The rmgarch models: Background and properties. (Version 1.3-0)

Alexios Ghalanos

September 12, 2019

Contents

1 Introduction

The ability to dynamically and jointly model the full multivariate density dynamics has very important implications for risk and portfolio management, and more generally economic policy decision making. However, feasible large-scale multivariate GARCH modelling has proved very challenging since the direct extension of the univariate models to a vector representation by Bollerslev et al. (1988). The **rmgarch** package aims to provide a subset of multivariate GARCH models which can handle large scale estimation through separation of the dynamics so that parallel processing may be used. Methods for fitting, filtering, forecasting and simulation are included were applicable with some interesting additional methods aimed at portfolio and risk applications. This document provides for a summarized theoretical background of the models and their properties.
While there are a number of open source and commercial packages implementing the DCC based models, the **rmgarch** package uniquely implements and introduces the GO-GARCH model with ICA using the multivariate affine Generalized Hyperbolic distribution and the relevant methods for working with this model in an applied setting.
The **rmgarch** package is on CRAN and the development version on bitbucket (`https://bitbucket.org/alexiosg`). Some online examples and demos are available on my website (`http://www.unstarched.net`).
The package is provided AS IS, without any implied warranty as to its accuracy or suitability. A lot of time and effort has gone into the development of this package, and it is offered under the GPL-3 license in the spirit of open knowledge sharing and dissemination. If you do use the model in published work DO remember to cite the package and author (type **citation("rmgarch")** for the appropriate BibTeX entry) , and if you have used it and found it useful, drop me a note and let me know.
IMPORTANT :
The package is still in development and some functions/methods MAY change over time, and bugs are certain to exist. Please report any suspected bugs in the code, mistakes in the models or general questions to the R-SIG-FINANCE mailing list and not directly to my email, unless solicited. I maintain a blog (`http://www.unstarched.net`) which contains some examples and posts which I update when time permits.

2 Multivariate GARCH Models

The generalization of univariate GARCH models to the multivariate domain is conceptually simple. Consider the stochastic vector process, $\boldsymbol{x}_t$ $\{t = 1, 2, ...T\}$ of financial returns with dimension $N \times 1$ and mean vector $\boldsymbol{\mu}_t$[1], given the information set $\boldsymbol{I}_{t-1}$:

$$\boldsymbol{x}_t \,|\boldsymbol{I}_{t-1} = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where the residuals of the process are modelled as:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t, \tag{2}$$

and $\boldsymbol{H}_t^{1/2}$ is an $N \times N$ positive definite matrix such that $\boldsymbol{H}_t$ is the conditional covariance matrix of $\boldsymbol{x}_t$[2], and $\boldsymbol{z}_t$ an $N \times 1$ i.i.d. random vector, with centered and scaled first 2 moments:

$$\mathrm{E}\left[\boldsymbol{z}_{\mathrm{t}}\right] = 0,$$

[1]The mean vector may for example be derived from a VAR model or may simply represent the unconditional means of the financial returns.

[2]One way to obtain the square root matrix is through the singular value decomposition of $\boldsymbol{H}_t$.

$$\mathrm{Var}[\boldsymbol{z}_{\mathrm{t}}] = \boldsymbol{I}_{\mathrm{N}}, \tag{3}$$

with $\boldsymbol{I}_N$ denoting the identity matrix of order N. The conditional covariance matrix $\boldsymbol{H}_t$ of $\boldsymbol{x}_t$ may be defined as:

$$\begin{aligned}\mathrm{Var}\left(\boldsymbol{x}_{\mathrm{t}}\,|\boldsymbol{I}_{\mathrm{t}-1}\right) = \mathrm{Var}_{\mathrm{t}-1}(\boldsymbol{x}_{\mathrm{t}}) &= \mathrm{Var}_{\mathrm{t}-1}(\boldsymbol{\varepsilon}_{\mathrm{t}}) \\ &= \boldsymbol{H}_t^{1/2}\mathrm{Var}_{\mathrm{t}-1}(\boldsymbol{z}_{\mathrm{t}})(\boldsymbol{H}_{\mathrm{t}}^{1/2})' \\ &= \boldsymbol{H}_t. \end{aligned} \tag{4}$$

The literature on the different specifications of $\boldsymbol{H}_t$ may be broadly divided into direct multivariate extensions, factor models and the conditional correlation models. The usual trade-off of model parametrization and dimensionality clearly applies here, with the fully parameterized models offering the richest dynamics at the cost of increasing parameter size, making it unfeasible for modelling anything beyond a couple of assets. There is, also, a not so evident tradeoff between those models which allow flexible univariate dynamics (in the motion dynamics and the distributions) to enter the equation at the cost of some multivariate dynamics. The next sections will review these models and some of the tradeoffs they present for the decision maker. A more complete review of multivariate GARCH (*MGARCH*) models is provided by Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009).

2.1 Conditional Mean Dynamics

The **rmgarch** package allows for either a constant, univariate AR or Vector AR (*VAR*) model to be fit (or a pre-filtered residual series). The constant and AR models are already implemented and described in the **rugarch** package. For the DCC based models, the constant-AR model is jointly estimated with the first stage GARCH dynamics, while for the GO-GARCH models the univariate ARFIMAX model is used assuming constant variance to obtain the parameter estimates. In the case of the VAR model, external regressors are also allowed as is the possibility to use a robust version of the model based on the multivariate least trimmed squares approach of Croux and Joossens (2008). When using a constant or AR model with DCC based models, standard errors are calculated for all first stage parameters using a partitioned standard error matrix. In the case of a VAR model, this joint estimation of standard errors is not practical due to the dimensionality of the system. Finally, in the case of the GO-GARCH model, there is no joint estimation of parameters for the first (conditional mean) and second (factor dynamics) stage estimation.[3]

2.2 Dynamic Conditional Correlation Models

Conditional correlation models are founded on a decomposition of the conditional covariance matrix into conditional standard deviations and correlations, so that it may be expressed in such a way that the univariate and multivariate dynamics may be separated, thus easing the estimation process. This decomposition comes at a cost of some dynamic structure as well as severe restriction on the type of multivariate distribution which can usually be decomposed in such a way.[4] Recently, some of these models have been extended to allow for more flexible dynamic structure which unfortunately has led to significant loss in the ease of estimation.
In the constant conditional correlation model (CCC) of Bollerslev (1990), the covariance matrix can be decomposed into

$$\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{R}\boldsymbol{D}_t = \rho_{ij}\sqrt{h_{iit}h_{jjt}}, \tag{5}$$

[3] Additionally, since ICA is a linear noiseless model, there is no uncertainty assumed with regards to the mixing matrix A.

[4] This has implications both for the use of the 2-stage estimation as well as the form of the covariance matrix which may be a complicated function of the scaling matrix for non-elliptical distributions.

where $\boldsymbol{D}_t = diag(\sqrt{h_{11,t}}, ..., \sqrt{h_{nn,t}})$, and $\boldsymbol{R}$ is the positive definite constant conditional correlation matrix. The conditional variances, and $h_{ii,t}$, which can be estimated separately, can be written in vector form based on GARCH(p,q) models[5]

$$h_t = \omega + \sum_{i=1}^{p} \boldsymbol{A}_i \varepsilon_{t-i} \odot \varepsilon_{t-i} + \sum_{i=1}^{q} \boldsymbol{B}_i h_{t-i} \tag{6}$$

where $\omega \in \mathbb{R}^n$, $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are $N \times N$ diagonal matrices, and $\odot$ denotes the Hadamard operator. The conditions for the positivity of the covariance matrix $\boldsymbol{H}_t$ are that $\boldsymbol{R}$ is positive definite, and the elements of ω and the diagonal elements of the matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are positive. In the extended CCC model (E-CCC) of Jeantheau (1998), implemented in the **ccgarch** package, the assumption of diagonal elements on $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ was relaxed, allowing the past squared errors and variances of the series to affect the dynamics of the individual conditional variances, and hence providing for a much richer structure, albeit at the cost of a lot more parameters. The decomposition in (5), allows the log-likelihood at each point in time (LL_t), in the multivariate normal case, to be expressed as

$$\begin{aligned} LL_t &= \frac{1}{2}\left(\log(2\pi) + \log|\boldsymbol{H}_t| + \boldsymbol{\varepsilon}'_t \boldsymbol{H}_t^{-1} \boldsymbol{\varepsilon}_t\right) \\ &= \frac{1}{2}\left(\log(2\pi) + \log|\boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t| + \boldsymbol{\varepsilon}'_t \boldsymbol{D}_t^{-1} \boldsymbol{R}^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{\varepsilon}_t\right) \\ &= \frac{1}{2}\left(\log(2\pi) + 2\log|\boldsymbol{D}_t| + \log|\boldsymbol{R}| + \boldsymbol{z}'_t \boldsymbol{R}^{-1} \boldsymbol{z}'_t\right) \end{aligned} \tag{7}$$

where $\boldsymbol{z}_t = \boldsymbol{D}_t^{-1} \boldsymbol{\varepsilon}_t$. This can be described as a term ($\boldsymbol{D}_t$) for the sum of the univariate GARCH model likelihoods, a term for the correlation ($\boldsymbol{R}$) and a term for the covariance which arises from the decomposition.
Because the restriction of constant conditional correlation may be unrealistic in practice, a class of models termed Dynamic Conditional Correlation (DCC) due to Engle (2002) and Tse and Tsui (2002) where introduced which allow for the correlation matrix to be time varying with motion dynamics, such that

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t. \tag{8}$$

In these models, apart from the fact that the time varying correlation matrix, $\boldsymbol{R}_t$, must be inverted at every point in time (making the calculation that much slower), it is also important to constrain it to be positive definite. The most popular of these DCC models, due to Engle (2002), achieves this constraint by modelling a proxy process, $\boldsymbol{Q}_t$ as:

$$\begin{aligned} \boldsymbol{Q}_t &= \bar{\boldsymbol{Q}} + a\left(\boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1} - \bar{\boldsymbol{Q}}\right) + b\left(\boldsymbol{Q}_{t-1} - \bar{\boldsymbol{Q}}\right) \\ &= (1 - a - b)\bar{\boldsymbol{Q}} + a\boldsymbol{z}_{t-1}\boldsymbol{z}'_{t-1} + b\boldsymbol{Q}_{t-1} \end{aligned} \tag{9}$$

where a and b are non negative scalars, with the condition that $a + b < 1$ imposed to ensure stationarity and positive definiteness of $\boldsymbol{Q}_t$. $\bar{Q}$ is the unconditional matrix of the standardized errors $\boldsymbol{z}_t$ which enters the equation via the covariance targeting part $(1 - a - b)\bar{\boldsymbol{Q}}$, and $\boldsymbol{Q}_0$ is positive definite. The correlation matrix $\boldsymbol{R}$ is then obtained by rescaling $\boldsymbol{Q}_t$ such that,

$$\boldsymbol{R}_t = diag(\boldsymbol{Q}_t)^{-1/2} \boldsymbol{Q}_t diag(\boldsymbol{Q}_t)^{-1/2}. \tag{10}$$

[5]The GARCH models are not restricted to be of one particular 'flavor', allowing to mix different GARCH models in the univariate stage.

The log-likelihood function in equation (6) can be decomposed more clearly into a volatility and correlation component by adding and subtracting $\boldsymbol{\varepsilon'}_t \boldsymbol{D}_t^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{\varepsilon}_t = \boldsymbol{z'}_t \boldsymbol{z}_t$,

$$\begin{aligned} LL &= \frac{1}{2} \sum_{i=1}^{T} \left(N \log(2\pi) + 2\log|\boldsymbol{D}_t| + \log|\boldsymbol{R}_t| + \boldsymbol{z'}_t \boldsymbol{R}_t^{-1} \boldsymbol{z'}_t \right) \\ &= \frac{1}{2} \sum_{i=1}^{T} \left(N \log(2\pi) + 2\log|\boldsymbol{D}_t| + \boldsymbol{\varepsilon'}_t \boldsymbol{D}_t^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{\varepsilon}_t \right) - \frac{1}{2} \sum_{i=1}^{T} \left(\boldsymbol{z'}_t \boldsymbol{z}_t + \log|\boldsymbol{R}_t| + \boldsymbol{z'}_t \boldsymbol{R}_t^{-1} \boldsymbol{z'}_t \right) \\ &= LL_V(\theta_1) + LL_R(\theta_1, \theta_2) \end{aligned} \tag{11}$$

where $LL_V(\theta_1)$ is the volatility component with parameters θ_1, and $LL_R(\theta_1, \theta_2)$ the correlation component with parameters θ_1 and θ_2. In the Multivariate Normal case, where no shape or skew parameters enter the density, the volatility component is the sum of the individual GARCH likelihoods which can be jointly maximized by separately maximizing each univariate model. In other distributions, such as the multivariate Student, the existence of a shape parameter means that the estimation must be performed in one step so that the shape parameter is jointly estimated for all models. Separation of the likelihood into 2 parts provides for feasible large scale estimation. Together with the use of variance targeting, very large scale systems may be estimated in a matter of seconds with the use of parallel and grid computing. Yet as the system becomes larger and larger, it becomes questionable whether the scalar parameters can adequately capture the dynamics of the underlying process. As such, Cappiello et al. (2006) generalize the DCC model with the introduction of the Asymmetric Generalized DCC ($AGDCC$) where the dynamics of $\boldsymbol{Q}_t$ are:

$$\boldsymbol{Q}_t = \left(\bar{\boldsymbol{Q}} - \boldsymbol{A'} \bar{Q} \boldsymbol{A} - \boldsymbol{B'} \bar{\boldsymbol{Q}} \boldsymbol{B} - \boldsymbol{G'} \bar{\boldsymbol{Q}}^{-} \boldsymbol{G} \right) + \boldsymbol{A'} \boldsymbol{z}_{t-1} \boldsymbol{z'}_{t-1} \boldsymbol{A} + \boldsymbol{B'} \boldsymbol{Q}_{t-1} \boldsymbol{B} + \boldsymbol{G'} \boldsymbol{z}_t^{-} \boldsymbol{z'}_t^{-} \boldsymbol{G} \tag{12}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{G}$ are the $N \times N$ parameter matrices, $\boldsymbol{z}_t^{-}$ are the zero-threshold standardized errors which are equal to $\boldsymbol{z}_t$ when less than zero else zero otherwise, $\bar{\boldsymbol{Q}}$ and $\bar{\boldsymbol{Q}}^{-}$ the unconditional matrices of $\boldsymbol{z}_t$ and $\boldsymbol{z}_t^{-}$ respectively. Because of its high dimensionality, restricted models have been used including the scalar, diagonal and symmetric versions with the specifications nested being

- DCC : $\boldsymbol{G} = [0], \boldsymbol{A} = \sqrt{a}, \boldsymbol{B} = \sqrt{b}$
- ADCC : $\boldsymbol{G} = \sqrt{g}, \boldsymbol{A} = \sqrt{a}, \boldsymbol{B} = \sqrt{b}$
- GDCC : $\boldsymbol{G} = [0]$.

Variance targeting in such high dimensional models where the parameters are no longer scalars, creates difficulties in imposing positive definiteness during estimation while at the same time guaranteeing a global optimum solution. Methods which directly check and penalize the eigenvalues of the intercept matrix introduce non-smoothness and discontinuities into the likelihood surface for which inference is likely to be difficult.[6] More substantially, Aielli (2009) points out that the estimation of $\bar{\boldsymbol{Q}}_t$ as the empirical counterpart of the correlation matrix of $\boldsymbol{z}_t$ in the DCC model is inconsistent since $\mathrm{E}[\boldsymbol{z}_{\mathrm{t}} \boldsymbol{z}_t] = \mathrm{E}[\boldsymbol{R}_{\mathrm{t}}] \neq \boldsymbol{R}[\bar{\boldsymbol{Q}}_{\mathrm{t}}]$. He proposes instead the cDCC model which includes a corrective step which eliminates this inconsistency, albeit at the cost of targeting which is not allowed.

One model which tries to balance dimensionality with more realistic dynamics is the Flexible

[6]This has not prevented a plethora of paper using these models and making inference based on questionable convergence criteria.

DCC (FDCC) model of Billio et al. (2006) which allows groups [7] of securities to have the same dynamics. The model may parsimoniously be represented as:

$$\boldsymbol{Q}_t = cc' + \sum_{j=1}^{P} (I_g a_j)(I_g a_j)' \odot \varepsilon_{t-j}\varepsilon'_{t-j} + \sum_{j=1}^{Q} (I_g b_j)(I_g b_j)' \odot \boldsymbol{Q}_{t-j} \tag{13}$$

where I_g is the $assets \times groups$ logical matrix of group exclusive membership. This is a very flexible representation allowing a large range of representations, from a single group driving all dynamics (like the DCC), to each asset having its own group (like the GDCC). Unfortunately, without specialized restrictions correlation targeting is lost, but the model still remains feasible for a not too large number of groups. In the **rmgarch** package, the intercept is estimated *using* correlation targeting with the intercept set to $(\mathbf{11'} - aa' - bb') \odot \bar{\boldsymbol{Q}}$ and the restriction that $a_i a_j + b_i b_j < 1, \forall i, j$ in order to avoid explosive patterns. Positive definiteness of the matrices is achieved by construction subject to a suitable starting point for $\boldsymbol{Q}_t$. Also note that only the FDCC(1,1) model is allowed (i.e. P=1, Q=1) because of the large number of pairwise constraints needed which make higher order models prohibitively expensive to calculate (and in any case it is quite rare to use anything beyond this for DCC type models).
In the **rmgarch** package, the DCC, aDCC and FDCC models are implemented using the 2-stage approach, with a choice of 3 distributions, the multivariate Normal (MVN), Student (MVT) and Laplace (MVL). For the MVT distribution, it is understood that this is based on known shape parameter (which may be fixed for the first and second stage estimation using the fixed.pars method on the specification object), else that the first stage estimation is QML based as in Bauwens and Laurent (2005).

2.2.1 Forecasting

Because of the nonlinearity of the DCC evolution process, the multi-step ahead forecast of the correlation cannot be directly solved, and is instead based on the approximation suggested in Engle and Sheppard (2001). Consider the multi-step ahead evolution of the proxy process Q_{t+n}:

$$Q_{t+n} = (1 - \alpha - \beta)\bar{Q} + \alpha E_t \left[z_{t+n-1} {z'}_{t+n-1} \right] + \beta Q_{t+n-1} \tag{14}$$

where $E_t \left[z_{t+n-1} {z'}_{t+n-1} \right] = R_{t+n-1}$ and $R_{t+n} = diag(Q_{t+n})^{-1/2} Q_{t+n} diag(Q_{t+n})^{-1/2}$. Engle and Sheppard (2001) suggest 2 types of approximations possible to solve for R_{t+n}, and the package adopts the one which, based on their findings, provides for the least bias. That is, set $\bar{Q} \approx R$ and $E_t[Q_{t+1}] = E_t[R_{t+1}]$, so that:

$$E_t[R_{t+n}] = \sum_{i=0}^{n-2} (1 - \alpha - \beta)\bar{R}(\alpha + \beta)^i + (\alpha + \beta)^{n-1} R_{n+1} \tag{15}$$

Importantly, for the rolling 1-ahead method, the estimate of $\bar{Q}$ at time T+n is updated from data up to time T+n-1, which will lead to small differences in results from applying the DCC filter method on new data (the difference will grow with n) which uses a fixed value for $\bar{Q}$ (and which can be controlled by the n.old option).

2.3 The GARCH-Copula Model

Copula functions were introduced by Sklar (1959) as a tool to connect disparate marginal distribution together to form a joint multivariate distribution. They were extensively used in survival

[7] Groups here is used liberally, and meant to denote membership in a set with common dynamics.

analysis and the actuarial sciences for many years before being introduced in the finance literature more than a decade ago by Frey and McNeil (2000) and Li (2000). They have since been very popular in investigating the dependence of financial time series of various assets classes and frequencies. Breymann et al. (2003) investigate bivariate hourly FX spot returns finding that the Student Copula best fit the data at all horizons (with the shape parameter increasing with the time horizon), while Malevergne and Sornette (2003) find that the Normal Copula fits pairs of currencies and equities well on the whole but unsurprisingly fails to capture tail events where the Student Copula does best.[8] Junker and May (2005) use a Frank copula with a transformation generator and GARCH dynamics for the margins using the empirical distribution, to analyze the bivariate dependency of the daily returns of 6 stocks and 3 Euro swap rates (with horizons 2,5, and 10 Years). The comparison with a range of popular copulas including the Normal and Student, in a risk exercise shows that asymmetric tail dependency is important and usually not accommodated by the Student distribution[9] While most studies are predominantly focused on bivariate copulas, the extension to n-variate models is not overtly challenging particularly for elliptical distributions, or the use of the more recent Vine pair copulas (see for example Joe et al. (2010)).

2.3.1 Copulas

An n-dimensional copula $C\left(u_1,\ldots,u_n\right)$ is an n-dimensional distribution in the unit hypercube $[0,1]^n$ with uniform margins. Sklar (1959) showed that every joint distribution F of the random vector $\boldsymbol{X}=\left(x_1,\ldots,x_n\right)$ with margins $F_1\left(x_1\right),\ldots,F_n\left(x_n\right)$, can be represented as:

$$F\left(x_1,\ldots,x_n\right)=C\left(F_1\left(x_1\right),\ldots,F_n\left(x_n\right)\right) \tag{16}$$

for a copula C, which is uniquely determined in $[0,1]^n$ for distributions F under absolutely continuous margins and obtained as:

$$C\left(u_1,\ldots,u_n\right)=F\left(F_1^{-1}\left(u_1\right),\ldots,F_n^{-1}\left(u_n\right)\right) \tag{17}$$

The density function may conversely be obtained as :

$$f\left(x_1,\ldots,x_n\right)=c\left(F_1\left(x_1\right),\ldots,F_n\left(x_n\right)\right)\prod_{i=1}^{n}f_i\left(x_i\right) \tag{18}$$

where f_i are the marginal densities and c is the density function of the copula given by:

$$c\left(u_1,\ldots,u_n\right)=\frac{f\left(F_1^{-1}\left(u_1\right),\ldots,F_n^{-1}\left(u_n\right)\right)}{\prod_{i=1}^{n}f_i\left(F_i^{-1}\left(u_i\right)\right)}. \tag{19}$$

with F_i^{-1} being the quantile function of the margins. A key property of copulas is their invariance under strictly increasing transformations of the components of the $\boldsymbol{X}$, so that for example the copula of the multivariate Normal distribution $F_n\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ is the same as that of $F_n\left(0,\boldsymbol{R}\right)$ where $\boldsymbol{R}$ is the correlation matrix implied by the covariance matrix, and the same for the copula of the multivariate Student distribution reviewed in detail in Demarta and McNeil (2005). The density

[8]Interestingly the authors argue that since such events are rare, the goodness of fit test they use cannot always reject the Normal Copula.

[9]An alternative would be to use the skew Generalized Hyperbolic Student distribution analyzed in Aas and Haff (2006) which allows for the modelling of one heavy (with polynomial behavior) and one semi-heavy (with exponential behavior) tail.

of the Normal Copula, of the n-dimensional random vector $\boldsymbol{X}$ in terms of the correlation matrix $\boldsymbol{R}$, is then:

$$c\,(u;\boldsymbol{R}) = \frac{1}{|\boldsymbol{R}|^{1/2}} e^{-\frac{1}{2}u'(\boldsymbol{R}'-\boldsymbol{I})u} \tag{20}$$

where $\boldsymbol{u}_i = \boldsymbol{\Phi}^{-1}\left(F_i\left(bfx_i\right)\right)$ for $i = 1,\ldots,n$, representing the quantile of the Probability Integral Transformed (PIT) values of $\boldsymbol{X}$, and $\boldsymbol{I}$ is the identity matrix. Because the Normal Copula cannot account for tail dependence, the Student Copula has been more widely used for modelling of financial assets. The density of the Student Copula, of the n-dimensional random vector $\boldsymbol{X}$ in terms of the correlation matrix $\boldsymbol{R}$ and shape parameter ν, can be written as:

$$c\,(\boldsymbol{u};\boldsymbol{R},\nu) = \frac{\boldsymbol{\Gamma}\left(\frac{\nu+n}{2}\right)\left(\boldsymbol{\Gamma}\left(\frac{\nu}{2}\right)\right)^n \left(1+\nu^{-1}\boldsymbol{u}'\boldsymbol{R}^{-1}\boldsymbol{u}\right)^{-(\nu+n)/2}}{|\boldsymbol{R}|^{1/2}\left(\boldsymbol{\Gamma}\left(\frac{\nu+n}{2}\right)\right)^n \boldsymbol{\Gamma}\left(\frac{\nu}{2}\right)\prod_{i=1}^{n}\left(1+\frac{\boldsymbol{u}_i^2}{\nu}\right)^{-(\nu+1)/2}} \tag{21}$$

where $\boldsymbol{u}_i = t_\nu{}^{-1}\left(F\left(x_i;\nu\right)\right)$, where t_ν^{-1} is the quantile function of the student distribution with shape parameter ν.

2.3.2 Correlation and Kendall's $\boldsymbol{\tau}$

Pearson's product moment correlation $\boldsymbol{R}$ totally characterizes the dependence structure in the multivariate Normal case, where zero correlation also implies independence, but can only characterize the ellipses of equal density when the distribution belongs to the elliptical class. In the latter case for instance, with a distribution such as the multivariate Student, the correlation cannot capture tail dependence determined by the shape parameter. Furthermore, it is not invariant under monotone transformations of original variables making it inadequate in many cases. An alternative measure which does not suffer from this is Kendall's $\boldsymbol{\tau}$ (see Kruskal (1958)) based on rank correlations which makes no assumption about the marginal distributions but depends only on the copula C. It is a pairwise measure of concordance calculated as:

$$\boldsymbol{\tau}\left(x_i, x_j\right) = 4\int_0^1\int_0^1 C\left(u_i,u_j\right)dC\left(u_i,u_j\right) - 1. \tag{22}$$

For elliptical distributions, Lindskog et al. (2003) proved that there is a one-to-one relationship between this measure and Pearson's correlation coefficient ρ given by:

$$\boldsymbol{\tau}\left(x_i, x_j\right) = \left(1 - \sum_{x\in\mathbb{R}}\left(\mathbb{P}\{X_i = x\}^2\right)\right)\frac{2}{\pi}\arcsin\rho_{ij} \tag{23}$$

which under certain assumptions (such as in the case of the multivariate Normal) simplifies to $\frac{2}{\pi}\arcsin\rho_{ij}$.[10] Kendall's $\boldsymbol{\tau}$ is also invariant under monotone transformations making it rather more suitable when working with non-elliptical distributions. A useful application arises in the case of the multivariate Student Distribution, where a maximum likelihood approach for the estimation of the Correlation matrix $\boldsymbol{R}$ becomes unfeasible for large dimensions. In this case, an alternative approach is to estimate the sample counterpart of Kendall's $\boldsymbol{\tau}$[11] from the transformed margins and then translate that into the correlation matrix as detailed in (23), providing for a method of moments type estimator.[12] The shape parameter ν may then be estimated keeping

[10] Another popular measure is Spearman's correlation coefficient ρ_s which under Normality equates to $\frac{6}{\pi}\arcsin\frac{\rho_{ij}}{2}$, and it is usually very close in result to Kendall's meaure.

[11] The matrix is build up from the pairwise estimates.

[12] It may be the case that the resultant matrix is not positive definite, in which case a variety of methods exist to tweak it into one.

the correlation matrix constant, with little loss in efficiency vis-a-vis the full maximum likelihood method.[13]

2.3.3 Transformations and Consistency

The estimation and PIT transformation of the margins provides for a great deal of flexibility, with the possibility of adopting a parametric, semi-parametric or empirical approach. The first method, whereby the margins and transformation are performed using a parametric density, was termed the Inference-Functions-for-Margins (*IFM*) by Joe (1997) who also established the asymptotic theory for it. The semi-parametric method (*SPD*) uses a distribution which couples together generalized Pareto distribution (*GPD*) fitted tails[14] with a kernel based interior and described in Davison and Smith (1990), and offers a rather flexible method for capturing fat tails observed in practise. Finally, the empirical approach, also called pseudo-likelihood, was investigated by Genest et al. (1995) and asymptotic properties established under the assumption that the sequence of $\boldsymbol{X}$ is i.i.d.(see Durrleman et al. (2000) for an excellent summary of the different methods and their properties.)
In the **rmgarch** package, all 3 choices of transformations are available with the SPD method using the **spd** package of Ghalanos (2012).

2.3.4 The Student Copula AGDCC

The extension of the static copula approach to dynamic models, and in particular GARCH, was investigated by Patton (2006) who extended and proved the validity of Sklarś theorem for the conditional case. Jondeau and Rockinger (2006) combine the ACD model of Hansen (1994) with skewed Student distribution to model time-varying or regime switching Student Copula for the dependence between pairs of countries, while Chollete et al. (2009) use a GARCH with skewed Student distribution in the first stage and a regime switching model with a Canonical vine copula for the high dependence regime and a Normal copula for the low dependence regime. The use of the skewed Student distribution in such models, beyond its tractability and desirable features, according to Chollete et al. (2009) is so as to ensure that the asymmetry in the dependence structure is purely the result of multivariate asymmetry and not an artifact of poor modelling of the margins. Demarta and McNeil (2005) describe a skewed Student copula derived from the Normal Mean Variance Mixture distribution (described in the next section), with margins univariate skewed student distributions with common shape (ν) but separate skewness (γ) parameters.[15]
In an elliptical distribution setting, adding dynamics to the correlation matrix of the copula seems a natural extension of the 2-stage DCC model, and allows the estimation of a Student copula with disparate shape parameters for the first stage, where this was not possible using the standard DCC model (unless estimated jointly). Let the n-dimensional random vector of asset returns $\boldsymbol{r}_t = r_{it}, \ldots, r_{nt}$ follow a copula GARCH model with joint distribution given by:

$$F\left(\boldsymbol{r}_t|\boldsymbol{\mu}_t, \boldsymbol{h}_t\right) = C\left(F_1\left(r_{1t}|\mu_{1t}, h_{1t}\right), \ldots, F_n\left(r_{nt}|\mu_{nt}, h_{nt}\right)\right) \tag{24}$$

where F_i, $i = 1, \ldots, n$ is the conditional distribution of the i^{th} marginal series density, C is the n-dimensional Copula. The conditional mean $\mathrm{E}\left[\mathrm{r_{it}}\,|\Im_{\mathrm{t-1}}\right] = \mu_{\mathrm{it}}$, where $\Im_{t-1}$ is the σ-field

[13] According to at least one study of Zeevi and Mashal (2002).
[14] For which a Probability Weighted Moment approach exists which is quite robust.
[15] The reason for the common shape parameter is that the mixing variable $\boldsymbol{W}$ in the Normal Mean Variance mixture is Inverse Gaussian distribution, $W \sim Ig\left(\nu/2, \nu/2\right)$. A grouped type copula whereby the shape parameter is also allowed to vary is also described by Demarta and McNeil (2005), in which case each variable has a different value for the mixing variable $\boldsymbol{W}$, so that $\boldsymbol{W}_j \sim Ig\left(\nu_j/2, \nu_j/2\right)$, for $j = 1, \ldots, n$, and the $\boldsymbol{W}_j$ are now perfectly correlated.

generated by the past realization of r_t, and the conditional variance h_{it} follows a GARCH(1,1) process[16]:

$$r_{it} = \mu_{it} + \varepsilon_{it}, \varepsilon_{it} = \sqrt{h_{it}} z_{it}, \tag{25}$$

$$h_{it} = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{it-1} \tag{26}$$

where z_{it} are i.i.d. random variables which conditionally follow a standardized skew Student distribution, $z_{it} \sim f_i(0, 1, \xi_i, \nu_i)$, of Fernandez and Steel (1998) with skew and shape parameters ξ and ν respectively and derived in the **rugarch** vignette.[17] The dependence structure of the margins is then assumed to follow a Student copula with conditional correlation $\boldsymbol{R}_t$ and constant shape parameter η. The conditional density at time t is given by:

$$c_t \left(u_{it}, \ldots, u_{nt} \left| \boldsymbol{R}_t, \eta \right. \right) = \frac{f_t \left(F_i^{-1} \left(u_{it} \left| \eta \right. \right), \ldots, F_i^{-1} \left(u_{nt} \left| \eta \right. \right) \left| \boldsymbol{R}_t, \eta \right. \right)}{\prod\limits_{i=1}^{n} f_i \left(F_i^{-1} \left(u_{it} \left| \eta \right. \right) \left| \eta \right. \right)} \tag{27}$$

where $u_{it} = F_{it} \left(r_i t \left| \mu_{it}, h_{it}, \xi_i, \nu_i \right. \right)$ is the PIT transformation of each series by its conditional distribution F_{it} estimated via the first stage GARCH process, $F_i^{-1} \left(u_{it} \left| \eta \right. \right)$ represents the quantile transformation of the uniform margins subject to the common shape parameter of the multivariate density, $f_t \left(. \left| \boldsymbol{R}_t, \eta \right. \right)$ is the multivariate density of the Student distribution with conditional correlation R_t and shape parameter η and $f_i \left(. \left| \eta \right. \right)$ is the univariate margins of the multivariate Student distribution with common shape parameter η. The dynamics of R_t are assumed to follow an AGDCC model as described in the previous section, though it is more common to use a restricted scalar DCC model for not too large a number of series. Finally, the joint density of the 2-stage estimation is written as:

$$f \left(\boldsymbol{r}_t \left| \boldsymbol{\mu}_t, \boldsymbol{h}_t, \boldsymbol{R}_t, \eta \right. \right) = c_t \left(u_{it}, \ldots, u_{nt} \left| \boldsymbol{R}_t, \eta \right. \right) \prod_{i=1}^{n} \frac{1}{\sqrt{h_{it}}} f_{it} \left(z_{it} \left| \nu_i, \xi_i \right. \right) \tag{28}$$

where it is clear that the likelihood is composed of a part due to the joint DCC copula dynamics and a part due to the first stage univariate GARCH dynamics.
A similar model, with Student margins, was estimated by Ausin and Lopes (2010) using a Bayesian setup, and an empirical risk management application, albeit once again using only a bivariate series (DAX and Dow Jones indices), used to illustrate its applicability and appropriateness.
In the **rmgarch** package, the Normal and Student copulas are implemented, with either a static or dynamic correlation model (aDCC).

2.3.5 Forecasting

Because of the nonlinear transformation of the margins, there is no closed form solution for the multi-step ahead forecast. As such, the *cgarchsim* method must be used. The inst folder of the package contains a number of examples.

2.4 The GO-GARCH Model

Factor ARCH models, originally introduced by Engle et al. (1990) and with foundations in the Arbitrage Pricing Theory of Ross (1976), are based on the assumption that returns are generated

[16]For simplicity of exposition, a simple GARCH model is chosen, but in fact any combination of GARCH models may be used.

[17]Any combination of conditional distributions can be used in the first stage. The skew-student is used here for illustration.

by a set of unobserved underlying factors that are conditionally heteroscedastic. The dependence framework is non-dynamic as a consequence of large scale estimation in a multivariate setting. The dependence structure of the unobserved factors then determines the type of factor model it belongs to, with correlated factors making up the F-ARCH type models, while uncorrelated and independent factors comprise the Orthogonal and Generalized Orthogonal Models respectively.[18] Because one can always re-discover uncorrelated or independent sources by certain statistical transformation, the correlated factor assumption of F-ARCH models does appear to be restrictive. GO-GARCH models on the other hand make use of those transformations to place the factors in an independence framework with unique benefits such as separability and weighted density convolution giving rise to truly large scale, real-time and feasible estimation. Consider a set of N assets whose returns $\boldsymbol{r}_t$ are observed for T periods, with conditional mean $E[\boldsymbol{r}_t|\mathfrak{F}_{t-1}] = \boldsymbol{m}_t$, where $\mathfrak{F}_{t-1}$ is the σ-field generated by the past realizations of $\boldsymbol{r}_t$, i.e. $\mathfrak{F}_{t-1} = \sigma(\boldsymbol{r}_{t-1}, \boldsymbol{r}_{t-2}, \ldots)$. The GO-GARCH Model of van der Weide (2002) maps $\boldsymbol{r}_t - \boldsymbol{m}_t$ onto a set of unobserved independent factors $\boldsymbol{f}_t$ (or "structural errors"),

$$\begin{aligned} \boldsymbol{r}_t &= \boldsymbol{m}_t + \boldsymbol{\epsilon}_t \quad t = 1, \ldots, T & (30) \\ \boldsymbol{\epsilon}_t &= \boldsymbol{A}\boldsymbol{f}_t, & (31) \end{aligned}$$

where $\boldsymbol{A}$ is invertible and constant over time and may be decomposed into the de-whitening matrix $\boldsymbol{\Sigma}^{1/2}$, representing the square root of the unconditional covariance, and an orthogonal matrix, $\boldsymbol{U}$, so that:

$$\boldsymbol{A} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{U}, \tag{32}$$

and $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{Nt})'$. The rows of the mixing matrix $\boldsymbol{A}$ therefore represent the independent source factor weights assigned to each asset (i.e. rows are the assets and columns the factors). The factors have the following specification:

$$\boldsymbol{f}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t, \tag{33}$$

where $\boldsymbol{H}_t = E[\boldsymbol{f}_t\boldsymbol{f}_t'|\mathfrak{F}_{t-1}]$ is a diagonal matrix with elements $(h_{1t}, \ldots, h_{Nt})$ which are the conditional variances of the factors, and $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{Nt})'$. The random variable z_{it} is independent of z_{jt-s} $\forall j \neq i$ and $\forall s$, with $E[z_{it}|\mathfrak{F}_{t-1}] = 0$ and $E[z_{it}^2] = 1$, this implies that $E[\boldsymbol{f}_t|\mathfrak{F}_{t-1}] = \boldsymbol{0}$ and $E[\boldsymbol{\epsilon}_t|\mathfrak{F}_{t-1}] = \boldsymbol{0}$. The factor conditional variances, $h_{i,t}$, can be modelled as a GARCH-type process. The unconditional distribution of the factors is characterized by:

$$E[\boldsymbol{f}_t] = \boldsymbol{0} \quad E[\boldsymbol{f}_t\boldsymbol{f}_t'] = \boldsymbol{I}_N \tag{34}$$

which, in turn, implies that:

$$E[\boldsymbol{\epsilon}_t] = \boldsymbol{0} \quad E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \boldsymbol{A}\boldsymbol{A}'. \tag{35}$$

It follows that the returns can be expressed as:

$$\boldsymbol{r}_t = \boldsymbol{m}_t + \boldsymbol{A}\boldsymbol{H}_t^{1/2}\boldsymbol{z}_t. \tag{36}$$

[18]It should be noted, that most of these factor models may be seen as special cases of the BEKK model. The GO-GARCH model has the following restricted BEKK representation:

$$H_t = C + \sum_{i=1}^{m} A_i x_{t-1} {x'}_{t-1} {A'}_i + BH_{t-1}B'. \tag{29}$$

Under the assumption that all A_i and B are restricted to have the same eigenvector Z, with the eigenvalues of A being all zero except the i^{th} one, and the C can be decomposed into ZDZ' where D is some positive definite diagonal matrix, then this is a GO-GARCH (with GARCH(1,1) univariate dynamics) model where Z is the linear ICA map. However, GO-GARCH model is not limited to GARCH(1,1) or any particular process for the factors.

The conditional covariance matrix, $\boldsymbol{\Sigma}_t \equiv E[(\boldsymbol{r}_t - \boldsymbol{m}_t)(\boldsymbol{r}_t - \boldsymbol{m}_t)'|\mathfrak{F}_{t-1}]$ of the returns is given by:

$$\boldsymbol{\Sigma}_t = \boldsymbol{A}\boldsymbol{H}_t\boldsymbol{A}'. \tag{37}$$

The Orthogonal Factor model of Alexander (2001)[19] which uses only information in the covariance matrix, leads to uncorrelated components but not necessarily independent unless assuming a multivariate normal distribution. However, while whitening is not sufficient for independence, it is nevertheless an important step in the preprocessing of the data in the search for independent factors, since by exhausting the second order information contained in the covariance matrix it makes it easier to infer higher order information, reducing the problem to one of rotation (orthogonalization). The original procedure of van der Weide (2002) used a 1-step maximum likelihood approach to jointly estimate the rotation matrix and dynamics making the procedure infeasible for anything other than a few assets. Alternative approaches such as nonlinear least squares and method of moments for the estimation of $\boldsymbol{U}$ have been proposed in van der Weide (2004) and Boswijk and van der Weide (2011), respectively. In the **rmgarch** package, I estimate the matrix $\boldsymbol{U}$ by ICA as in Broda and Paolella (2009) and Zhang and Chan (2009). One of the computational advantages offered by the Generalized Orthogonal approach is that following the estimation of the independent factors, the dynamics of the marginal density parameters of those factors may be estimated separately.

2.4.1 ICA

The estimation of the factor loading matrix $\boldsymbol{A}$ exploits the decomposition in (32). The estimation of $\boldsymbol{\Sigma}^{1/2}$, representing the square root of the unconditional covariance matrix, is usually obtained from the OLS residuals $\widehat{\boldsymbol{\epsilon}}_t = \boldsymbol{r}_t - \widehat{\boldsymbol{m}}_t$, while the orthogonal matrix $\boldsymbol{U}$ can be estimated using ICA (see Broda and Paolella (2009), Zhang and Chan (2009)). ICA is a computational method for separating multivariate mixed signals, $\boldsymbol{x} = [x_1, ..., x_n]'$, into additive statistically independent and non-Gaussian components, $\boldsymbol{s} = [s_1, ..., s_n]'$, such that $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{s}$. The objective is to decompose the observed $\boldsymbol{x} = [x_1, ..., x_n]'$, into independent factors $\boldsymbol{s} = [s_1, ..., s_n]'$ and a linear matrix $\boldsymbol{B}$, such that $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{s}$. The independent source vector $\boldsymbol{s} \in \mathbb{R}^n$, is assumed to be sampled from a joint distribution $f(\boldsymbol{s})$,

$$f(s_1, ..., s_n) = f(s_1)f(s_2)...f(s_n), \tag{38}$$

where $\boldsymbol{s}$ is not directly observable, nor is the particular form of the individual distributions, $f(s_i)$, usually known.[20] This forms the key property of independence, namely that the joint density of independent signals is simply the product of their margins. The estimate of the linear mixing matrix $\boldsymbol{B}$ can be obtained via estimation methods based on a choice of criteria for measuring independence which include the maximization of non-Gaussianity through measures such as kurtosis and negentropy, minimization of mutual information, likelihood and infomax. This follows from the Central Limit Theorem which states that mixtures of independent variables tend to become more Gaussian in distribution when they are mixed linearly, hence maximizing non-Gaussianity leads to independent components (see Hyvärinen and Oja (2000) for more details).[21] Entropy may be thought of as the amount of information inherent within a random variable, being an increasing function of the amount of randomness in that variable. For a

[19] When $\boldsymbol{U}$ is restricted to be an identity matrix, the model reduces to the Orthogonal Factor model.

[20] If the distributions are known the problem reduces to a classical maximum likelihood parametric estimation.

[21] Estimation by minimization of the mutual information was first proposed by Comon (1994) who derived a fundamental connection between cumulants, negentropy and mutual information. The approximation of negentropy by cumulants was originally considered much earlier in Jones and Sibson (1987), while the connection between infomax and likelihood was shown in Pearlmutter and Parra (1997), and the connection between mutual information and likelihood was explicitly discussed in Cardoso (2000)

discrete random variable X it is defined as,

$$H(X) = -\sum_i P(X = b_i) \log P(X = b_i), \tag{39}$$

with b_i denoting the possible values of X. In the continuous case, for a continuous random variable X with density $f_X(x)$, the entropy[22] H is defined as,

$$H(X) = -\int f_X(x) \log f_X(x) \mathrm{d}x. \tag{40}$$

A key result from information theory states that among all random variables of equal variance, a Gaussian variable has the largest entropy. Hence entropy could be used as a measure of non-Gaussianity. A related measure of non-Gaussianity is the negentropy which is always non-negative and zero for a Gaussian variable. It is defined as,

$$J(X) = H(X_{gauss}) - H(X), \tag{41}$$

where $H(X_{gauss})$ is the entropy of a Gaussian random variable having the same covariance matrix as X. As shown by Comon (1994), negentropy is invariant for invertible linear transformations and is an optimal estimator of non-Gaussianity with regards to its statistical properties (i.e. consistency, asymptotic variance and robustness). In practice, because we do not know the density, approximations of negentropy are used such as the one by Hyvärinen and Oja (2000),

$$J(X) \approx \sum_{i=1}^{p} k_i [E(G_i(X)) - E(G_i(V))]^2, \tag{42}$$

where k_i are positive constants, V is a standardized Gaussian variable and G_i are non-quadratic functions. The choice of the non-quadratic function has an impact on the robustness of the estimators of negentropy. with $G(x) = x^4$ (kurtosis based) being the least robust while more robust measures would include,

$$g_1(u) = \frac{1}{a_1} \log\cosh a_1 u, \qquad g_2(u) = -\exp(-0.5u^2). \tag{43}$$

Because these non-quadratic functions present a complex nonlinear optimization problem, sophisticated numerical algorithms are usually necessary. Two main algorithms are used, the online and batch methods, with the former based on stochastic gradient methods while in the latter case a popular choice is the natural gradient ascend of likelihood. The FastICA of Hyvärinen and Oja (2000) is a very efficient batch algorithm with a range of options for the non-quadratic functions. It can be used to estimate the components either one at a time by finding maximally non-Gaussian directions or in parallel by maximizing non-Gaussianity or the likelihood. The estimation procedure of the GO-GARCH model can be summarized as follows. First, the FastICA is applied to the whitened data $\boldsymbol{z}_t = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\epsilon}}_t$, where $\widehat{\boldsymbol{\Sigma}}^{1/2}$ is obtained from the eigenvalue decomposition of the OLS residual covariance matrix, returning an estimate of $\boldsymbol{f}_t$, i.e., $\boldsymbol{y}_t = \boldsymbol{W}\boldsymbol{z}_t$. Second, because of the assumption of independence, the likelihood function of the GO-GARCH model is greatly simplified so that the conditional log-likelihood function is expressed as the sum of the individual conditional log-likelihoods, derived from the conditional marginal densities of the factors, i.e., $GH_{\lambda_i}(y_{it}) \equiv GH(y_{it}; \lambda_i, \mu_i\sqrt{h_{it}}, \delta_i\sqrt{h_{it}}, \alpha_i/\sqrt{h_{it}}, \beta_i/\sqrt{h_{it}})$, plus a term for the mixing matrix $\boldsymbol{A}$, estimated in the first step by FastICA:

$$L(\widehat{\boldsymbol{\epsilon}}_t \,|\boldsymbol{\theta}, \boldsymbol{A}) = T \log\left|\boldsymbol{A}^{-1}\right| + \sum_{t=1}^{T}\sum_{i=1}^{N} \log\left(GH_{\lambda_i}(y_{it}|\theta_i)\right) \tag{44}$$

[22] In the continuous case this is usually called *differential entropy.*

where $\boldsymbol{\theta}$ is the vector of unknown parameters in the marginal densities. Because ICA is a linear noiseless model,[23] the implication for this 2 stage estimation in the GO-GARCH model is that uncertainty plays no part in the derivation of the mixing matrix $\boldsymbol{A}$ and hence does not affect the standard errors of the independent factors.
The possibility of modelling the independent factors separately not only increases the flexibility of the model but also its computational feasibility, since the multivariate estimation reduces to N univariate optimization steps plus a term which depends on the factor loading matrix. Thus the independence property of the model allows the estimation of very large scale systems on modern computational grids with the time required to calculate any n-dimensional model equivalent to the time it takes to estimate one single factor in this framework.
In the **rmgarch** package, 2 algorithms for ICA are implemented locally. The FastICA of Hyvärinen and Oja (2000), based on a direct translation of their Matlab code and the RADICAL of Learned-Miller and Fisher III (2003) which offers a robust alternative. Both models allow a choice of common options such as the type of covariance estimator to use for the whitening stage (e.g. Ledoit-Wolf, EWMA) as well as the possibility of dimensionality reduction during the PCA stage. In the latter case, some results for the model are still to be derived and it is therefore considered experimental at this stage.

2.4.2 Conditional Co-Moments

It seems to be a well-established, stylized fact that the unconditional security return distribution is not normal and the mean and variance of returns alone are insufficient to characterize the return distribution completely. This has led researchers to pay attention to the third moment - skewness - and the fourth moment - kurtosis. The validity of the CAPM in the presence of higher-order co-moments and their effects on asset prices has been thoroughly investigated. The simple, single-factor, CAPM only holds under very specific conditions. When asset prices are non-normal and investors have non-quadratic preferences, then they will care about all return moments and not only mean and variance, as in the standard CAPM. There are a number of extensions to the basic two-moments CAPM which predict a linear relationship in which terms like co-skewness and co-kurtosis are priced. For example, Kraus and Litzenberger (1976), Sears and Wei (1985) extended the CAPM to incorporate skewness in asset valuation models but provided mixed results. A few studies have shown that non-diversified skewness and kurtosis play an important role in determining security valuations. Fang and Tsong-Yue (1997), derived a four-moment CAPM where it was shown that systematic variance, systematic skewness and systematic kurtosis contribute to the risk premium of an asset. Harvey and Siddique (2000) examined an extended CAPM, including systematic co-skewness, reporting that conditional skewness explains the cross-sectional variation of expected returns across assets and is significant even when factors based on size and book-to-market are included. As skewness of a portfolio matters to investors, an asset's contribution to the skewness of a broadly diversified portfolio, referred to as "co-skewness" with the portfolio, may also be rewarded. Skewness preference further suggests that the representative investor may adjust his diversified portfolio such that an individual security's contribution to the skewness of the market portfolio may become a component of the security's expected returns. Mathematically, as demonstrated in Conine and Tamarkin (1981), both individual assets' skewness and co-skewness between assets contribute to the skewness of the portfolio which is composed of these assets. Intuitively, as positive (negative) skewness implies a probability of obtaining a large positive (negative) return (relative

[23] According to Hyvärinen and Oja (2000), this can be partially justified by the fact that most of the research on ICA has also concentrated on the noise free model and it has been shown with overwhelming empirical support across a number of different disciplines to be a very good approximation to a more complex model with noise added. Because the estimation of the noise-free model has proved to be a very difficult task in itself, the noise-free model may also be considered a tractable approximation of the more realistic noisy model.

to a benchmark such as the normal distribution), a positive co-skewness of an asset with another asset means that, when the price volatility goes up the return of this asset also goes up. The general acceptance that the conditional density of asset returns is not completely and adequately characterized by the first two moments, implies that the derivation of any measure of risk from that density requires estimates for the higher order co-moments of the return distribution if one is work within a multivariate setting. The linear affine representation of the GO-GARCH model allows to identify closed-form expression for the conditional co-skewness and co-kurtosis of asset returns[24], as described in de Athayde and Flôres Jr (2000). The conditional co-moments of $\boldsymbol{r}_t$ of order 3 and 4 are represented as tensor matrices,

$$\begin{aligned} \boldsymbol{M}_t^3 &= \boldsymbol{A}\boldsymbol{M}_{f,t}^3(\boldsymbol{A}' \otimes \boldsymbol{A}'), \\ \boldsymbol{M}_t^4 &= \boldsymbol{A}\boldsymbol{M}_{f,t}^4(\boldsymbol{A}' \otimes \boldsymbol{A}' \otimes \boldsymbol{A}'), \end{aligned} \tag{45}$$

where $\boldsymbol{M}_{f,t}^3$ and $\boldsymbol{M}_{f,t}^4$ are the $(N \times N^2)$ conditional third co-moment matrix and the $(N \times N^3)$ conditional fourth co-moment matrix of the factors, respectively. $\boldsymbol{M}_{f,t}^3$ and $\boldsymbol{M}_{f,t}^4$, defined as are given by

$$\boldsymbol{M}_{f,t}^3 = \left[\boldsymbol{M}_{1,f,t}^3, \boldsymbol{M}_{2,f,t}^3, \ldots, \boldsymbol{M}_{N,f,t}^3\right] \tag{46}$$

$$\boldsymbol{M}_{f,t}^4 = \left[\boldsymbol{M}_{11,f,t}^4, \boldsymbol{M}_{12,f,t}^4, \ldots, \boldsymbol{M}_{1N,f,t}^4 | \ldots | \boldsymbol{M}_{N1,f,t}^4, \boldsymbol{M}_{N2,f,t}^4, \ldots, \boldsymbol{M}_{NN,f,t}^4\right] \tag{47}$$

where $\boldsymbol{M}_{k,f,t}^3, k = 1, \ldots, N$ and $\boldsymbol{M}_{kl,f,t}^4, k, l = 1, \ldots, N$ are the $(N \times N)$ submatrices of $\boldsymbol{M}_{f,t}^3$ and $\boldsymbol{M}_{f,t}^4$, respectively, with elements

$$\begin{aligned} m_{ijk,f,t}^3 &= E[f_{i,t}f_{j,t}f_{k,t}|\mathfrak{F}_{t-1}] \\ m_{ijkl,f,t}^4 &= E[f_{i,t}f_{j,t}f_{k,t}f_{l,t}|\mathfrak{F}_{t-1}]. \end{aligned}$$

Since the factors f_{it} can be decomposed as $z_{it}\sqrt{h_{it}}$, and given the assumptions on z_{it}, then $E[f_{i,t}f_{j,t}f_{k,t}|\mathfrak{F}_{t-1}] = 0$. It is also true that for $i \neq j \neq k \neq l E[f_{i,t}f_{j,t}f_{k,t}f_{l,t}|\mathfrak{F}_{t-1}] = 0$ and when $i = j$ and $k = l$,

$$E[f_{i,t}f_{j,t}f_{k,t}f_{l,t}|\mathfrak{F}_{t-1}] = h_{it}^2 h_{kt}^2.$$

Thus, under the assumption of mutual independence, all elements in the conditional co-moments matrices with at least 3 different indices are zero. Finally, standardizing the conditional co-moments one obtains conditional co-skewness and co-kurtosis of $\boldsymbol{r}_t$,

$$\begin{aligned} \boldsymbol{S}_{ijk,t} &= \frac{m_{ijk,t}^3}{(\sigma_{i,t}\sigma_{j,t}\sigma_{k,t})}, \\ \boldsymbol{K}_{ijkl,t} &= \frac{m_{ijkl,t}^4}{(\sigma_{i,t}\sigma_{j,t}\sigma_{k,t}\sigma_{l,t})}, \end{aligned} \tag{48}$$

where $\boldsymbol{S}_{ijk,t}$ represents the asset co-skewness between elements i, j, k of $\boldsymbol{r}_t$, $\sigma_{i,t}$ the standard deviation of $\boldsymbol{r}_{i,t}$, and in the case of $i = j = k$ represents the skewness of asset i at time t, and similarly for the co-kurtosis tensor $\boldsymbol{K}_{ijkl,t}$. Two natural applications of return co-moments matrices are Taylor type utility expansions in portfolio allocation and higher moment news impact surfaces. In the **rmgarch** package the covariance, correlation, coskewness and cokurtosis can be extracted from any of the returned GO-GARCH objects (*goGARCHfit*, *goGARCHfilter*, *goGARCHforecast*, *goGARCHsim goGARCHroll*) by using the methods *rcov*, *rcor*, *rcoskew* and *rcokurt*, respectively. Additional arguments to these methods are clearly detailed in the help files. To obtain the weighted portfolio moments, using the geometric properties of the model, the method *gportmoments* can be called on any of the GO-GARCH objects together with a weighting matrix.

[24]It is possible to go beyond these moments but the notation becomes cumbersome and the benefits likely to be marginal.

2.4.3 The Portfolio Conditional Density

An important question that can be addressed in this framework is the determination of the portfolio conditional density, an issue of vital importance in risk management application. The N-dimensional NIG distribution, closed under convolution, is suited to problems in portfolio and risk management where a weighted sum of assets is considered. However, when the distributional parameters α and β, representing skew and shape, are allowed to vary per asset, as in the GO-GARCH case, this property no longer holds and numerical methods such as that of the Fast Fourier Transform (FFT) are needed to derive the weighted density by inversion of the characteristic function of the scaled parameters[25]. In the case of the NIG distribution, this is greatly simplified because of the representation of the modified Bessel function for the GIG shape index (λ) with value -0.5 which was derived in Barndorff-Nielsen and Bläsild (1981), otherwise the characteristic function of the GH involves the evaluation of the modified Bessel function with complex arguments, which complicates the inversion. Appendix 3 derives the characteristic functions used in the case of independent margins for both the NIG and full GH distributions. Let R_t be the portfolio return:

$$R_t = \boldsymbol{w}_t'\boldsymbol{r}_t = \boldsymbol{w}_t'\boldsymbol{m}_t + (\boldsymbol{w}_t'\boldsymbol{A}\boldsymbol{H}_t^{1/2})\boldsymbol{z}_t \tag{49}$$

where $\boldsymbol{H}_t^{1/2}$ is estimated from the GARCH dynamics of $\boldsymbol{y}_t$. The model allows to express the portfolio variance, skewness and kurtosis in closed form,

$$\begin{aligned} \sigma_{p,t}^2 &= \boldsymbol{w}_t'\boldsymbol{\Sigma}_t\boldsymbol{w}_t, \\ s_{p,t} &= \frac{\boldsymbol{w}_t'\boldsymbol{M}_t^3(\boldsymbol{w}_t \otimes \boldsymbol{w}_t)}{(\boldsymbol{w}_t'\boldsymbol{\Sigma}_t\boldsymbol{w}_t)^{3/2}}, \\ k_{p,t} &= \frac{\boldsymbol{w}_t'\boldsymbol{M}_t^4(\boldsymbol{w}_t \otimes \boldsymbol{w}_t \otimes \boldsymbol{w}_t)}{(\boldsymbol{w}_t'\boldsymbol{\Sigma}_t\boldsymbol{w}_t)^2}, \end{aligned} \tag{50}$$

where $\boldsymbol{\Sigma}_t$, $\boldsymbol{M}_t^3$ and $\boldsymbol{M}_t^4$ are derived in (45). The portfolio conditional density may be obtained via the inversion of the characteristic function through the FFT method as in Chen et al. (2007) (see Appendix 3 for details) or by simulation. The former is used in this package for its accuracy and speed. Provided that $\boldsymbol{z}_t$ is a N-dimensional vector of innovations, marginally distributed as 1-dimensional standardized GH, the density of weighted asset return, $w_{it}r_{it}$, is

$$w_{i,t}r_{i,t} = (w_{i,t}m_{i,t} + \overline{w}_{i,t}z_{i,t}) \sim GH_{\lambda_i}\left(\overline{w}_{i,t}\mu_i + w_{i,t}m_{i,t}, \left|\overline{w}_{i,t}\right|\delta_i, \frac{\alpha_i}{\left|\overline{w}_{i,t}\right|}, \frac{\beta_i}{\left|\overline{w}_{i,t}\right|}\right) \tag{51}$$

where $\overline{\boldsymbol{w}}_t'$ is equal to $\boldsymbol{w}_t'\boldsymbol{A}\boldsymbol{H}_t^{1/2}$, and $\overline{w}_{i,t}$ is the i-th element of $\overline{\boldsymbol{w}}_t$, $m_{i,t}$ the conditional mean of the i-th underlying asset. In order to obtain the density of the portfolio, we must sum the individual weighted densities of $z_{i,t}$. The characteristic function of the portfolio return R_t is

$$\varphi_R(u) = \prod_{i=1}^{n}\varphi_{\bar{w}Z_i}(u) = \exp\left(\mathrm{i}u\sum_{j=1}^{d}\bar{\mu}_j + \sum_{j=1}^{d}\left(\frac{\lambda_j}{2}\log\left(\frac{\gamma}{\upsilon}\right) + \log\left(\frac{\boldsymbol{K}_{\lambda_j}(\bar{\delta}_j\sqrt{\upsilon})}{\boldsymbol{K}_{\lambda_j}(\bar{\delta}_j\sqrt{\gamma})}\right)\right)\right) \tag{52}$$

where, $\gamma = \bar{\alpha}_j^2 - \bar{\beta}_j^2$, $\upsilon = \bar{\alpha}_j^2 - (\bar{\beta}_j + \mathrm{i}u)^2$, and $(\bar{\alpha}_j, \bar{\beta}_j, \bar{\delta}_j, \bar{\mu}_j)$ are the scaled versions of the parameters $(\alpha_i, \beta_i, \delta_i, \mu_i)$ as shown in (51). The density may be accurately approximated by FFT as follows,

$$f_R(r) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{(-\mathrm{i}ur)}\varphi_R(u)\mathrm{d}u \approx \frac{1}{2\pi}\int_{-s}^{s} e^{(-\mathrm{i}ur)}\varphi_R(u)\mathrm{d}u. \tag{53}$$

[25] This effectively means that the weighted density is not necessarily NIG distributed.

Once the density is formed by FFT inversion of the characteristic function, distribution, quantile and sampling functions can be created. In the **rmgarch** package these are represented are *dfft*, *pfft*, *qfft* and *rfft*, which operate on the point in time conditional density approximation, an object of class *goGARCHfft*, returned from calling the *convolution* method on a fitted (*goGARCHfit*), filtered (*goGARCHfilter*), forecasted (*goGARCHforecast*), simulated (*goGARCHsim*) or rolling (*goGARCHroll*) object. Finally, the *nportmoments* method applied to a *goGARCHfft* object will return the FFT-based semi analytic portfolio moments.

2.4.4 Forecasting

The multi-step ahead forecast of the GO-GARCH model is based completely on the univariate factor dynamics, already covered in the **rugarch** package. Additionally, all methods available for working with a fitted (*goGARCHfit*) object are also available for the resulting forecast (*goGARCHforecast*) object and covered in detail in the help file, and the examples in the inst folder of the package.

3 Miscellaneous

Like the **rugarch** package, parallel functionality is implemented by passing a pre-created cluster object from the parallel package. Unlike the **rugarch** package, there is a much higher cost to the use of a socket (snowfall) rather than fork (multicore) based setup, and depending on the number of sockets used, it may be the case that the data communication overhead is so high that non-parallel estimation is faster.
A comprehensive set of examples is available in the rmgarch.tests folder of the source. There are 5 main files, covering the Copula, DCC, FDCC and GO-GARCH models and the fScenario and fMoments methods for use in portfolio and risk management applications (see the **parma** package).

Appendices

The GH characteristic function

The moment generating function (MGF) of the GH Distribution is,

$$\begin{aligned} M_{GH(\lambda,\alpha,\beta,\delta,\mu)}(u) &= e^{\mu u} M_{GIG\left(\lambda,\delta\sqrt{\alpha^2-\beta^2}\right)}\left(\frac{u^2}{2}+\beta u\right), \\ &= e^{\mu u}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^{\lambda/2} \frac{\boldsymbol{K}_\lambda\left(\delta\sqrt{\alpha^2-(\beta+u)^2}\right)}{\boldsymbol{K}_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)} \end{aligned} \tag{54}$$

where M_{GIG} represents the moment generating function of the Generalized Inverse Gaussian which forms the mixing distribution in this variance-mean mixture subclass. Powers of the MGF, $M_{GH}(u)^p$, only have the representation in (54) for $p = 1$, which means that GH distributions are not closed under convolution with the exception of the NIG, and only in the case when the shape and skew parameters are the same. The MGF of the NIG is,

$$M_{NIG(\alpha,\beta,\delta,\mu)}(u) = e^{\mu u}\frac{e^{\delta\sqrt{\alpha^2-\beta^2}}}{e^{\delta\sqrt{\alpha^2-(\beta+u)^2}}}. \tag{55}$$

Powers of p are equivalent in this case to multiplication by p of δ and μ, so that,

$$NIG(\alpha,\beta,\delta_1,\mu_1)\times ... \times NIG(\alpha,\beta,\delta_n,\mu_n) = NIG(\alpha,\beta,\delta_1+...+\delta_n,\mu_1+...+\mu_n). \tag{56}$$

When the distribution is not closed under convolution, numerical methods are required such as the inversion of the characteristic function by FFT. Because the MGF is a holomorphic function for complex z, with $|z| < \alpha-\beta$, we can obtain the characteristic function of the GH distribution, using the following representation,

$$\phi_{GH}(u) = M_{GH}(\mathrm{i}u), \tag{57}$$

so that the characteristic function may be written as,

$$\phi_{GH(\lambda,\alpha,\beta,\delta,\mu)}(u) = e^{\mu \mathrm{i}u}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+\mathrm{i}u)^2}\right)^{\lambda/2} \frac{\boldsymbol{K}_\lambda\left(\delta\sqrt{\alpha^2-(\beta+\mathrm{i}u)^2}\right)}{\boldsymbol{K}_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}. \tag{58}$$

and for the NIG this is simplified to,

$$\phi_{NIG(\alpha,\beta,\delta,\mu)}(u) = e^{\mu \mathrm{i}u}\frac{e^{\delta\sqrt{\alpha^2-\beta^2}}}{e^{\delta\sqrt{\alpha^2-(\beta+\mathrm{i}u)^2}}}. \tag{59}$$

In order to find the portfolio density in the case of the GO-GARCH (maGH/maNIG) model, the characteristic function required for the inversion of the NIG density was already used in Chen et al. (2010) and given below,

$$\phi_{port}(u) = \exp\left\{\mathrm{i}u\sum_{j=1}^{d}\bar{\mu}_j + \sum_{j=1}^{d}\bar{\delta}_j\left(\sqrt{\bar{\alpha}_j^2-\bar{\beta}_j^2}-\sqrt{\bar{\alpha}_j^2-\left(\bar{\beta}_j+\mathrm{i}u\right)^2}\right)\right\} \tag{60}$$

where $\bar{\alpha}_j$, $\bar{\beta}_j$, $\bar{\delta}_j$ and $\bar{\mu}_j$ represent the parameters scaled as described in the main text of the paper. In the case of the GH characteristic function, this is a little more complicated as it involves the evaluation of modified Bessel function of the third kind with complex arguments.[26] Taking logs and summing,

$$\phi_{port}(u) = \exp\Bigg\{ \mathrm{i}u \sum_{j=1}^{d} \Big(\bar{\mu}_j + \frac{\lambda_j}{2} \log\left(\bar{\alpha}_j^2 - \bar{\beta}_j^2\right) - \frac{\lambda_j}{2} \log\left(\bar{\alpha}_j^2 - (\bar{\beta}_j + \mathrm{i}u)^2\right) + \log\left(\boldsymbol{K}_{\lambda_j}\left(\bar{\delta}_j\sqrt{\bar{\alpha}_j^2 - (\bar{\beta}_j + \mathrm{i}u)^2}\right)\right) - \log\left(\boldsymbol{K}_{\lambda_j}\left(\bar{\delta}_j\sqrt{\bar{\alpha}_j^2 - \bar{\beta}_j^2}\right)\right)\Big)\Bigg\} \tag{61}$$

which is more than 30 times slower to evaluate than the equivalent NIG function because of the Bessel function evaluations.

[26]The Bessel package of Maechler (2012) is used for this purpose.

References

K. Aas and I.H. Haff. The generalized hyperbolic skew student's t-distribution. *Journal of Financial Econometrics*, 4(2):275–309, 2006.

G. Aielli. Dynamic conditional correlations: On properties and estimation. In *Available at SSRN: http://ssrn. com/abstract*, volume 1507743, 2009.

C. Alexander. Orthogonal garch. In *Mastering Risk*, volume 2, pages 21–38. Financial Times-Prentice Hall, 2001.

M.C. Ausin and H.F. Lopes. Time-varying joint distribution through copulas. *Computational Statistics and Data Analysis*, 54(11):2383–2399, 2010.

O.E. Barndorff-Nielsen and P. Bläsild. Hyperbolic distributions and ramifications: Contributions to theory and application. In C. Taillie, G.P. Patil, and B. Baldessari, editors, *Statistical distributions in scientific work*, volume 4, pages 19–44. Reidel, 1981.

L. Bauwens and S. Laurent. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics*, 23(3):346–354, 2005.

L. Bauwens, S. Laurent, and J.V.K. Rombouts. Multivariate garch models: A survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

M. Billio, M. Caporin, and M. Gobbo. Flexible dynamic conditional correlation multivariate garch models for asset allocation. *Applied Financial Economics Letters*, 2(2):123–130, 2006.

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

T. Bollerslev, R.F. Engle, and J.M. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116, 1988.

P.H. Boswijk and R. van der Weide. Method of moments estimation of go-garch models. *Journal of Econometrics*, 163(1):118–126, 2011.

W. Breymann, A. Dias, and P. Embrechts. Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1):1–14, 2003.

S.A. Broda and M.S. Paolella. Chicago: A fast and accurate method for portfolio risk calculation. *Journal of Financial Econometrics*, 7(4):412, 2009.

L. Cappiello, R.F. Engle, and K. Sheppard. Asymmetric correlations in the dynamics of global equity and bond returns. *Journal of Financial Econometrics*, 4(4):537–572, 2006.

J.F. Cardoso. Entropic contrasts for source separation: geometry and stability. In Simon Haykin, editor, *Unsupervised adaptive filters*, pages 139–190. John Wiley & sons, 2000.

Y. Chen, W. Härdle, and V. Spokoiny. Portfolio value at risk based on independent component analysis. *Journal of Computational and Applied Mathematics*, 205(1):594–607, 2007.

Y. Chen, W. Härdle, and V. Spokoiny. Ghica–risk analysis with gh distributions and independent components. *Journal of Empirical Finance*, 17(2):255–269, 2010.

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

P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

T. E. Conine and M. J. Tamarkin. On diversification given asymmetry in returns. *Journal of Finance*, 36(5):1143–1155, 1981.

Christophe Croux and Kristel Joossens. Robust estimation of the vector autoregressive model by a least trimmed squares procedure. In Paula Brito, editor, *COMPSTAT 2008*, pages 489–501. Physica-Verlag HD, 2008. ISBN 978-3-7908-2084-3. URL `http://dx.doi.org/10.1007/978-3-7908-2084-3_40`.

A.C. Davison and R.L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–442, 1990.

G.M. de Athayde and R.G. Flôres Jr. Introducing higher moments in the capm: Some basic ideas. In C. Dunis, editor, *Advances in Quantitative Asset Management*. Springer, 2000.

S. Demarta and A.J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1):111–129, 2005.

V. Durrleman, A. Nikeghbali, and T. Roncalli. Which copula is the right one. *http://ssrn.com/abstract=1032545*, 2000. URL `http://ssrn.com/abstract=1032545`.

R.F. Engle. Dynamic conditional correlation. *Journal of Business and Economic Statistics*, 20 (3):339–350, 2002.

R.F. Engle and K. Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate garch. *NBER Working Paper*, 2001.

R.F. Engle, V. Ng, and M. Rothschild. Asset pricing with a factor arch covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45:213–237, 1990.

Hsing Fang and Lai Tsong-Yue. Co-kurtosis and capital asset pricing. *Financial Review*, 32(2): 293–307, 1997.

C. Fernandez and M.F. Steel. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371, 1998.

R. Frey and A.J. McNeil. Modelling dependent defaults. *http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=273*, 2000.

C. Genest, K. Ghoudi, and L.P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.

A. Ghalanos. *spd: The Semi Parametric Distribution*, 2012. URL `http://CRAN.R-project.org/package=spd`. R package version 1.7.

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

C. Harvey and A. Siddique. Conditional skewness in asset pricing tests. *Journal of Finance*, 55 (3):1263–1295, 2000.

A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

T. Jeantheau. Strong consistency of estimators for multivariate arch models. *Econometric Theory*, 14(1):70–86, 1998.

H. Joe. *Multivariate models and dependence concepts*, volume 73. Chapman & Hall/CRC, 1997.

H. Joe, H. Li, and A.K. Nikoloulopoulos. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270, 2010.

E. Jondeau and M. Rockinger. The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25(5):827–853, 2006.

M.C. Jones and R. Sibson. What is projection pursuit? *Journal of the Royal Statistical Society: Series A (General)*, 150(1):1–37, 1987.

M. Junker and A. May. Measurement of aggregate risk with copulas. *Econometrics Journal*, 8 (3):428–454, 2005.

A. Kraus and R. Litzenberger. Skewness preference and the valuation of risk assets. *Journal of Finance*, 31(4):1085 – 1100, 1976.

W.H. Kruskal. Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861, 1958.

Erik G. Learned-Miller and John W. Fisher III. Ica using spacings estimates of entropy. *Journal of Machine Learning Research*, 4:1271–1295, December 2003. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=945365.964306`.

DX Li. On default correlation: A copula function approach. *Journal of Fixed Income*, 9(4): 43–54, 2000.

F. Lindskog, A. Mcneil, and U. Schmock. Kendall's tau for elliptical distributions. In *Credit Risk: Measurement, Evaluation and Management*. Physica-Verlag, 2003.

M. Maechler. *Bessel: Bessel – Bessel Functions Computations and Approximations*, 2012. URL `http://CRAN.R-project.org/package=Bessel`. R package version 0.5-4.

Y. Malevergne and D. Sornette. Testing the gaussian copula hypothesis for financial assets dependences. *Quantitative Finance*, 3(4):231–250, 2003.

A.J. Patton. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556, 2006.

B. A. Pearlmutter and L. C. Parra. Maximum likelihood blind source separation a context-sensitive generalization of ica. *Advances in Neural Information Processing Systems*, 9:613–619, 1997.

S.A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3): 341–360, 1976.

R.S. Sears and K.C.J. Wei. Asset pricing, higher moments, and the market risk premium: A note. *Journal of Finance*, 40:1251–1253, 1985.

A. Silvennoinen and T. Teräsvirta. Multivariate garch models. In *Handbook of Financial Time Series*, pages 201–229. Springer, 2009.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8(1):11, 1959.

Y.K. Tse and A.K.C. Tsui. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economic Statistics*, 20 (3):351–362, 2002.

R. van der Weide. Go-garch: a multivariate generalized orthogonal garch model. *Journal of Applied Econometrics*, 17(5):549–564, 2002.

R. van der Weide. Wake me up before you go-garch. In *Computing in Economics and Finance*. Society for Computational Economics, 2004.

A. Zeevi and R. Mashal. Beyond correlation: Extreme co-movements between financial assets. *Columbia University*, 2002.

K. Zhang and L. Chan. Efficient factor garch models and factor-dcc models. *Quantitative Finance*, 9(1):71–91, 2009.