

Models for construction of multivariate dependence

Daniel Berg*

University of Oslo and Norwegian Computing Center

Kjersti Aas†

Norwegian Computing Center

Abstract

In this article we review models for construction of higher-dimensional dependence that have arisen recent years. A multivariate data set, which exhibit complex patterns of dependence, particularly in the tails, can be modelled using a cascade of lower-dimensional copulae. We examine two such models that differ in their construction of the dependency structure, namely the nested Archimedean constructions and the pair-copula constructions (also referred to as vines). The constructions are compared, and estimation- and simulation techniques are examined. The fit of the two constructions is tested on two different four-dimensional data sets; precipitation values and equity returns, using a state of the art copula goodness-of-fit procedure. The nested Archimedean construction is strongly rejected for both our data sets, while the pair-copula construction provides an appropriate fit. Through VaR calculations, we show that the latter does not overfit data, but works very well even out-of-sample.

KEY WORDS: Multivariate models, Nested Archimedean copulas, Pair-copula decompositions

*PhD-student University of Oslo and Norwegian Computing Center, daniel@danielberg.no

†Assistant research director, The Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo, Norway, Kjersti.Aas@nr.no (address for correspondence)

Contents

1	Introduction	3
2	Constructions of higher dimensional dependence	3
2.1	The nested Archimedean constructions (NAC's)	3
2.1.1	The exchangeable multivariate Archimedean copula (EAC)	4
2.1.2	The fully nested Archimedean construction (FNAC)	4
2.1.3	The partially nested Archimedean construction (PNAC)	4
2.1.4	The hierarchically nested Archimedean construction (HNAC)	5
2.1.5	Parameter estimation	6
2.1.6	Simulation	7
2.2	The pair-copula constructions (PCC)	7
2.2.1	Parameter estimation	8
2.2.2	Simulation	8
2.3	Comparison	9
3	Applications	10
3.1	Goodness-of-fit	10
3.2	Application 1: Precipitation data	10
3.2.1	Hierarchically nested Archimedean construction	10
3.2.2	Pair-copula construction	12
3.3	Application 2: Equity returns	12
3.3.1	Hierarchically nested Archimedean construction	13
3.3.2	Pair-copula construction	13
3.4	Validation	15
4	Summary and Conclusions	16
A	Parameters for GARCH-NIG model	20

1 Introduction

A copula is a multivariate distribution function with standard uniform marginal distributions. While the literature on copulae is substantial, most of the research is still limited to the bivariate case. Building higher-dimensional copulae is a natural next step, however, this is not an easy task. Apart from the Gaussian and Student copulae, the set of higher-dimensional copulae proposed in the literature is rather limited.

The class of Archimedean copulae (see e.g. Joe (1997) for a review) is a class that has attracted particular interest due to numerous properties which make them simple to analyse. The most common multivariate extension, the exchangeable multivariate Archimedean copula (EAC), is extremely restrictive, allowing the specification of only one distributional parameter, regardless of dimension. There have been some attempts at constructing more flexible multivariate Archimedean copula extensions. As far as we know, all of them were originally proposed by Joe (1993). In this paper we discuss three such extensions. The first two, henceforth denoted the fully nested Archimedean construction (FNAC) and the partially nested Archimedean construction (PNAC), are discussed in Embrechts et al. (2003) and Whelan (2004), while the third, henceforth denoted the hierarchically nested Archimedean construction (HNAC), is discussed in Whelan (2004) and McNeil (2007) and extensively treated in Savu and Trede (2006). As a group, the FNAC, PNAC and HNAC will be denoted nested Archimedean constructions (NAC's) in the remaining of this paper. For the d -dimensional case, all NAC's allow for the modelling of up to $d - 1$ bivariate Archimedean copulae, i.e. the free specification of up to $d - 1$ distributional parameters.

For a d -dimensional problem there are in general $d(d - 1)/2$ pairings of variables. Hence, while the NAC's constitute a huge improvement compared to the EAC, they are still not rich enough to model all possible mutual dependencies amongst the d variates. An even more flexible structure, here denoted the pair-copula construction (PCC) allows for the free specification of $d(d - 1)/2$ copulae. This structure was originally proposed by Joe (1996), and later discussed in detail by Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) (simulation) and Aas et al. (2007) (inference). Similar to the NAC's, the PCC is hierarchical in nature. The modelling scheme is based on a decomposition of a multivariate density into a cascade of bivariate copulae. In contrast to the NAC's, the PCC is not restricted to Archimedean copulae. The bivariate copulae may be from any family and several families may well be mixed in one PCC.

This paper has several contributions. In Section 2 we compare the two ways of constructing higher dimensional dependency structures, the NAC's and the PCC's. We examine properties and estimation- and simulation techniques, focusing on the relative strengths and weaknesses of the different constructions. In Section 3 we apply the HNAC and the PCC to two four-dimensional data sets; precipitation values and equity returns. We examine the goodness-of-fit and validate the PCC out-of-sample with respect to one day value at risk (VaR) for the equity portfolio. Finally, Section 4 provides some summarizing comments and conclusions.

2 Constructions of higher dimensional dependence

2.1 The nested Archimedean constructions (NAC's)

In this section we review three types of NAC's. We describe procedures for estimating parameters for, and simulating from, such constructions. The biggest challenge with the NAC's lies in checking that they lead to valid multivariate copulae. We will not go into details but only indicate the nature of the necessary conditions. For a comprehensive reference we refer the reader to Joe (1997). Before reviewing the three types of NAC's, we give a short description of the EAC, which serves as a baseline.

2.1.1 The exchangeable multivariate Archimedean copula (EAC)

The most common way of defining a multivariate Archimedean copula is the EAC, defined as

$$C(u_1, u_2, \dots, u_d) = \varphi^{-1} \{ \varphi(u_1) + \dots + \varphi(u_d) \}, \quad (1)$$

where the function φ is a decreasing function known as the generator of the copula and φ^{-1} denotes its inverse (see e.g. Nelsen (1999)). We assume that the generator has only one parameter, θ . For $C(u_1, u_2, \dots, u_d)$ to be a valid d -dimensional Archimedean copula, φ^{-1} has to be completely monotonic on $[0, \infty)$, and one usually also assumes that $\varphi(0) = \infty$, i.e. that the Archimedean copula is strict. Consider for example the popular Gumbel-Hougaard (Gumbel, 1960) and Clayton (Clayton, 1978) copulae. The generator functions for these two copulae are given by $\varphi(t) = (-\log(t))^\theta$ and $\varphi(t) = (t^{-\theta} - 1)/\theta$, respectively.

The copula in (1) suffers from a very restricted dependence structure, since all k -dimensional marginal distributions ($k < d$) are identical. For several applications, one would like to have multivariate copulae which allows for more flexibility. In the following sections we review three such extensions; the FNAC, PNAC and HNAC.

2.1.2 The fully nested Archimedean construction (FNAC)

A simple generalization of (1) can be found in Joe (1997) and is also discussed in Embrechts et al. (2003), Whelan (2004), Savu and Trede (2006) and McNeil (2007). The structure, which is shown in Figure 1 for the four-dimensional case, is quite simple, but notationally complex. As seen from the figure, one simply adds a dimension step by step. The nodes u_1 and u_2 are coupled through copula C_{11} , node u_3 is coupled with $C_{11}(u_1, u_2)$ through copula C_{21} , and finally node u_4 is coupled with $C_{21}(u_3, C_{11}(u_1, u_2))$ through copula C_{31} . Hence, the copula for the 4-dimensional case requires three bivariate copulae C_{11} , C_{21} , and C_{31} , with corresponding generators φ_{11} , φ_{21} , and φ_{31} :

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= C_{31}(u_4, C_{21}(u_3, C_{11}(u_1, u_2))) \\ &= \varphi_{31}^{-1} \{ \varphi_{31}(u_4) + \varphi_{31}(\varphi_{21}^{-1} \{ \varphi_{21}(u_3) + \varphi_{21}(\varphi_{11}^{-1} \{ \varphi_{11}(u_1) + \varphi_{11}(u_2) \}) \}) \}. \end{aligned}$$

For the d -dimensional case, the corresponding expression becomes

$$\begin{aligned} C(u_1, \dots, u_d) &= \varphi_{d-1,1}^{-1} \{ \varphi_{d-1,1}(u_d) + \varphi_{d-1,1} \circ \varphi_{d-2,1}^{-1} \{ \varphi_{d-2,1}(u_{d-1}) + \varphi_{d-2,1} \\ &\quad \circ \dots \circ \varphi_{11}^{-1} \{ \varphi_{11}(u_1) + \varphi_{11}(u_2) \} \} \}. \end{aligned} \quad (2)$$

In this structure, which Whelan (2004) refers to as fully nested, all bivariate margins are themselves Archimedean copulae. It allows for the free specification of $d-1$ copulae and corresponding distributional parameters, while the remaining $(d-1)(d-2)/2$ copulae and parameters are implicitly given through the construction. More specifically, in Figure 1, the two pairs (u_1, u_3) and (u_2, u_3) both have copula C_{21} with dependence parameter θ_{21} . Moreover, the three pairs (u_1, u_4) , (u_2, u_4) and (u_3, u_4) all have copula C_{31} with dependence parameter θ_{31} . Hence, when adding variable k to the structure, we specify the relationships between k pairs of variables.

The FNAC is a construction of partial exchangeability and there are some technical conditions that needs to be satisfied for (2) to be a proper d -dimensional copula. The consequence of these conditions for the FNAC is that the degree of dependence, as expressed by the copula parameter, must decrease with the level of nesting, i.e. $\theta_{11} \geq \theta_{21} \geq \dots \geq \theta_{d-1,1}$, in order for the resulting d -dimensional distribution to be a proper copula.

2.1.3 The partially nested Archimedean construction (PNAC)

An alternative multivariate extension is the PNAC. This structure was originally proposed by Joe (1997) and is also discussed in Whelan (2004), McNeil et al. (2006) (where it is denoted partially exchangeable) and McNeil (2007).

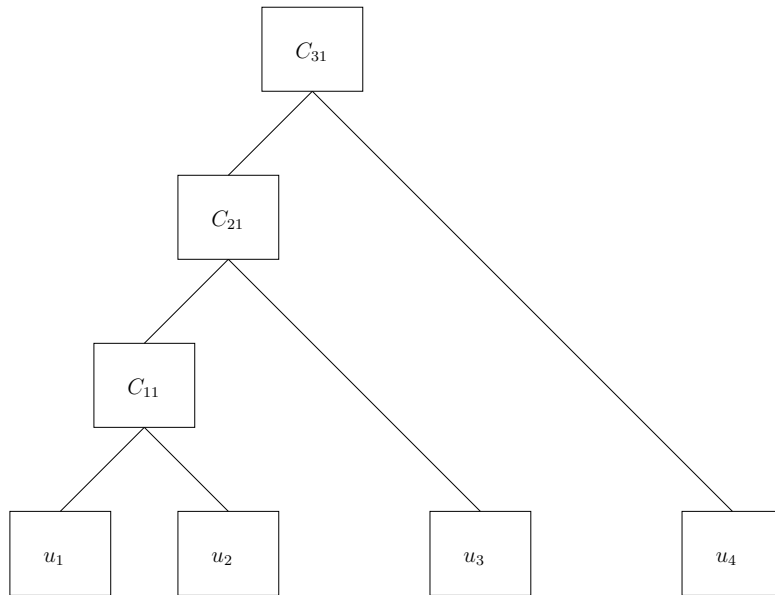


Figure 1: Fully nested Archimedean construction.

The lowest dimension for which there is a distinct structure of this class is four, when we have the following copula:

$$\begin{aligned}
 C(u_1, u_2, u_3, u_4) &= C_{21}(C_{11}(u_1, u_2), C_{21}(u_3, u_4)) \\
 &= \varphi_{21}^{-1}\{\varphi_{21}(\varphi_{11}^{-1}\{\varphi_{11}(u_1) + \varphi_{11}(u_2)\}) + \varphi_{21}(\varphi_{12}^{-1}\{\varphi_{12}(u_3) + \varphi_{12}(u_4)\})\}.
 \end{aligned}
 \tag{3}$$

Figure 2 illustrates this structure graphically. Again the construction is notationally complex although the logic is straightforward. We first couple the two pairs (u_1, u_2) and (u_3, u_4) with copulae C_{11} and C_{12} , having generator functions φ_{11} and φ_{12} , respectively. We then couple these two copulae using a third copula C_{21} . The resulting copula is exchangeable between u_1 and u_2 and also between u_3 and u_4 . Hence, it can be understood as a composite of the EAC and the FNAC.

For the PNAC, as for the FNAC, $d - 1$ copulae and corresponding distributional parameters are freely specified, while the remaining copulae and parameters are implicitly given through the construction. More specifically, in Figure 2, the four pairs (u_1, u_3) , (u_1, u_4) , (u_2, u_3) and (u_2, u_4) will all have copula C_{21} , with dependence parameter θ_{21} . Similar constraints on the parameters are required for the PNAC's as for the FNAC's.

2.1.4 The hierarchically nested Archimedean construction (HNAC)

The HNAC was originally suggested by Joe (1997, Chapter 4.2), and is also mentioned in Whelan (2004). However, Savu and Tiede (2006) were the first to work the idea out in full generality. Among other contributions, they derive the density of HNAC's in general.

The difference between this structure and the PNAC is that a copula at a specific level in the hierarchy does not have to be bivariate. Figure 3 shows an example for the 12-dimensional case (Savu and Tiede, 2006). At level one, there are three copulae. The first, C_{11} , is a three-dimensional EAC joining the variables u_1 - u_3 . The second, C_{12} , is a six-dimensional EAC joining the variables u_4 - u_9 , and the last, C_{13} , is a three-dimensional EAC joining the variables u_{10} - u_{12} . At the second level, the three copulas from the first level are joined by C_{21} , a 3-dimensional EAC.

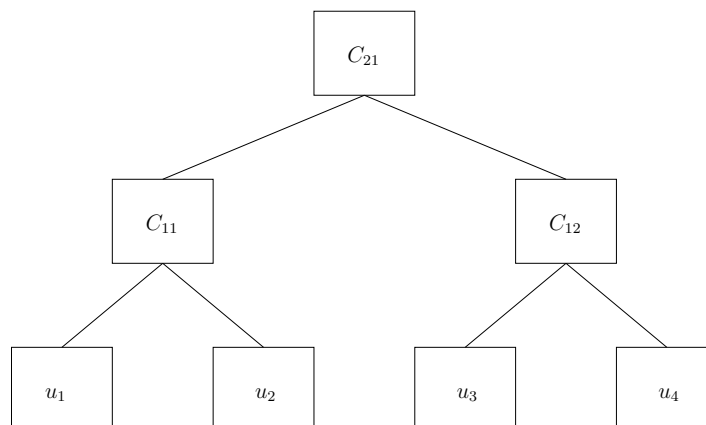


Figure 2: Partially nested Archimedean construction.

The resulting, 12-dimensional copula, can be expressed as

$$\begin{aligned}
 C(u_1, \dots, u_{12}) &= C_{21}(C_{11}(u_1, \dots, u_3), C_{12}(u_4, \dots, u_9), C_{13}(u_{10}, \dots, u_{12})) \\
 &= \phi_{21}^{-1} \left\{ \phi_{21}(\phi_{11}^{-1} \{ \phi_{11}(u_1) + \dots + \phi_{11}(u_3) \}) \right. \\
 &\quad + \phi_{21}(\phi_{12}^{-1} \{ \phi_{12}(u_4) + \dots + \phi_{12}(u_9) \}) \\
 &\quad \left. + \phi_{21}(\phi_{13}^{-1} \{ \phi_{13}(u_{10}) + \dots + \phi_{13}(u_{12}) \}) \right\}.
 \end{aligned}$$

The HNAC $C(u_1, \dots, u_{12})$ is a partially exchangeable copula. All bivariate margins of this 12-dimensional copula that is not directly specified through the copulae at the lower level will have copula C_{21} with dependence parameter θ_{21} .

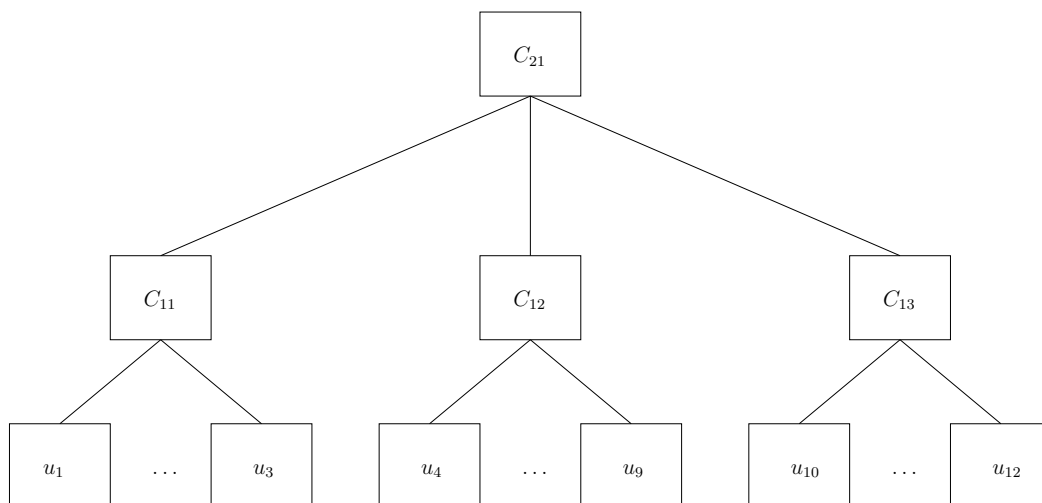


Figure 3: Hierarchically nested Archimedean construction.

2.1.5 Parameter estimation

For all the NAC's, as for the EAC's, the parameters may be estimated by maximum likelihood estimation. However, it is not straightforward to derive the density in general for all parametric families, not even for the EAC. For instance, for the Gumbel-Hougaard family, one has to resort

to a computer algebra system, such as *Mathematica*, or the function D in R, to derive a the d -dimensional density in general.

The expression for the density of one specific 4-dimensional HNAC can be found in Savu and Tiede (2006). The density is obtained using a recursive approach. Hence, the number of computational steps to evaluate the density increases rapidly with the complexity of the copula, and parameter estimation becomes time consuming in high dimensions.

2.1.6 Simulation

Simulating from the higher-dimensional constructions is a very important and central practical task. Simulating from EAC's is usually rather simple, and several algorithms exist. A popular algorithm utilizes the representation of the Archimedean copula generator using Laplace transforms (see e.g. Frees and Valdez (1998)). Simulating from the higher-dimensional NAC's is not that straightforward. Some algorithms have been proposed, of which most include higher-order derivatives of the generator-, inverse generator- or copula functions. These higher-order derivatives are usually extremely complex expressions for high dimensions (see e.g. Savu and Tiede (2006)).

McNeil (2007) shows how to use the Laplace-transform method also for the FNAC, in the case where all generators are taken from either the Gumbel-Hougaard- or the Clayton family. Moreover, he gives a recursive algorithm for sampling from the 4-dimensional PNAC shown in (3). However, this algorithm is not extended to higher dimensions, or to the more general case where any Archimedean copula is allowed at any level. Another problem with the Laplace transform method is that it is limited to copulae for which we can find a distribution that equals the Laplace transform of the inverse generator function and from which we can easily sample. Hence, the Laplace transform method does not in general apply to PNAC's of arbitrary dimensions (McNeil, 2007) or to HNAC's (Savu and Tiede, 2006). For such structures, there is, as of now, no alternative to the conditional inversion method described in e.g. Embrechts et al. (2003). This procedure involves the $d - 1$ first derivatives of the copula function and, in most cases, numerical inversion. Hence, simulation can become very inefficient for high dimensions.

2.2 The pair-copula constructions (PCC)

While the HNAC's constitute a large improvement compared to the EAC's, they still only allow for the modelling of up to $d - 1$ copulae. An even more flexible structure, the PCC, allows for the free specification of $d(d - 1)/2$ copulae. This structure was originally proposed by Joe (1996), and it has later been discussed in detail by Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) (simulation) and Aas et al. (2007) (inference). Similar to the NAC's, the PCC's are hierarchical in nature. The modelling scheme is based on a decomposition of a multivariate density into $d(d - 1)/2$ bivariate copula densities, of which the first $d - 1$ are unconditional, and the rest are conditional.

While the NAC's are defined through their distribution functions, the PCC's are usually represented in terms of the density. Two main types of PCC's have been proposed in the literature; canonical vines and D-vines (Kurowicka and Cooke, 2004). Here, we concentrate on the D-vine representation, for which the density is (Aas et al., 2007)

$$f(x_1, \dots, x_d) = \prod_{k=1}^d f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,i} \{F(x_i|x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})\}. \quad (4)$$

The conditional distribution functions are computed using (Joe, 1996)

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}\{F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})\}}{\partial F(v_j|\mathbf{v}_{-j})}, \quad (5)$$

where $C_{ij|\mathbf{k}}$ is a bivariate copula distribution function. To use this construction to represent a dependency structure through copulas, we assume that the univariate margins are uniform in $[0,1]$.

One 4-dimensional case of (4) is

$$\begin{aligned} c(u_1, u_2, u_3, u_4) &= c_{11}(u_1, u_2) \cdot c_{12}(u_2, u_3) \cdot c_{13}(u_3, u_4) \\ &\cdot c_{21}(F(u_1|u_2), F(u_3|u_2)) \cdot c_{22}(F(u_2|u_3), F(u_4|u_3)) \\ &\cdot c_{31}(F(u_1|u_2, u_3), F(u_4|u_2, u_3)), \end{aligned}$$

where $F(u_1|u_2) = \partial C_{11}(u_1, u_2)/\partial u_2$, $F(u_3|u_2) = \partial C_{12}(u_2, u_3)/\partial u_2$, $F(u_2|u_3) = \partial C_{12}(u_2, u_3)/\partial u_3$, $F(u_4|u_3) = \partial C_{13}(u_3, u_4)/\partial u_3$, $F(u_1|u_2, u_3) = \partial C_{21}(F(u_1|u_2), F(u_3|u_2))/\partial F(u_3|u_2)$ and $F(u_4|u_2, u_3) = \partial C_{22}(F(u_4|u_3), F(u_2|u_3))/\partial F(u_2|u_3)$. Figure 2.2 illustrates this structure.

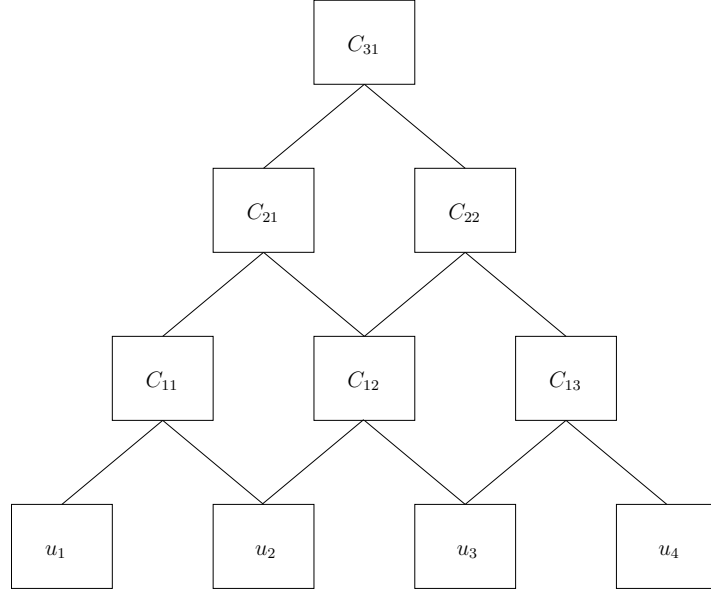


Figure 4: Pair-copula construction.

The copulae involved in (4) do not have to belong to the same family. In contrast to the NAC's they do not even have to belong to the same class. The resulting multivariate distribution will be valid even if we choose, for each pair of variables, the parametric copula that best fits the data. As seen from (4) the PCC consists of $d(d-1)/2$ bivariate copulae of known parametric families, of which $d-1$ are copulae of pairs of the original variables, while the remaining $(d-1)(d-2)/2$ are copulae of pairs of variables constructed using (5) recursively. This means that in contrast to the NAC's, neither the unspecified bivariate margins nor the d -dimensional distribution will belong to a known parametric family in general.

2.2.1 Parameter estimation

The parameters of the PCC may be estimated by maximum likelihood. In contrast to the NAC's, the density is explicitly given in general. However, also for this construction, a recursive approach is used (see Aas et al. (2007, Algorithm 4)). Hence, the number of computational steps to evaluate the density increases rapidly with the complexity of the copula, and parameter estimation becomes time consuming in high dimensions.

2.2.2 Simulation

The simulation algorithm for a D-vine is straightforward and simple to implement, see Aas et al. (2007, Algorithm 2). Like for the HNAC, the conditional inversion method is used. However, to determine each of the conditional distribution functions involved, only the first partial derivative of a bivariate copula needs to be computed (see Aas et al. (2007)). Hence, the simulation procedure for the PCC is in general much simpler and faster than for the NAC's.

Table 1: Summary of construction properties for the EAC, HNAC and PCC constructions.

<i>Construction</i>	<i>Max no. of copulae freely specified</i>	<i>Parameter constraints</i>	<i>Copula class</i>
EAC	1	none	Archimedean
HNAC	$d - 1$	$\theta_{ij} \geq \theta_{kj}, i < k < d, \forall j$	Archimedean
PCC	$d(d - 1)/2$	none	Any

Table 2: Computational times in sec. for different constructions and copulae.

<i>Method</i>	<i>Likelihood</i>	<i>Estimation</i>	<i>Simulation</i>
<i>Gumbel-Hougaard</i>			
HNAC	13.00	1187.08	457.35
PCC	0.20	63.00	4.89
<i>Clayton</i>			
HNAC	1.31	111.94	96.30
PCC	0.04	17.51	0.71

2.3 Comparison

In this section we summarize the differences between the NAC's and the PCC's with respect to ease of interpretation, applicability and computational complexity. Since the FNAC's and PNAC's can be viewed as special cases of the HNAC's, we only concentrate on the latter.

First, the main advantage of the PCC's is the increased flexibility compared to the HNAC's. While the HNAC's only allow for the free specification of $d - 1$ copulae, $d(d - 1)/2$ copulae may be specified in a PCC. Further, the HNAC's have another even more important restriction in that the degree of dependence must decrease with the level of nesting. When looking for appropriate data sets for the applications in Section 3, it turned out to be quite difficult to find real-world data sets satisfying this restriction. Hence, this feature of the HNAC's might prevent them from being extensively using in real-world applications. For the PCC's, on the other hand, one is always guaranteed that all parameter combinations are valid. In addition, the NAC's are constrained to the Archimedean class while the PCC's can be built using copulae from different classes. Table 1 summarizes these properties.

It is our opinion that another advantage of the PCC's is that they are represented in terms of the density and hence easier to handle than the HNAC's that are defined through their distribution functions. The PCC's are also in general more computationally efficient than the HNAC's. Table 2 shows computational times (s) in R ¹ for likelihood evaluation, parameter estimation and simulation for different structures. The parameter estimation is done for the data set described in Section 3.2, and the simulation is performed using the parameters in Table 4 (based on 1000 samples). The values for HNAC were computed using density expressions found in Savu and Trede (2006) but general expressions may also easily be obtained symbolically using the function `D` in R. The estimation times in Table 2 are only indicative and included as examples since they are very dependent on size and structure of the data set. It is more appropriate to study the times needed to compute one evaluation of the likelihood given in the leftmost column. As can be seen from the table, the PCC is superior to both implementations of the HNAC for both likelihood evaluation and simulation.

The multivariate distribution defined through a HNAC will always by definition be an Archimedean copula, and all bivariate margins will belong to the same parametric family. This is not the case for the PCC's, for which neither the multivariate distribution nor the unspecified bivariate margins will belong to a known parametric family in general. However, we not view this as a problem, since both might easily be obtained by simulation.

¹The experiments were run on a Intel(R) Pentium(R) 4 CPU 2.80GHz PC.

3 Applications

The fit of the HNAC and the PCC is assessed for two different four-dimensional data sets; precipitation values and equity returns. Appropriate modelling of precipitation is of great importance to insurance companies which are exposed to a growth in damages to buildings caused by external water exposition. Modelling precipitation and valuing related derivative contracts are also indeed a frontier in the field of weather derivatives. The dependencies within an equity portfolio can have enormous impacts on e.g. capital allocation and the pricing of collateralized debt obligations. Before these two applications are further treated, we describe the test that is used to validate goodness-of-fit in our study.

3.1 Goodness-of-fit

To evaluate whether a construction appropriately fits the data, a GoF test is necessary. We use the following test observator suggested by Genest and Rémillard (2005) and Genest et al. (2007)

$$S = n \int_{[0,1]^d} \left\{ \hat{C}(\mathbf{u}) - C_{\hat{\theta}}(\mathbf{u}) \right\}^2 d\hat{C}(\mathbf{u}) = \sum_{j=1}^n \left\{ \hat{C}(\mathbf{U}_j) - C_{\hat{\theta}}(\mathbf{U}_j) \right\}^2. \quad (6)$$

Here, $C_{\hat{\theta}}$ is the null hypothesis copula with a parameter estimate $\hat{\theta}$ and \hat{C} is the empirical copula distribution, defined by

$$\hat{C}(\mathbf{u}) = \frac{1}{n+1} \sum_{j=1}^n \mathbf{1}(U_{j1} \leq u_1, \dots, U_{jd} \leq u_d), \quad \mathbf{u} = (u_1, \dots, u_d) \in (0, 1)^d. \quad (7)$$

$\mathbf{U}_j = (U_{j1}, \dots, U_{jd})$ are the $U(0, 1)^d$ pseudo-observations, defined as normalized ranks.

Large values of the statistic S lead to the rejection of the null hypothesis copula. In practice, the limiting distribution of S depends on θ . Hence, approximate p -values for the test must be obtained through a parametric bootstrap procedure. We adopt the procedure in Appendix A in Genest et al. (2007), setting the bootstrap parameters m and N to 5000 and 2000, respectively. The validity of this bootstrap procedure is established in Genest and Rémillard (2005).

3.2 Application 1: Precipitation data

In this section we study daily precipitation data (mm) for the period 01.01.1990 to 31.12.2006 for 4 meteorological stations in Norway; Vestby, Ski, Nannestad and Hurdal, obtained from the Norwegian Meteorological Institute. Before further processing, we remove days with non-zero precipitation values for at least one station, resulting in 2065 observations for each variable. Figures 5-6 show the daily precipitation values and corresponding copulae for pairs of meteorological stations. Since we are mainly interested in estimating the dependence structure of the stations, the precipitation vectors are converted to uniform pseudo-observations before further modelling. In light of recent results due to Chen and Fan (2006), the method of maximum pseudo-likelihood is consistent even when time series models are fitted to the margins.

The Kendall's tau values for the different pairs of variables are shown in Table 3. They confirm the intuition that the magnitudes of the Kendall's tau values correspond to the distances between the stations. Ski and Vestby are closely located and so is Hurdal and Nannestad, while the distance from Ski/Vestby to Hurdal/Nannestad is larger.

3.2.1 Hierarchically nested Archimedean construction

From the appearance of the pseudo-observations in Figure 6, a PNAC like the one shown in Figure 2 seems like an appropriate choice, with C_{11} , C_{12} and C_{21} all from the Gumbel-Hougaard family. Here C_{11} is the copula of Vestby and Ski, C_{12} is the copula of Nannestad and Hurdal, and C_{21} is

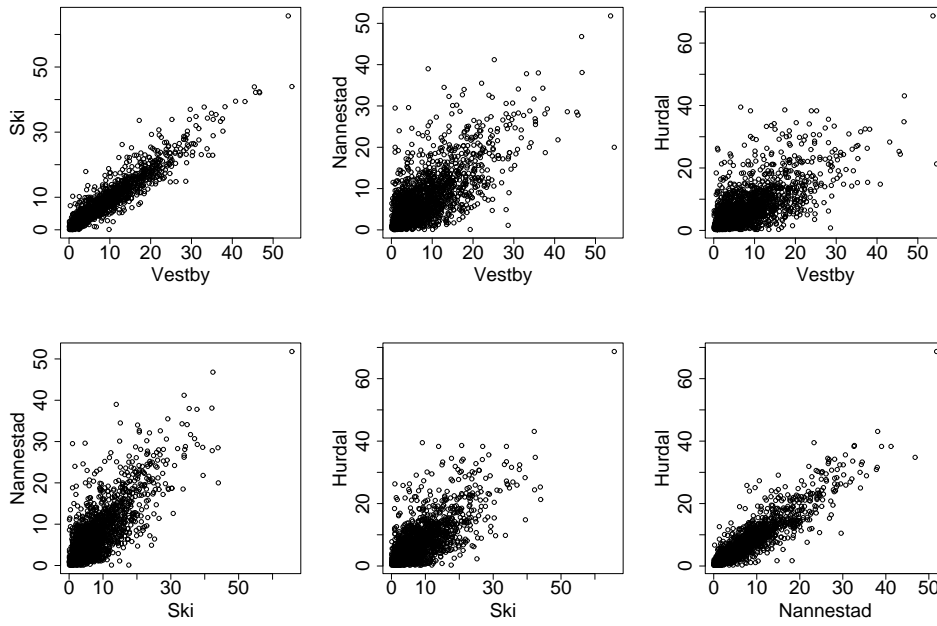


Figure 5: Daily precipitation (mm) for pairs of meteorological stations for the period 01.01.1990 to 31.12.2006, zeros removed.

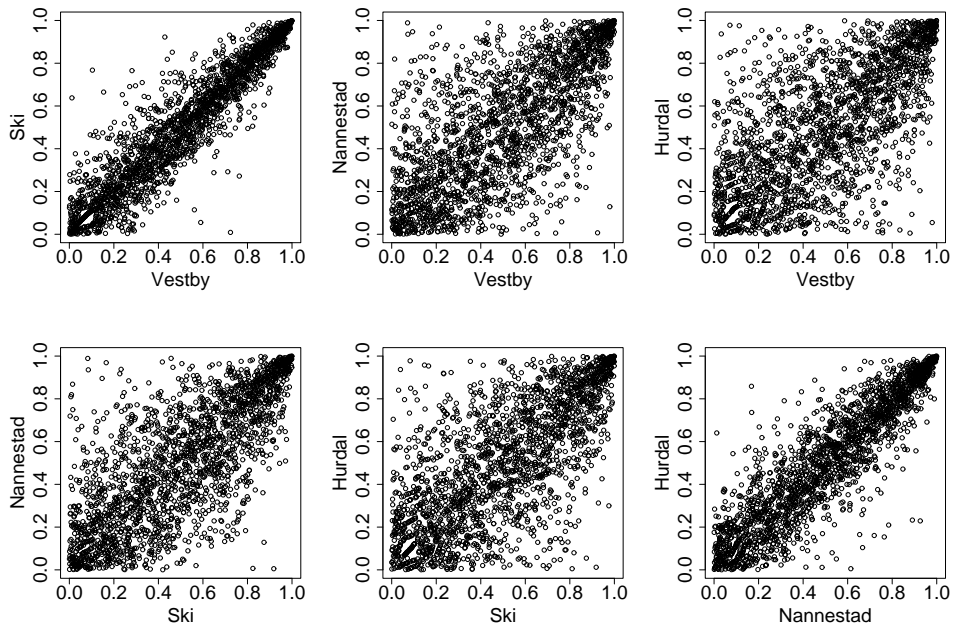


Figure 6: Pseudo-observations corresponding to Figure 5.

Table 3: Estimated Kendall's tau for pairs of variables.

<i>Between</i>	Ski	Nannestad	Hurdal
Vestby	0.79	0.52	0.49
Ski		0.57	0.52
Nannestad			0.73

Table 4: Estimated parameters of four-dimensional HNAC and PCC (D-vine) for the precipitation data.

<i>Param</i>	<i>HNAC estimate</i>	<i>PCC estimate</i>
δ_{11}	4.32	4.34
δ_{12}	3.45	2.24
δ_{13}	-	3.45
δ_{21}	1.97	1.01
δ_{22}	-	1.02
δ_{31}	-	1.03
Log-likelihood	4741.05	4842.25

the copula of C_{11} and C_{12} as well as of the remaining pairs. Table 4 shows the estimated parameter values and the resulting log-likelihood.

While Figure 6 may indicate that the HNAC is appropriate, the goodness-of-fit test strongly rejects the HNAC for the precipitation data, with an estimated p-value of 0.000. Hence, we can conclude that the Gumbel-Hougaard HNAC does not fit the precipitation data appropriately.

3.2.2 Pair-copula construction

We also fitted a four-dimensional PCC (D-vine) like the one shown in Figure 4, with Gumbel-Hougaard copulae for all pairs. The variables are ordered such that the three copulae fitted at level 1 in the pair-copula decomposition are those corresponding to the three largest tail dependence coefficients. Since there is a one-to-one correspondence between the tail dependence coefficient, λ , and Kendall's tau, τ , for the Gumbel-Hougaard copula, namely $\lambda = 2 - 2^{1-\tau}$, the largest tail dependence coefficients correspond to the largest Kendall's tau values. Hence, C_{11} is the copula of Vestby and Ski, C_{12} is the copula of Ski and Nannestad, and C_{13} is the copula of Ski and Hurdal. The parameters of the PCC are estimated using Algorithm 4 in Aas et al. (2007). Table 4 shows the estimated parameter values and the resulting log-likelihood.

The goodness-of-fit test does not reject the Gumbel-Hougaard PCC. The estimated p-value is 0.1635, i.e. the PCC fits the precipitation data reasonably well.

3.3 Application 2: Equity returns

In this section, we study an equity portfolio. The portfolio is comprised of four time series of daily log-return data from the period 14.08.2003 to 29.12.2006 (852 observations for each firm). The data set was downloaded from <http://finance.yahoo.com>. The firms are British Petroleum (BP), Exxon Mobile Corp (XOM), Deutsche Telekom AG (DT) and France Telecom (FTE). Financial log-returns are usually not independent over time. Hence, the original vectors of log-returns are processed by a GARCH filter before further modelling. We use the GARCH(1,1)-model (Bollerslev, 1986):

$$\begin{aligned}
 r_t &= c + \sigma_t z_t \\
 E[z_t] &= 0 \text{ and } \text{Var}[z_t] = 1 \\
 \sigma_t^2 &= a_0 + a \epsilon_{t-1}^2 + b \sigma_{t-1}^2.
 \end{aligned} \tag{8}$$

It is well recognised that GARCH models, coupled with the assumption of conditionally normally distributed errors are unable to fully account for the tails of the distributions of daily returns. Hence, we follow Venter and de Jongh (2002) and use the normal inverse gaussian (NIG) distribution (Barndorff-Nielsen, 1997) as the conditional distribution. In a study performed by Venter and de Jongh (2004) the NIG distribution outperforms a skewed Student's t-distribution and a non-parametric kernel approximation as the conditional distribution of a one-dimensional GARCH process. After filtering the original returns with (8) (estimated parameter values are shown in

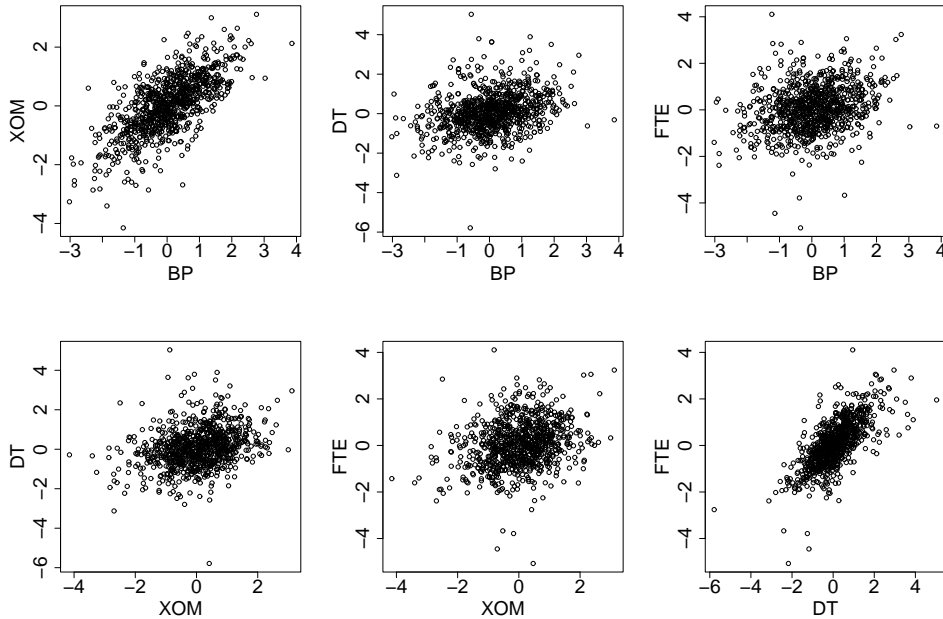


Figure 7: GARCH-filtered daily log-returns for our four stocks for the period from 14.08.2003 to 29.12.2006.

Table 5: Estimated Kendall's tau for pairs of variables for our four stocks.

<i>Firm</i>	XOM	DT	FTE
BP	0.491	0.210	0.182
XOM		0.207	0.172
DT			0.530

Appendix A), the standardised residual vectors are converted to uniform pseudo-observations. Figures 6-8 show the filtered daily log-returns and pseudo-observations for each pair of assets.

The Kendall's tau values for the different pairs of filtered variables in Table 5 show that the stocks within one industrial sector are more dependent than stocks from different sectors.

3.3.1 Hierarchically nested Archimedean construction

Based on previous studies (Aas et al., 2007), we believe that the most appropriate dependence structure of a pair of equities is a Student copula. However, since the HNACs are constrained to the Archimedean class we have to resort to the best alternative from this class, which we believe is the Gumbel-Hougaard copula. Hence, we use a PNAC with C_{11} , C_{12} and C_{21} all Gumbel-Hougaard, where C_{11} is the copula of BP and XOM , C_{12} is the copula of DT and FTE , and C_{21} is the copula of the remaining pairs. Table 6 shows the estimated parameter values and the resulting log-likelihood for the PNAC.

The goodness-of-fit test strongly rejects the Gumbel-Hougaard PNAC also for the equity data, again with an estimated p-value of 0.000. So we conclude that a Gumbel-Hougaard HNAC does not provide an appropriate fit to the equity data.

3.3.2 Pair-copula construction

As stated in the previous section, we believe that the most appropriate dependence structure of a pair of equities is a Student copula. Hence, here we have fitted a Student PCC (D-vine) to the

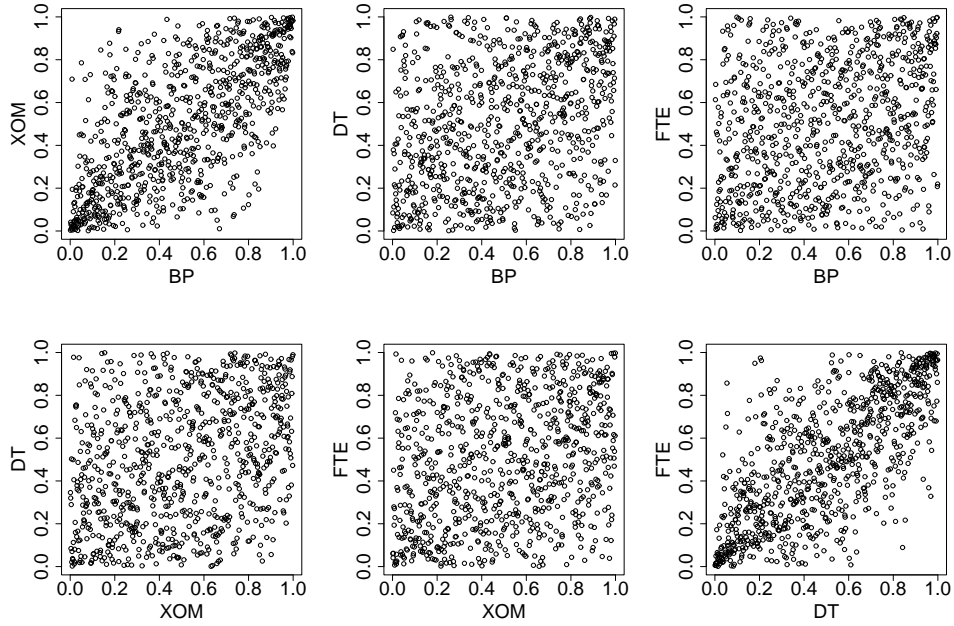


Figure 8: Pseudo-observations corresponding to Figure 7.

Table 6: Estimated parameters of a four-dimensional Gumbel-Hougaard HNAC, fitted to the filtered equity portfolio.

<i>Parameter</i>	<i>Estimate</i>
δ_{11}	1.854
δ_{12}	2.004
δ_{21}	1.223
Log-likelihood	610.685

Table 7: Pairwise tail dependence coefficients of the Student copula for the filtered equity portfolio.

<i>Firm</i>	XOM	DT	FTE
BP	0.125	0.007	0.022
XOM		0.000	0.001
DT			0.306

Table 8: Estimated parameters of Student PCC for the equity portfolio.

<i>Param</i>	<i>Estimate</i>
ρ_{11}	0.70
ν_{11}	13.76
ρ_{12}	0.32
ν_{12}	300
ρ_{13}	0.73
ν_{13}	6.45
ρ_{21}	0.14
ν_{21}	13.42
ρ_{22}	0.06
ν_{22}	23.43
ρ_{31}	0.07
ν_{31}	20.44
Log-likelihood	668.50

four equities. The most appropriate ordering of the variates in the decomposition is done by first fitting a bivariate Student copula to each pair of variates, obtaining estimated degrees of freedom for each pair. For this we use the two-step maximum likelihood method described in broad terms by Oakes (1994) and later formalised and studied by Genest et al. (1995) and Shih and Louis (1995).

Having fitted a bivariate Student copula to each pair, the variates are ordered such that the three copulae fitted at level 1 in the PCC are those corresponding to the three largest tail dependence coefficients. For the Student copula, the tail dependence coefficient is given as by (Embrechts et al., 2001)

$$\lambda(X, Y) = 2 t_{\nu+1} \left(-\sqrt{\nu+1} \sqrt{\frac{1-\rho}{1+\rho}} \right),$$

where $t_{\nu+1}$ denotes the distribution function of a univariate Student's t-distribution with $\nu + 1$ degrees of freedom. The tail dependence coefficients from our case are shown in Table 7. Based on this table, we choose C_{11} as the copula of *BP* and *XOM*, C_{12} as the copula of *XOM* and *DT*, and C_{13} as the copula of *DT* and *FTE*. The parameters of the PCC are estimated by maximum likelihood, see Algorithm 4 in Aas et al. (2007). Table 8 shows the estimated parameter values and the resulting log-likelihood.

The goodness-of-fit test does not reject the Student PCC for the equity data. The estimated p-value is 0.3142, i.e. the Student PCC fits the data well. For the sake of comparison, we have also fitted a Gumbel-Hougaard PCC to this data set. In contrast to the Gumbel-Hougaard HNAC it was not rejected at a 5% significance level, but the p-value was 0.0885, hence significantly lower than for the Student PCC and rejected at a 10% significance level.

3.4 Validation

With the increasing complexity of models there is always the risk of overfitting the data. To examine whether this is the case, we validate the PCC out-of-sample for the equity portfolio. More specifically, we use the PCC GARCH model described in Section 3.3.2 to determine the risk of the return distribution for an equally weighted portfolio of *BP*, *XOM*, *DT*, and *FTE* over a

Table 9: Number of violations of VaR, expected number of violations and p-values for the Kupiec test.

	0.005	0.01	0.05
<i>Observed</i>	2	2	9
<i>Expected</i>	0.55	1.1	5.5
<i>P-value</i>	0.13	0.44	0.16

one-day horizon. The equally-weighted portfolio is only meant as an example. In practice, the weights will fluctuate unless the portfolio is rebalanced every day.

The model estimated from the period 14.03.2003 to 29.12.2006 is used to forecast 1-day VaR at different confidence levels for each day in the period from 30.12.2006 to 11.06.2007 (110 days). The test procedure is as follows: For each day t in the test set:

1. For each variabel $j = 1, \dots, 4$, compute the one-step ahead forecast of $\sigma_{j,t}$, given information up to time t .
2. For each simulation $n = 1, \dots, 10,000$
 - Generate a sample u_1, \dots, u_4 from the estimated Student PCC.
 - Convert u_1, \dots, u_4 to NIG(0,1)-distributed samples z_1, \dots, z_4 using the inverses of the corresponding NIG distribution functions.
 - For each variabel $j = 1, \dots, 4$, determine the log-return $r_{j,t} = c_{j,t} + \sigma_{j,t} z_j$. (Here $c_{j,t}$ is computed as the mean of the last 100 observed log-returns.)
 - Compute the return of the portfolio as $r_{p,t} = \sum_{j=1}^4 \frac{1}{4} r_{j,t}$.
3. For confidence levels $q \in \{0.005, 0.01, 0.05\}$
 - Compute the 1-day VaR_t^q as the q th quantile of the distribution of $r_{p,t}$.
 - If VaR_t^q is greater than the observed value of $r_{p,t}$ this day, a violation is said to occur.

Figure 9 shows the actual log-returns for the portfolio in the period 30.12.2006 to 11.06.2007 and the corresponding VaR levels obtained from the procedure described above. Further, the two upper rows of Table 9 gives the number of violations x , of VaR for each confidence level and with the expected values, respectively. To test the significance of the differences between the observed and the expected values, we use the likelihood ratio statistic by Kupiec (1995). The null hypothesis is that the expected proportion of violations is equal to α . Under the null hypothesis, the likelihood ratio statistic given by

$$2\ln \left(\left(\frac{x}{N} \right)^x \left(1 - \frac{x}{N} \right)^{N-x} \right) - 2\ln \left(\alpha^x (1 - \alpha)^{N-x} \right),$$

where N is the length of the sample, is asymptotically distributed as $\chi^2(1)$. We have computed p-values of the null hypothesis for each quantile. The results are shown in the lower row of Table 9. If we use a 5% level for the Kupiec LR statistic, the null hypothesis is not rejected for any of the three quantiles. Hence, the GARCH-NIG Student PCC model seems to work very well out-of-sample.

4 Summary and Conclusions

In this paper we have reviewed two classes of structures for construction of higher-dimensional dependence; the nested Archimedean constructions (NAC's) and the pair-copula constructions (PCC's). For both structures a multivariate data set is modelled using a cascade of lower-dimensional copulae. They differ however in their construction of the dependence structure, the

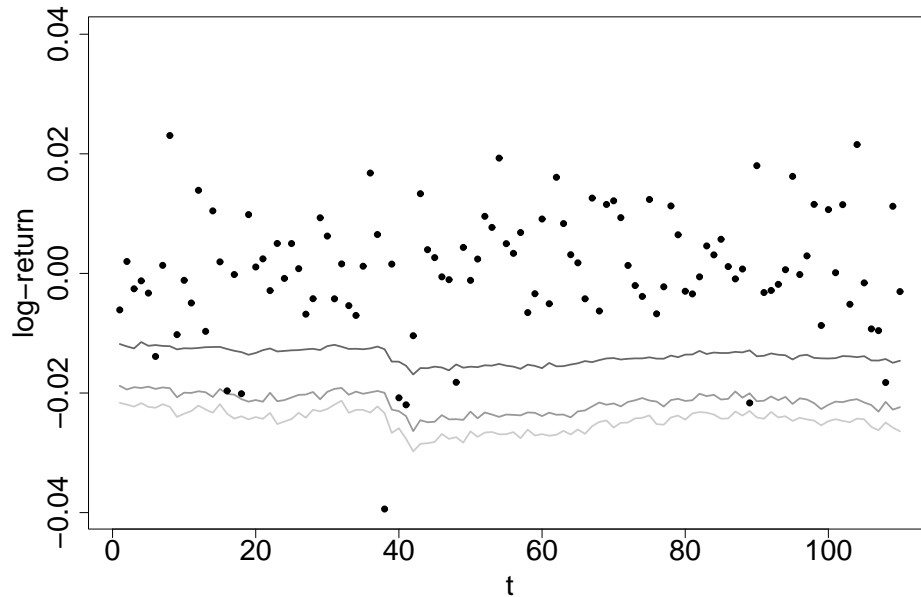


Figure 9: Log-returns for the equity portfolio for the period 30.12.2006 - 11.06.2007 along with 0.5%, 1%, 5% VaR simulated from the estimated Student PCC-GARCH-NIG model.

PCC being more flexible in that it allows for the free specification of $d(d-1)/2$ copulae, while the NAC's only allow for $d-1$.

Simulation and estimation techniques for the two structures have been examined, and we have shown that the PCC's are in general more computationally efficient than the NAC's. The fit of the two constructions has been tested on two different four-dimensional data sets; precipitation values and equity returns using a state of the art copula goodness-of-fit procedure. The NAC is strongly rejected for both our data sets, while the PCC provides an appropriate fit. We believe there are two main reasons for this. First, the NAC's have an important restriction in that the degree of dependence must decrease with the level of nesting. In addition, the NAC's are restricted to the Archimedean class, while the PCC's can be built using copulas from any class. These restrictions of the NAC's might prevent them from being used extensively in real-world applications. As far as we are concerned, the only potential disadvantage of the PCC's compared to the NAC's is that neither the unspecified bivariate margins nor the multivariate distribution will belong to a known parametric family in general. However, we do not view this as a big problem since these distributions might easily be obtained by simulation.

Finally, through VaR calculations for the equity returns, we have shown that the PCC does not overfit data, but works very well even out-of-sample.

Acknowledgements

Daniel Bergs work is supported by the Norwegian Research Council, grant number 154079/420 and Kjersti Aas' part is sponsored by the Norwegian fund Finansmarkedsfondet. We are very grateful to Cornelia Savu, Institute for Econometrics, University of Münster, Germany, for providing us with her code for the NAC's along with helpful comments. In addition we would like to express our deep gratitude, for assistance on the GARCH-NIG filtration, to Professor J.H. Venter, Centre of Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa.

References

- Aas, K., C. Czado, A. Frigessi, and H. Bakken (2007). Pair-copula constructions of multiple dependence. *To appear in Insurance: Mathematics and Economics*.
- Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24, 1–13.
- Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* 32, 245–268.
- Bedford, T. and R. M. Cooke (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics* 30, 1031–1068.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Chen, X. and Y. Fan (2006). Estimation and model selection of semi-parametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* 135, 125–154.
- Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65, 141–151.
- Embrechts, P., F. Lindskog, and A. McNeil (2003). Modelling dependence with copulas and applications to risk management. In S.T.Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*. North-Holland: Elsevier.
- Embrechts, P., A. J. McNeil, and D. Straumann (2001). Correlation and dependency in risk management: Properties and pitfalls. In *Value at Risk and Beyond*. Cambridge University Press.
- Frees, E. W. and E. A. Valdez (1998). Understanding relationships using copulas. *North American Actuarial Journal* 2(1), 1–25.
- Genest, C., K. Ghoudi, and L.-P. Rivest (1995). A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82, 543–552.
- Genest, C. and B. Rémillard (2005). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Technical Report G-2005-51, GERAD, Montreal, Canada.
- Genest, C., B. Rémillard, and D. Beaudoin (2007). Omnibus goodness-of-fit tests for copulas: A review and a power study. Submitted for publication.
- Gumbel, E. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris* 9, 171–173.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis* 46, 262–282.
- Joe, H. (1996). Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In L. Rüschendorf and B. Schweizer and M. D. Taylor (Ed.), *Distributions with Fixed Marginals and Related Topics*.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of derivatives* 2, 173–184.

- Kurowicka, D. and R. M. Cooke (2004). Distribution - free continuous bayesian belief nets. In *Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice, Santa Fe, New Mexico*.
- Kurowicka, D. and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. New York: Wiley.
- McNeil, A. J. (2007). Sampling nested archimedean copulas. *To appear in Journal of Statistical Computation and Simulation*.
- McNeil, A. J., R. Frey, and P. Embrechts (2006). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
- Nelsen, R. (1999). *An Introduction to Copulas*. New York: Springer.
- Oakes, D. (1994). Multivariate survival distributions. *Journal of Nonparametric Statistics* 3, 343–354.
- Savu, C. and M. Tiede (2006). Hierarchical Archimedean copulas. In *International Conference on High Frequency Finance, Konstanz, Germany, May*.
- Shih, J. H. and T. A. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* 51, 1384–1399.
- Venter, J. H. and P. J. de Jongh (2002). Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk* 4(2), 1–24.
- Venter, J. H. and P. J. de Jongh (2004). Selecting an innovation distribution for garch modelsto improve efficiency of risk and volatility estimation. *Journal of Risk* 6(3), 27–53.
- Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance* 4, 339–352.

Table 10: Estimated GARCH and NIG parameters for our four stocks.

<i>Parameter</i>	<i>BP</i>	<i>XOM</i>	<i>DT</i>	<i>FTE</i>
a_0	1.598e-06	1.400e-06	1.801e-06	1.231e-06
a	0.010	0.023	0.025	0.028
b	0.978	0.968	0.963	0.966
β	-0.357	-0.577	0.105	0.037
ψ	3.686	2.293	1.173	1.670

A Parameters for GARCH-NIG model

Table 10 shows the estimated parameters for the GARCH-NIG model used in Section 3.3. For further details of the estimation procedure see Venter and de Jongh (2002).