed above and below the corresponding (predictive) conditional mode up to and including time $t-1$, i.e.

$$\begin{array}{ll} n_t^> = n_{t-1}^> + I(y_t > \gamma_t) & \\ n_t^< = n_{t-1}^< + I(y_t < \gamma_t) & \end{array} \qquad (t \quad \geq \quad 1) \tag{19}$$

with (say)

$$n_0^> \quad = \quad n_0^< = 1$$

to start the recursion and $I(y_t > \gamma_t)$ (resp. $I(y_t < \gamma_t)$) being one when $y_t > \gamma_t$ (resp. $y_t < \gamma_t$) and 0 otherwise.

Note that an observation larger than the mode at some time t_1 contributes in the same way to $n_t^>$ than an observation above the mode at time t_2 with $t_1 < t_2 < t$. Thus the information brought by an observation on skewness does not get older with time. This is the translation of the constant skewness hypothesis made above.

This hypothesis can be relaxed and the skewness allowed to change with time by generalizing Equation (19) to

$$\begin{array}{l} n_t^>(\rho) = \rho\; n_{t-1}^>(\rho) + I(y_t > \gamma_t) \\ n_t^<(\rho) = \rho\; n_{t-1}^<(\rho) + I(y_t < \gamma_t) \end{array} \qquad (t \quad \geq \quad 1) \tag{20}$$

with $0 < \rho \leq 1$. Values of ρ close to 1 yield slowly changing empirical measures of skewness whereas smaller values yield more dynamic estimates. In practice, we propose to estimate ρ using its MLE.

5.2 Use of covariates to model skewness

Constant and time varying covariates can also be included to model skewness and change in skewness. If ψ_{3t} denotes the logarithm of the resulting unconditional skewness possibly defined as a function of covariates, then

$$\log(\xi_t^2) \quad = \quad \psi_{3t} + \phi_{31} \left\{ \log\left(\frac{n_{t-1}^>(\rho)}{n_{t-1}^<(\rho)} \right) - \psi_{3,t-1} \right\} \tag{21}$$

is a possible extension of the basic dynamic model defined above where ξ_t^2 is the conditional skewness parameter. When the considered covariates are constant over time, Equation (21) can be rewritten as

$$\log(\xi_t^2) \quad = \quad (1-\phi_{31})\psi_{30} + \phi_{31} \log\left(\frac{n_{t-1}^>(\rho)}{n_{t-1}^<(\rho)} \right)$$

In that particular case, we see that ϕ_{31} weights the contribution of exogenous information (summarized in ψ_{30}) and empirical information on skewness. A small value for ϕ_{31} indicates that covariates are more predictive of skewness at time t than its empirical conditional estimate.

5.3 General dynamic model for skewness

Using the same types of arguments as in the ARMA model, we could define a general dynamic model for skewness, GDMS(p_3, q_3), by replacing μ_t and y_t in Equation (11) by $\log(\xi_t^2)$ and $\log\left(\frac{n_t^>(\rho)}{n_t^<(\rho)}\right)$ respectively, yielding

$$\log(\xi_t^2) = \psi_{3t} + \sum_{i=1}^{p_3} \phi_{3i} \left\{ \log\left(\frac{n_{t-i}^>(\rho)}{n_{t-i}^<(\rho)} \right) - \psi_{3,t-i} \right\} + \sum_{j=1}^{q_3} \theta_{3j} \left\{ \log\left(\frac{n_{t-j}^>(\rho)}{n_{t-j}^<(\rho)} \right) - \log(\xi_{t-j}^2) \right\}$$

This extension is suggested by the clear parallel that can be drawn between moving average models for the conditional mean and Equation (21) where the roles of the conditional mean and innovation are played by $\log(\xi_t^2)$ and $\log\left(\frac{n_{t-1}^{>}(\rho)}{n_{t-1}^{<}(\rho)}\right) - \psi_{3,t-1}$ respectively. Although theoretically appealing, it is likely that only very simple versions of that model will be relevant in practice.

6 Application

The analyzed dataset consists of 4313 observations of the DEM-USD exchange rate returns from January 1980 till December 1996 (on a daily basis). Exchange rate returns have been analysed by multiple authors using mainly the normal and Student ARMA-GARCH model and normal finite mixture models. Recently, Beine, Laurent, and Lecourt (2000) showed (using the same dataset) that fractional differencing in the variance in combination with a day of the week effect and Student errors can improve the fit.

Here, we propose to analyse that long time series using the skewed Student (cf. Section 2.2) and the skewed stable (cf. Section 2.4). Dynamics will be introduced in the location and the dispersion parameters using the models of Section 4. A time varying skewness will also be allowed using the GDMS model of Section 5.3.

Whatever the considered distribution, both likelihood ratio tests and information criterions select the AR(1)-APARCH(1,1) in the ARMA-APARCH family of models. As mentioned in Section 3.3, no asymmetry has been found in the conditional second moment, i.e. $\tau_1 = 0$.

The MLEs (standard errors or s.e.'s) of the parameters in the AR(1)-APARCH(1,1)-GDMS(1,0) models for the above two families of distributions can be found in Table 1.

The two models in Table 1 have the same number of parameters. A better fit is obtained with the skewed Student distribution with a minus log-likelihood of 4546.0 against 4556.0 for the skewed stable.

The location parameters for the skewed Student and the skewed stable distributions are very similar. This is not surprising as, in both cases, we are dealing with the mode of the distribution. The autoregressive parameter is negative for the two distributions, indicating that a positive (negative) return tends to be followed by a negative (positive) on the next day. This is illustrated in the first part of Figures 3 and 4 where the predicted (one-step ahead) mode γ_t is plotted for the year 1987. Note that no difference can be visually detected when one superposes the skewed Student and the skewed stable predictions for the mode.

The dispersion parameters of the skewed Student and of the skewed stable distributions are very similar. This is not surprising as they measure the inverse curvature of their respective density at the mode. Note also that residuals to the power 1.5 were chosen instead of the "traditional" squared residuals in GARCH models. The second parts of Figures 3 and 4 give the predicted (one-step ahead) dispersion parameters σ_t and δ_t (respectively) for the year 1987. We clearly see the influence of the October crash around day 200 where a brutal (predicted) volatility surge is observed. Note that volatility was even larger in the mid-January–February period.

The unconditional skewness parameter $\log(\xi^2)$ cannot be assumed to be zero as shown by ψ_{30} and its standard error. Thus, "on average", there is evidence for mild negative skewness. However, the significantly non zero values for ϕ_{31} and ρ show that skewness is changing over time. We conclude that skewness alternates positive and negative periods. The third, fourth and fifth parts of Figures 3 and 4 give the predicted (one-step ahead) skewness parameter $\log(\xi_t)$, the corresponding (predictive) probability π_t to be over the mode and the predictive probability to have a positive return for the year 1987. Like for the dispersion, we clearly see the influence of the October crash around day 200 which is followed by a period of negative skewness. Note that it was preceded in early September by a (relatively) largely negative predicted skewness. Apart from the mid-June–mid-August period, negative skewness was usually expected, particularly in January.

Note that, here, the conclusions drawn from the fitted models are the same for the skewed Student and the skewed stable distributions. This was confirmed by a plot of the predictive densities for the year 1987. As nearly no difference was found out between the obtained plots for the two considered distributions, we only reproduce the graph related to the skewed Student in Figure 5. The solid line gives the predicted modes γ_t throughout 1987, while the dotted lines give the predicted qth (with $q \in \{.99, .95, .75, 25, .05, .01\}$) quantiles (computed using the expressions in Section 2.6) of the predictive skewed Student density $g_{\hat{\upsilon}}(y_t|\hat{\gamma}_t, \hat{\delta}_t, \hat{\xi}_t)$. We clearly see the large uncertainty associated to the predictions in January-February and during the weeks following the October crash.

7 Discussion

We have shown how the ARMA-APARCH model could be viewed as an explicit dynamic specification of the conditional mean and of the conditional variance. These specifications were proposed for the location and for the dispersion parameters of skewed distributions by considering y_t and $z_0^{-1}\xi_t^{-\text{sign}(y_t-\gamma_t)}|y_t - \gamma_t|$ as the empirical counterparts of the mode and of the dispersion parameters respectively.

Note that, as the skewed stable distribution has only moments of order less that α defined (except in the normal distribution case $\alpha = 2$ where moments of all orders are defined), it was not possible to reparameterize it in terms of the first two moments. However, it is possible to do so with the skewed Student distribution and thus to consider the traditional ARMA-APARCH model for the mean and for the variance directly for that distribution choice (this is however beyond the scope of the paper). But, then, comparing the results provided by the reparameterized skewed Student and the skewed stable (for which no variance is available) distributions would have been difficult.

We have also proposed a general dynamic model for the skewness parameter. The ARMA-APARCH-GDMS model was applied to the analysis of the DEM-USD exchange rates over the 1980-1996 period. We were able to show that periods of light and moderate (usually negative) skewness alternate. The definition of $\log(\xi^2)$ as the log-odds ratio of a return larger than the mode was found very useful to interpret not only the sign but also the value of the predicted skewness parameter. Thus, in addition to a risk measure as provided by the conditional variance, we are now able to give a probability to the sign of that risk.

Reporting the fitted model by a plot of several conditional quantiles over the targeted period was found very useful both to understand and to communicate the obtained results. It also puts in context the difference in log-likelihood between the skewed Student and the skewed stable based models. As mentioned earlier, no important difference was found between the plots of their conditional quantiles. But note that the difference between these two distributions must be looked for in the tails. These differences are expected as their Pareto indices are respectively estimated to be $\upsilon = 7.19$ and $\alpha = 1.87$ with no overlapping asymptotic 95% confidence intervals (cf. estimate $\pm$ 1.96 s.e.). Virtually no difference was found between the plotted conditional quantiles for the skewed Student (see Figure 5) and the skewed stable distributions, even for the 1% and 99% quantiles. This is illustrated on the upper part of Figure 6 where the predictive probabilities to observe a return larger than x (with $x \in [0, 3]$) on January 1st, 1987 under the fitted skewed Student and the fitted skewed stable are plotted. These probabilities are nearly equal. But when we compute the ratio of these probabilities (as shown on the lower part of Figure 6), we see that a large positive return of 3% (say) is about 3.5 times more likely under the skewed stable than under the skewed Student. However the upper part of the graph clearly shows that such a large return is a rare event, even under the skewed stable distribution. Thus, we should be very careful when inferring on the probabilities of extreme events as models yielding comparable fits can provide (relatively) very different values for these (extremely small) probabilities.

Acknowledgment

The authors would like to thank Jean-Pierre Urbain for useful remarks and suggestions.

References

Adler, J., R. Feldman, and M. Taqqu (1998): *A Practical Guide to Heavy Tails.* Birkhäuser, Basel.

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

Beine, M., and S. Laurent (1999): "Central bank interventions and jumps in double long memory models of Daily Exchange Rates," Mimeo, University of Liège.

Beine, M., S. Laurent, and C. Lecourt (2000): "Accounting for Conditional Leptokurtosis and Closing Days Effects in FIGARCH Models of Daily Exchange Rates," *Forthcoming in Applied Financial Economics.*

Bidarkota, P., and J. McCulloch (1998): "Optimal univariate inflation forecasting with symmetric stable shocks," *Journal of Applied Econometrics*, 13, 659–670.

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 condtional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

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

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

Campbell, R., and A. Siddique (1999): "Autoregressive conditional skewness," *Forthcoming in Journal of Financial and Quantitative Analysis.*

Chung, C.-F. (1999): "Estimating the fractionnally intergrated GARCH model," National Taïwan University working paper.

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.

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.

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.

Engle, R., V. K. Ng, and M. Rothschild (1990): "Asset pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills," *Journal of Econometrics*, 45, 213–238.

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

GAWRONSKI, W. (1984): "On the bell-shape of stable densities," *Annals of Probability*, 12, 230–242.

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

GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1993): "On the relation between expected return on stocks," *Journal of Finance*, 48, 1779–1801.

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

HE, C., AND T. TERASVIRTA (1999a): "Higher-order dependence in the general Power ARCH process and a special case," Stockholm School of Economics, Working Paper Series in Economics and Finance, No. 315.

——— (1999b): *Statistical properties of the asymmetric power ARCH process*chap. 19, pp. 462–474, Cointegration, causality, and forecasting. Festschrift in honour of Clive W.J. Granger. in Engle, Robert F. and Halbert White, oxford university press edn.

HOFFMANN-JØRGENSEN, J. (1994): *Probability with a View Toward Statistics*. Chapman and Hall, London.

HOSKING, J. (1981): "Fractional differencing," *Biometrika*, 68, 165–176.

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.

LAMBERT, P., AND J. LINDSEY (1999): "Analysing financial returns using regression models based on non-symmetric stable distributions," *Applied Statistics*, 48, 409–424.

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

MCCULLOCH, J. (1996): *Financial Applications of Stable Distributions*chap. 14, Statistical Methods in Finance (Handbook of Statistics). in G.S. Maddala and C.R. Rao, elsevier edn.

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.

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

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.

SAMORODNITSKY, G., AND M. TAQQU (1994): *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, London.

TAYLOR, S. (1986): *Modelling financial time series*. Wiley, New York.

TSE, Y. (1998): "The conditional heteroscedasticity of the yen-dollar exchange rate," *Journal of Applied Econometrics*, 193, 49–55.

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.

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

Yamazato, M. (1978): "Unimodality of infinitely divisible distribution functions of class l," *Annals of Probability*, 6, 523–531.

Figure 1: Skewness (measured by $\log \xi^2$) as a function of skewness and tail parameters β and α.

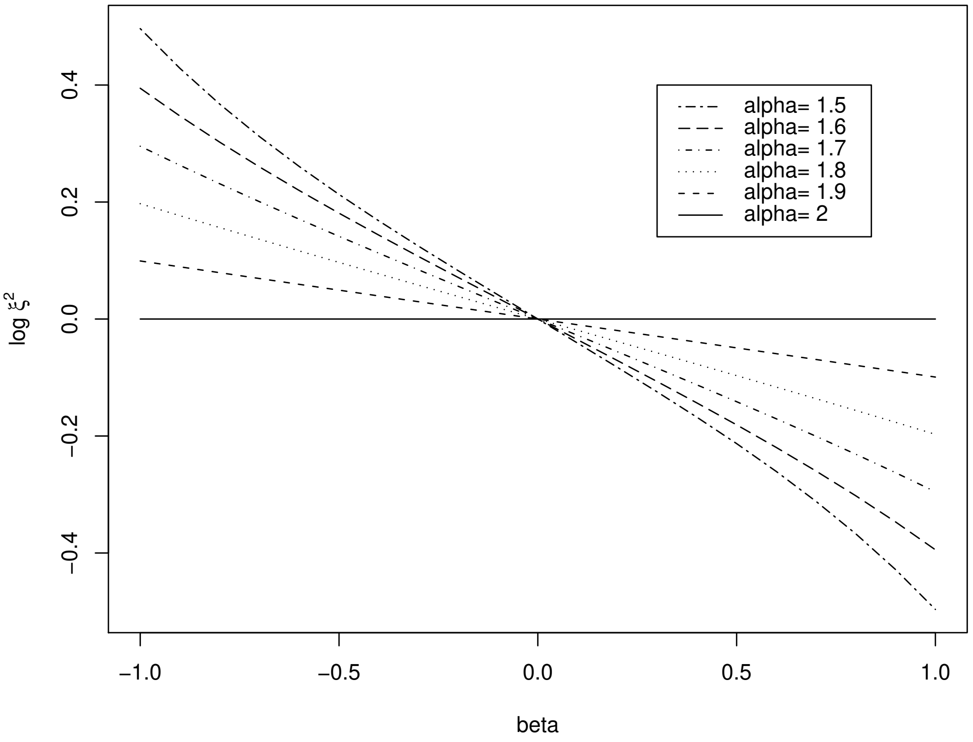


Figure 2: Values of $z_0(\alpha)$ for values of α in $(1, 2)$.

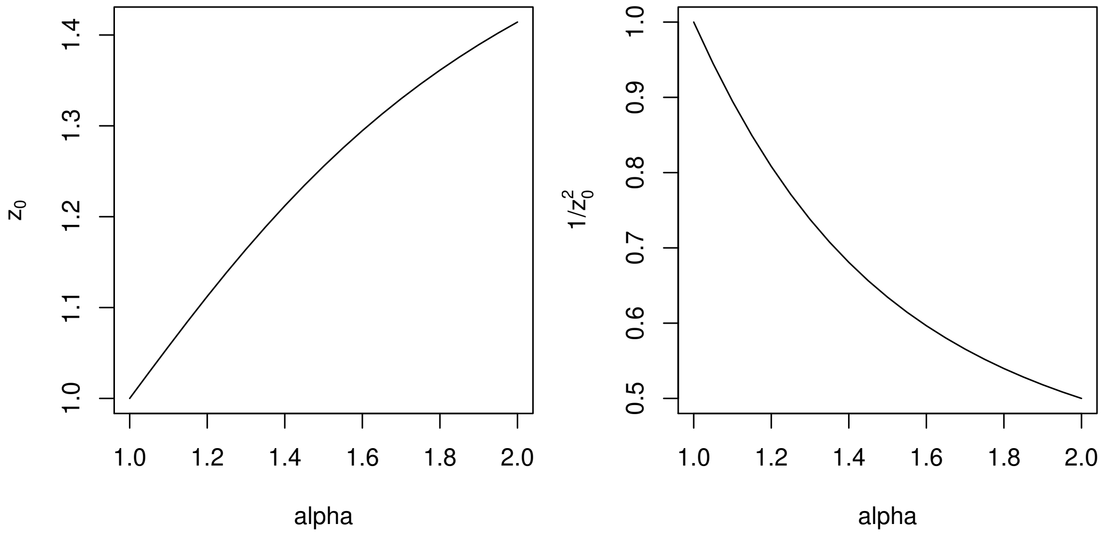


Figure 3: Predicted (one-step ahead) mode, dispersion and log skewness parameters for the ARMA-APARCH-GDMS model for the skewed Student distribution in 1987.

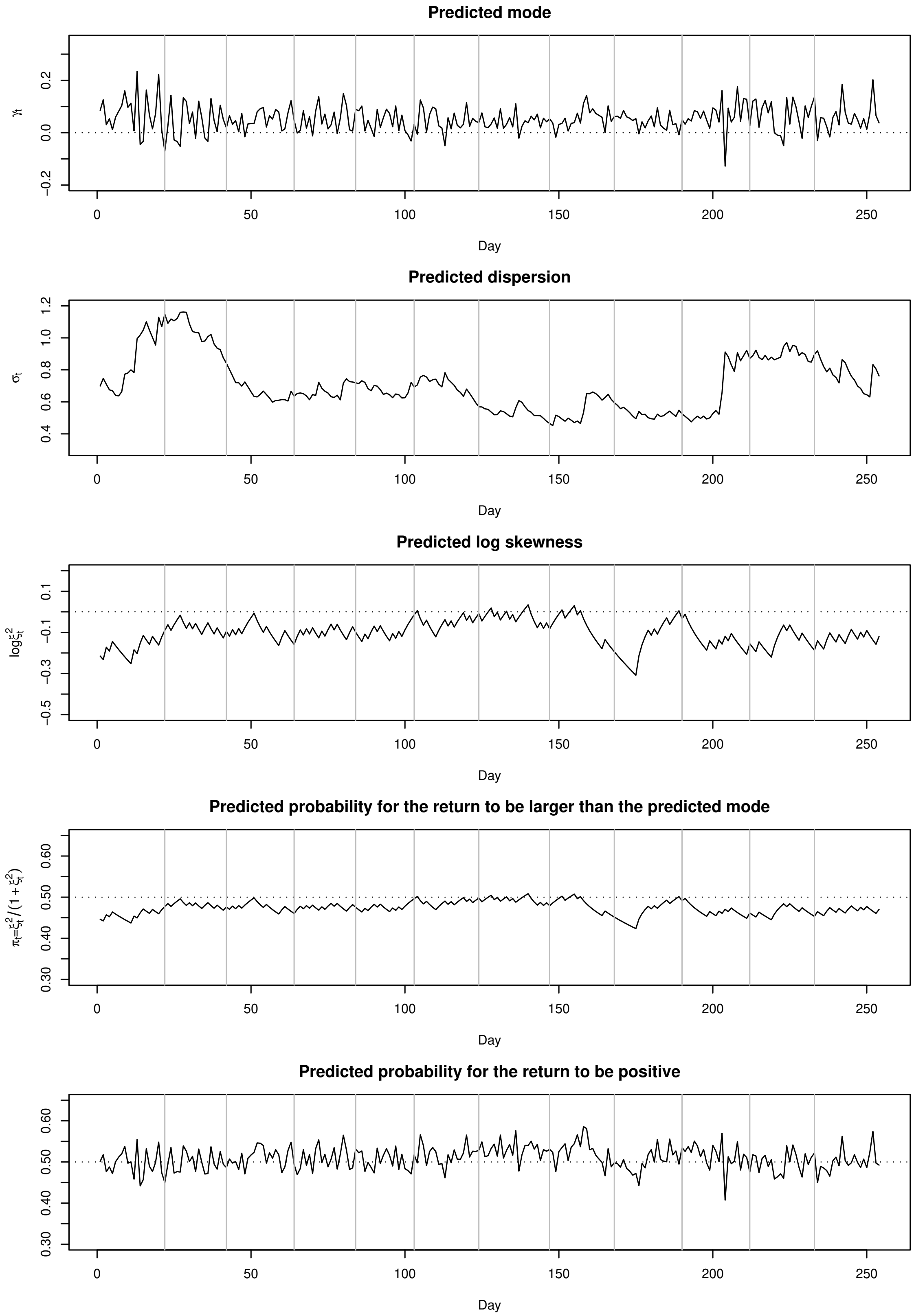
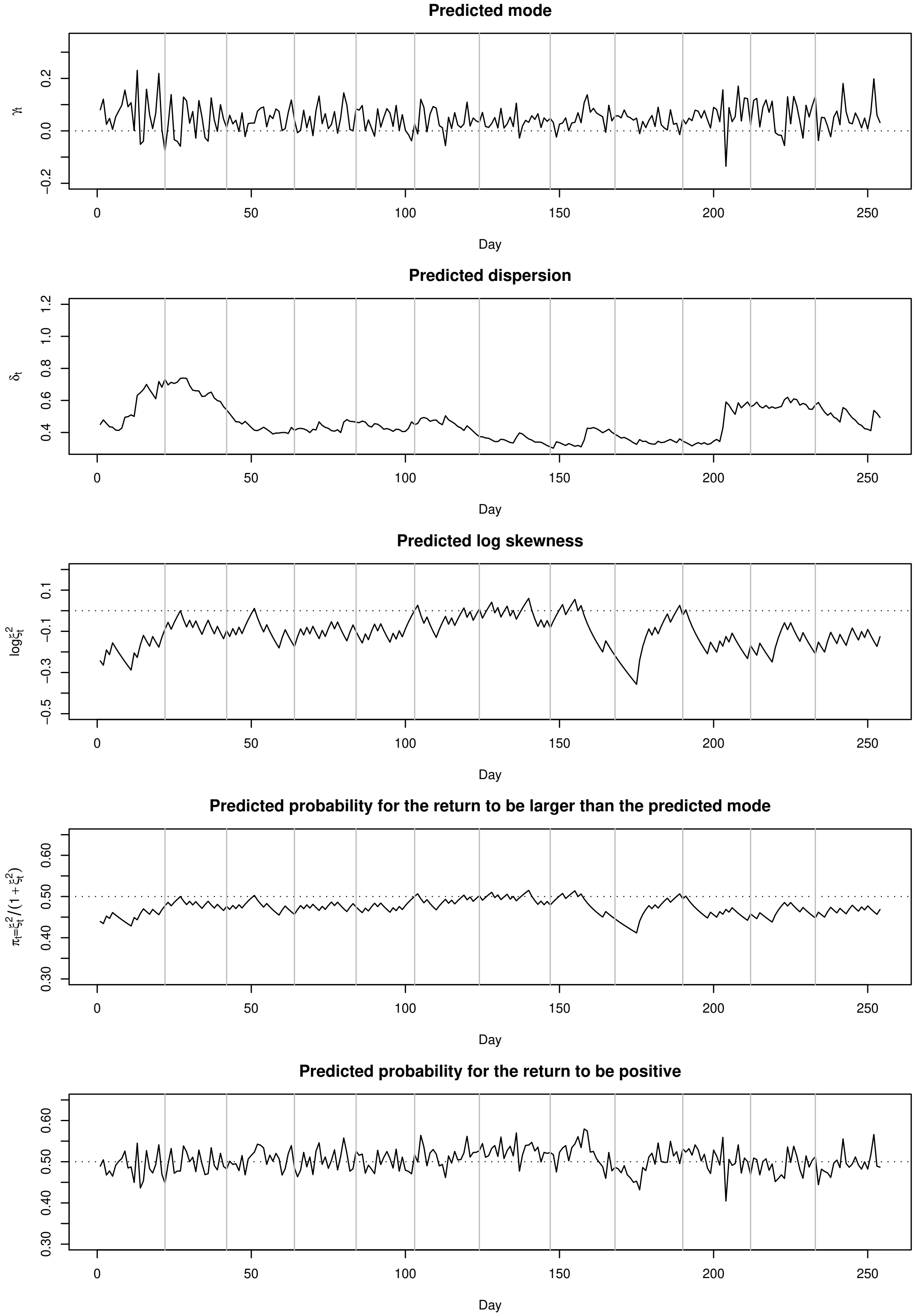


Figure 4: Predicted (one-step ahead) mode, dispersion and log skewness parameters for the ARMA-APARCH-GDMS model for the skewed stable distribution in 1987.



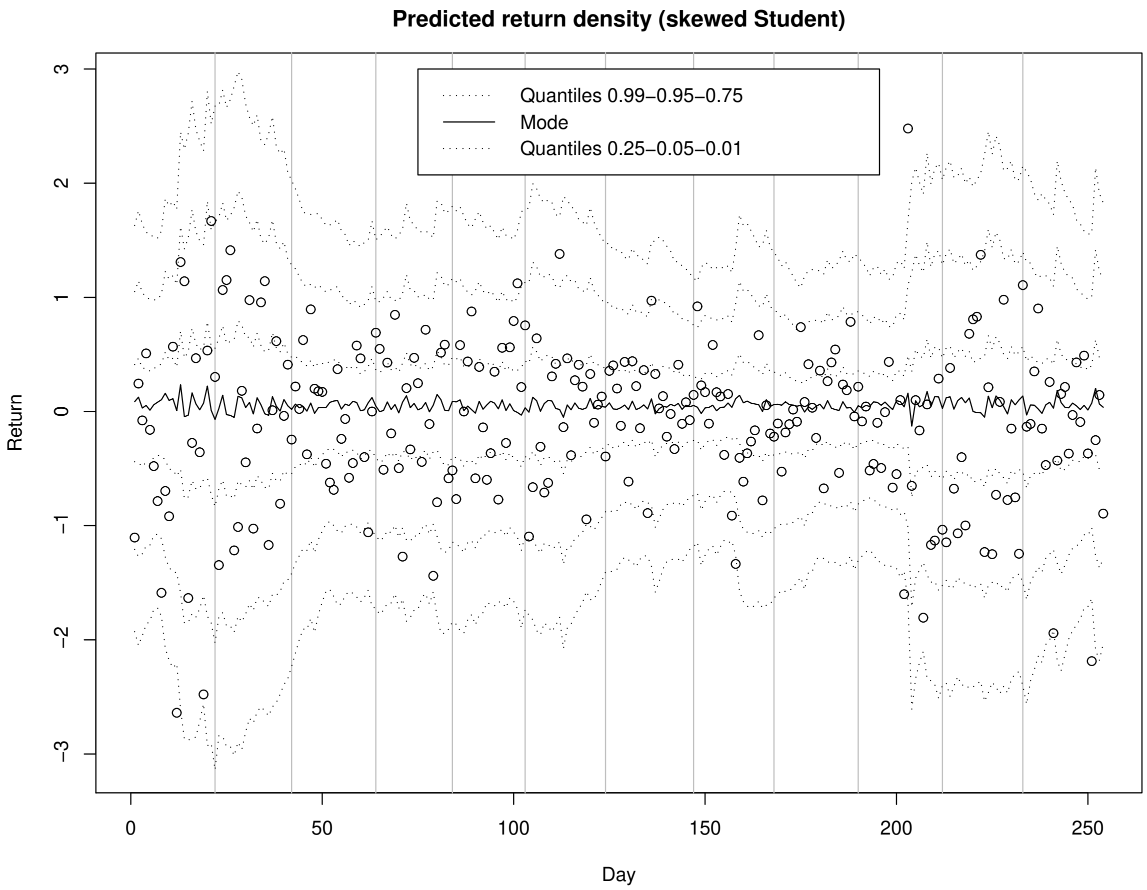


Figure 5: Predicted (one-step ahead) density for the 1987 returns using the ARMA-APARCH-GDMS model with the skewed Student distribution. The solid line gives the predicted modes while the dotted lines give the 0.99, 0.95, 0.75, 0.25, 0.05 and 0.01 quantiles.

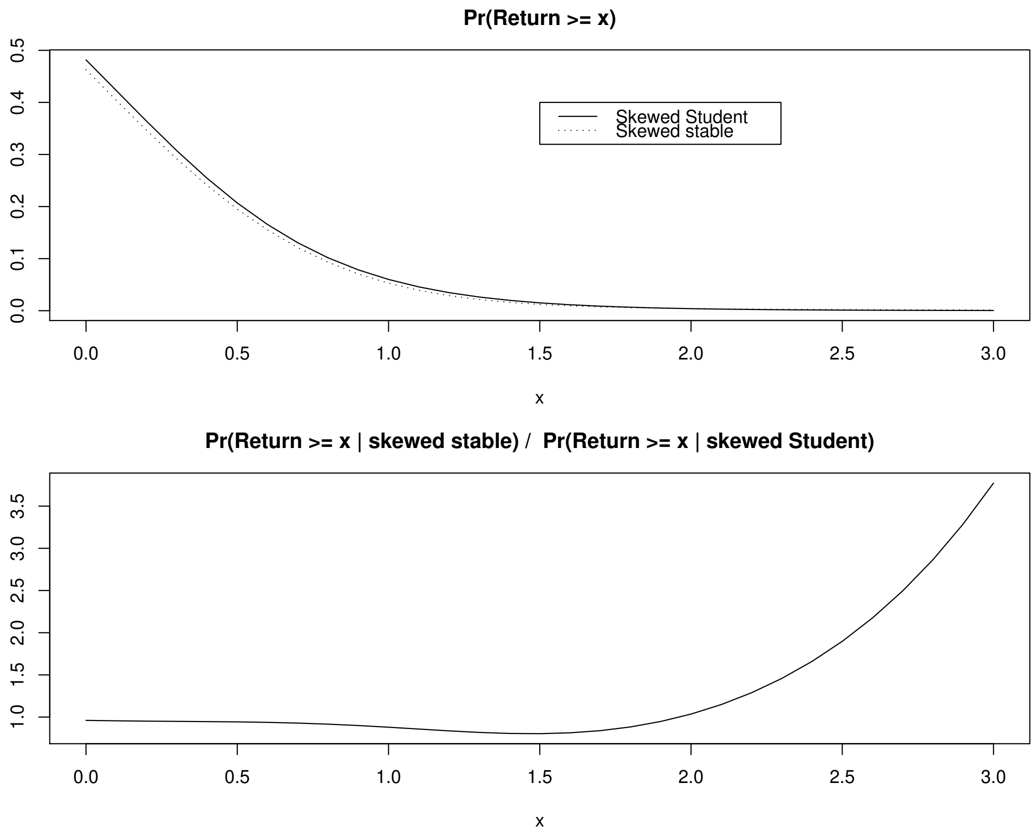


Figure 6: Graph 1: conditional probabilities to observe a return larger than x under the predictive skewed Student (solid line) and predictive the skewed stable (dotted line) on January 1st, 1987. Graph 2: ratio of the conditional probabilities to observe a return larger than x under the predictive skewed Student (solid line) and the predictive skewed stable (dotted line) on the January 1st, 1987.

	Skewed Student	Skewed stable
	Location	
ψ_{10}	0.044 (0.009)	0.046 (0.010)
ϕ_{11}	-0.071 (0.015)	-0.069 (0.015)
	Dispersion	
ψ_{20}	0.057 (0.130)	0.345 (0.103)
ϕ_{21}	0.968 (0.008)	0.970 (0.007)
θ_{21}	-0.893 (0.012)	-0.893 (0.012)
ζ	1.480 (0.178)	1.519 (0.183)
	Skewness	
ρ	0.874 (0.061)	0.873 (0.056)
ψ_{30}	-0.086 (0.022)	-0.086 (0.022)
ϕ_{31}	0.102 (0.044)	0.123 (0.047)
	Tail	
υ/α	7.19 (0.64)	1.87 (0.018)
Par.	10	10
$-\log(L)$	4546.0	4556.0

Table 1: Estimates (standard errors) of the parameters in AR(1)-APARCH(1,1)-GDMS(1,0) models with skewed Student and skewed stable distributions.

$$
\begin{aligned}
\gamma_t &= \psi_{10} + \phi_{11}\,(y_{t-1} - \psi_{10}) \\
\delta_t^{\zeta} &= \psi_{20} + \phi_{21}\,\left\{|d_{t-1}|^{\zeta} - \psi_{20}\right\} + \theta_{21}\left\{|d_{t-1}|^{\zeta} - \delta_{t-1}^{\zeta}\right\} \\
n_t^{>} &= \rho\, n_{t-1}^{>} + I(y_t > \gamma_t) \;\;;\;\; n_t^{<} = \rho\, n_{t-1}^{<} + I(y_t < \gamma_t) \\
\log(\xi_t^2) &= \psi_{30} + \phi_{31}\left\{\log\left(\frac{n_{t-1}^{>}(\rho)}{n_{t-1}^{<}(\rho)}\right) - \psi_{30}\right\} \\
\tau_1 &= 0
\end{aligned}
$$