Available online at www.sciencedirect.com
ScienceDirect
Insurance: Mathematics and Economics 44 (2009) 182–198
INSURANCE MATHEMATICS & ECONOMICS
www.elsevier.com/locate/ime

Pair-copula constructions of multiple dependence

Kjersti Aas[a,*], Claudia Czado[b], Arnoldo Frigessi[c], Henrik Bakken[d]

[a] *The Norwegian Computing Centre, P.O. Box 114, Blindern, N-0314 Oslo, Norway*
[b] *Technische Universität, München, Germany*
[c] *Centre for Statistics for Innovation, University of Oslo and the Norwegian Computing Centre, Oslo, Norway*
[d] *The Norwegian University of Science and Technology, Trondheim, Norway*

Received February 2007; accepted 6 February 2007

Abstract

Building on the work of Bedford, Cooke and Joe, we show how multivariate data, which exhibit complex patterns of dependence in the tails, can be modelled using a cascade of pair-copulae, acting on two variables at a time. We use the pair-copula decomposition of a general multivariate distribution and propose a method for performing inference. The model construction is hierarchical in nature, the various levels corresponding to the incorporation of more variables in the conditioning sets, using pair-copulae as simple building blocks. Pair-copula decomposed models also represent a very flexible way to construct higher-dimensional copulae. We apply the methodology to a financial data set. Our approach represents the first step towards the development of an unsupervised algorithm that explores the space of possible pair-copula models, that also can be applied to huge data sets automatically.

Keywords: Pair-copulae; Vines; Conditional distribution; Decomposition; Multivariate distribution

1. Introduction

Inspired by the work of Joe (1996), Bedford and Cooke (2001b, 2002), and Kurowicka and Cooke (2006), we show how multivariate data can be modelled using a cascade of simple building blocks called *pair-copulae*. This probabilistic construction represents a radically new way of constructing complex multivariate highly dependent models, which parallels classical hierarchical modelling (Green et al., 2003). There, the principle is to model dependency using simple local building blocks based on conditional independence, e.g., cliques in random fields. Here, the building blocks are pair-copulae. The modelling scheme is based on a decomposition of a multivariate density into a cascade of pair copulae, applied on original variables and on their conditional and unconditional distribution functions.

In this paper, we show that the pair-copula decomposition of Bedford and Cooke (2002) can be a simple and powerful tool for model building. While it maintains the logic of building complexity using simple elementary bricks, it does not require conditional independence assumptions when these are not natural. We present some of the theory of Bedford and Cooke (2001b, 2002) from a practical point of view, as a general modelling approach, concentrating on likelihood-based inference based on n variables repeatedly observed, say over time.

Kurowicka and Cooke (2006) approach model inference using partial correlations and the determinant of the correlation matrix as a measure of linear dependence. As an alternative, we propose to rely on a maximum pseudo-likelihood approach for parameter estimation of the pair-copula decomposition. An algorithm is given for evaluating the pseudo-likelihood efficiently based on any combination of pair-copulae. This pseudo-likelihood is based on the ranks of the observations. We illustrate this approach for a four-dimensional financial data set for bivariate Student and/or Clayton copulae as building blocks.

Building higher-dimensional copulae is generally recognised as a difficult problem. There are a huge number of parametric bivariate copulas, but the set of higher-dimensional copulae is rather limited. There have been some attempts to construct multivariate extensions of Archimedean bivariate copu-

* Corresponding author. Tel.: +47 22 85 25 00; fax: +47 22 69 76 60.
E-mail address: Kjersti.Aas@nr.no (K. Aas).

0167-6687/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2007.02.001

lae; see, e.g., Bandeen-Roche and Liang (1996), Joe (1997), Embrechts et al. (2003), Whelan (2004), Savu and Trede (2006) and McNeil (in press). Meta-elliptical copulae (Fang et al., 2002) also offer some flexibility for multivariate modelling. However, it is our opinion that the pair-copula decomposition treated in this paper represents a more flexible and intuitive way of extending bivariate copulae to higher dimensions.

The paper is organised as follows. In Section 2 we introduce the pair-copula decomposition of a general multivariate distribution and illustrate this with some simple examples. In Section 3 we see the effect of the conditional independence assumption on the pair-copula construction. Section 4 describes how to simulate from pair-copula decomposed models. In Section 5 we describe our estimation procedure, while in Section 6 we discuss aspects of the model selection process. In Section 7 we apply the methodology and discuss limitations and difficulties in the context of a financial data set. Finally, Section 8 contains some concluding remarks.

2. A pair-copula decomposition of a general multivariate distribution

Consider a vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ of random variables with a joint density function $f(x_1, \ldots, x_n)$. This density can be factorised as

$$f(x_1, \ldots, x_n) = f_n(x_n) \cdot f(x_{n-1}|x_n) \cdot f(x_{n-2}|x_{n-1}, x_n) \cdots f(x_1|x_2, \ldots, x_n), \quad (1)$$

and this decomposition is unique up to a re-labelling of the variables.

In a sense every joint distribution function implicitly contains both a description of the marginal behaviour of individual variables and a description of their dependency structure. Copulae provide a way of isolating the description of their dependency structure. A copula is a multivariate distribution, C, with uniformly distributed marginals $U(0, 1)$ on $[0, 1]$. Sklar's theorem (Sklar, 1959) states that every multivariate distribution F with marginals $F_1(x_1), \ldots, F_n(x_n)$ can be written as

$$F(x_1, \ldots, x_n) = C\{F_1(x_1), \ldots, F_n(x_n)\}, \quad (2)$$

for some appropriate n-dimensional copula C. In fact, the copula from (2) has the expression

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

where the $F_i^{-1}(u_i)$'s are the inverse distribution functions of the marginals.

Passing to the joint density function f, for an absolutely continuous F with strictly increasing, continuous marginal densities $F_1, \ldots, F_n$ using the chain rule we have

$$f(x_1, \ldots, x_n) = c_{1\cdots n}\{F_1(x_1), \ldots, F_n(x_n)\} \cdot f_1(x_1) \cdots f_n(x_n) \quad (3)$$

for some (uniquely identified) n-variate copula density $c_{1\cdots n}(\cdot)$. In the bivariate case (3) simplifies to

$$f(x_1, x_2) = c_{12}\{F_1(x_1), F_2(x_2)\} \cdot f_1(x_1) \cdot f_2(x_2),$$

where $c_{12}(\cdot, \cdot)$ is the appropriate *pair-copula density* for the pair of transformed variables $F_1(x_1)$ and $F_2(x_2)$. For a conditional density it easily follows that

$$f(x_1|x_2) = c_{12}\{F_1(x_1), F_2(x_2)\} \cdot f_1(x_1),$$

for the same pair-copula. For example, the second factor, $f(x_{n-1}|x_n)$, in the right-hand side of (1) can be decomposed into the pair-copula $c_{(n-1)n}\{F_{n-1}(x_{n-1}), F_n(x_n)\}$ and a marginal density $f_{n-1}(x_{n-1})$. For three random variables X_1, X_2 and X_3 we have that

$$f(x_1|x_2, x_3) = c_{12|3}\{F(x_1|x_3), F(x_2|x_3)\} \cdot f(x_1|x_3), \quad (4)$$

for the appropriate pair-copula $c_{12|3}$, applied to the transformed variables $F(x_1|x_3)$ and $F(x_2|x_3)$. An alternative decomposition is

$$f(x_1|x_2, x_3) = c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot f(x_1|x_2), \quad (5)$$

where $c_{13|2}$ is different from the pair-copula in (4). Decomposing $f(x_1|x_2)$ in (5) further, leads to

$$f(x_1|x_2, x_3) = c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot c_{12}\{F(x_1), F(x_2)\} \cdot f_1(x_1),$$

where two pair-copulae are present.

It is now clear that each term in (1) can be decomposed into the appropriate pair-copula times a conditional marginal density, using the general formula

$$f(x|\boldsymbol{v}) = c_{xv_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\} \cdot f(x|\boldsymbol{v}_{-j}),$$

for a d-dimensional vector $\boldsymbol{v}$. Here v_j is one arbitrarily chosen component of $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ denotes the v-vector, excluding this component. In conclusion, under appropriate regularity conditions, a multivariate density can be expressed as a product of pair-copulae, acting on several different conditional probability distributions. It is also clear that the construction is iterative by nature, and that given a specific factorisation, there are still many different re-parametrisations.

The pair-copula construction involves marginal conditional distributions of the form $F(x|\boldsymbol{v})$. Joe (1996) showed that, for every j,

$$F(x|\boldsymbol{v}) = \frac{\partial\, C_{x,v_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\}}{\partial F(v_j|\boldsymbol{v}_{-j})}, \quad (6)$$

where $C_{ij|\boldsymbol{k}}$ is a bivariate copula distribution function. For the special case where v is univariate, we have

$$F(x|v) = \frac{\partial\, C_{xv}\{F(x), F(v)\}}{\partial F(v)}.$$

In Sections 4–6 we will use the function $h(x, v, \Theta)$ to represent this conditional distribution function when x and v are uniform, i.e., $f(x) = f(v) = 1$, $F(x) = x$ and $F(v) = v$. That is,

$$h(x, v, \Theta) = F(x|v) = \frac{\partial\, C_{x,v}(x, v, \Theta)}{\partial v}, \quad (7)$$

where the second parameter of $h(\cdot)$ always corresponds to the conditioning variable and Θ denotes the set of parameters for the copula of the joint distribution function of x and v. Further,