**Mannheimer Manuskripte zu Risikotheorie,
Portfolio Management und Versicherungswirtschaft**

Nr. 182

A g-and-h Copula Approach to Risk Measurement in Multivariate Financial Models

von
Markus Huggenberger, Timo Klett

Version vom
15. Dezember 2010

A g-and-h Copula Approach to Risk Measurement in Multivariate Financial Models

Markus Huggenberger*, Timo Klett

University of Mannheim, Germany

This version: December 15, 2010

Abstract: We propose and backtest a multivariate Value-at-Risk model for financial returns based on Tukey's g-and-h distribution. This distributional assumption is especially useful if (conditional) asymmetries as well as heavy tails have to be considered and fast random sampling is of importance. To illustrate our methodology, we fit copula GARCH models with g-and-h distributed residuals to three European stock indices and provide results of out-of-sample Value-at-Risk backtests. We find that our g-and-h model outperforms models with less flexible residual distributions and attains similar results as a benchmark model based on Hansen's skewed-t distribution.

Keywords: g-and-h distribution, copula, GARCH, Value-at-Risk, stock indices, skewed-t distribution

JEL classifications: C16, C32, C46, C51, G10

1. Introduction

Accurate Value-at-Risk (VaR) estimates for portfolios are essential for internal risk management and regulatory frameworks. In the past years, traditional multivariate VaR models like the variance-covariance method, multivariate GARCH or historical simulation have been supplemented by more flexible models based on copula functions. Whereas earlier copula applications focused on the unconditional distribution of risk factors, more recently a series of papers[1] have analyzed the potential of combining time series techniques with the copula approach to obtain highly flexible time varying multivariate models. The foundation of these models is Sklar's theorem for conditional distributions (Patton, 2004, 2006). Recent applications of this approach to portfolio risk

* Corresponding author. Tel. +49 621 181 16 79. E-mail: huggenberger@bwl.uni-mannheim.de.

[1] This approach is e.g. used in Jondeau/Rockinger (2006), Rodriguez (2007), Bartram et al. (2007), Sun et al. (2009) and Fischer et al. (2009).

measurement combined GARCH-filtered margins with various copula functions. Huang et al. (2009), for example, used Student-t distributed marginals in combination with implicit and Archimedean copulas. Ghorbel/Trabelsi (2009) proposed copula VaR models with EVT margins. Chollette et al. (2009) presented VaR backtests for models with regime-switching copulas and Skewed-t residuals.

Our main contribution to this literature is the introduction of a residual distribution that combines a flexible shape with good simulation performance. In particular, we propose a copula GARCH model using a slight modification of Tukey's distribution (Martinez/Iglewicz, 1984). This transformation of the normal distribution introduces two shape parameters that allow for the calibration of skewness and kurtosis. Due to this flexibility, Badrinath/Chatterjee (1988), Mills (1995) as well as Dutta/Babbel (2002) found a good fit of g-and-h models to the univariate unconditional return distributions of several asset classes. More recent studies by Dutta/Perry (2006) and Degen et al. (2007) discussed the g-and-h distribution in the context of modeling operational risks.

The application of the g-and-h distribution to copula risk models is very promising for two reasons: First, recent studies found that the simultaneous capture of shape asymmetries and fat tails is of high importance for good density forecasts in general and VaR forecasts in particular (Kuester et al., 2006; Bao et al., 2007). Second, the g-and-h distribution is defined by transforming the quantiles of a Gaussian random variable, which makes simulation from this model nearly as fast as drawing normally distributed random numbers. This is important because integrals with respect to the probability distribution of multivariate copula models usually do not have closed form solutions. Therefore, applications – including risk measurement – rely on efficient simulation techniques.

In contrast to simulation, likelihood based parameter estimation in g-and-h based models is somewhat difficult because calculating the model density involves the numerical inversion of the g-and-h transformation function. This may be the reason why the g-and-h distribution has not yet been used in copula or time series models.[2] To overcome or at least ease this problem, we contribute a (semi)-closed form representation of the score vector of a g-and-h based GARCH model, which can be used in the optimization of the likelihood.[3]

In the empirical part of this paper, we compare copula GARCH models based on the g-and-h distribution and Hansen's (1994) skewed-t distribution. The latter distribution provides a similar flexibility and has been used in several copula and VaR studies before.

---

[2] The only multivariate generalization we know about was proposed by Field/Genton (2006) who described wind speed data. Fischer (2010) used a restricted version of the univariate g-and-h distribution to model the residuals of a GARCH process.

[3] "Semi-closed form" means that the derivatives still contain the inverse of the transformation function. However, supplying this expression avoids numerical differentiation of the likelihood, which would require several evaluations of the inversion.

Moreover, we consider restricted cases of the g-and-h and skewed-t distributions, which include the Gaussian case. After comparing the in-sample fit to three major European stock indices, we conduct in- and out-of-sample VaR backtests for all models. In these backtests, g-and-h and skewed-t based copula GARCH models outperform models with less flexible residual distributions based on formal statistical tests on VaR accuracy.

The outline of this paper is as follows: In section 2 we describe the construction of our multivariate g-and-h GARCH model, introducing the univariate g-and-h distribution and briefly reviewing GARCH and copula theory. Parameter estimation is discussed in section 3. Section 4 explains how to derive VaR estimates. In section 5, we report empirical results and section 6 concludes.

2. Model

2.1. Margins

For the marginal return distributions, we suggest the combination of the popular GARCH-filter (Bollerslev, 1986; Taylor, 1986) with a standardized version of Tukey's g-and-h distribution.

According to Martinez/Iglewicz (1984), the g-and-h distribution is derived from a Gaussian random variable using the following transformation function

$$T_{g,h}(y) = \begin{cases} \frac{\exp(gy)-1}{g} \cdot \exp\left(\frac{hy^2}{2}\right) & \text{if } g \neq 0, \\ y \cdot \exp\left(\frac{hy^2}{2}\right) & \text{if } g = 0. \end{cases} \tag{1}$$

The continuity of this transformation in $g = 0$ can be seen by applying L'Hôpital's rule or by using the series expansion of the exponential function. Its differentiability is considered in the appendix. A g-and-h distributed random variable $X_{g,h}$ is defined by transforming a standard normal random variable with $T_{g,h}$. That is

$$X_{g,h} = T_{g,h}(Y), \quad \text{where} \quad Y \sim \mathcal{N}(0,1). \tag{2}$$

This transformation allows for asymmetry and heavy tails. The parameter g determines the direction and the amount of asymmetry. A positive value of g corresponds to a positive skewness. The special symmetric case, which is obtained for $g = 0$, is known as h distribution. For $h > 0$ the distribution is leptokurtic with the mass in the tails increasing in h. If $h = 0$ and $g = 0$, $T_{g,h}$ degenerates to the identity. In this case, $X_{g,h}$ is standard normally distributed. If we require $h \geq 0$, then $T_{g,h}$ is strictly monotonically increasing. Since $T_{g,h}$ is also continuous in y, its inverse function exists, although it is not available in closed form. Denoting this inverse by $I_{g,h}$, the c.d.f. and p.d.f. of $X_{g,h}$

are

$$F_{X_{g,h}}(x) = \Phi_{0,1}\left(I_{g,h}(x)\right), \tag{3}$$
$$f_{X_{g,h}}(x) = \phi_{0,1}(I_{g,h}(x)) \cdot \left(T'_{g,h}(I_{g,h}(x))\right)^{-1}, \tag{4}$$

where $\Phi_{0,1}$ and $\phi_{0,1}$ are the c.d.f. and p.d.f. of the standard normal distribution.

To standardize $X_{g,h}$, its first two moments have to be finite, which can be attained by a further restriction of the parameter space. If $h < 0.5$, both moments exist and have closed form expressions (Hoaglin, 1985). In this case, the mean of $X_{g,h}$ is

$$\mu_{g,h} = \frac{1}{g\sqrt{1-h}}\left[\exp\left(\frac{g^2}{2(1-h)}\right) - 1\right] \tag{5}$$

if $g \neq 0$ and $\mu_h = 0$ otherwise. The variance is given by

$$\sigma^2_{g,h} = \frac{1}{g^2\sqrt{1-2h}} \cdot \left[\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1\right] - \mu^2_{g,h} \tag{6}$$

for $g \neq 0$. If $g = 0$, we obtain $\sigma^2_h = (1-2h)^{-\frac{3}{2}}$. With these expressions a standardized g-and-h distributed random variable $Z_{g,h}$ is defined by

$$Z_{g,h} = \frac{X_{g,h} - \mu_{g,h}}{\sigma_{g,h}}. \tag{7}$$

Obviously, (7) guarantees $\mathbb{E}\left[Z_{g,h}\right] = 0$ and $\text{var}\left[Z_{g,h}\right] = 1$. The density of $Z_{g,h}$ is

$$f_{Z_{g,h}}(z) = \sigma_{g,h} \cdot f_{X_{g,h}}\left(\mu_{g,h} + \sigma_{g,h} \cdot z\right). \tag{8}$$

The effects of different parameter values on the skewness and kurtosis of the standardized g-and-h distribution are illustrated in Figure 1.[4] Illustrations of possible distributions shapes are given in figure 2, where we compare the g-and-h model and its benchmark, the skewed-t model. The skewed generalization of the Student's t distribution that we use dates back to Hansen (1994). For $\lambda \in (-1, 1)$ and $\nu \in (2, \infty)$, a (standardized) skewed-t random variable Z_{st} can be defined by its density

$$f_{Z_{st}}(z) = \begin{cases} bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-\frac{(\nu+1)}{2}} & \text{if } z < -\frac{a}{b}, \\ bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-\frac{(\nu+1)}{2}} & \text{if } z \geq -\frac{a}{b}, \end{cases} \tag{9}$$

[4] To derive skewness and kurtosis of a standardized g-and-h distributed random variable, we use the results from Martinez/Iglewicz (1984) together with the binomial formula.

where

$$a = 4\lambda c \frac{\nu - 2}{\nu - 1}, \qquad b = \sqrt{1 + 3\lambda^2 - a^2} \tag{10}$$

and

$$c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}. \tag{11}$$

For a review of this distribution, see also Jondeau/Rockinger (2003, 2006).

Next, we briefly summarize the time series structure of our marginal model. Let R_t denote the log return of an investment at time t. We assume that

$$R_t = \mu + \varepsilon_t, \tag{12}$$

where μ describes the location of the return distribution[5] and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a series of innovations which follows a GARCH(1,1) process. Hence,

$$\varepsilon_t = \sigma_t Z_t, \qquad t \in \mathbb{Z}, \tag{13}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \qquad t \in \mathbb{Z}, \tag{14}$$

with $(Z_t)_{t\in\mathbb{Z}}$ being a white noise process, i.e. a series of independent, identically distributed random variables that satisfy $\mathbb{E}[Z_t] = 0$ and $\mathbb{E}\left[Z_t^2\right] < \infty$ for all $t \in \mathbb{Z}$. In particular, we use i.i.d. standardized g-and-h and skewed-t random variables for $(Z_t)_{t\in\mathbb{Z}}$. The property $\mathbb{E}[Z_t] = 0$ guarantees that μ corresponds to the unconditional mean of the return distribution. Requiring that $\operatorname{var}[Z_t] = 1$, σ_t^2 corresponds to the conditional variance of ε_t, based on the information available until $t-1$, denoted by $\mathcal{F}_{t-1}$[6].

From (14) it is obvious that $\alpha_0, \alpha_1, \beta_1 > 0$ are sufficient conditions for the positivity of the variance. The second restriction of the parameter space is introduced by the requirement of stationarity. A basic result for the GARCH(1,1) model with normally distributed white noise was obtained by Bollerslev (1986), who showed that $\alpha_1 + \beta_1 < 1$ is sufficient for the weak stationarity of such a process. Nelson (1990) showed that this condition is also sufficient for strict stationarity in more general settings, including models like the one presented in this paper. According to He/Teräsvirta (1999), $\mathbb{E}\left[(\beta_1 + \alpha_1 Z_0^2)^m\right] < 1$ is necessary and sufficient for the existence of $\mathbb{E}\left[\varepsilon^{2m}\right]$. Hence, $\alpha_1 + \beta_1 < 1$ implies strict and weak stationarity of $(\varepsilon_t)_{t\in\mathbb{Z}}$.

---

[5] This model can easily be supplemented by a conditional mean specification.

[6] $\mathcal{F}_{t-1}$ denotes the natural filtration of the return series until $t-1$, i.e. $\mathcal{F}_{t-1} := \sigma(R_s;\ s = 1, \ldots, t-1)$.

Figure 1: Skewness and kurtosis of the standardized g-and-h distribution

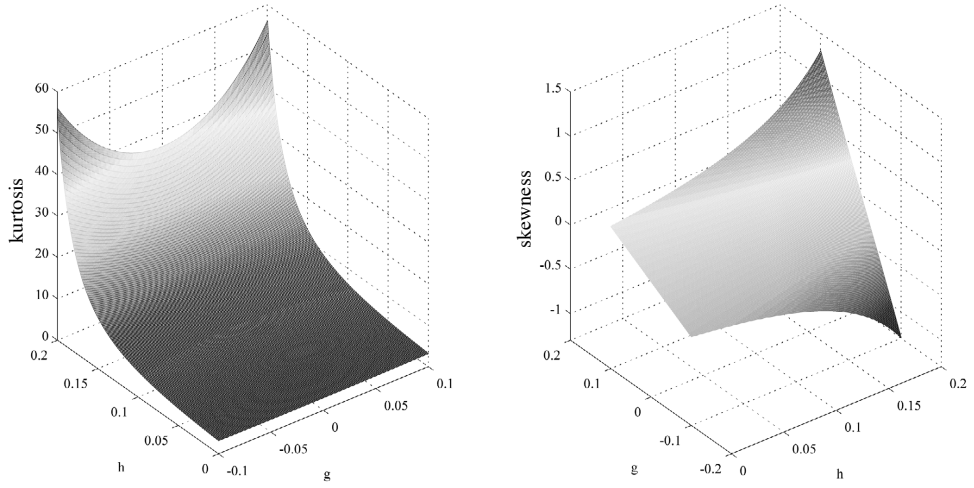
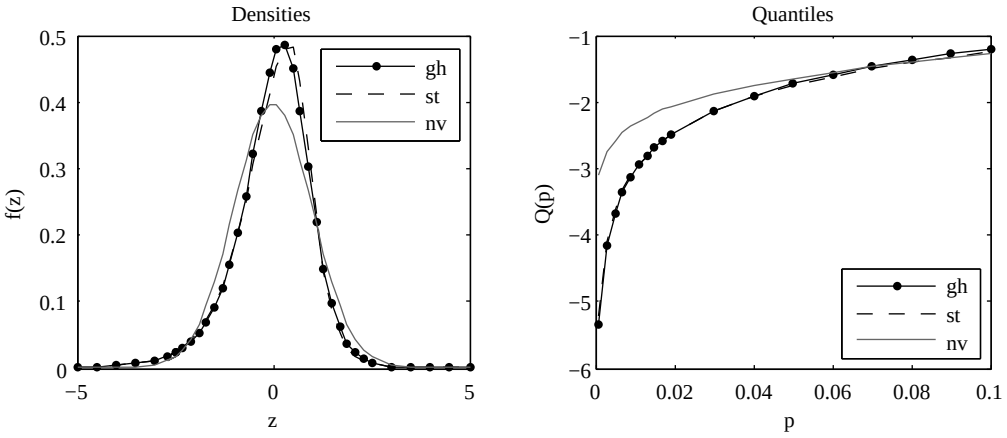


Figure 2: A comparison between the skewed-t and g-and-h distribution



The parameters of the g-and-h and skewed-t distributions are chosen such that both distributions have the same skewness (-1) and kurtosis (7.4).

2.2. Dependence Structure

Let $(\boldsymbol{R}_t)_{t\in\mathbb{Z}}$ be an n-dimensional discrete time stochastic process, which describes the log returns of n assets. The law of the process is given by the conditional c.d.f.s $F_{\boldsymbol{R}_t|\mathcal{F}_{t-1}}$ for $t \in \mathbb{Z}$. The marginal distributions of $\boldsymbol{R}_t$ conditional on $\mathcal{F}_{t-1}$ are denoted by $F_{t,i}(r) = F_{R_{t,i}|\mathcal{F}_{t-1,i}}(r|\mathcal{F}_{t-1,i})$ for $i = 1, \ldots, n$.[7] Models for these distributions were discussed in the last subsection. According to *Sklar's* theorem for conditional distributions (Patton, 2004, Theorem 1), $F_{\boldsymbol{R}_t|\mathcal{F}_{t-1}}$ can be decomposed into its marginal distributions and a copula function $C_{t|\mathcal{F}_{t-1}}$:[8]

$$\begin{aligned} &F_{\boldsymbol{R}_t|\mathcal{F}_{t-1}}(r_1, \ldots, r_n|\mathcal{F}_{t-1}) \\ &\qquad = C_{t|\mathcal{F}_{t-1}}\left(F_{t,1}(r_1), \ldots, F_{t,n}(r_n)|\mathcal{F}_{t-1}\right). \end{aligned} \tag{15}$$

As our focus is primarily on a new distributional assumption for the margins, we restrict our attention to unconditional copula models by setting $C_{t|\mathcal{F}_{t-1}} \equiv C$. In particular, we use the very popular Gaussian and Student's-t copulas. The n-dimensional Gaussian Copula C^n_{Ga} is obtained from the standard normal multivariate c.d.f. $\Phi^n_{\boldsymbol{P}}$ with correlation matrix $\boldsymbol{P}$ by

$$C^n_{Ga}(u_1, \ldots, u_n) := \Phi^n_{\boldsymbol{P}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)), \tag{16}$$

where Φ^{-1} is the quantile function of the one-dimensional standard normal distribution. Let $t^n_{\boldsymbol{P},\nu}$ be the c.d.f. of an n-dimensional t distribution and let t^{-1}_ν be the quantile function of a one-dimensional t distribution. The t copula is then defined by

$$C^n_t(u_1, \ldots, u_n) = t^n_{\boldsymbol{P},\nu}(t^{-1}_\nu(u_1), \ldots, t^{-1}_\nu(u_n)). \tag{17}$$

Just like in the case of the Gaussian copula the $(n \times n)$-matrix $\boldsymbol{P}$ is a correlation parameter. The so-called degrees of freedom parameter ν determines the probability of joint extreme realizations. This probability can be measured by $\lambda_l := \lim_{\alpha\to 0} \mathrm{P}\left(U_2 < \alpha | U_1 < \alpha\right)$ for the lower tail of a bivariate copula C. If this limit exists, λ_l is referred to as lower tail-dependence coefficient. A two-dimensional Gaussian copula cannot exhibit tail dependence and λ_l is zero if $\rho < 1$. In contrast, the tail dependence coefficient of a t copula is increasing in the correlation and decreasing in the degrees of freedom parameter. Therefore, the t copula is often used to capture the possibility of joint extremes.

---

[7] Here, $\mathcal{F}_t$ denotes the natural filtration of the multivariate stochastic process $\boldsymbol{R}_t$ and $\mathcal{F}_{t,i}$ denotes the filtration generated by the i-th margin of $\boldsymbol{R}_t$. We assume that the marginal distributions depend only on their own history, that is $F_{R_{t,i}|\mathcal{F}_{t-1}}(r|\mathcal{F}_{t-1}) := F_{R_{t,i}|\mathcal{F}_{t-1,i}}(r|\mathcal{F}_{t-1,i})$.

[8] In the case of absolutely continuous random variables, this decomposition is unique.

3. Parameter Estimation

Let $\boldsymbol{\Theta}$ denote the parameter space of our model and let $(\boldsymbol{r}_t)_{t=1,\dots,T}$ denote the sample of n-dimensional log returns, which we use for estimation. We assume that $(\boldsymbol{r}_t)_{t=1,\dots,T}$ is generated by the multivariate conditional density $f_{\boldsymbol{R}_t|\mathcal{F}_{t-1}}(\boldsymbol{r}_t|\mathcal{F}_{t-1},\boldsymbol{\theta}_o)$ with $\boldsymbol{\theta}_o \in \boldsymbol{\Theta}$. We suggest to use a conditional[9] maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ that is defined by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{t=1}^{T} l_t(\boldsymbol{r}_t,\boldsymbol{\theta}), \tag{18}$$

where

$$l_t(\boldsymbol{r}_t,\boldsymbol{\theta}) = \log f_{\boldsymbol{R}_t|\mathcal{F}_{t-1}}(\boldsymbol{r}_t|\mathcal{F}_{t-1},\boldsymbol{\theta}). \tag{19}$$

$f_{\boldsymbol{R}_t|\mathcal{F}_{t-1}}(\boldsymbol{r}_t|\mathcal{F}_{t-1},\boldsymbol{\theta})$ is obtained by differentiating the right-hand side of (15). With $c_{\boldsymbol{\theta}}$ denoting the copula density, we see that

$$\begin{aligned} &f_{\boldsymbol{R}_t|\mathcal{F}_{t-1}}(r_{t1},\dots,r_{tn}|\mathcal{F}_{t-1},\boldsymbol{\theta}) \\ &\quad = c_{\boldsymbol{\theta}}\left(F_{t,1}(r_{t1},\boldsymbol{\theta}),\dots,F_{t,n}(r_{tn},\boldsymbol{\theta})\right)\cdot\prod_{i=1}^{n} f_{t,i}(r_{ti},\boldsymbol{\theta}). \end{aligned} \tag{20}$$

The marginal p.d.f.s and c.d.f.s in this equation follow from the results presented in subsection 2.1. We simplify the solution of the optimization problem given in (18), applying the IFM-method proposed by Joe (1997). For this purpose we partition the parameter vector $\boldsymbol{\theta}_o$ into $(\boldsymbol{\theta}_C,\boldsymbol{\theta}_{R_1},\dots,\boldsymbol{\theta}_{R_n})$. $\boldsymbol{\theta}_{R_i}$ contains the parameters of the i-th marginal distribution and $\boldsymbol{\theta}_C$ consists of the copula parameters. The IFM-method is a two step procedure. In the first step,

$$\hat{\boldsymbol{\theta}}_{R_i} = \underset{\boldsymbol{\theta}_{R_i}}{\operatorname{argmax}} \sum_{t=1}^{T} \log f_{t,i}(r_{it}|\boldsymbol{\theta}_{R_i}) \tag{21}$$

is solved for $i = 1,\dots,n$. This means that we consecutively estimate the parameters of the marginal distributions. Given these estimates, we determine the copula parameters by

$$\hat{\boldsymbol{\theta}}_C = \underset{\boldsymbol{\theta}_C}{\operatorname{argmax}} \sum_{t=1}^{T} \log c_{\boldsymbol{\theta}_C}\left(F_{t,1}(r_{t1}|\hat{\boldsymbol{\theta}}_{R_i}),\dots,F_{t,n}(r_{tn}|\hat{\boldsymbol{\theta}}_{R_n})\right). \tag{22}$$

[9] We condition on presample values of $\sigma^2_{t,i}$ and $\varepsilon^2_{t,i}$ for $i = 1,\dots,n$. The choice of these values is detailed in the appendix.

Asymptotic properties of such IFM estimators were derived by Joe (1997). Patton (2006) extends these results to a time series context.[10] Under the standard regularity conditions summarized in Patton (2006, Appendix A) the two step estimator $\hat{\boldsymbol{\theta}}_{2s}$ is consistent and asymptotically normal with

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{2\mathrm{s},T} - \boldsymbol{\theta}_o) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}^{-1}) \qquad \text{for } T \to \infty, \tag{23}$$

where

$$\boldsymbol{B} = -\mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} l_t(\boldsymbol{R}_t, \boldsymbol{\theta}_o)\right] \tag{24}$$

and

$$\boldsymbol{A} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{R}_t, \boldsymbol{\theta}_o) \frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{R}_t, \boldsymbol{\theta}_o)'\right]. \tag{25}$$

The negative expectation of the Hessian $\boldsymbol{B}$ and the expected outer product of the scores $\boldsymbol{A}$ are consistently estimated by the corresponding sample averages, evaluated at $\hat{\boldsymbol{\theta}}_{2\mathrm{s},T}$. Hence,

$$\hat{\boldsymbol{B}}_T = -T^{-1}\sum_{t=1}^{T} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} l_t(\boldsymbol{r}_t, \hat{\boldsymbol{\theta}}_{2\mathrm{s},T}) \tag{26}$$

and

$$\hat{\boldsymbol{A}}_T = T^{-1}\sum_{t=1}^{T} \frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{r}_t, \hat{\boldsymbol{\theta}}_{2\mathrm{s},T}) \frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{r}_t, \hat{\boldsymbol{\theta}}_{2\mathrm{s},T})' \tag{27}$$

can be used to calculate standard errors of the estimated parameters.

For a fast implementation of the proposed estimation algorithm, we derive an analytical expression of the score vector for the univariate g-and-h model in the appendix. Using this result avoids numerical differentation in the optimization of the likelihood, which would require several time consuming inversions of the g-and-h transformation. This problem is probably the reason why our multistage MLE approach, which is very common in the field of copula modeling, has very rarely been used for g-and-h based models.[11] Nearly all former papers, including Dutta/Perry (2006) as well as Field/Genton (2006), exclusively use the quantile method proposed by Hoaglin (1985). This approach has the advantage of easy implementation and low computational cost. However, the application of this simple method involves a number of disadvantages: First and most important, this method was designed for unconditional univariate distributions and there is no direct generalization to the time series context presented here. Moreover, the estimation of g and h proceeds sequentially. Furthermore, the choice of quantiles used for estimation is arbitrary and the restriction of the parameter space ($0 < h < 0.5$) cannot be

[10] Patton (2006) considers a more general setting in which the copula may be time dependent and the samples used for estimating the margins may be of different lengths.

[11] Rayner/MacGillivray (2002) do a Monte Carlo study on the in-sample efficiency of MLE for generalized g-and-k distributions. To our knowledge, the first applied paper using this methodology is Fischer (2010), who considers the symmetric case ($g \equiv 0$).

implemented.

4. Risk Measurement

In this section, we show how to apply the return model developed so far to risk measurement, where we focus on VaR (Dowd, 1998; Jorion, 2001). Although, this risk measure is critized due to conceptional deficiencies (Tasche, 2002; Szegö, 2002), it remains the industry standard for institutional investors. Since we use conditional return distributions to derive our VaR estimates, these estimates will be time-varying. The relevance of such estimates for risk management is e.g. discussed in McNeil/Frey (2000). In brief, a VaR derived from the unconditional return distribution could be used to determine longterm risk capital requirements, whereas the VaR based on the conditional distribution may give important indications on reducing risky exposures in times of high volatility.

4.1. Estimation of Value at Risk

Let L denote an absolutely continuous random variable which measures the loss of a position during the period under consideration. For $\alpha \in (0,1)$, we define VaR by

$$\mathrm{VaR}_{\alpha,L} = Q_L(1-\alpha) \qquad \Leftrightarrow \qquad F_L(\mathrm{VaR}_{\alpha,L}) = 1-\alpha, \tag{28}$$

where Q_L is the quantile function of L. Hence, the probability of a loss higher than the VaR is equal to α. To avoid assumptions concerning the initial capital invested, we set $L_t = -R_t$, i.e. we consider negative returns as (relative) losses. We are particularly interested in the one-step ahead predictive loss distribution. This is stressed by writing $\mathrm{VaR}_{\alpha,L_{t+1}|\mathcal{F}_t}$.

In the univariate case, a closed form expression of $\mathrm{VaR}_{\alpha,L_{t+1}|\mathcal{F}_t}$ is easily obtained because R_t is an increasing, bijective transformation of Z_t. Thus,

$$\mathrm{VaR}_{\alpha,L_{t+1}|\mathcal{F}_t} = -\mu + \sigma_{t+1} \cdot Q_{-Z}(1-\alpha) = -\mu - \sigma_{t+1} \cdot Q_Z(\alpha), \tag{29}$$

where σ_{t+1} is calculated based on $\mathcal{F}_t$ according to the GARCH(1,1) equation.

We now consider a multivariate setting, where we derive the VaR of a portfolio based on the n-dimensional return density $f_{\boldsymbol{R}_{t+1}|\mathcal{F}_t}$ of the assets contained in this portfolio. If we choose a vector of portfolio weights $\boldsymbol{w}$, the portfolio return is given by $R_{p,t+1} := g(\boldsymbol{R}_{t+1}, \boldsymbol{w})$ with

$$g(\boldsymbol{x}, \boldsymbol{w}) := \log\left(\exp(\boldsymbol{x}') \cdot \boldsymbol{w}\right). \tag{30}$$

Its distribution is

$$F_{R_{p,t+1}|\mathcal{F}_t}(r_p) = \mathbb{E}\left[\mathbb{1}\left\{g(\boldsymbol{R}_{t+1}, \boldsymbol{w}) \leq r_p\right\}|\mathcal{F}_t\right]. \tag{31}$$

In many cases, including the models discussed in section 2, a closed form expression of this expectation is not available. One way to evaluate the c.d.f. of the portfolio return is a Monte Carlo simulation, that is generating a sample of m random vectors $(\boldsymbol{r}_j)_{j=1,\ldots,m}$ from the density $f_{\boldsymbol{R}_{t+1}|\mathcal{F}_t}$.[12] By the strong law of large numbers

$$\frac{1}{n}\sum_{j=1}^{m} \mathbb{1}\left\{r_{p,j} \leq r_p\right\} \to F_{R_{p,t+1}|\mathcal{F}_t}(r_p), \qquad \text{P-a.s.}, \tag{32}$$

where $r_{p,j} := g(\boldsymbol{r}_j, \boldsymbol{w})$. The α-quantile of the simulated sample $(r_{p,j})_{j=1,\ldots,m}$ is an estimator of the α-quantile of $R_{p,t+1}$. Noting that $Q_{-X}(1-\alpha) = -Q_X(\alpha)$, we set

$$\widehat{\mathrm{VaR}}_{\alpha,L_{p,t+1}|\mathcal{F}_t} = -r_{p,(\lfloor \alpha m \rfloor)}, \tag{33}$$

where $r_{(i)}$ denotes the i-th oberservation of an ordered sample.

To implement the described method, we must be able to draw random samples from the presented copula GARCH models. The corresponding simulation procedure works as follows:

1. Simulate a random vector $(u_1, \ldots, u_n)$ from the chosen copula.
2. Transform u_i into a corresponding log return r_i by applying the quantile functions of the marginal models for $i = 1, \ldots, n$.

In the case of a Gaussian copula with g-and-h margins, this procedure reduces to the simulation of a random vector from a multivariate normal distribution[13] and the componentwise application of the appropriate g-and-h as well as location scale transformations to this vector. This explains why simulation from a g-and-h model is numerically very similar to simulation from a multivariate normal distribution. Moreover, the discussion reveals that the Gauss copula case of a g-and-h model can be seen as a CCC-GARCH model (Bollerslev, 1990) with a more complex residual distribution.[14]

[12] An alternative method would be numerical integration. This alternative is less suitable here because we are not interested in a direct evaluation of the c.d.f. but in evaluating its inverse function. Thus, we would have to combine two numerical procedures: inversion and integration.

[13] With standardized margins and the estimated correlation matrix.

[14] Finally, the connection to the multivariate g-and-h model proposed by Field/Genton (2006) becomes apparent. These authors propose to transform the uncorrelated vector of standard normal random variables, whereas we consider correlation first and subsequently apply the g-and-h transformations.

4.2. Evaluation of Value at Risk

To assess the accuracy of VaR forecasts, we use the ideas developed by Christoffersen (1998). Given a sequence of losses $(L_t)_{t=1,\dots,T}$ and VaR forecasts $(\mathrm{VaR}_{t|\mathcal{F}_{t-1}})_{t=1,\dots,T}$, we define the corresponding hit (violation) series by

$$H_t = \mathbb{1}\left[L_t > \mathrm{VaR}_{\alpha,t|\mathcal{F}_{t-1}}\right], \qquad t = 1, \dots, T. \tag{34}$$

According to Christoffersen (1998, Definition 2), a VaR series is efficient if

$$\mathbb{E}\left[H_t | H_{t-1}, \dots, H_1\right] = \alpha, \qquad \text{for all } t = 1, \dots, T. \tag{35}$$

Thus, efficiency means that the conditional probability of a hit is equal to α for all $t = 1, \dots, T$. By the law of iterated expectations, this requirement implies that the unconditional hit probability corresponds to α as well. Beyond that, (35) ensures that hits do not occur in clusters. Efficiency, as definied in (35), implies that the hit series is i.i.d. Bernoulli with parameter α (Christoffersen, 1998, Lemma 1).

This observation is used to design a two step test on VaR-efficiency. In the first step, we assume independence and test on the parameter of the Bernoulli distribution. This corresponds to testing the null hypothesis $H_0 : \mathbb{E}[H_t] = \alpha$ against $H_1 : \mathbb{E}[H_t] \neq \alpha$. Then, the likelihood ratio test statistic LR_{uc} corresponds to the classical proportion of failure VaR-test proposed by Kupiec (1995). The second step is to test the independence assumption. Christoffersen (1998) constructs a likelihood ratio test statistic LR_{ind} assuming an explicit alternative hypothesis, a first order Markov chain. If we ignore the first observation in the calculation of LR_{uc}, both statistics can be combined to a test statistic on correct conditional coverage, which is $LR_{cc} := LR_{uc} + LR_{ind}$. LR_{cc} has an asymptotic χ^2 distribution with two degrees of freedom.

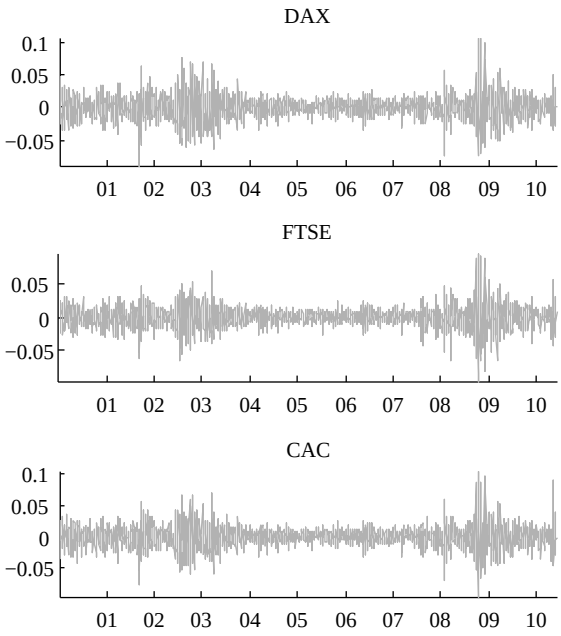
5. Empirical Results

5.1. The Data Set

We fit the models proposed in section 2 to three major European stock indices, the German DAX 30, the British FTSE 100 and the French CAC 40. Our sample covers a period from January 2000 to May 2010. We collected price data from Datastream to calculate continuously compounded daily returns. After eliminating observations that are affected by non-trading days, we are left with a three dimensional return sample of 2638 observations, which is depicted in Figure 3.

Descriptive statistics and standard test results for this sample are summarized in Table 1.

Figure 3: Return series



According to the results of Jarque/Bera (1987) tests, the univariate samples are far from being normally distributed. This is mainly due to the very high sample kurtosis. High Ljung Box test statistics for the series of squared returns give an indication for the existence of volatility clusters. The descriptive statistics of the standardized residuals of a normal GARCH(1,1) model show that a large amount of excess kurtosis is removed by the GARCH filter. However, sample kurtosis is still significantly larger than three and sample skewness is distinctively negative. Therefore, the assumption of normality is also rejected for the residual series at the 1% significance level. This observation motivates the application of return models that allow for conditional skewness and excess kurtosis.

5.2. Estimation Results

First, we present the estimation results of the marginal return models.[15] In particular, we compare the fit of GARCH(1,1) models with g-and-h and skewed-t distributed residuals. We also report estimation results for restricted versions of these models, that is for GARCH(1,1) models with h, t and normally distributed residuals. These specifications

[15] We implemented the estimation procedure described in section 3 in Matlab 2010b. The numerical maximization of the log-likelihood is performed using the Matlab optimizer `fmincon` (with active set algorithm and user supplied gradient as described in the appendix). The starting values for the parameters of the volatility model and residual distributions are determined by pre-estimations. Matlab codes implementing g-and-h related functions (including the derivatives presented in the appendix) are available from the authors.

Table 1: Descriptive statistics

	DAX	FTSE	CAC
data			
start	05/01/2000	05/01/2000	05/01/2000
end	31/05/2010	31/05/2010	31/05/2010
T	2638	2638	2638
log returns r_t			
mean	-3.8e-5	-7.4e-5	-6.0e-5
median	7.4e-4	3.6e-4	2.5e-4
min [%]	-8.87	-9.48	-9.47
max [%]	10.80	9.65	10.59
std [%]	1.67	1.46	1.59
skew	0.05	-0.06	0.09*
kurt	7.20***	8.69***	8.07***
JB-test	1940***	3557***	2825***
LB	1389***	2000***	1350***
GARCH(1,1) residuals $\hat{z}_t$			
skew $\hat{z}$	-0.30***	-0.21***	-0.25***
kurt $\hat{z}$	3.88***	3.64***	3.92***
JB-test	123.5***	63.6***	120.6***
Sample correlations			
DAX	1.00		
FTSE	0.78	1.00	
CAC	0.88	0.86	1.00

skew *and* kurt *are sample skewness and kurtosis.* JB *is the value of the Jarque/Bera (1987) test statistic.* LB *is the test statistics proposed in Ljung/Box (1978) applied to the squared log returns for 10 lags.*
*The marks ***, ** and * denote significance at 1%, 5% and 10% levels. The same marks will be used in the following tables.*

are labeled gh-, st-, h-, t- and nv-GARCH. Table 2 summarizes several goodness of fit statistics for these models. We observe that the likelihoods of the models with skewed-t and the g-and-h distributed residuals are very close together. Moreover, the full models are preferred to the restricted versions for all three time series according to likelihood ratio tests and AIC based model selection.

Table 2: Margins – goodness of fit

		DAX	FTSE	CAC
S1: nv-GARCH	$\mathcal{L}$	7581.1	7985.1	7710.7
	AIC	-15154.2	-15962.2	-15413.4
S2: t-GARCH	$\mathcal{L}$	7597.2	7999.1	7729.7
	LR(S2/S1)	32.2***	27.9***	38.1***
	AIC	-15184.5	-15988.1	-15449.5
S3: st-GARCH	$\mathcal{L}$	7604.3	8002.4	7735.5
	LR(S3/S2)	14.1***	6.6***	11.6***
	AIC	-15196.6	-15992.8	-15459.1
S4: h-GARCH	$\mathcal{L}$	7597.1	7999.0	7729.5
	LR(S4/S1)	32.0***	27.8***	37.6***
	AIC	-15184.2	-15988.0	-15449.0
S5: gh-GARCH	$\mathcal{L}$	7604.6	8002.8	7735.2
	LR(S5/S4)	15.0***	7.6***	11.5***
	AIC	-15197.2	-15993.6	-15458.5

$\mathcal{L}$ is the attainted value of the sample log-likelihood. LR(Si/Sj) *is the test statistics of a likelihood ratio test between the specifications* Si *and* Sj. AIC *is the value of the Akaike Information Criterion, which is defined as* $AIC = 2k - 2\mathcal{L}$*, where* k *is the number of model parameters.*

Table 3 shows the parameter estimates for the g-and-h and skewed-t models.[16] Note that the mean and variance parameters of both specifications are very similar and that the estimates of the kurtosis (ν, h) and the skewness (λ, g) parameters are significant for all indices.

The results of estimating the copula parameters for g-and-h distributed margins are presented in Table 4.[17] For both copula models and all stock index combinations, the dependency parameters are very high. The low values of ν imply that the European stock markets exhibit non-negligible tail-dependence. This explains the much better fit of the t copula according to the attained log-likelihoods and AICs.

5.3. Value-at-Risk Backtests

In this section, we present the results of backtesting one step ahead VaR forecasts based on our uni- and multivariate models. We apply the formal tests from section 4.2 to assess

[16] Estimation results for the restricted models are available from the authors upon request.

[17] The results for st-GARCH margins are very similar and can be obtained from the authors upon request.

Table 3: Margins – parameter estimates (st- and gh-GARCH)

st-GARCH	μ	α_0	α_1	β_1	ν	λ
DAX	5.8e-4	1.6e-6	0.089	0.906	14.5	-0.102
s.e.	(2.2e-4)	(0.6e-6)	(0.012)	(0.011)	(4.27)	(0.025)
FTSE	4.7e-4	1.4e-6	0.097	0.898	14.3	-0.073
s.e.	(1.9e-4)	(0.5e-6)	(0.014)	(0.014)	(3.39)	(0.028)
CAC	5.1e-4	1.4e-6	0.084	0.912	13.2	-0.097
s.e.	(2.1e-4)	(0.5e-6)	(0.012)	(0.011)	(3.21)	(0.029)

gh-GARCH	μ	α_0	α_1	β_1	g	h
DAX	6.0e-4	1.5e-6	0.090	0.906	-0.071	0.034
s.e.	(2.2e-4)	(0.6e-6)	(0.012)	(0.011)	(0.018)	(0.011)
FTSE	4.6e-4	1.4e-6	0.097	0.898	-0.051	0.036
s.e.	(1.9e-4)	(0.5e-6)	(0.014)	(0.014)	(0.019)	(0.010)
CAC	5.0e-4	1.4e-6	0.085	0.911	-0.063	0.040
s.e.	(2.1e-4)	(0.5e-6)	(0.012)	(0.011)	(0.019)	(0.011)

This table shows parameter estimates that were obtained in the first estimation step of the MLE approach described in section 3 and the appendix. Moreover, robust standard errors (s.e.) for these estimates are given.

Table 4: Copula – estimation results (gh-GARCH margins)

Gaussian copula	ρ	s.e.	ν	s.e.	$\mathcal{L}$ cop	$\mathcal{L}$	# par	AIC
DAX/FTSE	0.77	(0.008)	-	-	1197.6	16805.0	15	-33579.9
DAX/CAC	0.90	(0.004)	-	-	2013.2	17353.1	15	-34676.2
FTSE/CAC	0.81	(0.007)	-	-	1581.4	17319.4	15	-34608.9

t copula	ρ	s.e.	ν	s.e.	$\mathcal{L}$ cop	$\mathcal{L}$	# par	AIC
DAX/FTSE	0.78	(0.006)	3.2	(0.1)	1341.9	16949.3	16	-33866.6
DAX/CAC	0.90	(0.003)	2.5	(0.1)	2326.9	17666.7	16	-35301.4
FTSE/CAC	0.84	(0.004)	3.0	(0.1)	1718.5	17456.5	16	-34881.1

This table reports parameter estimates that were obtained in the second step of the MLE approach described in section 3 and the appendix. Robust standard errors (s.e.) for these estimates are shown. ρ is the correlation parameter of the bivariate copulas. ν corresponds to the estimated degrees of freedom for the t copula. $\mathcal{L}$ cop and $\mathcal{L}$ are the sample log-likelihoods of the copula and the full model. # par is the number of parameters of the full bivariate specifications.

the accuracy of in- and out-of-sample VaR-forecasts. For the in-sample tests, we use the full length time series and the parameter estimates presented in the last subsection. For the out-of-sample tests, we allow model parameters to change over time by using moving estimation windows. The size of the estimation windows is 1000 observations, so that the first window ends on December 16, 2003. Rolling our estimation window forward by one trading day, we re-estimate the model 1637 times and obtain as many VaR-predictions. We calculate VaR series for $\alpha = 5\%, 1\%, 0.1\%$.

We first present the backtest results for the nine VaR-series (3 indices, 3 α-levels) for each of the five univariate models. In this case, the VaR can be obtained in closed form using (29). In- and out-of-sample results are summarized in Table 5, where hit ratios and p-values for tests on correct unconditional and conditional coverage are reported. Large differences between in- and out-of-sample hit ratios are observed, which emphasizes the importance of out-of-sample testing. Hit ratios are higher than the target values across all models and indices, especially the out-of-sample hit ratios.[18] The degree of deviation substantially differs between the tested VaR-models. At the 5% significance level, the nv-GARCH forecasts have to be rejected for eight of nine combinations according to the conditional coverage LR-test. VaR-forecasts from models with leptocurtic residuals (h- and t-GARCH) only fail the test for a third of the VaR series. A further improvement is attained by allowing for conditional kurtosis and skewness. The hypothesis of correct conditional coverage cannot be rejected in eight of nine times for the gh- and st-GARCH models. These observations support the findings of Kuester et al. (2006) and Giot/Laurent (2004), who report a high forecasting performance of GARCH-filtered skewed-t models. In our case, these findings apply to the g-and-h specification as well.

Next, we report backtest results for VaR-predictions from our multivariate models. We build three different two-index portfolios from the univariate series by mixing these with equal weights.[19] In this case, the computations are considerably more expensive because all VaR-forecasts have to be obtained using the Monte Carlo approach described in (33). The simulations are performed with a sample size of one million random numbers for each forecast. An illustration of selected VaR series is given in Figure 4. Numerical results of the backtests are reported in Table 6. We first focus on the impact of the choice between Gaussian and t copula. Counting rejections of the correct conditional coverage hypothesis at the 5% significance level, differences due to the copula specification are only observed in three cases. These results add to the evidence provided by Ané/Kharoubi (2003) and Rosenberg/Schuermann (2006), who discovered that the copula specification has a limited impact on the VaR-estimates compared to the marginal distribution.[20] A comparison of the results from gh- or st-GARCH and nv-GARCH spec-

[18] This may be due to extreme stock price movements during the 2008 financial crisis.
[19] The relative weights of the indices are kept constant during the testing period.
[20] For a recent simulation study on the topic see Fantazzini (2009).

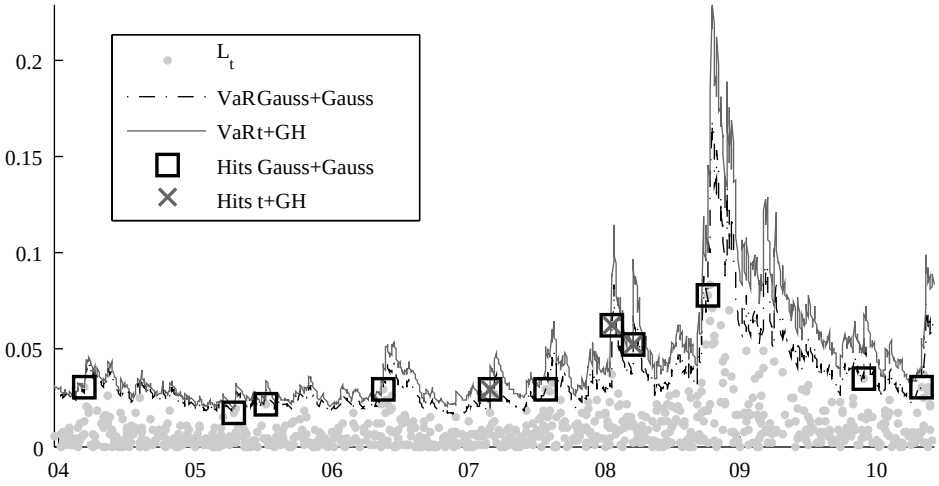
Table 5: Univariate Value-at-Risk backtest

	In-sample										Out-of-sample									
	DAX			FTSE			CAC				DAX			FTSE			CAC			
100α	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	#	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	#
nv-GARCH																				
5.0	6.7	0.00	0.00	5.5	0.25	0.51	6.1	0.01	0.02	2	6.5	0.01	0.00	5.5	0.36	0.43	6.6	0.00	0.02	2
1.0	1.3	0.15	0.27	1.8	0.00	0.00	1.6	0.01	0.01	2	2.0	0.00	0.00	2.1	0.00	0.00	2.2	0.00	0.00	3
0.1	0.3	0.03	0.08	0.5	0.00	0.00	0.3	0.00	0.01	2	0.5	0.00	0.00	0.6	0.00	0.00	0.5	0.00	0.00	3
										6										**8**
t-GARCH																				
5.0	6.8	0.00	0.00	5.7	0.11	0.28	6.4	0.00	0.00	2	6.8	0.00	0.00	6.1	0.05	0.10	6.8	0.00	0.00	2
1.0	1.1	0.49	0.51	1.4	0.07	0.12	1.2	0.29	0.40	0	1.6	0.03	0.06	1.8	0.00	0.01	1.6	0.03	0.06	1
0.1	0.1	0.83	0.97	0.2	0.20	0.43	0.2	0.20	0.43	0	0.3	0.03	0.11	0.2	0.12	0.29	0.2	0.12	0.29	0
										2										**3**
st-GARCH																				
5.0	5.9	0.04	0.04	5.3	0.53	0.79	5.5	0.28	0.28	1	6.1	0.06	0.06	5.7	0.21	0.23	6.5	0.01	0.03	1
1.0	0.9	0.64	0.42	1.0	0.90	0.75	0.8	0.38	0.56	0	1.2	0.38	0.34	1.5	0.05	0.09	1.3	0.27	0.30	0
0.1	0.1	0.83	0.97	0.2	0.20	0.43	0.2	0.44	0.73	0	0.2	0.12	0.29	0.2	0.12	0.29	0.2	0.34	0.63	0
										1										**1**
h-GARCH																				
5.0	6.8	0.00	0.00	5.7	0.11	0.28	6.4	0.00	0.00	2	6.8	0.00	0.00	6.1	0.05	0.10	6.8	0.00	0.00	2
1.0	1.1	0.49	0.51	1.4	0.07	0.12	1.2	0.29	0.40	0	1.5	0.05	0.10	1.8	0.00	0.01	1.5	0.05	0.10	1
0.1	0.1	0.83	0.97	0.2	0.20	0.43	0.2	0.20	0.43	0	0.3	0.03	0.11	0.3	0.03	0.11	0.2	0.12	0.29	0
										2										**3**
gh-GARCH																				
5.0	5.9	0.04	0.04	5.3	0.53	0.79	5.5	0.25	0.37	1	6.0	0.07	0.08	5.7	0.21	0.23	6.4	0.01	0.04	1
1.0	0.9	0.50	0.35	1.0	0.94	0.77	0.8	0.28	0.47	0	1.2	0.52	0.38	1.5	0.08	0.14	1.3	0.27	0.30	0
0.1	0.1	0.83	0.97	0.2	0.44	0.73	0.2	0.44	0.73	0	0.2	0.34	0.63	0.2	0.12	0.29	0.2	0.34	0.63	0
										1										**1**

hr denotes the attained hit ratio in percent, i.e. the relative number of VaR *violations.* p_{uc} *and* p_{cc} *are p-values of tests on correct unconditional and conditional coverage that we discussed in section 5.3. In the last column, labeled by #, we count the number of rejections of the hypothesis of correct conditional coverage at a 5% significance level.*

ifications emphasizes that the modeling of conditional skewness and kurtosis distinctively improves the forecasting performance. The differences between the g-and-h and skewed-t specifications are only very small.

Figure 4: Value-at-Risk backtest DAX/FTSE ($\alpha = 0.1\%$)



So far, we have seen that g-and-h and skewed-t models provide a very similar fit to stock market data and highly comparable VaR forecasts. However, they differ significantly concerning the computational cost of their implementation. Table 7 provides the computing times of the different specifications for the out-of-sample VaR-backtest. These results emphasize that the computational cost of estimating a skewed-t model is only little higher than in the case of Gaussian residuals, whereas estimating the g-and-h model takes much longer. This relation turns into the opposite when considering simulation from the models. In this case, differences between gh- and nv-GARCH models are negligible, whereas the computational cost for the skewed-t approach is much higher.

To understand the reported differences, we recall that the inversion of the g-and-h transformation is not available in closed form and thus neither is its density. This is the reason for which we provide analytical scores of the likelihood. The differences in simulation performance can be traced back to the time for evaluating the quantile functions of the margins.[21]

[21] The absolute values of these differences are hardware- and implementation-specific. In our test setting (bwGRiD (2010), one node, Matlab 2010b) applying the quantile function of the g-and-h distribution to a $10^6 \times 1$ random vector from a uniform distribution takes 0.07 seconds whereas the same calculations for the skewed-t model take 3.37 seconds.

Table 6: Multivariate Value-at-Risk backtest

	In-sample										Out-of-sample									
	DAX/FTSE			DAX/CAC			FTSE/CAC				DAX/FTSE			DAX/CAC			FTSE/CAC			
100α	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	#	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	hr	p_{uc}	p_{cc}	#
Gauss copula, nv-GARCH																				
5.0	6.1	0.01	0.03	6.6	0.00	0.00	5.9	0.04	0.10	2	6.3	0.02	0.06	6.6	0.00	0.01	6.4	0.02	0.03	2
1.0	1.5	0.02	0.06	1.6	0.00	0.02	1.4	0.03	0.09	1	2.4	0.00	0.00	2.2	0.00	0.00	2.2	0.00	0.00	3
0.1	0.3	0.00	0.01	0.3	0.01	0.03	0.4	0.00	0.00	3	0.7	0.00	0.00	0.5	0.00	0.00	0.6	0.00	0.00	3
										6										**8**
Gauss copula, st-GARCH																				
5.0	5.7	0.09	0.15	5.7	0.09	0.09	5.7	0.11	0.24	0	6.0	0.07	0.19	6.3	0.02	0.04	6.2	0.03	0.05	2
1.0	1.0	0.90	0.55	1.0	0.90	0.55	0.8	0.38	0.56	0	1.3	0.18	0.24	1.5	0.05	0.10	1.4	0.12	0.22	0
0.1	0.2	0.44	0.73	0.2	0.44	0.73	0.2	0.44	0.73	0	0.2	0.12	0.29	0.3	0.03	0.11	0.2	0.12	0.29	0
										0										**2**
Gauss copula, gh-GARCH																				
5.0	5.7	0.09	0.15	5.9	0.04	0.05	5.7	0.11	0.24	0	5.9	0.09	0.23	6.3	0.02	0.04	6.3	0.02	0.04	2
1.0	1.0	0.94	0.53	1.0	0.94	0.53	0.8	0.38	0.56	0	1.3	0.27	0.30	1.4	0.12	0.19	1.3	0.18	0.31	0
0.1	0.2	0.44	0.73	0.1	0.83	0.97	0.1	0.83	0.97	0	0.2	0.12	0.29	0.3	0.03	0.11	0.2	0.12	0.29	0
										0										**2**
t-copula, nv-GARCH																				
5.0	6.1	0.01	0.03	6.4	0.00	0.00	5.9	0.04	0.08	2	6.3	0.02	0.06	6.6	0.00	0.01	6.4	0.02	0.03	2
1.0	1.4	0.07	0.16	1.4	0.03	0.09	1.2	0.38	0.47	0	2.1	0.00	0.00	2.2	0.00	0.00	2.1	0.00	0.00	3
0.1	0.3	0.03	0.08	0.3	0.01	0.03	0.3	0.00	0.01	2	0.4	0.00	0.01	0.5	0.00	0.00	0.5	0.00	0.00	3
										4										**8**
t-copula, st-GARCH																				
5.0	5.8	0.07	0.14	5.6	0.16	0.16	5.8	0.08	0.17	0	6.1	0.06	0.15	6.3	0.02	0.04	6.2	0.03	0.05	2
1.0	0.9	0.79	0.48	0.9	0.64	0.42	0.8	0.28	0.47	0	1.3	0.18	0.24	1.3	0.18	0.24	1.3	0.18	0.31	0
0.1	0.2	0.44	0.73	0.1	0.83	0.97	0.1	0.83	0.97	0	0.2	0.34	0.63	0.3	0.03	0.11	0.2	0.12	0.29	0
										0										**2**
t-copula, gh-GARCH																				
5.0	5.8	0.08	0.17	5.7	0.09	0.09	5.7	0.09	0.20	0	6.1	0.06	0.15	6.2	0.04	0.07	6.3	0.02	0.04	1
1.0	0.9	0.79	0.48	0.9	0.64	0.42	0.8	0.19	0.37	0	1.2	0.38	0.34	1.3	0.27	0.30	1.3	0.18	0.31	0
0.1	0.2	0.44	0.73	0.1	0.83	0.97	0.1	0.83	0.97	0	0.2	0.34	0.63	0.2	0.34	0.63	0.2	0.12	0.29	0
										0										**1**

See Table 5 for an explanation of the symbols.

Table 7: Computing time – out-of-sample backtests

	estimation time [in min]			simulation time [in min]		
	DAX	FTSE	CAC	DAX FTSE	DAX CAC	FTSE CAC
nv-GARCH	8	7	8	34	33	33
st-GARCH	13	12	12	316	314	311
gh-GARCH	538	544	575	39	38	38

Columns 2 - 4 report the computing time for the estimation step of the backtests, i.e. the time for 1638 estimations of the marginal model given a sample size of 1000 observations. Columns 5-7 show the computing time for the VaR*-simulations in the case of the Gaussian copula. The computations were performed using Matlab 2010b on bwGRiD (2010) with a single node.*

6. Conclusion

In the first part of this paper, we discussed the usage of the g-and-h distribution in multivariate VaR models. In particular, we suggested building copula GARCH models with residuals from the standardized g-and-h distribution. This construction allows the capture of conditional fat tails and asymmetries, which are important stylized facts of financial return series. We adressed the problem of parameter estimation for this distribution and demonstrated that standard likelihood based estimation can be used. To increase the performance in numerical implementations, we derived the gradient of the density of a g-and-h GARCH(1,1) model. As an important application of this model, we discussed how to derive and evaluate risk forecasts.

In the empirical section, we fitted copula models with GARCH filtered margins to the return series of major European stock indices and compared the model fit for different residual distributions. Our g-and-h approach and its skewed-t benchmark provided a similar fit. These models were preferred to restricted versions according to likelihood ratio tests and AIC-based model selection. We presented backtest results for one-day ahead VaR forecasts based on our bivariate return models. In this context, we found further evidence for the importance of capturing shape asymmetries. The g-and-h and skewed-t models clearly outperformed the specifications with symmetric, leptokurtic residuals according to formal tests on correct conditional coverage in univariate and multivariate settings. Furthermore, our empirical study confirmed implementation specific advantages of the g-and-h and skewed-t based copula models. The likelihood based fitting of the latter is faster, the former profits from a better simulation performance.

Further research concerning g-and-h based copula models could address other applications like integrated risk management (Rosenberg/Schuermann, 2006) or high dimensional portfolio selection problems, which would benefit from the good simulation performance. Finally, we would like to reiterate that it is quite easy to upgrade existing implementations of Gaussian time series models using the g-and-h approach due to its definition based on a quantile transformation.

A. Gradient of the log-likelihood

Derivatives of the Inverse g-and-h Transformation: First, we use implicit function arguments to derive the gradient of the inverse g-and-h transformation I.[22] Recall that I is only defined implicitly by

$$T(I(x,g,h),g,h) = x, \tag{36}$$

[22] In this part text, we adopt a more flexible notation for the g-and-h transformation. Most of the time, we consider g and h as variables by writing $T(y,g,h)$ and respectively $I(x,g,h)$. We sometimes suppress the dependence on g and/or h to simplify notation. In those cases, we denote the functions by $T(y)$ or $T(y,\tau)$ and accordingly $I(x)$ or $I(x,\tau)$ for $\tau \in \{\, g,h \,\}$.

for $x \in \mathbb{R}$, $g \in \mathbb{R}$ and $h \in (0, 0.5)$ with T given in (1). Therefore, we cannot explicitly compute its derivatives. However, by differentiating (36) with respect to x and rearranging, we obtain the well-known result

$$\frac{dI}{dx}(x) = \frac{1}{\frac{dT}{dy}(I(x))}. \tag{37}$$

In the same way, we calculate the partial derivatives with respect to g and h. For $\tau \in \{ g, h \}$, it follows that

$$\frac{\partial I}{\partial \tau}(x,\tau) = -\frac{\frac{\partial T}{\partial \tau}(I(x,\tau),\tau)}{\frac{\partial T}{\partial y}(I(x,\tau),\tau)}, \qquad \text{for } \tau \in \{ g, h \}. \tag{38}$$

The implementation of (37) and (38) requires the computation of $\frac{\partial T}{\partial y}$, $\frac{\partial T}{\partial g}$ and $\frac{\partial T}{\partial h}$, which we provide in Table 8. Applying L'Hôpital's rule, it can easily be checked that these derivatives are continuous in $g = 0$. Moreover, $\frac{\partial T}{\partial y}$ is positive for $h \geq 0$, so that (37) and (38) are always well defined. To derive the gradient of the density of a g-and-h distributed random variable, we need second order derivatives of the inverse transformation that are of the form $\frac{\partial^2 I}{\partial \tau \partial x}$ for $\tau \in \{ x, g, h \}$. By applying the chain rule of differentiation, we conclude from (37) that

$$\begin{aligned}\frac{\partial^2 I}{\partial \tau \partial x}(x,\tau) = &-\left(\frac{\partial T}{\partial y}(I(x,\tau),\tau)\right)^{-2} \\ &\cdot \left(\frac{\partial^2 T}{\partial y^2}(I(x,\tau),\tau) \cdot \frac{\partial I(x,\tau)}{\partial \tau} + \frac{\partial^2 T}{\partial \tau \partial y}(I(x,\tau),\tau)\right),\end{aligned} \tag{39}$$

for $\tau \in \{ g, h \}$. The second x-derivative is

$$\frac{\partial^2 I}{\partial x^2}(x) = -\left(\frac{1}{\frac{\partial T}{\partial y}(I(x))}\right)^2 \cdot \frac{\partial^2 T}{\partial y^2}(I(x)) \cdot \frac{\partial I}{\partial x}(x). \tag{40}$$

The implementation of these expressions requires the computation of $\frac{\partial^2 T}{\partial y^2}$, $\frac{\partial^2 T}{\partial g \partial y}$ and $\frac{\partial^2 T}{\partial h \partial y}$. These derivatives can also be found in Table 8. Just like the first order derivatives, they are continuous in $g = 0$, which can again be seen by an application of L'Hôpital's rule.

Gradient of the g-and-h Density: Using the notation defined in section 2.1, the first derivative of the density of a g-and-h distributed random variable $X_{g,h}$ can be written as

$$\begin{aligned}\frac{\partial f_{X_{g,h}}}{\partial \tau}(x,g,h) = \phi'(I(x,\tau)) \cdot &\frac{\partial I}{\partial \tau}(x,\tau) \cdot \frac{\partial I}{\partial x}(x,\tau) \\ &+ \phi(I(x,\tau)) \cdot \frac{\partial^2 I}{\partial x \partial \tau}(x,\tau).\end{aligned} \tag{41}$$

For $\tau = x$ this yields

$$\frac{\partial f_{X_{g,h}}}{\partial x}(x,g,h) = \phi'(I(x)) \cdot \left(\frac{\partial I}{\partial x}(x)\right)^2 + \phi(I(x)) \cdot \frac{\partial^2 I}{\partial x^2}(x). \tag{42}$$

Next, we derive the gradient of a standardized g-and-h distributed random variable $Z_{g,h}$. From

$$f_{Z_{g,h}}(z,g,h) = \sigma_{g,h} \cdot f_{X_{g,h}}(S(z,g,h),g,h) \tag{43}$$

with $S(z,g,h) := \mu_{g,h} + \sigma_{g,h} \cdot z$, we conclude

$$\begin{aligned}\frac{\partial f_{Z_{g,h}}}{\partial \tau}(z,g,h) = &\frac{\partial \sigma_{g,h}}{\partial \tau} \cdot f_{X_{g,h}}(S(z,g,h),g,h) \\ &+ \sigma_{g,h} \cdot \left[\frac{\partial f_{X_{g,h}}}{\partial \tau}(S(z,g,h),g,h)\right. \\ &\left. + \frac{\partial f_{X_{g,h}}}{\partial x}(S(z,g,h),g,h) \cdot \frac{\partial S(x,g,h)}{\partial \tau}\right],\end{aligned} \tag{44}$$

for $\tau \in \{ g, h \}$ and

$$\frac{\partial f_{Z_{g,h}}}{\partial z}(z,g,h) = \sigma_{g,h} \cdot \frac{\partial f_{X_{g,h}}}{\partial x}(S(z,g,h),g,h) \cdot \frac{\partial S}{\partial z}(z,g,h). \tag{45}$$

For the implementation of these equations we have to compute the derivatives of the mean and variance of a g-and-h distributed random variable. The corresponding results are also provided in Table 8. Checking that these derivatives are continuous is again a L'Hôpital exercise. The combination of these expressions with $\frac{\partial \sigma}{\partial \tau}(\tau) = (2\sigma(\tau))^{-1} \cdot \frac{\partial \sigma^2}{\partial \tau}(\tau)$ completes the derivation of the gradient of the standardized g-and-h density.

Gradient of the log-Likelihood: As in section 3, l_t denotes the log-likelihood of a single observation r_t given $\mathcal{F}_{t-1}$. By setting $z_t := \frac{r_t - \mu}{\sigma_t}$, we can write $l_t := \log f_Z(z_t) - \log \sigma_t$. Collecting the parameters of the white noise density and the variance model in $\boldsymbol{\theta}_Z$ and respectively $\boldsymbol{\theta}_\sigma$, we obtain

$$\frac{\partial l_t}{\partial \boldsymbol{\theta}_Z} = \frac{1}{f_Z(z_t)} \cdot \frac{\partial f_Z}{\partial \boldsymbol{\theta}_Z}(z_t), \tag{57}$$

$$\frac{\partial l_t}{\partial \boldsymbol{\theta}_\sigma} = \frac{1}{f_Z(z_t)} \cdot \frac{\partial f_Z}{\partial z}(z_t) \cdot \frac{\partial z_t}{\partial \sigma_t} \cdot \frac{\partial \sigma_t}{\partial \boldsymbol{\theta}_\sigma} - \frac{1}{\sigma_t} \cdot \frac{\partial \sigma_t}{\partial \boldsymbol{\theta}_\sigma}, \tag{58}$$

$$\frac{\partial l_t}{\partial \mu} = \frac{1}{f_Z(z_t)} \cdot \frac{\partial f_Z}{\partial z}(z_t) \cdot \left[\frac{\partial z_t}{\partial \mu} + \frac{\partial z_t}{\partial \sigma_t} \cdot \frac{\partial \sigma_t}{\partial \mu} \right] - \frac{1}{\sigma_t} \cdot \frac{\partial \sigma_t}{\partial \mu}. \tag{59}$$

Noting that $\frac{\partial z_t}{\partial \sigma_t} = -\frac{r_t - \mu}{\sigma^2}$ and $\frac{\partial z_t}{\partial \mu} = -\frac{1}{\sigma_t}$, it follows that[23]

$$\frac{\partial l_t}{\partial \mu} = -\frac{1}{\sigma_t} \cdot \left[(f_Z(z_t))^{-1} \cdot \frac{\partial f_Z}{\partial z}(z_t) \cdot \left(1 + \frac{r_t - \mu}{2\sigma_t^2} \cdot \frac{\partial \sigma_t^2}{\partial \mu} \right) + \frac{1}{2\sigma_t} \cdot \frac{\partial \sigma_t^2}{\partial \mu} \right], \tag{60}$$

$$\frac{\partial l_t}{\partial \boldsymbol{\theta}_\sigma} = -\frac{1}{2\sigma_t^2} \cdot \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}_\sigma} \cdot \left[(f_Z(z_t))^{-1} \cdot \frac{\partial f_Z}{\partial z}(z_t) \cdot \frac{r_t - \mu}{\sigma_t} + 1 \right]. \tag{61}$$

In the case of the GARCH(1,1) specification, we have

$$\frac{\partial \sigma_t^2}{\partial \alpha_0} = 1 + \beta_1 \cdot \frac{\partial \sigma_{t-1}^2}{\partial \alpha_0}, \tag{62}$$

$$\frac{\partial \sigma_t^2}{\partial \alpha_1} = (r_{t-1} - \mu)^2 + \beta_1 \cdot \frac{\partial \sigma_{t-1}^2}{\partial \alpha_1}, \tag{63}$$

$$\frac{\partial \sigma_t^2}{\partial \beta_1} = \sigma_{t-1}^2 + \beta_1 \cdot \frac{\partial \sigma_{t-1}^2}{\partial \beta_1}, \tag{64}$$

$$\frac{\partial \sigma_t^2}{\partial \mu} = -2 \cdot \alpha_1 \cdot (r_{t-1} - \mu) + \beta_1 \cdot \frac{\partial \sigma_{t-1}^2}{\partial \mu}. \tag{65}$$

These derivatives have to be calculated iteratively starting from $t = 1$. The start values of the iteration depend on the assumptions on the presample values of σ_t^2 and ε_t^2. Following a suggestion in Bollerslev (1986), we use estimates of the unconditional variance. In particular, we set

$$h_0^2 = \varepsilon_0^2 = \frac{1}{T-1} \sum_{t=1}^{T} (r_t - \hat{\mu})^2, \tag{66}$$

[23] We prefer to rewrite σ_t-derivatives in terms of the σ_t^2-derivatives because we model the variance explicitly.

Table 8: Derivatives of $T_{g,h}$, $\mu_{g,h}$ and $\sigma^2_{g,h}$

transformation function:

$$\frac{\partial T}{\partial y} = \begin{cases} \exp\left(\frac{hy^2}{2}\right) \cdot [g\exp(gy) + hy(\exp(gy) - 1)] \cdot g^{-1} & \text{if } g \neq 0, \\ \exp\left(\frac{hy^2}{2}\right) \cdot \left(1 + hy^2\right) & \text{if } g = 0 \end{cases} \tag{46}$$

$$\frac{\partial T}{\partial g} = \begin{cases} \exp\left(\frac{hy^2}{2}\right) \cdot [\exp(gy)gy - (\exp(gy) - 1)] \cdot g^{-2} & \text{if } g \neq 0, \\ \exp\left(\frac{hy^2}{2}\right) \cdot \frac{1}{2}y^2 & \text{if } g = 0 \end{cases} \tag{47}$$

$$\frac{\partial T}{\partial h} = \begin{cases} \exp\left(\frac{hy^2}{2}\right) \cdot \frac{1}{2}y^2 \cdot \frac{\exp(gy)-1}{g} & \text{if } g \neq 0, \\ \exp\left(\frac{hy^2}{2}\right) \cdot \frac{1}{2}y^3 & \text{if } g = 0 \end{cases} \tag{48}$$

$$\frac{\partial^2 T}{\partial y^2} = \begin{cases} hy\frac{\partial T}{\partial y} + g^{-1}\exp\left(\frac{hy^2}{2}\right)\left(g^2\exp(gy) + h\left(\exp(gy) - 1\right) + hyg\exp(gy)\right) & \text{if } g \neq 0, \\ \exp\left(\frac{hy^2}{2}\right) \cdot \left(3hy + h^2y^3\right) & \text{if } g = 0 \end{cases} \tag{49}$$

$$\frac{\partial^2 T}{\partial g \partial y} = \begin{cases} -\frac{1}{g}\frac{\partial T}{\partial y} + g^{-1}\exp\left(\frac{hy^2}{2}\right)\left(\exp(gy) + gy\exp(gy) + hy^2\exp\left(gy\right)\right) & \text{if } g \neq 0, \\ \exp\left(\frac{hy^2}{2}\right) \cdot \left(y + \frac{1}{2}hy^3\right) & \text{if } g = 0 \end{cases} \tag{50}$$

$$\frac{\partial^2 T}{\partial h \partial y} = \begin{cases} \frac{y^2}{2}\frac{\partial T}{\partial y} + g^{-1}\exp\left(\frac{hy^2}{2}\right)\left(y\exp(gy) - y\right) & \text{if } g \neq 0, \\ \exp\left(\frac{hy^2}{2}\right) \cdot \left(\frac{3}{2}y^2 + \frac{1}{2}hy^4\right) & \text{if } g = 0 \end{cases} \tag{51}$$

mean:

$$\frac{\partial \mu(g,h)}{\partial g} = \begin{cases} -\frac{1}{g^2\sqrt{1-h}}\left[\exp\left(\frac{g^2}{2(1-h)}\right) - 1\right] + \frac{1}{g\sqrt{1-h}}\exp\left(\frac{g^2}{2(1-h)}\right)\frac{g}{1-h}, & g \neq 0, \\ \frac{1}{2}(1-h)^{-\frac{3}{2}}, & g = 0 \end{cases} \tag{52}$$

$$\frac{\partial \mu(g,h)}{\partial h} = \begin{cases} \frac{1}{2g} \cdot (1-h)^{-\frac{3}{2}} \cdot \left[\exp\left(\frac{g^2}{2(1-h)}\right) - 1\right] + \exp\left(\frac{g^2}{2(1-h)}\right) \cdot \frac{g}{2} \cdot (1-h)^{-\frac{5}{2}}, & g \neq 0, \\ 0, & g = 0 \end{cases} \tag{53}$$

variance:

$g \neq 0$:

$$\begin{aligned} \frac{\partial \sigma^2}{\partial g} = {} & \frac{-2}{g^3\sqrt{1-2h}}\left[\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1\right] \\ & + \frac{1}{g^2\sqrt{1-2h}}\left[\exp\left(\frac{2g^2}{1-2h}\right)\left(\frac{4g}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right)\frac{2g}{2(1-2h)}\right] \\ & - 2 \cdot \mu(g,h) \cdot \frac{\partial \mu(g,h)}{\partial g} \end{aligned} \tag{54}$$

$$\begin{aligned} \frac{\partial \sigma^2}{\partial h} = {} & \frac{1}{g^2}(1-2h)^{-\frac{3}{2}}\left[\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1\right] \\ & + \frac{1}{g^2\sqrt{1-2h}}\left[\exp\left(\frac{2g^2}{1-2h}\right) \cdot \frac{4g^2}{(1-2h)^2} - 2\exp\left(\frac{g^2}{2(1-2h)}\right) \cdot \frac{g^2}{(1-2h)^2}\right] \\ & - 2 \cdot \mu(g,h) \cdot \frac{\partial \mu(g,h)}{\partial h} \end{aligned} \tag{55}$$

$g = 0$:

$$\frac{\partial \sigma^2_{g,h}}{\partial h} = 3(1-2h)^{-\frac{5}{2}}, \qquad \frac{\partial \sigma^2_{g,h}}{\partial g} = 0 \tag{56}$$

with $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t$[24]. In this case, the prevalues are independent of the parameters and thus our starting values are given by $\frac{\partial \sigma_1^2}{\partial \alpha_0} = 1$, $\frac{\partial \sigma_1^2}{\partial \alpha_1} = (r_0 - \mu)^2$, $\frac{\partial \sigma_1^2}{\partial \beta_1} = \sigma_0^2$ and $\frac{\partial \sigma_1^2}{\partial \mu} = -2 \cdot \alpha_1 \cdot (r_0 - \mu)$.

[24] This estimator slighty deviates from the original suggestion in Bollerslev (1986, p. 316), where $\frac{1}{T}\sum_{t=1}^{T}(r_t - \mu)$ is used to estimate the unconditional variance.

References

Ané, T. and Kharoubi, C. (2003). Dependence structure and risk measure. *The Journal of Business*, 76(3), 411–438.

Badrinath, S. G. and Chatterjee, S. (1988). On measuring skewness and elongation in common stock return distributions: The case of the market index. *The Journal of Business*, 61(4), 451–472.

Bao, Y., Lee, T. and Saltoglu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, 26(3), 203–225.

Bartram, S. M., Taylor, S. and Wang, Y.-H. (2007). The euro and european financial market dependence. *Journal of Banking and Finance*, 31(5), 1461–1481.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics*, 72(3), 498–505.

bwGRiD (2007-2010). *http://www.bw-grid.de/, Member of the German D-Grid initiative, funded by the Ministry of Education and Research and the Ministry for Science, Research and Arts Baden-Wuerttemberg.* Technical report, Universities of Baden-Württemberg.

Chollette, L., Heinen, A. and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime-switching copula. *Journal of Financial Econometrics*, 7(4), 437–480.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.

Degen, M., Embrechts, P. and Lambrigger, D. D. (2007). The quantitative modeling of operational risk: Between g-and-h and EVT. *Astin Bulletin*, 37(2), 265–291.

Dowd, K. (1998). *Beyond value at risk - the new science in risk management.* Chichester: Wiley.

Dutta, K. K. and Babbel, D. F. (2002). On measuring skewness and kurtosis in short rate distributions: The case of the us dollar london inter bank offer rates. *Wharton Financial Institutions Center Working Paper*, 02-25.

Dutta, K. K. and Perry, J. (2006). A tale of tails: An empirical analysis of loss distribution models for estimating operational risk. *Federal Reserve Bank of Boston Working Papers*, 06-12.

Fantazzini, D. (2009). The effects of misspecified marginals and copulas on computing the value at risk: A monte carlo study. *Computational Statistics & Data Analysis*, 53(6), 2168–2188.

Field, C. and Genton, M. (2006). The multivariate g-and-h distribution. *Technometrics*, 48(1), 104–111.

Fischer, M. (2010). Generalized tukey-type distributions with application to financial and teletraffic data. *Stat Papers*, 51(1), 41–56.

Fischer, M., Kock, C., Schluter, S. and Weigert, F. (2009). An empirical analysis of multivariate copula models. *Quantitative Finance*, 9(7), 839–854.

Ghorbel, A. and Trabelsi, A. (2009). Measure of financial risk using conditional extreme value copulas with evt margins. *Journal of Risk*, 11(4), 51–85.

Giot, P. and Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance*, 11(3), 379–398.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705–730.

He, C. and Teräsvirta, T. (1999). Properties of moments of a family of GARCH processes. *Journal of Econometrics*, 92(1), 173–192.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D. C. Hoaglin, F. Mosteller and J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes* chapter 11, (pp. 461–511). New York: Wiley.

Huang, J.-J., Lee, K.-J., Liang, H. and Lin, W.-F. (2009). Estimating value at risk of portfolio by conditional copula-GARCH method. *Insurance: Mathematics and Economics*, 45(3), 315–324.

Jarque, C. M. and Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10), 1699–1737.

Jondeau, E. and Rockinger, M. (2006). The Copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25, 827–853.

Jorion, P. (2001). *Value At Risk : The new benchmark for managing financial risk*. New York: McGraw-Hill, 2nd. ed.

Kuester, K., Mittnik, S. and Paolella, M. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1), 53.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2), 73–84.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.

Martinez, J. and Iglewicz, B. (1984). Some properties of the tukey g and h family of distributions. *Computational Statistics & Data Analysis*, 13(3), 353–369.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4), 271 – 300.

Mills, T. C. (1995). Modelling skewness and kurtosis in the london stock exchange FT-SE index return distributions. *The Statistician*, 44(3), 323–332.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(3), 318–334.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1), 130–168.

Patton, A. J. (2006). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics*, 21(2), 147–173.

Rayner, G. D. and MacGillivray, H. (2002). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing*, 12, 57–75.

Rodriguez, J. (2007). Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, 14(3), 401–423.

Rosenberg, J. and Schuermann, T. (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics*, 79, 569–614.

Sun, W., Rachev, S., Fabozzi, F. J. and Kalev, P. S. (2009). A new approach to modeling co-movement of international equity markets: evidence of unconditional copula-based simulation of tail dependence. *Empirical Economics*, 36(1), 201–229.

Szegö, G. (2002). Measures of risk. *Journal of Banking and Finance*, 26(7), 1253–1272.

Tasche, D. (2002). Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7), 1519–1533.

Taylor, S. J. (1986). *Modelling financial time series*. Chichester: Wiley.