An extension of the NIG-S&ARCH model with an application to Value at Risk

Abstract

A new model for conditional skewness and kurtosis based on the Normal Inverse Gaussian (NIG) distribution is proposed. The new model and two previously used NIG models are evaluated by their Value at Risk (VaR) forecasts on a long series of daily Standard and Poor’s 500 returns. All three models perform very well compared to extant models and clearly outperform a Gaussian GARCH model. Moreover, the results show that only the new model produces satisfactory VaR forecasts for both 1% and 5% VaR.

Anders Wilhelmsson
Swedish School of Economics and Business Administration
Department of Finance and Statistics
and
Aarhus School of Business
Department of Marketing and Statistics
Haslegaardsvej 10
DK-8210 Aarhus V
Denmark

Email: Anders.Wilhelmsson@hanken.fi
Phone: +45 89 48 68 84

1 Financial support from the Research and Training Network “Microstructure of Financial Markets in Europe” is gratefully acknowledged. The author would like to thank Ole Barndorff-Nielsen, Asger Lunde, Esben Høg, Anders Ekholm, Peter Nyberg, participants at the Centre for Analytical Finance (CAF) members meeting 2006 and participants at the “Symposium för anvendt statistik” Copenhagen 2006, for helpful comments and suggestions.
I. Introduction

Realistic modeling of financial time series is of utmost importance in asset pricing and risk management. Empirical “facts” for equity returns that should be accounted for include left skewed leptokurtic return distributions and dependence in second moments. The second moment dependence, and to some extent the leptokurtosis, is addressed in the seminal article of Engle (1982). Among the models that account for the excess kurtosis not captured by the Gaussian GARCH (GARCH-n) model is the model of Barndorff-Nielsen (1997) based on the Normal Inverse Gaussian (NIG) distribution. This distribution, in addition to having nice analytical properties, can also be theoretically motivated from the mixture of distribution hypothesis of Clark (1973). Extensions of Barndorff-Nielsen’s model that allow for complex dynamics in the variance equation have been proposed by Andersson (2001), Jensen and Lunde (2001), as well as Forsberg and Bollerslev (2002). Jensen and Lunde (2001) also allow for leverage effects and nonzero skewness in the innovation distribution. However, recent studies by for example Harvey and Siddique (1999; 2000) indicate that there is also dependence in the skewness and possibly in the kurtosis of stock returns. This study therefore extends the NIG-S&ARCH model of Jensen and Lunde (2001) to model the dynamics not only in the variance but also in the skewness and kurtosis of the return distribution. Alternative models for conditional skewness and/or kurtosis are proposed by Hansen (1994), Harvey and Siddique (1999), Guermat and Harris (2002), Mittnik and Paolella (2003), Brännäs

The model proposed in this study has several advantages over previous models. The parameters that govern the shape of the distribution need not be restricted as opposed to Hansen (1994). Both the skewness and kurtosis are time varying, where in Harvey and Siddique (1999) and Lanne and Saikkonen (2005) only the skewness and in Guermat and Harris (2002) only the kurtosis is allowed to vary over time. The model has a closed form likelihood, making estimation easy, compared to e.g. Mittnik and Paolella (2003) and Níguez and Perote (2004) where the likelihood lacks an analytical expression. Furthermore, the NIG distribution can be motivated from economic theory and is in that sense not an ad hoc choice such as the Student t distribution.

In the initial estimation sample of daily Standard and Poor’s 500 returns ranging from July 3, 1962 to July 11, 1974, the new model shows a dramatic improvement in log likelihood value (-2,541.45) compared to -2,621.94 for the NIG-S&ARCH model. This is achieved at the cost of only three additional parameters.

The new model as well as the models of Jensen and Lunde (2001) and Forsberg and Bollerslev (2002) are then applied to compute Value at Risk (VaR) forecasts. VaR is the maximum loss expected to incur over a certain time period (h) with a given probability $\alpha$. With the adoption of Basel II, which allows banks to use internal VaR models for the purpose of regulating capital requirements there is much academic interest in the measure. For a survey see for example Duffie and Pan (1997) or the textbook treatment in Jorion (2000).
The VaR forecasts in this study are computed by rolling the estimation sample forward one day and re-estimating the parameters each day from July 12, 1974 to September 20, 2005, giving 7,878 forecasts, each based on the 3000 latest observations. A rolling scheme is preferred to extending the estimation sample due to possible shifts in the unconditional variance, see for example Mikosch and Starica (2004).

Since the capital requirements for a bank are directly effected by the number of VaR exceptions, i.e. the number of occasions when the actual loss is larger than predicted by the VaR model, evaluating VaR models by their ability to produce a correct number of exceptions (correct unconditional coverage) seems natural. However, Christoffersen (1998) points out that the exceptions from a correctly specified model should also be independently distributed over time. Using the terminology of Christoffersen (1998), a VaR model that has the correct number of exceptions that are also independent, is said to have correct conditional coverage.

The VaR forecasts in this study are hence examined using the testing methodology of Christoffersen (1998) and the recent advances by Christoffersen and Pelletier (2004) to evaluate both the conditional and unconditional coverage of the models. I find that the models based on the NIG-distribution perform very well with an almost perfect unconditional coverage both for 1% and 5% VaR. The NIG-S&ARCH-tv model proposed in this study as well as the models of Jensen and Lunde (2001) and Forsberg and Bollerslev (2002) are in the green zone as defined in Basel (1996), whereas a GARCH-n model is in the red zone. The green zone means that no additional capital requirements are necessary. The capital requirements given by a model in red zone have to be scaled upwards and also measures to improve the model must be taken immediately.
Comparing the results to Kuester et al. (2006) who evaluate 23 VaR models, including GARCH models with different error distributions and the CAViaR model of Engle and Manganelli (2004), on a NASDAQ sample of comparable size, I find that the GARCH-NIG, NIG-S&ARCH and NIG-S&ARCH-tv models in this study outperform all 23 models considered in Kuester et al. (2006) with regard to the conditional coverage for the 1% VaR. Furthermore the NIG-S&ARCH-tv model proposed in this study also beats all 23 models in Kuester et al. regarding conditional coverage for 5% VaR. The NIG-S&ARCH-tv is the only model that can not be rejected as providing the correct number of independent VaR exceptions for both 1% and 5% VaR. It should however be kept in mind that the Kuester et al. study is conducted on NASDAQ data, applying the NIG-S&ARCH-tv model also to this data would be interesting future work.

The rest of this article is structured as follows. Section II presents the theoretical foundation for describing financial returns using the NIG distribution. Section III presents the models whereas Section IV gives a brief introduction to Value at Risk and backtesting of VaR models. Section V describes the data and estimation. The results are presented in Section VI and Section VII summarizes and discusses the findings.
II. A theoretical motivation for the NIG-distribution

To capture the conventional characteristics of financial returns such as non-normality, conditional heteroscedasticity (Mandelbrot, (1963) and Fama, (1965)) and leverage effects for stocks (Black (1976)) a vast number of models have been proposed. Amongst the most successful are the GARCH-type models. For a review of these models see e.g. Bollerslev et al. (1994) or the collection of articles in Engle (1995). Previous research has shown these models to capture the persistence in volatility well. They also capture some but not all of the excess kurtosis in the data. To remedy this problem, alternative error distributions have been proposed. Among these are the Student-t (Bollerslev, (1987)), Generalized error distribution (Nelson, (1991)) and the skewed Student-t (Hansen, (1994)). The effect of different error distributions for estimation efficiency has recently been investigated in a simulation setting by Venter and de Jongh (2004), their results favored the NIG distribution for most of the data generation processes used.

Since the number of possible distributions to choose from is very large and since results are also dependent on the formulation of the mean and variance equation the number of possible combinations is daunting. This is true even if we restrict ourselves to the GARCH class models. For example Hansen and Lunde (2006) examine 330 different model specifications; despite this impressive number of models their study is far from being exhaustive. In view of this an alternative to an empirical or simulation based hunt for the best distribution is needed. The current study therefore pursues the use of a distribution that can be motivated from economic theory.
Consider the time $t$ price of a financial asset, for example a stock price, denoted by $P_t$, whose continuously compounded return over the unit time interval is given by $r_t = \log(P_t / P_{t-1})$ assuming possible dividends being added to the price. The mixture of distribution hypothesis (MDH) of Clark (1973) states that the conditional distribution of $r_t$ given a latent information arrival process (directing process) is normal. Traditionally the directing process has been assumed to follow a lognormal distribution resulting in a lognormal normal mixture distribution for the returns which unfortunately cannot be written in closed form.\(^2\) Instead, I follow Barndorff-Nielsen (1997) and assume the conditional mixing distribution, which is the distribution of the directing process, to be the Inverse Gaussian. That is, $\sigma_t^2 \mid \Omega_{t-1} \sim IG(\delta, \gamma)$ with $\Omega_t$ being the information set up to and including time $t$ information. Forsberg (2002) tests this assumption empirically on ECU/USD data and find that the inverse Gaussian distribution provides an even better fit for the variance than the log normal distribution. The density function of an IG-distributed variable $x$, is given by

$$f_\delta(x; \delta, \gamma) = \frac{\delta x^{-3/2}}{\sqrt{2\pi}} \exp \left[ - \frac{\delta^2}{2} \left( \frac{1}{\delta x^{-1} + \gamma^2 x} \right) \right].$$

\(^2\) Clark (1973) assumed an $iid$ log normal distribution. Later Taylor (1986) relaxed this assumption and let the variance, which proxies for the information arrival, follow an auto regression resulting in the Stochastic Volatility model.
The results in Barndorff-Nielsen (1977, 1978) then give that the unconditional distribution of \( r_t \) must be Normal Inverse Gaussian. In contrast to the lognormal normal mixture distribution the density of the NIG-distribution can be expressed in closed form. Ease of estimation is thus greatly enhanced and can be done by straightforward (numerical) maximum likelihood. The density function of the NIG distribution is given by

\[
f_x(x; \alpha, \beta, \mu, \delta) = \frac{\alpha}{\pi} \exp\left[\delta \sqrt{\alpha^2 - \beta^2} - \beta \mu\right] q\left(\frac{x - \mu}{\delta}\right)^{-1} \times K_1\left(\delta \alpha q\left(\frac{x - \mu}{\delta}\right)\right) \exp(\beta x),
\]

with \( 0 < |\beta| < \alpha \), \( \delta > 0 \) and \( q(x) = \sqrt{1 + x^2} \). \( K_1(\cdot) \) is the modified Bessel function of third order and index one. \( \alpha \) controls the kurtosis of the distribution and \( \beta \) the asymmetry. The location and scale of the distribution is decided by \( \mu \) and \( \delta \), respectively. The attractive features of the NIG distribution include the ability to fit leptokurtic and skewed data combined with nice analytical properties. In particular the NIG distribution is closed under convolution, for fixed values of \( \alpha \) and \( \beta \), meaning that if for example daily returns are NIG distributed then weekly returns will also be NIG distributed. For more details about the distribution including the moment generating function see Barndorff-Nielsen (1997) and references therein.

III. Presentation of models
The current study will mainly consider three models based on the NIG distribution. However a Gaussian GARCH (GARCH-n) and a Student-t Asymmetric Power GARCH (APARCH-t) model will also be included to facilitate comparison.

A. The NIG-S&ARCH-tv and NIG-S&ARCH models

To specify the dynamics of the variance, it is convenient to have the variance depend on a single parameter. This is done by using the location-scale invariant parameterization in Jensen and Lunde (2001) \( \alpha = a\delta, \beta = b\delta \), resulting in the density

\[
f_{x,t}(x;\alpha_t,\beta_t,\mu_t,\delta_t) = \frac{\alpha_t}{\pi\delta_t} \exp \left( \frac{\alpha_t^2}{\beta_t^2} + \beta_t \frac{(x-\mu)}{\delta_t} \right) q \left( \frac{(x-\mu)}{\delta_t} \right)^{-1} \times K_i \left( \frac{x-\mu}{\delta_t} \right)
\]

(3)

with \( 0 \leq \beta_t < \alpha_t, \delta_t > 0 \). Let \( \gamma_t = \sqrt{\alpha_t^2 - \beta_t^2} \epsilon_t \in \mathbb{R}^+ \) and \( \beta_t = (\beta_t / \alpha_t) \in [0,1] \).

Now specify the mean equation according to

\[
r_t = \mu + \gamma_t^{1/2} \delta_t \rho_t + \delta_t \eta_t,
\]

(4)

with \( \delta_t \eta_t = \epsilon_t \) and the distribution of \( \eta_t \) is NIG \( \left( \alpha_t, \beta_t, -\gamma_t^{1/2}, \beta_t / \alpha_t \right) \). The t subscripts on the parameters are added to indicate parameters that can vary over time. The purpose of the above specification used by Jensen and Lunde (2001) is that the mean and variance of
\( \eta \), will equal 0 and 1 respectively. Moreover, conditional mean and variance of the returns will be given by 
\[
E(r_t | \Omega_{t-1}) = \delta \gamma \rho_t + \mu \quad \text{and} \quad Var(r_t | \Omega_{t-1}) = \delta^2_t .
\]
This means that the return can be divided into three parts, a constant mean \( \mu \), a compensation for risk \( \delta \gamma \rho_t \) and a return innovation \( \varepsilon_t \). This sign of the risk compensation is given by the \( \rho_t \) parameter which as pointed out in Lanne and Saikkonen (2005) is a limitation since a positive compensation for risk is expected but to model negative skewness the \( \rho_t \) parameter must be negative. However, the specification in (4) is necessary to be able to model the conditional variance within the GARCH framework. Moreover there is no theoretical reason for why the risk compensation must be positive in an inter-temporal setting as pointed out in e.g. Glosten et al. (1993) and Abel (1988). The return innovations can in turn be divided into two parts following the basic idea in Engle (1982) to divide the innovation into one heteroskedastic part, \( \delta \), and one iid part, \( \eta \). In the NIG-S&ARCH-tv model the rescaled innovations \( \eta_t = \varepsilon_t / \delta_t \) are uncorrelated but not independent to each other since higher order dependence is present. This places the model in the semi-strong family of GARCH models in the terminology of Drost and Nijman (1993). The dynamics of \( \delta_t \) are given by

\[
(5) \quad \delta_t^\nu = c + b \delta_{t-1}^\nu + a \left[ \left| \varepsilon_{t-1} \right| - \tau \varepsilon_{t-1} \right]^{\nu},
\]

with \( c > 0, \quad b, a, \nu \geq 0 \quad \text{and} \quad -1 < \tau < 1 \). The specification of \( \delta_t \) in (5) is the Asymmetric Power ARCH specification proposed in Ding et al. (1993) and the sufficient condition for covariance stationarity is given by \( a E \left[ \left| \eta_t \right| - \tau \eta_t \right]^{\nu} + b < 1 \). This specification allows for different responses to positive and negative returns innovations of the same magnitude by
the \( \tau \) parameter. Furthermore, since there is no clear reason to either model the conditional standard deviation or the conditional variance, the \( \nu \) parameter decides what power of the conditional standard deviation, \( \delta_t \), to model. This specification of the variance dynamics is found successful for VaR forecasting in Mittnik and Paolella (2000) as well as in Giot and Laurent (2003) used with Student t and skewed Student t error distributions.

The current study adds to the literature by extending the original NIG-S&ARCH model of Jensen and Lunde (2001) to incorporate time variation in skewness and kurtosis. Recent interest in conditional higher moments (see inter alia, Harvey and Siddique, (1999, 2000), Dittmar, (2002), Guermat and Harris, (2002) as well as Christoffersen et al., (2006)) motivates this extension. This is done by the steepness and asymmetry parameters, given in Barndorff-Nielsen and Prause (2001), \( \xi = \left(1 + \overline{\gamma}\right)^{-1/2} \) which is closely related to the kurtosis and \( \chi = \rho \xi \) which is related to the skewness. The restriction \( 0 \leq |\overline{\beta}| < \overline{\alpha} \) makes the region for the attainable steepness and asymmetry a triangle in \( \mathbb{R}^2 \) given by \( \{(\chi, \xi) : -1 < \chi < 1, 0 < \xi < 1\} \) which is called the Normal Inverse Gaussian shape triangle. I make the steepness and asymmetry of the distribution conditional on the dataset according to

\[
\begin{align*}
\overline{\gamma}_t &= \exp\left(d + e \varepsilon_{t-1}^2 + f \log(\overline{\gamma}_{t-1})\right), \\
\overline{\rho}_t &= g + h \varepsilon_{t-1},
\end{align*}
\]
with $\tilde{\rho}_t = \log\left(\frac{1 + \rho_t}{1 - \rho_t}\right)$. The exponential form in (6) is used to guarantee that $y_t$ is positive without having to impose any restrictions on the estimated parameters $d$, $e$ and $f$. Similarly since $\rho \in [0,1]$ then $\tilde{\rho}_t = \log\left(\frac{1 + \rho_t}{1 - \rho_t}\right) \in \mathbb{R}$ so also (7) can be estimated without any restrictions on the parameters $g$ and $h$. Equation (6) for the steepness can be seen similar to the EGARCH model Nelson (1991) proposed for the variance but (6) does not allow for different responses in steepness to positive and negative returns innovations of the same magnitude. Also a specification for (7) that allows for more persistence by adding $\tilde{\rho}_{t-1}$ as an explanatory variable was tried but the coefficient in front of $\tilde{\rho}_{t-1}$ turned out to be insignificant.

The full Normal Inverse Gaussian stochastic volatility and autoregressive conditional heteroskedasticity model with time variation (or in short the NIG-S&ARCH-tv model) proposed in this study is given by equations 4-7. The NIG-S&ARCH model of Jensen and Lunde (2001) is given by (4) and (5) meaning that the steepness and asymmetry of the NIG distribution is forced to be constant in their specification.

### B. The GARCH-NIG model

---

3 The model can also be derived from the stochastic volatility literature (see Jensen and Lunde (2001)) which explains why the name NIG-S&ARCH (Stochastic volatility and ARCH) was chosen by Jensen and Lunde.
Forsberg (2002) and Forsberg and Bollerslev (2002) present the GARCH-NIG model by the parameterization $\alpha = \alpha \delta , \sigma^2 = \delta / \alpha$ and by setting $\beta = 0$. This results in the density

$$f_{z;FB}(z;\alpha, \sigma_i^2) = \frac{\sqrt{\alpha}}{\pi \sigma_i} \exp(\alpha) q \left( \frac{z}{\sigma \sqrt{\alpha}} \right)^{-1} K_1 \left( \alpha q \left( \frac{z}{\sigma \sqrt{\alpha}} \right) \right),$$

for the zero mean variable $z$. This parameterization sets the second central moment equal to $\sigma^2$ making it straightforward to incorporate and evaluate temporal dependence.

The model is given by the mean equation

$$r_i = \mu + \sigma \eta_i,$$

and variance equation

$$\sigma_i^2 = c + a \epsilon_{i-1}^2 + b \sigma_{i-1}^2,$$

with $\epsilon_i = \sigma \eta_i$, $c, a, b > 0$ and $\epsilon_i \sim NIG(\alpha, 0, 0, \sigma_i^2)$. The sufficient stationary condition is given by $a+b<1$. The $\sigma_i^2$ parameter, which is the mean of the IG distribution and the conditional variance of the returns, follows the variance equation from Bollerslev (1986).

The GARCH-NIG model was estimated on the ECU/US dollar exchange rate in Forsberg and Bollerslev (2002) and was shown to provide very promising unconditional one day ahead VaR predictions. However, setting $\beta$ to zero can be too restrictive for equity returns considering the often reported skewness, and a more flexible model may be needed. The GARCH-NIG model is nested by the NIG-S&ARCH model, as can be seen by setting $\beta = 0$, $\tau = 0$, $\nu = 2$ and $\delta^2 / \alpha = \sigma^2$. 

13
C. **The APARCH-t model**

To accommodate for the often observed leptokurtic residuals obtained from different GARCH model the Student t distribution has frequently been used, as first proposed by Bollerslev (1987), as the error distribution. To be able to compare the NIG distribution with the Student t distribution I estimate the Asymmetric Power ARCH (APARCH) model of Ding et al. (1993) with the error distribution being Student t. This specification is also used in Giot and Laurent (2003) for VaR modeling. The model is given by the mean equation

\[
(11) \quad r_t = \mu + \sigma_t \eta_t,
\]

and the dynamics of \( \sigma_t \) follow (5). The distribution of \( \varepsilon_t = \sigma_t \eta_t \) is Student t. A zero mean unit variance variable \( z \) follows a Student t distribution if its density function is given by

\[
(12) \quad f_{z, \text{Student} t}(z; df) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\sqrt{\pi(df-2)\Gamma\left(\frac{df}{2}\right)}} \left(1 + \frac{z^2}{df-2}\right)^(-\frac{(df+1)}{2}),
\]

with \( df \) being the degrees of freedom parameter and \( \Gamma(\cdot) \) is the gamma function. The sufficient stationary condition is given by,

\[
(13) \quad a \left\{ (1+\tau)^\nu + (1-\tau)^\nu \right\} \frac{\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{df-\nu}{2}\right)(df-2)^{1+\nu/2}}{2\sqrt{\pi(df-2)\Gamma\left(\frac{df}{2}\right)}} + b < 1,
\]

as shown in Giot and Laurent (2003).

D. **Conditional moments**
The first four moments for the Jensen and Lunde (2001) parameterization of the NIG distribution are given by:

\[ E(x) = \mu + \frac{\rho_i \delta_i}{\sqrt{1 - \rho_i^2}}, \]  
\[ \text{(14)} \]

\[ \text{Var}(x) = \frac{\delta_i^2}{\alpha_i \left(1 - \rho_i^2\right)^{3/2}}, \]
\[ \text{(15)} \]

\[ \text{Skew}(x) = 3 \frac{\rho_i}{\sqrt{1 - \rho_i^2} \sqrt{\alpha_i}}, \]
\[ \text{(16)} \]

\[ \text{Kurt}(x) = 3 \left(1 + \frac{4\rho_i^2 + 1}{\alpha_i \sqrt{1 - \rho_i^2}}\right), \]
\[ \text{(17)} \]

with \( \rho_i = \beta_i / \alpha_i = \bar{\beta}_i / \bar{\alpha}_i \). The moments for the NIG-S&ARCH-tv model are obtained by noticing that, \( \tilde{\rho}_i = \log \left(\frac{1 + \rho_i}{1 - \rho_i}\right) \Leftrightarrow \rho_i = -\frac{\exp(\tilde{\rho}_i) - 1}{\exp(\tilde{\rho}_i) + 1}, \)
\( \bar{\beta}_i = \rho_i \bar{\gamma}_i \left(1 - \rho_i^2\right)^{-1/2} \) and \( \bar{\alpha}_i = \sqrt{\bar{\beta}_i^2 + \bar{\gamma}_i^2} \) with \( \bar{\gamma}_i \) and \( \tilde{\rho}_i \) given in (6) and (7) respectively. For the GARCH-NIG model moments are obtained by noticing that this model has \( \rho_i = 0 \ \forall t, \mu = 0 \) and that

\[ \frac{\delta_i^2}{\alpha_i} = \delta_i = \sigma_i^2. \]  

In particular (17) will reduce to \( \text{Kurt}(x) = 3 + \frac{3}{\alpha_i} \) which clearly shows that any level of excess kurtosis is attainable. The normal distribution is nested by the NIG so the moments for the normal distribution can be obtained by setting

\( \bar{\beta}_i = 0 \ \forall t, \bar{\alpha}_i \rightarrow \infty \ \forall t \) such that

\[ \frac{\delta_i^2}{\alpha_i} = \frac{\delta_i}{\alpha_i} = \sigma_i^2. \]
To illustrate what levels and combinations of skewness and kurtosis are attainable by the NIG-distribution the results in Jondaeau and Rockinger (2003) are used which show that the skewness $\mu_3$ is always bounded for a given level of kurtosis $\mu_4$ by $\mu_3 < \mu_4 - 1$ assuming zero mean and unit variance. The results for the NIG-distribution are compared with the Generalized skewed t distribution of Hansen (1994). For the NIG-distribution the bounds are found by setting $\alpha = 0.9999\beta$ and computing the skewness and kurtosis for a fine grid of values that corresponds to levels of kurtosis ranging from 3.01 to 30. The bounds for the Generalized skewed t distribution are computed as in Jondaeau and Rockinger (2003).

[Insert Figure 1 about here.]

As can be seen from Figure 1 the NIG-distribution is generally more flexible in accommodating varying combinations of skewness and kurtosis than the generalized skewed t distribution, making it a strong candidate distribution for financial modeling.

**IV. Value at risk**

Value at Risk is the maximum loss expected to incur over a certain time period (h) for a given probability $\alpha$. Equivalently, it can be stated that the loss will be less than $\text{VaR}(\alpha, h)$ dollars, $(1-\alpha)*100\%$ of the time. Statistically $\text{VaR}_t(\alpha, h) = F_{t+h}^{-1}(\alpha) | \Omega_t$, where $F_{t+h}^{-1}$ is the h-step conditional forecast of the inverse cumulative distribution
The function (cdf) of the return \( r_t = \log \left( \frac{P_t}{P_{t-1}} \right) \) and \( \Omega_t \) is the information set up to and including time \( t \) information.

With the adoption of Basel II, which allows banks to use internal VaR models for the purpose of regulating capital requirements there has been much academic interest in the measure. For a survey see for example Duffie and Pan (1997) or the textbook treatment in Jorion (2000).

\textbf{A. Backtesting VaR models}

The Basel committee on banking supervision state in their 2004 “International convergence of capital measurement and capital standards” (page 39) that

\textit{“Internal models will only be accepted when a bank can prove the quality of its model to the supervisor through the backtesting of its output using one year of historical data.”}

The exact method for backtesting is not prescribed. However, the number of exceptions, i.e. the number of occasions when the actual loss is larger than predicted by the VaR model, is used to determine a multiplier that directly affects the capital requirements. Following the terminology of Christoffersen (1998), the Basel committee is only concerned with the unconditional coverage of the models. However, a model might have the correct average coverage even though it can be miss-specified at a given point of time. Christoffersen (1998) derives a test for correct conditional coverage that will be presented below.

Define the indicator variable \( I_t \) with \( t \) being a time subscript according to

\[ \]
$I_t = \begin{cases} 
1, & \text{if } r_t > F_{t-1}^{-1}(\alpha) | \Omega_{t-1}, \\
0, & \text{Otherwise} 
\end{cases}$

were $F_{t-1}^{-1}(\alpha) | \Omega_{t-1}$ is the conditional VaR forecast (the inverse of the cumulative distribution function evaluated at $\alpha$) from the particular model being evaluated. This means that a series that constitutes of zeros whenever there is an exception, i.e. when a return is below the value at risk given by the model, and ones otherwise is constructed.

To test if the number of exceptions is correct is called to test for correct unconditional coverage. If we have correct unconditional coverage alpha percent of the returns will be lower than the VaR prediction, $F_{t-1}^{-1}(\alpha) | \Omega_{t-1}$, so under the null we will have

$$E \left[ \frac{1}{T} \sum_{t=1}^{T} I_t \right] = 1 - \alpha.$$

The test for correct conditional coverage can be divided into two separate parts; one part tests for correct unconditional coverage and one part tests for independence in the sequence of exceptions. This is very useful since it can then be investigated if a model rejection is due to unconditional coverage failure, clustering of the exceptions, or both.

The null hypothesis for correct unconditional coverage gives that $I_t \sim Bern(1 - \alpha)$ which can be tested by a likelihood ratio test of the form

$$LR_{UC} = 2 \left( \log \left( \hat{\pi}_1^T \left( 1 - \hat{\pi}_1 \right)^{T-\hat{\pi}_1} \right) - \log \left( (1 - \alpha)^T \alpha^{T-\pi_1} \right) \right).$$
The number of observations is given by $T$, the number of ones is given by $T_1$ and

$$\hat{\pi}_i = T_i / T.$$

To see if the exceptions tend to cluster together over time Christoffersen (1998) suggests testing for independence with first order Markov dependence used as an alternative. The test statistic is given by

$$LR_{\text{IND}} = 2 \left( \log \left( \left( 1 - \hat{\pi}_{00} \right)^{T_{00} - T_{01}} \hat{\pi}_{01}^{T_{01}} \left( 1 - \hat{\pi}_{11} \right)^{T_{11} - T_{10}} \hat{\pi}_{10}^{T_{10}} \right) - \log \left( \hat{\pi}_i \left( 1 - \hat{\pi}_i \right)^{T - T_i} \right) \right).$$

$T_{ij}$ is the number of observations valued $i$ followed by observations valued $j$. The maximum likelihood estimates of $\hat{\pi}_{ij}$ are simply $\hat{\pi}_{01} = T_{01} / T_0$ and $\hat{\pi}_{11} = T_{11} / T_1$.

The joint test of correct conditional coverage means that $I_i \sim iid Bern(1-\alpha) \forall t$. The test statistic is simply given as the sum of the two individual tests in equations 20 and 21.

$$LR_{\text{CC}} = LR_{\text{UC}} + LR_{\text{IND}}$$

Christoffersen and Pelletier (2004) suggest that the $LR_{\text{IND}}$ test has poor power against general forms of dependence, and as an alternative, propose a duration based test for independence. The duration $D_i = t_i - t_{i-1}$, with $t_i$ being the day of exception number $i$, is defined as the time (in days) between two exceptions. The expected duration should
always be \(1/\alpha\) for a correctly specified model and the most promising implementation of the test\(^4\) is based on the Weibull distribution

\[
f_D(D; a, b) = a^b b D^{b-1} \exp(-(aD))^b.
\]

A model with independent exceptions will have a flat hazard function. For the Weibull distribution a \(b\) parameter equal to 1 gives a flat hazard so testing if \(b=1\) can be seen as a test for independence. The test statistic is given as a log likelihood test by

\[
LR_{\text{weibull}} = 2\left(\text{Logl}_{\text{max},a,b} f_D(D; a, b) - \text{Logl}_{\text{max},a} f_D(D; a, 1)\right),
\]

which is the difference between the unrestricted log likelihood and the log likelihood when \(b\) is restricted to the value of 1.

Christoffersen (1998) uses the asymptotic distribution results for the test statistics in equation 21-23. However, I will follow the recommendation in Christoffersen and Pelletier (2004) and simulate the distribution of the test statistics since the effective sample sizes are rather small in typical VaR settings.

**B. VaR using hyperbolic distributions**

The generalized hyperbolic distribution, of which the NIG is a special case, was introduced to the field of finance by Eberlein and Keller (1995) and Barndorff-Nielsen (1995). It has shown promising results for computing VaR in Eberlein et al. (1998),

\(^4\)Most promising in the sense that the Weibull implementation of the test had the highest power for all sample sizes over 750 in the simulation study in Christoffersen and Pelletier (2004). However, they only consider one DGP which is a GARCH model with t-distributed errors.
Bauer (2000), Forsberg and Bollerslev (2002) as well as Venter and de Jongh (2002). However the previously used models lack the flexible variance dynamics and inclusion of non-zero skewness provided by the NIG-S&ARCH model of Jensen and Lunde (2001) and they all assume constant conditional skewness and kurtosis. Furthermore no rigorous backtesting of the models has been conducted in the above studies.

V. Data and estimation

In this section we take a look at the data and describe the estimation and forecasting procedures. Estimation results are presented and discussed, this section ends with model diagnostics and density forecasting results.

A. Description of data

The descriptive statistics in Table 1 show, as is usual for daily stock return data, that normality is overwhelmingly rejected with p-values of the Jarque and Bera (1987) statistic less than 0.001. The forecasting part of the sample is slightly more volatile with a yearly average standard deviation equal to 15.63% compared to the 11.37% in the initial estimation sample and with an excess kurtosis of 4.96 compared to the 3.37 in the estimation sample. The sample skewness is slightly positive in the estimation sample and slightly negative in the forecast sample. However the quartile based skewness measure,
\[
\frac{q_3 + q_{1-2}q_2}{q_3 - q_1} \text{ with } q_i \text{ being the } i\text{th quartile of the empirical distribution, of Kim and White (2004) which is more robust to outliers shows negative skewness in both periods.}
\]

[Insert Table 1 about here.]

**B. Estimation**

The empirical investigation of the models is conducted on the Standard and Poor’s 500 index from July 3, 1962 (t=0) to September 20, 2005 (t=11,877). Logarithmic dividend adjusted returns have been used and the return on October 19, 1987 has been excluded. The reason for the exclusion is that the stock market crash cannot be properly captured by any of the models used; a similar conclusion was drawn in Venter and de Jongh (2004). This is clearly a failure of all the considered models. It is the author’s belief that crashes of this magnitude need to be modeled separately and while an interesting topic, it is outside the scope of this paper. The initial model estimation uses the 3,000 first observations which will be called the estimation sample. The rest of the data are saved for forecasting purposes. The models are estimated using a rolling window so that the first parameters are estimated on observations t=0 to t=2,999 and the complete density is forecasted for t=3,000 for the four models. Then the estimation sample is rolled forward one day so that new parameters are estimated on observations t=1 to t=3,000 and new forecasts are calculated. This procedure is repeated for the whole sample resulting in 7,878 density forecasts for each model. The models are estimated by numerical maximum
likelihood using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. The Bessel function is numerically calculated using code from numerical recipes in C that uses the polynomial approximation formula in Olver (1964) p. 378. Since the model parameters are re-estimated each day the computing time is considerable. To speed up convergence of the BFGS algorithm the last period’s estimates are used as starting values for the new estimation resulting in more rapid convergence.

In case of parameter stability (i.e. constant unconditional variance) the rolling scheme is of course less efficient than extending the estimation sample to include more and more observations. However it is not feasible to assume away shifts in the unconditional variance during the sample period in this study that covers more than 40 years in length. Shifts in the unconditional variance will lead to spurious long memory in the GARCH variance equation, pushing the persistence towards one (see e.g., Mikosch and Starica, (2004)).

C. Estimation results

The NIG-S&ARCH-tv model is allowed to have different asymmetry and steepness parameters for each observation. This is illustrated by plotting the shape of the distribution for each day in the Normal Inverse Gaussian shape triangle.

[Insert Figure 2 about here.]

We can see from Figure 2 that the distribution most of the days is close to being symmetric although there is a slight negative skewness, both the mean and the median of
the asymmetry parameter \( \chi \) is -0.007. The steepness is rather clustered around 0.25 but a few observations show considerably higher steepness. The daily shapes of the distribution can be compared with the normal distribution that corresponds to the region near \( \chi = 0, \xi = 0 \) and to the Cauchy distribution which is the limiting case near \( \chi = 0, \xi = 1 \).

The parameter estimation results in Table 2 show that the shape parameter alpha in both the GARCH-NIG and NIG-S&ARCH displays thicker tails than the normal distribution with parameter estimates of 4.08 and 6.36 implying a conditional kurtosis of 3.74 and 3.48 respectively. All the models are stationary although only barley so for the GARCH-NIG model with the sum of the a and b parameters being equal to 0.9967. The unconditional kurtosis of the GARCH-NIG model does not exist (is not finite) for the values of the parameters estimated here\(^5\). For the NIG-S&ARCH and NIG-S&ARCH–tv they are 7.71 and 9.53 respectively compared to the sample kurtosis of 6.37. For the APARCH-t model the unconditional kurtosis is 21.16. The unconditional kurtosis for the GARCH-NIG model is calculated from the moment results in Forsberg (2002, pp.154-155).\(^6\) For the NIG-S&ARCH-tv, NIG-S&ARCH and APARCH-t models, simulations have been used. The simulated numbers must of course be treated with caution as there is no guarantee of finite kurtosis since the conditions for moment existence for these models have not been derived.

[Insert Table 2 about here.]

---

\(^5\) The unconditional kurtosis for the GARCH-NIG model exists only if \( b^2 + 2ab + a^2 (3 + 3/\alpha) < 1 \) with a and b given in (10) and \( \alpha \) in (8).

\(^6\) Notice the misprint, \( \sqrt{a} \) should be \( \sqrt{\alpha} \) in A.2 Forsberg (2002, p.154).
The significant \( \hat{\beta} \) estimate of -0.387 for the NIG-S&ARCH indicates that setting beta to zero is too restrictive. The \( e \) parameter of the NIG-S&ARCH-tv model shows that \( \bar{\gamma} \), which is a one to one transform of the steepness \( \xi = (1 + \bar{\gamma})^{-1/2} \) of the distribution, is significantly negatively related to previous periods squared return innovations and further the \( f \) parameter of 0.7085 shows that there is much persistence in the steepness. This means that a large return innovation will give not only higher future variance but also higher future kurtosis. This is an additional risk factor that can not be captured by models that assume constant higher moments. The positive and significant \( h \) parameter means that a negative return today will give lower (more negative) skewness tomorrow. The conditional kurtosis and skewness produced by the NIG-S&ARCH-tv model are shown in Figures 3 and 4. As can be seen, the skewness oscillates around zero with some periods of more pronounced skewness especially in the later part of the sample. The conditional kurtosis shows less day-to-day fluctuations except for a few sharp peaks in the kurtosis. The occasionally very high kurtosis speaks in favor of a time varying specification such as in the NIG-S&ARCH-tv model since a constant specification would overestimate the kurtosis for most days and underestimate during the periods of high kurtosis. In fact, 90% of the days have a kurtosis that is below the average conditional kurtosis of 3.30. The improved fit of the NIG-S&ARCH-tv model is also evident in the much improved log likelihood value of -2,541.45 compared to -2,621.94 when the higher moments are restricted to be constant.

[Insert Figures 3 and 4 about here.]
D. Model diagnostics

As a diagnostic measure for the models, the Q-test of Hong and Li (2005) is used. One of the advantages of this test is that both the in-sample fit as well as the density forecasts produced by the models can be evaluated with the same test. It can also be seen as a test for VaR performance. The tests given in equations 21-23 all test the predictions for a specific VaR percentage ($\alpha$) though a correctly specified model should have the right number of exceptions for all $\alpha$. To jointly test for all $\alpha$ is to test the density forecast of the model.

It is of interest to assess the out-of-sample performance of the models, especially for the NIG-S&ARCH-tv model that is similar in number of parameters to Hansen’s (1994) conditional autoregressive density model. For example Guermat and Harris (2002, p. 410) express concern that “…in Hansen’s interest rate model, there are 15 estimated parameters and hence its effectiveness for out-of-sample forecasting is likely to be limited”. Therefore the estimated as well as the predicted density for all models will be evaluated.

The Hong and Li (2005) Q-test is based on the result that the probability integral transform (PIT) of the correct density will be iid uniform on the unit interval. That is,

$$\hat{z}_i = \int_{-\infty}^{\hat{y}_i} \hat{p}_t(u) \, du \sim iid U(0,1)$$

if the probability density function (pdf) of the model ($\hat{p}_t$) is equal to the true pdf of the returns. Using the PIT to evaluate density forecasts has previously been proposed by Diebold et al. (1998) as well as by Kim et al. (1998). Hong
and Li (2005) build on these ideas and propose an omnibus test to jointly test for both the uniformity and the independence of \( \{\tilde{z}_t\}_{t=1}^T \) by comparing a kernel estimate of the joint distribution of \( \{\tilde{z}_t, \tilde{z}_{t-j}\}_{t=j+1}^T \) with a bivariate uniform distribution. The test is given by

\[
\hat{Q}(j) = \left[ (T-j)h\hat{M}(j) - A_{h,j}^0 \right] / V_{0}^{1/2},
\]

which is asymptotically normally distributed \((0,1)\) under the null hypothesis of a correctly specified density. The kernel estimate of the joint distribution of

\[
\hat{g}_j(z_1,z_2) = (T-j)^{-1} \sum_{t=j+1}^{T} K_h(z_1,\tilde{z}_t)K_h(z_2,\tilde{z}_{t-j}),
\]

with \( \tilde{z}_t \) and \( \tilde{z}_{t-j} \) being sample values of the PIT-series and \( z_1 \) and \( z_2 \) being evaluation points is given by

\[
K_h(x,y) = \begin{cases} 
    h^{-1}k\left(\frac{x-y}{h}\right) / \int_{-(x/h)}^{1} k(u) du & \text{if } x \in [0,h) \\
    h^{-1}k\left(\frac{x-y}{h}\right) & \text{if } x \in [h,1-h] \\
    h^{-1}k\left(\frac{x-y}{h}\right) / \int_{-1}^{(1-x)/h} k(u) du & \text{if } x \in (1-h,1],
\end{cases}
\]

for \( x \in [0,1] \). The bandwidth parameter, \( h \), is chosen to equal the sample standard deviation of the PIT-series times the number of observations raised to the power of \(-1/6\) in accordance with Hong and Li (2005). Here

\[
k(u) = \frac{15}{16} (1-u^2)^2 I(|u| \leq 1),
\]

and \( I \) is an indicator function taking the value 1 when \(|u| \leq 1\) and zero otherwise.

The distance between the estimated distribution and the product of two \( U(0,1) \) densities is given by
Subtraction of

\[ A_h^0 = \left[ (h^{-1} - 2) \int_{-1}^{1} k^2(u)du + 2 \int_{0}^{h} k_h(u)du \right]^2 - 1, \]

with \( k_h(u) = k(u)/\int_{-1}^{b} k(v)dv \) centers the test statistic around zero and division by the square root of

\[ V_0 = 2 \left[ \int_{-1}^{1} \int_{-1}^{1} k(u + v)k(v)dv \right]^2 du \]

gives unit variance. Since negative values of \( \hat{Q}(j) \) can only occur under the null for large samples, one-sided critical values should be used.\(^7\)

[Insert Figure 5 about here.]

Figure 5 shows the Q-statistics for the estimated densities produced by the models for values of \( j \) ranging from 1 to 20. The GARCH-n as well as the GARCH-NIG, APARCH-t and the NIG-S&ARCH models can clearly be rejected whereas the NIG-S&ARCH-tv model has Q-statistics very close to the 5% critical value of 1.64. For the GARCH-n,

\(^7\) The test is implemented by numerical integration in Gauss using computer code generously contributed by Professor Yongmiao Hong at Cornell University.
APARCH-t and GARCH-NIG models, the test rejects also for \( j > 2 \). This is indicative of poor modeling of both dynamic and static properties. The NIG-S&ARCH performs much better for \( j > 1 \) indicating that only (unaccounted) one day dependence is the cause of rejection. The reason for this interpretation is the joint testing for uniformity and dependence. If the test statistic drops below the rejection level after a certain value for \( j \) then the static properties (the marginal distribution) of the returns must be well modeled and the high Q-statistic for lower values of \( j \) must be due to misspecification of the dynamics, since poor modeling of the static properties will show for all values of \( j \).

[Insert Figure 6 about here.]

The out-of-sample density forecast ability of the models shown in figure 4 is rather similar with the exception of the GARCH-n model that clearly performs the worst with a Q(1) statistic of 22.61. The NIG-S&ARCH–tv is still the best model but it can be rejected with a Q(1) statistic of 7.29. Actually the NIG-S&ARCH–tv model is the only model that fares worse out-of-sample than in-sample which hints at possible over fitting and/or difficulty in estimating the dynamics of the higher moments.

**VI. Value at risk results**

Figure 7 shows how the 1% VaR varies over the forecasting sample for the NIG-S&ARCH–tv model. The circles indicate when a loss larger than predicted by the model occurs. Strong clustering effects of the VaR are evident and the period from 1997-2002
shows considerably higher risk than the rest of the period with the exception of the sharp spike during the October-87 crash. The effects of the -87 crash are, however, remarkably short-lived and the period following the crash is rather tranquil. However, including the return of October 19, 1987 could alter this conclusion.

[Insert Figure 7 about here.]

All models with the exception of the GARCH-n and APARCH-t show exceptionally good results for conditional coverage at the 1% VaR. The p-values reported in Tables 3 and 4 are calculated by simulating the distribution of the test statistic under the null as outlined by Christoffersen and Pelletier (2004). A sample size equal to the empirical sample size (7,878 observations) with 100,000 replications has been used. In addition to the simulated p-values, asymptotic p-values are reported in brackets to facilitate comparison with previous studies that have relied on these.

The empirical percentage of rejections is 0.9520% for the NIG-S&ARCH–tv model and 0.9647% for the NIG-S&ARCH model. These numbers can clearly not be rejected as being different from 1% with p-values from the $LR_{UNC}$ test of 0.65 and 0.73 respectively. The unconditional coverage of the GARCH-NIG model at the 1% level is actually perfect with 79 exceptions compared to the expected 78.78. The APARCH-t model can not be rejected as having correct unconditional coverage with a $LR_{UNC}$ p-value of 0.19 whereas the GARCH-n clearly underestimates the VaR with the empirical percentage of

---

8 In Table 3 the empirical size is given as 1.0028%, this is because 1% is impossible to obtain for the selected sample size since the expected unconditional coverage of 0.01*7,878 is not an integer number.
exceptions at 1.5486% which can be rejected as providing correct unconditional coverage
with a p-value of less than 0.01.

[Insert Tables 3 and 4 about here.]

with the internal models approach to market risk capital requirements” is only concerned
with unconditional coverage and divides models into three groups; green, yellow and red
depending on the number of violations. A model is in the green zone if a 95% confidence
interval around the correct number of exceptions covers the realized number of
exceptions, in the yellow zone if a 99.99% confidence interval covers and otherwise in
the red zone. The green zone requires no additional capital requirements, the yellow zone
can lead to additional capital requirements according to the judgment of the supervisor,
and the red zone, in additional to higher capital requirements, also requires that measures
to improve the model should be taken immediately. For the number of observations in
this study, models that have exceptions below 1.18% will be in the green zone, 1.18-
1.42% in the yellow zone and exceptions higher than 1.42% will be in the red zone. The
three models with NIG errors and the APARCH-t are clearly in the green zone but the
GARCH-n model is in the red zone and hence would not be accepted by the supervision
agency. For VaR at the 5% level, the unconditional coverage of all the models except the
APARCH-t is satisfactory, showing that the failure of the GARCH-n model is only

---

9 The limits for the zones are given by \(0.01 + 1.6449\sqrt{(0.01 \cdot 0.99 / 7878)} = 1.18\%\) and
\(0.01 + 3.7190\sqrt{(0.01 \cdot 0.99 / 7878)} = 1.42\%\) respectively.
apparent in the utmost parts of the tail. Furthermore it is seen that the APARCH-t model with empirical exceptions at 4.3031% overestimates the risk for the 5% VaR. The three models based on the NIG distribution in this study compare favorably, in terms of unconditional coverage, to mixture GARCH models with a stable Paretian error distribution that was proposed and evaluated by Haas et al. (2005)\textsuperscript{10} on a German Dax-30 sample of comparable size.

In terms of independence, which ignores the unconditional coverage and only test if the exceptions tend to cluster together, the GARCH-n clearly performs worse than the other models for 1% VaR. Independence can be rejected at the 5% confidence level for the GARCH-n model with a p-value less than 0.01. However, when looking at more general forms of clustering, the $LR_{\text{Weibull}}$ test cannot reject any of the models. This is somewhat surprising considering that the power of the $LR_{\text{Weibull}}$ test is found superior to the $LR_{\text{IND}}$ test in the simulation study of Christoffersen and Pelletier (2004). It appears the power of the $LR_{\text{Weibull}}$ test is lower when comparing different types of GARCH models then when comparing GARCH type models to historical VaR as done in the simulation study by Christoffersen and Pelletier (2004).

For VaR at 5% the temporal dependencies seem more severe and all the models except the NIG-S&ARCH–tv can be rejected as providing independent VaR exceptions by the $LR_{\text{IND}}$ test on the 1% level whereas the NIG-S&ARCH–tv model can clearly not be rejected with a p-value of 0.80. However, also here the $LR_{\text{Weibull}}$ test fails to reject the null

\textsuperscript{10} Conditional coverage is not investigated in Haas et al. (2005) limiting the possible comparison to unconditional coverage.
of independent exceptions for all four models. The large discrepancy between the simulated and asymptotic p-values for the Weibull test is due to a discreteness bias that will lead to overestimation of the $b$ parameter in the $LR_{weibull}$ test. The bias is increasing when the durations under the null are shorter and is hence more severe for VaR at the 5% than the 1% level. However, the simulated p-values are unaffected by this bias. The NIG-S&ARCH–tv model is the only model that is not rejected by the joint test, $LR_{cc}$ for both correct conditional and unconditional coverage. For the 1% VaR only the GARCH-n model is rejected in the joint test.

These results can be compared with the recent study by Kuester et al. (2006) who evaluate the performance of 18 different conditional and 5 unconditional VaR models. Their study includes GARCH models with different error distributions and the CAViaR model of Engle and Manganelli (2004). Models are evaluated on 6681 one-step-ahead VaR forecasts for the NASDAQ index. When comparing with the backtesting results in Kuester et al. (2006) the NIG-S&ARCH–tv, NIG-S&ARCH and GARCH-NIG models in this study beat all the 23 models considered in Kuester et al. (2006) when comparing the $L_{cc}$ statistic for 1% VaR. Furthermore the NIG-S&ARCH-tv model proposed in this study also beats all 23 models in Kuester et al. (2006) regarding conditional coverage for 5% VaR and it is the only model that can not be rejected as providing the correct number of independent VaR exceptions for both 1% and 5% VaR.

**VII. Conclusions**
This paper proposes a new model based on the NIG distribution for the modeling of conditional variance, skewness and kurtosis. The NIG distribution is very flexible in that it can fit the skewness and excess kurtosis of the data. Furthermore it has attractive analytical properties, such as being closed under convolution and having a closed form density. In addition, the distribution naturally arises from Clark’s (1973) mixture of distribution theory when the mixing variable is assumed to follow and inverse Gaussian distribution.

The new NIG-S&ARCH-tv model as well as the NIG-S&ARCH model of Jensen and Lunde (2001) and the GARCH-NIG model of Forsberg and Bollerslev (2002) are applied to VaR forecasting on the Standard and Poor’s 500 from July 12, 1974 to September 20, 2005. The models perform very well compared to extant models evaluated in for example Haas et al. (2005) and Kuester et al. (2006) on the German DAX and NASDAQ indices respectively. The NIG-S&ARCH-tv model proposed in this paper is the only model that can not be rejected as providing a correct number of independent VaR exceptions both for 1% and 5% VaR.

Interestingly enough it seems more difficult to find a suitable model for the 5% VaR than for the 1 % VaR. This is due to the unmodeled dependence of the exceptions for the 5% VaR. This finding is corroborated in Kuester et al. (2006) in which only two of the 18 conditional VaR models have better test statistics for the joint test of Christoffersen (1998) \( LR_{cc} \) on 5% than 1% VaR. One reason for this, however, can be the weaker inference possible for the 1% VaR, since the expected number of rejections is 80% lower than for the 5% VaR which reduces the effective sample size.
One somewhat unexpected finding is that the unconditional coverage for the GARCH-NIG model is even better than on the foreign exchange sample used in Forsberg and Bollerslev (2002). This is surprising since the model does not allow for the asymmetric responses in volatility to negative and positive returns that are well documented in the extant literature. Yet several previous studies have shown (see Poon and Granger (2003) and references therein) that while allowing for leverage improves the in-sample fit, the out-of-sample forecast performance is often not increased.

The GARCH-n model both significantly underestimates the 1% VaR and also fails to produce correct conditional coverage whereas the APARCH-t model also fails to produce independent exceptions (at 5% VaR) it overestimates the risk. As judged by the rules set out in Basel (1996) the three NIG based models and the APARCH-t are in the green zone, requiring no additional capital requirements, whereas the GARCH-n model is in the red zone, meaning that measures to improve the model must be taken.

The NIG based models show very promising results for use in risk management and this study proposes one possible way to generalize those models by incorporating time variation in skewness and kurtosis. The NIG-S&ARCH-tv model proposed in this study provides improvements in VaR and density forecasting applications and is the only model that shows satisfying results for both 1% and 5% VaR.
References


Hong, Y., and H. Li. “Nonparametric specification testing for continuous-time models with applications to term structure of interest rates.” *The review of Financial Studies*, 18 (2005), 37-84.


## TABLE 1

Descriptive statistics of the Standard and Poor’s 500 Index

<table>
<thead>
<tr>
<th>Measure</th>
<th>Initial estimation sample</th>
<th>Backtesting sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>3,000</td>
<td>7,878</td>
</tr>
<tr>
<td>Daily Mean</td>
<td>0.034%</td>
<td>0.037%</td>
</tr>
<tr>
<td>Yearly Standard deviation</td>
<td>11.37%</td>
<td>15.63%</td>
</tr>
<tr>
<td>Maximum</td>
<td>4.90%</td>
<td>8.71%</td>
</tr>
<tr>
<td>Minimum</td>
<td>-3.12%</td>
<td>-8.64%</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.104</td>
<td>-.100</td>
</tr>
<tr>
<td>Skewness robust</td>
<td>-0.020</td>
<td>-.0100</td>
</tr>
<tr>
<td>Excess Kurtosis</td>
<td>3.366</td>
<td>4.961</td>
</tr>
<tr>
<td>JB</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 1 shows the descriptive statistics for the daily Standard and Poor’s 500 percentage returns for the initial estimation sample July 3, 1962 to July 11, 1974 as well as for the out-of-sample period July 12, 1974 to September 20, 2005. Skewness robust is the skewness measure \( \frac{q_3 + q_1 - 2q_2}{q_3 - q_1} \), suggested in Kim and White (2004) due to its robustness against outliers. JB is the p-value from the Jarque and Bera (1987) test with the null hypothesis of normally distributed returns. The return of October 19, 1987 is excluded.
**Table 2**

Estimation results for daily percentage Standard and Poor’s 500 returns

<table>
<thead>
<tr>
<th>Parameter</th>
<th>NIG-S&amp;ARCH-tv</th>
<th>NIG-S&amp;ARCH</th>
<th>GARCH-NIG</th>
<th>APARCH-t</th>
<th>GARCH-n</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0842*</td>
<td>0.1033**</td>
<td>0.0529**</td>
<td>0.01254**</td>
<td>0.0510**</td>
</tr>
<tr>
<td></td>
<td>(0.0366)</td>
<td>(0.0019)</td>
<td>(0.0097)</td>
<td>(&lt;0.001)</td>
<td>(0.0103)</td>
</tr>
<tr>
<td>$a$</td>
<td>0.0707**</td>
<td>0.0705**</td>
<td>0.1296**</td>
<td>0.0760**</td>
<td>0.1294**</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td>(0.001)</td>
<td>(0.0173)</td>
<td>(0.0111)</td>
<td>(0.0188)</td>
</tr>
<tr>
<td>$b$</td>
<td>0.9303**</td>
<td>0.9353**</td>
<td>0.8671**</td>
<td>0.9416**</td>
<td>0.8597**</td>
</tr>
<tr>
<td></td>
<td>(0.0061)</td>
<td>(0.0112)</td>
<td>(0.0165)</td>
<td>(0.010)</td>
<td>(0.0192)</td>
</tr>
<tr>
<td>$c$</td>
<td>0.0093**</td>
<td>0.0085**</td>
<td>0.0069**</td>
<td>0.001***</td>
<td>0.0075**</td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td>(0.0024)</td>
<td>(0.0017)</td>
<td>(n/a)</td>
<td>(0.0023)</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.8965**</td>
<td>0.7975**</td>
<td></td>
<td>0.9783**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0796)</td>
<td>(0.0868)</td>
<td></td>
<td>(0.0131)</td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>1.0068**</td>
<td>0.8615**</td>
<td></td>
<td>0.3949</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0191)</td>
<td>(0.1328)</td>
<td></td>
<td>(0.0747)</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>6.3557**</td>
<td></td>
<td>4.0789**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.4876)</td>
<td>(0.4946)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.3877**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1052)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows the Maximum Likelihood parameter estimates with robust standard errors in parenthesis from the initial estimation of the models on Standard and Poor’s 500 index returns from July 3, 1962 to July 11, 1974 (3,000 observations). The NIG-S&ARCH-tv model is given by (4)-(7), the NIG-S&ARCH by (4) and (5), the GARCH-NIG by (9) and (10) with the error distribution given by (3). The APARCH-t is given by (10) and (11) and the error distribution by (12). The GARCH-n is given by (9) and (10) and the error distribution is Gaussian. $a+b$ is the stationary condition, $aE \left[ \left( \eta_t \right)^\nu \right] + b < 1$, calculated analytically for the GARCH-NIG, APARCH-t and GARCH-n models and by simulation for the NIG-S&ARCH-tv and NIG-S&ARCH models. LL gives the log likelihood values (constant terms included) of the models.

* Indicates significance at the 5% level.
** Indicates significance at the 1% level.
*** By restriction.
### TABLE 3

1 day-ahead Value at Risk predictions at the 1% level

<table>
<thead>
<tr>
<th>Measure</th>
<th>NIG-S&amp;ARCH-tv</th>
<th>NIG-S&amp;ARCH</th>
<th>GARCH-NIG</th>
<th>APARCH-t</th>
<th>GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>% below VaR</td>
<td>0.9520</td>
<td>0.9647</td>
<td>1.0028</td>
<td>0.8632</td>
<td>1.5486</td>
</tr>
<tr>
<td>UC</td>
<td>0.65 (0.67)</td>
<td>0.73 (0.75)</td>
<td>1.00 (0.98)</td>
<td>0.19 (0.21)</td>
<td>&lt;0.01 (0.01)</td>
</tr>
<tr>
<td>Ind</td>
<td>0.75 (0.74)</td>
<td>0.78 (0.77)</td>
<td>0.84 (0.82)</td>
<td>0.50 (0.28)</td>
<td>&lt;0.01 (0.01)</td>
</tr>
<tr>
<td>CC</td>
<td>0.86 (0.86)</td>
<td>0.89 (0.91)</td>
<td>1.00 (0.97)</td>
<td>0.32 (0.25)</td>
<td>&lt;0.01 (0.01)</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.76 (0.86)</td>
<td>0.39 (0.40)</td>
<td>0.72 (0.80)</td>
<td>0.13 (0.12)</td>
<td>0.73 (0.81)</td>
</tr>
</tbody>
</table>

Table 3 reports the percentage of exceptions from the 7,878 daily VaR forecasts calculated from July 12, 1974 to September 20, 2005. Also reported are the simulated p-values (asymptotic p-values in parentheses) for the unconditional (UC), independence (IND) and joint test for the null of correct conditional coverage (CC) introduced by Christoffersen (1998) as well as the duration based independence test (Weibull) of Christoffersen and Pelletier (2004). Also reported is the model adequacy as determined by the three zones in the Basel (1996) regulations (see main text for details).
1 day-ahead Value at Risk predictions at the 5% level

<table>
<thead>
<tr>
<th>Measure</th>
<th>NIG-S&amp;ARCH-tv</th>
<th>NIG-S&amp;ARCH</th>
<th>GARCH-NIG</th>
<th>APARCH-t</th>
<th>GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>% below VaR</td>
<td>4.8998</td>
<td>4.7728</td>
<td>5.0140</td>
<td>4.3031</td>
<td>4.7855</td>
</tr>
<tr>
<td>UC</td>
<td>0.65</td>
<td>0.35</td>
<td>0.96</td>
<td>&lt;0.01</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>(0.67)</td>
<td>(0.35)</td>
<td>(0.95)</td>
<td>(&lt;0.01)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Ind</td>
<td>0.80</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>(0.79)</td>
<td>(&lt;0.01)</td>
<td>(&lt;0.01)</td>
<td>(&lt;0.01)</td>
<td>(&lt;0.01)</td>
</tr>
<tr>
<td>CC</td>
<td>0.90</td>
<td>0.02</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.01)</td>
<td>(&lt;0.01)</td>
<td>(&lt;0.01)</td>
<td>(&lt;0.01)</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.83</td>
<td>0.99</td>
<td>0.59</td>
<td>0.60</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.96)</td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.07)</td>
</tr>
</tbody>
</table>

Table 4 reports the percentage of exceptions from the 7,878 daily VaR forecasts calculated from July 12, 1974 to September 20, 2005. Also reported are the simulated p-values (asymptotic p-values in parentheses) for the unconditional (UC), independence (IND) and joint test for the null of correct conditional coverage (CC) introduced by Christoffersen (1998) as well as the duration based independence test (Weibull) of Christoffersen and Pelletier (2004). No model adequacy as determined in Basel (1996) is reported since the regulations only apply to VaR at the 1% level.
Figures

Figure 1
Skewness-Kurtosis bounds

This Figure compares the skewness-kurtosis bounds of the NIG-distribution and the Generalized t distribution (Hansen 1994) to the maximum attainable bounds.
This Figure shows the asymmetry $\chi_t = \rho_t \xi_t$ and steepness $\xi_t = \left(1 + \gamma_t^2\right)^{-1/2}$ for the NIG-S%ARCH-tv model obtained from the initial estimation sample from July 3, 1962 to July 11, 1974 (3,000 observations). $\gamma_t$ is given by (6) and $\rho_t$ from (7) by using $\rho_t = -(\exp(\hat{\rho}) - 1)(-\exp(\hat{\rho}) - 1)^{-1}$. Each pair of asymmetry and steepness is plotted in the Normal inverse Gaussian shape triangle given by the region \[ \{(\chi, \xi) : -1 < \chi < 1, 0 < \xi < 1\} \].
This Figure shows the conditional kurtosis given by the NIG-S&ARCH-tv model from July 3, 1962 to July 11, 1974 (3,000 observations). The kurtosis is given by $3\left(1+(4\rho+1)/\left(\sigma\sqrt{1-\rho^2}\right)\right)$ with
\[\hat{\rho}_t = g + h\varepsilon_{t-1}, \quad \rho = - \frac{\exp(\hat{\rho}) - 1}{-\exp(\hat{\rho}) - 1}, \quad \bar{\gamma}_t = \exp\left(d + e\varepsilon_{t-1}^2 + f\log(\bar{\gamma}_{t-1})\right), \quad \tilde{\beta} = \frac{\rho\bar{\gamma}_t}{\sqrt{1-\rho^2}} \quad \text{and} \quad \sigma_t = \sqrt{\bar{\gamma}_t^2 + \rho_t^2}.\]
The estimates of the parameters $d, e, f, g$ and $h$ are given in Table 2.
This Figure shows the conditional skewness given by the NIG-S&ARCH-tv model from July 3, 1962 to July 11, 1974 (3,000 observations). The skewness is given by \( 3 \rho \left( \frac{\sqrt{1 - \rho^2}}{\sqrt{\alpha}} \right)^{-1} \) with
\[
\tilde{\rho}_t = g + h \varepsilon_{t-1}, \quad \rho_t = -\exp(\tilde{\rho})^{-1}, \quad \overline{\gamma}_t = \exp\left(d + ee_{t-1}^2 + f \log(\overline{\gamma}_{t-1})\right), \quad \overline{\beta}_t = \frac{\rho_t \overline{\gamma}_t}{\sqrt{1 - \rho_t^2}} \quad \text{and} \quad \overline{\alpha}_t = \sqrt{\overline{\beta}_t^2 + \overline{\gamma}_t^2}.
\] The estimates of the parameters d, e, f, g, and h are given in Table 2.
This Figure shows the Q-test from Hong and Li (2005) given by equation 24 with a lag length from 1 to 20. The solid line is the 5% one-sided critical value (1.64) for rejecting the null of the estimated density being the true density. The densities are estimated on daily Standard and Poor’s 500 index returns from July 3, 1962 to July 11, 1974 (3,000 observations).
This Figure shows the Q-test from Hong and Li (2005) given by equation 24 with a lag length from 1 to 20. The solid line is the 5% one-sided critical value (1.64) for rejecting the null of the predicted density being the true density. The predictions are made by using the first 3000 observations (t=0, t=2999) to estimate the models and then make a one-step ahead density forecasts for t=3000. Then I roll the estimation window forward one period and re-estimate on t=1 to t=3000 data and make density forecasts for t=3001. This is repeated for the full sample giving 7,878 density forecasts for each model from July 12, 1974 to September 20, 2005.
This Figure shows the 1% 1 day ahead predicted VaR on a 100$ investment given by the NIG-S&ARCH–tv model, from July 12, 1974 to September 20, 2005 (7,878 predictions). Losses that are larger than predicted by the model are indicated by a circle.