



**EUROPEAN UNIVERSITY INSTITUTE**  
**DEPARTMENT OF ECONOMICS**

EUI Working Paper **ECO** No. 2005 /14

## Modeling Conditional Skewness in Stock Returns

MARKKU LANNE

and

PENTTI SAIKKONEN

BADIA FIESOLANA, SAN DOMENICO (FI)

All rights reserved.  
No part of this paper may be reproduced in any form  
Without permission of the author(s).

©2005 Markku Lanne and Pentti Saikkonen  
Published in Italy in October 2005  
European University Institute  
Badia Fiesolana  
I-50016 San Domenico (FI)  
Italy

# Modeling Conditional Skewness in Stock Returns<sup>1</sup>

Markku Lanne

University of Jyväskylä, RUESG and HECER

Pentti Saikkonen

University of Helsinki, RUESG and HECER

<sup>1</sup>We are grateful to the Research Unit of Economic Structures and Growth (RUESG) in the University of Helsinki and the Yrjö Jahnsson Foundation for financial support. The second author also acknowledges financial support from the Academy of Finland, and the Alexander von Humboldt Foundation under a Humboldt research award. Part of his research was done while he was visiting the Department of Economics, European University Institute, Florence, whose hospitality is gratefully acknowledged. Correspondence to: Markku Lanne, School of Business and Economics, P.O. Box 35, FIN-40014 University of Jyväskylä, FINLAND. Tel. +358142603190, Fax +358142603331, E-mail: markku.lanne@econ.jyu.fi.

## **Abstract**

In this paper we propose a new GARCH-in-Mean (GARCH-M) model allowing for conditional skewness. The model is based on the so-called z distribution capable of modeling moderate skewness and kurtosis typically encountered in stock return series. The need to allow for skewness can also be readily tested. Our empirical results indicate the presence of conditional skewness in the postwar U.S. stock returns. Small positive news is also found to have a smaller impact on conditional variance than no news at all. Moreover, the symmetric GARCH-M model not allowing for conditional skewness is found to systematically overpredict conditional variance and average excess returns.

JEL Classification: C16, C22, G12

Key words: Conditional skewness, GARCH-in-Mean, Risk-return tradeoff

## 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 explicitly 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. The model is inspired by the so-called volatility feedback effect (Campbell and Hentschel, 1992) that has been offered as an explanation to the presence of conditional left-skewness observed in stock returns. In line with this effect, the model imposes comovement of conditional skewness and conditional variance. Volatility feedback 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. This effect is also capable of explaining the observed left-skewness of unconditional return distributions.

Although the literature is not voluminous, there are a number of recent papers focusing on conditional and unconditional skewness in stock returns. It has even been suggested that conditional skewness is a priced risk factor (see Harvey and Siddique, 2000, and the references therein), while we merely argue that unmodeled skewness may affect inference on other parameters of the model, leading to biased pricing implications. In addition to asset pricing, another field where it is important to take potential conditional skewness properly into account, is risk management, i.e., risk measurement and pricing of derivative securities. These applications often rely on simulation methods that require data generating processes accurately describing the behavior of asset returns (see, e.g., Kalimipalli and Sivakumar, 2003, and Christoffersen et al., 2003).

In this paper we consider a GARCH-in-Mean (GARCH-M) model based on 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. It 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. Our results indicate the presence of conditional skewness in U.S. stock returns. It is also found that the news impact curve is not minimized at zero, but small positive news seem to have the smallest impact on the conditional variance. This goes contrary to most previous results according to which 'no news is good news', with Anderson et al. (1999) as a notable exception. Moreover, allowing for conditional skewness seems to greatly affect the magnitude of the conditional variance and risk premia predicted by GARCH models. In our data set, the GARCH-M model based on the symmetric t distribution is shown to yield systematically too high values of both of these, whereas the model based on the z distribution is strikingly accurate. As a potential explanation the results suggest that the GARCH-M model based on a symmetric error distribution is driven by highly volatile observations, and hence, tends to overprice assets.

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

## 2 Asset Pricing and Conditional Skewness

Several studies have examined the relationship between the expected return and conditional variance of stock returns with Mertons's (1973) Intertemporal Capital Asset Pricing Model (ICAPM) as a starting point. According to this model the expectation of the excess return on the stock market,  $r_t$ , depends positively on its conditional variance:

$$E_t(r_t) = \delta \text{Var}_{t-1}(r_t), \quad (1)$$

where  $\delta$  is assumed positive and can be interpreted as 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 allowing for the leverage effect. Even though, theoretically, there should be no intercept term in equation (1), virtually all previous studies have included one, which may explain the ambiguous results. Namely, Lanne and Saikkonen (2005) have recently shown that the unnecessary inclusion of an intercept term leads to inaccurate estimation and very low power in Wald tests of the null hypothesis  $\delta = 0$ .

In the empirical part of the paper we show that conditional skewness is present in stock returns and explicitly allowing for it has a big effect on the estimates of  $\delta$ . The presence of conditional and unconditional skewness has been documented in a number of previous empirical studies. 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. Like our model below, 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, which may be attributed to the inclusion of the intercept term in the mean equation, as discussed above. Harvey and Siddique (1999) employed variants of Hansen’s (1994) autoregressive conditional density model, which is probably the most prominent GARCH specification allowing for conditional skewness in the previous literature. 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, and, hence, it lends itself to clear economic interpretation, in line with the volatility feedback effect discussed above. A potential drawback of Hansen’s (1994) model is that it is not very parsimonious and it may be difficult to find an adequate specification for the degrees of freedom parameter, as the empirical examples of Hansen (1994) illustrate. Moreover, transformations due to the parameter constraints imposed by the t distribution may not facilitate straightforward interpretation of the connection between the conditioning variables and time-varying parameters.

## 2.1 GARCH-M-z Model

As a starting point for modeling the excess stock return  $r_t$  we have the following general GARCH-M model

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \delta h_t + h_t^{1/2} \varepsilon_t, \quad (2)$$

where  $\phi_0, \dots, \phi_p$  and  $\delta$  are real valued parameters,  $\varepsilon_t$  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  $r_t$ . Independence of  $h_{t-j}$  ( $j > 0$ ) and  $\varepsilon_t$  is also assumed and, for stationarity, the

roots of the polynomial  $1 - \phi_1 z - \dots - \phi_p z^p$  are required to lie outside the unit circle. Different versions of this model have been employed in the previous empirical literature. In applications to low-frequency data lagged returns are rarely needed. Also, as mentioned above, according to the ICAPM there should be no intercept term and unnecessarily including one may obscure the results. Therefore, from now on, we restrict  $\phi_0$  to zero. 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  $\varepsilon_t$ .

The distribution we are going to apply is the so-called  $z$  distribution. This distribution has been studied by Barndorff-Nielsen et al. (1982) who show that it can be represented as a normal variance-mean mixture with the mixing distribution an infinite convolution of exponential distributions. Other members of the family of variance-mean mixtures of normal distributions are the ordinary (symmetric)  $t$  distribution and its skewed version as well as the normal inverse Gaussian distribution which has recently been applied by Andersson (2001) and Jensen and Lunde (2001) to model conditional heteroskedasticity in stock returns. We refer to Barndorff-Nielsen et al. (1982) for more details of these distributions.

Except for the ordinary  $t$  distribution, the density functions of the distributions discussed above depend on a modified Bessel function. The  $z$  distribution, denoted by  $z(a, b, \sigma, \mu)$ , is analytically simpler and characterized by the density function

$$f(x) = \frac{1}{\sigma B(a, b)} \frac{\{\exp[(x - \mu)/\sigma]\}^a}{\{1 + \exp[(x - \mu)/\sigma]\}^{a+b}} \quad (x \in \mathbb{R}; a, b, \sigma > 0; \mu \in \mathbb{R}), \quad (3)$$

where  $B(\cdot, \cdot)$  is the beta function. Clearly,  $\mu$  is a location parameter and  $\sigma$  is a scale parameter. If  $a = b$  the distribution is symmetric whereas it is positively (negatively) skewed if  $a > b$  ( $b > a$ ). 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  $a = b = 1$ . The characteristic function of the

$z(a, b, \sigma, \mu)$  distribution is

$$\chi(s) = \frac{e^{it\mu} B(a + i\sigma s, b - i\sigma s)}{B(a, b)}. \quad (4)$$

We shall now consider moments of the  $z$  distribution. First, suppose that the random variable  $x$  has a  $z(a, b, 1, 0)$  distribution. From the characteristic function (4) 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  $\Psi^{(n)}(s) = d^n \Psi(s) / ds^n$  ( $n = 1, 2, \dots$ ). Then, the  $n$ th cumulant of  $x$ , denoted by  $\varkappa_n$ , is

$$\varkappa_n = \Psi^{(n-1)}(a) + (-1)^n \Psi^{(n-1)}(b), \quad n = 1, 2, \dots, \quad (5)$$

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

$$\begin{aligned} Ex &= \Psi(a) - \Psi(b) \stackrel{def}{=} \mu(a, b), \\ Var(x) &= \Psi'(a) + \Psi'(b) \stackrel{def}{=} \sigma^2(a, b), \\ E(x - Ex)^3 &= \Psi''(a) - \Psi''(b), \end{aligned}$$

and

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

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

To get an idea of the possible shapes of the  $z$  distribution, consider the symmetric  $z(\lambda, \lambda, 1, 0)$  distribution and note that the function  $\Psi^{(n)}(s)$  has the series representation  $\Psi^{(n)}(s) = (-1)^{n+1} n! \sum_{j=0}^{\infty} (s+j)^{-n-1}$  ( $n = 1, 2, \dots$ ) (see Abramowitz and Stegun (1972, result 6.4.10)). Using this result and the preceding expression of the fourth central moment of the  $z(a, b, 1, 0)$  distribution it is not difficult to show that the excess kurtosis of the  $z(\lambda, \lambda, 1, 0)$  distribution is a decreasing function of  $\lambda$  and approaches three as  $\lambda$  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  $z(a, b, 1, 0)$  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 the  $z$  distribution. It is worth noting, however, that for us these limits of the skewness and kurtosis are only relevant for the conditional but not for the unconditional distribution of the considered series. Indeed, because a  $z$  distribution is specified for the error term  $\varepsilon_t$  in (2) the unconditional skewness and kurtosis of  $r_t$  are generally larger than their conditional counterparts, and therefore, the limits are unlikely to be restrictive in applications to typical financial time series.<sup>1</sup>

As already mentioned, we shall assume that the error term  $\varepsilon_t$  in (2) has a  $z$  distribution. Because  $\varepsilon_t$  is an error term we want it to have zero mean and, as common in GARCH and GARCH-M models, unit variance. Thus, we assume that

$$\varepsilon_t \sim z(a, b, 1/\sigma(a, b), -\mu(a, b)/\sigma(a, b)). \quad (6)$$

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  $\text{Var}(\varepsilon_t) = 1$ . Thus, the model we wish to consider is defined by (2) and (6). An alternative possibility to define the model is to specify the conditional distribution of  $r_t$  given its past. The result can be obtained from (2) and (6). In symbols we have

$$r_t \mid \mathcal{F}_{t-1} \sim z\left(a, b, h_t^{1/2}/\sigma(a, b), \mu_t(\varphi) - h_t^{1/2}\mu(a, b)/\sigma(a, b)\right), \quad (7)$$

where  $\mathcal{F}_{t-1} = \{r_{t-1}, r_{t-2}, \dots\}$  and  $\mu_t(\varphi) = \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \delta h_t$  with  $\varphi =$

---

<sup>1</sup>As an illustration, consider the standard GARCH(1,1) model  $y_t = h_t^{1/2}\varepsilon_t$  where  $h_t = \omega + 0.87h_{t-1} + 0.10y_{t-1}^2$  and  $\varepsilon_t \sim i.i.d(0, 1)$  has a symmetric distribution with excess kurtosis 2.5. The (positive) constant term  $\omega$  has no effect on the excess kurtosis of  $y_t$  which, from equation (8) of He and Teräsvirta (1999), is found to be 20.05. In this example the sum  $\alpha + \beta = 0.87 + 0.10 = 0.97$  which is quite relevant for many financial time series and even its small increase can lead to a large increase in the excess kurtosis of  $y_t$ .

$[\phi_1 \cdots \phi_p \delta]'$ . Clearly,  $\mu_t(\varphi)$  and  $h_t$  are the conditional mean and variance of  $r_t$ , respectively. If the distribution of  $\varepsilon_t$  is skewed, it is obvious from equation (2) that the conditional skewness of  $r_t$ , measured by the third central moment, increases with its conditional variance. 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 consider a slight extension of the standard GARCH model given by

$$h_t = \omega + \sum_{j=1}^r \beta_j h_{t-j} + \sum_{j=1}^q \alpha_j u_{t-j}^2, \quad (8)$$

where

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

with  $\kappa$  a real valued parameter. As usual, the parameters in (8) are supposed to satisfy  $\omega > 0$ ,  $\beta_j \geq 0$  and  $\alpha_j \geq 0$ . Because  $\mu_t(\varphi)$  is the conditional mean of  $r_t$  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 it 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 (7)). Compared to the standard specification  $u_t = r_t - \mu_t(\varphi)$  this choice of  $\kappa$  shifts the distribution of  $u_t$  to the left (right) when the skewness is negative (positive), implying that negative (positive) values of  $u_t$  contribute more to conditional heteroskedasticity than in the standard case. Of course, one can also specify  $\kappa$  as a free parameter and let the data decide its most appropriate value.

If the value of the parameter  $\kappa$  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

$$E \left( \beta_1 + \alpha_1 (\varepsilon_t - \kappa)^2 \right)^k < 1, \quad k \geq 1, \quad (9)$$

where  $k$  is an integer. Then, from Corollary 6 of Carrasco and Chen (2002) it follows that the process  $h_t$  ( $t = 1, 2, \dots$ ) can be given an initial distribution which makes it stationary and strong mixing (or even  $\beta$ -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  $r_t$  can be treated as a stationary process with  $E|r_t|^k < \infty$ . 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.

## 2.2 Parameter Estimation and Statistical Inference

Maximum likelihood (ML) estimation of the parameters of the model defined by equations (2), (6) and (8) is, in principle, straightforward. Suppose we have an observed time series  $r_t$ ,  $t = -l + 1, \dots, T$  where  $l$  denotes the required number of initial values. Then the conditional density of  $r_t$  ( $t \geq 1$ ) given the past values of the series can be obtained from (3) and (7). The result is

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

where, for simplicity,  $m_t(\theta) = \mu_t(\varphi) - \mu(a, b) h_t^{1/2} / \sigma(a, b)$  and  $\theta = [\varphi' \ \gamma' \ a \ b]'$  with  $\gamma = [\omega \ \beta_1 \ \dots \ \beta_r \ \alpha_1 \ \dots \ \alpha_q \ \kappa]'$ . Here  $\kappa$  is treated as a free parameter. The restrictions discussed after equation (8) can be handled in an obvious way. Conditional on the initial values, the logarithm of the likelihood function can thus be written as

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

The maximization of  $l_T(\theta)$  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  $r_t$  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  $\theta$ , denoted by  $\hat{\theta}$ , can be treated as approximately normally distributed with mean value  $\theta$  and covariance matrix  $-(E\partial^2 l_T(\theta)/\partial\theta\partial\theta')^{-1}$ . Approximate standard errors of the components of  $\hat{\theta}$  can therefore be obtained by taking the square roots of the diagonal elements of  $-(\partial^2 l_T(\hat{\theta})/\partial\theta\partial\theta')^{-1}$ . Likelihood ratio, Wald, and Lagrange multiplier tests with approximate chi square distributions can also be performed in the usual way.

### 3 Conditional Skewness in U.S. Stock Returns

We estimate GARCH-M models based on the implication of the ICAP model in equation (1) 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. In particular, we consider the following special case of model (8) for the excess return  $r_t$ ,

$$\begin{aligned} r_t &= \delta h_t + \kappa h_t^{1/2} + u_t \\ h_t &= \omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned} \tag{10}$$

where  $u_t = h_t^{1/2}(\varepsilon_t - \kappa)$ . In line with the previous literature, GARCH(1,1) specification turns out to be adequate. The innovation  $\varepsilon_t$  is assumed to follow either the  $t$  distribution with  $\nu$  degrees of freedom or the  $z$  distribution (6). In the former case the value of  $\kappa$  is assumed to be zero, but in the case of the skewed  $z$  distribution, this is not done. Specifically, in that case we set  $\kappa = -\mu(a, b)/\sigma(a, b)$  which means that in the model for conditional variance the observed series is centered by using the location parameter of the  $z$  distribution assumed for the error term  $\varepsilon_t$  (see Section 2). We also estimated the model with  $\kappa$  as a free parameter, but its estimate turned out to be very close to  $-\mu(a, b)/\sigma(a, b)$  and the results remained virtually intact (the

p-value of a LR test for this restriction is 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.

Table 1 contains the estimation results of two GARCH-M specifications corresponding to equation (1). In each case the estimate of  $\delta$  is positive and significant as implied by Merton's ICAPM. In the symmetric GARCH-t model the estimate of  $\delta$  is considerably greater (4.584) than in the GARCH-z specification allowing for conditional skewness (3.377). Both estimates are reasonable compared to most previous estimates of the relative risk aversion of the representative agent (see Hall, 1988). However, as will be shown below, the estimated models are quite different in terms of fit and average predicted excess returns.

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  $\hat{a} < \hat{b}$  the conditional skewness is negative as expected based on the discussion on the volatility feedback effect in Section 2. 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. The point estimate of the coefficient of skewness of the error term  $\varepsilon_t$  is  $-0.43$  which is well within reach of the z distribution (see Section 2.1). The same can be said about the kurtosis. The estimated excess kurtosis of the error term  $\varepsilon_t$  is only about 0.8 (the corresponding figure implied by the estimated t distribution barely exceeds unity).

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 2.1) but their model for conditional mean was different from 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. The main advantage of our model over that of Jensen and Lunde (2001) is that it allows for separately estimat-

ing the conditional skewness and GARCH-in-mean effect. In Jensen and Lunde’s (2001) model a single parameter determines both. Because conditional skewness is expected to be negative and the GARCH-in-mean effect positive for stock returns, interpreting their estimation results is ambiguous. They explain the negative estimate as “the need to account for negative skewness in the return distribution dominating the GARCH-in-mean effect”. Moreover, from economic point of view, the obtained result cannot be interpreted in the same way as ours because the conditional mean was specified differently and because pure returns instead of excess returns were used.

According to the diagnostic tests in Table 1 the specification is adequate: there is no unmodeled autocorrelation or autoregressive conditional heteroskedasticity in the residuals. The estimated model is also stationary as the estimated value of the left hand side of inequality (9) with  $k = 1$  is 0.883. Further evidence on the fit is provided in Figure 1 that depicts a plot of the logarithmic density of the residuals against its theoretical counterpart. The logarithmic scale is useful in detecting deviations on the extreme tails. With the exception of the left tail the differences are minor, and the discrepancy is caused by a single observation (October 1987 stock market crash). The model was also estimated without this exceptional observation, but the results remained virtually the same. Also, the conclusions are not reversed by using robust standard errors.

To further check the fit of the GARCH-z and GARCH-t specifications, we compared the conditional variance series predicted by each model to the monthly realized variance obtained by summing squared daily returns over each month. This comparison is restricted to the period beginning in July 1963 due to availability of data. Two criteria, the mean square error (MSE) and mean absolute error (MAE) are employed, and the significance of the differences is tested by means of the Diebold-Mariano (1995) test. As the results in Table 2 show, the GARCH-z specification is more accurate in terms of both criteria at least at the 5% level of significance. In addition to the overall comparison comprising all observations, figures for three categories of equal size sorted by the magnitude of the realized variance are reported. It is seen that,

even though the GARCH-z model has a better fit in each category, the differences are not significant when the realized variance is high. It is thus only in the less volatile periods, especially in the medium volatility category, that there are significant differences between the two specifications. In the low and middle variance ranges the conditional variance predictions of the GARCH-t model are systematically too high compared both to realized variance and the predictions of the GARCH-z model.<sup>2</sup> This follows because the symmetric model predicts high conditional variance after large shocks, irrespective of their sign, and is thus likely to predict too high variance after large positive shocks. Because of volatility clustering, this is a somewhat lesser problem when volatility is low. Systematically higher conditional variance coupled with the higher estimate of  $\delta$  in the GARCH-t model translates into too high risk premia in the middle volatility range. It is likely that the estimation is driven by the influential observations in the high-volatility range, and for the GARCH-t model to be able to capture the risk-return tradeoff correctly there,  $\delta$  must be great. Hence, the estimate of  $\delta$  is far too great for the observations in the low and middle volatility ranges. These findings indicate that the GARCH-z model is more flexible due to its ability to predict high volatility only following a large negative shock, in line with the volatility feedback effect.

In addition to conditional variance comparisons, also the predicted average excess returns can be compared to the actual average excess return. Already French et al. (1987) pointed out that the GARCH-t predictions of average excess returns are far too high. This finding is verified in Table 3 which shows that the average excess return predicted by the GARCH-t model is over 42% higher than the actual value. For the

---

<sup>2</sup>Out of the 158 observations in the middle range, for 152 the conditional variance predicted by the GARCH-t model is greater than the realized variance and for 93 of these greater than the conditional variance predicted by the GARCH-z model. In the low variance range the GARCH-t model always predicts too high conditional variance and for 87 observations the prediction exceeds that of the GARCH-z specification. In the high volatility range the corresponding figures are 69 and 78, indicating no systematic pattern.

GARCH-z model the corresponding difference is negligible (5%). As a robustness check the sample period was divided into two periods of approximately equal length. As far as the GARCH-t model is concerned, the relative difference is virtually the same in both subsample periods, whereas the GARCH-z model yields a better fit in the 1975–2002 period compared to the 1946–1974 period or the entire sample.

Because of the asymmetry inherent in the GARCH-z model, shocks of different size and sign have different effects on the conditional variance. This is revealed by the news impact curve (NIC) of the estimated model. Originally Engle and Ng (1993) defined the NIC as

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

i.e., the expectation of the conditional variance next period conditional on a current shock of size  $\lambda$ , 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

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

i.e., similar to the NIC of the GARCH-t model. However, following Anderson et al. (1999), we find it more natural to define the shock as the innovation  $\varepsilon_t$  in which case the NIC of the GARCH-z model becomes

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

This expression shows that if the innovation is defined as standardized news, the NIC is asymmetric. The news impact curves of the estimated model specifications computed with  $\varepsilon_t$  as the shock are depicted in Figure 2. The NIC of the GARCH-t model is, of course, symmetric around zero, while in the GARCH-z model large negative shocks have greater impact on the conditional variance than large positive shocks. Moreover, the NIC does not take a minimum at zero but at 0.8, suggesting that slightly positive news is required for the market to be as tranquil as possible while 'no news' causes higher volatility. This is in line with the findings of Anderson et al. (1999) who fitted a smooth transition GARCH model to daily US stock returns from

January 1990 to October 1995. It is worth pointing out that this result is not obtained by using Hansen's (1994) autoregressive conditional density model with a skewed version of the  $t$  distribution. Even though that model is capable of capturing conditional skewness, its NIC has properties similar to those of the symmetric GARCH- $t$  model. In Figure 2 the NIC's are plotted for  $h = 0.01$ . A change in the value of  $h$  would only move the NIC of the symmetric GARCH model vertically, whereas the NIC of the GARCH- $z$  model becomes flatter with decreasing  $h$ . In other words, according to that model the impact of any kind of news is nearly the same in very tranquil times, while the discrepancy between the impact of different kinds of shocks increases with increasing volatility.

It is noteworthy that even though some previous GARCH specifications, such as the GJR-GARCH (Glosten et al., 1993) and EGARCH (Nelson, 1991), also imply asymmetric news impact curves, it is only the so-called leverage effect (Black, 1976, and Christie, 1982) that they are capable of capturing. This effect was first pointed out by Black (1976) and Christie (1982) who suggested that an increase in risk and, hence, volatility is followed by a negative shock causing an increase in financial leverage due to a drop in the stock price. The general conclusion in the previous literature is, however, that the leverage effect does not play a major role in explaining the observed asymmetry (see, e.g., Bae et al., 2004, and the references therein). As a check, we also fitted a GJR-GARCH-M- $t$  model to the excess return data, but the results did not deviate much from those of the GARCH-M- $t$  model, and are thus not reported. This outcome lends support to the finding in the previous literature that the volatility feedback effect is more important than the leverage effect in explaining asymmetries in conditional volatility.

#### **4 Conclusion**

This paper has clearly demonstrated the importance of allowing for conditional skewness when modeling stock returns. We modeled skewness 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. At least in the postwar U.S. stock returns the conditional variance and risk premia predicted by the GARCH-M model based on the symmetric  $t$  distribution are systematically too high. In contrast, the conditional variance predicted by the GARCH-M model based on the skewed  $z$  distribution turned out to be much closer to the realized variance and the deviation of the average predicted excess return from the actual value is strikingly small. These findings lend support to the usefulness of the more flexible specification for allowing conditional skewness. The results can also be interpreted in favor of volatility feedback as the form of conditional skewness incorporated in our model is in line with that effect.

## References

- Abramowitz, M., and I. Stegun (1972), *Handbook of Mathematical Functions*. Dover Publications. New York.
- Anderson, H.M., K. Nam, and F. Vahid (1999), Asymmetric nonlinear smooth transition GARCH models. In Rothman, P. (ed.), *Nonlinear Time Series Analysis of Economic and Financial Data*, Kluwer.
- Andersson, J. (2001), On the normal inverse Gaussian stochastic volatility model. *Journal of Business and Economic Statistics* 19, 44-54.
- Bae, J., C.-J. Kim, and C.R. Nelson (2004), Why are stock returns and volatility negatively correlated. Working paper, Yeungnam University, Korea University and University of Washington.
- Barndorff-Nielsen, O., J. Kent, and M. Sørensen (1982), Normal variance-mean mixtures and  $z$  distributions. *International Statistical Review* 50, 145–159.

Black, F. (1976), Studies of stock price volatility changes. Proceedings of the 1976 Meeting of the American Statistical Association, Business and Economical Statistics Section, 177–181.

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.

Christie, A.A. (1982), The stochastic behavior of common stock variances: Value, leverage, and interest rate effects. *Journal of Financial Economics* 10, 407–432.

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.

Davidson, J. (1994), *Stochastic Limit Theory*. Oxford University Press. Oxford.

Diebold, F.X., and R.S. Mariano (1995), Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–265.

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.

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.
- He, C., and T. Teräsvirta (1999), Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92, 173–192.
- 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.
- Lanne, M., and P. Saikkonen (2005), Why is it so difficult to uncover the risk-return tradeoff in stock returns? University of Jyväskylä, School of Business and Economics Working Paper 294/2005.
- 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.
- Nelson, D.B. (1991), Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
- Wooldridge, J.M. (1990), A unified approach to robust, regression-based specification tests. *Econometric Theory* 6, 17–43.

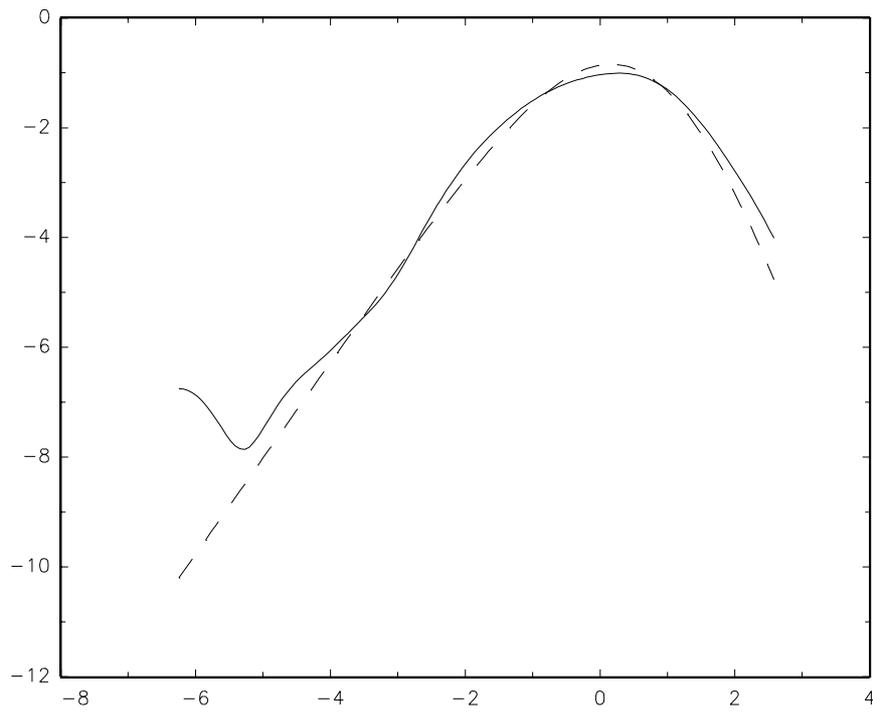


Figure 1: Plot of the logarithmic density of the residuals of the GARCH(1,1)-M-z model (solid line) against the theoretical logarithmic density (dashes).

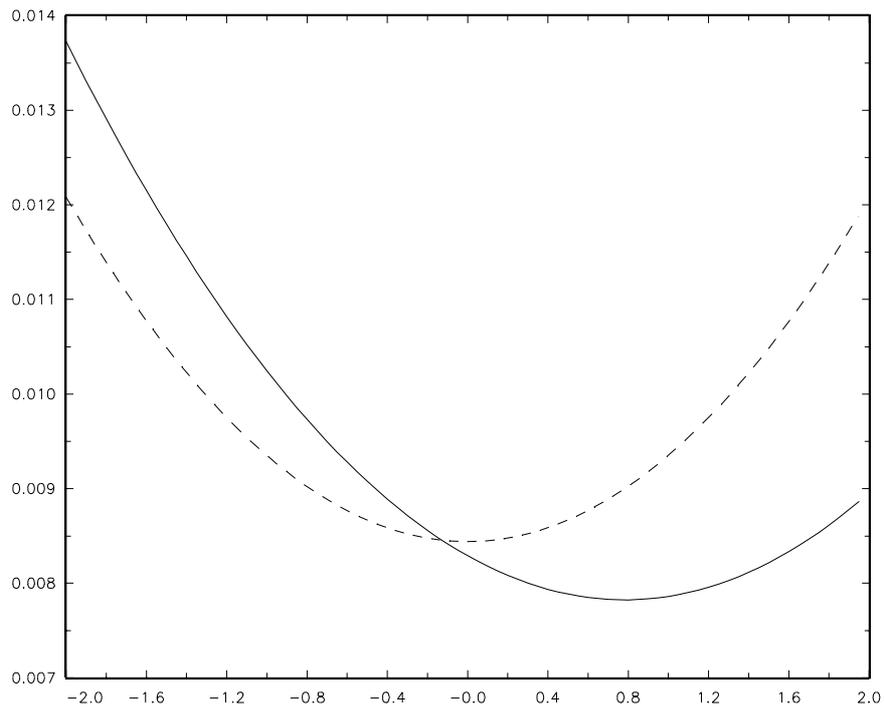


Figure 2: News impact curves of the GARCH(1,1)-M-z (solid line) and GARCH(1,1)-M-t (dashes) models.

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
$\delta$	3.377 (0.966)	4.584 (0.939)
$\omega$	0.0002 (0.0001)	0.0001 (6.483e-5)
$\alpha_1$	0.076 (0.021)	0.091 (0.028)
$\beta_1$	0.761 (0.065)	0.834 (0.050)
$\gamma_1$		
$a$	1.564 (0.599)	
$b$	3.128 (1.197)	
$\nu$		10.218 (3.158)
log likelihood	1222.54	1209.53
AR(1) <sup>a</sup>	0.852	0.189
ARCH(10) <sup>b</sup>	0.420	0.520

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.

<sup>a</sup>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).

<sup>b</sup>A test for remaining ARCH of order 10. For details see Lundbergh and Teräsvirta (2002).

Table 2: Comparison of the predictive accuracy of the GARCH(1,1)-M-z and GARCH(1,1)-M-t models for the excess stock return series.

Criterion	Model	Realized variance			
		All	Low	Medium	High
MSE	GARCH-z	8.01e-6*	1.14e-6*	9.10e-7**	2.19e-5
	GARCH-t	8.15e-6	1.08e-6	1.03e-6	2.21e-5
MAE	GARCH-z	0.00121**	0.00100*	7.92e-4**	0.00175
	GARCH-t	0.00118	0.00103	8.53e-4	0.00176

Symbols \* and \*\* denote significance in the Diebold-Mariano test at the 5% and 1% levels, respectively.

Table 3: Average excess returns computed from the data and predicted by the GARCH(1,1)-M-z and GARCH(1,1)-M-t models.

	Actual	GARCH(1,1)-M-z	GARCH(1,1)-M-t
1946–2002	0.0059 (0.0429)	0.0062 (0.0029)	0.0084 (0.0034)
1946–1974	0.0053 (0.0396)	0.0057 (0.0026)	0.0076 (0.0026)
1975–2002	0.0066 (0.0461)	0.0066 (0.0030)	0.0093 (0.0038)

Standard errors in parentheses.