On the structure and estimation of hierarchical Archimedean copulas

Ostap Okhrin [a], Yarema Okhrin [b,*], Wolfgang Schmid [c]

[a] Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, D-10099 Berlin, Germany
[b] Department of Statistics, University of Augsburg, D-86135 Augsburg, Germany
[c] Department of Statistics, European University Viadrina, D-15230 Frankfurt (Oder), Germany

1. Introduction

The modelling and estimation of multivariate distributions is one of the most critical issues in financial and economic applications. The distributions are usually restricted to the class of multivariate elliptical distributions. The main drawback of this procedure is the small number of alternative distributions and the large number of parameters for a given family. The seminal result of Sklar (1959) provides a partial solution to this problem. It allows separation of the marginal distributions from the dependence structures between the random variables. Since the theory of modelling and estimation of univariate distributions is well established, compared to the multivariate case, the initial problem reduces to modelling the dependence by copulas.

Hence the theoretical properties and application of copulas have recently attracted much attention in the academic literature. The semiparametric estimation of copula functions and parameters is discussed by Chen and Fan (2006), Chen et al. (2006), Genest et al. (1995), Joe (2005), Wang and Wells (2000), etc. Nonparametric estimation is discussed by Fermanian and Scaillet (2003), Chen and Huang (2007), etc. Goodness-of-fit tests are developed in the papers of Chen and Fan (2005), Chen et al. (2004), Fermanian (2005), Genest et al. (2006), and Genest and Rémillard (2008). A detailed review and discussion of Archimedean copulas is given in Joe (1997) and Nelsen (2006). Genest and Rivest (1993) derive important properties and provide an estimation technique. An in-depth discussion of simulation methodologies for Archimedean copulas is provided in Whelan (2004) and McNeil (2008). Copulas have found an increasing number of applications, especially in finance. An application to risk management is discussed by Embrechts et al. (2002) and Junker and May (2005). Portfolio selection problems are considered by Hennessy and Lapan (2002) and Patton (2004). Theoretical foundations of using copulas for GARCH models and applications are provided by Chen and Fan (2005). Lee and Long (2009) and Giacomini et al. (2009) consider time varying copulas.

In principle, in all the mentioned papers, the authors consider either elliptical or Archimedean copulas. However, the number of parameters by elliptical copulas increases quadratically and the postulated dependence is symmetric. The simple Archimedean copulas put too much structure on the dependence between the variables. One possibility to generalize the simple bivariate copulas to multivariate framework offer the vine-copulas, suggested by Bedford and Cooke (2002) and popularized by Aas et al. (2009). The vine copulas allow cascade-type decomposition of a multivariate distribution into a product of conditional bivariate copulas, graphically shown as a sequence of trees. The model is extremely flexible and special types of vines, like *C*-, *D*- or *R*-vines, were suggested to reduce the number of potential alternative models. Acar et al. (2012) discuss generalization of vines using nonparametric smoothing to overcome the potentially misleading results due to simplifying assumptions. The key advantage of the vines is that at each node an arbitrary bivariate copula can be used to model the conditional dependence. This implies that, on

* Corresponding author. Tel.: +49 17622041901.
E-mail addresses: yarema.okhrin@wiwi.uni-augsburg.de, okhrin@gmail.com (Y. Okhrin).

the one hand, that vines can be fitted in very high dimensions; while, on the other hand, the number of potential candidate models increases dramatically. Another class of multivariate model copulas constitute the skew-elliptical copulas popularized by Smith et al. (2012).

In this paper we advocate the Hierarchical Archimedean copulas (HAC), which allow a more flexible and intuitive dependence structure compared with the simple Archimedean copulas, and need a smaller number of parameters compared to elliptical copulas. Joe (1997) introduced this class of copulas, while Whelan (2004) and McNeil (2008) provide simulation techniques for HAC. Savu and Trede (2008) and Savu and Trede (2010) discuss goodness-of-fit tests and estimations of HACs with a known structure. Okhrin et al. (forthcoming) provide detailed statistical properties, like tail dependence, distribution of the copula function, etc.

Despite their advantages, little research has been done on determining the best grouping of the variables in hierarchical copulas, nor on their estimation and simulation. In a few empirical papers, such as Savu and Trede (2008) and Savu and Trede (2010), the structure of the copula was motivated economically. This approaches fails in higher dimensions or for very heterogeneous data sets. This paper attempts to fill this gap and provides a method for estimating the structure of an HAC. This can be seen as a part of copula selection problem. However, we concentrate not on the empirical comparison of alternative copulas, but on the estimation. We propose several approaches to determine the optimal aggregation for the next step at each level of the hierarchy. The first approach is based on a dimension free goodness-of-fit test. The second approach is based on binary trees and aggregated binary trees.

The subset of variables with the strongest fit are joined with the chosen copula function. We estimate the parameters by a multi-stage maximum-likelihood procedure and provide the relevant asymptotic theory. In the simulation study, this method of grouping dominates the goodness-of-fit approach.

This paper is structured as follows. Section 2 contains a short review of copulas. The main results of the present paper concern the determination of the structure and the estimation of copulas. These are presented in Section 3. Section 4 provides a simulation study and empirical applications of the methods developed. Section 5 concludes and all proofs are given in the Appendix.

2. Theoretical background: Archimedean and hierarchical Archimedean copulas

For an arbitrary k-dimensional continuous distribution function F, the associated *copula* is unique and defined as the continuous function $C : [0, 1]^k \to [0, 1]$ such that $F(x_1, \ldots, x_k) = C(F_1(x_1), \ldots, F_k(x_k)), x_1, \ldots, x_k \in \mathbb{R}$, where $F_1(x_1), \ldots, F_k(x_k)$ are the respective continuous marginal distributions.

To formalize the definition of the new class of copulas, let us introduce two classes of functions. The first is the family of strictly decreasing differentiable functions

$$\mathcal{L} = \{\phi : [0, \infty) \to [0, 1] \mid \phi(0) = 1,\ \phi(\infty) = 0,\ (-1)^j \phi^{(j)} \geq 0,\ j = 1, \ldots, \infty\}.$$

This family is also known as the class of Laplace transforms of strictly positive random variables. The second class $\mathcal{L}^*$ comprises functions with similar properties

$$\mathcal{L}^* = \{\omega : [0, \infty) \to [0, \infty) \mid \omega(0) = 0,\ \omega(\infty) = \infty,\ (-1)^{j-1} \omega^{(j)} \geq 0,\ j = 1, \ldots, \infty\}.$$

The function $C : [0, 1]^k \to [0, 1]$ defined as

$$C(u_1, \ldots, u_k) = \phi(\phi^{-1}(u_1) + \cdots + \phi^{-1}(u_k))$$

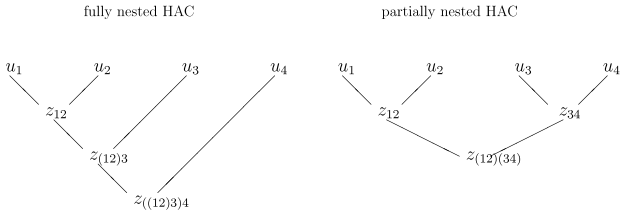

Fig. 1. Examples of HAC $k = 4$.

is called a k-dimensional *Archimedean copula* if $\phi \in \mathcal{L}$. The function ϕ is referred to as the generator of the copula. McNeil and Nešlehová (2009) give less restrictive conditions on the generator ϕ, which is required to be k-*monotone*, i.e., $k-2$ times differentiable with $(-1)^i \phi^{(i)}(x) \geq 0$, $i = 0, \ldots, k-2$ for any $x \in [0, \infty)$ and with $(-1)^{k-2} \phi^{(k-2)}(x)$ being nondecreasing and convex on $[0, \infty)$. However, here we restrict ourselves to the completely monotone case. In general the generator ϕ depends on a vector of parameters $\boldsymbol{\theta}$.

A simple multivariate Archimedean copula implies that the variables are exchangeable. This means that the distribution of $(u_1, \ldots, u_k)$ is the same as that of $(u_{j_1}, \ldots, u_{j_k})$ for all $j_\ell \neq j_v$. This is rarely a feasible assumption in practical applications. A much more flexible method is provided by *hierarchical Archimedean copulas* (HACs). In the special case of fully nested copulas, the copula function is given by

$$\begin{aligned} C(u_1, \ldots, u_k) = \phi_{k-1}(\phi_{k-1}^{-1} \circ \{\phi_{k-2}[\ldots(\phi_2^{-1} \circ \phi_1[\phi_1^{-1}(u_1) \\ + \phi_1^{-1}(u_2)] + \phi_2^{-1}(u_3)) \\ + \cdots + \phi_{k-2}^{-1}(u_{k-1})]\} + \phi_{k-1}^{-1}(u_k)). \end{aligned} \tag{1}$$

This is illustrated together with a partially nested copula in Fig. 1.

To guarantee that C is a multivariate distribution function we assume that $\phi_{k-i}^{-1} \circ \phi_{k-j} \in \mathcal{L}^*$, $i < j$, see Schweizer and Sklar (1983). Unlike the usual AC, HACs define the whole dependence structure in a recursive way. At the lowest level of a fully nested copula, the dependence between the first two variables is modelled by a copula function with generator ϕ_1, i.e., $z_1 = C(u_1, u_2) = \phi_1[\phi_1^{-1}(u_1) + \phi_1^{-1}(u_2)]$. At the second level, another copula function is used to model the dependence between z_1 and u_3, etc. Note that generators ϕ_i can come from the same family and differ only through the parameter or, to introduce more flexibility, can come from different generator families. Fully nested HACs were also discussed by Joe (1997), Whelan (2004), Savu and Trede (2010), and Embrechts et al. (2002).

In contrary to other copula classes, like elliptical or vines, the density of the HAC cannot be given explicitly and should be derived from the copula function by taking derivatives. It can be potentially a tedious task, but should not be seen as a drawback. Note that the generator functions are usually relatively simple functions allowing us to compute analytic (see Hofert and Pham (0000)) or numeric derivatives even in high dimensions. Furthermore, due to numerical integration it is commonly more difficult to compute the distribution function using the density function than vice versa.

HACs have a list of interesting properties. First, since the structure is hierarchical, at each node of the tree we have a marginal distribution which is also given by an HAC. Taking into account the structure of the copula function and the properties of the generator functions, we see that this also holds for any subset of the variables. This implies that the HAC family is closed under taking multivariate margins. For example, by setting $u_k = 1$ in (1) and using the fact that $\phi_{k-1}^{-1}(1) = 0$ we obtain

$$\begin{aligned} C(u_1, \ldots, u_{k-1}, 1) = \phi_{k-2}[\ldots(\phi_2^{-1} \circ \phi_1[\phi_1^{-1}(u_1) + \phi_1^{-1}(u_2)] \\ + \phi_2^{-1}(u_3)) + \cdots + \phi_{k-2}^{-1}(u_{k-1})]. \end{aligned}$$

Thus the marginal distribution of the first $k-1$ variables is also an HAC.

Second, if we consider copulas based on a single generator function, for most cases the copula parameter should increase from the lowest to the highest level. For a fully nested copula this corresponds to $\theta_1 > \theta_2 > \cdots > \theta_{k-1}$. This follows from the condition that $\phi_{k-i}^{-1} \circ \phi_{k-j} \in \mathcal{L}^*$, $i < j$. In the case of an HAC with different generator functions, the bounds on the parameters should be determined individually (see Hofert (2012)). Third, the flexibility of the structure and the fact that there are at most $k-1$ parameters make it a very flexible, but at the same time parsimonious, distribution model. From this point of view, the HAC approach can be seen as a flexible alternative to Gaussian models and to clustering techniques applied to the correlation matrix, see Mantegna (1999) and Tola et al. (2007). Correlation selection techniques for Gaussian copulas were also elaborated by Pitt et al. (2006) within a Bayesian framework for multivariate regression.

The recursive structure of an HAC may be particularly useful in modelling the time-dependence of a univariate time series. In contrary to the cross-sectional dependence, the time ordering imposes a natural structure on the hierarchy and allows for interesting alternatives to autoregressive models. The advantages of vines for modelling longitudinal data are discussed in Smith et al. (2010). These problems go, however, far beyond the issues addressed in this paper.

Despite its importance and flexibility, no studies are available on the estimation of the structure of HACs. Theoretically, for given k and ϕ we can enumerate all possible structures and choose the structure with the best fit. The number of different structures is equal to the number of phylogenetic trees with k nodes. For $k = 10$ this exceeds $2.8 \cdot 10^8$ and simple enumeration is unrealistic. Our aim is to develop a technique for determining the structure of an HAC leading to the best fit of the copula-based distribution to a given data set. We follow in this paper frequentist approach, while alternative procedures for determination of the structure can be developed in Bayesian framework, similarly as it is suggested for drawable vines by Smith et al. (2010) and Min and Czado (2010).

3. Determination of the structure

Let $\mathcal{M} = \{\phi_j\}_{j=1,\dots,M}$ denote the finite set of known generator functions under consideration. We assume that each function ϕ_j depends on a vector of parameters $\boldsymbol{\theta}_j$. Let $X_1, \dots, X_k$ denote the vector of iid random variables whose dependence structure we want to determine and let $\mathbf{X} = \{x_{ji}\}'$ be the respective samples for $i = 1, \dots, n$, $j = 1, \dots, k$. To simplify the presentation we introduce some further notation. Let s denote the structure of the HAC. It is a sequence of reordered indices $\{1, \dots, k\}$ grouped using parentheses to mark the variables joined at a single node. For example, the structures of the copulas in Fig. 1 are $((12)3)4$ and $(12)(34)$ respectively. The quantity s_j is the structure of subcopulas with $s_k = s$ and s_1 is the identity function. We assume that the variables X_j for $j = 1, \dots, k$ follow arbitrary continuous marginal distributions, which can be estimated either parametrically or nonparametrically. Further let the k-dimensional HAC be denoted by $C(\boldsymbol{\phi}, \boldsymbol{\theta}; s)(u_1, \dots, u_k)$, where $\boldsymbol{\phi}$ denotes the set of generating functions for all levels and $\boldsymbol{\theta}$ the set of copula parameters. For example, the fully nested HAC in (1) can be expressed as

$$
\begin{aligned}
&C(\boldsymbol{\phi}, \boldsymbol{\theta}; s = s_k)(u_1, \dots, u_k)\\
&\quad = C(\{\phi_1, \dots, \phi_{k-1}\}, \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{k-1}\}; \{(s_{k-1})k\})(u_1, \dots, u_k)\\
&\quad = \phi_{k-1,\boldsymbol{\theta}_{k-1}}(\phi_{k-1,\boldsymbol{\theta}_{k-1}}^{-1} \circ C(\{\phi_1, \dots, \phi_{k-2}\}, \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{k-2}\};\\
&\qquad ((s_{k-2})(k-1)))(u_1, \dots, u_{k-1}) + \phi_{k-1,\boldsymbol{\theta}_{k-1}}^{-1}(u_k)),
\end{aligned}
$$

where $s = \{(\dots(12)3)\dots k\}$.

Instead of enumerating and testing all possible structures we select at each level j of the hierarchy the best subset of variables I^j to be grouped. For this purpose we estimate for some subsets of the variables the copula parameters and choose the subset with the best fit. We denote this subset by I^j and the optimal generator function and its parameters at level j by ϕ_j and $\boldsymbol{\theta}_j$ respectively. The estimation of the parameters is discussed in Section 4. The selected subset is then used to define the pseudo-variables $C(\phi_j, \hat{\theta}_j; (I^j))$, which are treated as if variables in the usual sense at further levels. The properties of the pseudo-variables were discussed in the case of a bivariate Archimedean copula by Genest and Rivest (1993).

The estimation of the copula for a 4-dimensional data set is illustrated in the case of a single generator function in Diagram 1. At the first step we estimate the copula parameters for all subsets of the variables and choose the best set to be grouped. Assume it is the couple (X_1, X_3). At the next level of the hierarchy we define the sub-copula $Z_{13} = \phi_{\theta_{13}}(\phi_{\theta_{13}}^{-1}(\hat{F}_1(X_1)) + \phi_{\theta_{13}}^{-1}(\hat{F}_3(X_3)))$ and keep it fixed at the subsequent steps. At the next level we estimate the parameters of all subsets of X_2, X_4, $Z_{(13)}$ and assume that the best fit is exhibited by $Z_{(13)}$ and X_4. Therefore, we define the next-level sub-copula as $Z_{((13)4)} = \hat{C}(\phi_2, \theta_2; Z_{(13)}, \hat{F}_4(X_4))$ and it is joined at the highest level of the hierarchy with X_2. The estimated structure is then $s = (((13)4)2)$. If we restrict ourselves only to binary trees, then we consider only the bivariate subsets listed above the horizontal lines.

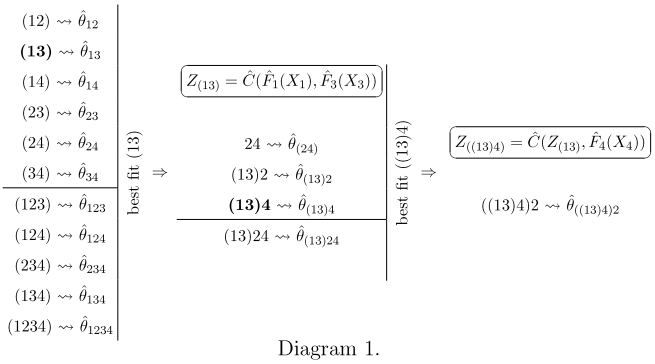

Diagram 1.

In the general case the algorithm can be sketched as follows:

Step 1. Estimate the parameters $\boldsymbol{\theta}_j$ using the ML technique for each set $\{X_i\}_{i \in I_{kj}}$, $j = 1, \dots, 2^k - k - 1$ and each $\phi \in \mathcal{M}$. In the nonparametric setup we first estimate the marginal distributions nonparametrically. Then the copula parameters are determined by solving (4) with $p = 1$. For parametric margins the estimator $\hat{\boldsymbol{\theta}}_j$ is obtained by solving (3) below.

Step 2. Determine the best subset of variables I^1 to be grouped by the given copula. Next, introduce a pseudo-variable defined by $Z_1 = \hat{C}(I^1)$ and consider the set $Z_1 \cup \{X_j\}_{j \in \{1,\dots,k\} \setminus I^1}$ of size $k_2 = k - dim(I^1) + 1$.

Step 3. With the new set, proceed similarly as in Step 1 and consider all subsets of $Z_1 \cup \{X_j\}_{j \in \{1,\dots,k\} \setminus I^1}$ of the variables at level two. Next, for each subset estimate the parameters of the copula by solving (3) or (4) below with $p = 2$.

Step 4. Determine the set of variables with the best copula fit at the second level. Let the obtained grouping be denoted by I^2. It may contain the original variables as well as pseudo-variables obtained from grouping at lower levels. Then, introduce a new variable defined by $Z_2 = \hat{C}(I^2)$ and consider the set $Z_2 \cup \{X_j\}_{j \in \{1,\dots,k_2\} \setminus I^2}$ of size $k_3 = k_2 - dim(I^2) + 1$.

Step 5. ... continue until $k_j = 1$.

Note that the pseudo-variables are only seen as functions of the original variables $X_1, \ldots, X_k$ and are not used explicitly in the estimation procedure. This implies, that while computing the copula density for the ML estimation, we have to take the derivatives with respect to the original variables $X_1, \ldots, X_k$ and not with respect to the pseudo-variables Z. Proceeding in the latter way produces incorrect estimates, as is shown in the simulation study. The correct procedure we call hereafter the recursive maximum likelihood estimation.

Two further important issues arise from the proposed approach. First, the usual maximum likelihood estimation is hindered by the fact that the stochastic properties of estimators at higher levels depend on the fact that the parameters at lower levels are estimated too. A detailed analysis of this issue is provided in Section 4. The second problem is the choice of the subset of variables which can be modelled by a copula in the best way. In Section 4.2 we consider following alternatives: goodness-of-fit, grouping based on binary copulas, and parameter proximity. Although the overall procedure seems to be computationally intensive, the hierarchical structure of the copula leads to substantial numerical simplifications.

The suggested algorithms resembles the procedure suggested for vines in (Dissmann et al. (0000)). In the case of vines the described methodology is formally based on analysing the conditional distributions and the set of potential candidates is enormous. In the case of HAC, however, we concentrate on the unconditional distributions, since the conditional structure is imposed by the form of the HAC. Also the strategy for selecting the optimal subset for grouping is theoretically unambiguous, while vines allow, due to their flexibility, a variety of approaches (see Dissmann et al. (0000) and Aas et al. (2009)).

4. Estimation issues

In this section we discuss the technical aspects of the estimation procedure for the parameters of an HAC. Let $H(x_1, \ldots, x_k)$ denote the true distribution function on a measurable k-dimensional Euclidean space and let g denote its measurable Radon–Nikodým density. Since H is unknown we specify a parametric family of distribution functions $F(x_1, \ldots, x_k, \boldsymbol{\eta})$ with densities $f(x_1, \ldots, x_k, \boldsymbol{\eta})$. It is assumed that the densities are measurable in $(x_1, \ldots, x_k)$ for every $\boldsymbol{\eta}$ from a compact subspace of an Euclidean space and continuous in $\boldsymbol{\eta}$ for every $(x_1, \ldots, x_k)$. The quasi-log-likelihood function of the sample is defined by

$$\mathcal{L}(\boldsymbol{\eta}, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} \log f(x_{1i}, \ldots, x_{ki}, \boldsymbol{\eta}).$$

Then the quasi-ML estimator $\hat{\boldsymbol{\eta}}$ is obtained by maximizing $\mathcal{L}(\boldsymbol{\eta}, \mathbf{X})$ wrt to $\boldsymbol{\eta}$ over a compact subspace of an Euclidean space. This estimator always exists under suitable regularity conditions, but is not necessarily unique.

The linkage between H and F is crucial for the asymptotic properties of the estimator. If F is correctly specified, i.e., if there exists a vector $\boldsymbol{\eta}_0$ such that $F(x_1, \ldots, x_k, \boldsymbol{\eta}_0) = H(x_1, \ldots, x_k)$ for all $(x_1, \ldots, x_k)$, then the estimator $\hat{\boldsymbol{\eta}}$ is consistent for $\boldsymbol{\eta}_0$. Moreover, it is asymptotically normal and achieves asymptotically the Cramer–Rao lower bound. However, if the model is misspecified, i.e., H does not belong to the family F, then the interpretation of the estimator and its properties is not straightforward. The vector $\hat{\boldsymbol{\eta}}$ is an estimator for $\boldsymbol{\eta}_*$ which minimizes the Kullback–Leibler divergence between the true and misspecified models defined by

$$\mathcal{K}(h, f, \boldsymbol{\eta}) = \mathrm{E}_h\{\log[h(x_1, \ldots, x_k)/f(x_1, \ldots, x_k, \boldsymbol{\eta})]\},$$

where the expectation is taken with respect to the true model. The Kullback–Leibler divergence measures the information content of the misspecified new model relative to the true density function. Assume that $\mathrm{E}[\log h(x_1, \ldots, x_k)]$ exists and $|\log f(x_1, \ldots, x_k, \boldsymbol{\eta})| \leq m(x_1, \ldots, x_k)$ for all $\boldsymbol{\eta}$ with m integrable with respect to H. If $\mathcal{K}$ has a unique minimum at $\boldsymbol{\eta}_*$, then $\hat{\boldsymbol{\eta}}$ converges almost surely to $\boldsymbol{\eta}_*$. Thus even if the model is misspecified the ML estimation still provides a consistent estimator of the parameters, but the limit is not the true model H.

This discussion is particularly relevant while estimating the HAC. The first form of misspecification arises if the family of the distribution is specified incorrectly, for example HAC vs. Gaussian models. This is a common form of misspecification. The second form is specific to HAC. It arises if we specify the HAC family correctly, but the structure is misspecified. Next we discuss the estimation of HAC paying particular attention to misspecification issues.

4.1. Estimation of HACs

The properties and goodness of the fit of estimators of the copula parameters $\boldsymbol{\theta}$ depend on the estimators $\hat{F}_j$ for $j = 1, \ldots, k$ of the marginal distributions. We distinguish between a parametric and a nonparametric specification of the margins. On the one hand, since we are interested in the estimator of the dependence structure, the estimator of $\boldsymbol{\theta}$ should be independent of any parametric models for the margins. On the other hand, in practical applications usually parametric models for individual variables are preferred, see Joe (1997). To simplify the exposition we assume that the margins are specified correctly. This can be justified by the fact that we analyse the margins separately and determine the model with the best fit independently of the copula.

Let $\boldsymbol{\alpha}_j$ denote the parameters of the j-th marginal distribution. The ML estimator is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log \quad c(\boldsymbol{\phi}, \boldsymbol{\theta}; s)(F_1(x_{1i}, \boldsymbol{\alpha}_1), \ldots, F_k(x_{ki}, \boldsymbol{\alpha}_k)) \prod_{j=1}^{k} f_j(x_{ji}, \boldsymbol{\alpha}_j) \Bigg), \tag{2}$$

where $c(\cdot, \cdot; \cdot)$ is the corresponding copula density.

If the margins are estimated nonparametrically, Genest et al. (1995) derive the asymptotic moments of the ML estimators and show their asymptotic normality and consistency. In the case of parametric specification of the marginal distribution, the ML estimation can be performed simultaneously for the parameters of the margins and of the copula function or in a two-stage procedure, see Joe (1997).

The ML estimation can only be applied to HACs with known structures. In this paper the structure is determined using a multistage ML procedure which can be sketched as follows. At the first stage we estimate the marginal distributions. At the next stage we estimate the parameter of the copula at the first level by taking the marginal distributions as known. At further stages the next level copula parameter is estimated assuming that the margins as well as the copula parameters at lower levels are known. In the next sections we provide a detailed description of this procedure and the relevant asymptotic theory.

Parametric margins

Assume that k variables are joined within p hierarchical levels, which means that if $s = ((123)(45))$, then $s = s_3 = ((s_1)(s_2))$ where s_1, s_2, s_3 correspond to the levels 1, 2 and 3 respectively. Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_k')'$ denote the vector of parameters of the marginal distributions and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_p')'$ the parameters of the copulas,

starting with the lowest up to the highest level. The ML for the estimator $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta} = (\boldsymbol{\alpha}', \boldsymbol{\theta}')'$ with parametric margins is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\eta}, \mathbf{X}) &= \sum_{i=1}^{n} \log \Bigg(c(\boldsymbol{\phi}, \boldsymbol{\theta}; s)(F_1(x_{1i}, \boldsymbol{\alpha}_1), \ldots, \\ &\qquad F_k(x_{ki}, \boldsymbol{\alpha}_k)) \prod_{j=1}^{k} f_j(x_{ji}, \boldsymbol{\alpha}_j) \Bigg) \\ &= \sum_{i=1}^{n} \Bigg(\log c(\boldsymbol{\phi}, \boldsymbol{\theta}; s)(F_1(x_{1i}, \boldsymbol{\alpha}_1), \ldots, \\ &\qquad F_k(x_{ki}, \boldsymbol{\alpha}_k)) + \sum_{j=1}^{k} \log f_j(x_{ji}, \boldsymbol{\alpha}_j) \Bigg).\end{aligned}$$

Recall that the $\boldsymbol{\theta}$ are not free parameters and, therefore, the ML procedure requires constrained optimization

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \mathcal{L}(\boldsymbol{\eta}, \mathbf{X}) \quad \text{s.t.} \quad \phi_j^{-1} \circ \phi_i \in \mathcal{L}^*, \quad 0 \le i < j \le p.$$

The structure of the log-likelihood function allows to separate the parameters of margins and of the copula in an additive way. The optimization problem can be solved using the maximization by parts technique of Song et al. (2005). The authors suggest a recursive procedure. In the first step of the recursion we estimate the parameters of margins; in the second step the copula parameters are estimated, by assuming the marginal parameters being constant. The procedure can formally be generalized and applied to estimate the HACs. However, we concentrate on a single loop of the recursion, since the major aim of the paper is a computationally efficient estimator of the copula parameters and of the structure. Because of the constraints on the parameters, the estimation and, particularly, the asymptotic theory for these estimators is very demanding. Multistage estimation eliminates these problems. We obtain the estimators by solving

$$\left(\frac{\partial \mathcal{L}_1}{\partial \boldsymbol{\alpha}_1'}, \ldots, \frac{\partial \mathcal{L}_k}{\partial \boldsymbol{\alpha}_k'}, \frac{\partial \mathcal{L}_{k+1}}{\partial \boldsymbol{\theta}_1'}, \ldots, \frac{\partial \mathcal{L}_{k+p}}{\partial \boldsymbol{\theta}_p'} \right)' = \mathbf{0}, \tag{3}$$

where

$$\begin{aligned}&\mathcal{L}_j = \sum_{i=1}^{n} l_j(\mathbf{X}_i), \quad \text{for } j = 1, \ldots, k+p, \\ &l_j(\mathbf{X}_i) = \log f_j(x_{ji}, \boldsymbol{\alpha}_j), \ \text{for } j = 1, \ldots, k, i = 1, \ldots, n, \\ &l_{j+k}(\mathbf{X}_i) = \log \Bigg[c(\{\phi_\ell, \boldsymbol{\theta}_\ell\}_{\ell=1,\ldots,j}; s_j) \\ &\qquad \left(\{F_m(x_{mi}, \boldsymbol{\alpha}_m)\}_{m \in s_j}\right) \prod_{m \in s_j} f_m(x_{mi}, \boldsymbol{\alpha}_m) \Bigg] \\ &\qquad \text{for } j = 1, \ldots, p, \ i = 1, \ldots, n.\end{aligned}$$

The first k components in (3) correspond to the usual ML estimation of the parameters of the marginal distributions. The next p components reflect the recursive estimation at each level of hierarchy, with each likelihood function depending on the copula parameters at lower levels. This implies that the constraints on the parameters are fulfilled by construction.

Let the solution of this system be denoted by $\tilde{\boldsymbol{\eta}}$. Following Joe (2005) we collect the derivatives of the likelihoods ℓ_j on the LHS of (3):

$$\begin{aligned}\sum_{i=1}^{n} \mathbf{g}(\mathbf{X}_i, \boldsymbol{\eta}) &= \sum_{i=1}^{n} (\mathbf{g}_1'(\mathbf{X}_i, \boldsymbol{\eta}), \ldots, \mathbf{g}_{k+p}'(\mathbf{X}_i, \boldsymbol{\eta}))' \\ &= \sum_{i=1}^{n} \left(\frac{\partial l_1(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_1'}, \ldots, \frac{\partial l_k(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_k'}, \frac{\partial l_{k+1}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_1'}, \ldots, \frac{l_{k+p}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_p'} \right)'.\end{aligned}$$

For $r = 1, \ldots, k+p$ denote by $\mathcal{I}_{rj}^* = -\mathrm{E}\frac{\partial \mathbf{g}_r(\cdot, \boldsymbol{\eta})}{\partial \boldsymbol{\alpha}_j}$ for $j = 1, \ldots, k$ and $\mathcal{I}_{r,k+j}^* = -\mathrm{E}\frac{\partial \mathbf{g}_r(\cdot, \boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j}$ for $j = 1, \ldots, p$. Note that $\mathcal{I}_{rj}^* = \mathbf{0}$ if $p \ne j \le k$ or $j > r$. The same holds if $j \notin s_r$ for $j \le k$ and $r > k$. Furthermore let $\mathcal{J}_{rj}^* = \mathrm{E}[\mathbf{g}_r(\cdot, \boldsymbol{\eta})\mathbf{g}_j'(\cdot, \boldsymbol{\eta})] - \mathrm{E}[\mathbf{g}_r(\cdot, \boldsymbol{\eta})] \ \mathrm{E}[\mathbf{g}_j'(\cdot, \boldsymbol{\eta})]$. If both the margins and the copula are correctly specified, then $\mathcal{I}_{rr}^* = \mathcal{I}_{rr}$ are the Fisher information matrices of the individual ML problems $\frac{\partial \mathcal{L}_r}{\partial \boldsymbol{\alpha}_r} = 0$ for $r = 1, \ldots, k$ and $\frac{\partial \mathcal{L}_r}{\partial \boldsymbol{\theta}_r} = 0$ for $r = k+1, \ldots, k+p$. Furthermore, $\mathcal{J}_{rj}^* = \mathcal{J}_{rj} = \mathrm{E}[\mathbf{g}_r(\cdot, \boldsymbol{\eta})\mathbf{g}_j'(\cdot, \boldsymbol{\eta})]$.

Theorem 1 is a generalization of Proposition A.1 in Joe (2005) which gives the asymptotic behaviour of the multi-stage estimator both in the case of a correctly specified and of a misspecified copula function.

Theorem 1. (a) *Suppose the model is misspecified, and let each marginal density f_r for $r = 1, \ldots, k$ and the copula density at each level of the hierarchy satisfy the regularity conditions A1–A8 (see Appendix A). Then* $\sqrt{n}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \overset{a}{\sim} \mathcal{N}(0, \mathbf{B}^{*-1}\boldsymbol{\Sigma}^*(\mathbf{B}^{*-1})')$, *with* $\boldsymbol{\Sigma}^* = Var[\mathbf{g}(\mathbf{X}, \boldsymbol{\eta})] = \{\mathcal{J}_{rj}^*\}_{r,j=1,\ldots,k+p}$ *and* $\mathbf{B}^* = \{\mathcal{I}_{rj}^*\}_{r,j=1,\ldots,k+p}$ *and* $\boldsymbol{\eta}^*$ *minimizes the Kullback–Leibler divergence between the HAC and the true model.*

(b) *Now suppose the model is correctly specified and let each marginal density f_r for $r = 1, \ldots, k$ and the copula density at each level of the hierarchy satisfy the regularity conditions A1–A4, A7, A9. Then* $\sqrt{n}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \overset{a}{\sim} \mathcal{N}(0, \mathbf{B}^{-1}\boldsymbol{\Sigma}(\mathbf{B}^{-1})')$, *where* $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_{rj}\}_{r,j=1,\ldots,k+p}$ *and* $\mathbf{B} = \{\mathbf{B}_{rj}\}_{r,j=1,\ldots,k+p}$ *with*

$$\begin{aligned}&\boldsymbol{\Sigma}_{rj} = \mathcal{J}_{rj}, \quad \text{for } r, j = 1, \ldots, k, \\ &\boldsymbol{\Sigma}_{rj} = 0, \quad \text{for } r, j = k+1, \ldots, k+p, r \ne j, \\ &\boldsymbol{\Sigma}_{rr} = \mathcal{I}_{rr}, \quad \text{for } r = k+1, \ldots, k+p, \\ &\boldsymbol{\Sigma}_{rj} = 0, \quad \text{for } r = 1, \ldots, k, j = k+1, \ldots, k+p, \\ &\mathbf{B}_{rr} = \mathcal{I}_{rr}, \quad \text{for } r = 1, \ldots, k, \\ &\mathbf{B}_{rj} = 0, \quad \text{for } r, j = k+1, \ldots, k+p, r \ne j, \\ &\mathbf{B}_{rj} = 0, \quad \text{for } r, j = 1, \ldots, k, r \ne j, \\ &\mathbf{B}_{rj} = 0, \quad \text{for } r = 1, \ldots, k, j = k+1, \ldots, k+p, \\ &\mathbf{B}_{rr} = \mathcal{I}_{rr}, \quad \text{for } r = k+1, \ldots, k+p,\end{aligned}$$

$$\mathbf{B}_{rj} = \begin{cases} \mathcal{I}_{rj}, & \text{for } r = k+1, \ldots, k+p, k+1 \le j < r \\ & \text{if } s_j \subset s_r \\ 0, & \text{for } r = k+1, \ldots, k+p, k+1 \le j < r \\ & \text{if } s_j \not\subset s_r, \end{cases},$$

$$\mathbf{B}_{rj} = \begin{cases} \mathcal{I}_{rj}, & \text{for } r = k+1, \ldots, k+p, j = 1, \ldots, k \\ & \text{if } j \in s_r \\ 0, & \text{for } r = k+1, \ldots, k+p, j = 1, \ldots, k \\ & \text{if } j \notin s_r. \end{cases}.$$

Note that the regularity assumptions of the theorem are fulfilled by virtually any common parametric HAC. The correct specification of the copula imposes interesting constraints on the structure of the covariance matrix and, therefore, on the asymptotic dependence between the marginal parameters and copula parameters at different levels. In particular, as follows from the form of the components of $\boldsymbol{\Sigma}$, we obtain that the gradients are asymptotically independent. It holds that $\boldsymbol{\Sigma}^* = \{\mathcal{J}_{rj}\}$ with $\mathcal{J}_{rj}^* \ne \mathbf{0}$ in general, but

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathcal{J}_{11} & \ldots & \mathcal{J}_{1k} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{J}_{k1} & \ldots & \mathcal{J}_{kk} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathcal{I}_{k+1,k+1} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \ldots & \mathcal{I}_{k+p,k+p} \end{pmatrix}.$$

Similarly

$$\mathbf{B} = \begin{pmatrix} \mathcal{J}_{11} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathcal{J}_{kk} & \mathbf{0} & \dots & \mathbf{0} \\ & & & \mathcal{I}_{k+1,k+1} & \dots & \mathbf{0} \\ & \neq\mathbf{0} & & & \ddots & \vdots \\ & & & \neq\mathbf{0} & & \mathcal{I}_{k+p,k+p} \end{pmatrix},$$

with $\mathbf{B}^*$ having the same structure with the elements $\mathcal{I}^*_{rj}$. In the applications, the numerical discrepancy between misspecified and correctly specified models arises from two factors: the first factor is the setting of some components of $\boldsymbol{\Sigma}^*$ to zero; the second factor is the reliance on the information inequality for the individual ML problems. This problem will be assessed in the simulation study below.

Depending on the structure of the hierarchical copula, the derivatives needed to compute the asymptotic covariance matrix may be difficult to obtain. Joe (1997) suggests a jackknife approach to obtain an approximation. In this case only the univariate and multivariate likelihoods have to be programmed. This approach can also be extended to multi-stage estimation following the technique in Joe (1997).

Nonparametric margins

In order to estimate the marginal distributions in a nonparametric way we use the modification of the empirical distribution family

$$\hat{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{I}(x_{ji} \le x), \quad j = 1, \dots, k,$$

where $\mathbf{I}$ denotes the indicator function. Since the results established in this section do not depend on the explicit form of the nonparametric kernel, but on its convergence properties, any other suitable estimator can be used as well. Let us denote $\mathbf{F} = (F_1, \dots, F_k)$ and $\hat{\mathbf{F}} = (\hat{F}_1, \dots, \hat{F}_k)$. The canonical ML estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ solves the system by maximizing the pseudo log-likelihood with nonparametric margins

$$\left(\frac{\partial \mathcal{L}_1}{\partial \boldsymbol{\theta}_1'}, \dots, \frac{\partial \mathcal{L}_p}{\partial \boldsymbol{\theta}_p'}\right)' = \mathbf{0}, \tag{4}$$

where

$$\mathcal{L}_j = \sum_{i=1}^{n} l_j(\mathbf{X}_i) \text{ for } j = 1, \dots, p,$$

$$l_j(\mathbf{X}_i) = \log\left[c(\{\phi_\ell, \boldsymbol{\theta}_\ell\}_{\ell=1,\dots,j};\ s_j)\left(\{\hat{F}_m(x_{mi})\}_{m\in s_j}\right) \prod_{m\in s_j} \hat{f}_m(x_{mi}) \right]$$
$$\text{for } j = 1, \dots, p,$$

where $\hat{f}_j(x)$ is the kernel density estimator with $j = 1, \dots, k$. As in the parametric case, here also we show that the semiparametric estimator $\hat{\boldsymbol{\theta}}$ is asymptotically normal and find its asymptotic moments. The asymptotic theory for nonparametric estimators is usually based on the rank theory, especially the rank statistics considered by Hájek and Šidák (1967) and Bhuchongkul (1964). Here we closely follow the approach of Genest et al. (1995). The next theorem provides intermediate results used further in Theorem 3.

Theorem 2. *Let $C_n(\mathbf{u})$ be a empirical copula*

$$C_n(u_1, \dots, u_k) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{k} \mathbf{I}\{\hat{F}_j(x_{ji}) \le u_j\}$$

and $\mathbf{J}(\mathbf{u})$ be a continuous function on $(0,1)^k \to \mathbb{R}^{\widetilde{p}}$, where $\widetilde{p} \ge 1$ is such that

$$\begin{aligned} \boldsymbol{\mu} &= \mathrm{E}[\mathbf{J}(F_1(X_1), \dots, F_k(X_k))] \\ &= (\mathrm{E}[J_1(F_1(X_1), \dots, F_k(X_k))], \dots, \mathrm{E}[J_{\widetilde{p}}(F_1(X_1), \dots, F_k(X_k))])' \\ &= \int \cdots \int \mathbf{J}(\mathbf{F})\, dC(\mathbf{F}). \end{aligned}$$

Consider the statistic

$$\begin{aligned} \mathbf{R}_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{J}(\hat{F}_1(x_{1i}), \dots, \hat{F}_k(x_{ki})) \\ &= \left\{ \frac{1}{n}\sum_{i=1}^{n} J_1(\hat{F}_1(x_{1i}), \dots, \hat{F}_k(x_{ki})), \dots, \right. \\ &\quad \left. \frac{1}{n}\sum_{i=1}^{n} J_{\widetilde{p}}(\hat{F}_1(x_{1i}), \dots, \hat{F}_k(x_{ki})) \right\}' \\ &= \int \dots \int \mathbf{J}(\hat{\mathbf{F}})\, dC_n(\mathbf{F}). \end{aligned}$$

If $J_{p^*}(\mathbf{u}) \le M\prod_{r=1}^{k} h(u_r)^{a_r}$, for $p^* = 1, \dots, \widetilde{p}$, with $a_r = (-1+\delta)/p_r$, for all $r = 1, \dots, k$, $h(u) = u(1-u)$, $\delta > 0$, $p, q > 0$, $\sum_{r=1}^{k} \frac{1}{p_r} = 1$, then $\mathbf{R}_n \xrightarrow{a.s} \boldsymbol{\mu}$.

If $b_r = (-0.5+\delta)/p_r$, for all $r = 1, \dots, k$, and $\mathbf{J}$ admits continuous partial derivatives on $(0,1)^k$ such that $\partial J_{p^*}(\mathbf{u})/\partial u_r \le M\left[\prod_{j=1, j\ne r}^{k} h(u_j)^{b_j}\right] h(u_r)^{b_r - 1}$ for $p^* = 1, \dots, \widetilde{p}$, then $n^{\frac{1}{2}}(\mathbf{R}_n - \boldsymbol{\mu}) \to \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = Cov\left[\mathbf{J}(\hat{\mathbf{F}}) + \sum_{r=1}^{k} \int \cdots \int \mathbf{I}\{F_r(x_r) \le u_r\} - u_r \right] \times \frac{\partial \mathbf{J}(\mathbf{u})}{\partial u_r}\, dC(\mathbf{u}) \Bigg].$$

Theorem 3. *Suppose the model is misspecified and suppose $\frac{\partial l_j}{\partial \boldsymbol{\theta}_j}$ and $\frac{\partial l_j}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_j'}$ satisfy the conditions of the $\mathbf{J}$-function from Theorem 2. Then the estimator $\hat{\boldsymbol{\theta}}$ is consistent and*

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{d}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{B}^{-1}\boldsymbol{\Sigma}\mathbf{B}^{-1}),$$

where the components $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_{rj}\}_{r,j=1,\dots,p}$ and $\mathbf{B} = \{\mathbf{B}_{rj}\}_{r,j=1,\dots,p}$ can be determined from

$$\begin{aligned} \boldsymbol{\Sigma}_j &= Cov\left[\frac{\partial l_j(\hat{F}_1(x_1), \dots, \hat{F}_k(x_k))}{\partial \boldsymbol{\theta}_j} \right. \\ &\quad \left. + \sum_{r=1}^{k} \int \cdots \int \mathbf{I}\{F_r(x_r) \le u_r\} - u_r \Big] \frac{\partial^2 l_j(\mathbf{u})}{\partial \boldsymbol{\theta}_j \partial u_r}\, dC(\mathbf{u}) \right] \\ &\quad \text{for } j = 1, \dots, p, \\ \boldsymbol{\Sigma}_{rj} &= \mathbf{0} \quad \text{for } r, j = 1, \dots, p, \\ \mathbf{B}_{rj} &= -\mathrm{E}\left[\frac{\partial^2 l_j(\mathbf{x})}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_j'} \right] \quad \text{for } r, j = 1, \dots, p. \end{aligned}$$

The second part of the expression for the $\boldsymbol{\Sigma}_j$ arises due to the estimation of the marginal distributions. The correctly specified case is obtained by taking the moments with respect to the true measures.

Discussion of the estimation issues

In general the classical single-step ML estimators are usually obtained as the solution of a system of highly non-linear equations.

This requires sophisticated numerical techniques to avoid local optima. This is of special importance in problems with many parameters. In the case of HAC we have an additional restriction which reflects the monotonicity of the parameters from the bottom to the top of the hierarchy. This complicates even further the estimation and particularly the asymptotic theory.

The multistage procedure considered here lacks these problems, because the systems in (3) and (4) are solved equation by equation. In each equation $\partial\mathcal{L}/\partial\boldsymbol{\theta}_j$ (or a system of equations if we have a generator function with several parameters) the parameters estimated from previous equations $(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{j-1})$ are taken as known quantities. This leads to a dramatic reduction in the computations and provides a more stable solution. Furthermore, the monotonicity of the parameters is automatically fulfilled and need not be enforced in the optimization, nor taken into account in the asymptotic theory.

After the grouping, the parameters of the copula can be in general reestimated. Since the structure of the copula is fixed we can use the single-step MLE to achieve efficiency. Nevertheless, it is still computationally demanding in higher dimensions due to several reasons. First, the copula density and the derivatives for the ML function have to be determined numerically. Second, the maximization is performed with respect to many parameters and subject to the above mentioned monotonicity constrains. Third, because of the constrains the asymptotic theory is difficult to establish. This diminishes the advantages of the reestimation.

The estimation of the copula parameter can be performed using Kendall's τ and the method of moments as proposed by Genest and Rivest (1993). This approach has several disadvantages. By estimating the θ from Kendall's τ, we cannot guarantee that $\theta_1 > \cdots > \theta_{k-1}$ (or $\phi_{k-i} \circ \phi_{k-j} \in \mathcal{L}^*_\infty$ $(i < j)$ for copulas based on different generator functions). Furthermore, the functional relationship between τ and θ in the bivariate case is strongly nonlinear for most of the popular generator functions. Thus minor changes in θ may lead to strong changes in τ and vice versa. The method, however, works well with different generator functions and with multiparameter copulas.

4.1.1. Extension to SCOMDY models

Let us assume that the process $\mathbf{X}_t$ is formed by the innovations of the normalized semiparametric copula-based multivariate dynamic models (SCOMDY) process $\mathbf{Y}_t$, i.e.,

$$\mathbf{Y}_t = \boldsymbol{\mu}_t(\boldsymbol{\beta}_0) + \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)\mathbf{X}_t,$$

where

$$\begin{aligned}\boldsymbol{\mu}_t(\boldsymbol{\alpha}_0) &= \mathrm{E}(\mathbf{Y}_t|\mathcal{F}_{t-1}),\\ \boldsymbol{\Sigma}_t(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) &= \mathrm{diag}\{\sigma_{jt}^2(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)\}_{j=1,\ldots,k}\\ &= \mathrm{diag}\{\mathrm{E}[(Y_{jt} - \mu_{jt}(\boldsymbol{\beta}_0))^2|\mathcal{F}_{t-1}]\}_{j=1,\ldots,k}.\end{aligned}$$

The SCOMDY parametrization is very flexible and allows for a variety of model specifications such as VAR and GARCH, see Chen and Fan (2006). The likelihood function of $\mathbf{Y}_t$ is given by

$$\begin{aligned}\mathcal{L}(\{\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\eta}\}, \mathbf{Y}) = &-\frac{1}{2n}\sum_{i=1}^{n}\log|\boldsymbol{\Sigma}_t^2(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)|\\ &+\frac{1}{n}\sum_{i=1}^{n}\log f(x_{1i}, \ldots, x_{ki}, \boldsymbol{\eta}).\end{aligned}$$

We maximize the likelihood function with respect to the parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, the joint density f, and the parameters $\boldsymbol{\eta}$. Since the aim of this paper is the estimation of HAC we assume for simplicity that $\boldsymbol{\mu}_t(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ are correctly specified. To estimate the structural part we follow the approach of Chen and Fan (2006). The estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ can be consistently estimated using a two-stage procedure as follows.

$$\hat{\boldsymbol{\beta}}_0 = \arg\min_{\boldsymbol{\beta}_0}\left\{-\frac{1}{2n}\sum_{i=1}^{n}(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0))'(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0))\right\},$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{\beta}_1}\Bigg[-\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{k}\Bigg\{&\frac{(Y_{ji} - \mu_{ji}(\hat{\boldsymbol{\beta}}_0))^2}{\sigma_{ji}^2(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_1)}\\ &+ \log\sigma_{ji}^2(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_1)\Bigg\}\Bigg].\end{aligned}$$

At the second stage we estimate the distribution of the innovations $\mathbf{X}_t$ given by an HAC. In general the estimation of the copula parameters is asymptotically independent of the estimation of the structural parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$. This follows from Chen and Fan (2006, Proposition 3.2). For this reason we proceed further only with the estimation of the distribution of $\mathbf{X}_t$ given by an HAC.

4.2. Criteria for grouping

At each level of the hierarchy we decide which variables should be gathered into a copula at the next level. Next we discuss several potential alternatives for grouping.

4.2.1. Grouping based on goodness-of-fit tests

The classical goodness-of-fit tests (GOF), for example, Kolmogorov–Smirnov or χ^2-type tests, cannot be applied directly to copula functions, because copulas depend on the estimated marginal distributions. Furthermore, in most cases the distribution of the test statistic is not a standard one and depends on the dimension. A dimension-free test is proposed in Chen et al. (2004). As argued by Chen et al. (2004), the power and size of the test are comparable with other more sophisticated tests. The test does not depend explicitly on the type of the non-parametric estimator of the marginals F_i, but uses the order of $\hat{F}_j(x_{ji}) - F_j(x_{ji})$ as a function of n. Thus we apply the test of Chen et al. (2004) as one of the criterion for grouping the variables at each level of the hierarchical copula. The set of variables with the smallest test statistic is joined by a copula at the next level.

Note that the GOF method appears to lead frequently to misleading results due to the low power of the test against alternative structures that are close to the true structure. Thus this method of grouping has several drawbacks, however, we still apply in the numerical studies for comparison purposes.

4.2.2. Grouping based on binary structures

A computationally simple approach is to determine a binary copula. First we consider a single generator function. Let us denote as $\{I_{ki}\}_{i=1,\ldots,2^k-k-1}$ the subsets of the initial set of size k, excluding the empty set and single element sets. At each level of a binary hierarchy we join the two variables with the strongest dependence. This implies in most cases the couple with the highest value of the copula parameter. More formally, the pair I^j to be grouped at level j is determined from

$$I^j = \arg\max_{I_{k-j+1,i}, |I_{k-j+1,i}|=2}\theta(I_{k-j+1,i}).$$

By $|I_{ki}|$ we denote the cardinality of the set I_{ki}. This approach always leads to a feasible copula function with $k-1$ parameters. Note that the procedure is not a specification test at each level of the hierarchy, but merely a deterministic procedure, which selects the couple with the strongest fit. This procedure also guarantees that the copula parameters are monotone as argued above.

If the true copula is not binary, the procedure leads to a potentially misspecified model. Despite a difference in structure,

the difference in the distribution functions is in general minor. To illustrate this point, we consider the following binary HAC $C^1(\phi, (\theta_1, \theta_2); ((12)3))(x_1, x_2, x_3)$. If the parameters are close then the dependence structure imposed by C^1 is very close to the dependence structure imposed by $C^2(\phi, \theta_1; (123))(x_1, x_2, x_3)$. This property is referred to as associativity of Archimedean copulas, see Theorem 4.1.5 of Nelsen (2006).

This fact additionally provides us with tools for aggregation of binary copulas to general HACs. To join three variables into a single copula the corresponding parameters of the binary copulas should be close. We can measure the distance in two ways. One approach uses the asymptotic test with $H_0 : \theta_j = \theta_{j+1}$, where θ_j and θ_{j+1} are copula parameters at two subsequent levels. The test can be implemented using the asymptotic results from Section 4. Alternatively, to make the procedure similar to the tools of the cluster analysis, we may use the threshold approach and join the variables if, for example, $(\theta_1 - \theta_2)/\theta_1 < \alpha$.

The aggregation of binary trees can be augmented by considering the parameters of subcopulas of dimension higher than two. To group ℓ variables, we compare the parameters of ℓ-variate copulas with the parameters of all possible $\ell - 1$-variate subcopulas. We formalize this idea using the minimax principle. Let

$$\Delta = \min_{I_{ki},\, |I_{ki}|\geq 3} \max_{J\subset I_{ki},\, |J|=k-1} |\theta(I_{ki}) - \theta(J)|$$

and the minimum is achieved at the subset I_Δ. For each subset of variables we find the largest distance between its θ and the θs of its own subsets with a dimension of $k-1$. Then the subset with the smallest maximal distance is taken as I_Δ. This procedure, however, does not allow of selecting a bivariate copula as the best grouping. To improve this we determine the subset of variables to be grouped using the criterion

$$I^1 = \begin{cases} I_\Delta, & \Delta \leq \delta \\ \underset{I_{ki},\, |I_{ki}|=2}{\operatorname{argmax}}\ \theta(I_{ki}), & \Delta > \delta \end{cases}.$$

If Δ is larger than some predetermined constant δ, then the parameters of the copulas with higher dimension are too far from the parameters of the lower dimensional copulas. In this case we take the couple with the largest copula parameter as the best grouping. If we choose δ small, then we prefer binary copulas, with only couples of variables joined at each level. For larger values of δ, joining several variables at once becomes more probable. The actual choice of δ depends on the parameter domain of θ. We can either fix the percentage deviation of the subset parameters or use confidence intervals based on the asymptotic theory above.

If the multivariate copula is constructed using different generator functions, we cannot directly apply the above method of grouping and of aggregating. However, at each level we can choose the couple of variables to be joined for each generator function. The KL divergence is used to select the couple with the better fit. Note that if the generator function is different at two subsequent levels of the binary copula, the corresponding variables cannot be joined.

5. Simulation study

The aim of this simulation study is the comparison of the discussed grouping methods for an example of simulated data. We assess the goodness-of-fit of the estimated models and the ability to capture the dependence correctly. Furthermore, we assess the impact of misspecification by analysing the fit of HAC to Gaussian data.

5.1. Comparison of the grouping techniques

To compare the grouping methods we consider two different true structures $s = (123)(45)$ and $s = (12(34))5$ with the Gumbel generator function given by $\phi^{-1} = (-\log(u))^\theta$ and $\phi = \exp\{-u^{1/\theta}\}$. The parameters for the first structure are set equal to $\theta_{123} = 4, \theta_{45} = 3$, and $\theta_{(123)(45)} = 2$, and for the second structure, to $\theta_{34} = 4, \theta_{12(34)} = 3$, and $\theta_{(12(34))5} = 2$. Without loss of generality the marginal distributions are taken to be uniform on [0, 1] and are not estimated. We simulate 1000 samples of size 500.

For the simulation we use the method based on Laplace transforms suggested in McNeil (2008) and further assessed in Hofert (2011), Hofert (2012). To implement the ML maximization, we need the corresponding density functions. These are determined explicitly by taking analytic derivatives of the copula using R software. The computation is fast even in big dimensions. For each simulated data set of the study and each structure we compared the fit, the structure and the ability to capture the dependence for the suggested grouping procedures. We consider the simple Archimedean copula (sAC) and groupings based on the Chen et al. (2004) test statistics (Chen), on the θ, binary copulas (binary), aggregated binary copulas (binary aggr.). The binary aggregated grouping with recursive estimation is denoted by RML. As benchmark models we consider the 5-dimensional Gaussian copula (Gauss) with $\hat{\boldsymbol{\Sigma}}$ estimated from the data and the t copula.

The results of this part of the simulation study are summarized in Table 1 for the first structure and in Table 2 for the second structure. The second column contains the three most frequent estimated structures with average parameter values. The frequency of the true structure for binary groupings is relative low. However, since the parameters are close, the values of the copula function for other structures will be numerically close to the value for the true structure. This is additionally illustrated in the next two columns, which contain the average Kullback–Leibler divergence and the average test statistic of Chen et al. (2004). The frequency of the true structure is much higher for the grouping based on Kendall's τ and for the recursive reestimation. Note, however, that the estimators are clearly biased in the former case. We observe that for the KL loss the grouping based on θ's with recursive reestimation clearly dominates the alternatives, followed by both grouping methods based on τ, the binary and the aggregated binary trees. The results for the Chen statistics obviously put more weights on the corresponding grouping method, but still show clear outperformance of HAC compared to the Gaussian and t-models.

To assess the ability of the estimated model to capture the dependence structure of the data, we evaluate the discrepancy between the pairwise true and the estimated Kendall's τ and the lower and upper tail indices. More precisely, we calculate the Frobenius norm of the difference between the Kendalls matrix for the true model and the Kendalls matrix for the model under consideration, say P_{model}. The smallest error exhibits again the grouping based on θ's with recursive reestimation. The grouping based on τ's is the second best option, followed by groupings based on θ's. Note that the Kendall τ's are well estimated by the Gaussian and t-copulas. The Frobenius norms of the differences in the true and estimated upper and lower tail indices are given in the columns λ_U and λ_L respectively. However, frequently recommended t-copula clearly fails to capture the zero lower-tail dependence and heavily biases the upper part. This is an important drawback, while the aim of copulas is the modelling the dependence and tail dependence. The last column of the tables contains the average computing times needed for a single replications. The groupings based on τ and on binary trees are computationally very efficient.

Next we address the issue of efficiency loss due to multistage ML estimation. Table 3 provides the true parameters, the average parameters from the multistage and the average parameters from single step ML estimation. Since the variance of the asymptotic distribution for the full ML is unknown due to the imposed inequality constraints on the parameters, we provide the corresponding empirical standard deviations. The efficiency loss at all levels of the hierarchy is generally small. A stronger jump in the standard deviation is observed if more than two variables are joint together at a single level, as we observe for the first structure. If, however, only two variables are joined, the loss is minor.

Table 1
Model fit for the true structure (12(34))5. The columns contain the average values of Kullback–Leibler divergence between the estimated and the true model (KL); the value of the test statistics of the GoF test of Chen et al. (2004) (Chen); the Frobenius norms of the differences between estimated and true Kendall matrices and upper/lower tail indices; the estimation time in s. The values in parenthesis are the corresponding standard deviations.

Method	Copula structure(s)	%	KL	CF	Av. error in τ	λ_U	λ_L	time (in s)
Gauss			0.2896 (0.0418)	1021.0584 (86.1239)	0.6417 (0.0298)	3.0882 (0.0000)	0.0000 (0.0000)	0.0003 (0.0004)
t			0.1992 (0.0281)	1068.3437 (85.0886)	0.6442 (0.0279)	1.4610 (0.1430)	1.8445 (0.1469)	37.2612 (22.0799)
sAC	$(12345)_{2.37}$	100.0	0.4963 (0.0664)	55.0961 (6.6516)	0.4338 (0.0147)	0.3938 (0.0127)	0.0000 (0.0000)	1.1138 (0.8659)
$\tau_{\triangle\tau>0}$	$(5(12(34)_{4.03})_{3.33})_{2.18}$	98.8	0.0318 (0.0219)	126.1234 (24.9078)	0.1627 (0.0582)	0.1488 (0.0539)	0.0000 (0.0000)	0.5360 (0.4191)
	$(5(1234)_{3.58})_{2.17}$	1.1						
	$(5(1(2(34)_{3.97})_{3.74})_{3.50})_{2.24}$	0.1						
τ_{binary}	$(5(2(1(34)_{4.02})_{3.34})_{3.33})_{2.18}$	51.3	0.0312 (0.0207)	125.9834 (24.8084)	0.1624 (0.0581)	0.1487 (0.0538)	0.0000 (0.0000)	0.5235 (0.3619)
	$(5(1(2(34)_{4.02})_{3.35})_{3.34})_{2.17}$	47.6						
	$(5((34)_{3.93}(1.20)_{3.21})_{3.21})_{2.22}$	1.1						
Chen	$(15(234)_{3.20})_{2.10}$	12.1	0.4512 (0.1377)	42.4741 (16.4991)	0.4939 (0.0891)	0.4503 (0.082)	0.0000 (0.0000)	382.2255 (116.2116)
	$(13(245)_{2.18})_{2.09}$	11.0						
	$(34(125)_{2.18})_{2.08}$	10.9						
θ	$(5(3(124)_{2.99})_{2.63})_{1.73}$	46.4	0.1464 (0.0288)	49.3676 (7.9451)	0.2983 (0.0504)	0.2828 (0.0512)	0.0000 (0.0000)	6.0718 (3.0262)
	$(5(4(123)_{3.01})_{2.65})_{1.73}$	44.6						
	$(5(1234)_{3.06})_{1.76}$	9.0						
θ_{binary}	$(5((34)_{4.00}(12)_{3.07})_{3.07})_{1.78}$	38.9	0.0312 (0.0163)	65.9947 (12.4950)	0.2196 (0.0562)	0.2152 (0.0562)	0.0000 (0.0000)	0.7069 (0.3823)
	$(5(2(1(34)_{4.03})_{3.06})_{2.59})_{1.75}$	32.3						
	$(5(1(2(34)_{4.03})_{3.06})_{2.59})_{1.75}$	28.8						
$\theta_{binary\ aggr.}$	$(5(12(34)_{4.00})_{3.07})_{1.78}$	38.9	0.0312 (0.0163)	66.0036 (12.5048)	0.2196 (0.0562)	0.2152 (0.0562)	0.0000 (0.0000)	0.6652 (0.3404)
	$(5(2(1(34)_{4.03})_{3.06})_{2.59})_{1.75}$	32.3						
	$(5(1(2(34)_{4.03})_{3.06})_{2.59})_{1.75}$	28.8						
θ_{RML}	$(5(12(34)_{4.01})_{3.02})_{2.00}$	81.1	−0.0025 (0.0032)	84.1208 (7.7031)	0.0509 (0.0253)	0.0472 (0.0244)	0.0000 (0.0000)	3.2885 (2.1005)
	$(5((12)_{3.12}(34)_{4.03})_{2.87})_{1.99}$	17.5						
	$(5(1(2(34)_{4.17})_{3.19})_{2.97})_{2.02}$	0.9						

Table 2
Model fit for the true structure (123)(45). The columns contain the average values of Kullback–Leibler divergence between the estimated and the true model (KL); the value of the test statistics of the GoF test of Chen et al. (2004) (Chen); the Frobenius norms of the differences between estimated and true Kendall matrices and upper/lower tail indices; the estimation time in s. The values in parenthesis are the corresponding standard deviations.

Method	Copula structure(s)	%	KL	CF	Av. error in τ	λ_U	λ_L	time (in s)
Gauss			0.2847 (0.0403)	1145.3618 (96.1629)	0.0789 (0.0299)	3.0261 (0.0000)	0.0000 (0.0000)	0.0003 (0.0005)
t			0.1989 (0.0283)	1214.3247 (93.5416)	0.0682 (0.0268)	1.3398 (0.1500)	1.7787 (0.1334)	40.0308 (24.2342)
sAC	$(12345)_{2.32}$	100.0	0.8038 (0.0711)	62.2205 (7.0608)	0.5249 (0.0097)	0.4710 (0.0078)	0.0000 (0.0000)	1.2514 (1.0350)
$\tau_{\Delta\tau>0}$	$((45)_{3.01}(123)_{4.14})_{2.23}$	100.0	0.0183 (0.0124)	109.0668 (22.0437)	0.1825 (0.0710)	0.1740 (0.0675)	0.0000 (0.0000)	0.5911 (0.3154)
τ_{binary}	$((45)_{3.01}(3(12)_{4.13})_{4.13})_{2.23}$	35.1	0.0182 (0.0124)	109.053 (22.0418)	0.1825 (0.0710)	0.1740 (0.0675)	0.0000 (0.0000)	0.5905 (0.3496)
	$((45)_{3.02}(1(23)_{4.14})_{4.14})_{2.23}$	32.6						
	$((45)_{3.01}(2(13)_{4.15})_{4.15})_{2.24}$	32.3						
Chen	$(45(123)_{4.01})_{2.02}$	18.3	0.7261 (0.2880)	43.7516 (12.3469)	0.5460 (0.1505)	0.4929 (0.1337)	0.0000 (0.0000)	419.2090 (122.4066)
	$(12(345)_{2.18})_{2.01}$	14.8						
	$(23(145)_{2.19})_{2.03}$	13.3						
θ	$(2(1345)_{2.23})_{1.78}$	29.1	0.8196 (0.3398)	50.233 (23.7257)	0.6528 (0.1860)	0.6011 (0.1689)	0.0000 (0.0000)	7.4331 (3.8592)
	$(1(2345)_{2.24})_{1.79}$	28.2						
	$(3(1245)_{2.24})_{1.78}$	27.9						
θ_{binary}	$((45)_{3.01}(3(12)_{4.11})_{3.91})_{2.27}$	34.7	0.0215 (0.0073)	102.6946 (9.1997)	0.2117 (0.0550)	0.2024 (0.0520)	0.0000 (0.0000)	0.8801 (0.4933)
	$((45)_{3.01}(1(23)_{4.11})_{3.91})_{2.28}$	33.4						
	$((45)_{3.01}(2(13)_{4.11})_{3.90})_{2.28}$	31.9						
$\theta_{binary\ aggr.}$	$((45)_{3.00}(123)_{4.02})_{2.28}$	51.4	0.0214 (0.0071)	103.2285 (9.3303)	0.2116 (0.0551)	0.2023 (0.0520)	0.0000 (0.0000)	0.8475 (0.4614)
	$((45)_{3.02}(1(23)_{4.15})_{3.84})_{2.28}$	16.5						
	$((45)_{3.02}(2(13)_{4.16})_{3.85})_{2.27}$	16.3						
θ_{RML}	$((123)_{4.04}(45)_{3.01})_{1.97}$	82.2	−0.0029 (0.0030)	84.3295 (7.2043)	0.0542 (0.0310)	0.0515 (0.0312)	0.0000 (0.0000)	0.5376 (0.0826)
	$((1(23)_{4.2})_{3.94},(4.5)_{3.03})_{1.99}$	6.4						
	$((3(12)_{4.23})_{3.96},(4.5)_{2.99})_{1.97}$	5.8						

Table 3
The average parameters and computational times for multistage ML and full ML estimation based on 1000 simulated samples of size 500. Standard errors are provided in brackets.

	$\hat{\theta}_3,\ (\theta_3 = 4.0)$	$\hat{\theta}_2,\ (\theta_2 = 3.0)$	$\hat{\theta}_1,\ (\theta_1 = 2.0)$	Time (in s)
		Structure $((123)_4(45)_3)_2$		
Multistage	4.028 (0.103)	3.010 (0.112)	1.967 (0.058)	0.496 (0.032)
Full	4.002 (0.100)	3.010 (0.111)	2.002 (0.058)	0.949 (0.060)
		Structure $((12(34)_4)_35)_2$		
Multistage	3.983 (0.148)	2.995 (0.078)	2.003 (0.061)	1.995 (0.372)
Full	3.980 (0.141)	3.004 (0.070)	2.005 (0.061)	2.740 (0.326)

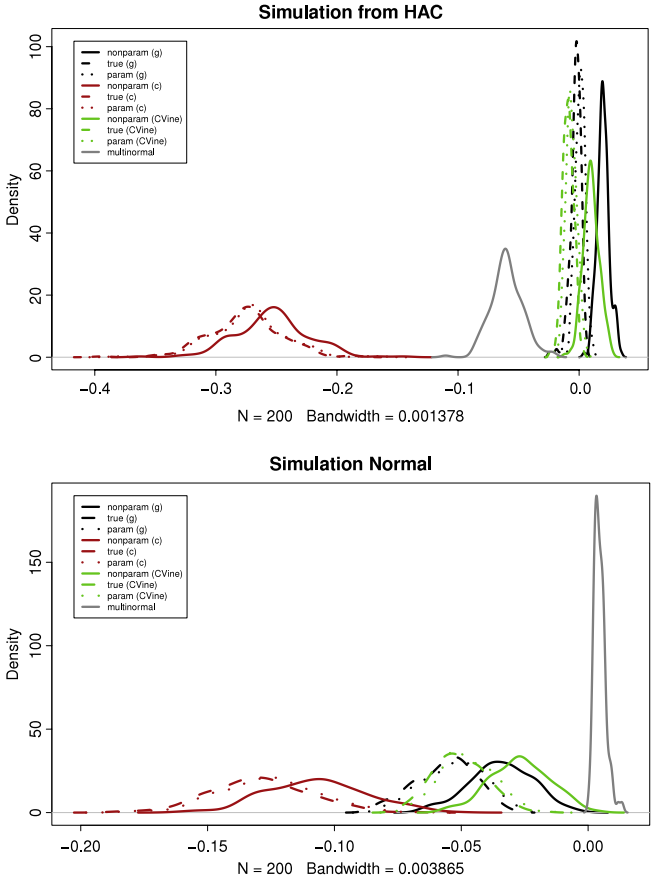

Fig. 2. Illustration of the misspecification problem: kernel density estimators of the distribution of the KL divergence between the estimated and empirical distributions. Simulation from the HAC with Gumbel generators (upper pictures), simulation from the Gauss copula (lower pictures) with 200 runs. In the legend: (g) and (c) refer to the fitted HAC with Gumbel and Clayton copulas respectively.

5.2. Misspecification

In this section we assess the impact of model misspecification on the estimation results. We simulate the data from two models. The first is a three dimensional HAC with Gumbel generator and parameters 1.5 and 2. The marginal distributions are standard normal. The second model is the normal distribution, with correlation matrix identical to the one imposed by the HAC in the first model. To each of the simulated datasets we fit one of four models: HAC with Gumbel generators, HAC with Clayton generators, the multivariate normal distribution and the *C*-vine with mixed generators. For the copula-based models we distinguish between estimation with nonparametric, true, and parametrically estimated margins. In Fig. 2 we plot the nonparametric density estimators of the KL divergence over the replications. The upper figures refer to the data simulated from the HAC model, the bottom figures show the results for the normal data. The model with the best fit should exhibit a narrow density of the KL criteria centred around zero.

If the data is generated from the HAC-based model, the model with the true known margins and correct generator function outperforms the alternatives, followed by the models with the parametric and nonparametric margins. The fitted normal distribution shows a stronger variation and a bias of the KL criteria if the sample size increases. The fit of the HAC with the misspecified Clayton generator is extremely poor. The *C*-vine shows results close to the results from HAC. We conclude that the misspecification caused by the fit of normal distribution to HAC-data is stronger that the misspecification caused by the fit of an HAC model to normal data for smaller samples. Having more parameters and different generators functions, the vines show slightly better performance compared to HAC for Gaussian samples.

6. Empirical example

In this subsection we apply the proposed estimation technique to real data. The purpose of the study is twofold. First, we assess the potential of the HAC model for modelling the returns of the given assets. Second, we evaluate the economic profit of applying the HAC by calculating the VaR for the profit and loss function of a portfolio. We consider daily log-returns of Apple (AAPL), Ford (F), Google (GOOG), Microsoft (MSFT) and Toyota Motors (TM). The sample period covers 1008 observations from 03.01.2007 to 31.12.2010. To overcome the time dependence of the data we fit all five time series to univariate AR(1)–GARCH(1,1) processes with generalized error distributed (GED) residuals (see Nelson (1991)). The estimation results are given in Table 4. We conclude that the fitted process successfully eliminates the autocorrelation (Box–Ljung test) and the hypothesis of GED residuals cannot be rejected by the Kolmogorov–Smirnov test. The subsequent modelling uses these residuals.

The pairwise residuals and the residuals mapped on the unit square by the empirical distribution functions show clear asymmetric behaviour, for example, GOOG vs. AAPL, APPL vs. F, MSFT vs. F, etc. We skip the corresponding pictures for brevity. This fact advocates the attempt to fit the copula-based distribution with GED margins to the five dimensional time series of the residuals. Table 5 contains details on the estimated binary HAC with Gumbel and Clayton generators. Note that the estimated structure are close for both generator functions. For both models first GOOG and AAPL are linked, with MSFT added at the next level. The automotive companies are added either jointly for the Clayton generator or linked sequentially with the Gumbel generator.

In the subsequent analysis we compare the fit of the HAC model with Clayton or Gumbel generators to the fits of several benchmarks. For simplicity we do not consider mixed generators. For possible constellations of different generators within a single HAC refer to Hofert (2011). The first two benchmarks are the Gaussian and t-copula. The simple Archimedean copulas with the same generators build the second class of alternatives.

To assess the dynamics of the estimated parameters and the estimated structure we perform a moving window estimation. At each step 100 observations are used for estimation purposes. The respective time series of the BIC values, Kendall's τ and of

Table 4
Estimation results of fitting univariate AR(1)–GARCH(1,1) processes to the data with the volatility equation $\sigma_t^2 = \gamma_0 + \gamma_1(r_{t-1} - \mu - \omega_1 r_{t-2})^2 + \delta_1 \sigma_{t-1}^2$ and GED residuals. The columns BL and KS contain the p-values of the Box–Ljung for the autocorrelation of the squared residuals and the Kolmogorov–Smirnov test for GED. Second lines contain the standard deviations of the parameters.

	μ	α_1	γ_0	γ_1	δ_1	shape	BL	KS
GOOG	8.585e−04	0.03429	3.323e−06	0.04525	0.94720	1.08482	0.3344	0.1038
	4.271e−04	0.02121	1.809e−06	0.01152	0.01250	0.05705		
AAPL	2.454e−03	−0.00964	6.075e−06	0.08711	0.90523	1.37463	0.1506	0.2886
	5.282e−04	0.03262	4.088e−06	0.02115	0.02279	0.08272		
MSFT	7.803e−05	−0.05871	5.153e−06	0.08253	0.90647	1.21966	0.6493	0.8201
	4.676e−04	0.02550	2.793e−06	0.02396	0.02570	0.07182		
TM	−5.223e−04	−0.08890	4.697e−06	0.08965	0.89870	1.52981	0.9418	0.7060
	4.693e−04	0.03221	2.404e−06	0.01932	0.02068	0.09729		
F	7.963e−04	0.02220	1.483e−05	0.12247	0.87397	1.35513	0.3228	0.1423
	7.356e−04	0.03145	8.834e−06	0.02945	0.02837	0.07890		

Table 5
Estimation results for the fit of the HAC with Gumbel and Clayton generators to the residuals. The standard errors of the parameters are given in the parenthesis.

Generator	estimated structure and parameters
Gumbel	$((((GOOG.AAPL)_{1.692(0.078)}.MSFT)_{1.490(0.046)}.TM)_{1.344(0.028)}.F)_{1.280(0.025)}$
Clayton	$(((GOOG.AAPL)_{1.072(0.120)}.MSFT)_{0.762(0.089)}.(TM.F)_{0.554(0.058)})_{0.548(0.051)}$

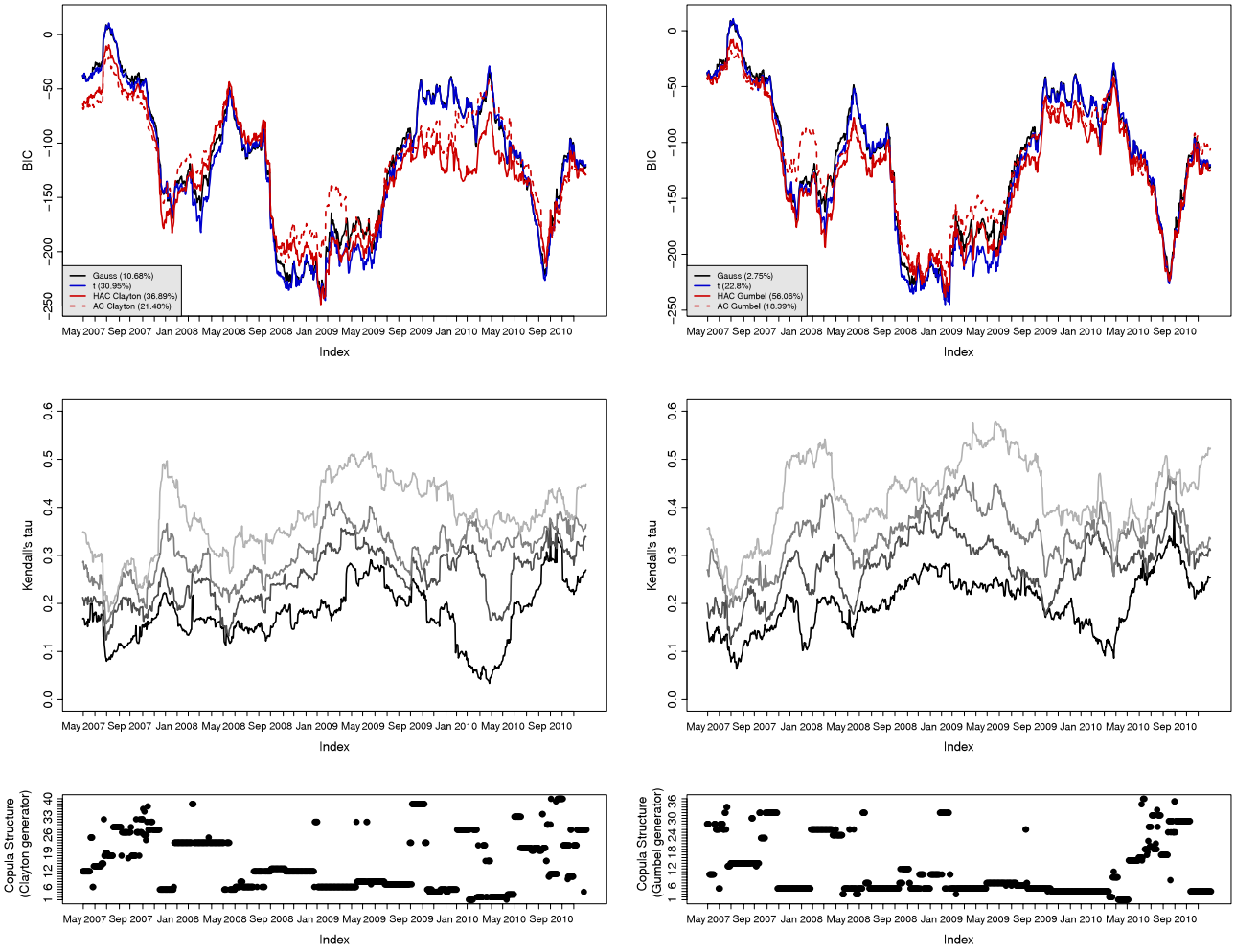

Fig. 3. BIC, the parameters transformed to Kendall's τ and the structure from moving window estimation with window length 100 using Clayton (left) and Gumbel (right) generators respectively.

the structure are shown in Fig. 3. The BIC reveals clear periods where one of the models dominates the alternatives; particularly important are the periods where the HAC-based distribution with Gumbel generator outperforms the Gaussian, t and Clayton benchmarks. The percentage of the days, when one of the models is dominating, is given in the legend. The Kendall's τ and the correspondingly the parameters exhibit clear co-movement with periodic divergences. There is no visual relationship between the performance and the dynamics of the parameters. In the periods with close parameters, the copula functions with different structures are numerically indistinguishable despite possessing different trees. For this reason we observe frequent jumps in the

structure only in the periods with similar parameters, leading to the conclusion that a few different structure are sufficient to model the data over the whole estimation period. This also illustrates the robustness of the estimation procedure: distant parameters necessarily lead to a stable copula structure.

To compare the performance of HAC we rely on the economic significance of the alternative distributions. For the analysis we chose the Gaussian distribution, simple Archimedean copulas and different vines. The vine-based benchmarks consist of six alternative models with three models arising from C-vines and three models from D-vines. Each triple consists of a Clayton based vine, Gumbel based vine and the mixed vine, which allows for a mixture of Clayton, Gumbel, Gaussian and t copulas. Particularly we consider the Value-at-Risk (VaR) of a portfolio of five assets with weights $\mathbf{w}$, assuming individual AR(1)–GARCH(1,1) data generating processes with GED residuals for each asset and the three alternative joint distributions for the residuals. The profit and loss function of the portfolio is defined as $L_{t+1} = \sum_{i=1}^{5} w_i P_{it}(e^{R_{i,t+1}} - 1)$, where P_{it} and R_{it} are the price and the return of the asset i at the time point t respectively. We drop the time index and denote by F_L the distribution function of L_{t+1}. The VaR of the portfolio at level α is given by $VaR(\alpha) = F_L^{-1}(\alpha)$. Since F_L depends on the marginal distributions of the returns as well as on the joint distribution of the residuals, it cannot be derived explicitly. The distribution function F_L is estimated by simulating the paths of the asset returns from the above marginal processes and alternative joint distributions of the residuals. The $\widehat{VaR}(\alpha)$ is computed as the corresponding empirical quantile.

Backtesting allows us to assess the economic significance of the chosen copula model. We estimate the realized α as a relative fraction of the exceedances in the time series, i.e.,

$$\hat{\alpha}_{\mathbf{w}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}\{L_t < \widehat{VaR}_t(\alpha)\}.$$

The relative distance between estimated $\hat{\alpha}$ and true α has been calculated by

$$e_{\mathbf{w}} = |\hat{\alpha}_{\mathbf{w}} - \alpha|/\alpha.$$

Similarly to Giacomini et al. (2009), we compute $\hat{\alpha}_{\mathbf{w}}$ and $e_{\mathbf{w}}$ for a set $W = \{\mathbf{w}^*, \mathbf{w}_n; n = 1, \ldots, 999\}$ of portfolios, for each $\mathbf{w}_n = (w_{n,1}, \ldots, w_{n,5})'$ is the realization of a random vector uniformly distributed on $S = \{(x_1, \ldots, x_5) \in \mathbb{R}^5 : \sum_{i=1}^{5} x_i = 1, x_i \geq 0.1\}$ and $\mathbf{w}^* = (1/5, \ldots, 1/5)'$ is the equally weighted portfolio. Also similarly to Giacomini et al. (2009), to measure the performance of each model we used the average relative exceedance over portfolios and its corresponding standard deviation

$$A_W = \frac{1}{|W|}\sum_{\mathbf{w}\in W} e_{\mathbf{w}}, \qquad D_W = \left\{\frac{1}{|W|}\sum_{\mathbf{w}\in W}(e_{\mathbf{w}} - A_W)^2\right\}^{1/2}.$$

The results of the backtesting are summarized in Tables 6 and 7. Table 6 contains the empirical quantiles $\widehat{VaR}(\alpha)$ and the corresponding standard deviations in parentheses for three different levels of the Value-at-Risk 10%, 5% and 1%. The closer the empirical level to the true level is, the better is the performance of the model. The best model for the 10% level is the mixed C-vine, the mixed D-vine optimal for 5% and, finally, for the 1% level the best model is the D-vine with Clayton generators. The HAC with Clayton generators is not the best model, however, in contrary to the benchmarks, it provides very robust and good performance. For 1% and 10% the empirical level for HAC model deviates from the best model only in the fourth digit, while for 5% it shows the fifth best results among 12 models. Note that is not completely fair to select as an alternative the mixed vine models, since these use explicitly a mixture of Clayton, Gumbel, Gaussian and t copulas within a single structure and depend on at least 10 parameters, compared to four parameters in a HAC. Very similar conclusions can be drawn from Table 7, where the relative exceedances and their standard deviations are presented. Summarizing, from the economic perspective the vine models are favourites, but it is not possible to suggest a single dominating vine model. The HAC with Clayton generator, however, exhibits very good performance, which is robust with respect to the choice of the VaR level.

Table 6
The empirical quantiles $\widehat{VaR}(\alpha)$ and the standard deviation in parenthesis.

	10%	5%	1%
HAC (Gumbel)	0.1070 (0.0048)	0.0586 (0.0061)	0.0173 (0.0020)
AC (Gumbel)	0.1094 (0.0070)	0.0639 (0.0067)	0.0186 (0.0028)
HAC (Clayton)	0.1017 (0.0050)	0.0473 (0.0034)	0.0107 (0.0012)
AC (Clayton)	0.1098 (0.0072)	0.0545 (0.0042)	0.0118 (0.0014)
Gauss	0.1015 (0.0053)	0.0542 (0.0034)	0.0139 (0.0016)
t	0.1015 (0.0048)	0.0530 (0.0033)	0.0128 (0.0015)
vineC (mixed)	0.1010 (0.0042)	0.0528 (0.0033)	0.0139 (0.0015)
vineC (Gumbel)	0.1084 (0.0053)	0.0614 (0.0049)	0.0157 (0.0029)
vineC (Clayton)	0.1028 (0.0044)	0.0492 (0.0033)	0.0119 (0.0016)
vineD (mixed)	0.1034 (0.0049)	0.0503 (0.0032)	0.0125 (0.0011)
vineD (Gumbel)	0.1074 (0.0054)	0.0602 (0.0057)	0.0157 (0.0028)
vineD (Clayton)	0.1060 (0.0047)	0.0496 (0.0030)	0.0105 (0.0016)

Table 7
The average exceedance A_W over all portfolios and its standard deviation D_W.

	10%	5%	1%
HAC (Gumbel)	0.0720 (0.0451)	0.1768 (0.1168)	0.7332 (0.2037)
AC (Gumbel)	0.0952 (0.0681)	0.2793 (0.1329)	0.8591 (0.2797)
HAC (Clayton)	0.0387 (0.0358)	0.0721 (0.0500)	0.1112 (0.0798)
AC (Clayton)	0.1002 (0.0679)	0.1024 (0.0691)	0.1957 (0.1174)
Gauss	0.0437 (0.0331)	0.0935 (0.0553)	0.3876 (0.1593)
t	0.0415 (0.0277)	0.0705 (0.0543)	0.2854 (0.1462)
vineC (optimal)	0.0347 (0.0257)	0.0733 (0.0467)	0.3898 (0.1488)
vineC (Gumbel)	0.0847 (0.0516)	0.2280 (0.0985)	0.5710 (0.2880)
vineC (Clayton)	0.0419 (0.0313)	0.0550 (0.0397)	0.2151 (0.1164)
vineD (optimal)	0.0451 (0.0389)	0.0517 (0.0383)	0.2529 (0.1029)
vineD (Gumbel)	0.0758 (0.0517)	0.2064 (0.1108)	0.5691 (0.2808)
vineD (Clayton)	0.0650 (0.0398)	0.0494 (0.0334)	0.1401 (0.0864)

7. Summary

In many cases multivariate elliptical distributions are considered. However, its flexibility is rather limited and the number of parameters increases dramatically with the dimension. This problem can be partially solved by modelling the dependence with hierarchical Archimedean copulas.

This paper considers the problem of estimating and determining the structure of hierarchical Archimedean copulas. We develop a hierarchical estimation technique, which determines the best grouping of the variables at each level of the hierarchy. We provide an asymptotic theory for the estimated copula parameters and distinguish between parametrically and nonparametrically estimated marginal distributions. Particular attention is paid to misspecified and correctly specified models. Several criteria for grouping are proposed. In practical applications we recommend constructing the HAC using the recursive estimation method or, for computational efficiency, either a binary tree or an aggregated binary tree. For the simulated data the method developed provides good results in determining the true structure. Moreover, the copula-based distributions show a good fit even in modelling normal samples.

Acknowledgements

The authors are grateful to Claudia Czado, Wolfgang K. Härdle, Friedrich Schmid and Vladimir Spokoiny for helpful discussions and suggestions. The first author acknowledges financial support of CRC 649 "Economic Risk".

Appendix A

Let $f(\boldsymbol{\eta})$ denote either the marginal density or the copula density for arbitrary subcopula. The elements of the vector of parameters $\boldsymbol{\eta}$ vary depending on the specification of f.

1. Let $\mathbf{X}_i$ be a stochastic process on a complete probability space with an absolutely continuous distribution.
2. $f : \mathbb{R}^{dim(\mathbf{X})} \times \Pi \longrightarrow \mathbb{R}^+$ are measurable and continuous for each $\boldsymbol{\eta} \in \Pi$.
3. f is continuously twice-differentiable on Π (a.s.).
4. $\mathrm{E}[\log f(\mathbf{X}, \boldsymbol{\eta})]$ exists and is finite for each $\boldsymbol{\eta}$. Additionally $\mathrm{E}[\log f(\mathbf{X}, \boldsymbol{\eta})]$ is continuous on Π (a.s.).
5. $\mathrm{E}[\nabla \log f(\mathbf{X}, \boldsymbol{\eta})]$ exists and is finite for each $\boldsymbol{\eta}$.
6. $\mathrm{E}[\nabla^2 \log f(\mathbf{X}, \boldsymbol{\eta})]$ exists, is finite, positive definite and continuous on Π.
7. $\mathcal{L}_j$ have identifiably unique maximizers on the interior of a subspace of Π for $j = 1, \ldots, p + k$.
8. $n^{-1/2} f(\mathbf{X}_n, \hat{\boldsymbol{\eta}}^*)$ obeys, uniformly in n, the CLT, with a positive definite covariance matrix.
9. For all $\boldsymbol{\eta}$ we have $\int \nabla f(\mathbf{X}, \boldsymbol{\eta}) dH(\mathbf{X}) = 0$ and $\int \nabla^2 f(\mathbf{X}, \boldsymbol{\eta}) dH(\mathbf{X}) = 0$.

Appendix B

Proof of Theorem 1. The asymptotic results follow directly from Theorem 6.2 of White (1982) for misspecified and from Theorem 6.4 of White (1982) for correctly specified models. Here we derive only the expressions for the components of the asymptotic covariance matrix of the correctly specified ML estimators. We have $\mathbf{B} = -\mathrm{E}\big(\partial \mathbf{g}(\mathbf{X}_i, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}'\big)$. The last derivative can be written explicitly in terms of the components of $\boldsymbol{\eta}$

$$\frac{\partial \mathbf{g}(\mathbf{X}_i, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}'} = \begin{pmatrix} \frac{\partial^2 l_1(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_1 \partial \boldsymbol{\alpha}_1'} & \cdots & \frac{\partial^2 l_1(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_1 \partial \boldsymbol{\alpha}_k'} & \frac{\partial^2 l_1(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_1 \partial \boldsymbol{\theta}_1'} & \cdots & \frac{\partial^2 l_1(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_1 \partial \boldsymbol{\theta}_p'} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l_k(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_k \partial \boldsymbol{\alpha}_1'} & \cdots & \frac{\partial^2 l_k(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_k \partial \boldsymbol{\alpha}_k'} & \frac{\partial^2 l_k(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_k \partial \boldsymbol{\theta}_1'} & \cdots & \frac{\partial^2 l_k(\mathbf{X}_i)}{\partial \boldsymbol{\alpha}_2 \partial \boldsymbol{\theta}_p'} \\ \frac{\partial^2 l_{k+1}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\alpha}_1'} & \cdots & \frac{\partial^2 l_{k+1}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\alpha}_k'} & \frac{\partial^2 l_{k+1}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'} & \cdots & \frac{\partial^2 l_{k+1}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_p'} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l_{k+p}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_p \partial \boldsymbol{\alpha}_1'} & \cdots & \frac{\partial^2 l_{k+p}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_p \partial \boldsymbol{\alpha}_k'} & \frac{\partial^2 l_{k+p}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_p \partial \boldsymbol{\theta}_1'} & \cdots & \frac{\partial^2 l_{k+p}(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_p \partial \boldsymbol{\theta}_p'} \end{pmatrix}.$$

The elements on the main diagonal are parts of the information matrices from the individual ML procedures. We denote them by $-\mathrm{E}\frac{\partial^2 \mathcal{L}_r}{\partial \boldsymbol{\alpha}_r \partial \boldsymbol{\alpha}_r'} = \mathcal{I}_{rr}$, for $r = 1, \ldots, k$ and $-\mathrm{E}\frac{\partial^2 \mathcal{L}_{k+j}}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} = \mathcal{I}_{jj}$, for $j = 1, \ldots, p$. Since the likelihood function $\mathcal{L}_r$ for $r = 1, \ldots, k$ is independent of $\boldsymbol{\alpha}_j$ for $j = 1, \ldots, k, r \neq j$ and of $\boldsymbol{\theta}_\ell$ for $\ell = 1, \ldots, p$, this implies that

$$\frac{\partial^2 \mathcal{L}_r}{\partial \boldsymbol{\alpha}_r \partial \boldsymbol{\alpha}_j'} = \frac{\partial^2 \mathcal{L}_r}{\partial \boldsymbol{\alpha}_j \partial \boldsymbol{\alpha}_r'} = \frac{\partial^2 \mathcal{L}_r}{\partial \boldsymbol{\alpha}_r \partial \boldsymbol{\theta}_\ell'} = 0,$$
$$r, j = 1, \ldots, k; r \neq j, \ \ell = 1, \ldots, p.$$

The elements in the last p rows below the main diagonal depend on the structure of the HAC. If the element j has been grouped at the level r, i.e., $j \in s_r$, then $-\mathrm{E}\frac{\partial^2 \mathcal{L}_r}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\alpha}_j'} = \mathcal{I}_{rj}$ for $j = 1, \ldots, k$ and $r = k + 1, \ldots, p$.

The matrix $\boldsymbol{\Sigma}$ is equal to $Var(\mathbf{g}) = \mathrm{E}\mathbf{g}\mathbf{g}'$. From Joe (1997) it follows that $\mathrm{E}\frac{\partial \mathcal{L}_r}{\partial \boldsymbol{\alpha}_r} \frac{\partial \mathcal{L}_{k+1}}{\partial \boldsymbol{\theta}_1'} = 0$ for $r = 1, \ldots, k$. We have to show that also $\mathrm{E}\frac{\partial \mathcal{L}_r}{\partial \boldsymbol{\alpha}_r} \frac{\partial \mathcal{L}_{k+j}}{\partial \boldsymbol{\theta}_j'} = \mathrm{E}\frac{\partial \mathcal{L}_{k+j}}{\partial \boldsymbol{\theta}_j} \frac{\partial \mathcal{L}_{k+\ell}}{\partial \boldsymbol{\theta}_\ell'} = 0$ for $j, \ell = 1, \ldots, p$ and $j \neq \ell$. Let $r \in s_j$, then it holds that

$$\begin{aligned}
&\mathrm{E}\frac{\partial l_r}{\partial \boldsymbol{\alpha}_r} \frac{\partial l_{k+j}}{\partial \boldsymbol{\theta}_j'} \\
&= \int \frac{\partial l_r(\mathbf{y}_r)}{\partial \boldsymbol{\alpha}_r} \frac{\partial l_{k+j}(\{\mathbf{y}_\ell\}_{\ell \in s_j})}{\partial \boldsymbol{\theta}_j'} c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j) \\
&\quad \times (\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) d\mathbf{y}_\ell \\
&= \int \frac{\partial l_r(\mathbf{y}_r)}{\partial \boldsymbol{\alpha}_r} \frac{\partial}{\partial \boldsymbol{\theta}_j'} \log\big(c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j)(\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j})\big) \\
&\quad \times c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j)(\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) d\mathbf{y}_\ell \\
&= \int \frac{\partial l_r(\mathbf{y}_r)}{\partial \boldsymbol{\alpha}_r} \frac{\partial}{\partial \boldsymbol{\theta}_j'} c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j)(\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) d\mathbf{y}_\ell \\
&= \int_{\mathbf{y}_r} \frac{\partial l_r(\mathbf{y}_r)}{\partial \boldsymbol{\alpha}_r} \left[\int_{\mathbf{y}_{-r}} \frac{\partial}{\partial \boldsymbol{\theta}_j'} c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j)(\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \right. \\
&\quad \left. \times \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) \prod_{\ell \in s_j, \ell \neq r} d\mathbf{y}_\ell \right] d\mathbf{y}_r
\end{aligned}$$

(where $\int_{\mathbf{y}_{-r}}$ denotes integration over all variables except for $\mathbf{y}_r$)

$$\begin{aligned}
&= \int_{\mathbf{y}_r} \frac{\partial l_r(\mathbf{y}_r)}{\partial \boldsymbol{\alpha}_r} \left[\frac{\partial}{\partial \boldsymbol{\theta}_j'} \int_{\mathbf{y}_{-r}} c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j)(\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \right. \\
&\quad \left. \times \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) \prod_{\ell \in s_j, \ell \neq r} d\mathbf{y}_\ell \right] d\mathbf{y}_r.
\end{aligned}$$

The internal integral is equal to the marginal density of the r-th variable, and thus it is independent of $\boldsymbol{\theta}_j$. This implies that the derivative and the whole expression are zero. For the second term $\mathrm{E}\frac{\partial \mathcal{L}_{k+j}}{\partial \boldsymbol{\theta}_j} \frac{\partial \mathcal{L}_{k+\ell}}{\partial \boldsymbol{\theta}_\ell'} = 0$ we proceed in a similar way. Let $r \in s_r \subset s_j$. Otherwise the expression is obviously zero, since $\mathrm{E}\mathbf{g}(\mathbf{X}_r, \boldsymbol{\eta}) = 0$. Now we have

$$\begin{aligned}
&\mathrm{E}\frac{\partial l_{k+r}}{\partial \boldsymbol{\theta}_r} \frac{\partial l_{k+j}}{\partial \boldsymbol{\theta}_j'} \\
&= \int \frac{\partial l_{k+r}(\{\mathbf{y}_\ell\}_{\ell \in s_r})}{\partial \boldsymbol{\theta}_r} \frac{\partial l_{k+j}(\{\mathbf{y}_\ell\}_{\ell \in s_j})}{\partial \boldsymbol{\theta}_j'} c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j) \\
&\quad \times (\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) d\mathbf{y}_\ell \\
&= \int \frac{\partial l_{k+r}(\{\mathbf{y}_\ell\}_{\ell \in s_r})}{\partial \boldsymbol{\theta}_i} \frac{\partial}{\partial \boldsymbol{\theta}_j'} c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j) \\
&\quad \times (\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) d\mathbf{y}_\ell \\
&= \int_{\mathbf{y}_\ell, \ell \in s_r} \frac{\partial l_{k+r}(\{\mathbf{y}_\ell\}_{\ell \in s_r})}{\partial \boldsymbol{\theta}_r} \\
&\quad \times \left[\frac{\partial}{\partial \boldsymbol{\theta}_j'} \int_{\mathbf{y}_\ell, \ell \in s_j / s_r} c(\boldsymbol{\phi}_j, \boldsymbol{\theta}_j; s_j) \right. \\
&\quad \left. \times (\{F_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j)\}_{\ell \in s_j}) \prod_{\ell \in s_j} f_\ell(\mathbf{y}_\ell, \boldsymbol{\alpha}_j) \prod_{\ell \in s_j / s_r} d\mathbf{y}_\ell \right] \prod_{\ell \in s_r} d\mathbf{y}_\ell.
\end{aligned}$$

The inner integral is equal to the joint distribution of the variables in s_r and is, therefore, independent of $\boldsymbol{\theta}_j$. This implies that the derivative with respect to $\boldsymbol{\theta}_j$ and the whole expression is zero. This completes the proof of the theorem.

Proof of Theorem 2. By the Lemma of Glivenko–Cantelli, $C_n(\mathbf{F})$ converges almost surely to $C(\mathbf{F})$. $\mathbf{R}_n$ can be written in the following way

$$\begin{aligned}\mathbf{R}_n &= \mathrm{E}(\mathbf{V}_n) = \mathrm{E}(V_{1,n}, \ldots, V_{\widetilde{p},n}) \\ &= \mathrm{E}(\mathbf{J}(v_{1n}, \ldots, v_{kn})) \\ &= (\mathrm{E}(J_1(v_{1n}, \ldots, v_{kn})), \ldots, \mathrm{E}(J_{\widetilde{p}}(v_{1n}, \ldots, v_{kn})))', \\ &\quad \text{where } v_{1n}, \ldots, v_{kn} \sim C_n(\mathbf{F}).\end{aligned}$$

Observe that $(v_{1n}, \ldots, v_{kn}) \xrightarrow{\mathcal{D}} (v_1, \ldots, v_k)$, where $(v_1, \ldots, v_k)$ are random variables distributed according to $C(\mathbf{F})$. Because of the continuity of $\mathbf{J}$, $\mathbf{V}_n$ converges to $\mathbf{V} = \mathbf{J}(v_1, \ldots, v_k)$. To complete the proof we have to show that all components of $\mathbf{V}_n$ are uniformly integrable, i.e., that there exists an $\epsilon > 0$ such that $\mathrm{E}(|V_{p^*,n}|^{1+\epsilon})$ for $p^* = 1, \ldots, \widetilde{p}$ are bounded. To prove this, we use a multivariate extension of Hölder's inequality. For positive functions $f_1(x), \ldots, f_k(x) : (a, b) \to \mathbb{R}$,

$$\int_a^b \prod_{j=1}^k f_j(x)\, dx \le \prod_{j=1}^k \left(\int_a^b f_j(x)^{p_j} dx \right)^{1/p_j}.$$

From now on, we fix such p^* that $1 \le p^* \le \widetilde{p}$. To show that $\mathrm{E}(|V_{p^*,n}|^{1+\epsilon})$ is bounded we use its integral representation, the inequality given in Theorem 2 for J, and the extension of Hölder's inequality.

$$\begin{aligned}\mathrm{E}(|V_{p^*,n}|^{1+\epsilon}) &= \int \cdots \int |J_{p^*}(\mathbf{F})|^{1+\epsilon}\, dC_n(\mathbf{F}) \\ &\le M \int \cdots \int \prod_{j=1}^k h(F_j)^{a_j(1+\epsilon)}\, dC_n(\mathbf{F}) \\ &\le M \prod_{j=1}^k \left\{ \frac{1}{n} \sum_{i=1}^n h\left(\frac{i}{n+1}\right)^{(-1+\delta)(1+\epsilon)} \right\}^{1/p_j} \\ &= \frac{M}{n} \sum_{i=1}^n h\left(\frac{i}{n+1}\right)^{(-1+\delta)(1+\epsilon)} \\ &\le M \int_0^1 \frac{1}{[u(1-u)]^{(1-\delta)(1+\epsilon)}}\, du.\end{aligned}$$

The last integral is finite for $\epsilon < \delta$, which completes the proof of the convergency of $\mathbf{R}_n$ to $\boldsymbol{\mu}$. To prove the existence of the variance we split the elements of our statistic into the following summands:

$$n^{\frac{1}{2}}(R_{p^*,n} - \mu_{p^*}) = H_{p^*,1n} + H_{p^*,2n} + G_{p^*,n},$$

where

$$H_{p^*,1n} = n^{\frac{1}{2}} \int \cdots \int J_{p^*}(\mathbf{F})\, d\{C_n(\hat{\mathbf{F}}) - C(\mathbf{F})\}$$

$$H_{p^*,2n} = n^{\frac{1}{2}} \int \cdots \int [J_{p^*}(\hat{\mathbf{F}}) - J_{p^*}(\mathbf{F})] d\{C_n(\hat{\mathbf{F}}) - C(\mathbf{F})\}$$

$$G_{p^*,n} = \int \cdots \int J_{p^*}(\hat{\mathbf{F}}) - J_{p^*}(\mathbf{F})\, dC(\mathbf{F}) = \sum_{r=1}^k G_{p^*,rn}$$

$$= \sum_{r=1}^k n^{\frac{1}{2}} \int \cdots \int (\hat{F}_r - F_r) \times \frac{\partial J_{p^*}(F_1, \ldots, F_{r-1}, \Phi_r, F_{r+1}, \ldots, F_k)}{\partial \Phi_r}\Bigg|_{\Phi_r = F_r + \triangle(\hat{F}_r - F_r)} dC(\mathbf{F}),$$

where $G_{p^*,n}$ is transformed by the mean value theorem and $\triangle \in [0, 1]$. Each G_{rn} could be rewritten as

$$G_{p^*,rn} = n^{\frac{1}{2}} \int \cdots \int (\hat{F}_r - F_r) \times \left[\frac{\partial J_{p^*}(F_1, \ldots, F_{r-1}, \Phi_r, F_{r+1}, \ldots, F_k)}{\partial \Phi_r}\Bigg|_{\Phi_r = F_r + \triangle(\hat{F}_r - F_r)} - \frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r} \right] dC(\mathbf{F}) + n^{\frac{1}{2}} \int \cdots \int (\hat{F}_r - F_r) \frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}\, dC(\mathbf{F}).$$

While the first derivative of the function $\mathbf{J}$ is uniformly continuous and bounded by the assumptions, and since $|\Phi_r - F_r| \le |\hat{F}_r - F_r|$, by the Lemma of Glivenko–Cantelli $\frac{\partial J_{p^*}(F_1, \ldots, F_{r-1}, \Phi_r, F_{r+1}, \ldots, F_k)}{\partial \Phi_r}\Big|_{\Phi_r = F_r + \triangle(\hat{F}_r - F_r)}$ converges in probability to $\frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}$. Using the definition of the empirical distribution function which for simplicity is multiplied by $\frac{n+1}{n}$, the last summand has following representation:

$$\begin{aligned}&n^{\frac{1}{2}} \int \cdots \int (\hat{F}_r - F_r) \frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}\, dC(\mathbf{F}) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \int \cdots \int (\mathbf{I}\{x_{ri} \le x_r\} - F_r) \frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}\, dC(\mathbf{F}) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n \int \cdots \int [\mathbf{I}\{x_{ri} \le x_r\} - F_r] \frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}\, dC(\mathbf{F}) \\ &= n^{\frac{1}{2}} \int \cdots \int [\mathbf{I}\{y_r \le x_r\} - F_r] \frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}\, dC(\mathbf{F}).\end{aligned}$$

Note that $\frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}$ has no influence on the variance but contributes to the covariance between the elements of the vector. Applying Hölder's inequality and the assumption, that $\frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r} \le M \left[\prod_{j=1, j\ne r}^k h(u_j)^{b_j} \right] h(u_r)^{b_r - 1}$ provides the existence of the second moment of $G_{p^*,n}$:

$$Var(G_{p^*,n}) = Var\left[\sum_{r=1}^k \int \cdots \int \mathbf{I}\{y_r \le x_r\} \frac{\partial J_{p^*}(\mathbf{F})}{\partial F_r}\, dC(\mathbf{F}) \right].$$

In the main difference, $H_{p^*,1n}$ can be rewritten in the following form

$$H_{p^*,1n} = n^{-\frac{1}{2}} \sum_{i=1}^n H_{p^*,1in},$$

where $H_{p^*,1in} = J_{p^*}(F_1(x_{1i}), \ldots, F_k(x_{ki})) - \mu_{p^*}$ are iid with zero means. Using that $J_{p^*}(\mathbf{F}) \le M \prod_{r=1}^k h(u_r)^{a_r}$, the application of the multivariate extension of the Hölder's inequality for the same set of p_r, $r = 1, \ldots, k$ as in the first part of the theorem, shows that $H_{p^*,1in}$ has finite absolute moment of order $2 + \delta_0$, for some $\delta_0 > 0$. By the Glivenko–Cantelli Lemma $H_{p^*,2n}$ converges almost surely to zero and from Bhuchongkul (1964) we get the resulting covariance

$$\begin{aligned}&\mathrm{Cov}[n^{\frac{1}{2}}(\mathbf{R}_n - \boldsymbol{\mu})] \\ &= Cov\Bigg[\mathbf{J}(F_1(y_1), \ldots, F_k(y_k)) \\ &\quad + \sum_{r=1}^k \int \cdots \int \mathbf{I}\{y_r \le x_r\} - F_r(y_r)\Big] \\ &\quad \times \frac{\partial \mathbf{J}(F_1(y_1), \ldots, F_k(y_k))}{\partial F_r(y_r)}\, dC(F_1(y_1), \ldots, F_k(y_k))\Bigg]\end{aligned}$$

Proof of Theorem 3. In this proof we use similar methods as in the proof of Theorem 2.1 of Ruymgaart et al. (1972) and Proposition 2.1 of Genest et al. (1995). Expanding $\frac{\partial \mathcal{L}_j(\boldsymbol{\theta},\mathbf{X})}{\partial \boldsymbol{\theta}}$ for $j = 1,\ldots,p$ in a Taylor series we obtain

$$\left.\frac{\partial \mathcal{L}_j(\boldsymbol{\theta},\mathbf{X})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \approx \mathbf{a}_n - \mathbf{B}_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

where

$$\mathbf{a}_{jn} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial l_j(\hat{F}_1(x_{1i}),\ldots,\hat{F}_k(x_{ki}))}{\partial \boldsymbol{\theta}_j}, \quad \text{for } j = 1,\ldots,p$$

$$\mathbf{a}_n = (\mathbf{a}'_{1n},\ldots,\mathbf{a}'_{pn})'$$

$$\mathbf{B}_n = \{\mathbf{B}_{jrn}\}_{r,j=1,\ldots,p} = \left\{-\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 l_j(\hat{F}_1(x_{1i}),\ldots,\hat{F}_k(x_{ki}))}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_r}\right\}_{r,j=1,\ldots,p},$$

where $\mathbf{B}_n$ is a block matrix. From the theory of rank statistics we conclude that $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \approx n^{\frac{1}{2}}\mathbf{B}_n^{-1}\mathbf{a}_n$. Taking $\mathbf{J}$ in Theorem 2 equal to $\frac{\partial l_j}{\partial \boldsymbol{\theta}_j}$ and $\frac{\partial^2 l_j}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_r}$ where $\widetilde{p}$ will be the number of parameters to be estimated on the current level, we get

$$\mathbf{B}_{rjn} \xrightarrow{a.s.} \boldsymbol{\Upsilon}_{rj} = -\mathrm{E}\left[\frac{\partial^2 l_j(\hat{F}_1(x_1),\ldots,\hat{F}_k(x_k))}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}'_j}\right]$$

and $n^{\frac{1}{2}}\mathbf{a}_{jn} \overset{a}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_j)$, where

$$\boldsymbol{\Sigma}_j = Cov\left[\frac{\partial l_j(\hat{F}_1(x_1),\ldots,\hat{F}_k(x_k))}{\partial \boldsymbol{\theta}_j} + \sum_{r=1}^{k} \int\cdots\int \mathbf{I}\{F_r(x_r) \le u_r\} - u_r \Big]\frac{\partial^2 l_j(\mathbf{u})}{\partial \boldsymbol{\theta}_j \partial u_r}\, dC(\mathbf{u})\right].$$

We assemble the variances $\boldsymbol{\Sigma}_j$ in the block matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Sigma}_p \end{pmatrix},$$

as well as $\mathbf{B} = \{\boldsymbol{\Upsilon}_{jr}\}_{j,r=1,\ldots,p}$. Putting them together we can conclude that

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{a}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{B}^{-1}\boldsymbol{\Sigma}\mathbf{B}^{-1}).$$

References

Aas, K., Czado, C., Frignessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics 44, 182–198.

Acar, E., Genest, C., Nešlehová, J., 2012. Beyond simplified pair-copula constructions. Journal of Multivariate Analysis 110, 74–90.

Bedford, T., Cooke, R.M., 2002. Vines—a new graphical model for dependent random variables. Annals of Statistics 30 (4), 1031–1068.

Bhuchongkul, S., 1964. A class of nonparametric tests for independence in bivariate populations. The Annals of Mathematical Statistics 35, 138–149.

Chen, X., Fan, Y., 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. Journal of Econometrics 135, 125–154.

Chen, X., Fan, Y., 2005. Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. The Canadian Journal of Statistics 33 (2), 389–414.

Chen, X., Fan, Y., Patton, A., 2004. Simple tests for models of dependence between multiple financial time series, with applications to U.S. Equity returns and exchange rates, Discussion paper 483, Financial Markets Group, London School of Economics.

Chen, X., Fan, Y., Tsyrennikov, V., 2006. Efficient estimation of semiparametric multivariate copula models. Journal of the American Statistical Association 101 (475), 1228–1240.

Chen, S.X., Huang, T., 2007. Nonparametric estimation of copula functions for dependence modeling. The Canadian Journal of Statistics 35, 265–282.

Dissmann, J., Brechmann, E., Czado, C., Kurowicka, D., Selecting and estimating regular vine copulae and application to financial returns, arXiv:1202.2002.

Embrechts, P., McNeil, A.J., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. In: Risk Management: Value at Risk and Beyond. Cambridge University Press, pp. 176–223.

Fermanian, J.-D., 2005. Goodness-of-fit tests for copulas. Journal of Multivariate Analysis 95 (1), 119–152.

Fermanian, J.-D., Scaillet, O., 2003. Nonparametric estimation of copulas for time series. Journal of Risk 5, 25–54.

Genest, C., Ghoudi, K., Rivest, L.-P., 1995. A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82, 543–552.

Genest, C., Quessy, J.-F., Rémillard, B., 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. Scandinavian Journal of Statistics 33, 337–366.

Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88, 1034–1043.

Genest, C., Rémillard, B., 2008. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Annales de l'Institut Henri Poincaré. Probabilités et Statistiques 44, 1096–1127.

Giacomini, E., Härdle, W.K., Spokoiny, V., 2009. Inhomogeneous dependence modeling with time-varying copulae. Journal of Business and Economic Statistics 27 (2), 224–234.

Hennessy, D.A., Lapan, H.E., 2002. The use of Archimedean copulas to model portfolio allocations. Mathematical Finance 12, 143–154.

Hájek, J., Šidák, Z., 1967. Theory of Rank Test. Academic Press, New York.

Hofert, J.M., 2012. A stochastic representation and sampling algorithm for nested Archimedean copulas. Journal of Statistical Computation and Simulation 82 (9), 1239–1255.

Hofert, J.M., 2011. Efficiently sampling Archimedean copulas. Computational Statistics & Data Analysis 55, 57–70.

Hofert, M., Pham, D., Densities of nested Archimedean copulas, arXiv:1204.2410v2.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analisys 94, 401–419.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Junker, M., May, A., 2005. Measurement of aggregate risk with copulas. Econometrics Journal 8, 428–454.

Lee, T.-H., Long, X., 2009. Copula-based multivariate GARCH model with uncorrelated dependent errors. Journal of Econometrics 150, 207–218.

Mantegna, R.N., 1999. Hierarchical structure in financial markets. European Physical Journal B 11, 193–197.

McNeil, A.J., 2008. Sampling nested Archimedean copulas. Journal Statistical Computation and Simulation 78, 567–581.

McNeil, A.J., Nešlehová, J., 2009. Multivariate Archimedean copulas, d-monotone functions and l_1 norm symmetric distributions. Annals of Statistics 37 (5b), 3059–3097.

Min, A., Czado, C., 2010. Bayesian inference for multivariate copulas using pair-copula constructions. Journal of Financial Econometrics 8, 511–546.

Nelsen, R.B., 2006. An Introduction to Copulas. Springer, Berlin.

Nelson, D., 1991. Conditional heteroscedasticity in asset returns: a new approach. Econometrica 59, 347–370.

Okhrin, O., Okhrin, Y., Schmid, W., Properties of the hierarchical Archimedean copulas, Statistics & Risk Modelling, forthcoming, (http://dx.doi.org/10.1524/strm.2013.1071).

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

Pitt, M., Chan, D., Kohn, R., 2006. Efficient Bayesian inference for Gaussian copula regression models. Biometrika 93, 537–554.

Ruymgaart, F., Shorak, G., Zwet, W., 1972. Asymptotic normality of nonparametric tests for independence. The Annals of Mathematical Statistics 43, 1122–1135.

Savu, C., Trede, M., 2008. Goodness-of-fit tests for parametric families of Archimedean copulas. Quantitative Finance 8, 109–116.

Savu, C., Trede, M., 2010. Hierarchical Archimedean copulas. Quantitative Finance 10, 295–304.

Schweizer, B., Sklar, A., 1983. Probabilistic Metric Spaces. North Holland, Amsterdam.

Sklar, A., 1959. Fonctions dé repartition á n dimension et leurs marges. Publ. Inst. Stat. Univ. Paris 8, 229–231.

Smith, M., Gan, Q., Kohn, R., 2012. Modeling dependence using skew t copulas: bayesian inference and applications. Journal of Applied Econometrics 27, 500–522.

Smith, M., Min, A., Almeida, C., Czado, C., 2010. Modelling longitudinal data using a pair-copula decomposition of serial dependence. Journal of the American Statistical Association 105, 1467–1479.

Song, P.X.-K., Fan, Y., Kalbfleisch, J.D., 2005. Maximization by parts in likelihood inference. Journal of the American Statistical Association 100, 1145–1167.

Tola, V., Lillo, F., Gallegati, M., Mantegna, R., 2007. Cluster analysis for portfolio optimization. Journal of Economic Dynamics and Control 32, 235–258.

Wang, W., Wells, M., 2000. Model selection and semiparametric inference for bivariate failure-time data. Journal of the American Statistical Association 95, 62–76.

Whelan, N., 2004. Sampling from Archimedean copulas. Quantitative Finance 4, 339–352.

White, H., 1982. Estimation, Inference and Specification Analysis. Cambridge University Press.