Conditional risk measurement using conditional higher order moments: a dynamic application of the skew $t$-distribution

Malcolm J Faddy and Rodney C Wolff
School of Mathematical Sciences, Queensland University of Technology,
GPO Box 2434, BRISBANE QLD 4001, AUSTRALIA

December 12, 2008

Abstract

This paper presents a parsimonious approach to estimating conditional skewness and kurtosis, as well as conditional variance, in financial log-returns time series. Using a GARCH formulation of the skew $t$-distribution (Jones and Faddy, 2003), autoregressive relationships are developed for the conditional skewness and conditional kurtosis. A numerical example indicates that allowing for estimates of conditional skewness and kurtosis can improve the estimate of the conditional volatility, as well as provide quantitative indicators of epochs of high risk by identification of increased conditional skewness or kurtosis. The model enables Value-at-Risk to be estimated more conservatively by weakening the requirements of both Normality and independence, thus capturing time-varying risk from more general non-linear financial time series with greater accuracy.

Keywords: Conditional kurtosis; Conditional skewness; GARCH; Numerical maximum likelihood; Skew $t$-distribution; Value-at-Risk.

1 Introduction

The notion of risk plays a major part in financial markets, and risk management is therefore a crucial task for financial institutions. Risk may be defined as the potential loss arising from a probabilistic event; consequently, accurate computational statistical methods have a central role to play in risk management. Understanding the volatility of risk factors (or variables) in a portfolio is important, as large volatility implies heavy risk (Markowitz, 1959). However, it is not only the magnitude of volatility which is important, but also asymmetries: with a proper grasp of such properties, risk
calculations can be sharpened. To that end, the concept of Value-at-Risk (VaR) (RiskMetrics, 1994) is another way of quantifying risk. VaR is a cross-sectional measure in practice, considering an extreme loss at a given probability at a given time; that is,

$$\text{VaR}_\pi = \max \left[ - \inf \{ z \mid P(Z \leq z) \geq \pi \} , 0 \right],$$

where $Z$ is the risk factor at the stated time horizon, and $\pi$ is the parameter called the probability level. Statistically, VaR is given by the quantile of the theoretical risk factor distribution. For market risk measurement purposes, $\pi = 0.01$ is the most commonly used value, usually over a ten-day horizon. As is custom, losses are expressed as negative profits. See Jorion (2001) and Duffie and Pan (1997) for detailed discussion.

Calculation of volatility or VaR as risk measures would be elementary if the sequence of values of the risk factor, as a time series, were strictly stationary. It is widely observed that financial risk factors seldom have this property, which gave rise to the now well-known conditional volatility models of Engle (1982) and, in a generalised form, of Bollerslev (1986). Numerous extensions to these ARCH/GARCH models have been proposed. The classical GARCH-type models imply that volatility is symmetric, but a so-called stylised fact of traditional asset returns is that negative shocks tend to reach deeper compared with positive shocks.

Despite sophisticated models for conditional volatility, the treatment of VaR has been much more simplistic, usually assuming that conditionally standardised risk factors are independent and follow a Normal distribution. There are notable exceptions, such as that of Engle and Manganelli (2004), who studied this using quantile regression. Nevertheless, given another stylised fact, namely that returns are much more heavy-tailed than the Normal distribution, it is clear that using Normal-based error models has the potential to underestimate risk severely.

Among many parametric approaches, the Student $t$-distribution is a natural competitor of the Normal in this regard: firstly, it embraces the Normal distribution as a special case, when the degrees of freedom tend to infinity; and secondly, it is capable of capturing tails heavier than the Normal through finite values of the degrees of freedom parameter. Jones and Faddy (2003) discussed a generalisation of the symmetric Student $t$ which is capable of exhibiting skewness, and they elicited its probabilistic
and statistical properties. In particular, relatively straightforward and parsimonious parameterisations of the skew $t$-distribution allow skewness and kurtosis to be obtained. This model will be described fully later in the paper.

In the present paper, the skew $t$ will provide the route to estimate conditional skewness and kurtosis, based on autoregressive GARCH-style expressions. Likewise, conditional quantiles (VaR) will also be obtained from this model. The aim is to demonstrate improvements the skew $t$-distribution can make in risk measurement (as far as volatility and VaR are concerned) over the Normal and symmetric $t$, and the ease with which conditional skewness and kurtosis may be obtained in a natural way from simple extensions of the model.

Of course, the present paper is not alone in this endeavour. A relatively early attempt to incorporate arbitrary, but fixed, kurtosis in distribution fitting was a semi-parametric approach by Steinwolf (1996), who used phase variation in a spectral approach to join two shifted sections of Gaussian laws. Guermat and Harris (2002) showed that an exponentially weighted moving average estimate of the variance function was a special case of exponentially weighted maximum likelihood, which permitted time-variation in skewness and kurtosis to be calculated. In terms of direct parametric models, Harvey and Siddique (1999, 2000) employed the non-central $t$-distribution in a GARCH model to describe conditional skewness in asset returns at several levels of frequency, and found that allowing for conditional skewness impacted on estimates of conditional volatility. They also estimated the risk premium for accepting asymmetry in the risk, showing it to be significant. León et al. (2005) attempted to simplify Harvey and Siddique’s approach, by using a Gram-Charlier expansion of the Normal density. Reservations over this kind of approach relate to oscillations in estimators which arise as artefacts of the truncation of the expansion, rather than from the data (Marumo and Wolff, 2007). Christoffersen et al. (2006) used the inverse Gaussian distribution to estimate conditional skewness to aid option pricing on financial assets. Under their parameterisation, the kurtosis is a scalar multiple of the square of the skewness, which represents a more limited parameterisation than that used here. Lanne and Saikkonen (2007) used the so-called $z$ distributions within GARCH-in-mean models: they can be represented as Normal variance-mean mixtures, with mixing distribution an infinite convolution of exponentials. They estimate conditional skewness directly as a function of conditional volatility. By contrast, a more flexible parameterisation is used in the present paper,
allowing conditional skewness and kurtosis to have autoregressive evolutions distinct from that of the conditional second moment. Jondeau and Rockinger (2003) also used a generalised $t$-distribution, but captured asymmetry by controlling for the sign of the asset return directly in the model.

This paper is organised as follows. In Section 2, a brief account of the skew $t$-distribution is given, and probabilistic and statistical properties relevant to the present problem are summarised. A general dynamic model is presented, which nests the Normal and symmetric $t$ as special cases. Estimation procedures are specified. In Section 3, data for a case study application are presented, and results are shown, particularly the effects on the conditional variance of successively estimating conditional skewness and kurtosis. Conclusions are given in Section 4.

2 Econometric model

In this section, a brief account is given of the skew $t$-distribution, followed by its adaptation as an econometric model.

2.1 Properties of the skew $t$-distribution

Jones and Faddy (2003) introduced a tractable skew $t$-distribution as a flexible heavy-tailed alternative model to the Normal distribution, which has robustness to outliers as well as exhibiting skewness. The probability density function for this distribution is given by

\[ f(x; a, b) = C_{a,b}^{-1} \left\{ 1 + \frac{x}{(a + b + x^2)^{1/2}} \right\}^{a+1/2} \left\{ 1 - \frac{x}{(a + b + x^2)^{1/2}} \right\}^{b+1/2}, \quad -\infty < x < \infty, \ a, b > 0, \]

where $C_{a,b} = 2^{a+b-1}B(a,b)(a + b)^{1/2}$ and $B(.,.)$ is the Beta function. One obtains the usual $t$-distribution on $2a$ degrees of freedom when $a = b$. The distribution is negatively or positively skewed when $a < b$ or $a > b$, respectively. A number of other statistical properties, such as relations with other classical distributions, its behaviour for various $a$ and $b$, moments, tail behaviour, and properties of the likelihood, are derived in their paper.
It can be shown that the mean of the skew $t$-distribution is given by

$$m = \frac{(a-b)\sqrt{(a+b)\Gamma(a-\frac{1}{2})\Gamma(b-\frac{1}{2})}}{2\Gamma(a)\Gamma(b)}$$

and variance

$$v = \frac{(a+b)(a-b)^2 + a-1 + b-1}{4(a-1)(b-1)} - m^2.$$ 

A four-parameter version, allowing for arbitrary location and scale, is $f\{(x-\mu)/\sigma; a,b\}$, which will have mean $\mu + m\sigma$ and variance $v\sigma^2$.

While the luxury of orthogonal reparameterisations of the skew $t$-likelihood do not appear to exist, following Prentice (1975) the reparameterisations $p = 2/(a+b)$ and $q = (a-b)/\{ab(a+b)\}^{1/2}$ are used. These are related to skewness and kurtosis, respectively, in as far as

$$\lim_{p,q \to 0} \frac{\partial \ell}{\partial q} = -\frac{5}{12} (x_0^3 - 3x_0), \quad \lim_{p,q \to 0} \frac{\partial \ell}{\partial p} = \frac{1}{8} (x_0^4 - 6x_0^2 + 3),$$

where $\ell$ is the contribution to the log-likelihood based on the normalised density in Equation (1), for $x_0 = (x-\mu)/\sigma$, resulting in the score test for Normality within this family being based on $(ns^3)^{-1}\sum(X_i - \bar{X})^3$ and $(ns^4)^{-1}\sum(X_i - \bar{X})^4 - 3$, the usual sample skewness and kurtosis.

Note that $q = 0$ corresponds to the usual symmetric $t$-distribution, and $p = q = 0$ corresponds to the standard Normal distribution.

### 2.2 Model for conditional moments

Bollerslev (1986) introduced the Generalised Autoregressive Conditional Heteroskedastic (GARCH) model, as a finite-order generalisation of Engle’s (1982) ARCH model. The GARCH($m_1, m_2$) model is given by

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{j=1}^{m_1} \alpha_j X_{t-j}^2 + \sum_{j=1}^{m_2} \beta_j \sigma_{t-j}^2, \quad (2)$$

where $(\varepsilon_t - m)$ is a sequence of independent innovations which will be taken to follow the distribution in Equation (1). In the sequel, $X_t$ will be the continuously compounded log-returns of the data series in question. It is widely found that such series of asset prices and market indices, when they are not
under stress, follow a GARCH(1, 1) model, and so $m_1 = m_2 = 1$ will be assumed in Equation (2) for the sake of simplicity of presentation.

By analogy with the autoregressive representation of the scale parameter in the GARCH(1, 1) model, the conditional kurtosis may be represented as

$$p_t \sigma_t = \alpha_0^{(p)} + \alpha_1^{(p)} |X_{t-1}| + \beta_1^{(p)} p_{t-1} \sigma_{t-1}, \tag{3}$$

and the conditional skewness as

$$q_t \sigma_t = \alpha_0^{(q)} + \alpha_1^{(q)} X_{t-1} + \beta_1^{(q)} q_{t-1} \sigma_{t-1}. \tag{4}$$

Although $(q_t \sigma_t)^3$ and $(p_t \sigma_t)^4$ here are not the third and fourth central moments, they can be considered as mimicking these quantities, as they describe skewness and kurtosis, respectively. Notice that $|X_{t-1}|$ appears in the expression for $p_t$ because the $t$-distribution only admits kurtosis greater than that of a Normal distribution.

Observe further that both Equations (3) and (4) are dimensionally consistent, since $p_t$ and $q_t$ are dimensionless. It might be tempting to model these moments in terms of $q_t \sigma_t^3$ and $p_t \sigma_t^4$, but typical behaviour of regression models suggests that such an approach would not render materially different results.

3 Case study

This section gives an account of the data and illustrates estimation methods. Computational and econometric issues are discussed.

3.1 Data

The data used in this study are the Nikkei 225 index, the capitalisation-weighted index of the top 225 firms in Japan. Daily closing values of the index, $I_t$, in the period 16 January 2003 to 18 November
2005 were used, and continuously compounded log-returns \( X_t = \log \left( \frac{I_t}{I_{t-1}} \right) \) were modelled according to Equations (2), (3) and (4), using Equation (1) as the innovation distribution. Summary statistics are presented in Table 1, showing clear evidence of non-Normality and asymmetry. A time plot of \( X_t \) is given in Figure 1: volatility clustering is plainly evident, as is asymmetry in the shocks.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Mean</th>
<th>Std Dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Skew.</th>
<th>Kurt.</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.00076</td>
<td>0.012</td>
<td>−0.052</td>
<td>0.033</td>
<td>−0.50</td>
<td>4.46</td>
<td>91.5</td>
</tr>
</tbody>
</table>

Table 1: Table showing summary statistics for the log-returns on the Nikkei index: mean, standard deviation, minimum, maximum, skewness, kurtosis, and the null value of the Jarque-Bera test statistic for Normality (which has an asymptotic \( \chi^2_2 \) distribution, giving a \( p \)-value of approximately zero).

![Figure 1: Time plot of the log-returns of the daily Nikkei index, 16JAN03 to 18NOV05.](image)

Models for \( (X_t) \) were fitted using numerical maximum likelihood, with additional parameters of \( \sigma_1, p_1 \) and \( q_1 \), estimated to initialise the recursive fit of the GARCH(1,1) variance in Equation (2), and extensions in Equations (3) and (4).

Table 2 displays the maximum likelihood estimates in fitting the full and reduced models, along with the value of the maximised log-likelihood. To fit a Normal model, all parameters in Equations (3) and (4) are suppressed to be zero. By freeing the parameters in Equation (3), a fit of the symmetric \( t \)
was obtained. Results shown correspond to putting $\alpha_0 = \alpha_0^{(p)} = \beta_1^{(p)} = p_1 = 0$: allowing them all to be positive resulted in a marginally greater log-likelihood of 2168.70. Similarly, in the full model for the skew $t$, allowing $\alpha_0 > 0$, $\beta_1^{(q)} > 0$, $p_1 > 0$, $\beta_1^{(p)} \neq 0$ and $q_1 \neq 0$ resulted in a marginally greater log-likelihood of 2179.58. In all three models, volatility clustering is identified by the (traditionally) large estimates of $\beta_1$, close to unity. Allowing a more flexible model of the symmetric or skew $t$ accentuates the degree of volatility clustering, in that the estimated value of $\hat{\beta}_1$ was larger in the more general models.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>$\hat{\sigma}_1$</th>
<th>$\hat{\alpha}_0$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\alpha}_0^{(p)}$</th>
<th>$\hat{\alpha}_1^{(p)}$</th>
<th>$\hat{\beta}_1^{(p)}$</th>
<th>$\hat{\alpha}_0^{(q)}$</th>
<th>$\hat{\alpha}_1^{(q)}$</th>
<th>$\hat{\beta}_1^{(q)}$</th>
<th>$\ell_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.014</td>
<td>1.5 x 10^{-6}</td>
<td>0.069</td>
<td>0.921</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2159.94</td>
</tr>
<tr>
<td></td>
<td>(0.0037)</td>
<td>(8.2 x 10^{-8})</td>
<td>(0.016)</td>
<td>(0.015)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Symmetric t</td>
<td>0.014</td>
<td>0</td>
<td>0.037</td>
<td>0.957</td>
<td>0</td>
<td>0.19</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2167.60</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td></td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.057)</td>
<td>(0.016)</td>
<td>(0.057)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Skew t</td>
<td>0.015</td>
<td>0</td>
<td>0.037</td>
<td>0.949</td>
<td>0.00030</td>
<td>0.0070</td>
<td>0</td>
<td>–0.00094</td>
<td>0.088</td>
<td>0</td>
<td>2178.36</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td></td>
<td>(0.0074)</td>
<td>(0.0087)</td>
<td>(0.00020)</td>
<td>(0.0032)</td>
<td>(0.00024)</td>
<td>(0.019)</td>
<td>(0.032)</td>
<td>(0.00024)</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>

Table 2: Table showing maximum likelihood estimates for the full skew $t$ GARCH model and nested reduced models. Figures in parentheses show standard errors. The value for $\hat{\sigma}_1$ is the fitted initial standard deviation, estimated as a parameter in the model. All other parameter estimates correspond to Equations (2), (3) and (4). The maximised log-likelihood is denoted $\ell_{max}$, and the corresponding $p$-value shows the significance of the change in $\ell_{max}$ over its immediately reduced model.

Figure 2 shows the standard deviation estimates from all three fitted models. The broad trajectory of the volatility process is captured by each model, but with important substantive differences. The Normal model tends to oversmooth the trajectory relative to it parent models. While the degree of smoothing of the symmetric $t$ relative to the skew $t$ is far less than for the Normal relative to the symmetric $t$, some smoothing is evident: firstly, the skew $t$ model estimates the peaks of the volatility process at greater values than in the symmetric $t$; and secondly, the local maxima along the trajectory are greater for the skew $t$ than the symmetric $t$.

Estimates of the conditional skewness and kurtosis are plotted in Figure 3. While local maxima in the conditional kurtosis are identified equivalently in both the symmetric $t$ and skew $t$ distribution, the skew $t$ model exhibits values damped by a factor in excess of two. (This is also reflected in the autoregressive coefficients, $\alpha_1^{(p)}$ and $\alpha_1^{(q)}$, in Table 2.) This is not surprising, as skewness in the data, which could not be explicitly modelled by the symmetric $t$, no doubt contributed to variation in the conditional kurtosis in that fitted model. It can also be seen that clustering does not appear to be a feature of conditional skewness and kurtosis, at least for these data, and that the skew $t$ fit seems to identify slight evidence of contrary movements in the conditional skewness and conditional kurtosis.
Figure 2: Time plots of the conditional volatility (standard deviation) for the Normal, symmetric $t$ and skew $t$ innovation distributions, respectively. Note that the vertical scales are not identical for all plots.
Figure 3: Time plots of the conditional kurtosis for the symmetric $t$ model, and the conditional kurtosis and conditional skewness for the skew $t$ model, respectively. Note that the vertical scales are not identical for all plots.
Finally, Figure 4 shows the one-step-ahead lower 1% points calculated from fits of the Normal, symmetric \( t \) and skew \( t \) distributions. Only the last 101 epochs are shown, for clarity. The skew \( t \) uniformly estimates the VaR more conservatively than the symmetric \( t \), and likewise when comparing the symmetric \( t \) and the Normal. In particular, the estimate using the Normal distribution seems to be a smoothed version of the estimate from the symmetric \( t \). The skew \( t \)-distribution exhibits the greatest volatility in the VaR.

![Figure 4: Time plot of the 1% Value-at-Risk, calculated one-step ahead for each of the Normal (-.-.-), symmetric \( t \) (– – –) and skew \( t \) (——) distributions: last 101 epochs only.](image)

**4 Conclusion**

This paper has demonstrated applications of numerical maximum likelihood for the skew \( t \)-distribution in fitting conditional models for financial time series. Compared to the few other studies of this problem, simple reparameterisations were employed to obtain natural estimates of conditional skewness and kurtosis, following an autoregressive scheme. In the data analysis case study, it was illustrated how the flexibility of the skew \( t \)-distribution can sharpen estimates of volatility clustering, as well as
enable local peaks in volatility to be identified with more precision by removing the onus of it having to account for variation in the conditional skewness and kurtosis. It was also found that the conditional skewness and kurtosis were certainly autocorrelated, however weakly, and did not appear to display clustering in the same way that the conditional standard deviation did. Estimates of skewness and kurtosis enabled local maxima in those values to be identified, along with some apparent contrary movement in these quantities.

There are good implications for risk management. Implied skewness has been investigated recently as a feature of estimated conditional volatility trajectories, as in, for instance, Doran et al. (2008). The methods in the present paper enable skewness and kurtosis to be estimated directly, and their impact on risk measures to be explicitly identified. Moreover, the autoregressive nature of the estimated conditional higher moments presented in this paper provide estimates of VaR which are implicitly correlated, rather than estimated independently at each time step, as is wide practice.

References


Steinwolf, A (1996) Approximation and simulation of probability distributions with a variable kurto-