A FRAMEWORK FOR NORMAL MEAN VARIANCE MIXTURE INNOVATIONS WITH APPLICATION TO GARCH MODELLING

P. J. de Jongh [1]
North-West University, Potchefstroom, South Africa
e-mail: *riaan.dejongh@nwu.ac.za*
and
J. H. Venter
North-West University, Potchefstroom, South Africa

Key words: GARCH models, normal gamma, normal inverse gamma, normal inverse Gaussian, normal log-normal, normal mean variance mixtures, z distribution.

Summary: GARCH models are useful to estimate the volatility of financial return series. Historically the innovation distribution of a GARCH model was assumed to be standard normal but recent research emphasizes the need for more general distributions allowing both asymmetry (skewness) and kurtosis in the innovation distribution to obtain better fitting models. A number of authors have proposed models which are special cases of the class of normal mean variance mixtures. We introduce a general framework within which this class of innovation distributions may be discussed. This entails writing the innovation term as a standardised combination of two variables, namely a normally distributed term and a mixing variable, each with its own interpretation. We list the existing models that fit into this framework and compare the corresponding innovation distributions, finding that they tend to be quite similar. This is confirmed by an empirical illustration which fits the models to the monthly excess returns series of the US stocks. The illustration finds further support for the ICAPM model of Merton, thus supporting recent results of Lanne and Saikonnen (2006).

1. Introduction

The success of GARCH models in describing the volatility of financial time series is amply demonstrated by the huge growth in the literature on these models (see surveys by Bollerslev, Chou and Kroner, 1992; Bollerslev, Engle and Nelson, 1994; Li, Ling and McAleer, 2002; Engle, 2002; Engle, 2003). Traditionally these models are fitted to data by means of quasi-maximum likelihood, i.e. the innovation distribution is assumed to be a standard normal distribution and maximum likelihood estimation is carried out. Although consistent estimates of the parameters result from this procedure (Bollerslev and Wooldridge, 1992), it is presently well known that the normality assumption is often unrealistic. An innovation distribution with heavier tails than the normal, as well as possible

[1] Corresponding author.

AMS: 60E99, 62E15, 91B84

asymmetry, often provides a better description and therefore would lead to more efficient estimation results. Approaching possible alternatives to the normal distribution from an empirical point of view, many authors have suggested heavier tailed distributions: among others, the t-distribution (Bollerslev, 1987), variations of the t-distribution (Jondeau and Rockinger, 2003), the variance gamma (Madan and Seneta, 1990) the general error distribution (Nelson, 1991), the NIG-distribution (Barndorff-Nielsen, 1997; Andersson, 2001; Jensen and Lunde, 2001; Stentoft, 2003; Venter and de Jongh, 2004), the z-distribution (Lanne and Saikkonen, 2004) and discrete mixtures of normal distributions (Alexander and Lazar, 2004). Many of these choices are related to the mixture-of-distributions-hypothesis (MDH) of Clark (1973) according to which a suitable variance mixture of normal distributions is appropriate for the innovation distribution. This aspect is emphasized in the work of Forsberg (2002) who used the notion of realized volatility (Andersen, Bollerslev, Diebold and Labys, 2000; Andersson, 2001; Andersen, Bollerslev, Diebold and Ebens, 2001a; Andersen, Bollerslev, Diebold and Labys, 2001b; Andersen, Bollerslev, Diebold and Labys, 2003) to guide in the selection of a mixing distribution. They find that the distribution of realized volatility scaled by a GARCH type volatility estimate is well approximated by an Inverse Gaussian (IG) distribution. As a consequence, the NIG distribution should be particularly appropriate for the innovation distribution. This result elucidates the findings of Venter and de Jongh (2004) that using the NIG distribution in this context often leads to more efficient estimation of model parameters and related quantities such as volatility and risk measures.

In this paper we formulate a fairly general family of normal mean variance mixture (NMVM) distributions that may be used. In Section 2 we discuss GARCH models in which the innovation term is a standardized form of the random variable $\sqrt{W}Z + \beta W$, with Z standard normally distributed independently from W, a positive mixing variable which can be interpreted as an impact possibly arising from news noise effects. Section 3 reviews six special choices for the distribution of W that have been selected in the existing literature; these are the Inverse Gaussian (IG), Inverse gamma (IGam), gamma, log-normal (LN), discrete mixtures and infinite convolutions of exponential distributions. We compare these mixing distributions and the corresponding NMVM distributions and find that they can often approximate each other quite closely. This suggests that it would not matter much in practice which particular mixing distribution is selected: the mixing concept is what yields better modelling, not the particular implementation of this idea. This suggestion is confirmed empirically in Section 4 where we illustrate the NMVM models in terms of the monthly excess returns of US stocks. The GARCH specification used in this illustration follows that of Lanne and Saikkonen (2006) where strong evidence in support of the intertemporal capital asset pricing model (ICAPM) of Merton (1973) is provided. Our illustration adds further support to ICAPM in that the result is robust with respect to specification of the innovation distribution used in the GARCH model. Section 5 closes with a summary and an indication of further research issues.

2. NMVM GARCH models for return series

Let $Y_1, Y_2, \ldots$ denote the returns of a financial instrument over successive time periods. A GARCH model to describe the return series takes the form

$$Y_n = \mu_n + \sqrt{h_n} X_n, \qquad \text{for } n = 1, 2, \ldots \tag{1}$$

where μ_n represents an expected (or structural) component, h_n is the volatility and X_n the innovation over period n. It is assumed that μ_n and h_n are at most dependent on the past information set $\mathscr{F}_{n-1}$(i.e. the σ-field generated by $Y_1, Y_2, \ldots, Y_{n-1}$ or a larger data set available at time $n-1$). The specific choices to be used here will be indicated later. The focus in this paper is on the choice of the distribution of the innovations X_n in the model. To make μ_n and h_n identifiable, we assume that $EX_n = 0$ and $Var(X_n) = 1$ so that $\mu_n = E[Y_n \,|\mathscr{F}_{n-1}\,]$ and $h_n = Var[Y_n \,|\mathscr{F}_{n-1}\,]$. We suppose that the X_n's are independent of $\mathscr{F}_{n-1}$ and also independent and identically distributed (i.i.d.). The common distribution of the innovations is often assumed to be $N(0,1)$ in which case Equation (1) may be written as

$$Y_n = \mu_n + \sqrt{h_n} Z_n, \tag{2}$$

where $Z_1, Z_2, \ldots$ are i.i.d. $N(0,1)$. If this model is fitted and corresponding residuals are calculated and tested against the normality assumption by a normal QQ-plot or a normal probability integral transform (PIT) plot, it often appears that an innovation distribution with heavier tails and possibly also asymmetry is required to get a better fit (Forsberg, 2002; Lanne and Saikkonen, 2004). Among others, the t-distribution and variations of the t-distribution have been used for this purpose; see e.g., Bollerslev (1987), Jondeau and Rockinger (2003). Recently a number of authors have argued in favour of using the NIG distribution for this purpose; see e.g., Andersson (2001), Jensen and Lunde (2001), Venter and de Jongh (2004), Forsberg (2002) and especially the paper of Forsberg and Bollerslev (2002). Lanne and Saikkonen (2004) use the z-distributions while Madan and Seneta (1990) favour the use of the "variance gamma" distribution. These are all special members of the class of normal mean variance mixture (NMVM) distributions. This suggests that the NMVM class be considered in somewhat greater generality as models for the innovation distribution of a GARCH model and we do so in this paper.

Let $Z_1, Z_2, \ldots$ be i.i.d. $N(0,1)$ as above and let $W_1, W_2, \ldots$ be positive i.i.d. with common density g independent of $Z_1, Z_2, \ldots$. Then we assume that

$$X_n = (\sqrt{W_n} Z_n + \beta(W_n - EW_n)) \Big/ [EW_n + \beta^2 Var(W_n)]^{\frac{1}{2}}. \tag{3}$$

The random part of X_n is given by $\sqrt{W_n} Z_n + \beta W_n$ which has an NMVM distribution with W_n as the mixing variable. The distribution is negatively skewed if the parameter β is negative which often happens in practice, as will be illustrated in our example below. Subtraction of βEW_n in Equation (3) ensures that $EX_n = 0$ and scaling by the divisor $[EW_n + \beta^2 Var(W_n)]^{\frac{1}{2}}$ ensures that $Var(X_n) = 1$ as required above. Replacing W_n by $c\tilde{W}_n$ with $c > 0$ and β by $\tilde{\beta} \Big/ \sqrt{c}$ we can express X_n in the form $X_n = (\sqrt{\tilde{W}_n} Z_n + \tilde{\beta}(\tilde{W}_n - E\tilde{W}_n)) \Big/ [E\tilde{W}_n + \tilde{\beta}^2 Var(\tilde{W}_n)]^{\frac{1}{2}}$. This shows that we need a scale restriction on the distribution of the W_n's in order to make β identifiable, which we implement by assuming henceforth that $EW_n = 1$. We shall write $\gamma^2 = 1 + \beta^2 Var(W_n)$ so that Equation (3) may be written as

$$X_n = (\sqrt{W_n} Z_n + \beta(W_n - 1))/\gamma. \tag{4}$$

Clearly the conditional distribution of X_n given W_n is $N(\beta(W_n - 1)/\gamma, W_n/\gamma^2)$ so that the density of X_n is given by

$$f(x) = \int_0^\infty \frac{\gamma}{\sqrt{w}} \varphi\left(\frac{\gamma x + \beta - \beta w)}{\sqrt{w}}\right) dG(w), \tag{5}$$

with φ the standard normal density and G the distribution function of W_n. This explicitly shows that the density of the innovations X_n is in the NMVM class. We refer to model Equation (1) with innovation X_n given by Equation (4) as a NMVM-GARCH model. Note that when the variance of W_n approaches zero, the distribution of W_n becomes degenerate on 1 and Equation (5) simply becomes the N(0,1) density.

Given past information (so that μ_n and h_n are fixed) the innovation part of the return over the n-th period is driven by two variables, namely Z_n and W_n. We think of Z_n as "normal trading noise" in the sense that it represents the total of many individually small shocks originating from the give and take in the trading process and resulting in Z_n being normally distributed (by the central limit theorem). We think of W_n as a "news noise impact" factor in the sense that it represents the effects of important news (analyst reports and ratings, company actions, etc.) that appear after the end of the $(n-1)$-th period. These effects may change the overall tone of trading during the present (n-th) period, compared to what it would be if only data available up to the end of the previous period was taken into account. Notice that in addition to $E[Y_n\,|\mathscr{F}_{n-1}] = \mu_n$ and $Var[Y_n\,|\mathscr{F}_{n-1}] = h_n$ we have

$$E[Y_n\,|\mathscr{F}_{n-1}, W_n] = \mu_n + \sqrt{h_n}\beta(W_n - 1)/\gamma \text{ and } Var[Y_n\,|\mathscr{F}_{n-1}, W_n] = h_n W_n/\gamma^2. \tag{6}$$

Thus over period n, the outcome $W_n > 1$ corresponds to "bad news" in the sense that the conditional expected return is reduced to a lower level (accepting that $\beta < 0$). At the same time, conditional volatility increases compared to what it would be in the case where the news noise impact W_n has its average value of 1. Similarly the outcome $W_n < 1$ corresponds to "good news" in the sense that the conditional expected return is enhanced and the conditional volatility is reduced. While μ_n and h_n are the expected return and volatility (conditional on past data only), Equation (6) also points to a different expected return and a different volatility (conditional on past data as well as the market tone set by news noise) that may be operating over the period, depending on the value of W_n that actually realizes during that period. Of course, with only the returns Y_n observed, we do not have direct information on the W_n's (nor on the Z_n's).

To fit a NMVM-GARCH model we need its log-likelihood function. This may be written as

$$\sum_{n=1}^{N} \left\{ \log f\left((Y_n - \mu_n) \big/ \sqrt{h_n} \right) - \tfrac{1}{2}\log h_n \right\}, \tag{7}$$

where N is the total number of periods for which data is available. To proceed we must choose the mixing distribution G and then apply Equation (5) to calculate f. The next section lists a number of special cases and indicates how to handle these issues for each one.

3. Special NMVM innovation distributions

Normal inverse Gaussian (NIG) distributions

According to Jorgensen (1982) the density of the generalised inverse Gaussian (GIG) distribution may be written as $g(w) = \left(\delta_2/\delta_1\right)^{\lambda} w^{\lambda-1}\exp[-\frac{1}{2}(\delta_1^2 w^{-1} + \delta_2^2 w)]/2K_\lambda(\delta_1\delta_2)$ where $\delta_1, \delta_2 > 0$ and K_λ is the modified Bessel function of third order. If W has this distribution, then $EW = \delta_1 K_{\lambda+1}(\delta_1\delta_2)/\delta_2 K_\lambda(\delta_1\delta_2)$. The special choice $\lambda = -\frac{1}{2}$ yields the inverse Gaussian distribution, for which $EW = \delta_1 K_{1/2}(\delta_1\delta_2) \big/ \delta_2 K_{-1/2}(\delta_1\delta_2) = \delta_1 \big/ \delta_2$. To make $EW = 1$ we take $\delta_1 = \delta_2 = \psi > 0$, obtaining what we call the "**unit inverse Gaussian**" distribution (abbreviated to $UIG(\psi)$), having density

$$\begin{aligned} g_{UIG(\psi)}(w) &= \tfrac{1}{2K_{1/2}(\psi^2)} w^{-3/2} \exp\left[-\frac{1}{2}\psi^2(w^{-1}+w)\right] \\ &= \tfrac{\psi \exp(\psi^2)}{\sqrt{2\pi}} w^{-3/2} \exp\left[-\frac{1}{2}\psi^2(w^{-1}+w)\right], \quad \text{for } w > 0. \end{aligned} \tag{8}$$

If W is $UIG(\psi)$-distributed then $Var(W) = 1\big/\psi^2$ and for this choice of g we have $\gamma^2 = 1 + \beta^2\big/\psi^2$. The density of X in this case is given by

$$f(x) = \frac{\gamma\psi}{\pi}\left(\frac{\beta^2+\psi^2}{(\gamma x+\beta)^2+\psi^2}\right)^{\frac{1}{2}} e^{\psi^2+\beta(\gamma x+\beta)} K_1\left(\sqrt{\beta^2+\psi^2}\sqrt{(\gamma x+\beta)^2+\psi^2}\right). \tag{9}$$

This can be derived from the fact that GIG densities integrate to 1. The density in Equation (9) may be called the "standard normal inverse Gaussian" distribution with parameters β and ψ (abbreviated $SNIG(\beta,\psi)$). The general class of four parameter NIG distributions, as discussed e.g. in Barndorff-Nielsen (1997), can be obtained from the $SNIG(\beta,\psi)$ distribution by translation and scale changes. It can be shown that if X is $SNIG(\beta,\psi)$ distributed then $cX+d$ is $NIG(\psi\gamma^2/c, \beta\gamma/c, d-c\beta/\gamma, c\psi/\gamma)$ distributed. Computation of the corresponding likelihood function requires access to routines that can handle Bessel functions; fortunately these are widely available in standard software packages.

Normal inverse gamma and t distributions

If W is inverse gamma distributed with parameters r and λ, then it has the density

$$g(w) = \lambda^r w^{-(r+1)}\, e^{-\lambda/w} \Big/ \Gamma(r)$$

and we find $EW = \lambda\big/(r-1)$ (here we are assuming that $r > 1$, since otherwise this expectation does not exist). Hence the requirement $EW = 1$ forces the choice $r = \lambda + 1$. With this choice we get the "**unit inverse gamma**" distribution (abbreviated $UIGam(\lambda)$) with density

$$g_{UIGam(\lambda)}(w) = \frac{\lambda^{\lambda+1} w^{-(\lambda+2)} e^{-\lambda/w}}{\Gamma(\lambda+1)}, \qquad w > 0. \tag{10}$$

It is easy to show that $Var(W) = 1\big/(\lambda-1)$ where we must assume that $\lambda > 1$ to get a finite variance. For this choice of g in Equation (5) and for the special case $\beta = 0$ we obtain the innovation density

$$f(x) = \int_0^\infty \frac{1}{\sqrt{w}}\varphi\left(\frac{x}{\sqrt{w}}\right) g_{UIGam(\lambda)}(w)dw = \frac{\Gamma(\lambda+\frac{3}{2})}{\sqrt{2\pi}\sqrt{\lambda}\,\Gamma(\lambda+1)}\left(1+\frac{x^2}{2\lambda}\right)^{-(\lambda+\frac{3}{2})}. \tag{11}$$

This is a scaled $t_{2\lambda+2}$-density with unit variance. More generally, with $\beta \neq 0$ we get a skewed form of the t-densities given by

$$f(x) = \int_0^\infty \frac{\gamma}{\sqrt{w}} \varphi\left(\frac{\gamma x+\beta-\beta w}{\sqrt{w}}\right) g_{UIGam(\lambda)}(w)dw$$
$$= \sqrt{\frac{2}{\pi}} \frac{\gamma\lambda^{\lambda+1}}{\Gamma(\lambda+1)} \left(\frac{|\beta|}{\sqrt{(\gamma x+\beta)^2+2\lambda}}\right)^{\lambda+\frac{3}{2}} e^{\beta(\gamma x+\beta)} K_{\lambda+\frac{3}{2}}\left(|\beta|\sqrt{(\gamma x+\beta)^2+2\lambda}\right) \quad (12)$$

where now $\gamma^2 = 1+\beta^2/(\lambda-1)$. Again this follows from the GIG density integrating to 1. Various skewed forms of the t distributions have been proposed (Jondeau and Rockinger, 2003) but as far as we know Equation (12) has not been discussed before. It reverts to the symmetric $t_{2\lambda+2}$ in the limit when $\beta \to 0$. We refer to it by a more descriptive name as the "**standard normal inverse gamma**" distribution with parameters β and λ (abbreviated $SNIGam(\beta,\lambda)$).

Normal gamma distributions

If W is gamma distributed with parameters r and ζ, it has the density $g(w) = \zeta^r w^{r-1} e^{-\zeta w}/\Gamma(r)$ and $EW = r/\zeta$. Hence the requirement $EW = 1$ forces the choice $r = \zeta$ leading to the "**unit gamma**" distribution (abbreviated $UGam(\zeta)$) with density

$$g_{UGam(\zeta)}(w) = \frac{\zeta^\zeta w^{\zeta-1} e^{-\zeta w}}{\Gamma(\zeta)}, \qquad w > 0. \quad (13)$$

In this case $Var(W) = 1/\zeta$ and $\gamma^2 = 1+\beta^2/\zeta$. With this choice of g in Equation (5) we get the innovation density

$$f(x) = \int_0^\infty \frac{\gamma}{\sqrt{w}} \varphi\left(\frac{\gamma x+\beta-\beta w}{\sqrt{w}}\right) g_{UGam(\zeta)}(w)dw$$
$$= \sqrt{\frac{2}{\pi}} \frac{\gamma\zeta^\zeta}{\Gamma(\zeta)} \left(\frac{|\gamma x+\beta|}{\sqrt{\beta^2+2\zeta}}\right)^{\zeta-\frac{1}{2}} e^{\beta(\gamma x+\beta)} K_{\zeta-\frac{1}{2}}\left(|\gamma x+\beta|\sqrt{\beta^2+2\zeta}\right). \quad (14)$$

Again this follows from the GIG density integrating to 1. The special case $\beta = 0$ amounts to the "variance gamma" distribution, extensively discussed by Madan and Seneta (1990). We refer to Equation (14) as the "**standard normal gamma**" distribution with parameters β and ζ (abbreviated $SNGam(\beta,\zeta)$) which is in line with the naming conventions applied so far.

Normal log-normal distributions

If U is N(0,1) distributed then $W = \exp(\tau U - \frac{1}{2}\tau^2)$ has a "**unit log-normal**" ($ULN(\tau)$) distribution in the sense that it is log-normally distributed with $EW = 1$ for all values of τ. We can restrict τ to $\tau \geq 0$ since if $\tau < 0$ we may replace U by $-U$ and write $W = \exp(\tau(-U) - \frac{1}{2}\tau^2)$ =exp$((-\tau)U - \frac{1}{2}\tau^2)$. In this case $Var(W) = \exp(\tau^2) - 1$. The distribution in Equation (5) with g the $ULN(\tau)$density may be called the "**standard normal log-normal**" (abbreviated $SNLN(\beta,\tau)$) distribution. In this case it does not seem possible to write the integral in Equation (5) in a more explicit form and one has to resort to numerical methods to compute it. When writing Equation (5) in the form

$$f(x)=\int_{-\infty}^{\infty}\frac{\gamma}{\sqrt{w(u)}}\varphi\left(\frac{\gamma x+\beta-\beta w(u))}{\sqrt{w(u)}}\right)\varphi(u)du, \quad \text{where } w(u)=\exp(\tau u-\tfrac{1}{2}\tau^2), \tag{15}$$

the Gauss-Hermite quadrature method given by Liu and Pierce (1994) is quite effective. In fact we cross-checked this method by applying it to the integrals in Equations (12) and (14) and comparing the results to the explicit expressions available in those cases and always found excellent agreement.

z distributions

The $z(a,b,c,d)$ distribution has density function

$$f(x)=\frac{1}{cB(a,b)}\frac{\left\{\exp[(x-d)/c]\right\}^a}{\left\{1+\exp[(x-d)/c]\right\}^{a+b}}, \quad \text{with } a,b,c>0, \tag{16}$$

where $B(a,b)$ is the beta function; see e.g. Lanne and Saikkonen (2004) who use it as the innovation density in GARCH models. Using arguments similar to those in Barndorff-Nielsen, Kent and Sorensen (1982), it can be shown that if $\alpha=\sum_{k=0}^{\infty}2/(a+k)(b+k)$, $\beta=\frac{1}{2}\sqrt{\alpha}(a-b)$ and

$$W=\frac{1}{\alpha}\sum_{k=0}^{\infty}\frac{2}{(a+k)(b+k)}V_k, \tag{17}$$

with $V_0,V_1,\ldots$ i.i.d. exponentially distributed with expectation 1, then $X=(\sqrt{W}Z+\beta(W-1))/\gamma$ is $z(a,b,1/\sqrt{\alpha}\gamma,-\beta/\gamma)$ distributed. Thus in this case W is an infinite convolution of exponential distributions and the corresponding model is in the NMVM class. In this case $Var(W)=\left(4/\alpha^2\right)\sum_{k=0}^{\infty}1/(a+k)^2(b+k)^2$ and as above $\gamma^2=1+\beta^2 Var(W)$. We can also parameterise the distributions in terms of the parameters α and β, which is useful in view of the special role that β plays in our exposition. The equations above express α and β in terms of a and b. Conversely, for given values of α and β, we can write $b=a-2\beta/\sqrt{\alpha}$, then solve the equation $\alpha=\sum_{k=0}^{\infty}2/(a+k)(a-2\beta/\sqrt{\alpha}+k)$ for $a=a(\alpha,\beta)$ and finally set $b=b(\alpha,\beta)=a(\alpha,\beta)-2\beta/\sqrt{\alpha}$. An expression for the density of W can be obtained as in Barndorff-Nielsen et al. (1982), but we omit it here since we have the advantageous situation that the density of the z distribution can be computed directly from Equation (16).

Discrete normal mixtures

Each of the special cases above used continuous mixing distributions. In a related but different GARCH model context, Alexander and Lazar (2004) use discrete mixing distributions. In our context, many choices of discrete mixing distributions are possible but we restrict attention to the following three point choice. Take

$$W=\begin{cases} 1-w_1, & \text{with probability} \quad p_1 \\ 1, & \text{with probability} \quad 1-p_1-p_2 \\ 1+w_2, & \text{with probability} \quad p_2 \end{cases} \tag{18}$$

with the restrictions $0 < w_1 < 1$, $w_2 > 0$ and $p_1 w_1 = p_2 w_2$ (to ensure that $EW = 1$). Then $Var(W) = p_1 w_1^2 + p_2 w_2^2$ and $\gamma^2 = 1 + \beta^2(p_1 w_1^2 + p_2 w_2^2)$. From this result we obtain the innovation density

$$f(x) = \frac{\gamma p_1}{\sqrt{1-w_1}} \varphi\left(\frac{\gamma x + \beta w_1}{\sqrt{1-w_1}}\right) + \gamma(1-p_1-p_2)\varphi(\gamma x) + \frac{\gamma p_2}{\sqrt{1+w_2}} \varphi\left(\frac{\gamma x - \beta w_2}{\sqrt{1+w_2}}\right). \quad (19)$$

We refer to this as the "**standard normal discrete**" (abbreviated $SND(\beta, w_1, w_2, p_1, p_2)$) distribution. The outcome $W = 1 - w_1$ (resp. $W = 1 + w_2$) may be referred to as a "good news" (resp. "bad news") period and the outcome $W = 1$ is a "neutral news" period. This is a simple family to deal with from a computational point of view, although it has one more free parameter than the other models considered. Furthermore, the implied assumption that only certain specific news noise impact severities are possible may be unrealistic in many cases.

Comparisons of the distributions

The distribution of the W_n's was unitized to have expectation 1. However, since their variances depend on the remaining parameter(s), one way to compare them is to choose these parameter(s) such that they have the same variances. In the $UIG(\psi)$, $UIGam(\lambda)$, $UGam(\zeta)$ and $ULN(\tau)$ cases, the distribution of W_n depends only on one parameter and not on the skewness inducing parameter β of the normal mixing formula. The equations

$$\frac{1}{\psi^2} = \frac{1}{\lambda - 1} = \frac{1}{\zeta} = \exp(\tau^2) - 1 \quad (20)$$

ensure that the variances of these distributions are the same. However, it is more relevant to compare the NMVM distributions resulting from using these mixing distributions for W_n since it need not be the case that equalizing the variances would lead to the mixtures approximating each other well. Relative entropy (Soofi and Retzer, 2002) may be used to make a more relevant comparison. The entropy $\int_{-\infty}^{\infty} \log\{f_{SNIG(\beta,\psi)}(x)/f_{SNIGam(\beta,\lambda)}(x)\} f_{SNIG(\beta,\psi)}(x)dx$ measures the distance between the $SNIG(\beta, \psi)$ and $SNIGam(\beta, \lambda)$ distributions. For a given ψ we can choose $\lambda = \lambda(\psi)$ to minimize this entropy (keeping β fixed). The choice $\lambda = \lambda(\psi)$ may be interpreted as the MLE of λ when the $UIGam(\lambda)$ distribution is fitted to an infinite sample of observations from a $SNIG(\beta, \psi)$ distribution. Similar entropy minimizing choices $\zeta(\psi)$, $\tau(\psi)$ and $\{w_1(\psi), w_2(\psi), p_1(\psi), p_2(\psi)\}$ may be defined to approximate the $SNIG(\beta, \psi)$ by the $SNGam(\beta, \varsigma)$, $SNLN(\beta, \tau)$ and $SND(\beta, w_1, w_2, p_1, p_2)$ distributions respectively. As an illustration we used the $SNIG(-0.5, 2)$ distribution and found numerically that $SNIGam(-0.5, 4.8423)$, $SNGam(-0.5, 4.2306)$, $SNLN(-0.5, 0.4734)$ and $SND(-0.5, 0.5595, 1.2399, 0.2492, 0.1125)$ are the corresponding entropy minimizers. The case of the z distribution is somewhat more involved since it depends on both parameters a and b. It thus indirectly involves the mean mixing parameter β via the relationship $\beta = \frac{1}{2}\sqrt{\alpha}(a - b)$. In this case we can minimize the entropy over both a and b to obtain the best approximation among the z-densities to any one of the others. For the $SNIG(-0.5, 2)$ distribution the choices $a = 1.2664$ and $b = 2.0369$ (with $\alpha = 1.6962$, $\beta = -0.5016$) yield the best entropy approximation among the z-densities.

Figure 1 shows the graphs of the $SNIG(-0.5, 2)$ together with the best approximations from the other five densities. Figure 2 compares the logs of these densities. Evidently in both graphs the densities are so nearly equal as to be graphically indistinguishable, even far out in the tails. Similar results hold at other parameter choices of the parameters. All numerical results were done with SAS using Proc NLP for numerical optimisation work.

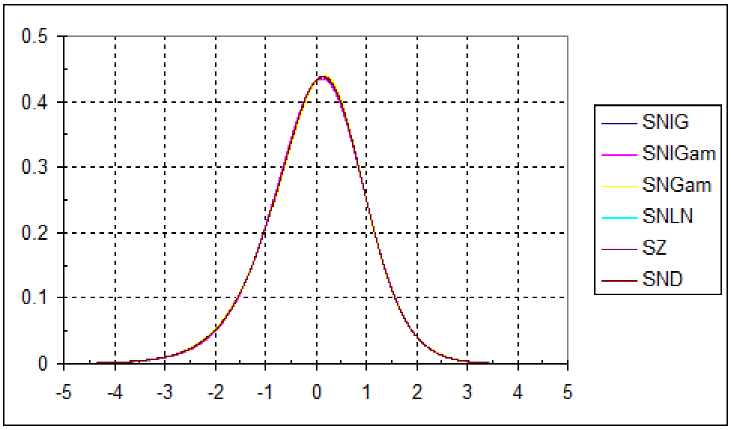

Figure 1: $SNIG(-0.5, 2)$ density and its best entropy approximations.

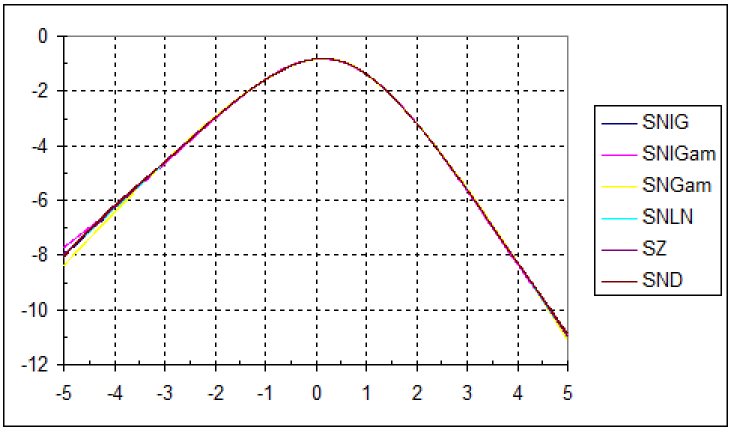

Figure 2: Logs of the $SNIG(-0.5, 2)$ density and its best entropy approximations.

4. Empirical illustration

In this section we illustrate the NMVM models by applying them to the monthly excess returns series of US stocks (downloaded from the website[2] of Kenneth French). The model specifications

[2] The series Mkt-Rf in the file `F-F_Research_Data_Factors.txt` at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

used here are motivated by a similar application in Lanne and Saikkonen (2006). According to the intertemporal capital asset pricing model (ICAPM) of Merton (1973) conditional expected excess return should be proportional to the conditional excess volatility, i.e. we should have $\mu_n = \delta h_n$ with $\delta > 0$. Lanne and Saikkonen (2004, 2006) and also Ghysels, Santa-Clara and Valkanov (2005) review this theory and the extensive literature that examined the empirical support of the anticipated ICAPM relation. Prior to their work, relatively unconvincing empirical support for the ICAPM was reported in the literature. Ghysels et al. (2005) introduce a new type of volatility estimator, referred to as the MIDAS estimator. In principle NMVM innovation distributions can be applied in the MIDAS context as well, but we do not pursue this here. Instead we follow only the GARCH work of Lanne and Saikkonen (2006), who fit an NMVM-GARCH model with the symmetric t-distribution for the innovations using the specifications $\mu_n = \delta_0 + \delta h_n$, $h_n = \alpha_0 + \alpha_1(Y_{n-1} - \mu_{n-1})^2 + \beta_1 h_{n-1}$. They find that inclusion of the unnecessary intercept δ_0 in the specification of μ_n makes estimation of δ highly inaccurate and forms part of the reason why support for the ICAPM was often not found in empirical work. With the specification $\mu_n = \delta h_n$ their estimate of δ is positive and significantly different from zero, confirming the ICAPM relation. They obtain this result for the series over the period 1928:1 to 2000:12 as well as over the sub-periods 1928:1 to 1963:12 and 1964:1 to 2000:12. We repeated their analysis under all choices of the NMVM innovation densities and found that their results continue to hold in all cases.

To save space we report here only the results for the ICAPM model over the second sub-period 1964:1 to 2006:11. We used all six NMVM innovation densities as well as the symmetric t choice for comparison purposes. Table 1 shows the corresponding MLEs (and their asymptotic standard errors) of the proportionality parameter δ and the GARCH parameters and Table 2 shows the MLEs of the skewness mixing parameter β and the parameters of the various innovation distributions.

Table 1: MLEs of the ICAPM and GARCH parameters of NMVM models for the US monthly excess returns series.

Model	δ	$\alpha_0 \times 10^4$	α_1	β_1
SNIG	3.2411 (1.0927)	0.9139 (0.3617)	0.0986 (0.0228)	0.8549 (0.0268)
SNIGam	3.2867 (1.0817)	0.9167 (0.3720	0.1002 (0.0234)	0.8538 (0.0275)
SNGam	3.2437 (1.0940)	0.9365 (0.3614)	0.0978 (0.0224)	0.8542 (0.0267)
SNLN	3.2578 (1.0863)	0.9223 (0.3647)	0.0989 (0.0228)	0.8542 (0.0271)
Sz	3.2314 (1.0753)	0.8600 (0.3648)	0.0965 (0.0230)	0.8598 (0.0291)
SND	3.3299 (1.0406)	1.1200 (0.4775)	0.1018 (0.0280)	0.8432 (0.0330)
$t(v)$	3.9259 (0.9926)	0.1080 (0.3990)	0.1053 (0.0237)	0.8419 (0.0270)

Table 2: MLEs of the innovation distribution parameters of NMVM models for the US monthly excess returns series.

Model	β	Other parameters
SNIG	-0.5581 (0.2137)	ψ: 1.7324 (0.3706)
SNIGam	-0.5787 (0.22673)	λ: 3.9678 (1.6968)
SNGam	-0.5948 (0.1872)	ζ: 3.1659 (1.0330)
SNLN	-0.5647 (0.2106)	τ: 0.5386 (0.1001)
Sz	-0.5729 (0.1585)	α: 2.3179 (0.3240)
SND	-0.7366 (0.3415)	w_1: 0.8982 (0.1141) w_2: 2.5045 (1.3201) p_1: 0.0936 (0.0609) p_2: 0.0336 (0.0251)
$t(v)$	0	v:7.4786 (2.3254)

It is clear that the estimates of the GARCH parameters are all very similar across the different innovation distributions, as anticipated, and also agree quite well with the values of Lanne and Saikkonen (2006) for their second sub-period. The estimates of δ for the first six innovation densities (which allow for possible skewness) are also very similar and somewhat smaller than that of the symmetric t – which is again very close to that of Lanne and Saikkonen (2006). These estimates are about three times their standard errors and the p-values for testing that $\delta = 0$ are less than 0.01, giving good support for the ICAPM theory. They are also in line with the results reported by Lanne and Saikkonen (2006) for the symmetric t. The estimates of β are also quite similar, except for the case of the SND model, which seems somewhat more negative. However, the differences do not appear excessive when taking the standard errors into account. The p-values of these estimates are of the order of 0.03 and less, suggesting that innovation distributions with negative skewness are required.

The estimates of the other parameters can be compared using the variance or entropy techniques discussed in the previous section. For example, taking the SNIG parameter values $\psi = 1.7324$ as given, the closest SNLN entropy approximation has $\tau = 0.5373$ which corresponds closely with the entry in the SNLN line in Table 2. This also happens for the other cases in the table, except for SND where the correspondence is poor. Perhaps the differences between the discrete mixture and the continuous mixtures are simply too large from a finite sample point of view to expect close agreement in that case.

Figure 3 shows PIT plots for the innovation residuals corresponding to the six NMVM models. To obtain these we first substitute the parameter estimates into the expressions for μ_n and h_n to obtain their estimates $\hat{\mu}_n$ and $\hat{h}_n$, compute the residuals $\hat{X}_n = (Y_n - \hat{\mu}_n) \Big/ \sqrt{\hat{h}_n}$ and their probability integral transforms (PITs) $PIT_n = \int_{-\infty}^{\hat{X}_n} \hat{f}(x)dx$ with $\hat{f}$ the relevant estimated innovation density, then order them into $PIT_{(1)} < \cdots < PIT_{(N)}$ and finally draw a scatter plot of the pairs $(n/(N+1), PIT_{(n)})$. In all cases the PIT plots (virtually indistinguishable on the graph) closely follow the equi-angular line corresponding to the uniform distribution, indicating excellent fits of the models, irrespective of which particular choice of the mixing distribution is implemented. Formal tests of fit can be done by applying tests of normality to the series $\Phi^{-1}(PIT_n)$. Shapiro-Wilk W-tests of this form have p-values of the order of 0.4, confirming the quality of the fits.

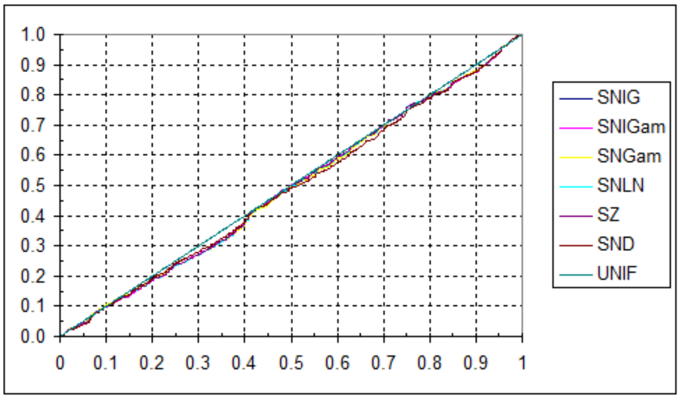

Figure 3: PIT plots of the residuals of the NMVM models.

In summary, the illustration shows that the five NMVM innovation distributions based on continuous mixing distributions tend to produce closely similar results. This is often also the case with the sixth distribution where the mixing distribution is discrete. The results of this illustration are themselves also of considerable interest, in the sense that only few instances of empirical support for the ICAPM theory have been reported in the literature so far. Here the recent positive contributions of Lanne and Saikkonen (2004, 2006) are shown to be robust with respect to the choice of innovation distribution used in the volatility model.

5. Summary and concluding remarks

Well-fitting GARCH models for financial series often require both skewed and heavy-tailed innovation distributions. This can be achieved by normal mean and variance mixture distributions. The literature already contains a number of proposals for this purpose. In this paper we presented a systematic and unified NMVM-GARCH framework and reviewed and compared the specific innovation distributions that have been used. Our experience based on the illustrations in Section 4 (as well as on other examples and simulation studies not reported here) suggest that the five NMVM innovation distributions based on continuous mixing distributions virtually always produce closely similar results. This is encouraging in the sense that it is the underlying phenomenon we are trying to model that is important, more so than the specific mathematical forms used in the process. As such the results should be stable when these forms are varied over reasonably possible alternatives.

A number of questions call for further research. Firstly, as the results are stable under the different mixing distribution choices, is a type of semi-parametric approach possible in which we do not need to make a specific parametric choice for the mixing distribution? Secondly, what information about the periodic news-noise impact factors W_n can be inferred from the observed returns? In principle it is possible to write down expressions for filtering estimators of the form $E[W_n|Y_1,\ldots,Y_n]$, but they tend to be inaccurate in that it is largely the single observation Y_n that plays a role in the estimation of W_n. However, additional data over each period is often available. For example, in addition to the closing values, open, high and low values are also available for many other financial series. In principle the use of this additional data should lead to more accurate modelling and estimation. An approach towards doing this for the case of a somewhat different but related form of the NIG innovation distribution is given in Venter, de Jongh and Griebenow (2005). Another instance is the case of daily returns, where intra-day data may be available in addition to the daily returns. An analogous approach towards dealing with this case is discussed in Venter, de Jongh and Griebenow (2006). We intend to extend this work to incorporate the approach of the present paper and its applications.

Acknowledgements

The authors sincerely thank the reviewers for very useful comments. This research was supported by the National Research Foundation and by industry (Absa and SAS Institute).

References

ALEXANDER, C. AND LAZAR, E. (2004). The equity index skew and asymmetric normal mixture GARCH. *ISMA Centre discussion papers in Finance 2004-14*.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND EBENS, H. (2001a). The distribution of stock return volatility. *Journal of Financial Economics*, **61**, 43–76.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND LABYS, H. (2000). Exchange rate returns standardized by realized volatility are (nearly) Gaussian. *Multinational Finance Journal*, **4**, 159–179.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND LABYS, H. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96**, 42–55.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND LABYS, H. (2003). Modelling and forecasting realized volatility. *Econometrica*, **71**, 579–625.

ANDERSSON, J. (2001). On the normal inverse Gaussian stochastic volatility model. *Journal of Business and Economic Statistics*, **19**, 44–54.

BARNDORFF-NIELSEN, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, **24**, 1–13.

BARNDORFF-NIELSEN, O. E., KENT, J., AND SORENSEN, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review*, **50**, 145–159.

BOLLERSLEV, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Econometrics and Statistics*, **69** (3), 542–547.

BOLLERSLEV, T., CHOU, R. Y., AND KRONER, K. F. (1992). ARCH modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, **39**, 5–59.

BOLLERSLEV, T., ENGLE, R. F., AND NELSON, D. B. (1994). ARCH models. *In* ENGLE, R. F. AND MCFADDEN, D. C. (Editors) *Handbook of Econometrics*. North-Holland: Amsterdam, pp. 2959–3038.

BOLLERSLEV, T. AND WOOLDRIDGE, D. B. (1992). Quasi-maximum likelihood estimation of dynamic models with time varying covariances. *Econometrics Reviews*, **11**, 143–172.

CLARK, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, **41**, 135–155.

ENGLE, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, **17**, 425–446.

ENGLE, R. F. (2003). Risk and volatility: Econometric models and financial practice. *In Nobel Lecture*.

FORSBERG, L. (2002). *On the Normal Inverse Gaussian Distribution in Modeling Volatility in the Financial Markets*. Ph.D. thesis, Uppsala University Library. Studia Statistica Upsaliensia 5, Christofferson, A. and Joreskog, K.G., Editors.

FORSBERG, L. AND BOLLERSLEV, T. (2002). Bridging the gap between the distribution of realized (Ecu) volatility and ARCH modelling (of the Euro): The GARCH-NIG Model. *Journal of Applied Econometrics*, **17**, 535–548.

GHYSELS, E., SANTA-CLARA, P., AND VALKANOV, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, **76**, 509–548.

JENSEN, M. B. AND LUNDE, A. (2001). The NIG-S&ARCH model: a fat-tailed, stochastic and autoregressive conditional heteroscedastic volatility model. *Econometrics Journal*, **4**, 319–342.

JONDEAU, E. AND ROCKINGER, M. (2003). Conditional volatility, skewness and kurtosis: existence, persistence and comovements. *Journal of Economic Dynamics and Control*, **27**, 1699–1737.

JORGENSEN, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*, volume 9 of *Lecture Notes in Statistics*. Springer: Heidelberg.

LANNE, M. AND SAIKKONEN, P. (2004). A skewed GARCH-in-mean model: an application to US stock returns. School of Business and Economics, University of Jyvaskyla, Finland.

LANNE, M. AND SAIKKONEN, P. (2006). Why is it so difficult to uncover the risk-return tradeoff in stock returns? *Economic Letters*, **92**, 118–125.

LI, W. K., LING, S., AND MCALEER, M. (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economics Surveys*, **16**, 245–269.

LIU, Q. AND PIERCE, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika*, **81**, 624–629.

MADAN, D. B. AND SENETA, E. (1990). The variance gamma (V.G.) model for share market returns. *The Journal of Business*, **63**, 511–524.

MERTON, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, **41**, 867–887.

NELSON, D. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica*, **59**, 347–370.

SOOFI, E. S. AND RETZER, J. J. (2002). Information indices: Unification and applications. *Journal of Econometrics*, **107**, 17–40.

STENTOFT, L. (2003). *Modelling the Volatility of Financial Assets using the Normal Inverse Gaussian Distribution: With an application to Option Pricing*. School of Economics & Management, University of Aarhus, Denmark.

VENTER, J. H. AND DE JONGH, P. J. (2004). Selecting an innovation distribution for GARCH models to improve efficiency of risk and volatility estimation. *The Journal of Risk*, **6** (3), 27–53.

VENTER, J. H., DE JONGH, P. J., AND GRIEBENOW, G. (2005). NIG-GARCH models based on open, close, high and low prices. *South African Statistical Jounal*, **39**, 173–195.

VENTER, J. H., DE JONGH, P. J., AND GRIEBENOW, G. (2006). GARCH-type volatility models based Brownian inverse Gaussian intra-day return processes. *The Journal of Risk*, **8** (4), 97–116.

Manuscript received, 2014-04-29, revised, 2015-02-16, accepted, 2015-02-20.