A Skewed GARCH-in-Mean Model:
An Application to U.S. Stock Returns

Markku Lanne
University of Jyväskylä

Pentti Saikkonen
University of Helsinki

January 31, 2004

Abstract

In this paper we consider a GARCH-in-Mean (GARCH-M) model based on the so-called z distribution. This distribution is capable of modeling moderate skewness and kurtosis typically encountered in financial return series, and the need to allow for skewness can be readily tested. We apply the new GARCH-M model to study the relationship between risk and return in monthly postwar U.S. stock market data. Our results indicate the presence of conditional skewness in U.S. stock returns, and, in contrast to the previous literature, we show that a positive and significant relationship between return and risk can be uncovered, once an appropriate probability distribution is employed to allow for conditional skewness.

1 Introduction

The presence of both conditional and unconditional skewness in financial market returns, especially stock returns, has been recognized in the empirical financial literature for decades, but only few attempts to model it have been made. In this paper we introduce a new kind of GARCH model that allows the error term to be conditionally skewed. Specifically, the model imposes comovement of conditional skewness and conditional variance, in line with the so-called volatility feedback effect (Campbell and Hetschel (1992)) that has been used to explain the presence of conditional left-skewness observed in stock returns. This effect amplifies the impact of bad news but dampens the impact of good news on returns through an increase in future volatility following all kinds of news. Under this effect also the unconditional return distribution tends to be left-skewed.

Properly capturing conditional skewness in financial returns is important at least for three reasons. First, unmodeled skewness may affect inference on other parameters in the model, and hence, misleading conclusions may be drawn, as our empirical application to stock returns illustrates. Second, data generating processes that accurately describe the return process are required in option pricing and risk management where simulation methods are employed. Recently Kalimipalli and Sivakumar (2003) and Christoffersen et al. (2003) have demonstrated the importance of incorporating conditional skewness in models used for option pricing. Finally, the results of Harvey and Siddique (2000) suggest that conditional skewness is also priced in the stock market.

Probably the most prominent specification incorporating skewness and GARCH in the empirical literature so far is Hansen's (1994) autoregressive conditional density model with a skewed version of the t distribution. In this paper we consider a GARCH-in-Mean (GARCH-M) model based on an alternative distribution, namely the so-called z distribution. This distribution was studied by Barndorff-Nielsen et al. (1982) who showed that it can be represented as a variance-mean mixture of normal distributions. The z distribution has an analytically simple density and its moments can be readily obtained. The z distribution is capable of modeling moderate skewness and kurtosis and the need to allow for skewness can be readily tested.

We apply the new GARCH-M model to study the relationship between risk and return in monthly postwar U.S. stock market data. Theoretically the relationship should be positive, but the voluminous empirical literature examining this issue is not unanimous. Different GARCH-M specifications have been considered, but to date there is very little empirical evidence of a positive relationship between risk and return. Recently Ghysels et al. (2003) even argued that monthly data are insufficient to accurately estimate the expected return–volatility trade-off and demonstrated the success of their new method combining data sampled at different frequencies. Our results indicate the presence of conditional skewness in U.S. stock returns, and, in contrast to the previous literature, we show that a positive and significant relationship between return and risk can be uncovered, once an appropriate probability distribution is employed to allow for conditional skewness.

The plan of the paper is as follows. In Section 2 the new GARCH-M specification is introduced and its properties are discussed, while Section 3 briefly deals with parameter estimation and statistical inference. In Section 4 the empirical results are presented. Finally, Section 5 concludes.

2 Model

Consider the GARCH-M model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \delta h_t + h_t^{1/2} \varepsilon_t, \tag{1}$$

where $\phi_0, ..., \phi_p$ and δ are real valued parameters, ε_t is a sequence of independent, identically distributed ($i.i.d.$) random variables, and $h_t^{1/2}$ is a (positive) volatility process which describes the conditional heteroskedasticity in the observed process y_t. Independence of h_{t-j} ($j > 0$) and ε_t is also assumed and, for stationarity, the roots of the polynomial $1 - \phi_1 z - \cdots - \phi_p z^p$ are required to lie outside the unit circle. Any available model can be used to model conditional heteroskedasticity. We shall return to this point later after discussing the distribution assumed for the error term ε_t.

We shall assume that the distribution of the error term is a certain mixture of normal distributions. In general, we say that the distribution of a random variable x is a normal variance-mean mixture with a nonnegative mixing variable ξ if, for a given ξ, the distribution of x is normal with mean $\mu + \nu\xi$ and variance ξ. If $\nu = 0$, the distribution is symmetric and called normal variance mixture. We refer to Barndorff-Nielsen et al. (1982) for a discussion of variance-mean mixtures of normal distributions.

The distribution specified for the mixing variable ξ determines the (unconditional) distribution of the random variable x. Various special cases can be obtained. For instance, assuming that the mixing variable is distributed as a reciprocal of a gamma random variable gives in the symmetric case an ordinary t distribution whereas a skewed version of the t distribution results in the asymmetric case. Another special case is obtained by assuming an inverse Gaussian distribution for the mixing variable. This special case has recently been applied by Andersson (2001) and Jensen and Lunde (2001) to model conditional heteroskedasticity. These examples are special cases of a more general specification which assumes that the mixing variable has a generalized inverse Gaussian distribution. Except for the ordinary t distribution, the density functions of these distributions depend on a modified Bessel function. An analytically simpler density is obtained by a specification to be discussed below.

The distribution we are going to apply is the so-called z distribution. This distribution is also studied by Barndorff-Nielsen et al. (1982) who show that it can be represented as a normal variance-mean mixture with the mixing distribution is an infinite convolution of exponential distributions. The z distribution, denoted by $\mathrm{z}(a, b, \sigma, \mu)$, is characterized by the density function

$$f(x) = \frac{1}{\sigma B(a, b)} \frac{\{\exp[(x-\mu)/\sigma]\}^a}{\{1+\exp[(x-\mu)/\sigma]\}^{a+b}} \quad (x \in \mathbb{R};\ a, b, \sigma > 0;\ \mu \in \mathbb{R}), \tag{2}$$

where $B(\cdot, \cdot)$ is the beta function. Clearly, μ is a location parameter and σ is a scale parameter. If $a = b$ the distribution is symmetric whereas it is positively (negatively) skewed if $a > b$ $(b > a)$. The characteristic function of the z(a, b, σ, μ) distribution is

$$\chi(s) = \frac{e^{it\mu} B(a + i\sigma s, b - i\sigma s)}{B(a, b)}. \tag{3}$$

It may be noted that the density function of the standard symmetric z$(\lambda, \lambda, 1, 0)$ distribution can also be written as

$$f(x) = \left[4^{\lambda} B(\lambda, \lambda)\right]^{-1} [\cosh(x/2)]^{-2\lambda} \qquad (x \in \mathbb{R};\ \lambda > 0).$$

The reason for the name z distribution is that the z-transformation of the sample correlation coefficient from a normal population is obtained as a special case. Another well-known special case is the logistic distribution which is obtained by assuming $a = b = 1$. Further relations to standard distributions can be obtained by observing that if the random variable w has a beta distribution with parameters α and β then $\log(w/(1-w)) \sim z(\alpha, \beta, 1, 0)$. Hence, if w has an F distribution with f_1 and f_2 degrees of freedom then $\log w \sim z\left(\frac{1}{2}f_1, \frac{1}{2}f_2, 1, \log(f_2/f_1)\right)$.

Now suppose that the random variable x has a z$(a, b, 1, 0)$ distribution. From the characteristic function (3) it is straightforward to obtain the cumulants of x. Let $\Psi(s) = d \log \Gamma(s)/ds$ signify the psi or digamma function and denote $\Psi^{(n)}(s) = d^n \Psi(s)/ds^n$ $(n = 1, 2, ...)$. Then, the nth cumulant of x, denoted by $\varkappa_n$, is

$$\varkappa_n = \Psi^{(n-1)}(a) + (-1)^n \Psi^{(n-1)}(b), \quad n = 1, 2, ..., \tag{4}$$

where $\Psi^{(0)}(s) = \Psi(s)$. From this expression and the well-known relations between cumulants and moments one can obtain the moments of x. The first four central moments are

$$Ex = \Psi(a) - \Psi(b) \stackrel{def}{=} \mu(a, b),$$

$$Var(x) = \Psi'(a) + \Psi'(b) \stackrel{def}{=} \sigma^2(a, b),$$

$$E(x - Ex)^3 = \Psi''(a) - \Psi''(b),$$

and

$$E(x - Ex)^4 = \Psi'''(a) + \Psi'''(b) + 3\sigma^4(a, b).$$

Because the transformed variable $\sigma x + \mu$ has the $\mathrm{z}(a, b, \sigma, \mu)$ distribution these results can readily be extended to any values of the parameters σ and μ.

To get an idea of the possible shapes of the z distribution, consider the symmetric $\mathrm{z}(\lambda, \lambda, 1, 0)$ distribution and note that the function $\Psi^{(n)}(s)$ has the series representation $\Psi^{(n)}(s) = (-1)^{n+1} n! \sum_{j=0}^{\infty} (s+j)^{-n-1}$ $(n = 1, 2, ...)$ (see Abramowitz and Stegun (1972, result 6.4.10)). Using this result and the preceding expression of the fourth central moment of the $\mathrm{z}(a, b, 1, 0)$ distribution it is not difficult to show that the excess kurtosis of the $\mathrm{z}(\lambda, \lambda, 1, 0)$ distribution is a decreasing function of λ and approaches three as λ approaches zero. In the asymmetric case the situation is different, however. Arguments similar to those in the symmetric case show that, for a fixed value of the parameter b, the excess kurtosis of the $\mathrm{z}(a, b, 1, 0)$ distribution is a decreasing function of a and approaches six as a approaches zero. The same result is obtained if the roles of the parameters a and b are reversed. In a similar way it can also be seen that the coefficient of skewness can be at most two in absolute value. Thus, data sets which require very strong kurtosis or skewness cannot be modeled by z distributions.

As already mentioned, we shall assume that the error term ε_t in (1) has a z distribution. Because ε_t is an error term we want it to have zero mean and, as common in GARCH and GARCH-M models, unit variance. Thus, we shall assume that

$$\varepsilon_t \sim \mathrm{z}\left(a, b, 1/\sigma(a,b), -\mu(a,b)/\sigma(a,b)\right). \tag{5}$$

Using the moments obtained for the z distribution above it is easy to check that this assumption really implies that $E\varepsilon_t = 0$ and $Var(\varepsilon_t) = 1$. Thus, the model we wish to consider is defined by (1) and (5). An alternative possibility to define the model is to specify the conditional distribution of y_t given its past. The result can be obtained from (1) and (5). In symbols we have

$$y_t \mid \mathcal{F}_{t-1} \sim \mathrm{z}\left(a, b, h_t^{1/2}/\sigma(a,b), \mu_t(\varphi) - h_t^{1/2}\mu(a,b)/\sigma(a,b)\right), \tag{6}$$

where $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$ and $\mu_t(\varphi) = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \delta h_t$ with $\varphi = \begin{bmatrix}\phi_0 & \cdots & \phi_p & \delta\end{bmatrix}'$. Clearly, $\mu_t(\varphi)$ and h_t are the conditional mean and variance of

y_t, respectively. To make the specification complete, we still have to specify a model for conditional heteroskedasticity.

As already mentioned, any available model can be used to model conditional heteroskedasticity. In this paper we shall consider a slight extension of the standard GARCH model given by

$$h_t = \omega + \sum_{j=1}^{r} \beta_j h_{t-j} + \sum_{j=1}^{q} \alpha_j u_{t-j}^2, \tag{7}$$

where

$$u_t = y_t - \mu_t(\varphi) - \kappa h_t^{1/2}$$

with κ a real valued parameter. As usual, the parameters in (7) are supposed to satisfy $\omega > 0$, $\beta_j \geq 0$ and $\alpha_j \geq 0$. Because $\mu_t(\varphi)$ is the conditional mean of y_t the choice $\kappa = 0$ corresponds to the standard GARCH specification. The motivation to allow for other possibilities is that in the case of skewed distributions is may not be clear whether the conditional mean provides the best way to center the observed series. For instance, choosing $\kappa = -\mu(a,b)/\sigma(a,b)$ means that the centering is performed by using the location parameter of the employed z distribution (see (6)). Compared to the standard specification $u_t = y_t - \mu_t(\varphi)$ this choice of κ shifts the distribution of u_t to the left when the skewness is negative, implying that negative values of u_t contribute more to conditional heteroskedasticity than in the standard case. When the skewness is positive the opposite happens. Of course, one can also specify κ as a free parameter and let the data decide its most appropriate value.

If the value of the parameter κ is nonzero the usual stationarity conditions of the GARCH process are not directly applicable. However, because $u_t = h_t^{1/2}(\varepsilon_t - \kappa)$ appropriate stationarity conditions can be readily concluded from results of Carrasco and Chen (2002). For simplicity, consider the important special case $p = q = 1$ and assume that

$$E\left(\beta_1 + \alpha_1 (\varepsilon_t - \kappa)^2\right)^k < 1, \quad k \geq 1, \tag{8}$$

where k is an integer. Then, from Corollary 6 of Carrasco and Chen (2002) it follows that the process h_t $(t = 1, 2, ...)$ can be given an initial distribution which makes it stationarity and strong mixing (or even β-mixing) with geometrically decaying mixing numbers. From the same result one also obtains that $Eh_t^k < \infty$ and that the process u_t is stationary with $Eu_t^{2k} < \infty$. This implies that y_t can be treated as a stationary process with $E\left|y_t\right|^k < \infty$. It is also near epoch dependent in L_k-norm and of any finite size (cf. Davidson (1994, Example 17.3.)). Thus, for $k \geq 2$, usual laws of large numbers and central limit theorems apply.

3 Parameter Estimation and Statistical Inference

ML estimation of the parameters of the model defined by equations (1), (5) and (7) is, in principle, straightforward. Suppose we have an observed time series y_t, $t = -l + 1, ..., T$ where $l = \max(p, q)$. Then the conditional density of y_t $(t \geq 1)$ given the past values of the series can be obtained from (2) and (6). The result is

$$f_{t-1}\left(y_t; \theta\right) = \frac{\sigma\left(a, b\right)}{h_t^{1/2} B\left(a, b\right)} \frac{\left\{\exp\left[\sigma\left(a, b\right)\left(y_t - m_t\left(\theta\right)\right)/h_t^{1/2}\right]\right\}^a}{\left\{1 + \exp\left[\sigma\left(a, b\right)\left(y_t - m_t\left(\theta\right)\right)/h_t^{1/2}\right]\right\}^{a+b}},$$

where, for simplicity, $m_t\left(\theta\right) = \mu_t\left(\varphi\right) - \mu\left(a, b\right) h_t^{1/2}/\sigma\left(a, b\right)$ and $\theta = \left[\varphi' \ \gamma' \ a \ b\right]'$ with $\gamma = \left[\omega \ \beta_1 \ \cdots \ \beta_r \ \alpha_1 \ \cdots \ \alpha_q \ \kappa\right]'$. Here κ is treated as a free parameter. The restrictions discussed after equation (7) can be handled in an obvious way. Conditional on the initial values $y_{-l+1}, ..., y_0$, the logarithm of the likelihood function can thus be written as

$$l_T\left(\theta\right) = \sum_{t=1}^{T} \log f_{t-1}\left(y_t; \theta\right).$$

The maximization of $l_T\left(\theta\right)$ is, of course, a highly nonlinear problem but can be carried out by standard numerical algorithms.

By the stationarity and near epoch dependence properties of the processes y_t and h_t discussed at the end of the previous section it is reasonable to apply conventional large sample results of ML estimation. Thus, a ML estimator of the parameter θ, denoted by $\widehat{\theta}$, can be treated as approximately normally distributed with mean value

θ and covariance matrix -$\left(\partial^2 l_T(\theta)/\partial\theta\partial\theta'\right)^{-1}$. Approximate standard errors of the components of $\widehat{\theta}$ can therefore be obtained by taking the square roots of the diagonal elements of -$\left(\partial^2 l_T(\widehat{\theta})/\partial\theta\partial\theta'\right)^{-1}$. Likelihood ratio, Wald, and Lagrange multiplier tests with approximate chi square distributions can also be performed in the usual way.

4 Application to U.S. Stock Returns

To illustrate the properties of the model presented in the previous section, we consider an application to U.S. stock returns. Several studies have examined the relationship between expected return and conditional variance with Mertons's (1973) Intertemporal Capital Asset Pricing Model (ICAPM) as a starting point. According to this model the expected excess return on the stock market depends positively on its conditional variance:

$$E_t\left(R_{t+1}\right) = \delta \mathrm{Var}_t\left(R_{t+1}\right), \tag{9}$$

where δ is the coefficient of relative risk aversion of the representative agent.

The empirical literature examining the expected return–volatility relationship is vast. Typically GARCH-M models have been employed, and depending on the market, the sample period, and the exact model specification, conflicting results have been obtained. For instance, using monthly U.S. data French et al. (1987) and Campbell and Hentschel (1992) found a predominantly positive but insignificant relationship, while Glosten et al. (1993) found a negative and significant relationship employing an extended GARCH-M model that allows negative and positive shocks to have different effect on the conditional variance. Recently, Ghysels et al. (2003) argued that monthly data are insufficient to accurately estimate the expected return–volatility trade-off and succeeded in uncovering a significantly positive relationship through a new method combining data sampled at different frequencies. Their mixed data sampling (MIDAS) estimator is, however, rather complicated, and as our empirical results below show, also models confined to monthly data can produce results in support of the ICAPM relationship.

In what follows we show that conditional skewness has a central role to play in uncovering the expected return–volatility relationship. The presence of conditional and unconditional skewness has been documented in a number of previous empirical studies.[1] Campbell and Hentschel (1992) and Harvey and Siddique (1999) also incorporated conditional skewness in various GARCH-M specifications to examine the expected return–volatility trade-off. Theoretically the conditional skewness can be explained by the so-called volatility feedback effect (Campbell and Hentschel (1992)) that relies on volatility persistence and a positive intertemporal relation between expected return and conditional variance. This effect arises as follows. Because of persistence, a large piece of news increases not only present but also future volatility, which in turn increases the required rate of return on stock and, hence, lowers the stock price. This effect amplifies the impact of bad news but dampens the impact of good news, and therefore, large negative stock returns tend to occur more frequently than large positive ones when volatility is high. As a result, also the unconditional return distribution tends to be left-skewed.

Of the studies mentioned above, the paper by Harvey and Siddique (1999) comes closest to our approach. Also their models allowed for time-varying conditional skewness in a GARCH-M model for stock returns, but they failed to find a significantly positive relationship between expected returns and conditional variance in U.S. data. Harvey and Siddique (1999) employed variants of Hansen's (1994) autoregressive conditional density model with a skewed version of the t distribution specified for the error term. The model extends the standard GARCH-M model by allowing the conditional skewness and degrees of freedom of the skewed t distribution to depend linearly on functions of lagged error terms. In our model, in contrast, the conditional skewness is directly dependent on conditional variance in line with the volatility feedback effect discussed above.

[1] Also theoretical asset pricing models explicitly incorporating conditional or unconditional skewness have been presented. See, e.g. Harvey and Siddique (2000) and references therein.

4.1 Empirical Results

We test the implication of the ICAP model given by equation (9) using monthly excess U.S. stock returns from January 1946 to December 2002. As a proxy for the market return we use the value-weighted CRSP index and the three-month Treasury bill rate as the risk-free interest rate. All the models for the excess return r_t to be estimated are obtained from the following general specification:

$$\begin{aligned} r_t &= \delta h_t + \kappa h_t^{1/2} + u_t \\ h_t &= \omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \gamma_1 I\left(\varepsilon_{t-1} < 0\right) u_{t-1}^2, \end{aligned} \tag{10}$$

where $u_t = h_t^{1/2}\left(\varepsilon_t - \kappa\right)$, $I(\cdot)$ is an indicator function and γ_1 deviates from zero only in the GJR-GARCH (Glosten et al. (1993)) specification where positive and negative shocks are allowed to have different effects on conditional variance. The innovation ε_t is assumed to follow either the t distribution with ν degrees of freedom or the z distribution (5). In the former case κ is set equal to zero, but in the case of the skewed z distribution, a nonzero κ centers the observed series such that δh_t can be interpreted as the conditional mean of r_t. In other words, in the case of the z distribution we set $\kappa = -\mu\left(a,b\right)/\sigma\left(a,b\right)$ (see Section 2). We also estimated the model with κ as a free parameter, but its estimate turned out to be very close to $-\mu\left(a,b\right)/\sigma\left(a,b\right)$ and the results hardly changed otherwise either (the p-value of a LR test for this restriction was 0.233). As far as the symmetric distributions are concerned, we also experimented with the standard normal distribution and the conclusions were qualitatively the same as with the t distribution, but the latter is preferred because of its ability to better capture the fat tails. As discussed in Section 2, the z distribution is not usable if kurtosis is extreme. This should not be any kind of limitation here, especially as we are dealing with monthly data; the excess kurtosis implied by the estimated t distribution barely exceeds unity and the corresponding figure for the z distribution is about 0.8.

Table 1 contains the estimation results of three GARCH-M specifications corresponding to equation (9). Note that in line with the theoretical ICAPM model, the specifications have no intercept in the conditional mean equation; models with a nonzero intercept were also estimated, but the additional parameter turned out to be insignificant at any reasonable significance level in all cases. The results for the GARCH-M-t and the corresponding asymmetric GJR-GARCH-t models confirm the findings in the previous literature. The estimates obtained for δ are positive as expected but, due to huge standard errors, clearly insignificant. In contrast, for the GARCH-z specification we obtain a positive and significant coefficient (p-value 0.0002 based on asymptotic normality). The magnitude of the estimate, 3.377, also falls within the range previously obtained for the coefficient of relative risk aversion of the representative agent (see, for instance, Hall (1988) and references therein). Furthermore, this result is in line with the recent MIDAS estimates of Ghysels et al. (2003), indicating that a significantly positive relation between risk and return in the stock market can be uncovered even from merely monthly data once the error distribution is appropriately specified.

Because the null hypothesis $a = b$ is clearly rejected by the LR test (p-value 3.123e-8) our model implies significant conditional skewness which increases with conditional volatility. Moreover, because $\widehat{a} > \widehat{b}$ the conditional skewness is negative as expected based on the discussion on the volatility feedback effect above the point estimate of the coefficient of skewness of the error term ε_t was –0.428. Thus, the GARCH-z model captures the feature that large negative shocks, and hence returns, are more likely than positive ones when conditional variance is high.[2]

[2] In a related application to daily U.S. stock returns from 1885 through 1997, significant negative skewness was also found by Jensen and Lunde (2001). These authors used a model based on the normal inverse Gaussian distribution (cf. section 2) but their model for conditional mean was different form ours. Instead of the conditional variance used here, it contained the conditional standard deviation whose estimated effect on expected returns turned out to be negative. This result is consistent with the fact that the sign of the related parameter is determined by the skewness of the conditional distribution and it can probably be attributed to the specification used for the conditional mean. From economic point of view, the obtained result cannot be interpreted in the same way as our result because the conditional mean was specified differently and because pure

The coefficient γ_1 is positive and significant (at the 5% level) in the GJR-GARCH specification, indicating that negative shocks have stronger impact on the conditional variance than positive shocks. Although perhaps not so obvious, the GARCH-z model is also capable of capturing similar asymmetry, which can be seen by examining the news impact curve (NIC) implied by the model. Originally Engle and Ng (1993) defined the NIC as

$$E\left(h_{t+1}|h_t = h, u_t = \lambda\right),$$

i.e., the expectation of the conditional variance next period conditional on a current shock of size λ, where the shock is taken to be the error term u_t. Using this definition we could write the NIC of the GARCH-z model as

$$NIC\left(h_{t+1}|h_t = h, u_t = \lambda\right) = \omega + \alpha_1 \lambda^2 + \beta_1 h,$$

i.e., similar to the NIC of the GARCH-t model. However, we find it more natural to define the shock as the innovation ε_t in which case the NIC of the GARCH-z model becomes

$$NIC\left(h_{t+1}|h_t = h, \varepsilon_t = \theta\right) = \omega + \alpha h \left(\theta - \kappa\right)^2 + \beta_1 h.$$

This expression shows that if the innovation is defined as news, this NIC is asymmetric in the same way as that of the GJR-GARCH model. The news impact curves of the three estimated model specifications computed with ε_t as the shock are depicted in Figure 1. The NIC's of the GARCH-t and GJR-GARCH models are as expected with negative shocks having greater impact on volatility in the asymmetric specification. The shape of the NIC of the GARCH-z model is similar but the difference between the effects of large negative and positive shocks is even greater than in the GJR-GARCH specification. Moreover, the NIC does not take minimum at zero but at 0.8, potentially suggesting that slightly positive news is required for the market to be as tranquil as possible while 'no news' causes higher volatility.

returns instead of excess returns were used. (The fact that Jensen and Lunde (2001) used a different specification for the conditional variance is hardly of any major importance in this respect.)

The dynamics of the different GARCH models can be studied by computing their cumulative impulse response functions

$$IRF\,(h, s, \theta) = E(h_{t+s}|h_t = h, \varepsilon_t = \theta) - E(h_{t+s}|h_t = h),$$

i.e., the effects of a shock of size θ s periods ahead for different values of s. These are depicted in Figure 2 for a unit shock ($\theta = 1$). For comparison, also the impulse responses of Hansen's (1994) skewed-t model are graphed.[3] For the symmetric specifications the IRF's are simple to compute recursively while for the skewed models simulation methods are required. Furthermore, in the latter case the functions are dependent on the initial level of conditional variance and in the case of the GARCH-z model on the sign of the shock as well. For these models we consider three different values of h: 0.001, 0.005 and 0.01 are close to the minimum, average and maximum of the conditional variance implied by the estimated GARCH-M-z model, respectively. As expected, the influence of a shock is very persistent in the GARCH-t and GJR-GARCH-t models and also in Hansen's (1994) skewed t model. The initial impact of a shock does not depend on the initial conditional variance in Hansen's (1994) model, but different values of h yield somewhat different impulse response functions. For the GARCH-z model the decay of the impulse response functions is clearly faster, with the impact of the shock being very close to zero after 30 months. For a positive shock, the initial impact is the higher the smaller the conditional variance initially is, whereas the reverse holds for a negative shock. Moreover, a negative shock always has a higher initial impact than a positive shock so that a negative shock in turbulent times has the greatest impact, while a positive shock in turbulent times has the smallest impact. The level of initial conditional variance has little effect on the speed of decay of the impulse response function, though.

[3] We ended up with a specification where the degrees of freedom parameter is time-varying while the skewness parameter is constant. In this specification the estimate of the coefficient of relative risk aversion (corresponding to δ in (10)) equals 3.499 with standard error 0.939. Detailed estimation results are available upon request.

5 Conclusion

This paper has clearly demonstrated the importance to allow for conditional skewness when modeling stock returns. The standard GARCH-M-t model and its asymmetric GJR-GARCH-t counterpart were totally incapable of uncovering the expected positive relationship between monthly excess U.S. stock returns and risk. A different result was obtained when a GARCH-M model based on a probability distribution capable of allowing for skewness was applied. Then the expected positive relationship was significant at all conventional significance levels and significant conditional skewness was also found.

In this paper skewness was modeled by using the z distribution which can be thought of as an analytically simple special case of the family of a variance-mean mixtures of normal distributions. As in Andersson (2001) and Jensen and Lunde (2001), one may also consider other members of this family. Care is needed in the specification of the conditional mean, however, because different specifications can lead to very different results and conclusions. Our specification for the conditional mean was guided by the ICAPM model whereas Jensen and Lunde (2001) used another specification and did not obtain results with economically meaningful interpretation.

References

Abramowitz, M., and I. Stegun (1972), Handbook of Mathematical Functions. Dover Publications. New York.

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

Barndorff-Nielsen, O., J. Kent, and M. Sørensen (1982), Normal variance-mean mixtures and z distributions. International Statistical Review 50, 145–159.

Campbell, J.Y., and L. Hentschel (1992), No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Carrasco, M., and X. Chen (2002), Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18, 17-39.

Christoffersen, P., S. Heston, and K. Jacobs (2003), Option valuation with conditional skewness. Working paper, McGill University.

Engle, R.F., and V.K. Ng (1993), Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

French, K.R., G.W. Schwert, and R.F. Stambaugh (1987), Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Ghysel, E., Santa-Clara, P., and R. Valkanov (2003), There is a risk-return tradeoff after all. Working paper 10-03 Finance Department, University of California, Los Angeles.

Glosten, L.R., R. Jagannathan, and D.E. Runkle (1993), On the relation between the expected value of the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Hall, R.E. (1988), Intertemporal substitution in consumption. Journal of Political Economy 96, 339–357.

Hansen, B.E. (1994), Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Harvey, C.R., and A. Siddique (1999), Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34, 465–487.

Harvey, C.R., and A. Siddique (2000), Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.

Jensen, M.B. and A. Lunde (2001), The NIG-S&ARCH model: A fat tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. Econometrics Journal 4, 319-342.

Kalimipalli, M., and R. Sivakumar (2003), Coes skewness matter? Evidence from the index options market. Working paper, Wilfrid Laurier University.

Lundbergh, S., and T. Teräsvirta (2002), Evaluating GARCH models. Journal of Econometrics 110, 417–435.

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

Wooldridge, J.M. (1990), A unified approach to robust, regression-based specification tests. Econometric Theory 6, 17–43.

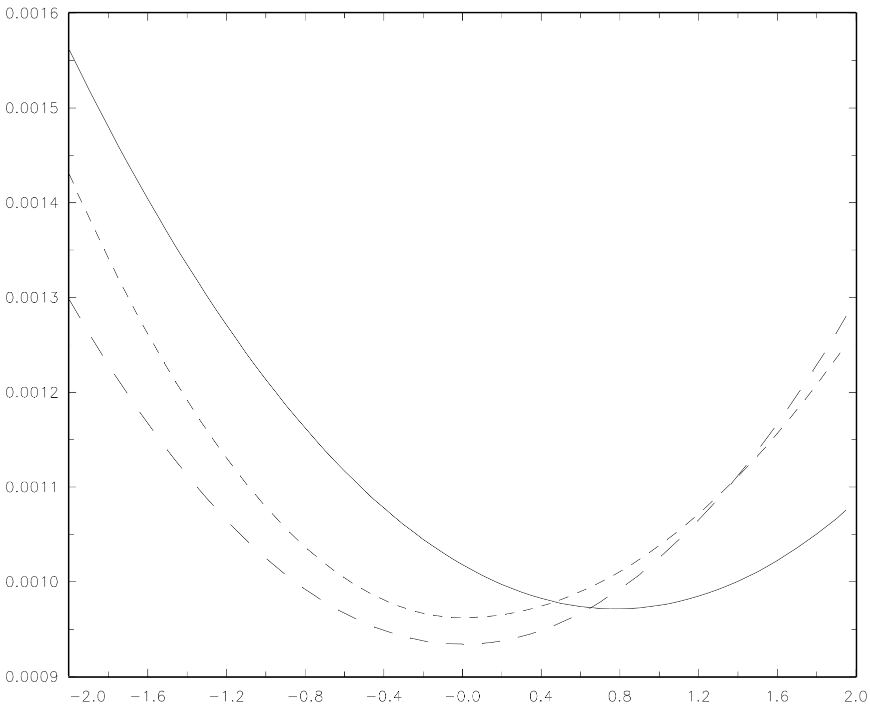

Figure 1: News impact curves of the GARCH(1,1)-M-z (solid line), GARCH(1,1)-M-t (long dashes) and GJR-GARCH(1,1)-M-t (short dashes) models.

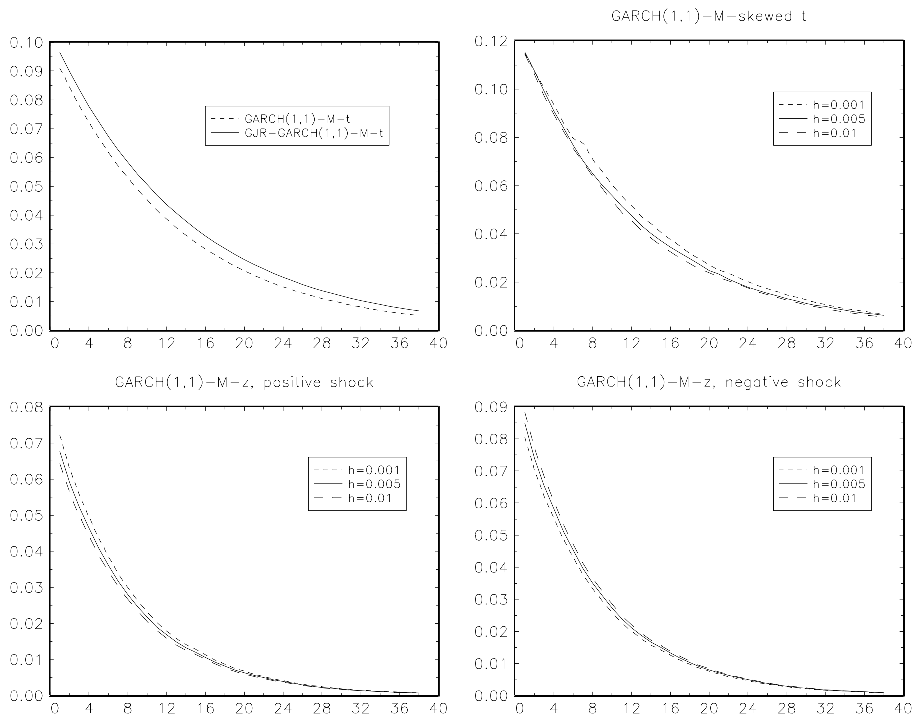

Figure 2: Impulse reponse functions implied by different GARCH models.

Table 1: Results of the GARCH(1,1)-M-z and GARCH(1,1)-M-t models for the excess stock return series.

	GARCH(1,1)-M-z	GARCH(1,1)-M-t	GJR-GARCH(1,1)-M-t
δ	3.377	4.584	3.936
	(0.966)	(87.88)	(47.46)
ω	0.0002	0.0001	0.0002
	(0.0001)	(6.483e-5)	(8.942e-5)
α_1	0.076	0.091	0.076
	(0.021)	(0.028)	(0.029)
β_1	0.761	0.834	0.762
	(0.065)	(0.050)	(0.074)
γ_1			0.041
			(0.019)
a	1.564		
	(0.599)		
b	3.128		
	(1.197)		
ν		10.218	11.731
		(3.158)	(3.776)
log likelihood	1222.54	1209.53	1216.65
AR(1)[a]	0.852	0.189	0.373
ARCH(10)[b]	0.420	0.520	0.232

The figures in the parentheses are standard errors computed from the inverse of the final Hessian matrix. The figures reported for the diagnostic tests are marginal significance levels.

[a]The alternative model is the corresponding AR(1)-GARCH(1,1)-M model, and under the null hypothesis of no remaining autocorrelation the coefficient of the AR(1) term equals zero. The test is robustified against misspecified conditional variance following Wooldridge (1990, Example 3.3).

[b]A test for remaining ARCH of order 10. For details see Lundbergh and Teräsvirta (2002).