Lecture Notes in Statistics 213
Proceedings

Piotr Jaworski
Fabrizio Durante
Wolfgang Karl Härdle *Editors*

Copulae in Mathematical and Quantitative Finance

Proceedings of the Workshop Held in Cracow, 10–11 July 2012

Springer

Lecture Notes in Statistics 213

For further volumes:
http://www.springer.com/series/694

Piotr Jaworski • Fabrizio Durante •
Wolfgang Karl Härdle
Editors

Copulae in Mathematical and Quantitative Finance

Proceedings of the Workshop Held in Cracow, 10-11 July 2012

Springer

Editors
Piotr Jaworski
Faculty of Mathematics, Informatics, and Mechanics
University of Warsaw
Warszawa
Poland

Fabrizio Durante
School of Economics and Management
Free University of Bozen-Bolzano
Bozen
Italy

Wolfgang Karl Härdle
L.v.Bortkiewicz Chair of Statistics, C.A.S.E. Centre for Applied Statistics and Economics
Humboldt-Universität zu Berlin
Berlin
Germany

ISSN 0930-0325
ISBN 978-3-642-35406-9 ISBN 978-3-642-35407-6 (eBook)
DOI 10.1007/978-3-642-35407-6
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013940256

Foreword

The workshop "Copulae in Mathematical and Quantitative Finance" took place in Cracow (Poland) on 10th and 11th July 2012. This meeting was honoured to be a satellite event of the 6th European Congress of Mathematics, which was held in the same city the week before (2.07.2012–7.07.2012).

The event gathered 65 participants from 18 countries, across Europe and other continents, in the old and prestigious city of Cracow, a pearl of Polish art and cultural heritage and also a great scientific centre. Inspired by the nice atmosphere of the venue all the participants were actively involved in interesting and stimulating discussions about copula theory and its applications.

The workshop was preceded by a short course "Copulae Calibration in Theory and Practise" consisting of two sections organized by Claudia Czado and Eike Brechmann (Technische Universität München, Germany) and by Wolfgang K. Härdle and Ostap Okhrin (C.A.S.E. Humboldt-Universität zu Berlin, Germany). The course was particularly devoted both to PhD students and young researchers, who have found challenging ideas about multivariate copula models, and to practitioners, who have particularly benefited of practical implementation of the proposed methodologies.

As members of the organizing committee of the workshop, we have the privilege and the great pleasure to present this volume collecting results and achievements discussed by the participants. It is another confirmation that the event was fruitful for further scientific developments. Therefore, we would like to express our gratitude to all the participants for their delightful combination of scholarly inquiry and cheerful conviviality which confirm copula theory being such an active area of research.

We also would like to acknowledge the support of the institutional organizers of the workshop: Polish Mathematical Society, C.A.S.E.—Center for Applied Statistics and Economics (Humboldt-Universität zu Berlin, Germany) and Stefan Banach International Mathematical Center (Institute of Mathematics of Polish

Academy of Sciences). Moreover, we are honoured to be supported by the Ministry of Science and Higher Education of the Republic of Poland.

The attendance of specialists from various research groups around the word as well as the support of the institutional organizers and sponsors made the workshop a very successful event.

Warszawa, Poland — Piotr Jaworski
Bolzano, Italy — Fabrizio Durante
Warszawa, Poland — Krystyna Jaworska
Berlin, Germany — Ostap Okhrin
January 2013

Preface

The notion of copula provides an efficient way to describe the interrelationships of random variables and offers a great flexibility in building multivariate stochastic models. Since its discovery in the early 1950s, copulas have contributed to understand better the various facets of stochastic dependence and have allowed to break away from the standard assumptions (like multivariate Gaussian distribution), which generally underestimate the probability of joint extreme risks.

Nowadays, copula-based dependence models are rapidly gaining considerable popularity in several fields and are becoming indispensable tools not only in finance, insurance, risk management and econometrics but also in biostatistics, hydrology or machine learning. For example, they are widely used for the modelling of market, credit and operational risk, as well as for the aggregation of risks and portfolio selection. Moreover, such a large interest in the applications of copulas has spurred researchers and scientists in investigating and developing new theoretical methods and tools for handling randomness and uncertainty in practical situations.

The workshop "Copulae in Mathematical and Quantitative Finance", which took place in Cracow (Poland) on 10th–11th July 2012, has represented a good opportunity for intensive exchange of ideas about recent developments and achievements that can contribute to the general development of the field. The talks presented at this event have focused on several interesting theoretical problems as well as empirical applications.

In order to make all these contributions available to a larger audience, we have prepared this volume collecting both surveys giving an up-to-date account of some aspects of copula models and extended versions of talks presented at the workshop in Cracow.

Our special thanks go to the authors for their willingness to contribute to this volume and to our colleagues whose contribution as reviewers was essential in the preparation of the volume.

The professional work of the scientific and organizing committees was greatly appreciated, as well as the support of the co-sponsors of this conference.

Finally, we are indebted to our publisher Springer, in particular to Alice Blanck for her assistance in the editorial process.

Bolzano, Italy — Fabrizio Durante
Berlin, Germany — Wolfgang Karl Härdle
Warszawa, Poland — Piotr Jaworski
January 2013

Contents

Contributors

Eike Christian Brechmann Zentrum Mathematik, Technische Universität München, Munich, Germany

Umberto Cherubini Department of Statistics, University of Bologna, Bologna, Italy

Claudia Czado Zentrum Mathematik, Technische Universität München, Munich, Germany

Gal Elidan Department of Statistics, Hebrew University, Jerusalem, Israel

Jean-David Fermanian Department of Finance, Ensae-Crest, Malakoff, France

Christian Genest Department of Mathematics and Statistics, McGill University, Montréal (Québec), Canada

Fabio Gobbi Department of Statistics, University of Bologna, Bologna, Italy

Lutz Gruber Zentrum Mathematik, Technische Universität München, Munich, Germany

Nikolaus Hautsch School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Piotr Jaworski Institute of Mathematics, Warsaw University, Warsaw, Poland

Dominic Lauterbach Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Dietmar Pfeifer Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Haijun Li Department of Mathematics, Washington State University, Pullman, WA, USA

Mingyao Li Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA

Jan-Frederik Mai Assenagon Credit Management GmbH, München, Germany

XAIA Investment GmbH, München, Germany

Johanna G. Nešlehová Department of Mathematics and Statistics, McGill University, Montréal (Québec), Canada

Aristidis K. Nikoloulopoulos School of Computing Sciences, University of East Anglia, Norwich, UK

Ostap Okhrin School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Alexander Ristig School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Matthias Scherer Technische Universität München, Munich, Germany

Peter X.-K. Song Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

Bertrand Tavin Department of Management and Business Economics, Université Paris 1 - Panthéon Sorbonne, Paris, France

Rudi Zagst Technische Universität München, Munich, Germany

Peng Zhang Department of Mathematics, Zhejiang University, Hongzhou, China

Chapter 1
A Convolution-Based Autoregressive Process

Umberto Cherubini and Fabio Gobbi

Abstract We propose a convolution-based approach to the estimation of nonlinear autoregressive processes. The model allows for state-dependent autocorrelation, that is different persistence of the shocks in different phases of the market and dependent innovations, that is drawn from different distributions in different phases of the market.

1.1 Introduction

In this paper we review the application of copula functions to Markov processes used in econometrics, as it was recently described in the book by Cherubini, Gobbi, Mulinacci, and Romagnoli [6]. Our contribution relies on the application of a particular family of copulas, which are generated by the convolution operator, to the design of time series processes. From this point of view, the paper contributes to the literature modeling time series with copulas [3–5]. While this literature builds on the pioneering paper by Darsow, Nguyen, and Olsen [8] on the link between copula functions and Markov processes, our paper introduces the concept of convolution-based copulas.

Beyond the Markov property, there is a long-standing and extremely vast literature on the fact that most of the changes of the processes, those that are called innovations, are not predictable on the basis of past information. In financial markets, this concept goes back to the work by Bachelier in 1900 [1] and the developments of what is known as the *Efficient Market Hypothesis* due to the contributions by Paul Samuelson and Eugene Fama in the 1960s and 1970s [11, 17, 18]. The natural representation of this theory is to assess that log-prices

U. Cherubini (✉) · F. Gobbi
Department of Statistics, University of Bologna, Bologna, Italy
e-mail: umberto.cherubini@unibo.it; fabio.gobbi@unibo.it

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_1,

of assets follow a random walk. Technically, this process is characterized by innovations that are permanent and independent of the level of the process. The same random walk hypothesis spreads into the literature in the field of macroeconomics in the 1980s, starting with the seminal paper by Nelson and Plosser [16]. Based on the first unit root tests, due to Dickey–Fuller [9,10], Nelson and Plosser found that most of the US macroeconomic time series included a random walk component, that is a shock, independent and persistent, that is bound to remain forever in the history of the time series itself. In this paper we propose an extension to this approach, which allows for dependent innovations and for nonlinear dependence between the value of a process and that of the previous period.

The plan of the contribution is as follows. In Sect. 1.2 we present our non linear autoregressive model and discuss the possible restrictions that may be implied for copula function applications. In Sect. 1.3 we present the general probability framework, based on the concept of *C-convolution*. In Sect. 1.4 we specialize the model to provide a generalization of the efficient market hypothesis beyond the *random walk* assumption. In Sects. 1.5 and 1.6 we discuss estimation and simulation of the model. Finally, in Sect. 1.7 we present an application of the analysis to three stock indexes. Section 1.8 concludes.

1.2 A Nonlinear Autoregressive Model

As a starting point, consider the standard autoregressive process of first order, AR(1). It has the form

$$Y_t = \phi Y_{t-1} + \varepsilon_t$$

with the constant parameter ϕ satisfying the condition $|\phi| < 1$ to ensure stationarity. Furthermore, Y_{t-1} is assumed independent of the innovation ε_t for every t. When $\phi = 1$, the AR(1) process is simply a random walk.

We may explore two ways to generalize the above relationship. The first is to assume that the strength of the autoregressive relationship varies with the level of the process. It may, for example, happen that the autoregressive parameter be much closer to 1 in the tails, rather than in the middle of the distribution. A natural way to account for this generalization would be to make the autoregressive parameter dependent on the level of the process, that is

$$Y_t = \phi(Y_{t-1})Y_{t-1} + \varepsilon_t$$

where $\phi(.)$ is a function to be specified.

The second extension of the general AR(1) model has to do with the independence of the innovations to the process with respect to the level of the process itself. In this case, a natural way to address the problem in full generality is to use a copula function to link the level Y_{t-1} and the innovation ε_t. The history of this tool

dates back to a theorem by Sklar from 1959 [13, 15, 19], which states that any d-dimensional distribution can be decomposed into d univariate marginal distributions and a copula function which describes the dependence among the d variables. For our case, that is limited to Markovian processes, we stick to the bivariate case, and set $d = 2$.

It is easy to see that this simple idea of applying a general and, by today, standard tool to represent the dependence structure of two variables, such as copulas, in this case gives rise to a methodological innovation. In fact, remember that one of the most appreciated advantages of copula functions is their full flexibility, and the possibility to model whatever dependence structure of variables arbitrarily chosen. More formally, if we have any two variables X and Z with distributions F_X and F_Z by Sklar's theorem, we may write

$$H(x, z) = C(F_X(x), F_Y(z)),$$

where H is the joint distribution of (X, Z) and C is the copula of (X, Z). We are completely free to choose whatever distribution for X, and whatever one for Y, and we are free to accommodate whatever dependence structure between the two variables. In our case, some of this flexibility is lost. As a matter of fact, we are still free to select the dependence structure between Y_{t-1} and ε_t in full generality. We are also free to model the marginal distribution of ε_t as we like. But we cannot, and this is the distinguishing feature of this application, freely choose any distribution for Y_{t-1}. The reason is twofold. First, Y_{t-1} is a function of past innovations. Second, the way in which innovations affect the level is given by a linear relationship. Then, Y_{t-1} is the convolution of past innovations. This directly leads to a class of dependence functions that we call convolution copulas.

1.3 Convolution-Based Copulas

Our purpose here is to address the distribution of Y_t. Technically, this is a convolution of a sequence of innovations. The problem is that the concept of convolution is limited to independent variables. In order to allow for dependent innovations, the convolution concept must be extended to allow for dependent marginals. This extension, called *C-convolution*, was introduced in a paper by Cherubini, Mulinacci, and Romagnoli in 2011 [7] and developed in the quoted book [6]. We report here the main result for reference.

Proposition 1.1. *Let X e Y be two real-valued random variables on the same probability space $(\Omega, \Im, \mathbb{P})$ with corresponding copula $C_{X,Y}$ and continuous marginals F_X and F_Y. With $D_1 C_{X,Y}(u, v)$ we denote $\frac{\partial C_{X,Y}(u,v)}{\partial u}$. Then*

$$F_{X+Y}(z) = \int_0^1 D_1 C_{X,Y}\left(w, F_Y(z - F_X^{-1}(w))\right) dw \tag{1.1}$$

and

$$C_{X,X+Y}(u,v) = \int_0^u D_1 C_{X,Y}\left(w, F_Y(F_{X+Y}^{-1}(v) - F_X^{-1}(w))\right) dw. \qquad (1.2)$$

Notice that the convolution concept, extended to allow for dependence between the variables, jointly determines the distribution of the sum of the variables and the copula function linking one of the two variables and the sum of the two variables.

Our proposal here is to apply this framework to study the case of a nonlinear autoregressive process. More explicitly, we assume an AR(1) process in which the autoregressive coefficient ϕ is a function of Y_{t-1}. The model is $Y_t = \phi(Y_{t-1})Y_{t-1} + \varepsilon_t$, for $t = 1, 2 \ldots$, where we may choose $\phi(\cdot)$ as a flexible functional form, such as the following bounded and continuous function

$$\phi(y) = \frac{1}{1 + \lambda e^{-\mu y}}.$$

This function is characterized by two parameters λ and μ which will be estimated. We assume the disturbance $(\varepsilon_t)_t$ to be a sequence of identically distributed random variables with zero means and finite second moments. Moreover, we assume that the dependence structure between Y_{t-1} and ε_t is modeled by a time-invariant copula function C.

We now show how to derive both the distribution of Y_t, that we denote F_t, and the temporal dependence structure, represented by the joint distribution of (Y_{t-1}, Y_t). We can write

$$\begin{aligned}
F_{Y_{t-1},Y_t}(x,y) &= \mathbb{P}(Y_{t-1} \le x, Y_t \le y) = \\
&= \mathbb{P}(Y_{t-1} \le x, \phi(Y_{t-1})Y_{t-1} + \varepsilon_t \le y) = \\
&= \int_{-\infty}^{x} \mathbb{P}(Y_{t-1} \le x, \phi(Y_{t-1})Y_{t-1} + \varepsilon_t \le y | Y_{t-1} = s) dF_{t-1}(s) = \\
&= \int_{-\infty}^{x} \mathbb{P}(s \le x, \varepsilon_t \le y - \phi(s)s | Y_{t-1} = s) dF_{t-1}(s) = \\
&\int_{-\infty}^{x} D_1 C\left(F_{t-1}(s), F_{\varepsilon_t}(y - \phi(s)s)\right) dF_{t-1}(s),
\end{aligned}$$

where $D_1 C$ denotes the derivative w.r.t. the first argument of the copula C. With the change of variable $w = F_{t-1}(s)$ we get

$$F_{Y_{t-1},Y_t}(x,y) = \int_0^{F_{t-1}(x)} D_1 C\left(w, F_{\varepsilon_t}(y - \phi(F_{t-1}^{-1}(w))F_{t-1}^{-1}(w))\right) dw.$$

The distribution of Y_t is then

$$F_t(y) = \lim_{x \to +\infty} F_{Y_{t-1},Y_t}(x, y) = \int_0^1 D_1 C \left(w, F_{\varepsilon_t}(y - \phi(F_{t-1}^{-1}(w))F_{t-1}^{-1}(w))\right) dw, \tag{1.3}$$

whereas its density is

$$f_t(y) = \int_0^1 c \left(w, F_{\varepsilon_t}(y - \phi(F_{t-1}^{-1}(w))F_{t-1}^{-1}(w))\right) f_\varepsilon(y - \phi(F_{t-1}^{-1}(w))F_{t-1}^{-1}(w))dw,$$

where $c(\cdot, \cdot)$ is the copula density of C and f_{ε_t} is the density function of ε_t. By Sklar's theorem we can recover the copula function between Y_{t-1} and Y_t which is

$$\begin{aligned} C_{Y_{t-1},Y_t}(u, v) &= F_{Y_{t-1},Y_t}(F_{t-1}^{-1}(u), F_t^{-1}(v)) = \\ &= \int_0^u D_1 C \left(w, F_{\varepsilon_t}(F_t^{-1}(v) - \phi(F_{t-1}^{-1}(w))F_{t-1}^{-1}(w))\right) dw. \end{aligned}$$

The copula density between Y_{t-1} and Y_t is

$$\begin{aligned} c_{Y_{t-1},Y_t}(u, v) &= \frac{\partial^2}{\partial u \partial v} C_{Y_{t-1},Y_t}(u, v) = \\ &= c(u, F_{\varepsilon_t}(F_t^{-1}(v) - \phi(F_{t-1}^{-1}(u))F_{t-1}^{-1}(u))) \frac{f_{\varepsilon_t}(F_t^{-1}(v) - \phi(F_{t-1}^{-1}(u))F_{t-1}^{-1}(u))}{f_t(F_t^{-1}(v))}, \end{aligned} \tag{1.4}$$

where c denotes the copula density of C.

1.4 Efficient Market Dynamics

A possible specialization of the dynamics described above is an extension of the standard statistical assumptions that describe the dynamics of speculative efficient markets. The requirements that are typically imposed in the standard literature are $\phi(y) = 1$ and $C(u, v) = uv$. This defines the so-called *random walk* model. In plain words, innovations are assumed to be independent of the level of the process, and increments are not a function of the level itself.

It is important to notice that the model allows extensions of the efficient market paradigm beyond the assumptions above. In particular, we could consider keeping the assumption $\phi(y) = 1$ and dropping independence between the innovation and the price. This was done in [7], and it is important to notice that this result paves the way to more robust tests of the efficient market hypothesis, extending beyond the *random walk* assumption. We report here the details of the result.

We first define the restrictions that must be included in a process with dependent increments to satisfy the martingale condition:

Theorem 1.1. *Let $X = (X_t)_{t\geq 0}$ be a Markov process and set $Y = X_t - X_s$. X is a martingale if and only if for all $t, s, s < t$:*

1. *F_Y has finite mean;*
2. *$\int_0^1 F_Y^{-1}(v)d(D_1 C_{X_s,Y}(u,v)) = 0, \quad \forall u \in [0,1]$ a.e..*

If we restrict the selection of distributions of increments to the set of symmetric distributions, we can formally work out the restriction that has to be imposed on the dependence of increments.

Proposition 1.2. *The martingale condition is satisfied for every symmetric distribution of increments F_Y if and only if the copula between the increments and the levels has the symmetry property*

$$\tilde{C}(u,v) \equiv u - C(u, 1-v) = C(u,v) \tag{1.5}$$

Proof. For simplicity we set $C_{X_s,Y} = C$ and $F_Y = F$, being F a symmetric distribution,

$$\begin{aligned}
\int_0^1 F^{-1}(v)d(D_1C(u,v)) &= \int_0^{\frac{1}{2}} F^{-1}(v)d(D_1C(u,v)) + \int_{\frac{1}{2}}^1 F^{-1}(v)d(D_1C(u,v)) \\
&= \int_0^{\frac{1}{2}} F^{-1}(v)d(D_1C(u,v)) + \int_{\frac{1}{2}}^0 F^{-1}(1-\rho)d(D_1C(u,1-\rho)) \\
&= \int_0^{\frac{1}{2}} F^{-1}(v)d(D_1C(u,v)) + \int_0^{\frac{1}{2}} F^{-1}(\rho)d(D_1C(u,1-\rho)) \\
&= \int_0^{\frac{1}{2}} F^{-1}(v)d(D_1C(u,v)) + d(D_1C(u,1-v)) \\
&= \int_0^{\frac{1}{2}} F^{-1}(v)d(D_1C(u,v)) + D_1C(u,1-v)) \\
&= \int_0^{\frac{1}{2}} F^{-1}(v)d(D_1(C(u,v) + C(u,1-v)) = 0, \qquad \forall u \in (0,1).
\end{aligned}$$

Last condition is satisfied for every symmetric distribution F if and only if (notice that, in last integral, $F^{-1}(v) < 0$ in a given not empty interval)

$$d(D_1(C(u,v) + C(u,1-v))) = 0 \quad \forall u, v \in (0,1).$$

It maybe easily verified that this condition is satisfied if and only if

$$C(u,v) + C(u,1-v) = u$$

which is the symmetry condition required for the copula □

Notice that there are almost infinite possibilities to build copulas with the symmetry described above. A simple technique would be to select any copula $A(u, v)$ and to define its symmetric counterpart $\tilde{A}(u, v) \equiv u - A(u, 1 - v)$. Then, construct a mixture copula $C(u.v) \equiv 0.5A + 0.5\tilde{A}$. It is an easy exercise to prove that $C(u, v)$ is endowed with the property required in the proposition above (see [14], for application of this technique to general concepts of symmetry). So, the choice is extremely vast but not exhaustive, however, as the counterexample below demonstrates.

Remark 1.1. Assume you want to construct a martingale process with symmetric and dependent increments, with dependence modeled by an FGM copula. It is easy to show that this is unfeasible. Select $A_\theta(u, v) = uv + \theta uv(1-u)(1-v)$ as required. Now compute

$$\tilde{A}_\theta(u, v) = u - u(1-v) - \theta uv(1-u)(1-v) = uv - \theta uv(1-u)(1-v) = A_{-\theta}(u, v).$$

But now, by the property of the FGM copula discussed above (or by direct computation) we have

$$0.5A_\theta(u, v) + 0.5\tilde{A}_\theta(u, v) = 0.5A_\theta(u, v) + 0.5A_{-\theta}(u, v) = A_0(u, v) = uv$$

and the increments mischievously turn out independent.

1.5 Simulating Convolution Based Processes

We now show how to simulate our nonlinear AR(1) process in full generality, that is taking into account the dependence structure between the innovation and the level of the process. This is simply obtained by applying the technique of conditional sampling as described in the quasi-algorithm reported below. The input is given by a sequence of distributions of innovations that for the sake of simplicity we assume stationary, $F_{\varepsilon_t} = F_\varepsilon$, and a temporal dependence structure that we consider stationary as well, $C_{Y_{t-1},\varepsilon_t}(u, v) = C(u, v)$. We also assume $Y_0 = 0$. The procedure to generate a trajectory of n points from the convolution-based process is the following:

1. $t = 1$.
2. Generate u from a uniform distribution.
3. Compute $Y_t = F_\varepsilon^{-1}(u)$.
4. Use conditional sampling to generate v from $D_1 C(u, v)$.
5. Compute $\varepsilon_{t+1} = F_\varepsilon^{-1}(v)$.
6. Compute $Y_{t+1} = Y_t + \varepsilon_{t+1}$.
7. Compute the distribution of Y_{t+1}, $F_{t+1}(y)$, by Eq. (1.1).
8. Compute $u = F_{t+1}(Y_{t+1})$.
9. $t = t + 1$.
10. If $t < n + 1$ go to step 4, else End.

Notice that the algorithm can be very simply extended to the simulation of martingale process. The key idea is that if we assume a symmetric distribution for innovation, the martingale requirement implies a corresponding symmetry of the copula function linking innovations and levels, along the lines described in Sect. 1.4. Drawing from a copula with such features only requires two lines of code more with respect to the pseudo-algorithm above. In particular, when conditionally on the level we draw a new innovation, it is sufficient to generate a binary random variable for the sign of the innovation itself. For clarity, here below we report the algorithm modified accordingly:

1. $t = 1$.
2. Generate u from a uniform distribution.
3. Compute $Y_t = F_{\varepsilon}^{-1}(u)$.
4. Generate s from a uniform distribution.
5. Use conditional sampling to generate v from $D_1 C(u, v)$.
6. Compute $\varepsilon_{t+1} = F_{\varepsilon}^{-1}(v)$.
7. If $s \leq 0.5$, $\varepsilon_{t+1} = -\varepsilon_{t+1}$.
8. Compute $Y_{t+1} = Y_t + \varepsilon_{t+1}$.
9. Compute the distribution of Y_{t+1}, $F_{t+1}(y)$, by Eq. (1.1).
10. Compute $u = F_{t+1}(Y_{t+1})$.
11. $t = t + 1$.
12. If $t < n + 1$ go to step 4, else End.

The new lines of code required are those in steps 6 and 7.

1.6 Maximum Likelihood Estimation

We now show how to recover the likelihood function for the estimation of our C-convolution-based Markov process. Since Y_t is Markov, the likelihood function needs the transition probability of Y_t at the points t and $t - 1$. For a detailed discussion of these results, the reader can refer to Basawa and Rao [2] for the general case Markov processes and Hamilton [12] for the particular case of autoregressive processes. Denote such a probability by $\mathbb{P}_t(y_t|y_{t-1}) = \mathbb{P}(Y_t \leq y_t|Y_{t-1} = y_{t-1})$. We assume that the disturbances are Gaussian with zero mean and standard deviation σ_ε. So, in our model we have four parameters: λ, μ, σ_ε, and the copula parameter θ. Let $\Theta = (\lambda, \mu, \sigma_\varepsilon, \theta)$. We write $\mathbb{P}_t(y_t|y_{t-1}; \Theta)$ to emphasize the dependence on parameters. The transition density $f_t(y_t|y_{t-1}; \Theta)$ exists for each Θ and for $t = 1, \ldots, n$. Denote by $(y_1, \ldots, y_n)$ a time series from Y_t. Then, it is easy to construct the log-likelihood function of $(y_1, \ldots y_n)$; it is given by

$$\ell(y_1, \ldots, y_n; \Theta) = \sum_{t=2}^{n} \log f_t(y_t|y_{t-1}; \Theta),$$

The transition density $f_t(y_t|y_{t-1}; \Theta)$ is the joint density between Y_{t-1} and Y_t divided by the marginal density of Y_{t-1}

$$f_t(y_t|y_{t-1};\Theta) = \frac{f_{Y_{t-1},Y_t}(y_{t-1},y_t;\Theta)}{f_{t-1}(y_{t-1};\Theta)}, \quad t = 2,\dots,n.$$

Moreover, the joint density $f_{Y_{t-1},Y_t}(y_{t-1},y_t;\Theta)$ is

$$f_{Y_{t-1},Y_t}(y_{t-1},y_t;\Theta) = c_{Y_{t-1},Y_t}(F_{t-1}(y_{t-1}),F_t(y_t);\Theta)\, f_{t-1}(y_{t-1};\Theta)\, f_t(y_t;\Theta),$$

and then

$$f_t(y_t|y_{t-1};\Theta) = c_{Y_{t-1},Y_t}(F_{t-1}(y_{t-1}),F_t(y_t);\Theta)\, f_t(y_t;\Theta), \quad t = 2,\dots,n,$$

and from Eq. (1.4)

$$f_t(y_t|y_{t-1};\Theta) = c(F_{t-1}(y_{t-1}),F_t(y_t);\Theta)\, f_{\varepsilon_t}(y_t - \phi(y_{t-1})y_{t-1};\Theta).$$

Finally, the log-likelihood of the process is the following

$$\begin{aligned}&\ell(y_1,\dots,y_n;\Theta)\\ &= \sum_{t=2}^{n} \log\left(c(F_{t-1}(y_{t-1}),F_t(y_t);\Theta)\right) + \sum_{t=2}^{n} \log\left(f_{\varepsilon_t}(y_t - \phi(y_{t-1})y_{t-1};\Theta)\right).\end{aligned}$$

The maximum likelihood estimate is then

$$\hat{\Theta} = \arg\max_{\Theta} \ell(y_1,\dots,y_n;\Theta).$$

1.7 Application

In this section we provide an empirical application of our convolution-based model to three stock indexes. We consider three daily time series of the three major European financial market indexes, i.e., Cac40 (France), Dax30 (Germany), and FtseMib (Italy). The data set employed runs from 1 January 2007 to 31 December 2011. The total number of observations is $n = 1{,}300$. Figure 1.1 displays the logarithm of the prices.

1.7.1 Estimation

For the estimation, we implement the following strategy. First, we fit a copula function between the sequence of increments and the corresponding level before the increment itself. Second, we perform maximum likelihood estimation of the

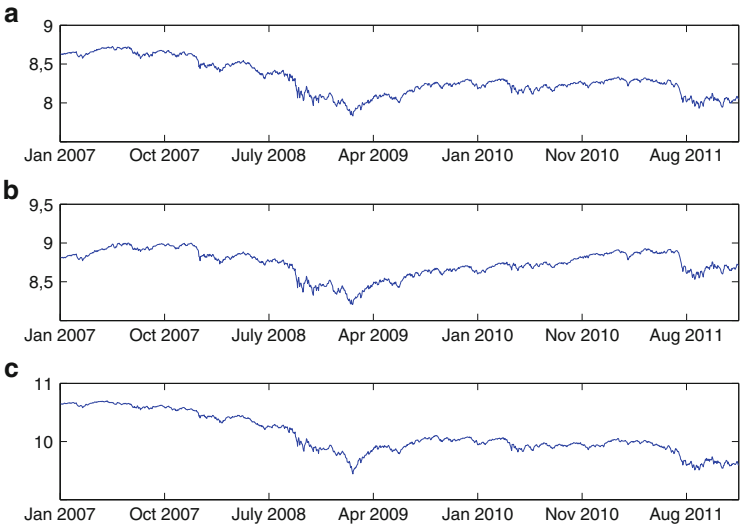

Fig. 1.1 Log-prices of indexes. (**a**) CaC40, (**b**) Dax30, (**c**) FtseMib

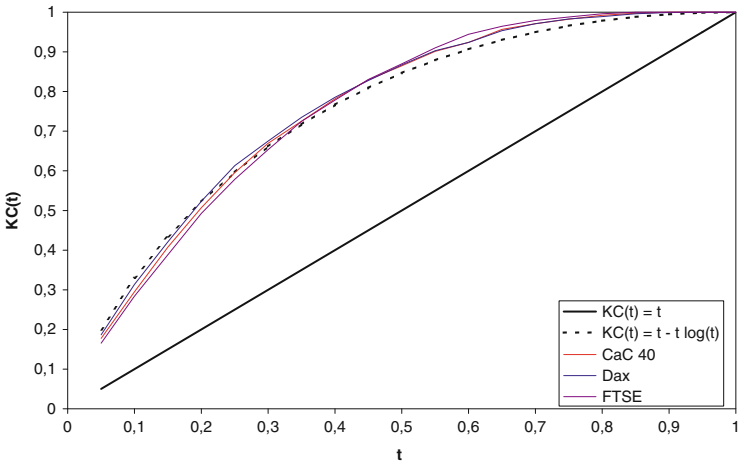

Fig. 1.2 Kendall function analysis of increments: CaC40, Dax30, FtseMib

C-convolution model. As for the first step, in this application we apply the standard technique typically used for copula fitting, that is the Kendall function approach. Figure 1.2 reports the empirical Kendall functions estimated for the three indexes. In order to give a visual representation of the actual degree of dependence between increments of the indexes and their levels, we also report what the Kendall function should be in case of perfect dependence and independence. While in case of perfect dependence the Kendall function should coincide with the straight line, we see that the empirical Kendall functions are very close to the curved schedule

Table 1.1 Maximum likelihood estimates of model parameters and relative standard errors

CaC40	DaX30	FtseMib
$\hat{\lambda} = 0.5942(0.9883)$	$\hat{\lambda} = 0.7947(0.9940)$	$\hat{\lambda} = 0.2318(5.1751)$
$\hat{\mu} = 4.2938^*(0.2725)$	$\hat{\mu} = 2.7573^*(0.4173)$	$\hat{\mu} = 1.7884(1.8785)$
$\hat{\sigma_\varepsilon} = 0.0177^*(3.46 \times 10^{-4})$	$\hat{\sigma_\varepsilon} = 0.0168^*(3.29 \times 10^{-4})$	$\hat{\sigma_\varepsilon} = 0.0185^*(3.63 \times 10^{-4})$

The asterisk denotes that the parameter is significantly different from zero at the 5 % level

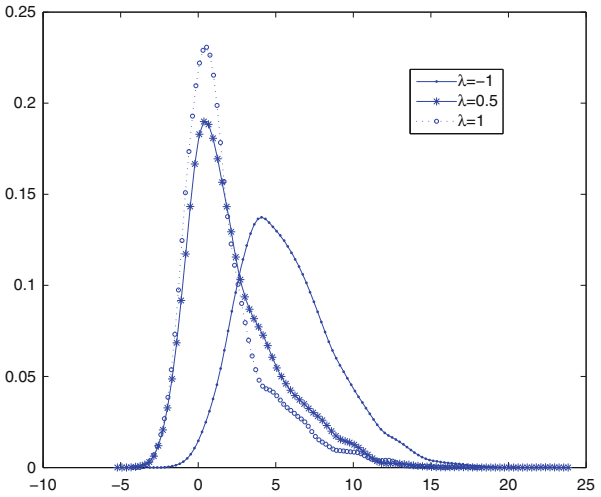

Fig. 1.3 The effect of the parameter λ on the Kernel density estimate relative to simulated values. Here $\mu = 1$ and $t = 25$ (1 month)

representing independence. Of course, this does not come as a surprise and testifies that the association that is typically made between efficient market and independent increments is borne out by the data, at least in this application. What remains to be estimated is the shape of the function $\phi(.)$. In other words, we find that the copula C is the product copula $\Pi(u, v) = uv$, so that three parameters are left to be estimated, namely λ, μ, and σ_ε.

Table 1.1 reports the maximum likelihood estimates of the three parameters of the model. We see that the λ parameters are not significant in any stock index.

1.7.2 Simulation

After estimating the process, the next step would be simulation and construction of the density of the process. Unfortunately, since in our case the estimates did not provide an interesting model to simulate, rather than a plain random walk, here we take the opportunity to perform the simulation on a sample of parameters that allow us to understand their impact on the density. We then ask the reader to pretend that these parameters were the outcome of an estimation procedure.

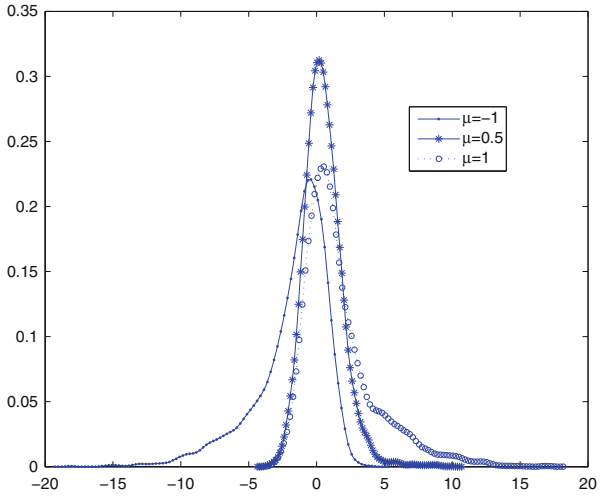

Fig. 1.4 The effect of the parameter μ on the Kernel density estimate relative to simulated values. Here $\lambda = 1$ and $t = 25$ (1 month)

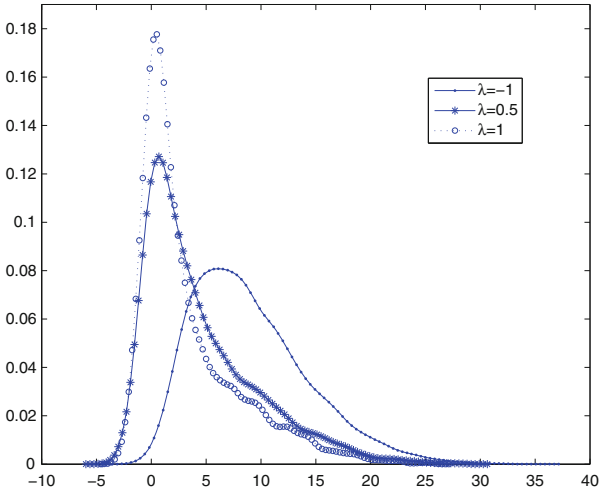

Fig. 1.5 The effect of the parameter λ on the Kernel density estimate relative to simulated values. Here $\mu = 1$ and $t = 75$ (3 months)

The simulation design is the following. We assume that the innovations $(\varepsilon_t)_t$ are Gaussian with zero mean and time-independent variance $\sigma_\varepsilon = 1$. Moreover, the starting point is at random from a standard Normal distribution, i.e. $Y_1 \sim N(0, 1)$. We assume that Y_{t-1} and ε_t are independent. We study 12 cases corresponding to the combinations of three values of λ (-1, 0.5, and 1) and four values of μ (-1, -0.5, 0.5, and 1). We generate 5,000 trajectories of 250 points for each case described above. The comparative study aims to capture the effect on the distribution of Y_t of

Table 1.2 Descriptive statistics relative to Monte Carlo simulations of our C-convolution-based model in the independent case where $t = 25$ (1 month)

	$\mu = -1$	$\mu = -0.5$	$\mu = 0.5$	$\mu = 1$
$\lambda = -1$	$m = -5.7230$	$m = -9.6669$	$m = 9.6872$	$m = 5.7518$
	$s = 3.0074$	$s = 2.9626$	$s = 2.9725$	$s = 3.0782$
	$\kappa = 3.5613$	$\kappa = 3.3526$	$\kappa = 3.6441$	$\kappa = 3.4260$
	$\gamma = -0.7322$	$\gamma = -0.5012$	$\gamma = 0.5200$	$\gamma = 0.6933$
$\lambda = 0.5$	$m = -2.1679$	$m = -0.9118$	$m = 0.9390$	$m = 2.2648$
	$s = 2.9415$	$s = 2.0527$	$s = 2.0446$	$s = 2.9183$
	$\kappa = 5.1456$	$\kappa = 8.3283$	$\kappa = 7.3815$	$\kappa = 4.4763$
	$\gamma = -1.2820$	$\gamma = -1.6606$	$\gamma = 1.5301$	$\gamma = 1.1458$
$\lambda = 1$	$m = -1.7532$	$m = -0.4469$	$m = 0.4932$	$m = 1.6792$
	$s = 2.6831$	$s = 1.4292$	$s = 1.4184$	$s = 2.7001$
	$\kappa = 6.0142$	$\kappa = 6.5753$	$\kappa = 6.3225$	$\kappa = 5.9665$
	$\gamma = -1.5003$	$\gamma = -0.9971$	$\gamma = 0.9897$	$\gamma = 1.5469$

Table 1.3 Descriptive statistics relative to Monte Carlo simulations of our C-convolution-based model in the independent case where $t = 75$ (3 months)

	$\mu = -1$	$\mu = -0.5$	$\mu = 0.5$	$\mu = 1$
$\lambda = -1$	$m = -9.0376$	$m = -13.7802$	$m = 13.7844$	$m = 9.0399$
	$s = 5.1798$	$s = 5.1562$	$s = 5.1717$	$s = 5.2144$
	$\kappa = 3.8614$	$\kappa = 3.7549$	$\kappa = 3.9126$	$\kappa = 3.6363$
	$\gamma = -0.8854$	$\gamma = -0.7901$	$\gamma = 0.7941$	$\gamma = 0.8532$
$\lambda = 0.5$	$m = -4.2716$	$m = -1.2801$	$m = 1.3897$	$m = 4.3986$
	$s = 4.9981$	$s = 2.8565$	$s = 2.9864$	$s = 5.0711$
	$\kappa = 4.9188$	$\kappa = 14.1282$	$\kappa = 12.4878$	$\kappa = 4.4420$
	$\gamma = -1.3682$	$\gamma = -2.7083$	$\gamma = 2.5545$	$\gamma = 1.2755$
$\lambda = 1$	$m = -3.3821$	$m = -0.4672$	$m = 0.5256$	$m = 3.3468$
	$s = 4.7324$	$s = 1.4817$	$s = 1.5886$	$s = 4.5971$
	$\kappa = 6.1422$	$\kappa = 11.4774$	$\kappa = 24.2268$	$\kappa = 5.2446$
	$\gamma = -1.7022$	$\gamma = -1.4904$	$\gamma = 2.5171$	$\gamma = 1.5381$

the two parameters. Tables 1.2 and 1.3 summarize the results of our Monte Carlo simulation in terms of descriptive statistics: the mean (m), the standard deviation (s), the kurtosis (κ), and the skewness (γ). The parameter λ clearly affects the kurtosis of the distribution of Y_t: it raises with λ and in particular it is higher as μ is lower. Moreover, the sign of μ affects the sign of the skewness, i.e., the distribution has a negative skewness as long as the parameter μ is negative and vice versa. Figures 1.3–1.6 display the estimated density functions for the simulated values when one of the two parameters is fixed and the other one varies to underline their role in the properties of the distribution of Y_t.

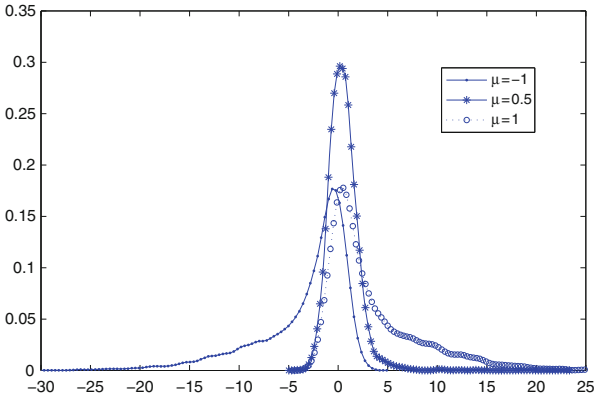

Fig. 1.6 The effect of the parameter μ on the Kernel density estimate relative to simulated values. Here $\lambda = 1$ and $t = 75$ (3 months)

1.8 Conclusion

In this paper we propose a convolution-based approach to the estimation of nonlinear autoregressive processes. The model allows for state-dependent autocorrelation, that is different persistence of the shocks in different phases of the market and dependent innovations, that is drawn from different distributions in different phases of the market. The model is well suited to address problems of persistent and unpredictable shocks, beyond the standard paradigm of linear models.

For what concerns the theory of copula functions, this provides an example in which the approach surrenders some of its flexibility. The idea is that once the distribution of innovations is specified, and the dependence structure between innovations and levels of the process is chosen, the distribution of the process can be automatically recovered.

References

1. Bachelier, L.: Théorie de la Speculation. Gauthier-Villars, Paris (1900)
2. Basawa, I.V., Prakasa Rao, B.L.S.: Statistical Inference for Stochastic Processes. Academic, New York (1980)
3. Beare, B.K.: Copulas and temporal dependence. Econometrica **78**(1), 395–410 (2010)
4. Chen X., Fan Y.: Estimation of copula-based semiparametric time series models. J. Econom. **130**, 307–335 (2006)
5. Chen, X., Wu, W.B., Yi Y.: Efficient estimation of copula-based semiparametric Markov models. Ann. Stat. **37**(6B), 4214–4253 (2009)
6. Cherubini, U., Gobbi, F., Mulinacci, S., Romagnoli, S.: Dynamic Copula Methods in Finance. Wiley, Chichester (2011)

7. Cherubini, U., Mulinacci, S., Romagnoli, S.: A copula-based model of speculative price dynamics in discrete time. J. Multivar. Anal. **102**(6), 1047–1063 (2011)
8. Darsow, W.F., Nguyen, B., Olsen, E.T.: Copulas and Markov processes. Illinois J. Math. **36**(4), 600–642 (1992)
9. Dickey, D.A., Fuller, W.A.: Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. **74**, 427–431 (1979)
10. Dickey, D.A., Fuller, W.A.: Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica **49**, 1057–1072 (1981)
11. Fama, E.F.: Efficient capital markets: a review of theory and empirical work. J. Financ. **25**(2), 383–417 (1965)
12. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)
13. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
14. Klement, E.P., Mesiar, R., Pap, E.: Invariant copulas. Kybernetika **38**, 275–285 (2002)
15. Nelsen, R.: An Introduction to Copulas. Springer, New York (2006)
16. Nelson, C., Plosser, C.: Trend and randon walk in macroeconomic time series. J. Monet. Econ. **10**, 139–169 (1982)
17. Samuelson, P.A.R.: Proof that properly anticipated prices fluctuate randomly. Ind. Manage. Rev. **6**, 41–50 (1963)
18. Samuelson, P.A.R.: Mathematics of speculative price. SIAM Rev. **15**(1), 1–42 (1973)
19. Sklar, A.: Fonctions de répartition à n dimension et leurs marges. Publ. Inst. Stat. Univ. Paris **8**, 229–231 (1959)

Chapter 2
Selection of Vine Copulas

Claudia Czado, Eike Christian Brechmann, and Lutz Gruber

Abstract Vine copula models have proven themselves as a very flexible class of multivariate copula models with regard to symmetry and tail dependence for pairs of variables. The full specification of a vine model requires the choice of a vine tree structure, the copula families for each pair copula term and their corresponding parameters. In this survey we discuss the different approaches, both frequentist and Bayesian, for these model choices so far and point to open problems.

2.1 Introduction

The analysis of high-dimensional data sets requires flexible multivariate stochastic models that can capture the inherent dependency patterns. The copula approach, which separates the modeling of the marginal distributions from modeling the dependence characteristics, is a natural one to follow in this context. This development has spawned a tremendous increase in copula-based applications in the last 10 years, especially in the areas of finance, economics, and hydrology.

Considerable efforts have been undertaken to increase the flexibility of multivariate copula models beyond the scope of elliptical and Archimedean copulas. Vine copulas are among the best-received of such efforts. Vine copulas use (conditional) bivariate copulas as the so-called pair copula building blocks to describe a multivariate distribution (see [37]). A set of linked trees—the "vine"—describes a vine copula's factorization of the multivariate copula density function into the density functions of its pair copulas (see [8, 9]). The article by [1] illustrates a first application of the vine copula concept using non-Gaussian pair copulas to financial data. The first comprehensive account of vine copulas is found in [45], a

C. Czado (✉) · E.C. Brechmann · L. Gruber
Zentrum Mathematik, Technische Universität München, Munich, Germany
e-mail: cczado@ma.tum.de; brechmann@ma.tum.de; lutz.gruber@ma.tum.de

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_2,

recent survey in [20], and the current developments of this active research area in [46].

Elliptical copulas as well as Archimedean copulas have been shown to be inadequate models to describe the dependence characteristics of real data applications (see, for example, [1, 23, 24]). As a pair copula construction, vine copulas allow different structural behaviors of pairs of variables to be modeled suitably, in particular so with regard to their symmetry, or lack thereof, strength of dependence, and tail dependencies. Such flexibility requires well-designed model selection procedures to realize the full potential of vine copulas as dependence models. Successful applications of vines can be found, amongst others, in [10, 12, 17, 22, 24, 32, 48, 51, 54, 61].

A parametric vine copula consists of three components: a set of linked trees identifying the pairs of variables and their conditioning variables, the copula families for each pair copula term given by the tree structure, and the corresponding copula parameters. The three-layered definition leads to three fundamental estimation and selection tasks: (1) Estimation of copula parameters for a chosen vine tree structure and pair copula families, (2) Selection of the parametric copula family for each pair copula term and estimation of the corresponding parameters for a chosen vine tree structure, and (3) Selection and estimation of all three model components. In this survey we address these tasks and give an overview of the statistical approaches taken so far. These range from frequentist to Bayesian methods.

The remainder of this paper is structured as follows. In Sect. 2.2 we provide the necessary methodical background on regular vines and regular vine copulas. We then discuss estimation of parameters of regular vine copulas in Sect. 2.3 and the selection of appropriate pair copulas in Sect. 2.4. Section 2.5 treats the joint selection of the regular vine tree structure, the copula families, and their parameters. Section 2.6 concludes with a discussion of available software and open problems.

2.2 Regular Vine Copulas

Copulas describe a statistical model's dependence behavior separately from its marginal distributions [60]. As such, a copula is a multivariate distribution function with all marginal distributions being uniform: i.e. the copula associated with an n-variate cumulative distribution function $F_{1:n}$ with univariate marginal distribution functions $F_1, \ldots, F_n$ is a distribution function $C : [0,1]^n \to [0,1]$ satisfying

$$F_{1:n}(\mathbf{x}) = C\left(F_1(x_1), \ldots, F_n(x_n)\right), \quad \mathbf{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n.$$

If C is absolutely continuous, its density is denoted by c.

The factorization of multivariate copula densities into (conditional) bivariate copula densities is due to [8, 37], which were developed independently. The details of these factorizations are represented by the graph theoretical construction called regular vine to organize different decompositions. Graphs are defined in terms of a set of nodes N and a set of edges E connecting these nodes, i.e. $E \subset N \times N$. Vines

are based on trees which are particular graphs where there is a unique sequence of edges between each two nodes.

Definition 2.1 (Regular Vine Tree Sequence). A set of linked trees $\mathcal{V} = (T_1, T_2, \ldots, T_{n-1})$ is a regular vine (R-vine) on n elements if

1. T_1 is a tree with nodes $N_1 = \{1, \ldots, n\}$ and a set of edges denoted by E_1.
2. For $i = 2, \ldots, n-1$, T_i is a tree with nodes $N_i = E_{i-1}$ and edge set E_i.
3. For $i = 2, \ldots, n-1$, if $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are two nodes in N_i connected by an edge, then exactly one of the a_i equals one of the b_i (proximity condition).

In other words, the proximity condition requires that the edges corresponding to two connected nodes in tree T_i share a common node in tree T_{i-1}. This ensures that the decomposition into bivariate copulas which is given below is well defined.

Two sub-classes of regular vines have been studied extensively in the literature: canonical vines (C-vines) and drawable vines (D-vines) (see [1, 45]). C-vines are characterized by a root node in each tree T_i, $i \in \{1, \ldots, n-1\}$, which has degree $n - i$; that means that the root node is connected to all other nodes of the tree. D-vines, on the other hand, are uniquely characterized through their first tree which is, in graph theoretical terms, a path; this means that each node has degree of at most 2. Therefore the order of variables in the first tree defines the complete D-vine tree sequence.

Some more definitions are needed to introduce regular vine copulas: the complete union A_e of an edge $e = \{a, b\} \in E_i$ in tree T_i of a regular vine $\mathcal{V}$ is defined by

$$A_e = \{v \in N_1 : \exists e_m \in E_m,\ m = 1, \ldots, i-1,\ \text{such that } v \in e_1 \in \cdots \in e_{i-1} \in e\}. \tag{2.1}$$

The conditioning set associated with $e = \{a, b\}$ is defined as $D_e := A_a \cap A_b$ and the conditioned sets associated with $e = \{a, b\}$ are defined as $\mathscr{C}_{e,a} := A_a \setminus D_e$ and $\mathscr{C}_{e,b} := A_b \setminus D_e$. Bedford and Cooke [8] showed that the conditioned sets are singletons, and we will therefore refer to edges by their labels $\{\mathscr{C}_{e,a}, \mathscr{C}_{e,b} | D_e\} \hat{=} \{i(e), j(e) | D(e)\}$. An exemplary regular vine on five elements is shown in Fig. 2.1.

Given these sets, we can specify a regular vine copula by associating a (conditional) bivariate copula with each edge of the regular vine, a so-called pair copula.

Definition 2.2 (Regular Vine Copula). A regular vine copula $C = (\mathcal{V}, \mathscr{B}(\mathcal{V}), \boldsymbol{\theta}(\mathscr{B}(\mathcal{V})))$ in n dimensions is a multivariate distribution function such that for a random vector $\mathbf{U} = (U_1, \ldots, U_n)' \sim C$ with uniform margins

1. $\mathcal{V}$ is a regular vine on n elements,
2. $\mathscr{B}(\mathcal{V}) = \left\{C_{i(e),j(e)|D(e)} | e \in E_m,\ m = 1, \ldots n-1\right\}$ is a set of $n(n-1)/2$ copula families identifying the conditional distributions of $U_{i(e)}, U_{j(e)} | \mathbf{U}_{D(e)}$,
3. $\boldsymbol{\theta}(\mathscr{B}(\mathcal{V})) = \left\{\boldsymbol{\theta}_{i(e),j(e)|D(e)} | e \in E_m,\ m = 1, \ldots n-1\right\}$ is the set of parameter vectors corresponding to the copulas in $\mathscr{B}(\mathcal{V})$.

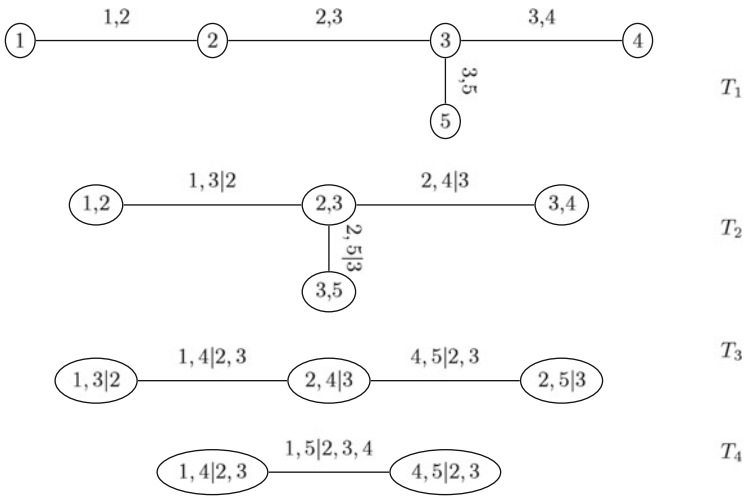

Fig. 2.1 Regular vine on five elements

Therefore the full specification of a regular vine copula consists of three layers: the regular vine tree structure $\mathcal{V}$, the pair copula families $\mathcal{B} = \mathcal{B}(\mathcal{V})$, and the pair copula parameters $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathcal{B}(\mathcal{V}))$. Regular vine copulas which differ in the tree structure or in at least one pair copula family represent in general different statistical models. Notable exceptions from this are the multivariate Gaussian, Student's t or Clayton copulas, which can be decomposed into bivariate Gaussian, Student's t or Clayton copulas, respectively, in multiple ways (see [64]). The probability density function $f_{1:n}$ at point $\mathbf{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n$ of an n-dimensional regular vine-dependent distribution $F_{1:n}$ is easily calculated as

$$f_{1:n}(\mathbf{x}|\mathcal{V}, \mathcal{B}, \boldsymbol{\theta}) = \left(\prod_{m=1}^{n-1} \prod_{e=\{a,b\} \in E_m} c_{i(e),j(e)|D(e)} \left(F_{i(e)|D(e)}, F_{j(e)|D(e)} | \boldsymbol{\theta}_{i(e),j(e)|D(e)} \right) \right) \times f_1(x_1) \cdots f_n(x_n), \tag{2.2}$$

where $F_{i(e)|D(e)} := F_{i(e)|D(e)}\left(x_{i(e)}|\mathbf{x}_{D(e)}\right)$ and $F_{j(e)|D(e)} := F_{j(e)|D(e)}\left(x_{j(e)}|\mathbf{x}_{D(e)}\right)$ (see [8]). These conditional distribution functions can be determined recursively tree-by-tree using the following relationship

$$\begin{aligned} F_{i(e)|D(e)}\left(x_{i(e)}|\mathbf{x}_{D(e)}\right) &= F_{\mathscr{C}_{e,a}|D_e}\left(x_{\mathscr{C}_{e,a}}|\mathbf{x}_{D_e}\right) \\ &= \frac{\partial C_{\mathscr{C}_{a,a_1},\mathscr{C}_{a,a_2}|D_a}\left(F_{\mathscr{C}_{a,a_1}|D_a}\left(x_{\mathscr{C}_{a,a_1}}|\mathbf{x}_{D_a}\right), F_{\mathscr{C}_{a,a_2}|D_a}\left(x_{\mathscr{C}_{a,a_2}}|\mathbf{x}_{D_a}\right)\right)}{\partial F_{\mathscr{C}_{a,a_2}|D_a}\left(x_{\mathscr{C}_{a,a_2}}|\mathbf{x}_{D_a}\right)}, \end{aligned} \tag{2.3}$$

where $e = \{a, b\}$ with $a = \{a_1, a_2\}$ as before (see [23] for details).

To facilitate inference, it is assumed that the pair copulas $C_{i(e),j(e)|D(e)}$ only depend on the variables with indices in $D(e)$ through the arguments $F_{i(e)|D(e)}$ and $F_{j(e)|D(e)}$. This so-called simplifying assumption has been investigated by [35, 64]. In particular, [35] show examples where the assumption is not severe and [64] show that the multivariate Clayton copula is the only Archimedean copula which can be represented as a simplified vine. A critical look on the subject can be found in [2] who take a first step in building vine copulas of non-simplified nature.

Following expression (2.2), the likelihood L of a regular vine copula $C = (\mathscr{V}, \mathscr{B}, \boldsymbol{\theta})$ given the observed data $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)' \in \mathbb{R}^{N \times n}$ can be calculated as

$$L(\mathscr{V}, \mathscr{B}, \boldsymbol{\theta} | \mathbf{x}) = \prod_{k=1}^{N} f_{1:n}(\mathbf{x}_k | \mathscr{V}, \mathscr{B}, \boldsymbol{\theta}). \tag{2.4}$$

The corresponding log-likelihood is denoted by ℓ.

A regular vine copula is said to be *truncated at level* M if all pair copulas conditioning on M or more variables are set to bivariate independence copulas [14]. As a result, the iteration index m of the first product in (2.2) runs only up to $m = M$ and the distribution is fully specified by $\big(\mathscr{V} = (T_1, \ldots, T_M), \mathscr{B}(\mathscr{V}), \boldsymbol{\theta}(\mathscr{B}(\mathscr{V}))\big)$.

In the following, we discuss in reverse order how the components of a regular vine copula can be selected and estimated. That is, we begin with estimation of the parameters $\boldsymbol{\theta}$, then treat the selection of appropriate copula families $\mathscr{B}$, and finally discuss the selection of vine trees $\mathscr{V}$.

2.3 Parameter Estimation for Given Vine Tree Structure and Pair Copula Families

Given a vine tree structure $\mathscr{V}$ and pair copula families $\mathscr{B} = \mathscr{B}(\mathscr{V})$, the challenge is to estimate the parameters $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathscr{B}(\mathscr{V}))$ of a regular vine copula for observed data $\mathbf{x} \in \mathbb{R}^{N \times n}$. The crucial point here is to evaluate the conditional distribution functions $F_{i(e)|D(e)}$, which depend on the copulas of previous trees [see (2.3)].

2.3.1 Maximum Likelihood Estimation

Classically, parameters of a statistical model are often estimated using maximum likelihood techniques. Here, this means that regular vine copula parameters $\boldsymbol{\theta}$ and parameters of the marginal distributions are estimated by maximizing the likelihood (2.4) in terms of these parameters. For copulas, in particular if $n > 2$, the number of parameters to be estimated may, however, be too large, so that one typically either uses empirical distribution functions for the margins as proposed by [25, 26] or estimates parameters of the marginal distributions in a first step and then fixes

these marginal parameters to their estimated values in the estimation of the copula parameters. The latter method is called inference functions for margins (IFM) by [38, 39].

But even when estimation of marginal and dependence parameters is separated, joint maximum likelihood estimation of regular vine copula parameters can be computationally intensive, since the vine decomposition involves $n(n-1)/2$ bivariate copulas with corresponding parameters. Aas et al. [1] therefore proposed a sequential method: starting with the copulas of the first tree, this method proceeds tree-wise and estimates the parameters of the copulas in a tree by fixing the parameters of copulas in all previous trees.

Example 2.1 (Sequential Estimation). Let the 5-dimensional regular vine $\mathscr{V}$ of Fig. 2.1 be given with copulas $\mathscr{B}(\mathscr{V})$. In the first step, we estimate the parameters of the copulas $C_{1,2}$, $C_{2,3}$, $C_{3,4}$, and $C_{3,5}$ using maximum likelihood based on the transformed observations $F_j(x_{kj})$ for x_{kj}, $k = 1, \ldots, N$, $j = 1, \ldots, 5$. In the second tree, we then have to estimate, for instance, the parameter(s) of the copula $C_{1,3|2}$. For this we form pseudo-observations $F_{1|2}(x_{k1}|x_{k2}, \hat{\boldsymbol{\theta}}_{1,2})$ and $F_{3|2}(x_{k3}|x_{k2}, \hat{\boldsymbol{\theta}}_{2,3})$, $k = 1, \ldots, N$, according to expression (2.3) and using the estimated parameters of copulas $C_{1,2}$ and $C_{2,3}$, respectively. Based on these pseudo-observations estimation of $\boldsymbol{\theta}_{1,3|2}$ is again straightforward. All other copulas are estimated analogously.

Note that this strategy only involves the estimation of bivariate copulas and therefore is computationally much simpler than joint maximum likelihood estimation of all parameters at once. For more details, see [1] as well as [33] who investigates the asymptotic properties of this sequential approach. A comparison study of estimators for regular vine copulas can be found in [34].

If joint maximum likelihood estimates are desired, the sequential method can be used to obtain starting values for the numerical optimization. A detailed discussion how to compute score functions and the observed information matrix for regular vine copulas is provided by [65].

2.3.2 Bayesian Posterior Estimation

Bayesian statistics considers parameters $\boldsymbol{\theta}$ as random variables. As such, inference focuses on estimating the entire distribution of the parameters instead of only finding a point estimate. In particular, the so-called posterior distribution $p(\boldsymbol{\theta}) := p(\boldsymbol{\theta}|\mathbf{x})$, the distribution of the parameters $\boldsymbol{\theta}$ given the observed data $\mathbf{x}$, is the main object of interest. The unnormalized posterior density factorizes into the product of the likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$ (2.4) and the prior density function $\pi(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}) := p(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\boldsymbol{\theta}|\mathbf{x}) \cdot \pi(\boldsymbol{\theta})}{f(\mathbf{x})} \propto L(\boldsymbol{\theta}|\mathbf{x}) \cdot \pi(\boldsymbol{\theta}).$$

The prior distribution incorporates a priori beliefs about the distribution of the parameters. It must not depend on, or be conditional upon, the observed data.

Markov chain Monte Carlo (MCMC) procedures clear the remaining obstacle of sampling from a distribution which is known only up to a constant. The trick is to simulate the run of a Markov chain whose equilibrium distribution is the targeted posterior distribution $p(\boldsymbol{\theta})$ of the parameters $\boldsymbol{\theta}$. Upon convergence of the Markov chain, its states represent draws from the desired distribution.

The Metropolis–Hastings algorithm [31, 49] implements the simulation of such a Markov chain through an acceptance/rejection mechanism for the updates of the chain: in the first step, an update of the chain from its current state $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^* \sim q(\cdot|\boldsymbol{\theta})$ is proposed. The proposal distribution q can be chosen almost arbitrarily. In the second step, the proposal is accepted with probability $\alpha := \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$. The acceptance probability α is chosen such that convergence of the Markov chain to the targeted distribution is guaranteed by Markov chain theory [50]. The Metropolis–Hastings acceptance probability for convergence to the target distribution $p(\cdot)$ is

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta})} \cdot \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}.$$

The target distribution goes into the Metropolis–Hastings algorithm only at the evaluation of the acceptance probability α. It can be easily seen that it is sufficient to know the density function of the target distribution only up to a constant for this algorithm to work, given that α only depends on the ratio of two densities.

The following Metropolis–Hastings scheme to sample from the posterior distribution of the parameters $\boldsymbol{\theta}$ of a regular vine copula is proposed in [28, 51, 62]. The authors suggest normally distributed random walk proposals be used in the update step.

Algorithm 2.1 (Metropolis–Hastings Sampler for Parameter Estimation).

Choose arbitrary starting values $\boldsymbol{\theta}^0 = (\theta_1^0, \ldots, \theta_D^0)'$, *where* $D := n(n-1)/2$.
for *each iteration* $r = 1, \ldots, R$ ***do***
  Set $\boldsymbol{\theta}^r = \boldsymbol{\theta}^{r-1}$.
  for $i = 1, \ldots, D$ ***do***
    Draw θ_i^r *from a* $N(\theta_i^{r-1}, \sigma_i^2)$ *distribution with probability density function* $\phi_{(\theta_i^{r-1}, \sigma_i^2)}(\cdot)$.
    Evaluate the Metropolis–Hastings acceptance probability of the proposal

$$\alpha := \alpha(\theta_i^{r-1}, \theta_i^r) = \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta})} \cdot \frac{\phi_{(\theta_i^r, \sigma_i^2)}(\theta_i^{r-1})}{\phi_{(\theta_i^{r-1}, \sigma_i^2)}(\theta_i^r)} = \frac{L(\boldsymbol{\theta}^*|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \cdot \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})},$$

    where $\boldsymbol{\theta}^* = (\theta_1^r, \ldots, \theta_i^r, \theta_{i+1}^{r-1}, \ldots, \theta_D^{r-1})'$
    and $\boldsymbol{\theta} = (\theta_1^r, \ldots, \theta_{i-1}^r, \theta_i^{r-1}, \ldots, \theta_D^{r-1})'$.
    Accept the proposal θ_i^r *with probability* α*; if rejected, set* $\theta_i^r = \theta_i^{r-1}$.
  end for

9: ***end for***
10: ***return*** $(\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^R)$

2.4 Selection of the Pair Copula Families and their Parameters for a Known Vine Tree Structure

An n-dimensional regular vine copula with tree structure $\mathcal{V}$ is based on a set $\mathcal{B}(\mathcal{V})$ of $n(n-1)/2$ bivariate copulas and their corresponding parameters. The copula families can be chosen arbitrarily, e.g., from the popular classes of Archimedean, elliptical or extreme-value copulas. Assuming that an appropriate vine structure $\mathcal{V}$ is chosen, the question is how to select adequate (conditional) pair copulas $C_{i(e),j(e)|D(e)}$ and their parameters for given data $\mathbf{x}$.

2.4.1 Sequential Selection

If a regular vine is truncated at level 1, the copulas in trees T_2 to T_{n-1} are set to independence. The corresponding regular vine copula density reduces to the product of the unconditional bivariate copula densities $C_{i(e),j(e)}$, which can be selected based on data $F_{i(e)}(x_{k,i(e)})$ and $F_{j(e)}(x_{k,j(e)})$, $k = 1, \ldots, N$. Typical criteria for copula selection from a given set of families are information criteria such as the AIC as proposed by [11, 47]. The latter compares AIC-based selection to three alternative selection strategies: selection of family with highest p-value of a copula goodness of fit test based on the Cramér–von Mises statistic, with smallest distance between empirical and modeled dependence characteristics (Kendall's τ, tail dependence), or with highest number of wins in pairwise comparisons of families using the test by [66]. In a large-scale Monte Carlo study the AIC turned out to be the most reliable selection criterion.

In a general regular vine, the selection of a pair copula $C_{i(e),j(e)|D(e)}$, however, depends on the choices for copulas in previous trees due to its arguments (2.3). Since a joint selection seems infeasible because of the many possibilities, one typically proceeds tree-by-tree as in the sequential estimation method. That is, instead of estimating the parameters $\boldsymbol{\theta}_{i(e),j(e)|D(e)}$ of $C_{i(e),j(e)|D(e)}$, the copula is selected first and then estimated, which usually coincides for most selection strategies. Given the selected and estimated copulas of previous trees, one then selects the copulas of the next tree.

Example 2.2 (Sequential Selection). Again consider the five-dimensional regular vine $\mathcal{V}$ of Fig. 2.1 but now with unknown copulas $\mathcal{B}(\mathcal{V})$. In the first tree, we select (and then estimate) the copulas $C_{1,2}$, $C_{2,3}$, $C_{3,4}$, and $C_{3,5}$ using our method of choice based on $F_j(x_{kj})$, $k = 1, \ldots, N$, $j = 1, \ldots, 5$. Given these copulas, we then have to select conditional copulas in the second tree. In case of the copula

$C_{1,3|2}$, we therefore again form the pseudo-observations $F_{1|2}(x_{k1}|x_{k2}, \hat{\boldsymbol{\theta}}_{1,2})$ and $F_{3|2}(x_{k3}|x_{k2}, \hat{\boldsymbol{\theta}}_{2,3})$, $k = 1, \ldots, N$, according to expression (2.3) and then select (and estimate) $C_{1,3|2}$ based on them. This can be iterated for all trees.

Clearly this sequential selection strategy accumulates uncertainty in the selection and therefore the final model has to be carefully checked and compared to alternative models. For the latter, the tests for non-nested model comparison by [18, 66] may be used.

2.4.2 Reversible Jump MCMC-Based Bayesian Selection

Bayesian copula family and parameter selection aims at estimating the joint posterior distribution of the pair copula families $\mathscr{B} = \mathscr{B}(\mathscr{V})$ and parameters $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathscr{B}(\mathscr{V}))$. The posterior density function factorizes into the product of the likelihood function L (2.4) and the prior density function π: $p(\mathscr{B}, \boldsymbol{\theta}) \propto L(\mathscr{B}, \boldsymbol{\theta}|\mathbf{x}) \cdot \pi(\mathscr{B}, \boldsymbol{\theta})$.

As in Sect. 2.4.1, family selection is understood in the context of choosing from a pre-specified set $\mathbf{B}$ of parametric pair copula families. Model sparsity can be induced through the choice of prior distributions which favor regular vine copulas with fewer parameters over those with more parameters.

Bayesian techniques to select the pair copula families of D-vine copulas are covered in [51, 52, 61]. The reversible jump MCMC algorithm presented in this section follows the latest developments of family selection methods for regular vine copulas laid out in [28].

Reversible jump MCMC is an extension of ordinary MCMC to sample from discrete-continuous posterior distributions. The sampling algorithm and the mathematics underpinning the convergence statements are an immediate generalization of the Metropolis–Hastings algorithm [27]. A reversible jump MCMC algorithm functions like an ordinary MCMC algorithm with one extra step built in: before, or after, updating the parameters, a "jump" to another model is attempted. That is called the "between-models move," while the updating of the parameters only is called the "within-model move" [16].

Algorithm 2.2 implements a reversible jump MCMC sampler which samples from the joint posterior distribution of the pair copula families $\mathscr{B}$ and the parameters $\boldsymbol{\theta}$, given a regular vine tree structure $\mathscr{V}$. In each iteration, it first updates all parameters as in Algorithm 2.1. Then the pair copula families are updated along with their parameters one-by-one. The pair copula family proposals are drawn from a uniform distribution over $\mathbf{B}$, while the parameter proposals are sampled from a (multivariate) normal distribution centered at the maximum likelihood estimates of the parameters.

Algorithm 2.2 (Reversible Jump MCMC Sampler for Family Selection).

Choose arbitrary copula family and parameter starting values $(\mathscr{B}^0, \boldsymbol{\theta}^0) = ((\mathscr{B}_1^0, \boldsymbol{\theta}_1^0), \ldots, (\mathscr{B}_D^0, \boldsymbol{\theta}_D^0))$, *where* $D := n(n-1)/2$.
for *each iteration* $r = 1, \ldots, R$ ***do***
  Set $(\mathscr{B}^r, \boldsymbol{\theta}^r) = (\mathscr{B}^{r-1}, \boldsymbol{\theta}^{r-1})$.
  Perform one update step of Algorithm 2.1 for the parameters $\boldsymbol{\theta}^r$*; denote the updated parameter entries by* $\boldsymbol{\theta}^+$.
  for $i = 1, \ldots, D$ ***do***
    Draw $\mathscr{B}_i^r$ *from a Unif*(**B**) *distribution.*
    Draw $\boldsymbol{\theta}_i^r$ *from a multivariate* $N(\hat{\boldsymbol{\theta}}_i^r, \hat{\Sigma}_i^r)$ *distribution with probability density function* $\phi_{(\hat{\boldsymbol{\theta}}_i^r, \hat{\Sigma}_i^r)}(\cdot)$. *Here* $\hat{\boldsymbol{\theta}}_i^r$ *denotes the MLE of the copula parameters of the pair copula* $\mathscr{B}_i^r$ *and* $\hat{\Sigma}_i^r$ *denotes the estimated approximative covariance matrix of the parameter estimates* $\hat{\boldsymbol{\theta}}_i^r$.
    Evaluate the generalized Metropolis–Hastings acceptance probability of the proposal

$$\begin{aligned}\alpha &:= \alpha((\mathscr{B}_i^{r-1}, \boldsymbol{\theta}_i^+), (\mathscr{B}_i^r, \boldsymbol{\theta}_i^r)) \\ &= \frac{p(\mathscr{B}^*, \boldsymbol{\theta}^*)}{p(\mathscr{B}, \boldsymbol{\theta})} \cdot \frac{\phi_{(\hat{\boldsymbol{\theta}}_i^{r-1}, \hat{\Sigma}_i^{r-1})}(\boldsymbol{\theta}_i^+)}{\phi_{(\hat{\boldsymbol{\theta}}_i^r, \hat{\Sigma}_i^r)}(\boldsymbol{\theta}_i^r)} \\ &= \frac{\pi(\mathscr{B}^*, \boldsymbol{\theta}^*)}{\pi(\mathscr{B}, \boldsymbol{\theta})} \cdot \frac{L(\mathscr{B}^*, \boldsymbol{\theta}^* | \mathbf{x})}{L(\mathscr{B}, \boldsymbol{\theta} | \mathbf{x})} \cdot \frac{\phi_{(\hat{\boldsymbol{\theta}}_i^{r-1}, \hat{\Sigma}_i^{r-1})}(\boldsymbol{\theta}_i^+)}{\phi_{(\hat{\boldsymbol{\theta}}_i^r, \hat{\Sigma}_i^r)}(\boldsymbol{\theta}_i^r)},\end{aligned}$$

    where $\mathscr{B}^* = (\mathscr{B}_1^r, \ldots, \mathscr{B}_i^r, \mathscr{B}_{i+1}^{r-1}, \ldots, \mathscr{B}_D^{r-1})$,
    $\mathscr{B} = (\mathscr{B}_1^r, \ldots, \mathscr{B}_{i-1}^r, \mathscr{B}_i^{r-1}, \ldots, \mathscr{B}_D^{r-1})$, $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^r, \ldots, \boldsymbol{\theta}_i^r, \boldsymbol{\theta}_{i+1}^+, \ldots, \boldsymbol{\theta}_D^+)$, *and* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^r, \ldots, \boldsymbol{\theta}_{i-1}^r, \boldsymbol{\theta}_i^+, \ldots, \boldsymbol{\theta}_D^+)$.
    Accept the proposal $(\mathscr{B}_i^r, \theta_i^r)$ *with probability* α*; if rejected, set* $(\mathscr{B}_i^r, \theta_i^r) = (\mathscr{B}_i^{r-1}, \theta_i^+)$.
  end for
end for
return $((\mathscr{B}^1, \boldsymbol{\theta}^1), \ldots, (\mathscr{B}^R, \boldsymbol{\theta}^R))$

2.5 Selection of Vine Tree Structure, Pair Copula Families, and Parameters

As pair copula families $\mathscr{B} = \mathscr{B}(\mathscr{V})$ and parameters $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathscr{B}(\mathscr{V}))$ both depend on the vine tree structure $\mathscr{V}$, the identification of adequate trees is crucial to the model selection of vine copulas. As it was already the case for pair copula selection in Sect. 2.4.1, it is again not feasible to simply try and fit all possible regular vine

copulas $\boldsymbol{C} = (\mathscr{V}, \mathscr{B}, \boldsymbol{\theta})$ and then choose the "best" one. In particular, the number of possible regular vines on n variables is $\frac{n!}{2} \times 2^{\binom{n-2}{2}}$ as shown by [53]. This means that even if pair copulas and parameters were known, the number of different models would still be enormous.

This remains true even when the selection is restricted to the sub-classes of C- and D-vines, since there are still $n!/2$ different C- and D-vines in n dimensions, respectively (see [1]). It should, however, be noted that C- and D-vine copulas are most appropriate if their structure is explicitly motivated by the data. In particular, C-vine copulas may be used if there is a set of pivotal variables such as stock indices (see [12, 32]) and D-vine copulas are particularly attractive to model variables with temporal order (see [13, 61]). Nevertheless we describe how C- and D-vines can be selected for arbitrary data sets.

2.5.1 Top-Down and Bottom-Up Selection

Due to the proximity condition (see Definition 2.2) regular vine trees are closely linked with each other and have to be constructed carefully. Two construction strategies have been proposed in the literature: a top-down approach by [23] and a bottom-up method by [43]. Both strategies proceed sequentially tree-by-tree and respect the proximity condition in each step. We first describe the top-down, then the bottom-up method.

2.5.1.1 Top-Down Selection

Selecting regular vine trees top-down means that one starts with the selection of the first tree T_1 and continues tree-by-tree up to the last tree T_{n-1}. The first tree T_1 can be selected as an arbitrary spanning tree. Given that a tree T_m, $m \in \{1, \ldots, n-2\}$, has been selected, the next tree T_{m+1} is chosen respecting the proximity condition (see Definition 2.2). In other words, T_{m+1} can only be formed by (conditional) pairs $\{i(e), j(e)|D(e)\}$ which fulfill the proximity condition.

Example 2.3 (Top-Down Tree Selection). Assume that we have selected the first tree T_1 as shown in Fig. 2.1. Then the question is which pairs $\{i(e), j(e)|D(e)\}$ are eligible for tree construction in the second tree T_2. According to the proximity condition these are $\{1, 3|2\}$, $\{2, 4|3\}$, $\{2, 5|3\}$, and $\{4, 5|3\}$. Obviously, the last three pairs form a cycle and therefore only two of them can be selected for T_2. One of the three possibilities is shown in Fig. 2.1.

To perform this iterative selection strategy a criterion is needed to select a spanning tree among the set of eligible edges, where a spanning tree simply denotes a tree on all nodes. Clearly, the log-likelihood ℓ_m of the pair copulas in tree T_m of a regular vine copula [see expression (2.4)] is given by

$$\ell_m\left(T_m, \mathscr{B}_{T_m}, \boldsymbol{\theta}_{T_m}|\mathbf{x}\right) = \sum_{k=1}^{N} \sum_{e\in E_m} \log\left(c_{i(e),j(e)|D(e)} \times \left(F_{i(e)|D(e)}, F_{j(e)|D(e)}|\boldsymbol{\theta}_{i(e),j(e)|D(e)}\right)\right), \tag{2.5}$$

where we write $\mathscr{B}_{T_m} := \mathscr{B}(T_m)$ and $\boldsymbol{\theta}_{T_m} := \boldsymbol{\theta}(\mathscr{B}(T_m))$.

A straightforward solution therefore would be to choose the tree such that (2.5) is maximized after having selected pair copulas with high log-likelihood for each (conditional) pair $\{i(e), j(e)|D(e)\}$ that fulfills the proximity condition. This solution, however, leads to highly over-parameterized models, since models with more parameters in which simpler models are nested will always give a higher likelihood. For instance, the Student's t copula always has a higher likelihood than the Gaussian copula which is a special case of the Student's t as the degrees of freedom tend to infinity.

Therefore we formulate the following algorithm in terms of a general weight ω assuming that we want to maximize it for each tree. The previously discussed strategy corresponds to choosing the pair copula log-likelihoods as weights.

Algorithm 2.3 (Sequential Top-Down Selection Based on Weights).

Calculate the weight $\omega_{i,j}$ for all possible variable pairs $\{i, j\}, 1 \le i < j \le n$.
Select the maximum spanning tree, i.e.

$$T_1 = \underset{T=(N,E)\ spanning\ tree}{\operatorname{argmax}} \sum_{e\in E} \omega_{i(e),j(e)}.$$

for** each edge $e \in E_1$ **do
  Select a copula $C_{i(e),j(e)}$.
  Estimate the corresponding parameter(s) $\boldsymbol{\theta}_{i(e),j(e)}$.
  For $k = 1, \ldots, N$ transform $F_{i(e)|j(e)}(x_{k,i(e)}|x_{k,j(e)}, \hat{\boldsymbol{\theta}}_{i(e),j(e)})$ and $F_{j(e)|i(e)}(x_{k,j(e)}|x_{k,i(e)}, \hat{\boldsymbol{\theta}}_{i(e),j(e)})$ using (2.3).
end for
for** $m = 2, \ldots, n-1$ **do
  Calculate the weights $\omega_{i(e),j(e)|D(e)}$ for all conditional variable pairs $\{i(e), j(e)|D(e)\}$ that can be part of tree T_m. We denote this set of edges which fulfill the proximity condition by E_P.
  Among these edges, select the maximum spanning tree, i.e.,

$$T_m = \underset{T=(N,E)spanning\ tree\ with\ E\subset E_P}{\operatorname{argmax}} \sum_{e\in E} \omega_{i(e),j(e)|D(e)}.$$

  for** each edge $e \in E_m$ **do
    Select a conditional copula $C_{i(e),j(e)|D(e)}$.
    Estimate the corresponding parameter(s) $\boldsymbol{\theta}_{i(e),j(e)|D(e)}$.

14: *For* $k = 1, \ldots, N$ *transform* $F_{i(e)|j(e)\cup D(e)}(x_{k,i(e)}|x_{k,j(e)}, \mathbf{x}_{k,D(e)}, \hat{\boldsymbol{\theta}}_{i(e),j(e)|D(e)})$ *and* $F_{j(e)|i(e)\cup D(e)}(x_{k,j(e)}|x_{k,i(e)}, \mathbf{x}_{k,D(e)}, \hat{\boldsymbol{\theta}}_{i(e),j(e)|D(e)})$ *using* (2.3).

15: ***end for***

16: ***end for***

17: ***return*** *the sequential model estimate* $(\hat{\mathscr{V}}, \hat{\mathscr{B}}, \hat{\boldsymbol{\theta}})$.

Clearly, this algorithm only makes a locally optimal selection in each step, since the impact on previous and subsequent trees is ignored. The strategy is, however, reasonable in light of the definition of regular vines (see Definition 2.2). The maximum spanning tree in lines 2 and 10 can be found, e.g., using the classical algorithms by Prim or Kruskal (see [19, Section 23.2]). Possible choices for the weight ω are, for example:

- The absolute empirical Kendall's τ as proposed by [22, 23].
- The AIC of each pair copula corresponding to the discussion in Sect. 2.4.1.
- The (negative) estimated degrees of freedom of Student's t pair copulas (see [48]).
- The p-value of a copula goodness of fit test and variants as proposed by [21].

Some remarks: First, taking the empirical Kendall's τ as weight does not require to select and estimate pair copulas prior to the tree selection step. The other three weights require this, so that lines 4–5 and 12–13 in Algorithm 2.3 may be redundant in this case. Second, AIC weights also maximize the AIC of the entire tree, since the individual AICs sum like the log-likelihood. Third, the strategy proposed by [48] concentrates on tail dependence, since the algorithm will select pairs with estimated small degrees of freedom corresponding to a stronger deviation from Gaussianity with no tail dependence.

Copula goodness of fit tests based on the Cramér–von Mises statistic are considered in [21] to take into account the uncertainty of the pair copula fit. Corresponding p-values are calculated using fast bootstrap methods based on the multiplier approach developed and implemented in [40–42].

Finally, we like to note that the empirical Kendall's τ as weight also approximatively maximizes (2.5), since the log-likelihoods of bivariate copulas tend to increase with increasing absolute values of Kendall's τ. It, however, does not lead to over-parameterization, since copula selection and tree selection in a particular tree are independent. The strategy by [23] has therefore already been used successfully in applications with up to 52 variables (see [12]).

Recently, [36] proposed the construction of vines with non-parametric pair copulas based on empirical copula estimators. For vine tree selection [36] also advocate the use of empirical dependence measures such as Spearman's ρ, since it does not require any parametric assumption on the pair copulas. Other dependence measures such as Blomqvist's β could of course be used instead in the approaches by [23, 36].

Especially the strategy based on Kendall's τ leads to regular vine copulas with decreasing dependence in higher order trees. This is advantageous to the vine

structure, since numerical errors may add up and estimates become less precise as the number of conditioning variables increases. This has been exploited by [14] to identify appropriate truncation levels of regular vine copulas.

Algorithm 2.3 can easily be modified to select C- or D-vines instead of general regular vines. For C-vines the root node in each tree can simply be identified as the node with maximal sum of weights to all other nodes (see [22]). In the case of D-vines only the order of variables in the first tree has to be chosen. Since D-vine trees are paths on all nodes, so-called Hamiltonian paths, a maximum Hamiltonian path has to be identified. This problem is equivalent to a traveling salesman problem as discussed by [11]. As an NP-hard problem, there is no known efficient algorithm to find a solution.

2.5.1.2 Bottom-Up Selection

Rather than beginning with the first tree T_1, [43] proposed a bottom-up selection strategy which starts with tree T_{n-1} and then sequentially selects trees T_m for $m = n-2, \ldots, 1$. Similar to top-down selection using Kendall's τ by [23], this approach is also motivated by choosing a regular vine with weaker dependence in later trees. Here, dependence is, however, measured in terms of partial correlations. More precisely, each tree is selected such that the sum of absolute partial correlations as edge weights is minimized. This is feasible, since partial correlation estimates for each combination of variables can be obtained from the data without any parametric assumptions (see, e.g., [67]).

The bottom-up strategy by [43] thus selects tree T_{n-1} as the edge corresponding to the pair of variable with lowest absolute partial correlation given all other variables. Kurowicka [43] then provides conditions such that the proximity condition is satisfied when a tree T_m, $m \in \{1, \ldots, n-2\}$, is selected given trees $T_{m+1}, \ldots, T_{n-1}$. As before there may be several choices possible in the set of eligible edges.

Example 2.4 (Bottom-Up Tree Selection). Assume in Fig. 2.1 that we have selected T_4 with edge $e = \{a, b\} = \{1, 5|2, 3, 4\}$ as shown. Then $A_a = \{1, 2, 3, 4\}$ and $A_b = \{2, 3, 4, 5\}$ [see (2.1)]. Any choice of edges from A_a and A_b in tree T_3 leads to a valid tree T_4. For instance, the edges $\{1, 2|3, 4\}$ and $\{4, 5|2, 3\}$ are compatible with A_a and A_b, respectively, and would lead to another vine than the one shown in Fig. 2.1. In fact, it could be constructed as a D-vine with order $1, 4, 3, 2, 5$.

Since partial correlations are equal to conditional correlations for elliptical distributions (see [5]), this strategy is particularly appropriate if all pair copula families are elliptical. In contrast to the top-down approach discussed above which requires the selection and estimation of pair copulas in each tree, this strategy chooses the vine tree structure without any assumption on the pair copula families. Having selected the vine tree structure, the selection of pair copulas can therefore proceed as discussed in Sect. 2.4.1.

2.5.2 Sequential Bayesian Tree Selection

Bayesian approaches to estimating the posterior distribution of the tree structure of a regular vine copula have only recently been developed. The sheer number of possible regular vine tree structures poses a relevant challenge to computationally efficient posterior evaluation.

This section presents a reversible jump MCMC-based approach proposed by [29] to obtain a sequential estimate of the posterior distribution of the regular vine tree structure $\mathscr{V}$, the pair copula families $\mathscr{B}$, and the copula parameters $\boldsymbol{\theta}$. The term "sequential estimation" is understood analogously to the notion discussed in Sect. 2.5.1.1 as a tree-by-tree procedure.

As in Sect. 2.4.2, model priors $\pi(\mathscr{V}, \mathscr{B}, \boldsymbol{\theta})$ which favor sparse models can serve to guard against selecting models with runaway complexity. On the other hand, the use of non-informative flat priors allows for tree-by-tree maximum likelihood estimation of the regular vine tree structure $\mathscr{V}$, the pair copula families $\mathscr{B}$, and the copula parameters $\boldsymbol{\theta}$. As a result, the posterior mode estimate of this procedure will agree with the model estimate of Algorithm 2.3, if flat priors are used. Again, the pair copula families are chosen from a set $\mathbf{B}$ of parametric pair copula families.

For the sake of notational convenience and enhanced readability, the procedure is presented in two algorithms. Algorithm 2.4 gives the general outline of the sampling procedure. Algorithm 2.5 details the reversible jump MCMC algorithm to sample from the posterior distribution of tree T_m of the regular vine tree structure $\mathscr{V} = (T_1, \ldots, T_{n-1})$.

Algorithm 2.4 implements the tree-by-tree selection procedure. As such, it calls Algorithm 2.5 for the actual posterior estimation. However, it condenses the posterior sample into the posterior mode estimate and organizes the move from selecting tree T_m to selecting tree T_{m+1}. The posterior mode estimate $(\hat{T}_m, \hat{\mathscr{B}}_{T_m}, \hat{\boldsymbol{\theta}}_{T_m})$ of tree T_m is the most frequently sampled combination of tree structure T_m and pair copula families $\mathscr{B}_{T_m}$ parameterized at the mode of the posterior sample of the parameters $\boldsymbol{\theta}_{T_m}$ of this model.

Algorithm 2.4 (Outline of the Tree-by-Tree Sampling Algorithm).

Sample from the posterior distribution of the first tree, $(T_1, \mathscr{B}_{T_1}, \boldsymbol{\theta}_{T_1})$*: see Algorithm 2.5.*
Set the tree estimate $(\hat{T}_1, \hat{\mathscr{B}}_{T_1}, \hat{\boldsymbol{\theta}}_{T_1})$ *to the posterior mode.*
for $m = 2, \ldots, n-1$ ***do***
    Sample from the posterior distribution of the m-th tree, $(T_m, \mathscr{B}_{T_m}, \boldsymbol{\theta}_{T_m})$*, given the previous trees' estimates* $(\hat{T}_1, \hat{\mathscr{B}}_{T_1}, \hat{\boldsymbol{\theta}}_{T_1}), \ldots, (\hat{T}_{m-1}, \hat{\mathscr{B}}_{T_{m-1}}, \hat{\boldsymbol{\theta}}_{T_{m-1}})$*: see Algorithm 2.5.*
    Set the tree estimate $(\hat{T}_m, \hat{\mathscr{B}}_{T_m}, \hat{\boldsymbol{\theta}}_{T_m})$ *to the posterior mode.*
end for
return *the sequential Bayesian model estimate*

$$(\hat{\mathscr{V}}, \hat{\mathscr{B}}, \hat{\boldsymbol{\theta}}) = ((\hat{T}_1, \ldots, \hat{T}_{n-1}), (\hat{\mathscr{B}}_{T_1}, \ldots, \hat{\mathscr{B}}_{T_{n-1}}), (\hat{\boldsymbol{\theta}}_{T_1}, \ldots, \hat{\boldsymbol{\theta}}_{T_{n-1}})).$$

Algorithm 2.5 samples from the posterior distribution of $(T_m, \mathscr{B}_{T_m}, \boldsymbol{\theta}_{T_m})$ given the previously selected trees $((\hat{T}_l, \hat{\mathscr{B}}_{T_l}, \hat{\boldsymbol{\theta}}_{T_l}), l = 1, \ldots, m-1)$. The within-model move of this algorithm to update the pair copula parameters $\boldsymbol{\theta}$ follows Algorithm 2.1. In the between-models move, the proposal trees are sampled as spanning trees which satisfy the proximity condition. Practitioners may note that this functionality is implemented, e.g., in the C++ `boost` library [59]. Edge weights can be used to fine-tune the variance of the proposal distribution: in our implementation, higher values for $p \in (0, 1)$ increase the probability that the proposals for the tree structure, T_m^r, are similar to the current state T_m^{r-1} of the sampling chain. As in Algorithm 2.2, the pair copula family proposals are sampled from a uniform distribution over $\mathbf{B}$, and the parameter proposals are drawn from a (multivariate) normal distribution centered at the maximum likelihood estimates of the parameters.

Algorithm 2.5 (Reversible Jump MCMC Algorithm to Sample $(T_m, \mathscr{B}_{T_m}, \boldsymbol{\theta}_{T_m})$).

Denote the pair copula families of tree $T_m = (N_m, E_m)$ by $\mathscr{B}_{T_m} = (\mathscr{B}_e, e \in E_m)$, and the corresponding pair copula parameters by $\boldsymbol{\theta}_{T_m} = (\boldsymbol{\theta}_e, e \in E_m)$.

1: *Choose arbitrary, but valid, tree structure, copula family and parameter starting values $(T_m^0, \mathscr{B}_{T_m}^0, \boldsymbol{\theta}_{T_m}^0)$. If $m \geq 2$, observe the proximity condition imposed on T_m^0.*
2: ***for** each iteration $r = 1, \ldots, R$ **do***
3: *Set $(T_m^r, \mathscr{B}_{T_m}^r, \boldsymbol{\theta}_{T_m}^r) = (T_m^{r-1}, \mathscr{B}_{T_m}^{r-1}, \boldsymbol{\theta}_{T_m}^{r-1})$.*
4: *Perform one update step of Algorithm 2.1 for the pair copula parameters of tree T_m, $\boldsymbol{\theta}_{T_m}^r$; denote the updated parameter entries by $\boldsymbol{\theta}_{T_m}^+$.*
5: *Draw a spanning tree T_m^r that satisfies the proximity condition from the proposal distribution*

$$q(T_m^{r-1} = (N_m, E_m^{r-1}) \to T_m^r = (N_m, E_m^r)) \propto p^{|E_m^r \cap E_m^{r-1}|}(1-p)^{|E_m^r \setminus E_m^{r-1}|}.$$

6: ***for** each edge $e \in E_m^r$ **do***
7: *Draw $\mathscr{B}_e^r$ from a Unif$(\mathbf{B})$ distribution.*
8: *Draw $\boldsymbol{\theta}_e^r$ from a multivariate $N(\hat{\boldsymbol{\theta}}_e^r, \hat{\Sigma}_e^r)$ distribution with probability density function $\phi_{(\hat{\boldsymbol{\theta}}_e^r, \hat{\Sigma}_e^r)}(\cdot)$. Here $\hat{\boldsymbol{\theta}}_e^r$ denotes the MLE of the copula parameters of the pair copula $\mathscr{B}_e^r$ and $\hat{\Sigma}_e^r$ denotes the estimated approximative covariance matrix of the parameter estimates $\hat{\boldsymbol{\theta}}_e^r$.*
9: ***end for***
10: *Set $\mathscr{B}_{T_m}^r = (\mathscr{B}_e^r, e \in E_m^r)$, $\boldsymbol{\theta}_{T_m}^r = (\boldsymbol{\theta}_e^r, e \in E_m^r)$.*
11: *Evaluate the generalized Metropolis–Hastings acceptance probability of the proposal*

$$\begin{aligned}\alpha &:= \alpha((T_m^{r-1}, \mathscr{B}_{T_m}^{r-1}, \boldsymbol{\theta}_{T_m}^+), (T_m^r, \mathscr{B}_{T_m}^r, \boldsymbol{\theta}_{T_m}^r)) \\ &= \frac{p(\mathscr{V}^*, \mathscr{B}^*, \boldsymbol{\theta}^*)}{p(\mathscr{V}, \mathscr{B}, \boldsymbol{\theta})} \cdot \frac{\prod_{e \in E_m^{r-1}} \phi_{(\hat{\boldsymbol{\theta}}_e^{r-1}, \hat{\Sigma}_e^{r-1})}(\boldsymbol{\theta}_e^+)}{\prod_{e \in E_m^r} \phi_{(\hat{\boldsymbol{\theta}}_e^r, \hat{\Sigma}_e^r)}(\boldsymbol{\theta}_e^r)}\end{aligned}$$

$$= \frac{\pi(\mathcal{V}^*, \mathcal{B}^*, \boldsymbol{\theta}^*)}{\pi(\mathcal{V}, \mathcal{B}, \boldsymbol{\theta})} \cdot \frac{L(\mathcal{V}^*, \mathcal{B}^*, \boldsymbol{\theta}^* | \mathbf{x})}{L(\mathcal{V}, \mathcal{B}, \boldsymbol{\theta} | \mathbf{x})} \cdot \frac{\prod_{e \in E_m^{r-1}} \phi_{(\hat{\boldsymbol{\theta}}_e^{r-1}, \hat{\Sigma}_e^{r-1})}(\boldsymbol{\theta}_e^+)}{\prod_{e \in E_m^r} \phi_{(\hat{\boldsymbol{\theta}}_e^r, \hat{\Sigma}_e^r)}(\boldsymbol{\theta}_e^r)},$$

where $\mathcal{V}^* = (\hat{T}_1, \ldots, \hat{T}_{m-1}, T_m^r)$, $\mathcal{V} = (\hat{T}_1, \ldots, \hat{T}_{m-1}, T_m^{r-1})$,
$\mathcal{B}^* = (\hat{\mathcal{B}}_{T_1}, \ldots, \hat{\mathcal{B}}_{T_{m-1}}, \mathcal{B}_{T_m}^r)$, $\mathcal{B} = (\hat{\mathcal{B}}_{T_1}, \ldots, \hat{\mathcal{B}}_{T_{m-1}}, \mathcal{B}_{T_m}^{r-1})$,
$\boldsymbol{\theta}^* = (\hat{\boldsymbol{\theta}}_{T_1}, \ldots, \hat{\boldsymbol{\theta}}_{T_{m-1}}, \boldsymbol{\theta}_{T_m}^r)$, *and* $\boldsymbol{\theta} = (\hat{\boldsymbol{\theta}}_{T_1}, \ldots, \hat{\boldsymbol{\theta}}_{T_{m-1}}, \boldsymbol{\theta}_{T_m}^+)$.

Accept the proposal $(T_m^r, \mathcal{B}_{T_m}^r, \boldsymbol{\theta}_{T_m}^r)$ *with probability* α*; if rejected, set* $(T_m^r, \mathcal{B}_{T_m}^r, \boldsymbol{\theta}_{T_m}^r) = (T_m^{r-1}, \mathcal{B}_{T_m}^{r-1}, \boldsymbol{\theta}_{T_m}^+)$.

end for

return $((T_m^1, \mathcal{B}_{T_m}^1, \boldsymbol{\theta}_{T_m}^1), \ldots, (T_m^R, \mathcal{B}_{T_m}^R, \boldsymbol{\theta}_{T_m}^R))$

2.6 Conclusions and Outlook

We have focused on the various model selection and estimation methods for regular vine copulas. Since a regular vine model is specified by three linked components, this results in three fundamental tasks with increasing complexity. We discussed frequentist and Bayesian approaches for each of these tasks. In particular this involved sequential approaches starting from the top tree until the last tree. The frequentist approaches are implemented in the R-packages `CDVine` [15, 57] for D- and C-vines and `VineCopula` [58] for regular vines, respectively.

In view of the linked nature of the vine tree structure the sequential approach is a natural approach. However as in the case of covariate selection in linear models, this might not yield the best fit to the data. In addition the approaches so far primarily considered in-sample fit and model comparisons. More empirical work is needed to validate the models in an out-of-sample performance study. However this is now feasible and tractable and is the subject of current investigations.

It has been recently recognized by [2] that even the flexible class of simplified regular vines might be insufficient for some data sets. This might be the result that the underlying joint density is not well approximated by a simplified regular vine density, where the conditional copula family terms are chosen to be independent of the conditioning value. Currently the proposed solution of [2] using non-parametric two-dimensional smoothing methods is limited to three dimensions and it will be a major challenge to extend the model and the selection methods to higher dimensions.

Other non-standard vine models occur when the pair copula terms depend on the conditioning time point, thus yielding models with time-varying copula parameters. These can also be seen as non-simplified vines in the special case that the conditioning variables follow a functional relationship to time. First parameter-driven time dependence was considered using an AR(1) dynamics in the copula parameters in the papers by [3, 4], while [62] follow a regime switching approach. Here only parameter estimation and assessment of the uncertainty of the parameters

are considered so far. It has to be investigated if the additional flexibility of the copula families and the different vine tree structures is needed here.

This survey primarily focused on the selection and estimation problem of regular vines with parametric pair copula families. The approach of [36] can be compared to ones based on kernel methods as proposed in [7, 56].

Other data structures than multivariate continuous data have also gained by models based on pair copula constructions. In particular network structures were considered by [30, 44] in a primarily Gaussian setup and by [7] in a non-Gaussian setting. Here the network is modeled by a directed acyclic graph (DAG) model. While learning the network structure from data is a very complex task even in Gaussian DAG models, non-Gaussian learning algorithms are currently developed and investigated in [6].

Another very interesting multivariate data structure are discrete and mixed discrete-continuous outcomes which occur most often in the life sciences. Pair copula constructions using D-vines for discrete outcomes were recently developed in [55] together with highly efficient parameter estimation techniques. Current research is conducted to allow for mixed outcomes [63].

While these extensions are still concerned primarily with parameter estimation, it remains an important open challenge to find non-sequential solutions to the selection of regular vines when all three components have to be selected.

Acknowledgments The second and third author gratefully acknowledge the support of the TUM Graduate School's International School of Applied Mathematics. The second author is also supported by a research stipend from Allianz Deutschland AG.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insur. Math. Econ. **44**, 182–198 (2009)
2. Acar, E.F., Genest, C., Nešlehová, J.: Beyond simplified pair-copula constructions. J. Multivar. Anal. **110**, 74–90 (2012)
3. Almeida, C., Czado, C.: Efficient Bayesian inference for stochastic time-varying copula models. Comput. Stat. Data Anal. **56**(6), 1511–1527 (2012)
4. Almeida, C., Czado, C., Manner, H.: Modeling high dimensional time-varying dependence using D-vine SCAR models. http://arxiv.org/abs/1202.2008 (2012, preprint)
5. Baba, K., Shibata, R., Sibuya, M.: Partial correlation and conditional correlation as measures of conditional independence. Aust. N. Z. J. Stat. **46(4)**, 657–664 (2004)
6. Bauer, A., Czado, C.: Pair copula Bayesian networks. http://arxiv.org/abs/1211.5620 (2012, preprint)
7. Bauer, A., Czado, C., Klein, T.: Pair-copula constructions for non-Gaussian DAG models. Can. J. Stat. **40**(1), 86–109 (2012)
8. Bedford, T., Cooke, R.M.: Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. Intell. **32**, 245–268 (2001)
9. Bedford, T., Cooke, R.M.: Vines: a new graphical model for dependent random variables. Ann. Stat. **30**(4), 1031–1068 (2002)

10. Berg, D., Aas, K.: Models for construction of higher-dimensional dependence: a comparison study. Eur. J. Financ. **15**, 639–659 (2009)
11. Brechmann, E.C.: Truncated and simplified regular vines and their applications. Master's thesis, Technische Universität München (2010)
12. Brechmann, E.C., Czado, C.: Risk management with high-dimensional vine copulas: An analysis of the Euro Stoxx 50. Stat. Risk Model. (2013, in press)
13. Brechmann, E.C., Czado, C.: COPAR - Multivariate time series modeling using the COPula AutoRegressive model. http://arxiv.org/abs/1203.3328 (2012, preprint)
14. Brechmann, E.C., Czado, C., Aas, K.: Truncated regular vines in high dimensions with application to financial data. Can. J. Stat. **40**(1), 68–85 (2012)
15. Brechmann, E.C., Schepsmeier, U.: Modeling dependence with C- and D-vine copulas: the R-package CDVine. J. Stat. Softw. **52**(3), 1–27 (2012)
16. Brooks, S., Gelman, A., Jones, G.L., Meng, X.L.: Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC, Boca Raton (2011)
17. Chollete, L., Heinen, A., Valdesogo, A.: Modeling international financial returns with a multivariate regime switching copula. J. Financ. Econom. **7**, 437–480 (2009)
18. Clarke, K.A.: A simple distribution-free test for nonnested model selection. Polit. Anal. **15**(3), 347–363 (2007)
19. Cormen, T., Leiserson, E., Rivest, R., Stein, C.: Introduction to Algorithms. The MIT Press, Cambridge (2009)
20. Czado, C.: Pair copula constructions of multivariate copulas. In: Jaworki, P., Durante, F., Härdle, W., Rychlik, W. (eds.) Workshop on Copula Theory and its Applications. Springer, Berlin (2010)
21. Czado, C., Jeske, S., Hofmann, M.: Selection strategies for regular vine copulae. (2012, preprint)
22. Czado, C., Schepsmeier, U., Min, A.: Maximum likelihood estimation of mixed C-vines with application to exchange rates. Stat. Modelling **12**(3), 229–255 (2012)
23. Dißmann, J., Brechmann, E.C., Czado, C., Kurowicka, D.: Selecting and estimating regular vine copulae and application to financial returns. Comput. Stat. Data Anal. **59**(1), 52–69 (2013)
24. Fischer, M., Köck, C., Schlüter, S., Weigert, F.: An empirical analysis of multivariate copula models. Quant. Financ. **9**(7), 839–854 (2009)
25. Genest, C., Ghoudi, K., Rivest, L.P.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika **82**, 543–552 (1995)
26. Genest, C., Werker, B.J.M.: Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In: Cuadras, C.M., Fortiana, J., Rodríguez Lallena, J.A. (eds.) Distributions with Given Marginals and Statistical Modelling, pp. 103–112. Kluwer, Dordrecht (2002)
27. Green, P.J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82**, 711–732 (1995)
28. Gruber, L.F.: Bayesian analysis of R-vine copulas. Master's thesis. Technische Universität München (2011)
29. Gruber, L.F., Czado C.: Sequential Bayesian model selection of regular vine copulas. (2013, preprint)
30. Hanea, A., Kurowicka, D., Cooke, R.: Hybrid method for quantifying and analyzing Bayesian belief nets. Qual. Reliab. Eng. Int. **22**(6), 613–729 (2006)
31. Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97–109 (1970)
32. Heinen, A., Valdesogo, A.: Asymmetric CAPM dependence for large dimensions: The Canonical Vine Autoregressive Model. CORE discussion papers 2009069. Université catholique de Louvain, Center for Operations Research and Econometrics (CORE) (2009)
33. Hobæk Haff, I.: Parameter estimation for pair-copula constructions. Bernoulli **19**(2), 462–491 (2013)
34. Hobæk Haff, I.: Comparison of estimators for pair-copula constructions. J. Multivar. Anal. **110**, 91–105 (2012)

35. Hobæk Haff, I., Aas, K., Frigessi, A.: On the simplified pair-copula construction - simply useful or too simplistic? J. Multivar. Anal. **101**(5), 1296–1310 (2010)
36. Hobæk Haff, I., Segers, J.: Nonparametric estimation of pair-copula constructions with the empirical pair-copula. http://arxiv.org/abs/1201.5133 (2012, preprint)
37. Joe, H.: Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In: Rüschendorf, L., Schweizer, B., Taylor, M.D. (eds.) Distributions with Fixed Marginals and Related Topics, pp. 120–141. Institute of Mathematical Statistics, Hayward (1996)
38. Joe, H.: Asymptotic efficiency of the two-stage estimation method for copula-based models. J. Multivar. Anal. **94**, 401–419 (2005)
39. Joe, H., Xu, J.: The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia (1996)
40. Kojadinovic, I., Yan, J.: Modeling multivariate distributions with continuous margins using the copula R package. J. Stat. Softw. **34(9)**, 1–20 (2010)
41. Kojadinovic, I., Yan, J.: A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. Stat. Comput. **21**, 17–30 (2011)
42. Kojadinovic, I., Yan, J., Holmes, M.: Fast large-sample goodness-of-fit test for copulas. Statistica Sinica **21**, 841–871 (2011)
43. Kurowicka, D.: Optimal truncation of vines. In: Kurowicka, D., Joe, H. (eds.) Dependence Modeling: Vine Copula Handbook. World Scientific, Singapore (2011)
44. Kurowicka, D., Cooke, R.: Distribution-free continuous Bayesian belief nets. In: Wilson, A., Limnios, N., Keller-McNulty, S., Armojo, Y. (eds.) Modern Statistical and Mathematical Methods in Reliability, pp. 309–323. World Scientific, Singapore (2005)
45. Kurowicka, D., Cooke, R.M.: Uncertainty Analysis with High Dimensional Dependence Modelling. Wiley, Chichester (2006)
46. Kurowicka, D., Joe, H.: Dependence Modeling: Vine Copula Handbook. World Scientific, Singapore (2011)
47. Manner, H.: Estimation and model selection of copulas with an application to exchange rates. METEOR research memorandum 07/056. Maastricht University (2007)
48. Mendes, B.V.d.M., Semeraro, M.M., Leal, R.P.C.: Pair-copulas modeling in finance. Financ. Markets Portfolio Manag. **24**(2), 193–213 (2010)
49. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. **21**, 1087–1092 (1953)
50. Meyn, S., Tweedie, R.L.: Markov Chain and Stochastic Stability, 2nd edn. Cambridge University Press, Cambridge (2009)
51. Min, A., Czado, C.: Bayesian inference for multivariate copulas using pair-copula constructions. J. Financ. Econom. **8**(4), 511–546 (2010)
52. Min, A., Czado, C.: Bayesian model selection for D-vine pair-copula constructions. Can. J. Stat. **39**, 239–258 (2011)
53. Morales-Nápoles, O.: Counting vines. In: Kurowicka, D., Joe, H. (eds.) Dependence Modeling: Vine Copula Handbook. World Scientific, Singapore (2011)
54. Nikoloulopoulos, A.K., Joe, H., Li, H.: Vine copulas with asymmetric tail dependence and applications to financial return data. Comput. Stat. Data Anal. **56**(11), 3659–3673 (2012)
55. Panagiotelis, A., Czado, C., Joe, H.: Pair copula constructions for multivariate discrete data. J. Am. Stat. Assoc. **107**(499), 1063–1072 (2012)
56. Schellhase, C., Kauermann, G.: Flexible pair-copula estimation in D-vines with penalized splines. (2012, preprint)
57. Schepsmeier, U., Brechmann, E.C.: CDVine: Statistical Inference of C- and D-Vine Copulas. R package version 1.1-12. http://CRAN.R-project.org/package=CDVine (2013)
58. Schepsmeier, U., Stoeber, J., Brechmann, E.C.: VineCopula: Statistical Inference of Vine Copulas. R package version 1.1-1. http://CRAN.R-project.org/package=VineCopula (2013)
59. Siek, J., Lee, L.Q., Lumsdaine, A.: Boost Graph Library: Random Spanning Tree - 1.49.0. http://www.boost.org/doc/libs/1_49_0/libs/graph/doc/random_spanning_tree.html (2012)

60. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
61. Smith, M., Min, A., Almeida, C., Czado, C.: Modeling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence. J. Am. Stat. Assoc. **105**(492), 1467–1479 (2010)
62. Stöber, J., Czado, C.: Detecting regime switches in the dependence structure of high dimensional financial data. http://arxiv.org/abs/1202.2009 (2011, preprint)
63. Stöber, J., Czado, C., Hong, H., Ghosh, P.: Interrelation among chronic diseases in the elderly. (2012, preprint)
64. Stöber, J., Czado, C., Joe, H.: Simplified pair copula constructions - limits and extensions. http://arxiv.org/abs/1205.4844 (2012, preprint)
65. Stöber, J., Schepsmeier, U.: Is there significant time-variation in multivariate copulas? http://arxiv.org/abs/1205.4841 (2012, preprint)
66. Vuong, Q.H.: Ratio tests for model selection and non-nested hypotheses. Econometrica **57**(2), 307–333 (1989)
67. Whittaker, J.: Graphical Models in Applied Multivariate Statistics. Wiley, New York (1990)

Chapter 3
Copulas in Machine Learning

Gal Elidan

Abstract Despite overlapping goals of multivariate modeling and dependence identification, until recently the fields of machine learning in general and probabilistic graphical models in particular have been ignorant of the framework of copulas. At the same time, the complementing strengths of the two fields suggests the great fruitfulness of a synergy. The purpose of this paper is to survey recent copula-based constructions in the field of machine learning, so as to provide a stepping stone for those interested in further exploring this emerging symbiotic research.

3.1 Introduction

Multivariate modeling is of fundamental interest in diverse complex domains ranging from computational biology to computer vision to astronomy. Unfortunately, high-dimensional modeling in the context of finite data and limited computational resources can be quite challenging and susceptible to the curse of dimensionality. Probabilistic graphical models [33], a marriage between probability and graph theory, is a general purpose framework aimed at coping with this task. These models are used to represent multivariate densities via a combination of a qualitative graph structure that encodes independencies and local quantitative parameters. The joint density has a decomposable form that corresponds to the intuitive graph structure. This, in turn, allows for relatively efficient methods for marginal and posterior computations (a task called *inference* in the field), estimation (parameter learning), and model selection (structure learning). Probabilistic graphical models have become a central axis of the field of machine learning, have made substantial

G. Elidan (✉)
Department of Statistics, Hebrew University, Jerusalem, Israel
e-mail: galel@huji.ac.il

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_3,

impact in related fields such as machine vision, natural language processing, and bioinformatics, and have become prevalent in uncountable applications.

It is somewhat remarkable that, until recently, researchers in the field of probabilistic graphical models were largely unaware of the multivariate modeling framework of copulas. This ignorance is even more perplexing when considering the limitations of graphical models in the context of real-valued measurements: while probabilistic graphical models are conceptually general, practical considerations almost always force the local quantitative part of the model to be of a simple form. In fact, when faced with data that cannot be captured well with multivariate Gaussians or mixtures thereof, the vast majority of works first discretize the data, and then take advantage of the impressive progress that has been made in the discrete case.

Much of the copula community has also been ignorant of the potential of a symbiosis with the field of machine learning. A decade ago, Kurowicka and Cooke [23] identified a relationship between vine models and Bayesian networks (a directed graphical model), and this was later generalized [16, 24] to yield high-dimensional copula constructions. However, no algorithmic innovation was borrowed from or inspired by machine learning, with the goal of, for example, automatically inferring the structure of such models from partially observed data.

There are fundamental reasons as to why a symbiosis between the two fields should be pursued. Graphical models are inherently aimed at high-dimensional domains, and substantial advances have been made in learning such models from data. Unfortunately, in real-valued scenarios the field is still largely handicapped. In contrast, copulas offer a flexible mechanism for modeling real-valued distributions. Yet, much of the field is still focused on the bivariate case or is limited in practice to few variables (exceptions are discussed later). The two frameworks thus complement each other in a way that offers opportunities for fruitful synergic innovations.

The need for a synergy between the copula framework and the field of machine learning goes further than probabilistic graphical models. Dependence measures, most notably Shannon's mutual information, are fundamental to numerous machine learning algorithms such as clustering, features selection, structure learning, causality detection, and more. As is well known, copulas are closely tied to such dependence concepts and the meeting of the two fields can give rise to new techniques for measuring dependance in high dimension.

It was only recently that the ignorance barrier between the two fields was broken by Kirshner's work [21] that generalizes Darsow's Markovian operator [7] for tree structured models. Since then, interest in copulas has been steadily growing and the last years have seen a range of innovative copula-based constructions in machine learning. The purpose of this paper is to survey these works. Rather than aiming at a complete coverage, the focus is on multivariate constructions as well as information estimation. For lack of space, additional works that, generally speaking, use copulas in a more plug-in manner are not discussed. For the interested reader, these include copula-based independent component analysis [35], component analysis [2, 27], mixture models (e.g., [14, 50]), dependency seeking clustering [40]. Also of great interest but not presented here is the use of copulas as a particular instance within the cumulative distribution network model [17, 45]. Finally, this survey does not cover

application papers or works that appeared in the computational statistics community and that are more likely to be familiar with copula researchers.

3.2 Background

To allow for reasonable accessibility to both copula and machine learning researchers, in this section we briefly review the necessary background material from both fields and set a common notation. We use capital letters X, Y to denote random variables, lowercase letters x, y to denote realizations of these variables, boldfaced letters to refer to set of variables $\mathbf{X}$ and their assignments $\mathbf{x}$.

3.2.1 Copulas

A copula function [47] links univariate marginal distributions to form a joint multivariate one. Formally,

Definition 3.1. Let $U_1, \ldots, U_n$ be real random variables marginally uniformly distributed on $[0, 1]$. A copula function $C : [0, 1]^n \rightarrow [0, 1]$ is a joint distribution

$$C(u_1, \ldots, u_n) = P(U_1 \leq u_1, \ldots, U_n \leq u_n).$$

We will use $C_\theta(\cdot)$ to denote a parameterized copula function where needed.

Sklar's seminal theorem [47] states that *any* joint distribution $F_{\mathbf{X}}(\mathbf{x})$ can be represented as a copula function $C(\cdot)$ of its univariate marginal distributions

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

When the marginals are continuous, $C(\cdot)$ is uniquely defined. The constructive converse, which is of interest from a modeling perspective, is also true: *any* copula function taking *any* univariate marginal distributions $\{F_i(x_i)\}$ as its arguments, defines a valid joint distribution with marginals $\{F_i(x_i)\}$. Thus, copulas are "distribution generating" functions that allow us to separate the choice of the univariate marginals and that of the dependence structure, encoded in the copula function $C(\cdot)$. Importantly, this flexibility often results in a construction that is beneficial in practice.

Assuming $C_\theta(\cdot)$ has n'th order partial derivatives, the joint density can be derived from the copula function using the derivative chain rule

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n C_\theta(F_1(x_1), \ldots, F_n(x_n))}{\partial F_1(x_1) \ldots \partial F_n(x_n)} \prod_i f_i(x_i) \equiv c_\theta(F_1(x_1), \ldots, F_n(x_n)) \prod_i f_i(x_i), \tag{3.1}$$

where $c_\theta(\cdot)$ is called the *copula density*.

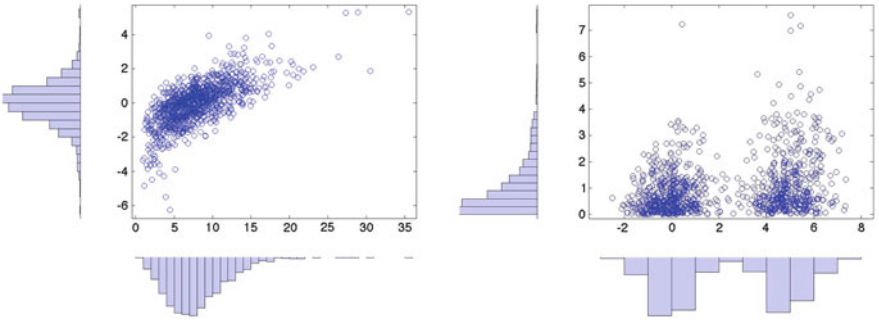

Fig. 3.1 Samples from the bivariate Gaussian copula with correlation $\theta = 0.25$. (*left*) with unit variance Gaussian and Gamma marginals; (*right*) with a mixture of Gaussian and exponential marginals

Example 3.1. Perhaps the most commonly used is the Gaussian copula [11]:

$$C_\Sigma(\{F_i(x_i)\}) = \Phi_\Sigma\left(\Phi^{-1}(F_1(x_1)), \ldots, \Phi^{-1}(F_n(x_n))\right), \tag{3.2}$$

where Φ is the standard normal distribution and Φ_Σ is a zero mean normal distribution with correlation matrix Σ.

Figure 3.1 shows samples from the bivariate Gaussian copula using two different marginal settings. As can be seen, even in this simple case, markedly different and multi-modal distributions can be constructed. More generally, and without any added computational difficulty, we can use different marginals for each variable and can also mix and match marginals of different forms with *any* copula function.

3.2.2 Probabilistic Graphical Models

In this section we briefly review probabilistic graphical models [33], a widely popular framework for representing and reasoning about high-dimensional densities.

A *directed graph* is a set of nodes connected by directed edges. A *directed acyclic graph* (DAG) $\mathcal{G}$ is a directed graph with no directed cycle. The *parents* of a node V in a directed graph is the set of all nodes U such that there exists a direct edge from U to V. A node U is an ancestor V in the graph if there is a directed path from U to V. Children and descendant are similarly defined.

Directed graphical models or *Bayesian networks* (BNs) use a DAG $\mathcal{G}$ whose nodes correspond to the random variables of interest $X_1, \ldots, X_n$ to encode the independencies $I(\mathcal{G}) = \{(X_i \perp \mathbf{ND}_i \mid \mathbf{Pa}_i)\}$, where $\perp$ denotes the independence relationship and $\mathbf{ND}_i$ are nodes that are not descendants of X_i in $\mathcal{G}$. Independencies that follow from $I(\mathcal{G})$ are easily identifiable via an efficient algorithm. If the

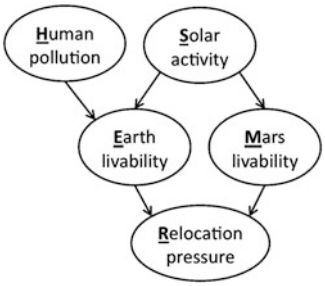

Fig. 3.2 A toy Bayesian network of a Mars relocation scenario where $f(\cdot) = f(H)f(S)f(E|S,H)f(M|S)f(R|E,M)$

independencies encoded by $\mathscr{G}$ hold in $f_{\mathbf{X}}$, then it is easy to show that the joint density decomposes into a product of local conditional densities

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i|\mathbf{Pa}_i}(x_i \mid \mathbf{pa}_i),$$

where $\mathbf{Pa}_i$ are the parents of node X_i in $\mathscr{G}$. The converse composition theorem states that a product of *any* local conditional densities defines a valid joint density and that the independencies encoded by $\mathscr{G}$ hold in this density.

As an example, Fig. 3.2 shows a plausible model that involves relocation of human population into Mars. Human pollution is unfortunately assumed independent of Solar activity. Yet, these two factors are dependent given evidence of livability conditions on earth. Similar deductions all follow from the independencies encoded in the graph. Note that inferences can be made in any direction, regardless of the direction of the edges, hence the name *Bayesian* networks.

Undirected graphical models, or *Markov Networks* (MNs), use an undirected graph $\mathscr{H}$ that encodes the independencies $I(\mathscr{H}) = \{(X_i \perp \mathbf{X} \setminus \{X_i\} \cup \mathbf{Ne}_i \mid \mathbf{Ne}_i)\}$, where $\mathbf{Ne}_i$ are the neighbors of X_i in $\mathscr{H}$. That is, each node is independent of all others given its neighbors in $\mathscr{H}$. Let $\mathscr{C}$ be the set of cliques in $\mathscr{H}$ (a clique is set of nodes such that each node is connected to all others in the set). As for directed models, the Hammersley–Clifford theorem [15] states that, for positive densities, if the independence statements encoded by $\mathscr{H}$ hold in $f_{\mathbf{X}}(\mathbf{x})$, then the joint density decomposes according to the graph structure:

$$f_{\mathbf{X}}(\mathbf{x}) = \tfrac{1}{Z} \textstyle\prod_{c \in \mathscr{C}} \phi_c(\mathbf{x}_c), \tag{3.3}$$

where $\mathbf{X}_c$ are the set of nodes in the clique c, and $\phi_c : \mathbb{R}^{|c|} \to \mathbb{R}^+$ is any positive function over the values of these nodes. Z is a normalizing constant called the partition function. The converse composition theorem also holds.

There are various generalization of the Bayesian and Markov network representations (which overlap only for tree structured models) including temporal,

relational, and mixed directionality models (chain graphs). The common theme is that of decomposition into local terms which, in additional to facilitating a compact representation, gives rises to relatively efficient marginal and conditional computations (a task called *inference* in the ML community), estimation (parameter learning), and high-dimensional model selection (structure learning). See [22, 33] for a comprehensive presentation of probabilistic graphical models.

3.3 Multivariate Copula-Based Construction

In this section we present several high-dimensional copula-based models recently developed in the machine learning community. As is common in the copula community [20], these works generally start with univariate estimation, and then plug in the "given" marginals into the copula function. Thus, except where essential, our exposition below does not cover the relatively straightforward and standard univariate estimation step. Instead, we focus on the multivariate construction. We end with a comparative summary in Sect. 3.3.5, which can also be read first.

3.3.1 Tree Structured Models

The first work in the machine learning community to combine ideas from the graphical models framework and copulas is that of Kirshner [21] (the earlier work of [24] independently developed in the copula community is discussed in Sect. 3.3.3). We start by describing the basic tree-structured copula construction and then present the tree-averaged density model. We conclude this section with a flexible Bayesian approach to a mixture of copula trees suggested by Silva and Gramacy [46].

3.3.1.1 Tree-Structured Copulas

Let T be an undirected tree structured graph (i.e., a graph with no cycles) and let $\mathscr{E}$ denote the set of edges in T that connect two vertices. From the Hammersley–Clifford decomposition of (3.3), it easily follows that, if the independencies $I(T)$ hold in $f_{\mathbf{X}}(\mathbf{x})$, then it can be written as

$$f_{\mathbf{X}}(\mathbf{x}) = \left[\prod_i f_i(x_i)\right] \prod_{(i,j)\in\mathscr{E}} \frac{f_{ij}(x_i, x_j)}{f_i(x_i) f_j(x_j)}.$$

Using (3.1), a decomposition of the joint copula also follows:

$$c_T(\cdot) = \frac{f_{\mathbf{X}}(\mathbf{x})}{\prod_i f_i(x_i)} = \prod_{(i,j)\in\mathscr{E}} \frac{f_{ij}(x_i, x_j)}{f_i(x_i) f_j(x_j)} = \prod_{(i,j)\in\mathscr{E}} c_{ij}(F_i(x_i), F_j(x_j)), \quad (3.4)$$

where $c_T(\cdot)$ is used to denote a copula density that corresponds to the structure T and $c_{ij}(\cdot)$ is used to denote the bivariate copula corresponding to the edge (i, j). The converse composition also holds: a product of local bivariate copula densities, each associated with an edge of T, defines a valid copula density. This result generalizes Darsow's operator [7] to the case of Markov trees. Indeed, it can be proved directly or by an inductive application of Darsow's product operator starting from the leaves of the trees and progressing inwards.

The main appeal of the above decomposition, as is the case for graphical models in general, is that estimation or learning also benefits from the compact representation. Given univariate marginals, (3.4) leads to a decomposition of the log-likelihood into independent terms, and estimation can be carried out by only considering bivariate statistics. This is in contrast to vine copula models [3] that also involve bivariate copulas but where (conditional) statistics over large sets of variables are required (see Sect. 3.3.5 for further discussion).

3.3.1.2 Tree-Averaged Copulas

As noted, the main appeal of the tree-structured copula is that it relies solely on bivariate estimation. However, this comes at the cost of firm independence assumptions. To relax these, Kirshner suggests the construction of a mixture of all copula trees model. On the surface, such a model may appear to be computationally prohibitive as the number of possible trees with n variables is n^{n-2}.

This difficulty is overcome by defining an appropriate decomposable prior over all spanning trees, as suggested by Meila and Jaakkola [28]. Let β be a symmetric $n \times n$ matrix with nonnegative entries and zero on the diagonal. Let $\mathscr{T}$ be the set of all spanning trees over $X_1, \ldots, X_n$. The probability of a spanning tree T is defined as

$$P(T \in \mathscr{T} \mid \beta) = \tfrac{1}{Z} \textstyle\prod_{(u,v)\in\mathscr{E}_T} \beta_{uv},$$

where Z is a normalization constant. Using a generalization of the Laplacian matrix:

$$L_{uv}(\beta) = \begin{cases} -\beta_{uv} & u \neq v \\ \sum_w \beta_{uw} & u = v, \end{cases}$$

it can be shown that the normalization constant Z is equal to the determinant $|L^*(\beta)|$, where $L^*(\beta)$ represents the first $(n-1)$ rows and columns of $L(\beta)$. This result can then be used to efficiently compute the density of the average of *all* copula spanning trees, which itself is also a copula density:

$$\sum_{T\in\mathscr{T}} P(T \mid \beta)c_T(\cdot) = \frac{1}{Z}\sum_{T\in\mathscr{T}}\left[\prod_{(u,v)\in\mathscr{E}_T}\beta_{uv}c_{uv}(F_u(x_u), F_v(x_v))\right] = \frac{|L^*(\beta\circ c_T(\cdot))|}{|L^*(\beta)|},$$

where $\circ$ denotes an element-wise product. The reader is referred to Kirshner [21] for additional details on the efficient EM method used for parameter estimation of the model and for appealing results when modeling multi-site precipitation data using an HMM-based construction.

3.3.1.3 Bayesian Mixtures of Copula Trees

The all tree mixture model described in the previous section overcomes some of the limitations imposed by a single tree model. However, to facilitate computational efficiency, the prior used involves heavy parameter sharing. Specifically, the set of all n^{n-2} trees is parameterized by only $n(n-1)$ parameters. Further, the approach relies on the assumption that there are no missing observations.

To offer more flexibility, Silva and Gramacy [46] suggest a Bayesian approach that allows for a mixture of *some* trees with flexible priors on all components of the model. The construction is based on the Bayesian nonparametric Dirichlet process infinite mixture model. This model, first formalized by Ferguson [12], is a distribution over discrete mixtures such that for every finite set of mixtures, its parameters have a Dirichlet prior. Following Silva and Gramacy, we present the model here as the limit as $K\to\infty$ of a finite mixture model with K components.

Let $\mathbf{X}$ be a set of random variables, z be an index of the set of all trees $\mathscr{T}$ over these variables, and Θ be the set of copula parameters, one for each pair of variables. The following model is a standard Bayesian mixture model, with the novelty that the parameters of the univariate marginals Λ are shared by all mixture components:

$$\begin{array}{ll} \Lambda \sim f_\Lambda & T_z \sim T_0(z) \\ \pi \sim \text{Dirichlet}(\alpha/K,\ldots,\alpha/K) & \Theta_z \sim f_\Theta \\ z \mid \pi \sim \text{Discrete}(\pi_1,\ldots,\pi_K) & \mathbf{X}|z,\mathscr{T},\Theta,\Lambda \sim f(X \mid T_z,\Theta_z,\Lambda). \end{array}$$

The first two lines on the left correspond to any general generating mechanisms for the univariate marginal parameters Λ and the mixture prior parameters π. Given these, a specific tree is selected by sampling z from any discrete distribution of the appropriate dimension parameterized by π. The parameters corresponding to the tree edges Θ_z are then sampled from a prior on the copula parameters. Finally, given a specific tree and previously sampled parameters, the density for a sample $f(X \mid T_z,\Theta_z,\Lambda)$ is constructed using a copula tree, as described in previous section.

Obviously, the above model offers great flexibility and using a Dirichlet process formulation where $K\to\infty$ allows for a variable number of components. The flexibility comes with a computational burden which is the central challenge addressed by Silva and Gramacy using a Markov chain Monte Carlo approach. The central difficulty is in the sampling of trees since, given a specific tree,

most parameters are redundant and sampling these naively will lead to useless computations in later iterations. The solution is a novel proposal distribution from which trees *and* parameters are sampled in a sensible way. The reader is referred to Silva and Gramacy [46] for the precise details. Experiments are carried out on several datasets from the UCI repository [31], as well as missing data scenarios using financial data.

3.3.2 Undirected Structure Learning

The *lasso* method of Tibshirani [51] extends linear regression to the high-dimensional case by including in the objective function an L1 norm sparsity constraint on the feature coefficients and proposing an efficient method for optimizing this objective. A nonparametric extension, called sparse additive models, was recently developed by Ravikumar et al. [38]. Orthogonally, the *graphical lasso* (glasso) [13] employs similar sparsity constraints to facilitate high-dimensional estimation of undirected Gaussian graphical models. In this section we present the work of Liu et al. [26] that fills the void of high-dimensional nonparametric structure estimation. Specifically, a theoretically founded structure estimator is developed based on the combination of the Gaussian copula and a specific form of nonparametric univariate marginals.

3.3.2.1 Parametric Undirected Graph Estimation

Let $\mathscr{H}$ be an undirected graph whose nodes correspond to real-valued random variables $X_1, \ldots, X_n$. For multivariate Gaussian distributions, the independencies between the random variables as encoded by the graph's structure are characterized by the inverse covariance matrix $\Omega = \Sigma^{-1}$. Specifically, X_i is independent of X_j given all other variables, denoted by $X_i \perp X_j \mid \mathbf{X}_{\setminus\{i,j\}}$ if and only if $\Sigma^{-1}_{ij} = 0$.

Given m samples of the random vector $\mathbf{X}$, estimation of Σ when $n > m$ cannot be carried out using a maximum likelihood estimator since the empirical covariance matrix is not full rank. Inspired by the success of L1 sparsity regularization for linear models, several authors suggested that Σ be estimated by finding the solution to the following regularized likelihood objective:

$$\hat{\Omega} = \min_{\Omega} -\frac{1}{2}\left(\log|\Omega| - tr(\Omega\hat{S})\right) + \lambda \sum_{j \neq k} |\Omega_{jk}|, \tag{3.5}$$

where $\hat{S}$ is the sample covariance matrix. The estimator $\hat{\Omega}$ can be computed efficiently by the glasso algorithm, which is simply a block coordinate descent that applies the standard lasso to a single row and column of Ω at each iteration. The resulting estimator has been shown to have appealing theoretical properties [39,41].

3.3.2.2 Nonparanormal Estimation

A real-valued random vector $\mathbf{X}$ is said to have a nonparanormal distribution, $\mathbf{X} \sim NPN(\mu, \Sigma, g)$, if there exist functions $\{g_i\}_{i=1}^n$ such that $(g_1(X_1), \ldots, g_n(X_n)) \sim N(\mu, \Sigma)$. When g_i are monotone and differentiable, this is simply the Gaussian copula. Now, define

$$h_i(x) = \Phi^{-1}(F_i(x_i)),$$

and let Λ be the covariance matrix of $h(X)$. The independence properties discussed above for the multivariate Gaussian hold so that $X_i \perp X_j \mid \mathbf{X}_{\setminus\{i,j\}}$ if and only if $\Lambda_{ij}^{-1} = 0$. Thus, to estimate the graph's structure, it is sufficient to identify Λ^{-1}.

Consider the obvious rank-based estimator for Λ that relies on the empirical marginal distribution function $\hat{F}_i(t) \equiv \frac{1}{m}\sum_{l=1}^{m} \mathbf{1}_{\{x_i[l] \le t\}}$, where $x_i[l]$ is used to denote the assignment to X_i in the l'th sample. Unfortunately, using this estimator as a plug-in to covariance estimation does not work well in high dimension since the variance of $\hat{F}_i$ can be large. Instead, the following Winsorized estimator is suggested

$$\tilde{F}_i(x) = \begin{cases} \delta_m & \text{if } \hat{F}_i(x) < \delta_m \\ \hat{F}_i(x) & \text{if } \delta_m \le \hat{F}_i(x) \le 1 - \delta_m \\ (1 - \delta_m) & \text{if } \hat{F}_i(x) > 1 - \delta_m, \end{cases}$$

where δ_m is a truncation parameter. Using $\delta_m \equiv \frac{1}{4m^{1/4}\sqrt{\pi \log m}}$ strikes the right bias-variance trade-off that leads to the desirable theoretical properties discussed below. Given this estimate for the distribution of X_i, and using $\tilde{h}_i(x) = \Phi^{-1}\left(\tilde{F}_i(x)\right)$, define the transformation functions by

$$\tilde{g}_i(x) \equiv \hat{\mu}_i + \hat{\sigma}_i \tilde{h}_i(x), \tag{3.6}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are sample mean and standard deviation of X_i, respectively. The sample covariance matrix $S_m(\tilde{g})$ can now be plugged in (3.5) in place of $\hat{S}$, defining a two-step estimation procedure for the estimator $\hat{\Omega}_m$:

1. Replace the observations with Winsorized normalized scores as defined in (3.6).
2. Use the graphical lasso to estimate the undirected graph.

Appealingly, the procedure is both easy to compute and makes little assumptions regarding the distribution of $\mathbf{X}$. The only tuning parameter is the regularization parameter λ that defines the objective minimized by the glasso algorithm. Next, we summarize the theoretical and empirical merits of this estimator.

3.3.2.3 Properties of the Estimator

Building on the analysis of Rothman et al. [41] and Ravikumar et al. [39], Liu et al. are able to show that their estimator has favorable persistency, norm consistency,

and model selection consistency properties. The main technical result is an analysis of the covariance of the Winsorized estimator. Specifically, under appropriate conditions,

$$\max_{i,j} \left| S_m(\tilde{g})_{ij} - S_m(g)_{ij} \right| = O_P\left(m^{-1/4}\right).$$

Using this result, norm consistency of $\hat{\Omega}$ with respect to the Frobenius and L2 norm follows, with a similar dependence on m. Using additional technical assumptions, a model selection consistency result (so that the true structure is recovered) is also provided. Further, Liu et al. also show that their estimator is consistent in risk, that is when the true distribution is not assumed to be nonparanormal.

Liu et al. demonstrate the ability of their method to accurately recover known structure in simulation experiments under different transformations that are applied to the univariate marginals, and various training sample sizes. The also apply their method to biological and financial data, leading to structures that are different than those learned with a purely Gaussian model, potentially revealing novel insights. The interested reader is referred to Liu et al. [26] for details.

3.3.3 Copula Bayesian Networks

Elidan [8] tackles the task of flexibly representing a multivariate real-valued distribution based on a directed graph representation.

3.3.3.1 The CBN Model

As discussed in Sect. 3.2, a joint distribution that relies on a DAG to encode independencies is quantified by local conditional densities. Accordingly, the construction starts with the following building block:

Lemma 3.1. *Let $f(x \mid \mathbf{y})$, with $\mathbf{y} = \{y_1, \ldots, y_k\}$, be a conditional density function. There exists a copula density function $c(F(x), F_1(y_1), \ldots, F_K(y_K))$ such that*

$$f(x \mid \mathbf{y}) = R_c(F(x), F_1(y_1), \ldots, F_K(y_K))\, f_X(x),$$

where R_c is the copula ratio

$$R_c(F(x), F_1(y_1), \ldots, F_K(y_K)) \equiv \frac{c(F(x), F_1(y_1), \ldots, F_K(y_K))}{\frac{\partial^K C(1, F_1(y_1), \ldots, F_K(y_K))}{\partial F_1(y_1) \ldots \partial F_K(y_K)}},$$

and R_c is defined to be 1 when $\mathbf{Y} = \emptyset$. The converse is also true: for any *copula, $R_c(F(x), F_1(y_1), \ldots, F_K(y_K))\, f_X(x)$ defines a valid conditional density.*

Note that the denominator of R_c is only seeming complex and is in fact a derivative of a lower order than the numerator copula density. Thus, whenever the copula density has a convenient form, so does R_c, and the conditional normalization does not involve any costly integration. With this building block in hand, the multivariate density model can be defined:

Definition 3.2. A Copula Bayesian Network (CBN) is a triplet $\mathscr{C} = (\mathscr{G}, \Theta_C, \Theta_f)$ that defines $f_{\mathbf{X}}(\mathbf{x})$. $\mathscr{G}$ encodes the independencies $\{(X_i \perp \mathbf{ND}_i \mid \mathbf{Pa}_i)\}$, assumed to hold in $f_{\mathbf{X}}(\mathbf{x})$. Θ_C is a set of local copula functions $\{C_i(F(x_i), F(\mathbf{pa}_{i1}), \ldots, F(\mathbf{pa}_{ik_i}))\}$ that are associated with the nodes of $\mathscr{G}$ that have at least one parent. In addition, Θ_f is the set of parameters representing the marginal densities $f_i(x_i)$ (and distributions $F_i(x_i)$). The joint density $f_{\mathbf{X}}(\mathbf{x})$ then takes the form

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} R_{c_i}\left(F(x_i), F(\mathbf{pa}_{i1}), \ldots, F(\mathbf{pa}_{ik_i})\right) f_i(x_i).$$

Elidan showed that if the independencies encoded in $\mathscr{G}$ hold in $f_{\mathbf{X}}(\mathbf{x})$, then the joint copula decomposes into a product of local copula ratio terms R_{c_i}. However, the converse is only partially true. The above product $\prod_i R_{c_i}(\cdot) f_i(x_i)$ *always* defines a valid joint density. However, the product $\prod_i R_{c_i}$, when each copula ratio is constructed independently, does not always define a valid copula. In this case, the marginals of the *valid* joint distribution do not necessarily equal to $F_i(x_i)$.

While this may seem unacceptable from a copula perspective, the model offers greater flexibility at the cost of marginal skewness, which in practice is not substantial. Moreover, when the structure of the graph $\mathscr{G}$ is a tree, the model collapses to the tree model described in Sect. 3.3.1, and the univariate marginals are preserved. Further, when using the Gaussian copula, the correct marginals can be maintained using an appropriate specification scheme, in which case the model is equivalent to a nonparametric BN model [24]. See Sect. 3.3.5 for further discussion.

Importantly, the above flexibility allows for the use of efficient algorithmic tools. Straightforwardly, assuming the marginals are estimated first, estimation of the entire CBN model decomposes into independent estimation of local copulas. Building on the same decomposability, standard greedy methods for structure learning can also be employed. More interestingly, the representation gives rise to approximate inference and structure learning innovations that are specifically tailored to the model. The latter is briefly described next while the interested reader is referred to Elidan [9] for details of the former.

3.3.3.2 Lightning-Speed Structure Learning

Elidan [10] tackles the challenge of automated structure learning of CBNs in a high-dimensional settings. When the graph $\mathscr{G}$ is constrained to be a tree, the optimal structure can be learned using a maximum spanning tree procedure [6]. More generally, as the number of possible graphs is super-exponential in the number

of variables, the common approach for structure learning is a greedy procedure that involves local structure modifications (e.g., single edge addition, delete, and reversal) and is guided by a model selection score. Typical scores, such as the Bayesian Information Criterion (BIC) [43], balance the likelihood of the model and its complexity. See, for example, Koller and Friedman [22] for details and variants.

The building block of essentially all score-based structure learning methods for graphical models is the evaluation of the merit of an edge in the network. This involves computing the likelihood gain that would result from adding an edge to the network, which in turn involves estimation of the bivariate maximum likelihood parameters. In the case of the CBN model, this involves computation of

$$\sum_{l=1}^{m} \log c_{\hat{\theta}}(F_X(x[l]), F_Y(y[l])),$$

where $\hat{\theta}$ are the estimated parameters, $x[l]$ is the value of X in the l'th instance, and the sum is over samples. Unfortunately, estimating $\hat{\theta}$, as well as the actual computation of the log-likelihood function can be difficult. In fact, for non-Gaussian real-valued models, even the learning of a tree structure can be prohibitive. Elidan [10] proposes an alternative that builds on the fact that as m grows, the above expression approaches the negative (differential) entropy:

$$-H(C_\theta(U,V)) = \int c_\theta(u,v) \log c_\theta(u,v) du dv, \tag{3.7}$$

with $U \equiv F_X, V \equiv F_Y$. However, computation of the copula entropy can also be difficult since for most copula families the above integral does not have a closed form. Instead, an efficient to compute proxy is proposed.

The relationship between Spearman's rho rank correlation measure of association $\rho_s(X,Y) \equiv \frac{cov(U,V)}{\sigma(U)\sigma(V)}$ and the copula function is well known: it can be easily shown (e.g., [30]) that for a distribution $f_{X,Y}(x,y)$ and its corresponding copula

$$\rho_s(X,Y) = \rho_s(C_\theta) \equiv 12 \iint C_\theta(U,V) du dv - 3. \tag{3.8}$$

Further, the vast majority of copula families define a concordance ordering where $\theta_2 > \theta_1$ implies $C_{\theta_2}(u,v) > C_{\theta_1}(u,v)$ for all u, v. Thus, for most copula families, Spearman's rho is monotonic in the dependence parameter θ.

Elidan [10] identifies a further intriguing relationship: it is conjectured that Spearman's rho is monotonic *in the copula entropy*, possibly given some weak necessary conditions. The result is proved for elliptical copulas and for the Farlie-Gumbel-Morgenstern family. In addition, the conjecture is demonstrated via simulation for varied families whose only known commonality is concordance ordering.

Thus, in many cases, the easy to compute Spearman's rho can be used as a proxy to the expected log-likelihood, and asymptotically consistent model selection can be carried out for tree models. For several real-life datasets, where the underlying

distribution in unknown, a near monotonic relationship is demonstrated in practice between the log-likelihood function and the empirical Spearman's rho. For more complex structures, Spearman's rho can be used to heuristically guide the learning procedure. The result is a lightning-speed procedure that learns structures that are as effective in terms of generalization to unseen test data as those learned by a costly exact procedure, with orders of magnitude improvement in running time. Appealingly, the running time improvement grows with the domain's complexity. A 100 variable structure, for example, is learned in essentially the same time that it takes to learn the structure of a naive Gaussian BN (less than a minute on a single CPU).

3.3.4 Copula Processes

Consider the problem of measuring the dependencies between real-valued measurements of a continuous process. For example, the dependence between a rocket's velocity at different times as it leaves earth and how it relates to the dependence between the rocket's distances. As Wilson and Ghahramani [52] observe, these quantities are naturally on different scales and have different marginal distributions. Thus, it is desirable to separate the univariate effect from the dependence structure. Toward this goal, they define a copula process which can describe the dependence between *arbitrarily* many random variables.

Definition 3.3. Let $\{X_t\}$ be a collection of random variables indexed by t with marginal distributions $U_t \equiv F_t(X_t)$. Let G_t be the marginal distributions of a base process, and let H be the base joint distribution. X_t is a *copula process* with G_t, H, denoted $X_t \sim \mathrm{CP}(G_t, H)$, if for every finite set of indices $\mathscr{I} = \{t_1, \ldots, t_n\}$

$$P\left(\cap_{i=1}^{n}\{G_{t_i}^{-1}(U_{t_i}) \leq a_i\}\right) = H_{t_1,\ldots,t_n}(a_1, \ldots, a_n),$$

where G_t^{-1} is the quasi-inverse of G_t. That is, for all $t_i \in \mathscr{I}$, H defines the joint distribution over $\{G_{t_i}^{-1}\}_{t_i \in \mathscr{I}}$.

As an example, consider the case where the base measure is a Gaussian process (GP). X_t is a GP if for every finite subset of indices $\mathscr{I}$, the set $\{X_{t_i}\}_{t_i \in \mathscr{I}}$ has a multivariate Gaussian distribution. To allow for a variable size set $\mathscr{I}$, a GP is parameterized by a mean function $m(t)$ that determines the expectation of the random variable X_t, and a kernel function $k(t, t')$ that determines the covariance of X_t and $X_{t'}$. GPs are widely used in machine learning to define distributions over an arbitrary number of random variables or functions (see Rasmussen [37]). When the base measure is chosen to be a GP, we say that X_t has a Gaussian copula process (GCP) distribution. This is equivalent to the existence of a mapping Ψ such that $\Psi(X_t)$ is a GP. We denote this by $X_t \sim \mathrm{GCP}(\Psi, m(t), k(t, t'))$.

In principle, given complete samples and a known mapping, one can estimate a GCP by simply transforming the data and using black box procedures for GP estimation, such as that of Snelson et al. [48]. Wilson and Ghahramani, however,

consider a more challenging application setting that requires further algorithmic innovation. Concretely, they introduce a volatility model where the unobserved standard deviations of the data follow a GCP distribution

$$\sigma_t \sim \text{GCP}(g^{-1}, 0, k(t, t')).$$

The observations $X_t \sim N(0, \sigma_t^2)$ are assumed to follow a normal distribution, though this assumption can easily be relaxed. The difficulty is rooted in the fact that the σ_t's are never observed and that the so-called *warping function* g is unknown.

Let θ be the parameters that define both the GP covariance function and the warping function. Further, using a different notation from Wilson and Ghahramani to maintain consistency, let $z_t = g^{-1}(\sigma_t)$ be the latent function values that have a GP distribution. The central components involved in estimating θ from samples x_t and making prediction at some unrealized time $t^\star$ are:

- A Laplace approximation for the posterior $f(f_Z(z_{t^\star}) \mid \mathbf{y}, \theta)$.
- A Markov Chain Monte Carlo technique to sample from this posterior, specifically the elliptical slice sampling method [29].
- A flexible parametric as well as nonparametric warping functions to transform the samples into standard deviation space.

We refer the interested reader to Wilson and Ghahramani [52] for the details, as well as favorable results relative to a GARCH model when applied to financial data.

3.3.5 Comparative Summary

In this section we summarize the relative merits of the different multivariate approaches presented in the previous sections. Also discussed is the relationship to vine models and a related BN-based construction. Table 3.1 summarizes the properties of each of the models discussed.

Vine models [3, 19] have become the dominant tool in the copula community for the construction of flexible multivariate copulas. The widely studied formalism builds on successive conditioning and the use of bivariate copulas to construct multivariate distributions. While the framework is quite general, the seemingly bivariate estimation relies on conditional terms of greater dimension that can be hard to estimate. In practice, most applications are computationally limited to less than 10 variables, with recent innovations (e.g., [5]) somewhat pushing this boundary.

The tree-average distribution model of Kirshner [21] described in Sect. 3.3.1 generalizes Darsow's Markovian operator and allows for the construction of high-dimensional copulas via a composition of (unconditional) bivariate copulas. Appealingly this requires only bivariate estimation but is hampered by the independence assumptions implied by the tree structure. These assumptions are relaxed by allowing for a mixture of all trees construction which is efficiently learned using a

Table 3.1 Summary of the different copula-based multivariate models

Model	References	Variables	Structure	Copula	Comments
Vines	[1, 3, 25]	< 10 in practice	Conditional dependence	Any bivariate	Well understood general purpose framework
Nonparametric BBN	[16, 24]	100s	BN + vines	Gaussian in practice	Mature application
Tree-averaged	[21, 46] Section 3.3.1	10s	Mixture of trees	Any bivariate	Requires only bivariate estimation
Nonparanormal	[26] Section 3.3.2	100–1000s	MN	Gaussian	High-dimensional estimation with theoretical guarantees
Copula networks	[8, 10] Section 3.3.3	100s	BN	Any	Flexible at the cost of partial control over marginals
Copula processes	[18, 52] Section 3.3.4	∞ (replications)	–	Multivariate	Nonparametric generalization of Gaussian processes

compactly represented prior. A Bayesian refinement of the work was later suggested by Silva and Gramacy [46]. The construction is practical for 10s of variables.

Distribution-free or nonparametric belief Bayesian networks (NPBBNs) [16, 24] are aimed at overcoming the limitations of simple vines by using a BN structure to encode a decomposition of the joint distribution and employing local vines to encode $f_{X_i|\mathbf{Pa}_i}$. In principle, the construction can be used with any copula for which the specified conditional rank correlations can be realized. In practice, this can be carried out easily only when using an elliptical copula. That said, NPBBNs have led to the most mature and large-scale copula constructions to date.

CBNs [8], developed in the machine learning community, also use a BN structure to encode independencies that are assumed to hold in the distribution. The local conditional density, however, is parameterized differently via a proper normalization of a joint local copula over a variable and its parents in the graph. For tree structured models, a CBN reduces to the tree construction suggested by Kirshner [21]. When using a Gaussian copula, as discussed, it is also possible to estimate the parameters of the entire model so to ensure preservation of the univariate marginals. Thus specified, the model is equivalent to NPBBNs using local Gaussian copulas. However, CBNs also allow for greater flexibility at the cost of "skewed" marginals. Intuitively, this results from overlapping influences of multiple parents of a variable. Practically, since each local density is parameterized via an estimated joint copula with the same marginals, the overall univariate marginals are quite accurate. From a *given marginals* viewpoint this may be unacceptable. However, from a broader modeling perspective, in the face of finite data and an unknown joint distribution,

the goals of maximum likelihood and full control over the univariate marginals are competing ones. In this light, a balance between flexible modeling and univariate control may be beneficial. Importantly, if one is willing to strike this balance, then the CBN construction opens the door for algorithmic advances from the field of probabilistic graphical models. Indeed, the experiments presented in Elidan [8] are the largest where the structure of the model was automatically learned. The construction also subsequently led to specifically tailored efficient inference [9] and structure learning methods [10].

The nonparanormal method of Liu et al. [26] tackles the problem of structure learning in the complementing representation of undirected graphs. While it is specifically focused on an Gaussian copula, it provides appealing theoretical guarantees of consistency when the data is generated from the model, as well as risk consistency guarantees when samples arise from a different distribution. Importantly, the method applies to the previously unstudied regime of nonparametric estimation in high-dimensions when the number of parameters exceeds that of the samples.

Finally, the copula process model of Wilson and Ghahramani [52] defines a distribution over an infinite number of random variables while allowing for the explicit control over the marginals, thus generalizing Gaussian processes. We note that "infinite" here may be misleading since a "variable" is a replication, and Gaussian processes can also suffer from computational limitations. An obvious but challenging future prospect is the combination of this construction with local decomposability.

3.4 Information Estimation

Estimation of the mutual information of a set of variables is a fundamental challenge in machine learning that underlies numerous tasks ranging from learning the structure of graphical models to independent component analysis to image registration. However, for real-valued non-Gaussian random variables, estimation of different information measures can be difficult. In particular, the plug-in approach of computing the information based on an estimated density is often ineffective due to the difficulty of constructing complex joint distributions. Fortunately, just as copulas are opening new frontiers for modeling high-dimensional complex densities, so do they offer new opportunities for estimation of information measures. In this section we describe a series of recent works that build on such opportunities.

For all works discussed below, let $X[1 : m] = \mathbf{X}[1], \ldots, \mathbf{X}[m]$ be m i.i.d. samples of $\mathbf{X}$. The first (obvious in the context of copulas) step of all works is a rank-based transform $Z_i[l] = \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}_{X_i[l] \leq X_i[k]}$. Asymptotically, Z_i will be uniformly distributed on $[0, 1]$. However, the random samples $\mathbf{Z}[1], \ldots, \mathbf{Z}[m]$ are no longer independent. The works below take advantage of the former property and overcome the limitations of the latter consequence to produce appealing information estimators.

3.4.1 Information Estimation Based on Graph Optimization

The goal of both Póczos et al. [36] and Pal et al. [32] is to effectively estimate the Rényi information defined as

$$I_\alpha(\mathbf{X}) = \frac{1}{\alpha} \log \int f_{\mathbf{X}}^{\alpha}(\mathbf{x}) \Big(\prod_i f_i(x_i) \Big)^{1-\alpha} d\mathbf{x}.$$

Note that when $\alpha \to 1$, Rényi information converges to the well-known Shannon's mutual information measure. Rather than attempting to estimate $f_{\mathbf{X}}(\mathbf{x})$ which is a nuisance parameter, both works perform direct nonparametric estimation of $I_\alpha(\mathbf{X})$ by combining copula-based tools and graph-based estimators for the Rényi entropy

$$H_\alpha(\mathbf{X}) = \frac{1}{\alpha} \log \int f_{\mathbf{X}}^{\alpha}(\mathbf{x}) d\mathbf{x}.$$

Although both works contain interesting contributions, for clarity of exposition we focus on the former and encourage the interested reader to explore the latter.

Let G be a graph with m nodes. Note that this is *not* a probabilistic graphical model over $\mathbf{X}$ but rather a graph whose nodes will index the training samples. Let $E(G)$ be the set of edges in G and let $\mathcal{G}$ be a family of such graphs. For example, $\mathcal{G}_{ST}$ will correspond to the family of all spanning trees over m nodes. Now define

$$L_m(\mathbf{X}[1:m]) = \min_{G \in \mathcal{G}} \sum_{l,k \in E(G)} \|\mathbf{X}[l] - \mathbf{X}[k]\|^p.$$

In words, $L_m(\cdot)$ is the minimum p-power weighted edge length of graphs in $\mathcal{G}$. For example, for $\mathcal{G}_{ST}$ and $p = 1$, $L_m(\cdot)$ is simply the length of the minimal spanning tree, a quantity readily found using efficient graph optimization. Remarkably, $L_m(\cdot)$ is also useful for entropy estimation:

Theorem 3.1 (Steele [49]). *Let $n \geq 2, 0 < \alpha < 1$, and let $\mathbf{X}[1:m]$ be i.i.d. random vectors supported on $[0,1]^n$ with density $f_{\mathbf{X}}$. Define the estimator*

$$H_m(\mathbf{X}[1:m]) = \frac{1}{1-\alpha} \log \frac{L_m(\mathbf{X}[1:m])}{\gamma_{n,\alpha} m^\alpha},$$

where $\gamma_{n,\alpha}$ is a constant that does not depend on $f_{\mathbf{X}}$. Then, $H_m(\mathbf{X}[1:m]) \to H_\alpha(\mathbf{X})$ almost surely as $m \to \infty$ (similar theorems exist for other graph families $\mathcal{G}$, see Póczos et al. [36] for details and references).

The first obstacle in using the above theorem is that it applies to variables that are supported on $[0,1]^n$. This is easily overcome by the rank-based transform that results in $\mathbf{Z}[1], \ldots, \mathbf{Z}[m]$. Now, since Z_i is defined via a measurable invertible mapping, $I_\alpha(\mathbf{Z}) = I_\alpha(\mathbf{X})$. Further, since the marginals of $\mathbf{Z}$ are uniform, we have $I_\alpha(\mathbf{Z}) = -H_\alpha(\mathbf{Z})$ so that an entropy estimator can used to estimate information (this generalizes the known fact that Shannon's information is equal to the negative

copula entropy). The transform, however, introduces a new difficulty since the samples $Z[m]$ are now dependent. Poczos et al. [36] shows that despite this the estimator has favorable strong consistency and robustness properties. They also demonstrate the advantage of their rank-based approach in practice, for an image registration task.

3.4.2 Kernel-Based Dependency Measures

Like the above works, Póczos et al. [34] also start with an empirical rank transformation of the data followed by the application of an existing distance measure between distributions. The combination, however, is quite different than the graph optimization-based approaches described above. Omitting most of the technical details, we briefly present the high level idea and the merits of the resulting estimator. We start with the definition of the maximum mean discrepancy (MMD) measure of distributions similarity, which can be efficiently estimated from i.i.d. samples:

Definition 3.4. Let $\mathscr{F}$ be a class of functions, P and Q be probability distributions. The MMD between P and Q on the function class $\mathscr{F}$ is defined as follows:

$$\mathscr{M}[\mathscr{F}, P, Q] \equiv \sup_{f \in \mathscr{F}} \left(\mathbb{E}_{\mathbf{X} \sim P}[f(\mathbf{x})] - \mathbb{E}_{\mathbf{Y} \sim Q}[f(\mathbf{y})] \right).$$

We will focus on functional spaces that are a reproducing kernel Hilbert Space (RKHS), a fundamental tool in machine learning [42]. Without going into the technical definition, the notion of RKHS is important since, assuming that $\mathscr{F}$ is a unit ball of RKHS $\mathscr{H}$, measures such as $\mathscr{M}(\mathscr{F}, P, Q)$ and related quantities can be estimated efficiently [4]. Building on this fact, consider the following dependence measure

$$I_k(X_1, \ldots, X_n) \equiv \mathscr{M}(\mathscr{F}, F_{\mathbf{X}}, F_{\mathbf{U}}),$$

where we use $F_{\mathbf{U}}$ to denote the n-dimensional uniform distribution. Póczos et al. [34] show that if one chooses $\mathscr{F}$ properly (a RKHS with an additional denseness requirement), then I_k is a proper dependence measure that follows Schweizer and Wolffs's intuitive axioms [44]. They suggest an empirical estimator for I_k that is based on an empirical MMD estimation of the rank transformed samples $\mathbf{Z}[m]$, and prove that their easy to compute estimator is almost surely consistent. Further, they provide upper bound convergence rates. Finally, they demonstrate the merit of the estimator in practice in the context of a feature selection task.

3.5 Summary

In the introduction it was argued that, in the context of multivariate modeling and information estimation, the complementing strengths and weaknesses of the fields of machine learning and that of copulas offer opportunities for symbiotic constructions.

This paper surveyed the main such synergic works that recently emerged in the machine learning community.

While discrete high-dimensional modeling has been studied extensively, real-valued modeling for more than a few dimensions is still in its infancy. There exists no framework that is as general and as flexible as copulas for multivariate modeling. Thus, it is inevitable that machine learning researchers who aim to stop discretizing data will have to pay serious attention to the power of copulas. Conversely, if researchers in the copula community aim to cope with truly high-dimensional challenges, algorithmic prowess, a focus of the machine learning community, will have to be used. True, impressive large-scale models have been built using NPBBNs. However, the multi-year endeavor supported by human expertise cannot be scaled up or easily applied to a broad range of problems. Consequently, automated learning of models that takes into account the difficulties presented by the high-dimensional and partially observed setting is clearly needed. The goal of this survey is to provide an entry point for those aiming to tackle this far from realized challenge.

Acknowledgments G. Elidan was partially supported by a Google research grant and by an Israeli Science Foundation Center of Research grant.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependencies. Insur. Math. Econ. **44**, 182–198 (2009)
2. Abayomi, K.: Diagnostics for Multivariate Imputation, Copula Based Independent Component Analysis and a Motivating Example. Columbia University, New York (2008)
3. Bedford, T., Cooke, R.: Vines - a new graphical model for dependent random variables. Ann. Stat. **30**, 1031–1068 (2002)
4. Borgwardt, K.M., Gretton, A., Rasch, M.J., Kriegel, H., Schölkopf, B., Smola, A.: Integrating structured biological data by Kernel maximum mean discrepancy. Bioinformatics **22**(14), e49–e57 (2006)
5. Brechmann, E.C., Czado, C., Aas, K.: Truncated regular vines in high dimensions with applications to financial data. Can. J. Stat. **40**(1), 68–85 (2012)
6. Chow, C.K., Liu, C.N.: Approximating discrete probability distributions with dependence trees. IEEE Trans. Inf. Theor. **14**, 462–467 (1968)
7. Darsow, W., Nguyen, B., Olsen, E.: Copulas and Markov processes. Illinois J. Math. **36**, 600–642 (1992)
8. Elidan, G.: Copula Bayesian networks. In: Neural Information Processing Systems (NIPS), Vancouver, 2010
9. Elidan, G.: Inference-less density estimation using copula Bayesian networks. In: Uncertainty in Artificial Intelligence (UAI), Catalina Island, 2010
10. Elidan, G.: Lightning-speed structure learning of nonlinear continuous networks. In: Proceedings of the AI and Statistics Conference (AISTATS), La Palma, 2012
11. Embrechts, P., Lindskog, F., McNeil, A.: Modeling dependence with copulas and applications to risk management. In: Handbook of Heavy Tailed Distributions in Finance. Elsevier, Amsterdam (2003)
12. Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. Ann. Stat. **1**(2), 209–230 (1973)

13. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2007)
14. Fujimaki, R., Sogawa, Y., Morinaga, S.: Online heterogeneous mixture modeling with marginal and copula selection. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, 2011
15. Hammersley, J.M., Clifford, P.E.: Markov fields on finite graphs and lattices. Unpublished manuscript (1971)
16. Hanea, A., Kurowicka, D., Cooke, R.M., Ababei, D.: Mining and visualising ordinal data with non-parametric continuous bbns. Comput. Stat. Data Anal. **54**(3), 668–687 (2010)
17. Huang, J., Frey, B.: Cumulative distribution networks and the derivative-sum-product algorithm. In: Uncertainty in Artificial Intelligence (UAI), Helsinki, 2008
18. Jaimungal, S., Ng, E.: Kernel-based copula processes. In: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part I, Bled, 2009
19. Joe, H.: Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. In: Distributions with Fixed Marginals and Related Topics. Institute of Mathematical Statistics, California (1996)
20. Joe, H., Xu, J.: The Estimation method of inference functions for margins for multivariate models. Technical report no. 166, Department of Statistics, University of British Columbia (1966)
21. Kirshner, S.: Learning with tree-averaged densities and distributions. In: Neural Information Processing Systems (NIPS), Vancouver, 2007
22. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. The MIT Press, Cambridge (2009)
23. Kurowicka, D., Cooke, R.: The vine copula method for representing high dimensional dependent distributions: Applications to cont. belief nets. In: Proceedings of the Simulation Conference, San Diego 2002
24. Kurowicka, D., Cooke, R.M.: Distribution-free continuous Bayesian belief nets. In: Modern Statistical and Mathematical Methods in Reliability. Selected Papers Based on the Presentation at the International Conference on Mathematical Methods in Reliability (MMR), Santa Fe, 2005
25. Kurowicka, D., Joe, H.: Dependence Modeling: Vine Copula Handbook. World Scientific, Singapore (2011)
26. Liu, H., Lafferty, J., Wasserman, L.: The nonparanormal: semiparametric estimation of high dimensional undirected graphs. J. Mach. Learn. Res. **10**, 2295–2328 (2010)
27. Ma, J., Sun, Z.: Copula component analysis. In: Proceedings of the International conference on Independent Component Analysis and Signal Separation, London, 2007
28. Meilă, M., Jaakkola, T.: Tractable Bayesian learning of tree belief networks. Stat. Comput. **16**(1), 77–92 (2006)
29. Murray, I., Adams, R., MacKay, D.: Elliptical slice sampling. In: Proceedings of the AI and Statistics Conference (AISTATS), Sardinia, 2010
30. Nelsen, R.: An Introduction to Copulas. Springer, New York (2007)
31. Newman, D., Hettich, S., Blake, C., Merz, C.: UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html (1998)
32. Pal, D., Póczos, B., Szepesvári, C.: Estimation of Rényi entropy and mutual information based on generalized nearest-neighbor graphs. In: Neural Information Processing Systems (NIPS), Vancouver, 2010
33. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Francisco (1988)
34. Póczos, B., Ghahramani, Z., Schneider, J.: Copula-based kernel dependency measures. In: Proceedings of the International Conference on Machine Learning (ICML), Edinburgh, 2012
35. Póczos, B., Kirshner, S.: ICA and ISA using Schweizer-Wolff measure of dependence. In: Proceedings of the International Conference on Machine Learning (ICML), Helsinki, 2008

36. Póczos, B., Kirshner, S., Szepesvári, C.: REGO: Rank-based estimation of Rényi info. using Euclidean graph optimization. In: Proceedings of the AI and Statistics Conference (AISTATS), Sardinia, 2010
37. Rasmussen, C.E.: Gaussian Processes for Machine Learning. MIT Press, Cambridge (2006)
38. Ravikumar, P., Lafferty, J., Liu, H., Wasserman, L.: Sparse additive models. J. R. Stat. Soc. Series B **71**(5), 1009–1030 (2009)
39. Ravikumar, P., Wainwright, M., Raskutti, G., Yu, B.: Model selection in Gaussian graphical models: high-dimensional consistency of l1-regularized mle. In: Neural Information Processing Systems (NIPS). MIT Press, Cambridge (2009)
40. Rey, M., Roth, V.: Copula mixture model for dependency-seeking clustering. In: International Conference on Machine Learning (ICML), Edinburgh, 2012
41. Rothman, A., Bickel, P., Levina, E., Zhu, J.: Sparse permutation invariant covariance estimation. Electron. J. Stat. **2**, 494–515 (2008)
42. Schölkopf, B., Smola, A.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge (2001)
43. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978)
44. Schweizer, B., Wolff, E.: On nonparameteric measures of dependence for random variables. Ann. Stat. **9**, 879–885 (1981)
45. Silva, R., Blundell, C., Teh, Y.: Mixed cumulative distribution networks. In: Proceedings of the AI and Statistics Conference (AISTATS), Ft. Lauderdale, 2011
46. Silva, R., Gramacy, R.: MCMC methods for Bayesian mixtures of copulas. In: Proceedings of the AI and Statistics Conference (AISTATS). Clearwater Beach (2009)
47. Sklar, A.: Fonctions de repartition a n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
48. Snelson, E., Rasmussen, C.E., Ghahramani, Z.: Warped Gaussian processes. In: Advances in Neural Information Processing Systems (NIPS). MIT Press, Cambridge (2003)
49. Steele, J.: Probability Theory and Combinatorial Optimization. Society for Industrial and Applied Mathematics, Philadelphia (1987)
50. Tewari, A., Giering, M.J., Raghunathan, A.: Parametric characterization of multimodal distributions with non-Gaussian modes. In: IEEE Conference on Data Mining Workshops, Vancouver, 2011
51. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B **58**(1), 267–288 (1996)
52. Wilson, A., Ghahramani, Z.: Copula processes. In: Neural Information Processing Systems (NIPS), Vancouver, 2010

Chapter 4
An Overview of the Goodness-of-Fit Test Problem for Copulas

Jean-David Fermanian

Abstract We review the main "omnibus procedures" for goodness-of-fit (GOF) testing for copulas: tests based on the empirical copula process, on probability integral transformations (PITs), on Kendall's dependence function, etc., and some corresponding reductions of dimension techniques. The problems of finding asymptotic distribution-free test statistics and the calculation of reliable p-values are discussed. Some particular cases, like convenient tests for time-dependent copulas, for Archimedean or extreme-value copulas, etc., are dealt with. Finally, the practical performances of the proposed approaches are briefly summarized.

4.1 Introduction

Once a model has been stated and estimated, a key question is to check whether the initial model assumptions are realistic. In other words, and even it is sometimes eluted, every modeler is faced with the so-called goodness-of-fit (GOF) problem. This is an old-dated statistical problem that can be rewritten as: denoting by F the cumulative distribution function (cdf hereafter) of every observation, we would like to test

$$\mathscr{H}_0 : F = F_0, \quad \text{against} \quad \mathscr{H}_a : F \neq F_0,$$

for a given cdf F_0, or, more commonly,

$$\mathscr{H}_0 : F \in \mathscr{F}, \quad \text{against} \quad \mathscr{H}_a : F \notin \mathscr{F},$$

J.-D. Fermanian (✉)
Department of Finance, Ensae-Crest, Malakoff, France
e-mail: jean-david.fermanian@ensae.fr

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_4,

for a given family of distributions $\mathscr{F} := \{F_\theta, \theta \in \Theta\}$. This distinction between simple and composite assumptions is traditional and we keep it. Nonetheless, except in some particular cases (test of independence, e.g.), the latter framework is a lot more useful than the former in practice.

Some testing procedures are "universal" (or "omnibus"), in the sense they can be applied for any underlying distribution. In other terms, they do not depend on some particular properties of F_0 or of the assumed family $\mathscr{F}$. Such tests are of primary interest for us. Note that we will not consider Bayesian testing procedures, as proposed in [54], for instance.

To fix the ideas, consider an i.i.d. sample $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of a d-dimensional random vector $\mathbf{X}$. Its joint cdf is denoted by F, and the associated marginal cdfs' by F_j, $j = 1, \ldots, d$. Traditional key quantities are provided by the empirical distribution functions of the previous sample: for every $\mathbf{x} \in \mathbb{R}^d$, set d marginal cdfs'

$$F_{n,k}(x_k) := n^{-1} \sum_{i=1}^{n} \mathbf{1}(X_{i,k} \leq x_k), \ k = 1, \ldots, d,$$

and the joint empirical cdf $F_n(\mathbf{x}) := n^{-1} \sum_{i=1}^{n} \mathbf{1}(\mathbf{X}_i \leq \mathbf{x})$. The latter inequality has to be understood componentwise. Most of the "omnibus" tests are based on transformations of the underlying empirical distribution function or of the empirical process $\mathbb{F}_n := \sqrt{n}(F_n - F_0)$ itself: $T_n = \psi_n(F_n)$ or $T_n = \psi_n(\mathbb{F}_n)$. It is the case of the famous Kolmogorov–Smirnov (KS), Anderson–Darling (AD), Cramer–von Mises (CvM), and chi-squared tests, for example.

Naively, it could be thought the picture is the same for copulas and that straightforward modifications of standard GOF tests should do the job. Indeed, the problem for copulas can be simply written as testing

$$\mathscr{H}_0 : C = C_0, \quad \text{against} \quad \mathscr{H}_a : C \neq C_0, \quad \text{or}$$

$$\mathscr{H}_0 : C \in \mathscr{C}, \quad \text{against} \quad \mathscr{H}_a : C \notin \mathscr{C},$$

for some copula family $\mathscr{C} := \{C_\theta, \theta \in \Theta\}$. Moreover, empirical copulas, introduced by Deheuvels in the 1980s (see [23–25]), play the same role for copulas as standard empirical cdfs' for general distributions. For any $\mathbf{u} \in [0,1]^d$, they can be defined by

$$C_n(\mathbf{u}) := F_n(F_{n,1}^{(-1)}(u_1), \ldots, F_{n,d}^{(-1)}(u_d)),$$

with the help of generalized inverse functions, or by

$$\bar{C}_n(\mathbf{u}) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(F_{n,1}(X_{i,1}) \leq u_1, \ldots, F_{n,d}(X_{i,d}) \leq u_d).$$

It can be proved easily that $\|C_n - \bar{C}_n\|_\infty \leq d\, n^{-1}$ (see [36]). Then, for the purpose of GOF testing, working with C_n or $\bar{C}_n$ does not make any difference asymptotically. In every case, empirical copulas are explicit functionals of the underlying empirical

cdf: $C_n = \zeta(F_n)$. Thus, any previous GOF test statistics for copulas could be defined as $T_n = \psi_n(C_n) = \psi_n \circ \zeta(F_n)$. But this functional ζ is sufficient to induce significant technical difficulties, when applied to standard statistical procedures.

Actually, the latter parallel applies formally, but strong differences appear in terms of the limiting laws of the "copula-related" GOF test statistics. Indeed, some of them are distribution-free in the standard case, i.e., their limiting laws under the null do not depend on the true underlying law F, and then, they can be tabulated: KS (in the univariate case), chi-squared tests, for example. Unfortunately, it is almost impossible to get such nice results for copulas, due to their multivariate nature and due to the complexity of the previous mapping between F_n and C_n. Only a few GOF test techniques for copulas induce distribution-free limiting laws. Therefore, most of the time, some simulation-based procedures have been proposed for this task.

In Sect. 4.2, we discuss the "brute-force" approaches based on some distances between the empirical copula C_n and the assumed copula (under the null), and we review the associated bootstrap-like techniques. We detail how to get asymptotically distribution-free test statistics in Sect. 4.3, and we explain some testing procedures that exploit the particular features of copulas. We discuss some ways of testing the belonging to some "large" infinite-dimensional families of copulas like Archimedean, extreme-value, vine, or HAC copulas in Sect. 4.4. Tests adapted to time-dependent copulas are introduced in Sect. 4.5. Finally, empirical performances of these GOF tests are discussed in Sect. 4.6.

4.2 The "Brute-Force" Approach: The Empirical Copula Process and the Bootstrap

4.2.1 Some Tests Based on Empirical Copula Processes

Such copula GOF tests are the parallels of the most standard GOF tests in the literature, replacing F_n (resp. F_0) by C_n (resp. C_0). These statistics are based on distances between the empirical copula C_n and the true copula C_0 (simple zero assumption), or between C_n and $C_{\hat{\theta}_n}$ (composite zero assumption), for some convergent and convenient estimator $\hat{\theta}_n$ of the "true" copula parameter θ_0. It is often reduced simply to the evaluation of norms of the empirical copula process $\mathbb{C}_n := \sqrt{n}(C_n - C_0)$, or one of its approximations $\hat{\mathbb{C}}_n := \sqrt{n}(C_n - C_{\hat{\theta}_n})$.

In this family, let us cite the KS type statistics

$$T_n^{KS} := \|\mathbb{C}_n\|_\infty = \sup_{\mathbf{u}\in[0,1]^d} |\sqrt{n}(C_n - C_0)(\mathbf{u})|,$$

and the AD type statistics

$$T_n^{AD} := \|\mathbb{C}_n\|_{L^2} = n \int (C_n - C_0)^2(\mathbf{u}) w_n(\mathbf{u})\, d\mathbf{u},$$

for some positive (possibly random) weight function w_n, and their composite versions. By smoothing conveniently the empirical copula process, [70] defined alternative versions of the latter tests.

In practice, the statistics T_n^{KS} seem to be less powerful than a lot of competitors, particularly of the type T_n^{AD} (see [46]). Therefore, a "total variation" version of T_n^{KS} has been proposed in [37] that appears significantly more powerful than the classical T_n^{KS}:

$$T_n^{ATV} := \sup_{B_1,\dots,B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{C}_n(B_k)|, \quad \text{or} \quad \hat{T}_n^{ATV} := \sup_{B_1,\dots,B_{L_n}} \sum_{k=1}^{L_n} |\hat{\mathbb{C}}_n(B_k)|,$$

for simple or composite assumptions, respectively. Above, the supremum is taken over all disjoint rectangles $B_1, \dots, B_{L_n} \subset [0,1]^d$, and $L_n \sim \ln n$.

Another example of distance is proposed in [71]: let two functions f_1 and f_2 in $\mathbb{R}^d$. Typically, they represent copula densities. Set a positive definite bilinear form as

$$< f_1, f_2 >:= \int \kappa_d(\mathbf{x}_1, \mathbf{x}_2)\, f_1(\mathbf{x}_1) f_2(\mathbf{x}_2)\, d\mathbf{x}_1\, d\mathbf{x}_2,$$

where $\kappa_d(\mathbf{x}_1, \mathbf{x}_2) := \exp(-\|\mathbf{x}_1 - \mathbf{x}_2\|^2/(2dh^2))$, for some Euclidian norm $\|\cdot\|$ in $\mathbb{R}^d$ and a bandwidth $h > 0$. A squared distance between f_1 and f_2 is given simply by $\mu(f_1, f_2) :=< f_1 - f_2, f_1 - f_2 >=< f_1, f_1 > -2 < f_1, f_2 > + < f_2, f_2 >$. When f_1 and f_2 are the copula densities of C_1 and C_2, respectively, the three latter terms can be rewritten in terms of copula directly. For instance, $< f_1, f_2 >= \int \kappa_d(\mathbf{x}_1, \mathbf{x}_2)\, C_1(d\mathbf{x}_1)\, C_2(d\mathbf{x}_2)$. Since such expressions have simple empirical counterparts, a GOF test for copulas can be built easily: typically, replace C_1 by the empirical copula C_n and C_2 by the true copula C_0 (or $C_{\hat{\theta}_n}$).

Closely connected to this family of tests are statistics T_n that are zero when the associated copula processes are zero, but not the opposite. Strictly speaking, this is the case of the CvM statistics

$$T_n^{CvM} := n \int (C_n - C_0)^2(\mathbf{u})\, C_n(d\mathbf{u}),$$

and of chi-squared type test statistics, like

$$T_n^{Chi} := n \sum_{k=1}^{p} w_k (C_n - C_0)^2(B_k),$$

where $B_1, \dots, B_p$ denote disjoint boxes in $[0,1]^d$ and w_k, $k = 1, \dots, p$ are convenient weights (possibly random). More generally, we can consider

$$T_n^{\mu} := \sum_{k=1}^{p} \mu(C_n(E_k), C_0(E_k)), \quad \text{or} \quad T_n^{\mu} := \sum_{k=1}^{p} \mu(C_n(E_k), C_{\hat{\theta}_n}(E_k)),$$

for any metric μ on the real line, and arbitrary subsets $E_1, \ldots, E_p$ in $[0,1]^d$. This is the idea of the chi-square test detailed in [30]: set the vectors of pseudo-observations $\hat{\mathbf{U}}_i := (F_{n,1}(X_{i,1}), \ldots, F_{n,d}(X_{i,d}))$, and a partition of $[0,1]^d$ into p disjoint rectangles B_j. The natural chi-square-style test statistics is

$$T_n^{\chi} := \sum_{k=1}^{p} \frac{\left(\hat{N}_k - p_k(\hat{\theta}_n)\right)^2}{n p_k(\hat{\theta}_n)}$$

where $\hat{N}_k$ denotes the number of vectors $\hat{\mathbf{U}}_i$, $i = 1, \ldots, n$ that belong to B_k, and $p_k(\theta)$ denotes the probability of the event $\{\mathbf{U} \in B_k\}$ under the copula C_θ. This idea of applying an arbitrary categorization of the data into contingency tables $[0,1]^d$ has been applied more or less fruitfully in a lot of papers: [4, 33, 41, 58, 59], etc.

Finally, note that a likelihood ratio test has been proposed in [30], based on a Kullback–Leibler pseudo distance between a "discrete" version of C_n and the corresponding estimated copula under the null:

$$T_n^{LR} := \sum_{k=1}^{p} N_k \ln p_k(\hat{\theta}_n).$$

To compare the fit of two potential parametric copulas, the same information criterion has been used in [28] to build a similar test statistics but based on copula densities directly.

The convergence of all these tests relies crucially on the fact that the empirical copula processes $\mathbb{C}_n$ and $\hat{\mathbb{C}}_n$ are weakly convergent under the null, and for convenient sequences of estimates $\hat{\theta}_n$: see [36, 38, 82]. Particularly, it has been proved that $\mathbb{C}_n$ tends weakly in $\ell^\infty([0,1]^d)$ (equipped with the metric induced by the sup-norm) to a Gaussian process $\mathbb{G}_{C_0}$, where

$$\mathbb{G}_{C_0}(\mathbf{u}) := \mathbb{B}_{C_0}(\mathbf{u}) - \sum_{j=1}^{d} \partial_j C_0(\mathbf{u}) \mathbb{B}_{C_0}(u_j, \mathbf{1}_{-j}), \ \forall \mathbf{u} \in [0,1]^d,$$

with obvious notations and for some d-dimensional Brownian bridge $\mathbb{B}_{C_0}$ in $[0,1]^d$, whose covariance is

$$\mathbb{E}\left[\mathbb{B}_{C_0}(\mathbf{u}) \mathbb{B}_{C_0}(\mathbf{v})\right] = C_0(\mathbf{u} \wedge \mathbf{v}) - C_0(\mathbf{u}) C_0(\mathbf{v}), \quad \forall (\mathbf{u}, \mathbf{v}) \in [0,1]^{2d}.$$

To get this weak convergence result, it is not necessary to assume that C_0 is continuously differentiable on the whole hypercube $[0,1]^d$, a condition that is often not fulfilled in practice. Recently, [87] has shown that such a result is true when, for every $j = 1, \ldots, d$, $\partial_j C_0$ exists and is continuous on the set $\{\mathbf{u} \in [0,1]^d, 0 < u_j < 1\}$.

Clearly, the law of $\mathbb{G}$ involves the particular underlying copula C_0 strongly contrary to usual Brownian bridges. Therefore, the tabulation of the limiting laws of T_n GOF statistics appears difficult. A natural idea is to rely on computer intensive methods to approximate these law numerically. The bootstrap appeared as a natural tool for doing this task

4.2.2 Bootstrap Techniques

The standard nonparametric bootstrap is based on resampling with replacement inside an original i.i.d. $\mathbf{X}$-sample $S_{\mathbf{X}}$. We get new samples $S_{\mathbf{X}}^* = (\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*)$. Associate to every new sample $S_{\mathbf{X}}^*$ its "bootstrapped" empirical copula C_n^* and its bootstrapped empirical process $\mathbb{C}_n^* := \sqrt{n}(C_n^* - C_n)$. In [36], it is proved that, under mild conditions, this bootstrapped process $\mathbb{C}_n^*$ is weakly convergent in $\ell^\infty([0,1]^d)$ towards the previous Gaussian process $\mathbb{G}_{C_0}$. Therefore, in the case of simple null assumptions, we can get easily some critical values or p-values of the previous GOF tests: resample M times, $M >> 1$, and calculate the empirical quantiles of the obtained bootstrapped test statistics. Nonetheless, this task has to be done for every zero assumption. This can become a tedious and rather long task, especially when d is "large" (> 3 in practice) and/or with large datasets ($> 1,000$, typically).

When dealing with composite assumptions, some versions of the parametric bootstrap are advocated, depending on the limiting behavior of $\hat{\theta}_n - \theta_0$: see the theory in [40], and the appendices in [46] for detailed examples. To summarize these ideas in typical cases, it is now necessary to draw random samples from $C_{\hat{\theta}_n}$. For every bootstrapped sample, calculate the associated empirical copula C_n^* and a new estimated value $\hat{\theta}_n^*$ of the parameter. Since the weak limit of $\sqrt{n}(C_n^* - C_{\hat{\theta}_n^*})$ is the same as the limit of $\hat{\mathbb{C}}_n = \sqrt{n}(C_n - C_{\hat{\theta}_n})$, the law of every functional of $\hat{\mathbb{C}}_n$ can be approximated. When the cdf $C_{\hat{\theta}_n}$ cannot be evaluated explicitly (in closed-form), a two-level parametric bootstrap has been proposed in [40], by bootstrapping first an approximated version of $C_{\hat{\theta}_n}$.

Instead of resampling with replacement, a multiplier bootstrap procedure can approximate the limiting process $\mathbb{G}_{C_0}$ (or one of its functionals), as in [80]: consider $Z_1, \ldots, Z_n$ i.i.d. real centered random variables with variance one, independent of the data $\mathbf{X}_1, \ldots, \mathbf{X}_n$. A new bootstrapped empirical copula is defined by

$$C_n^*(\mathbf{u}) := \frac{1}{n} \sum_{i=1}^{n} Z_i . \mathbf{1}(F_{n,1}(X_{i,1}) \leq u_1, \ldots, F_{n,d}(X_{i,d}) \leq u_d),$$

for every $\mathbf{u} \in [0,1]^d$. Setting $\bar{Z}_n := n^{-1} \sum_{i=1}^n Z_i$, the process $\beta_n := \sqrt{n}(C_n^* - \bar{Z}_n C_n)$ tends weakly to the Brownian bridge $\mathbb{B}_{C_0}$. By approximating (by finite differences) the derivatives of the true copula function, it is shown in [80] how to modify β_n to get an approximation of $\mathbb{G}_{C_0}$. To avoid this last stage, another

bootstrap procedure has been proposed in [14]. It applies the multiplier idea to the underlying joint and marginal cdfs' and invoke classical delta method arguments. Nonetheless, despite more attractive theoretical properties, the latter technique does not seem to improve the initial multiplier bootstrap of [80]. In [61], the multiplier approach is extended to deal with parametric copula families of any dimension, and the finite-sample performance of the associated CvM test statistics has been studied. A variant of the multiplier approach has been proposed in [60]. It is shown that the use of multiplier approaches instead of the parametric bootstrap leads to a strong reduction in the computing time. Note that both methods have been implemented in the copula R package.

Recently, in [37], a modified nonparametric bootstrap technique has been introduced to evaluate the limiting law of the previous KS-type test statistics T_n^{ATV} in the case of composite zero assumptions. In this case, the key process is still

$$\hat{\mathbb{C}}_n := \sqrt{n}(C_n - C_{\hat{\theta}_n}) = \mathbb{C}_n - \sqrt{n}(C_{\hat{\theta}_n} - C_{\theta_0}).$$

Generate a usual nonparametric bootstrap sample, obtained after resampling with replacement from the original sample. This allows the calculation of the bootstrapped empirical copula C_n^* and a new parameter estimate $\hat{\theta}_n^*$. Instead of considering the "intuitive" bootstrapped empirical copula process $\sqrt{n}(C_n^* - C_{\hat{\theta}_n^*})$, a new bootstrapped process is introduced:

$$\mathbb{Y}_n^* := \sqrt{n}(C_n^* - C_n) - \sqrt{n}(C_{\hat{\theta}_n^*} - C_{\hat{\theta}_n}).$$

Indeed, the process $\sqrt{n}(C_n^* - C_{\hat{\theta}^*})$, while perhaps a natural candidate, does not yield a consistent estimate of the distribution of $\hat{\mathbb{C}}_n$, contrary to $\mathbb{Y}_n^*$. For the moment, the performances of this new bootstrapped process have to be studied more in depth.

4.3 Copula GOF Test Statistics: Alternative Approaches

4.3.1 Working with Copula Densities

Even if the limiting laws of the empirical copula processes $\mathbb{C}_n$ and $\hat{\mathbb{C}}_n$ involve the underlying (true) copula in a rather complex way, it is still possible to get asymptotically distribution-free test statistics. Unfortunately, the price to be paid is an additional level of complexity.

To the best of our knowledge, there exists a single strategy. The idea is to rely on copula densities themselves, rather than copulas (cdfs'). Indeed, testing the identity $C = C_0$ is equivalent to studying the closeness between the true copula density τ_0 (w.r.t. the Lebesgue measure on $[0,1]^d$, that is assumed to exist) and one of its estimates τ_n. In [33], a L^2-distance between τ_n and τ_0 allows to build convenient

test statistics. To be specific, a kernel estimator of a copula density τ at point $\mathbf{u}$ is defined by

$$\tau_n(\mathbf{u}) = \frac{1}{h^d} \int K\left(\frac{\mathbf{u}-\mathbf{v}}{h}\right) C_n(d\mathbf{v}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{u}-\hat{\mathbf{U}}_i}{h}\right),$$

where $\hat{\mathbf{U}}_i := (F_{n,1}(X_{i,1}), \ldots, F_{n,d}(X_{i,d}))$ for all $i = 1, \ldots, n$. Moreover, K is a d-dimensional kernel and $h = h(n)$ is a bandwidth sequence, chosen conveniently. Under some regularity assumptions, for every m and every vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ in $]0,1[^d$, such that $\tau_0(\mathbf{u}_k) > 0$ for every k, the vector $(nh^d)^{1/2}((\tau_n - \tau_0)(\mathbf{u}_1), \ldots, (\tau_n - \tau_0)(\mathbf{u}_m))$ tends weakly to a Gaussian random vector, whose components are independent. Therefore, under the null, the test statistics

$$T_n^{\tau,0} = \frac{nh^d}{\int K^2} \sum_{k=1}^{m} \frac{(\tau_n(\mathbf{u}_k) - \tau_0(\mathbf{u}_k))^2}{\tau_0(\mathbf{u}_k)^2}.$$

tends in law towards a m-dimensional chi-squared distribution. This can be adapted easily for composite assumptions. The previous test statistics depend on a finite and arbitrary set of points $\mathbf{u}_k, k = 1, \ldots, m$. To avoid this drawback, [33] has introduced

$$J_n = \int (\tau_n - K_h * \hat{\tau})^2(\mathbf{u})\omega(\mathbf{u})\, d\mathbf{u},$$

for some nonnegative weight function ω. Here, $\hat{\tau}$ denotes τ_0 (simple assumption) or $\tau(\cdot, \hat{\theta}_n)$ (composite assumption), for sufficiently regular estimates $\hat{\theta}_n$ of θ_0. It is proved that

$$T_n^{\tau,1} := \frac{n^2 h^d \left(J_n - (nh^d)^{-1} \int K^2(\mathbf{t}).(\hat{\tau}\omega)(\mathbf{u}-h\mathbf{t})\, d\mathbf{t}\, d\mathbf{u} + (nh)^{-1} \int \hat{\tau}^2\omega . \sum_{r=1}^{d} \int K_r^2\right)^2}{2 \int \hat{\tau}^2\omega \cdot \int \left\{\int K(\mathbf{u})K(\mathbf{u}+\mathbf{v})\, d\mathbf{u}\right\}^2 d\mathbf{v}}$$

tends to a $\chi^2(1)$ under the null.

Even if the previous test statistics are pivotal, they are rather complex and require the choice of smoothing parameters and kernels. Nonetheless, such ideas have been extended in [86] to deal with the fixed design case. Moreover, the properties of these tests under fixed alternatives are studied in [13]. The impact of several choices of parameter estimates $\hat{\theta}_n$ on the asymptotic behavior of J_n is detailed too. Apparently, for small sample sizes, the normal approximation does not provide sufficiently exact critical values (in line with [51] or [32]), but it is still possible to use a parametric bootstrap procedure to evaluate the limiting law of T_n^{τ} in this case. Apparently, in the latter case, the results are as good as the main competitors (see [13], Sect. 5).

Since copula densities have a compact support, kernel smoothing can generate some undesirable boundary effects. One solution is to use improved kernel

estimators that take care of the typical corner bias problem, as in [70]. Another solution is to estimate copula densities through wavelets, for which the border effects are handled automatically, due to the good localization properties of the wavelet basis: see [45]. This idea has been developed in [39], in a minimax theory framework, to determine the largest alternative for which the decision remains feasible. Here, the copula densities under consideration are supposed to belong to a range of Besov balls. According to the minimax approach, the testing problem is then solved in an adaptive framework.

4.3.2 The Probability Integral Transformation

A rather simple result of probability theory, proposed initially in [81], has attracted the attention of authors for copula GOF testing purpose. Indeed, this transformation maps a general d-dimensional random vector $\mathbf{X}$ into a vector of d independent uniform random variables on $[0, 1]$ in a one-to-one way. It is known as Rosenblatt's or probability integral transformation (PIT). Once the joint law of $\mathbf{X}$ is known and analytically tractable, this is a universal way of generating independent and uniform random vectors without losing statistical information. Note that other transformations of the same type exist (see [22]).

To be specific, the copula C is the joint cdf of $\mathbf{U} := (F_1(X_1), \ldots, F_d(X_d))$. We define the d-dimensional random vector $\mathbf{V}$ by

$$V_1 := U_1 = F_1(Z_1),\ V_2 := C(U_2|U_1), \cdots, V_d := C(U_d|U_1, \ldots, U_{d-1}), \quad (4.1)$$

where $C(\cdot|u_1, \ldots, u_{k-1})$ is the law of U_k given $U_1 = u_1, \ldots, U_{k-1} = u_{k-1}$, $k = 2, \ldots, d$. Then, the variables V_k, $k = 1, \ldots, d$ are uniformly and independently distributed on $[0, 1]$. In other words, $\mathbf{U} \sim C$ iff $\mathbf{V} = \mathscr{R}(\mathbf{U})$ follows the d-variate independence copula $C_\perp(\mathbf{u}) = u_1.\cdots.u_d$.

The main advantage of this transformation is the simplicity of the transformed vector $\mathbf{V}$. This implies that the zero assumptions of a GOF test based on $\mathbf{V}$ are always the same: test the i.i.d. feature of $\mathbf{V}$, that is satisfied when C is the true underlying copula. A drawback is the arbitrariness in the choice of the successive margins. Indeed, there are at most $d!$ different PITs that induce generally different test statistics. Another disadvantage is the necessity of potentially tedious calculations. Indeed, typically, the conditional joint distributions are calculated through the formulas

$$\begin{aligned} &C(u_k|u_1, \ldots, u_{k-1}) \\ &= \partial^{k-1}_{1,2,\ldots,k-1} C(u_1, \ldots, u_k, 1, \ldots, 1) / \partial^{k-1}_{1,2,\ldots,k-1} C(u_1, \ldots, u_{k-1}, 1, \ldots, 1), \end{aligned}$$

for every $k = 2, \ldots, d$ and every $\mathbf{u} \in [0, 1]^d$. Therefore, with some copula families and/or with large dimensions d, the explicit calculation (and coding!) of the PIT can become unfeasible.

The application of such transformations for copula GOF testing appeared first in [12]. This idea has been reworked and extended in several papers afterwards: see [8, 10, 31, 43], etc. Several applications of such techniques to financial series modeling and risk management has emerged, notably [19, 27, 63, 65, 92], among others.

For copula GOF testing, we are only interested in the copula itself, and the marginal distributions $F_k, k = 1, \ldots, d$ are seen as nuisance parameters. Therefore, they are usually replaced by the marginal empirical cdfs' $F_{n,k}$. Equivalently, the observations $\mathbf{X}_i$, $i = 1, \ldots, n$ are often replaced by their pseudo-observations $\hat{\mathbf{U}}_i := (F_{n,1}(X_{i,1}), \ldots, F_{n,d}(X_{i,d}))$. Moreover, for composite zero assumptions, the chosen estimator $\hat{\theta}_n$ disturbs the limiting law of the test statistics most of the time. This difficulty is typical of the statistics of copulas, and it is a common source of mistakes, as pointed out in [34]. For instance, in [12], these problems were not tackled conveniently and the reported p-values are incorrect. Breymann et al. [12] noticed that the r.v. $\sum_{k=1}^{d}[\Phi^{-1}(V_k)]^2$ follows a $\chi^2(d)$. But it is no more the case of $\sum_{k=1}^{d}[\Phi^{-1}(\hat{V}_{n,k})]^2$, where $\hat{\mathbf{V}} = \mathscr{R}(\hat{\mathbf{U}})$. This point has been pointed out in [40]. A corrected test statistics with reliable p-values has been introduced in [31]. An extension of these tests has been introduced in [10]. It implies data-driven weight functions to emphasize some regions of the underlying copula possibly. Its comparative performances are studied in [8, 9].

Thus, to the best of our knowledge, all the previous proposed tests procedures have to rely on bootstrap procedures to evaluate the corresponding limiting laws under the null. This is clearly a shame, keeping in mind the simplicity of the law of $\mathbf{V}$, after a PIT of the original dataset (but with *known* margins). In practice, we have to work with (transformed) pseudo-observations $\hat{\mathbf{V}}_i$, $i = 1, \ldots, n$. As we said, they are calculated from formulas (4.1), replacing unobservable uniformly distributed vectors $\mathbf{U}_i$ by pseudo-observations $\hat{\mathbf{U}}_i$, $i = 1, \ldots, n$. The vectors $\hat{\mathbf{V}}_i$ are no longer independent and only approximately uniform on $[0, 1]^d$. Nonetheless, test statistics $T_n^{\psi,PIT} = \psi(\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_n)$ may be relevant, for convenient real functions ψ. In general and for composite zero assumptions, we are not insured that the law of $\hat{\mathbf{V}}$, denoted by $C_{\infty,\mathbf{V}}$, tends to the independence copula. If we were able to evaluate $C_{\infty,\mathbf{V}}$, a "brute-force" approach would still be possible, as in Sect. 4.2. For instance and naively, we could introduce the Kolmogorov-type statistics

$$T_n^{KM,PIT} := \sup_{\mathbf{u}\in(0,1)^d} |\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{\mathbf{V}}_i \leq \mathbf{u}) - C_{\infty,\mathbf{V}}(\mathbf{u})|.$$

Nonetheless, due to the difficulty to evaluate precisely $C_{\infty,\mathbf{V}}$ (by Monte-Carlo, in practice), most authors have preferred to reduce the dimensionality of the problem. By this way, they are able to tackle more easily the case $d \geq 3$.

4.3.3 Reductions of Dimension

Generally speaking, in a GOF test, it is tempting to reduce the dimensionality of the underlying distributions, for instance from d to one. Indeed, especially when $d >> 1$, the "brute-force" procedures based on empirical processes involve significant analytical or numerical difficulties in practice. For instance, a CvM necessitates the calculation of a d-dimensional integral.

Formally, a reduction of dimension means replacing the initial GOF problem "$\mathscr{H}_0$: the copula of $\mathbf{X}$ is C_0" by "$\mathscr{H}_0^*$: the law of $\psi(\mathbf{X})$ is $G_{\psi,0}$", for some transformation $\psi : \mathbb{R}^d \to \mathbb{R}^p$, with $p << d$, and for some p-dimensional cdf $G_{\psi,0}$. As $\mathscr{H}_0$ implies $\mathscr{H}_0^*$, we decide to reject $\mathscr{H}_0$ when $\mathscr{H}_0^*$ is not satisfied. Obviously, this reduction of the available information induces a loss of power, but the practical advantages of this trick often dominate its drawbacks.

For instance, when $p = 1$ and if we are able to identify $G_{\psi,0}$, it becomes possible to invoke standard univariate GOF test statistics, or even to use ad-hoc visual procedures like QQ-plots. Thus, by reducing a multivariate GOF problem to a univariate problem, we rely on numerically efficient procedures, even for high dimensional underlying distributions. However, we still depend on Monte-Carlo methods to evaluate the corresponding p-values. Inspired by [83], we get one of the most naive methods of dimension reduction: replace T_n^{KS} above by

$$\tilde{T}_n^{KS} := \sum_{\alpha\in(0,1)} |C_n(A_\alpha) - C_0(A_\alpha)|, \quad \text{or} \quad \tilde{T}_n^{KS} := \sum_{\alpha\in(0,1)} |C_n(\hat{A}_\alpha) - C_{\hat{\theta}_n}(\hat{A}_\alpha)|,$$

where $(A_\alpha)_{\alpha\in(0,1)}$ is an increasing sequence of subsets in $[0,1]^d$ s.t. $A_\alpha = \{\mathbf{u} \in [0,1]^d | C_0(\mathbf{u}) \leq \alpha\}$ and $\hat{A}_\alpha = \{\mathbf{u} \in [0,1]^d | C_{\hat{\theta}_n}(\mathbf{u}) \leq \alpha\}$.

To revisit a previous example and with the same notations, [31] considered particular test statistics $T_n^{\psi,PIT}$ based on the variables $\hat{Z}_i := \sum_{k=1}^d \Phi(\hat{V}_{i,k})^{-1}$, $i = 1, \ldots, n$. If the margins F_k, $k = 1, \ldots, d$, and the true copula C_0 were known, then we were able to calculate $Z_i := \sum_{k=1}^d \Phi(V_{i,k})^{-1}$ that follows a chi-square law of dimension d under the null. Since it is not the case in practice, the limiting law of $\hat{Z}_i$ is unknown, and it has to be evaluated numerically by simulations. It is denoted by $F_{\hat{Z}}$. Therefore, [31] propose to test

$$\mathscr{H}_0^* : \text{the asymptotic law of } T_n^{\psi,PIT} \text{ is a given cdf } F_\psi \text{ (to be estimated)},$$

where $T_n^{\psi,PIT}$ is defined by usual (univariate) KS, AD or CvM test statistics. For instance,

$$T_n^{AD,PIT} := n \int \frac{(F_{n,\hat{Z}} - F_{0,\hat{Z}})^2}{F_{0,\hat{Z}}(1 - F_{0,\hat{Z}})},$$

where $F_{n,\hat{Z}}$ is the empirical cdf of the pseudo sample $\hat{Z}_1, \ldots, \hat{Z}_n$. Note that $F_{n,\hat{Z}}$ and $F_{0,\hat{Z}}$ depend strongly on the underlying cdf of $\mathbf{X}$, its true copula C_0, the way marginal cdfs have been estimated to get pseudo-observations (empirical or parametric estimates) and possibly the particular estimate $\hat{\theta}_n$.

Beside the PIT idea, there exist a lot of possibilities of dimension reductions potentially. They will provide more or less relevant test statistics, depending on the particular underlying parametric family and on the empirical features of the data. For instance, in the bivariate case, Kendall's tau τ_K or Spearman's rho ρ_S may appear as nice "average" measures of dependence. They are just single numbers, instead of a true two-dimensional function like C_n. Therefore, such a GOF test may be simply

$$\mathscr{H}_0^* : \hat{\tau}_K = \tau_{K,C_0},$$

where $\tau_{K,C_0} = 4\mathbb{E}_{C_0}[C_0(\mathbf{U})] - 1$ is the Kendall's tau of the true copula C_0, and $\hat{\tau}_K$ is an estimate of this measure of dependence, for instance its empirical counterpart

$$\hat{\tau}_{K,n} := \frac{2[\text{number of concordant pairs of observations} - \text{number of discordant pairs}]}{n(n-1)}.$$

Here, we can set $T_n^{KTau} := n(\hat{\tau}_{K,n} - \tau_{C_0})^2$, or $T_n^{KTau} := n(\hat{\tau}_{K,n} - \tau_{C_{\hat{\theta}_n}})^2$ in the case of composite assumption. Clearly, the performances of all these tests in terms of power will be very different and there is no hope to get a clear hierarchy between all of them. Sometimes, it will be relevant to discriminate between several distributions depending on the behaviors in the tails. Thus, some adapted summaries of the information provided by the underlying copula C are required, like tail-indices for instance (see e.g., [68]). But in every case, their main weakness is a lack of convergence against a large family of alternatives. For instance, the previous test T_n^{KTau} will not be able to discriminate between all copulas that have the same Kendall's tau τ_{K,C_0}. In other words, this dimension reduction is probably too strong most of the time: we reduce a d-dimensional problem to a real number. It is more fruitful to keep the idea of generating a univariate process, i.e., going from a dimension d to a dimension one. This is the idea of Kendall's process (see below).

Another closely related family of tests is based on the comparison between several parameter estimates. They have been called "moment-based" GOF test statistics (see [11, 44, 88]). In their simplest form, assume a univariate unknown copula parameter θ, and two estimation equations ("moments") such that $m_1 = r_1(\theta)$ and $m_2 = r_2(\theta)$ (one-to-one mappings). Given empirical counterparts $\hat{m}_k$ of m_k, $k = 1, 2$, [88] has proposed the copula GOF test

$$T_n^{moment} := \sqrt{n}\left\{r_1^{-1}(\hat{m}_1) - r_2^{-1}(\hat{m}_2)\right\}.$$

Typically, some estimating equations are provided by Kendall's tau and Spearman's rho that have well-known empirical counterparts. Nonetheless, other estimates have been proposed, as the pseudo-maximum likelihood (also called "canonical

maximum likelihood"). To deal with multi-dimensional parameters θ, estimating equations can be obtained by the equality between the hessian matrix and minus the expected outer product of the score function. This is the idea of White's specification test (see [93]), adapted to copulas in [76].

4.3.4 Kendall's Process

This is another and well-known example of dimension reduction related to copula problems. Let C be the copula of an arbitrary random vector $\mathbf{X} \in \mathbb{R}^d$. Define the univariate cdf

$$K(t) := \mathbb{P}(C(\mathbf{U}) \leq t), \quad \forall t \in \mathbb{R},$$

where, as usual, we set $\mathbf{U} = (F_1(X_1), \ldots, F_d(X_d))$. The function K depends on C only. Therefore, this univariate function is a "summary" of the underlying dependence structure given by C. It is called the Kendall's dependence function of C. An empirical counterpart of K is the empirical Kendall's function

$$K_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(C_n(\hat{\mathbf{U}}_i) \leq t),$$

with pseudo-observations $\hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_n$. The associated Kendall's process is simply given by $\mathbb{K}_n = \sqrt{n}(K_n - K)$, or $\hat{\mathbb{K}}_n = \sqrt{n}(K_n - K(\hat{\theta}_n, \cdot))$ when the true copula is unknown but belongs to a given parametric family. The properties of Kendall's processes have been studied in depth in [5, 43, 48] particularly. In the later papers, the weak convergence of $\mathbb{K}_n$ towards a continuous centered Gaussian process in the Skorohod space of cadlag functions is proved, for convenient consistent sequences of estimates $\hat{\theta}_n$. Its variance-covariance function is complex and copula dependent. It depends on the derivatives of K w.r.t. the parameter θ and the limiting law of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Then, there are a lot of possibilities of GOF tests based on the univariate function K_n or the associated process $\mathbb{K}_n$. For instance, [90] introduced a test statistics based on the L^2 norm of $\mathbb{K}_n$. To be specific, they restrict themselves to bivariate Archimedean copulas, but allow censoring. That is why their GOF test statistics $T_n^{L2,Kendall} = \int_\xi^1 |\mathbb{K}_n|^2$ involves an arbitrary cutoff point $\xi > 0$. Nonetheless, the idea of such a statistics is still valid for arbitrary dimensions and copulas. It has been extended in [43] that considers

$$T_n^{L2,Kendall} := \int_0^1 |\mathbb{K}_n(t)|^2 k(\hat{\theta}_n, t)\, dt, \quad \text{and} \quad T_n^{KS,Kendall} := \sup_{t \in [0,1]} |\mathbb{K}_n(t)|,$$

where $k(\theta, \cdot)$ denotes the density of $C(\mathbf{U})$ w.r.t. to the Lebesgue measure (i.e., the derivative of K) and $\hat{\theta}_n$ is a consistent estimate of the true parameter under the null.

Nonetheless, working with $\mathbb{K}_n$ or $\hat{\mathbb{K}}_n$ instead of $\mathbb{C}_n$ or $\hat{\mathbb{C}}_n$, respectively, is not the panacea. As we said, the dimension reduction is not free of charge, and testing $\mathscr{H}_0^*$ instead of $\mathscr{H}_0$ reduces the ability to discriminate between copula alternatives. For instance, consider two extreme-value copulas C_1 and C_2, i.e., in the bivariate case,

$$C_j(u,v) = \exp\left(\ln(uv) A_j\left(\frac{\ln u}{\ln uv}\right)\right), \quad j = 1, 2,$$

for some Pickands functions A_1 and A_2 (convex functions on $[0,1]$, such that $\max(t, (1-t)) \leq A_j(t) \leq 1$ for all $t \in [0,1]$). As noticed in [48], the associated Kendall's functions are

$$K_j(t) = t - (1 - \tau_{K,j}) t \ln t, \quad t \in (0,1),$$

where $\tau_{K,j}$ denotes the Kendall's tau of C_j. Then, if the two Kendall's tau are the same, the corresponding Kendall's functions K_1 and K_2 are identical. Thus, a test of $\mathscr{H}_0^* : K = K_0$ will appear worthless if the underlying copulas are of the extreme-value type.

In practice, the evaluation of the true Kendall function K_0 under the null may become tedious, or even unfeasible for a lot of copula families. Therefore, [9] proposed to apply the previous Kendall process methodology to random vectors obtained through a PIT in a preliminary stage, to "stabilize" the limiting law under the null. In this case, K_0 is always the same: the Kendall function associated with the independence copula $C_\perp$. This idea has been implemented in [46], under the form of CvM GOF test statistics of the type

$$T_n^{CvM,PIT} := n \int \left(D_n(\mathbf{u}) - C_\perp(\mathbf{u})\right)^2 \, dD_n(\mathbf{u}) = \sum_{i=1}^{n} \left(D_n(\hat{\mathbf{U}}_i) - C_\perp(\hat{\mathbf{U}}_i)\right)^2,$$

where $D_n(\mathbf{u}) = n^{-1} \sum_{i=1}^{n} \mathbf{1}(\hat{\mathbf{U}}_i \leq \mathbf{u})$ is the empirical cdf associated with the pseudo-observations of the sample. Nonetheless, the limiting behavior of all these test statistics are not distribution-free for composite zero assumptions, and limiting laws have to be evaluated numerically by Monte-Carlo methods (as usual).

Note that [78] have proposed a similar idea, but based on Spearman's dependence function L instead of Kendall's dependence function. Formally, L is defined by

$$L(u) := \mathbb{P}\left(C_\perp(\mathbf{U}) \leq u\right) = \mathbb{P}\left(\prod_{k=1}^{d} F_k(X_k) \leq u\right), \quad \forall u \in [0,1].$$

When working with a random sample, the empirical counterpart of L is then

$$\hat{L}_n(u) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(C_\perp(\hat{\mathbf{U}}_i) \le u\right),$$

and all the previous GOF test statistics may be applied. For instance, [8] proposed to use the CvM statistic

$$T_n^{L,CvM} := \int_0^1 \left(\hat{L}_n - L_{\hat{\theta}_n}\right)^2 \hat{L}_n(du),$$

where $L(\theta)$ is the Spearman's dependence function of an assumed copula C_θ and $\hat{\theta}_n$ is an estimate of the true parameter under the zero assumption.

4.4 GOF Tests for Some Particular Classes of Copulas

Beside omnibus GOF tests, there exist other test statistics that are related to particular families of copulas only. We will not study such GOF tests when they are related to particular finite-dimensional parametric families (to decide whether C_0 is a Gaussian copula, for instance). Nonetheless, in this section, we will be interested in a rather unusual GOF problem: to say whether C_0 belongs to a particular infinite-dimensional parametric family of copulas. Among such large families, some of them are important in practice: the Archimedean family, the elliptical one, extreme-value copulas, vines, hierarchical Archimedean copulas, etc.

4.4.1 Testing the Archimedeanity

All the previously proposed test statistics can be applied when $\mathscr{C}$ is an assumed particular Archimedean family, as in [85, 90]. Other test statistics, that are based on some analytical properties of Archimedean copulas, have been proposed too (e.g., [52]). Interestingly, [41] proposed a graphical procedure for selecting a Archimedean copula (among several competitors), through a visual comparison between the empirical Kendall's function K_n and an estimated Kendall function obtained under a composite null hypothesis $\mathscr{H}_0$.

Now, we would like to test "$\mathscr{H}_0 : C$ is Archimedean" against the opposite, i.e., without any assumption concerning a particular parametric family. This problem has not received a lot of attention in the literature, despite its practical importance.

Consider first the (unknown) generator ϕ of the underlying bivariate copula C, i.e. $C(\mathbf{u}) = \phi^{-1}(\phi(u_1) + \phi(u_2))$ for every $\mathbf{u} = (u_1, u_2) \in [0,1]^2$. Genest and Rivest [41] proved that $V_1 := \phi(F_1(X_1))/\{\phi(F_1(X_1)) + \phi(F_2(X_2))\}$ is uniformly distributed on $(0,1)$ and that $V_2 := C(F_1(X_1), F_2(X_2))$ is distributed as the Kendall's dependence function $K(t) = t - \phi(t)/\phi'(t)$. Moreover, V_1 and V_2 are

independent. Since K can be estimated empirically, these properties provide a way of estimating ϕ itself (by ϕ_n). Therefore, as noticed in the conclusion of [41], if the underlying copula is Archimedean, then the r.v.

$$\hat{V}_1 := \phi_n(F_{1,n}(X_1))/\{\phi(F_{1,n}(X_1)) + \phi(F_{2,n}(X_2))\}$$

should be distributed uniformly on $(0, 1)$ asymptotically. This observation can lead to some obvious GOF test procedures.

Another testing strategy starts from the following property, proved in [68]: a bivariate copula C is Archimedean iff it is associative (i.e., $C(u_1, C(u_2, u_3)) = C(C(u_1, u_2), u_3)$ for every triplet (u_1, u_2, u_3) in $[0, 1]^3$) and satisfies the inequality $C(u, u) < u$ for all $u \in (0, 1)$. This property, known as Ling's Theorem (see [64]), has been extended in an arbitrary dimension $d > 2$ by [89]. Then, [56] proposed to test the associativity of C to check the validity of the Archimedean zero assumption. For every couple (u_1, u_2) in $(0, 1)^2$, he defined the test statistics

$$\mathscr{T}_n^J(u_1, u_2) := \sqrt{n}\,\{C_n(u_1, C_n(u_2, u_2)) - C_n(C_n(u_1, u_2), u_2)\}\,.$$

Despite its simplicity, the latter pointwise approach is not consistent against a large class of alternatives. For instance, there exist copulas that are associative but not Archimedean. Therefore, [15] revisited this idea, by invoking fully the previous characterization of Archimedean copulas. To deal with associativity, they introduced the trivariate process

$$\mathscr{T}_n(u_1, u_2, u_3) := \sqrt{n}\,\{C_n(u_1, C_n(u_2, u_3)) - C_n(C_n(u_1, u_2), u_3)\}\,,$$

and proved its weak convergence in $\ell^\infty([0, 1]^3)$. CvM T_n^{CvM} and KS T_n^{KS} test statistics can be built on $\mathscr{T}_n$. To reject associative copulas that are not Archimedean, these statistics are slightly modified to get

$$\tilde{T}_n^{CvM} := T_n^{CvM} + n^\alpha \psi\left(\max\left\{\frac{i}{n}(1 - \frac{i}{n}) : C_n(\frac{i}{n}, \frac{i}{n}) = \frac{i}{n}\right\}\right),$$

for some chosen constant $\alpha \in (0, 1/2)$ and some increasing function ψ, $\psi(0) = 0$. Therefore, such final tests are consistent against all departures from Archimedeanity.

Unfortunately, the two previous procedures are limited to bivariate copulas, and their generalization to higher dimensions d seems to be problematic.

4.4.2 Extreme-Value Dependence

As we have seen previously, bivariate extreme-value copulas are written as

$$C(u, v) = \exp\left\{\ln(uv) A(\frac{\ln(v)}{\ln(uv)})\right\}, \tag{4.2}$$

for every u, v in $(0,1)$, where $A : [0,1] \to [1/2,1]$ is convex and satisfies $\max(t, 1-t) \leq A(t) \leq 1$ for every $t \in [0,1]$. Therefore, such copulas are fully parameterized by the so-called Pickands dependence function A, that is univariate. Extreme-value copulas are important in a lot of fields because they characterize the large-sample limits of copulas of componentwise maxima of strongly mixing stationary sequences ([26, 53], and the recent survey [50]). Then, it should be of interest to test whether the underlying copula can be represented by (4.2), for some unspecified dependence function A.

Studying the Kendall's process associated with an extreme-value copula C, [49] have noticed that, by setting $W := C(U_1, U_2)$, we have $K(t) = P(W \leq t) = t - (1-\tau)t\ln(t)$, for every $t \in (0,1)$, where τ is the underlying Kendall's tau. Moreover, they show that the moments of W are $E[W^i] = (i\tau + 1)/(i+1)^2$, for all $i \geq 1$. Therefore, under $\mathscr{H}_0$, $-1 + 8E[W] - 9E[W^2] = 0$. Then they proposed a test (that the underlying copula is extreme-value) based on an empirical counterpart of the latter relation: set

$$T_n := -1 + \frac{8}{n(n-1)} \sum_{i \neq j} I_{ij} - \frac{9}{n(n-1)(n-2)} \sum_{i \neq j \neq k} I_{ij} I_{kj},$$

where $I_{ij} := \mathbf{1}(X_{i,1} \leq X_{j,1}, X_{i,2} \leq X_{j,2})$, for all $i, j \in \{1, \ldots, n\}$. Under $\mathscr{H}_0$, the latter test statistic is asymptotically normal. Its asymptotic variance has been evaluated in [7]. Quessy [77] has provided extensions of this idea towards more higher order moments of W.

These approaches rely on the so-called reduction of dimension techniques (see Sect. 4.3.3). To improve the power of GOF tests, it would be necessary to work in functional spaces, i.e. concentrate on empirical counterparts of extreme-value copulas, or, equivalently, of the functions A themselves. For instance, [77] proposed a CvM GOF test, based on the Kendall's function K above. More generally, several estimates of the Pickands dependence function are available, but most of them rely on the estimation of marginal distributions: see Sect. 9.3 in [6] or [2]. Nonetheless, [47] have built "pure" copula GOF test statistics, i.e. independent from margins, by invoking empirical counterparts of the Pickands function introduced in [42]: given our previous notations,

1. Define the pseudo-observations

$$\tilde{U}_i := nF_{n,1}(X_{i,1})/(n+1), \ \tilde{V}_i := nF_{n,1}(X_{i,2})/(n+1).$$

2. Define the r.v. $\hat{S}_i := -\ln \tilde{U}_i$ and $\hat{T}_i := -\ln \tilde{V}_i$.
3. For every $i = 1, \ldots, n$, set $\hat{\xi}_1(0) := \hat{S}_i$, and $\hat{\xi}_1(1) := \hat{T}_i$. Moreover, for every $t \in (0,1)$, set

$$\hat{\xi}_i(t) := \min\left(\frac{\hat{S}_i}{1-t}, \frac{\hat{T}_i}{t}\right).$$

4. Two estimates of A are given by

$$A_n^P(t) := \left[n^{-1} \sum_{i=1}^{n} \hat{\xi}_i(t) \right]^{-1} \quad \text{and} \quad A_n^{CFG}(t) := \exp\left(-\gamma - n^{-1} \sum_{i}^{n} \ln \hat{\xi}_i(t) \right),$$

where γ denotes the Euler constant.

The two latter estimates are the "rank-based" version of those proposed in [16,75], respectively.

There is an explicit one-to-one mapping between A_n^P (resp. A_n^{CFG}) and the empirical copula C_n. Therefore, after endpoint corrections, [42] have exhibited the weak limit of the corresponding processes $\mathbb{A}_n^P := \sqrt{n}(A_n^P - A)$ and $\mathbb{A}_n^{CFG} := \sqrt{n}(A_n^{CFG} - A)$. Working with the two latter processes instead of $\mathbb{C}_n$, a lot of GOF tests can be built. For instance , [47] have detailed an AD type test based on the L^2 norm of $\mathbb{A}_n^P$ and $\mathbb{A}_n^{CFG}$, even under composite null assumptions.

In the same vein, another strategy has been proposed in [62]: there is an equivalence between extreme-value copula C and max-stable copulas, i.e. copulas for which $C(\mathbf{u})^r = C(\mathbf{u}^r)$, for every $\mathbf{u} \in [0,1]^d$ and $r \in \mathbb{R}^+$. By setting $\mathbb{D}_{n,r}(\mathbf{u}) := \sqrt{n}(\{C_n(\mathbf{u}^{1/r})\}^r - C_n(\mathbf{u}))$, for all $\mathbf{u} \in [0,1]^d$ and every $r > 0$, [62] have built some tests based on the limiting law of the joint process $(\mathbb{D}_{n,r_1}, \ldots, \mathbb{D}_{n,r_p})$ for an arbitrary integer p.

4.4.3 Pair-Copula Constructions

In the recent years, a lot of effort has been devoted to the construction of d-dimensional copulas, $d > 2$, as combinations of several two-dimensional copulas. Some authors have enriched the Archimedean copula class: Hierarchical, nested or multiplicative Archimedean copulas. Among others, see [57,67,69,84,94]. Other authors have studied the large class of vines: D-vines, C-vines, regular vines more generally (see e.g., [1,20]). Inference, simulation, and specification techniques have made significant progress to deal with these families of models $\mathscr{F}$. These advances provide large classes of very flexible copulas.

We will not discuss in depth the way of choosing the best Hierarchical Archimedean copula or the best D-vine, for a given data. Apparently, every proposition in this stream of the literature follows the same steps:

1. Assume an underlying class of models $\mathscr{F}$ (D-vine, for instance);
2. Choose the potential bivariate families of copulas that may appear in the construction;
3. Evaluate the best structure (a network, or a tree) and estimate the associated bivariate copulas (simultaneously, in general).

Mathematically, we can nest this methodology inside the previous general GOF copula framework detailed above. Indeed, the copula candidates belong to a finite dimensional parametric family, even if the dimension of the unknown parameter θ can be very large. Obviously, authors have developed ad-hoc procedures to avoid such a violent approach of GOF testing: see [21] or [29] for vine selection, for instance.

At the opposite, there is no test of the slightly different and more difficult GOF problem

$$\mathscr{H}_0 : C \text{ belongs to a given class } \mathscr{F}.$$

For instance, a natural question would be to test whether an underlying copula belongs to the large (and infinite dimensional!) class of Hierarchical Archimedean copulas. To the best of our knowledge, this way of testing is still a fully open problem.

4.5 GOF Copula Tests for Multivariate Time Series

One limiting feature of copulas is the difficulty to use them in the presence of multivariate-dependent vectors $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, with $\mathbf{X}_n \in \mathbb{R}^d$. In general, the "modeler problem" is to specify the full law of this process, i.e., the joint laws $(\mathbf{X}_{n_1}, \ldots, \mathbf{X}_{n_p})$ for every p and every indices $n_1, \ldots, n_p$ and in a consistent way. Applying the copula ideas to such a problem seems to be rather natural (see [74] for a survey). Nonetheless, even if we restrict ourselves to stationary processes, the latter task is far from easy.

The first idea is to describe the law of the vectors $(\mathbf{X}_m, \mathbf{X}_{m+1}, \ldots, \mathbf{X}_n)$ with copulas directly, for every couple (m, n), $m < n$. This can be done by modeling separately (but consistently) $d(n-m+1)$ unconditional margins plus a $d(n-m+1)$-dimensional copula. This approach seems particularly useful when the underlying process is stationary and Markov (see [17] for the general procedure). But the conditions of Markov coherence are complex (see [55]), and there is no general GOF strategy in this framework, to the best of our knowledge.

A more usual procedure in econometrics is to specify a multivariate time-series model, typically a linear regression, and to estimate residuals, assumed serially independent: see [18] that deals with a GARCH-like model with diagonal innovation matrix. They showed that estimating the copula parameters using rank-based pseudo-likelihood methods with the ranks of the residuals instead of the (non-observable) ranks of innovations leads to the same asymptotic distribution. In particular, the limiting law of the estimated copula parameters does not depend on the unknown parameters used to estimate the conditional means and the conditional variances. This is very useful to develop GOF tests for the copula family of the innovations. Rémillard [79] extended these results: under similar technical assumptions, the empirical copula process has the same limiting distribution as

if one would have started with the innovations instead of the residuals. As a consequence, a lot of tools developed for the serially independent case remain valid for the residuals. However, that is not true if the stochastic volatility is genuinely non-diagonal.

A third approach would be to use information on the marginal processes themselves. This requires to specify conditional marginal distributions, instead of unconditional margins as above in the first idea. This would induce a richer application of the two-step basic copula idea, i.e., use "standard" univariate processes as inputs of more complicated multivariate models:

1. For every $j = 1, \ldots, d$, specify the law of $X_{n,j}$ knowing the past values $X_{n-1,j}$, $X_{n-2,j}, \ldots$;
2. Specify (and/or estimate) relevant dependence structures, "knowing" these univariate underlying processes, to recover the entire process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$.

Using similar motivations, Patton [72, 73] introduced the so-called conditional copulas, which are associated with conditional laws in a particular way. Specifically, let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector from $(\Omega, \mathscr{A}_0, \mathbb{P})$ to $\mathbb{R}^d$. Consider some arbitrary sub-σ-algebra $\mathscr{A} \subset \mathscr{A}_0$. A conditional copula associated with $(\mathbf{X}, \mathscr{A})$ is a $\mathscr{B}([0,1]^d) \otimes \mathscr{A}$ measurable function C such that, for any $x_1, \ldots, x_d \in \mathbb{R}$,

$$\mathbb{P}\left(\mathbf{X} \leq \mathbf{x} | \mathscr{A}\right) = C\left\{\mathbb{P}(X_1 \leq x_1 | \mathscr{A}), \ldots, \mathbb{P}(X_d \leq x_d | \mathscr{A}) | \mathscr{A}\right\}.$$

The random function $C(\cdot|\mathscr{A})$ is uniquely defined on the product of the values taken by $x_j \mapsto \mathbb{P}(X_j \leq x_j \mid \mathscr{A})(\omega)$, $j = 1, \ldots, d$, for every realization $\omega \in \mathscr{A}$. As in the proof of Sklar's theorem, $C(\cdot|A)$ can be extended on $[0,1]^d$ as a copula, for every conditioning subset of events $A \subset \mathscr{A}$.

In Patton's approach, it is necessary to know/model each margin, knowing all the past information, and not only the past observations of each particular margin. Nonetheless, practitioners often have good estimates of the conditional distribution of each margin, conditionally given its own past, i.e., $\mathbb{P}(X_{n,j} \leq x_j | \mathscr{A}_{n,j})$, $j = 1, \ldots, d$, by setting $\mathscr{A}_{n,j} = \sigma(X_{n-1,j}, X_{n-2,j}, \ldots)$. To link these quantities with the (joint) law of $\mathbf{X}_n$ knowing its own past, it is tempting to write

$$\mathbb{P}\left(\mathbf{X}_n \leq \mathbf{x} | \mathscr{A}_n\right) = C^*\left\{\mathbb{P}(X_{1,n} \leq x_1 | \mathscr{A}_{n,1}), \ldots, \mathbb{P}(X_{d,n} \leq x_d | \mathscr{A}_{n,d})\right\},$$

for some random function $C^* : [0,1]^d \longrightarrow [0,1]$ whose measurability would depend on $\mathscr{A}_n$ and on the $\mathscr{A}_{n,j}$, $j = 1, \ldots, d$. Actually, the latter function is a copula only if the process $(X_{k,n}, k \neq j)_{n\in\mathbb{Z}}$ does not "Granger-cause" the process $(X_{j,n})_{n\in\mathbb{Z}}$, for every $j = 1, \ldots, d$. This assumption that each variable depends on its own lags, but not on the lags of any other variable, is clearly strong, even though it can be accepted empirically; see the discussion in [74], pp. 772–773. Thus, [35] has extended Patton's conditional copula concept, by defining the so-called pseudo-copulas that are simply cdf on $[0,1]^d$ with arbitrary margins. They prove:

Theorem 4.1. *For any sub-algebras $\mathscr{B}, \mathscr{A}_1, \ldots, \mathscr{A}_d$ such that $\mathscr{A}_j \subset \mathscr{B}$, $j = 1, \ldots, d$, there exists a random function $C : [0,1]^d \times \Omega \longrightarrow [0,1]$ such that*

$$\begin{aligned}\mathbb{P}(\mathbf{X} \leq \mathbf{x} \mid \mathscr{B})(\omega) &= C\left\{\mathbb{P}(X_1 \leq x_1 \mid \mathscr{A}_1)(\omega), \ldots, \mathbb{P}(X_d \leq x_d \mid \mathscr{A}_d)(\omega), \omega\right\} \\ &\equiv C\left\{\mathbb{P}(X_1 \leq x_1 \mid \mathscr{A}_1), \ldots, \mathbb{P}(X_d \leq x_d \mid \mathscr{A}_d)\right\}(\omega),\end{aligned}$$

for every $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and almost every $\omega \in \Omega$. This function C is $\mathscr{B}([0,1]^d) \otimes \mathscr{B}$ measurable. For almost every $\omega \in \Omega$, $C(\cdot, \omega)$ is a pseudo-copula and is uniquely defined on the product of the values taken by $x_j \mapsto \mathbb{P}(X_j \leq x_j \mid \mathscr{A}_j)(\omega)$, $j = 1, \ldots, d$.

If C is unique, it is called the conditional $(\mathscr{A}, \mathscr{B})$-pseudo-copula associated with $\mathbf{X}$ and denoted by $C(\cdot|\mathscr{A}, \mathscr{B})$. Actually, $C(\cdot|\,\mathscr{A}, \mathscr{B})$ is a copula iff

$$\mathbb{P}(X_j \leq x_j \mid \mathscr{B}) = \mathbb{P}(X_j \leq x_j \mid \mathscr{A}_j) \quad \text{a.e.} \tag{4.3}$$

for all $j = 1, \ldots, d$ and $\mathbf{x} \in \mathbb{R}^d$. This means that $\mathscr{B}$ cannot provide more information about X_j than $\mathscr{A}_j$, for every j. Patton's conditional copula corresponds to the particular case $\mathscr{B} = \mathscr{A}_1 = \cdots = \mathscr{A}_d$, for which (4.3) is clearly satisfied.

One key issue is to state if pseudo-copulas depend really on the past values of the underlying process, i.e., to test their constancy, an assumption often made in practice. In [35], they estimate nonparametrically conditional pseudo-copulas, including Patton's conditional copulas as a special case and test their constancy with respect to their conditioning subsets. Here, we specify their technique.

For a stationary and strongly mixing process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, we restrict ourselves to conditional sub-algebras $\mathscr{A}_n$ and $\mathscr{B}_n$ that are defined by a finite number of past values of the process, typically $(\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \ldots, \mathbf{X}_{n-p})$ for some $p \geq 1$. The dependence of $\mathscr{A}$ and $\mathscr{B}$ with respect to past values $\mathbf{y}$ will be implicit hereafter. Formally, [35] consider the test of several null hypothesis:

(a)

$$\mathscr{H}_0^{(1)} : \text{ For every } \mathbf{y},\ C(\cdot \mid \mathscr{A}, \mathscr{B}) = C_0(\cdot),$$

against

$$\mathscr{H}_a : \text{ For some } \mathbf{y},\ C(\cdot \mid \mathscr{A}, \mathscr{B}) \neq C_0(\cdot),$$

where C_0 denotes a fixed pseudo-copula function. In this case, $\mathscr{H}_0^{(1)}$ means that the underlying conditional $(\mathscr{A}, \mathscr{B})$-pseudo-copula is in fact a true copula, independent of the past values of the process.

(b)

$$\mathscr{H}_0^{(2)} : \text{ There exists a parameter } \theta_0 \text{ such that } C(\cdot|\mathscr{A}, \mathscr{B}) = C_{\theta_0} \in \mathscr{C},\ \text{for every } \mathbf{y},$$

where $\mathscr{C} = \{C_\theta, \theta \in \Theta\}$ denotes some parametric family of pseudo-copulas.

(c)

$$\mathscr{H}_0^{(3)}: \text{ For some function } \theta(\mathbf{y}) = \theta(\mathscr{A}, \mathscr{B}), \text{ we have}$$
$$C(\cdot|\mathscr{A}, \mathscr{B}) = C_{\theta(\mathbf{y})} \in \mathscr{C}, \text{ for every } \mathbf{y}.$$

The latter assumption says that the conditional pseudo-copulas stay inside the same pre-specified parametric family of pseudo-copulas (possibly copulas), for different observed values in the past. Fermanian and Wegkamp [35] proposed a fully nonparametric estimator of the conditional pseudo-copulas, and derived its limiting distribution. This provides a framework for "brute-force" GOF tests of multivariate dynamic dependence structures (conditional copulas, or even pseudo-copulas), similar to what has been done in Sect. 4.2.

Fermanian and Wegkamp [35] stated the equivalent of the empirical processes $\mathbb{C}_n$ or $\hat{\mathbb{C}}_n$. Use the short-hand notation $\mathbf{X}_m^n$ for the vector $(\mathbf{X}_m, \mathbf{X}_{m+1}, \ldots, \mathbf{X}_n)$. Similarly, write $\mathbf{X}_{m,j}^n = (X_{m,j}, \ldots, X_{n,j})$. Assume that every conditioning set $\mathscr{A}_{n,j}$ (resp. $\mathscr{B}_n$) is related to the vector $\mathbf{X}_{n-p,j}^{n-1}$ (resp. $\mathbf{X}_{n-p}^{n-1}$). Specifically, consider the events $(\mathbf{X}_{n-p}^{n-1} = \mathbf{y}^*) \in \mathscr{B}_n$, with $\mathbf{y}^* = (\mathbf{y}_1, \ldots, \mathbf{y}_p)$, and $(\mathbf{X}_{n-p,j}^{n-1} = \mathbf{y}_j^*) \in \mathscr{A}_{n,j}$, with $\mathbf{y}_j^* = (y_{1j}, \ldots, y_{pj})$. Their nonparametric estimator of the pseudo-copula is based on a standard plug-in technique that requires estimates of the joint conditional distribution

$$m(\mathbf{x} \mid \mathbf{y}^*) = \mathbb{P}\left(\mathbf{X}_p \le \mathbf{x} \mid \mathbf{X}_0^{p-1} = \mathbf{y}^*\right),$$

and of conditional marginal cdfs

$$m_j(x_j \mid \mathbf{y}_j^*) = \mathbb{P}\left(X_{pj} \le x_j \mid \mathbf{X}_{0,j}^{p-1} = \mathbf{y}_j^*\right), \quad j = 1, \ldots, d.$$

Let F_{nj} be the (marginal) empirical distribution function of X_j, based on the $(X_{1,j}, \ldots, X_{n,j})$. For convenient kernels K and $\bar{K}$, set

$$K_h(\mathbf{x}) = h^{-pd} K\left(\frac{x_1}{h}, \cdots, \frac{x_{pd}}{h}\right), \text{ and } \quad \bar{K}_{\bar{h}}(\mathbf{x}) = \bar{h}^{-p} \bar{K}\left(\frac{x_1}{\bar{h}}, \cdots, \frac{x_p}{\bar{h}}\right).$$

For every $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y}^* \in \mathbb{R}^{pd}$, estimate the conditional distribution $m(\mathbf{x} \mid \mathbf{y}^*) = \mathbb{P}\left(\mathbf{X}_p \le \mathbf{x} \mid \mathbf{X}_0^{p-1} = \mathbf{y}^*\right)$ by

$$m_n(\mathbf{x} \mid \mathbf{y}^*) = \frac{1}{n-p} \sum_{\ell=0}^{n-p} K_n(\mathbf{X}_\ell^{\ell+p-1}) \mathbf{1}(\mathbf{X}_{\ell+p} \le \mathbf{x}),$$

where

$$K_n(\mathbf{X}_\ell^{\ell+p-1}) = K_h\{F_{n1}(X_{\ell 1}) - F_{n1}(y_{11}), \ldots, F_{nd}(X_{\ell d}) - F_{nd}(y_{1d}), \ldots,$$
$$\ldots, F_{n1}(X_{(\ell+p-1),1}) - F_{n1}(y_{p1}), \ldots, F_{nd}(X_{(\ell+p-1),d}) - F_{nd}(y_{pd})\}.$$

Similarly, for all $x_j \in \mathbb{R}$ and $\mathbf{y}_j^* \in \mathbb{R}^p$, the conditional marginal cdf's $m_j(x_j \mid \mathbf{y}_j^*)$ is estimated in a nonparametric way by

$$m_{n,j}(x_j \mid \mathbf{y}_j^*) = \frac{1}{n-p} \sum_{\ell=1}^{n-p} \bar{K}_{\bar{h}}\{F_{nj}(X_{\ell,j}) - F_{nj}(y_{1j}), \ldots,$$
$$F_{nj}(X_{\ell+p-1,j}) - F_{nj}(y_{pj})\}\mathbf{1}(\mathbf{X}_{\ell+p,j} \le x_j),$$

for every $j = 1, \ldots, d$. Fermanian and Wegkamp [35] proposed to estimate the underlying conditional pseudo-copula by

$$\hat{C}(\mathbf{u} \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}^*) = m_n\{m_{n,1}^{(-1)}(u_1 \mid \mathbf{y}_1^*), \ldots, m_{n,d}^{(-1)}(u_d \mid \mathbf{y}_d^*) \mid \mathbf{y}^*\},$$

with the use of pseudo-inverse functions. Then, under $\mathscr{H}_0^{(1)}$, for all $\mathbf{u} \in [0,1]^d$ and $\mathbf{y}^* = (\mathbf{y}_1, \ldots, \mathbf{y}_p) \in \mathbb{R}^{dp}$,

$$\sqrt{nh_n^{pd}}\{\hat{C}(\mathbf{u} \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}^*) - C_0(\mathbf{u})\} \xrightarrow{\mathrm{d}} \mathscr{N}[0, \sigma(\mathbf{u})]$$

as $n \to \infty$, where $\sigma(\mathbf{u}) = C_0(\mathbf{u})\{1-C_0(\mathbf{u})\}\int K^2(\mathbf{v})\,\mathrm{d}\mathbf{v}$. This result can be extended to deal with different vectors $\mathbf{y}^*$ simultaneously and with the null hypotheses $\mathscr{H}_0^{(2)}$ and $\mathscr{H}_0^{(3)}$: for all $\mathbf{u} \in \mathbb{R}^d$,

$$\sqrt{nh_n^{pd}}\{\hat{C}(\mathbf{u} \mid \mathbf{y}_1^*) - C_{\hat{\theta}_1}(\mathbf{u}), \ldots, \hat{C}(\mathbf{u} \mid \mathbf{y}_q^*) - C_{\hat{\theta}_q}(\mathbf{u})\} \xrightarrow{\mathrm{d}} \mathscr{N}[0, \Sigma(\mathbf{u}, \mathbf{y}_1^*, \ldots, \mathbf{y}_q^*)],$$

as $n \to \infty$, where

$$\Sigma(\mathbf{u}, \mathbf{y}_1^*, \ldots, \mathbf{y}_q^*) = \mathrm{diag}\left(C_{\theta(\mathbf{y}_k^*)}(\mathbf{u})\{1 - C_{\theta(\mathbf{y}_k^*)}(\mathbf{u})\}\int K^2(\mathbf{v})\,\mathrm{d}\mathbf{v},\ 1 \le k \le q\right),$$

for some consistent estimators $\hat{\theta}_k$ such that $\hat{\theta}_k = \theta(\mathbf{y}_k^*) + O_P(n^{-1/2})$, $k = 1, \ldots, q$.

Each kth term on the diagonal of Σ can be consistently estimated by

$$\hat{\sigma}_k^2(\mathbf{u}) = C_{\hat{\theta}_k}(\mathbf{u})\{1 - C_{\hat{\theta}_k}(\mathbf{u})\}\int K^2(\mathbf{v})\,\mathrm{d}\mathbf{v}.$$

Note that, in the corollary above, the limiting correlation matrix is diagonal because we are considering different conditioning values $\mathbf{y}_1^*, \ldots, \mathbf{y}_q^*$ but the same argument $\mathbf{u}$. At the opposite, an identical conditioning event but different arguments $\mathbf{u}_1, \mathbf{u}_2, \ldots$

would lead to a complex (nondiagonal) correlation matrix, as explained in [33]. The latter weak convergence result of random vectors allows the building of GOF tests as in Sect. 4.2. For instance, as in [33], a simple test procedure may be

$$T(\mathbf{u}, \mathbf{y}_1^*, \ldots, \mathbf{y}_q^*) = (nh_n^{pd}) \sum_{k=1}^{q} \frac{\{\hat{C}(\mathbf{u} \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}_k^*) - C_{\hat{\theta}_k}(\mathbf{u})\}^2}{\hat{\sigma}^2_{\mathbf{y}_k^*}(\mathbf{u})},$$

for different choices of $\mathbf{u}$ and conditioning values $\mathbf{y}_k^*$. Under $\mathscr{H}_0^{(1)}$, the term on the right-hand side tends to a $\chi^2(q)$ distribution under the null hypothesis. Note that this test is "local" since it depends strongly on the choice of a single $\mathbf{u}$. An interesting extension would be to build a "global" test, based on the behavior of the full process

$$\sqrt{nh_n^{pd}}\,\{\hat{C}(\cdot \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}_k^*) - C_{\hat{\theta}_k}(\cdot)\}.$$

But the task of getting pivotal limiting laws is far from easy, as illustrated in [33].

In practice, authors often restrict themselves to the case of time-dependent copula parameters instead of managing time-dependent multivariate cdfs nonparametrically. For instance, every conditional copula or pseudo-copula is assumed to belong to the Clayton family, and their random parameters θ depend on the past observations. Abegaz et al. [3] has proposed a non-parametric estimate $\hat{\theta}(\cdot)$ of the function θ, in the case of a univariate conditioning variable. It seems possible to build some GOF tests based on this estimate and its limiting behavior, at least for simple null hypothesis, but the theory requires more developments.

4.6 Practical Performances of GOF Copula Tests

Once a paper introduces one or several new copula GOF tests, it is rather usual to include an illustrative section. Typically, two characteristics are of interest for some tests in competition: their ability to maintain the theoretical levels powers and their power performances under several alternatives. Nonetheless, these empirical elements, even useful, are often partial and insufficient to found a clear judgment. Actually, only a few papers have studied and compared the performances of the main previous tests in depth. Indeed, the calculation power required for such a large analysis is significant. That is why a lot of simulation studies restrict themselves to bivariate copulas and small or moderate sample sizes (from $n = 50$ to $n = 500$, typically). The most extensive studies of finite sample performances are probably those of [8, 46]. In both papers, the set of tests under scrutiny contains the three main approaches:

1. "Brute-force" proposals like T_n^{KS} and/or T_n^{CvM}, as in Sect. 4.2;
2. Kendall's process-based tests;
3. Test statistics invoking the PIT (see Sect. 4.3).

These works found that a lot of tests perform rather well, even for small samples (from $n = 50$, e.g.). Moreover, it is difficult to exhibit clear hierarchy among all of these tests in terms of power performances. As pointed out by [46],

> No single test is preferable to all others, irrespective of the circumstances.

In their experiments, [46] restricted themselves to bivariate copulas and small sample sizes $n \in \{50, 150\}$. The statistics based on Kendall's dependence function are promoted, particularly when the underlying copula is assumed to be Archimedean. It appeared that CvM style test statistics are preferable to KS ones, all other things being equal, and whatever the possible transformations of the data and/or the reductions of information. Among the tests based on a CvM statistic, it is difficult to discriminate between the three main approaches.

The latter fact is confirmed in [8] that led some simulated experiments with higher dimensions $d \in \{2, 4, 8\}$ and larger sample sizes $n \in \{100, 500\}$. Berg [8] observed the particularly good performances of a new test statistic, calculated as the average of the three approaches. Moreover, he studied the impact of the variables ordering in the PIT. Even if estimated p-values may be different, depending on which permutation order is chosen, this does not seem to create worrying discrepancies.

Notably [11] led an extensive simulated experiment of the same type, but their main focus was related to detecting small departures from the null hypothesis. Thus, they studied the asymptotic behavior of some GOF test statistics under sequences of alternatives of the type

$$\mathscr{H}_{a,n} : C = (1 - \delta_n)C_0 + \delta_n D,$$

where $\delta_n = n^{-1/2}\delta$, $\delta > 0$, and D is another copula. They computed local power curves and compared them for different test statistics. They showed that the estimation strategy can have a significant impact on the power of CvM statistics and that some "moment-based" statistics provide very powerful tests under many distributional scenarios.

Despite the number of available tests in the literature, the usefulness of all these procedures in practice has to be proved more convincingly. Apparently, some authors have raised doubts about the latter point. For instance, [91] has evaluated the performances of value-at-risk or VaR (quantiles of loss) and expected shortfall or ES (average losses above a VaR level) forecasts, for a large set of portfolios of two financial assets and different copula models. They estimate static copula models on couples of asset return residuals, once GARCH(1,1) dynamics have been fitted for every asset independently. They applied three families of GOF tests (empirical copula process, PIT, Kendall's function) and five copula models. They found that

> Although copula models with GARCH margins yield considerably better estimates than correlation-based models, the identification of the optimal parametric copula form is a serious unsolved problem.

Indeed, none of the GOF tests is able to select the copula family that yields the best VaR- or ES-forecasts. This points out the difficulty of finding relevant and stable multivariate dynamics models, especially related to joint extreme moves. But, such results highlight the fact that it remains a significant gap between good performances with simulated experiments and trustworthy multivariate models, even validated formally by statistical tests.

Indeed, contrary to studies based on simulated samples drawn from an assumed copula family (the standard case, as in [46] or [8]), real data can suffer from outliers or measurement errors. This is magnified by the fact that most realistic copulas are actually time dependent [91] and/or are mixtures or copulas [63]. Therefore, [92] showed that even minor contamination of a dataset can lead to significant power decreases of copula GOF tests. He applied several outlier detection methods from the theory of robust statistics, as in [66], before leading the formal GOF test of any parametric copula family. Weiss [92] concluded that the exclusion of outliers can have a beneficial effect on the power of copula GOF tests.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula construction of multiple dependence. Insur. Mathe. Econ. **44**, 182–198 (2009)
2. Abdous, B., Ghoudi, K.: Non-parametric estimators of multivariate extreme dependence functions. J. Nonparametr. Stat. **17**, 915–935 (2005)
3. Abegaz, F., Gijbels, I., Veraverbeke, N.: Semiparametric estimation of conditional copulas. J. Multivariate Anal. **110**, 43–73 (2012)
4. Andersen, P.K., Ekstrøm, C.T., Klein, J.P., Shu, Y., Zhang, M.-J.: A class of goodness of fit tests for a copula based on bivariate right-censored data. Biometrical J. **47**, 815–824 (2005)
5. Barbe, P., Genest, C., Ghoudi, K., Rémillard, B.: On Kendall's process. J. Multivar. Anal. **58**, 197–229 (1996)
6. Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of Extremes: Theory and Applications. Wiley, Chichester (2004)
7. Ben Ghorbal, N., Genest, C., Nešlehová, J.: On the Ghoudi, Koudraji and Rivest test for extreme-value dependence. Can. J. Stat. **37**, 534–552 (2009)
8. Berg, D.: Copula goodness-of-fit testing: an overview and power comparison. Eur. J. Financ. **15**, 675–705 (2009)
9. Berg, D., Bakken, H.: Copula goodness-of-fit tests: A comparative study. Working Paper, Norwegian Computing Centre (2006)
10. Berg, D., Bakken, H.: A copula goodness-of-fit approach based on the conditional probability integral transformation. Norwegian Computing Centre, SAMBA/41/05 (2007)
11. Berg, D., Quessy, J.-F.: Local sensitivity analysis of goodness-of-fit tests for copulas. Scandin. J. Stat. **36**, 389–412 (2009)
12. Breymann, W., Dias, A., Embrechts, P.: Dependence structures for multivariate high-frequency data in finance. Quant. Financ. **1**, 1–14 (2003)
13. Bücher, A., Dette, H.: Some comments on goodness-of-fit tests for the parametric form of the copula based on L^2-distances. J. Multivar. Anal. **101**, 749–763 (2010)
14. Bücher, A., Dette, H.: A note on bootstrap approximations for the empirical copula process. Stat. Prob. Lett. **80**, 1925–1932 (2010)
15. Bücher, A., Dette, H., Volgushev, S.: A test for Archimedeanity in bivariate copula models. J. Multivar. Anal. **110**, 121–132 (2012)

16. Capéraà, P., Fougères, A.-L., Genest, C.: A nonparametric estimation procedure for bivariate extreme value copulas. Biometrika **84**, 567–577 (1997)
17. Chen, X., Fan, Y.: Estimation of copula-based semiparametric time series models. J. Econom. **130**, 307–335 (2006)
18. Chen, X., Fan, Y.: Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. J. Econom. **135**, 125–154 (2006)
19. Chen, X., Fan, Y., Patton, A.: Simple tests for models of dependence between multiple financial time series, with applications to U.S. equity returns and exchange rates. Discussion Paper 483, London School of Economics (2004)
20. Czado, C.: Pair-copula constructions of multivariate copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.): Copula Theory and Its Applications. Proceedings of the Workshop Held in Warsaw, 25–26 September 2009. Lecture Notes in Statistics – Proceedings, vol. 198. Springer, Dordrecht (2010)
21. Czado, C., Schepsmeier, U., Min, A.: Maximum likelihood estimation of mixed C-vines with applications to exchange rates. Statistical Modelling: An Int. J. **12**, 229–255 (2012)
22. D'Agostino, R.B., Stephens, M.A.: Goodness-of-Fit Techniques. Dekker, New York (1986)
23. Deheuvels, P.: La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. Acad. R. Belg. Bull. Cl. Sci. 5th Série **65**, 274–292 (1979)
24. Deheuvels, P.: A Kolmogorov-Smirnov type test for independence and multivariate samples. Rev. Roum. Math. Pures et Appl. XXVI **2**, 213–226 (1981)
25. Deheuvels, P.: A nonparametric test of independence. Publ. de l'ISUP **26**, 29–50 (1981)
26. Deheuvels, P.: Probabilistic aspects of multivariate extremes. In: Tiago de Oliveira, J. (ed.) Statistical Extremes and Applications, pp. 117–130. Reidel, Dordrecht (1984)
27. Dias, A., Embrechts, P.: Dynamic copula models for multivariate high–frequency data in finance. Working Paper, New University of Lisbon, Portugal (2004)
28. Diks, C., Panchenko, V., van Dijk, D.: Out-of-sample comparison of copula specifications in multivariate density forecasts. J. Econ. Dynam. Contr. **34**, 1596–1609 (2010)
29. Dissmann, J., Brechmann, E. C., Czado, C., Kurowicka, D.: Selecting and estimating regular vine copulae and application to financial returns. Working Paper (2012)
30. Dobric, J., Schmid, F.: Testing goodness-of-fit for parametric families of copulas. Applications to financial data. Comm. Stat. Simul. Comput. **34**, 1053–1068 (2005)
31. Dobric, J., Schmid, F.: A goodness of fit test for copulas based on Rosenblatt's transformation. Comput. Stat. Data Anal. **51**, 4633–4642 (2007)
32. Fan, Y., Linton, O.: Some higher order theory for a consistent nonparametric model specification test. J. Stat. Plan. Infer **109**, 125–154 (2003)
33. Fermanian, J-D.: Goodness-of-fit tests for copulas. J. Multivar. Anal. **95**, 119–152 (2005)
34. Fermanian, J.-D., Scaillet, O.: Some statistical pitfalls in copula modelling for financial applications. In: Klein, E. (ed.) Capital Formation, Governance and Banking, pp. 59–74. Nova Science Publishers, Inc., New York (2005)
35. Fermanian, J.-D., Wegkamp, M.: Time-dependent copulas. J. Multivar. Anal. **110**, 19–29 (2012)
36. Fermanian, J.-D., Radulovic, D., Wegkamp, M.: Weak convergence of empirical copula processes. Bernoulli **10**, 847–860 (2004)
37. Fermanian, J.-D., Radulovic, D., Wegkamp, M.: An asymptotc total variation test for copulas. Working Paper (2012)
38. Gaënsler, P., Stute, W.: Seminar on Empirical Processes. Birkhäuser, Boston (1987)
39. Gayraud, G.,Tribouley, K.: A test of goodness-of-fit for copulas. Test **20**, 549–573 (2011)
40. Genest, C., Rémillard, B.: Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Ann. Inst. Henri Poincaré Prob. et Stat. **44**, 1096–1127 (2008)
41. Genest C., Rivest, L.: Statistical inference procedures for bivariate Archimedean copulas. J. Am. Stat. Ass. **88**, 1034–1043 (1993)
42. Genest, C., Segers, J.: Rank-based inference for bivariate extreme-value copulas. Ann. Stat. **37**, 2990–3022 (2009)

43. Genest, C., Quessy J.-F., Rémillard, B.: Goodness-of-fit procedures for copula models based on the probability integral transform. Scandin. J. Stat. **33**, 337–366 (2006)
44. Genest, C., Quessy J.-F., Rémillard, B.: On the joint asymptotic behavior of two rank-based estimators of the association parameter in the gamma frailty model. Stat. Probab. Lett. **76**, 10–18 (2006)
45. Genest, C., Masiello, E., Tribouley, K.: Estimating copula densities through wavelets. Insur. Math. Econ. **44**, 170–181 (2009)
46. Genest, C., Rémillard, B., Beaudoin, D.: Goodness-of-fit tests for copulas: a review and a power study. Insur. Math. Econ. **44**, 199–213 (2009)
47. Genest, C., Kojadinovic, I., Nešlehová, J., Yan, J.: Bernoulli **17**, 253–275 (2011)
48. Ghoudi, K., Rémillard, B.: Empirical processes based on pseudo-observations. In: Szyskowitz, B. (ed.) Asymptotic Methods in Probability and Statistics: A Volume in Honor of Miklos Csörgö, pp. 171–197. Elsevier, Amsterdam (1998)
49. Ghoudi, K., Khoudraji, A., Rivest, L.-P.: Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. Canad. J. Stat. **26**, 187–197 (1998)
50. Gudendorf, G., Segers, J.: Extreme-value copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Proceedings of the Workshop Held in Warsaw, 25–26 September 2009. Lecture Notes in Statistics – Proceedings, vol. 198. Springer, Dordrecht (2010)
51. Härdle, W., Mammen, E.: Testing parametric versus nonparametric regression. Ann. Stat. **21**, 1926–1947 (1993)
52. Hering, C., Hofert, M.: Goodness-of-fit tests for Archimedean copulas in large dimensions. Working Paper (2010)
53. Hsing, T.: Extreme value theory for multivariate stationary sequences. J. Multivar. Anal. **29**, 274–291 (1989)
54. Huard, D., Evin, G., Favre, A.-C.: Bayesian copula selection. Comput. Stat. Data Anal. **51**, 809–822 (2006)
55. Ibragimov, R.: Copula-based characterizations for higher-order Markov processes. Econom. Theory **25**, 819–846 (2009)
56. Jaworski, P.: Testing archimedeanity. In: Borglet, C., Gonzalez-Rodrigez, G., Trutschnig, W., Lubiano, M., Gil, M., Grzegorzewski, P., Hryniewicz, O. (eds.) Combining Soft Comuting and Statistical Methods in Data Analysis. Advances in Intelligent and Soft Computing, vol. 77, pp. 353–360. Springer, Berlin (2010)
57. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
58. Junker, M., May, A.: Measurement of aggregate risk with copulas. Econom. J. **8**, 428–454 (2005)
59. Klugman, S., Parsa, R.: Fitting bivariate loss distributions with copulas. Insur. Math. Econ. **24**, 139–148 (1999)
60. Kojadinovic, I., Yan, J.: Fast large-sample goodness-of-fit for copulas. Stat. Sinica **21**, 841–871 (2011)
61. Kojadinovic, I., Yan, J.: A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. Stat. Comput. **21**, 17–30 (2011)
62. Kojadinovic, I., Segers, J., Yan, J.: Large-sample tests of extreme-value dependence for multivariate copulas. Canad. J. Stat. **39**, 703–720 (2011)
63. Kole, E., Koedijk, K.C.G., Verbeek, M.: Selecting copulas for risk management. J. Bank. Financ. **31**, 2405–2423 (2007)
64. Ling, C.M.: Representation of associative functions. Publ. Math. Debrecen **12**, 189–212 (1965)
65. Malevergne, Y., Sornette, D.: Testing the Gaussian copula hypothesis for financial assets dependence. Quant. Financ. **3**, 231–250 (2003)
66. Mendes, B., de Melo, E., Nelsen, R.: Robust fits for copula models Comm. Stat. Simul. Comput. **36**, 997–1017 (2007)
67. Morillas, P.: A method to obtain new copulas from a given one. Metrika **61**, 169–184 (2005)
68. Nelsen, R.B.: An introduction to copulas. Lecture Notes in Statistics, vol. 139. Springer, New York (1999)

69. Okhrin, O., Okhrin, Y., Schmid, W.: On the structure and estimation of hierarchical Archimedean copulas. Discussion Paper 2009–014, Humbolt-Univerzität zu Berlin, Germany (2009)
70. Omelka, M. Gijbels, I., Veraverbeke, N.: Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. Ann. stat. **37**, 3023–3058 (2009)
71. Panchenko, V.: Goodness-of-fit test for copulas. Physica A **335**, 176–182 (2005)
72. Patton, A.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**, 527–556 (2006)
73. Patton A.: Estimation of multivariate models for time series of possibly different lengths. J. Appl. Econom. **21**, 147–173 (2006)
74. Patton, A.: Copula based models for financial time series. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (eds.) Handbook of Financial Time Series, pp. 767–785. Springer, Berlin (2009)
75. Pickands, J.: Multivariate extreme value distributions (with a discussion). In: Proceedings of the 43rd Session of the International Statistical Institute, pp. 894–902. Bull. Inst. Int. Stat. **49**, 859–878 (1981)
76. Prokhorov, A.: A goodness-of-fit test for copulas. Working Paper, Columbia University (2008)
77. Quessy, J.-F.: Testing for bivariate extreme dependence using Kendall's process. Scandin. J. Stat. **39**, 497–514 (2011)
78. Quessy, J.-F., Mesfioui, M., Toupin, M.-H.: On a new goodness-of-fit process for families of copulas. Canad. J. Stat. **37**, 80–101 (2009)
79. Rémillard, B.: Goodness-of-fit tests for copulas of multivariate time series. Working Paper (2010)
80. Rémillard, B., Scaillet, O.: Testing the equality between two copulas. J. Multivar. Anal. **100**, 377–386 (2009)
81. Rosenblatt, M.: Remarks on a multivariate transformation. Ann. Math. Stat. **23**, 470–472 (1952)
82. Rüschendorf, L.: Asymptotic distributions of multivariate rank order statistics. Ann. Stat. **4**, 912–923 (1976)
83. Saunders, R., Laud, P.: The multidimensional Kolmogorov-Smirnov test. Biometrika **67**, 237 (1980)
84. Savu, C., Trede, M.: Hierarchical Archimedean copulas. In: International Conference on High Frequency Finance, Konstanz, Germany, May 2006
85. Savu, C., Trede, M.: Goodness-of-fit tests for parametric families of Archimedean copulas. Quant. Financ. **8**, 109–116 (2008)
86. Scaillet, O.: Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. J. Multivar. Anal. **98**, 533–543 (2007)
87. Segers, J.: Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. Bernoulli **26**, 288–305 (2012)
88. Shih, J.H.: A goodness-of-fit test for association in a bivariate survival model. Biometrika **85**, 189–200 (1998)
89. Stupňanová, A., Kolesárová, A.: Associative n-dimensional copulas. Kybernetika **47**, 93–99 (2011)
90. Wang,W.,Wells, M.T.: Model selection and semiparametric inference for bivariate failure-time data. J. Am. Stat. Ass. **95**, 62–72 (2000)
91. Weiss, G.: On the robustness of goodness-of-fit tests for copulas. Discussion Paper, Dortmund (2011)
92. Weiss, G.: Are copula-GOF-test of any practical use? Empirical evidence for stocks, commodities and FX futures. Q. Rev. Econ. Financ. **51**, 173–188 (2011)
93. White, H.: Maximum likelihood estimation in misspecified models. Econometrica **50**, 1–26 (1982)
94. Whelan, N.: Sampling from Archimedean copulas. Quant. Financ. **4**, 339–352 (2004)

Chapter 5
Assessing and Modeling Asymmetry in Bivariate Continuous Data

Christian Genest and Johanna G. Nešlehová

Abstract A bivariate copula is the cumulative distribution function of a pair (U, V) of uniform random variables. This copula is said to be symmetric if and only if (V, U) and (U, V) have the same distribution. Many standard bivariate parametric families of copulas have this property; Archimedean and meta-elliptical copulas are prime examples. In practice, however, dependence is often asymmetric. This paper revisits key aspects of this issue from a modeling perspective. Measures of asymmetry and rank-based estimators thereof are discussed, along with recently proposed tests of symmetry. Several techniques for the construction of asymmetric dependence structures are critically reviewed. A hydrological data set is used for illustration purposes.

5.1 Introduction

Let (X, Y) be a pair of continuous random variables and let its marginal and joint cumulative distribution functions be denoted, at each $x, y \in \mathbb{R}$, by $F(x) = \Pr(X \leq x)$, $G(y) = \Pr(Y \leq y)$, and $H(x, y) = \Pr(X \leq x, Y \leq y)$, respectively. Sklar's Representation Theorem states that there exists a unique copula C such that, for all $x, y \in \mathbb{R}$,

$$H(x, y) = C\{F(x), G(y)\}. \tag{5.1}$$

In fact, C is the cumulative distribution function of the pair $(U, V) = (F(X), G(Y))$ having uniform margins on the interval $(0, 1)$.

In applications, H is typically unknown. A convenient way to model it consists of expressing H in the form (5.1) and assuming that F, G, and C belong to specific

C. Genest (✉) · J.G. Nešlehová
Department of Mathematics and Statistics, McGill University, Montréal (Québec), Canada
e-mail: cgenest@math.mcgill.ca; neslehova@math.mcgill.ca

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_5,

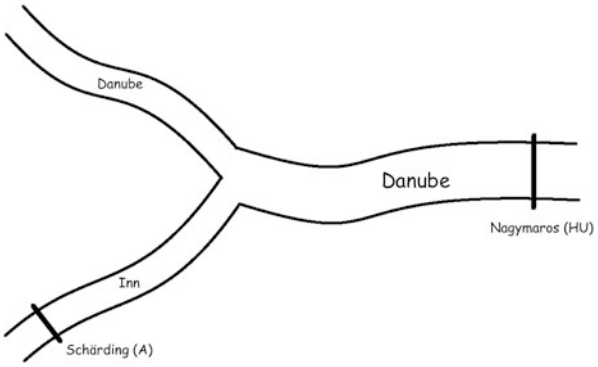

Fig. 5.1 Stylized map showing the location of the Schärding and Nagymaros measurement stations on the Inn and Danube rivers, respectively

parametric classes of distributions. For instance, one might take F Gaussian, G Pareto, and C from the Farlie–Gumbel–Morgenstern family of copulas .

Many parametric classes of bivariate copulas have been proposed; see, e.g., [26, 36] for extensive lists. Most classical copulas are symmetric, however. This property, also referred to as exchangeability, means that for all $u, v \in [0, 1]$,

$$C(u, v) = C(v, u). \tag{5.2}$$

When this condition fails for some $u, v \in [0, 1]$, C is said to be asymmetric, or non-exchangeable. Note that in the copula literature, the term "asymmetric" sometimes refers to the lack of radial symmetry, and particularly the presence of right-tail but no left-tail dependence; see, e.g., [39, 47] for applications in asset allocation and equity markets. These types of asymmetry are not further considered here.

Asymmetric dependence, i.e., failure of (5.2) for some $u, v \in [0, 1]$, typically occurs when there is a causality relationship between the variables. Consider, for instance, the monthly average flow rate (in m^3/s) of the Inn and Danube rivers [48], measured at the Schärding and Nagymaros stations, respectively. The stylized map in Fig. 5.1 suggests that high discharge X of the Inn (upstream) is likely to imply high discharge Y of the Danube (downstream), but not necessarily vice versa. Asymmetry in the dependence between X and Y is apparent from Fig. 5.2, which shows approximate scatter plots of the copulas of the raw (left) and de-trended data (right).

The purpose of this paper is to provide a critical review of the literature on asymmetric copulas and to contribute to it in modest ways. Its structure reflects the progression of a statistical analysis of asymmetric dependence; the river discharge data are used throughout for illustration. Measures of asymmetry and their estimates are discussed in Sect. 5.2. Tests of asymmetry based on these measures are presented in Sect. 5.3. Techniques for constructing asymmetric copulas are reviewed in Sect. 5.4. Concluding comments are given in Sect. 5.5.

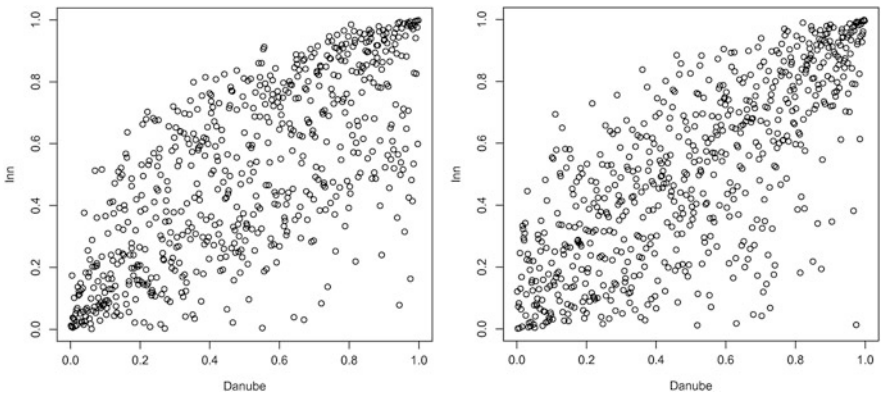

Fig. 5.2 Rank plots of pairs of raw data (*left panel*) and de-trended (*right panel*) monthly average flow rate at the Schärding and Nagymaros stations in the period 1936–1991

5.2 Measures of Asymmetry

Given $p \geq 1$, a natural measure of asymmetry in a copula C is given by

$$\mu_p(C) = \left\{ \int_0^1 \int_0^1 |C(u,v) - C(v,u)|^p \mathrm{d}v\mathrm{d}u \right\}^{1/p},$$

i.e., the L_p distance between C and its "transpose" $C^\top$ defined, for all $u, v \in [0,1]$, by $C^\top(u,v) = C(v,u)$. One could also consider letting $p \to \infty$, which leads to

$$\mu_\infty(C) = \sup_{(u,v)\in[0,1]^2} |C(u,v) - C(v,u)|.$$

These measures were first discussed in a copula context by Klement and Mesiar [28] and Nelsen [37]. Durante et al. [13] show that the measures μ_p, with $p \in [1,\infty]$, satisfy a series of natural axioms for measures of non-exchangeability. To be specific, let $\mathscr{C}$ be the class of bivariate copulas and, for any $C \in \mathscr{C}$ and $u, v \in [0,1]$, set

$$\bar{C}(u,v) = u + v - 1 + C(1-u, 1-v).$$

Durante et al. [13] then show that:

(B1) There exists $K > 0$ such that $0 \leq \mu_p(C) \leq K$ for all $C \in \mathscr{C}$.
(B2) $\mu_p(C) = 0$ if and only if $C = C^\top$.
(B3) $\mu_p(C) = \mu_p(C^\top)$ for all $C \in \mathscr{C}$.
(B4) $\mu_p(C) = \mu_p(\bar{C})$ for all $C \in \mathscr{C}$.
(B5) If $C_n \in \mathscr{C}$ and if $C_n \to C \in \mathscr{C}$ uniformly as $n \to \infty$, then $\mu_p(C_n) \to \mu_p(C)$.

Siburg and Stoimenov [43] consider alternative measures of the form

$$\delta(C) = \frac{\|C_s\|^2 - \|C_a\|^2}{\|C\|^2},$$

which take into account both the symmetric part $C_s = (C + C^\top)/2$ and the asymmetric part $C_a = (C - C^\top)/2$ of C. The norm $\|\cdot\|$ can be arbitrary, but the authors focus on a modified Sobolev norm. Given that their measure is equivalent to μ_∞ for the uniform norm and that the Sobolev norm seems to be of limited statistical use, their idea is not considered further here.

5.2.1 Maximal Asymmetry

It is clear that whatever $p \in [1, \infty]$, $\mu_p(C) = 0$ if and only if C is symmetric. Because $|C(u, v)| \leq 1$ for any $u, v \in [0, 1]$, it is immediate that $\mu_p(C) \leq 2$ for any choice of C and $p \in [1, \infty]$. This bound is, however, never attained and as such unsuitable for standardization purposes. To obtain sharper bounds, Klement and Mesiar [28] and Nelsen [37] prove that for any C and $u, v \in [0, 1]$, $|C(u, v) - C(v, u)| \leq \Delta(u, v)$ where

$$\Delta(u, v) = \min(u, v, 1 - u, 1 - v, |v - u|).$$

This implies that for any C,

$$\mu_\infty(C) \leq \kappa_\infty = \sup_{u,v\in[0,1]} |\Delta(u, v)| = \frac{1}{3},$$

while for any $p \in [1, \infty)$,

$$\mu_p(C) \leq \kappa_p = \left\{ \int_0^1 \int_0^1 \Delta(u, v)^p \mathrm{d}v\mathrm{d}u \right\}^{1/p} = \left\{ \frac{2 \times 3^{-p}}{(p + 1)(p + 2)} \right\}^{1/p}$$

as computed by Durante et al. [13]. Although Proposition 3.1 in [28] states that no copula exists such that $|C(u, v) - C(v, u)| = \Delta(u, v)$ for all $u, v \in [0, 1]$, the bound κ_∞ turns out to be sharp. The following result is excerpted from [37].

Proposition 5.1. *If C is an arbitrary copula, then $\mu_\infty(C) = 1/3$ if and only if either (i) $C(\frac{1}{3}, \frac{2}{3}) = \frac{1}{3}$ and $C(\frac{2}{3}, \frac{1}{3}) = 0$, or (ii) $C(\frac{1}{3}, \frac{2}{3}) = 0$ and $C(\frac{2}{3}, \frac{1}{3}) = \frac{1}{3}$.*

By Proposition 5.1, $\mu_\infty(C) = 1/3$ whenever C or $C^\top$ places one third of its probability mass in each of the rectangles $[0, 1/3] \times [1/3, 2/3]$, $[1/3, 2/3] \times [2/3, 1]$ and $[2/3, 1] \times [0, 1/3]$. There are infinitely many such copulas, one prime example being

$$C_\infty(u,v) = \max\{\max(u+v-1,0), \min(u, v-1/3)\},$$

which is a singular copula whose support consists of two line segments, one from $(0,1/3)$ to $(2/3,1)$ and the other from $(2/3,1/3)$ to $(1,0)$.

When $p \in [1,\infty)$, the matter is less clear, however. Although Proposition 4 of Durante et al. [13] guarantees that there exists a copula C_p for which μ_p is maximized, it is not known what C_p looks like and whether $\mu_p(C_p) = \kappa_p$.

Finally, there are several interesting relationships between symmetry and association. As explained by Nelsen [37], Proposition 5.1 leads to bounds on the Spearman and Kendall measures of association given, respectively, by

$$\rho(C) = -3 + 12\int_0^1\int_0^1 C(u,v)\mathrm{d}v\mathrm{d}u, \quad \tau(C) = -1 + 4\int_0^1\int_0^1 C(u,v)\mathrm{d}C(u,v).$$

Nelsen concludes that whenever $\mu_\infty(C) = 1/3$, $\rho(C) \in [-5/9,-1/3]$ while $\tau(C) \in [-5/9,1/9]$. This supports the statement of De Baets et al. [6] who write that "positive dependence plays in favor of symmetry." These authors find that when C is positive quadrant dependent, i.e., if for all $u,v \in [0,1]$, $C(u,v) \geq uv$, the measure of asymmetry $\mu_\infty(C)$ is at most $3-2\sqrt{2} \approx 0.172$. They also identify the copula for which the bound is achieved. When C has the stronger positive dependence property called stochastic increasingness in both arguments, Durante and Papini [10] show that the upper bound for μ_∞ is given by $(5\sqrt{5}-11)/2 \approx 0.09$.

When C is negative quadrant dependent, i.e., if for all $u,v \in [0,1]$, $C(u,v) \leq uv$, Durante and Papini [11] show that $\mu_\infty(C) \leq \sqrt{5}-2 \approx 0.236$. They also find that when C is stochastically decreasing in both arguments, $\mu_\infty(C) \leq 3-2\sqrt{2} \approx 0.172$ and conclude that symmetric concepts of strong negative association decrease the level of possible asymmetry.

5.2.2 Estimates

Given a random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ from an unknown distribution H with unique underlying copula C, the question arises as to how to estimate $\mu_p(C)$. To this end, let the (rescaled) empirical distribution of H be defined, for all $x,y \in \mathbb{R}$, by

$$H_n(x,y) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x, Y_i \leq y),$$

and let F_n and G_n denote its margins. The more traditional definition of H_n involving division by n rather than $n+1$ leads to slightly different expressions in finite samples, but this has no effect asymptotically.

For each $i \in \{1,\ldots,n\}$, the pair $(\hat{U}_i,\hat{V}_i) = (F_n(X_i), G_n(Y_i))$ may be viewed as a pseudo-observation from C. A natural estimate of C is then given by the empirical copula $\hat{C}_n$ defined, at every $u,v \in [0,1]$, by

$$\hat{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{U}_i \le u, \hat{V}_i \le v).$$

Note that $\hat{C}_n$ is rank-based because for all $i \in \{1,\dots,n\}$, $(n+1)\hat{U}_i$ is the rank of X_i among $X_1,\dots,X_n$ and similarly, $(n+1)\hat{V}_i$ is the rank of Y_i among $Y_1,\dots,Y_n$.

Plug-in estimates of $\mu_\infty(C)$ and $\mu_p(C)$ for $p \in [1,\infty)$ are then given by

$$\mu_\infty(\hat{C}_n) = \sup_{(u,v)\in[0,1]^2} \left|\hat{C}_n(u,v) - \hat{C}_n(v,u)\right|,$$

$$\mu_p(\hat{C}_n) = \left\{\int_0^1\int_0^1 \left|\hat{C}_n(u,v) - \hat{C}_n(v,u)\right|^p \mathrm{d}v\,\mathrm{d}u\right\}^{1/p}.$$

Their consistency stems from the fact that the empirical copula $\hat{C}_n$ is itself a consistent estimator of C, provided that C is regular in the following sense [41].

Definition 5.1. A bivariate copula C is said to be regular if

(i) the partial derivatives $\dot{C}_1(u,v) = \partial C(u,v)/\partial u$ and $\dot{C}_2(u,v) = \partial C(u,v)/\partial v$ exist everywhere on $[0,1]^2$, where by convention, one-sided derivatives are used at the boundary points;
(ii) $\dot{C}_1$ is continuous on $(0,1)\times[0,1]$ and $\dot{C}_2$ is continuous on $[0,1]\times(0,1)$.

As illustrated by Segers [41], all copulas commonly used in practice satisfy this condition. Assuming henceforth that C is regular and observing that μ_p is a continuous functional of C for each $p \in [1,\infty]$, the Continuous Mapping Theorem readily implies the following result.

Proposition 5.2. *If C is a regular copula, then for all $p \in [1,\infty]$, $\mu_p(\hat{C}_n)$ converges in probability to $\mu_p(C)$ as $n \to \infty$.*

The estimator $\mu_\infty(\hat{C}_n)$ is easier to compute than it appears because

$$\mu_\infty(\hat{C}_n) = \max_{i,j\in\{1,\dots,n\}} \left|\hat{C}_n\left(\frac{i}{n+1},\frac{j}{n+1}\right) - \hat{C}_n\left(\frac{j}{n+1},\frac{i}{n+1}\right)\right|.$$

As shown by Genest et al. [23], one has also

$$\mu_2^2(\hat{C}_n) = \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(1-\hat{U}_i\vee\hat{U}_j)(1-\hat{V}_i\vee\hat{V}_j) - (1-\hat{U}_i\vee\hat{V}_j)(1-\hat{V}_i\vee\hat{U}_j),$$

where for arbitrary $a,b \in \mathbb{R}$, $a\vee b = \max(a,b)$. For other values of p, $\mu_p(\hat{C}_n)$ may be tedious to compute. To circumvent this problem, one might consider an alternative empirical measure of asymmetry given by

$$\nu_p(\hat{C}_n) = \left\{ \int_0^1 \int_0^1 \left| \hat{C}_n(u,v) - \hat{C}_n(v,u) \right|^p \mathrm{d}\hat{C}_n(u,v) \right\}^{1/p}.$$

Because $\hat{C}_n$ is the joint distribution function of a counting probability measure,

$$\nu_p^p(\hat{C}_n) = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{C}_n(\hat{U}_i, \hat{V}_i) - \hat{C}_n(\hat{V}_i, \hat{U}_i) \right|^p.$$

Furthermore, the following result is showed in the Appendix.

Proposition 5.3. *For $p \in [1, \infty)$, $\nu_p(\hat{C}_n)$ converges in probability, as $n \to \infty$, to*

$$\nu_p(C) = \left\{ \int_0^1 \int_0^1 |C(u,v) - C(v,u)|^p \, \mathrm{d}C(u,v) \right\}^{1/p}. \tag{5.3}$$

Equation (5.3) defines a new population measure of asymmetry in the sense of Durante et al. [13]. This result, proved in the Appendix, is formally recorded below.

Proposition 5.4. *For any $p \in [1, \infty)$, ν_p satisfies the axioms (B1)–(B5).*

While ν_p is easier to estimate than μ_p, a realistic (let alone sharp) upper bound for its value is unknown at present. Such a bound would be useful to compare values of ν_p across copulas and with other standardized measures of asymmetry.

5.2.3 Illustration

Consider the monthly average flow rate (in m^3/s) of the Inn and Danube rivers, as observed at the Schärding and Nagymaros measurement stations in the 55-year period extending from 1936 to 1991. The 660 pairs of normalized ranks $(\hat{U}_i, \hat{V}_i)$ for the raw data are displayed in Fig. 5.2 (left). Though the plot suggests asymmetry, the estimates of μ_∞, μ_2 and ν_2 are not particularly large:

$$3\,\mu_\infty(\hat{C}_n) \approx 0.127, \quad \sqrt{54}\,\mu_2(\hat{C}_n) \approx 0.083, \quad \nu_2(\hat{C}_n) \approx 0.014.$$

For comparison purposes, μ_∞ and μ_2 were divided by $\kappa_\infty = 1/3$ and $\kappa_2 = 1/\sqrt{54}$, respectively. This means that $3\,\mu_\infty$ takes values in the entire interval $[0, 1]$, while it is only known that $\sqrt{54}\,\mu_2 \leq 1$. Although κ_2 may not be sharp, it seems to give a reasonable standardization in practice. As an upper bound for ν_2 is unknown, no standardization was made. Given that the values of μ_2 and ν_2 are often close, κ_2 can be used as a rule of thumb adjustment; this yields $\sqrt{54}\,\nu_2(\hat{C}_n) \approx 0.0997$.

Hydrological data typically exhibit time trends. Such is the case here and hence the relationship between the two variables may be confounded with serial dependence. To eliminate this effect, Bacigál et al. [2] suggest de-trending the series by AR(1) models. The resulting residuals exhibit no further significant

autocorrelation and can thus be used to study the time-invariant dependence between the variables (assuming there is one). The rank plot of the 659 pairs of residuals is given in Fig. 5.2 (right) and the standardized sample values of the asymmetry measures are

$$3\,\mu_\infty(\hat{C}_n) \approx 0.091, \quad \sqrt{54}\,\mu_2(\hat{C}_n) \approx 0.063, \quad \sqrt{54}\,\nu_2(\hat{C}_n) \approx 0.065.$$

Note that the small values of the asymmetry measures observed above do not come as a surprise given that the variables exhibit substantial positive correlation. For example, the sample value of Kendall's tau is approximatively 0.525 for the raw data and 0.548 for the de-trended data. Whether the above sample measures of asymmetry are significantly greater than zero can only be determined using formal statistical tests, which are described next.

5.3 Testing for Symmetry

Tests of the hypothesis

$$\mathscr{H}_0 : C = C^\top$$

of symmetry have been developed recently by Genest et al. [23]. Their procedures are based on the rank-based statistics $R_n = \mu_2^2(\hat{C}_n)$, $S_n = \nu_2^2(\hat{C}_n)$ and $T_n = \mu_\infty(\hat{C}_n)$, whose values tend to zero in probability as $n \to \infty$ under $\mathscr{H}_0$. Note that the empirical copula in [23] is based on the pairs $\{(n+1)/n\}(\hat{U}_i, \hat{V}_i)$. However, this slightly different definition is inconsequential asymptotically and is not adopted here.

To carry out these tests, the asymptotic null distribution of their corresponding statistic is needed. If C is regular in the sense of Definition 5.1, the so-called empirical copula process given, for all $u, v \in [0,1]$, by

$$\hat{\mathbb{C}}_n(u,v) = \sqrt{n}\,\{\hat{C}_n(u,v) - C(u,v)\}$$

converges weakly in the space $\ell[0,1]^2$ of bounded functions on $[0,1]^2$ equipped with the uniform norm [41]. In other words, one has $\hat{\mathbb{C}}_n \rightsquigarrow \hat{\mathbb{C}}$ as $n \to \infty$, where $\hat{\mathbb{C}}$ is a centered Gaussian process defined, for all $u, v \in [0,1]$, by

$$\hat{\mathbb{C}}(u,v) = \mathbb{C}(u,v) - \dot{C}_1(u,v)\mathbb{C}(u,1) - \dot{C}_2(u,v)\mathbb{C}(1,v).$$

Here, $\mathbb{C}$ is a tucked C-Brownian sheet, i.e., a centered Gaussian random field with covariance function given, for all $u, v, s, t \in [0,1]$, by

$$\Gamma_{\mathbb{C}}(u,v,s,t) = C(u \wedge s, v \wedge t) - C(u,v)\,C(s,t),$$

where for arbitrary $a, b \in \mathbb{R}$, $a \wedge b = \min(a,b)$. For variants, see [16, 18, 40, 44].

Using the continuous mapping theorem, Genest et al. [23] showed that if $\hat{\mathbb{D}}_n$ is the empirical process defined, for all $u, v \in [0,1]$, by

$$\hat{\mathbb{D}}_n(u,v) = \sqrt{n}\,\{\hat{C}_n(u,v) - \hat{C}_n(v,u)\},$$

then as $n \to \infty$, $\hat{\mathbb{D}}_n \rightsquigarrow \hat{\mathbb{D}}$ in $\ell[0,1]^2$, where $\hat{\mathbb{D}}$ admits the representation

$$\hat{\mathbb{D}}(u,v) = \mathbb{D}(u,v) - \dot{C}_1(u,v)\,\mathbb{D}(u,1) - \dot{C}_2(u,v)\,\mathbb{D}(1,v),$$

in terms of a centered Gaussian random field $\mathbb{D}$ with covariance function given, at each $u, v, s, t \in [0,1]$, by $\Gamma_{\mathbb{D}}(u,v,s,t) = 2\,\{\Gamma_{\mathbb{C}}(u,v,s,t) - \Gamma_{\mathbb{C}}(u,v,t,s)\}$. This observation leads to the following result, excerpted from [23].

Proposition 5.5. *If C is a regular symmetric copula, then as $n \to \infty$,*

$$nR_n = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n(u,v)\}^2 \mathrm{d}v\mathrm{d}u \quad \rightsquigarrow \quad \mathbb{D}_R = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}(u,v)\}^2 \mathrm{d}v\mathrm{d}u,$$

$$nS_n = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n(u,v)\}^2 \mathrm{d}\hat{C}_n(u,v) \quad \rightsquigarrow \quad \mathbb{D}_S = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}(u,v)\}^2 \mathrm{d}C(u,v),$$

$$n^{1/2}\,T_n = \sup_{(u,v)\in[0,1]^2} |\hat{\mathbb{D}}_n(u,v)| \quad \rightsquigarrow \quad \mathbb{D}_T = \sup_{(u,v)\in[0,1]^2} |\hat{\mathbb{D}}(u,v)|.$$

Unfortunately, the limiting distribution of all three statistics depends on the underlying copula C, which is unknown.

5.3.1 *p*-Value Computation

As shown in [23], valid p-values for the tests based on R_n, S_n, and T_n can be computed using a bootstrap approximation based on the Multiplier Central Limit Theorem of van der Vaart and Wellner [46]; see also [4, 41] for further details. The step-by-step description of this procedure is provided below; an implementation using the R Project for Statistical Computing is available from the authors.

Step 0. Compute the statistic R_n, S_n or T_n.

Step 1. Define P_n at any $u, v \in [0,1]$ as the $n \times 1$ vector with ith component

$$P_{in}(u,v) = \mathbf{1}(\hat{U}_i \le u, \hat{V}_i \le v) - \mathbf{1}(\hat{U}_i \le v, \hat{V}_i \le u).$$

Step 2. Fix a bandwidth $\ell_n \in (0, 1/2)$, typically $\ell_n \approx 1/\sqrt{n}$, and a large integer M. For each $h \in \{1, \ldots, M\}$, do the following.

Step 2a. For arbitrary $v \in [0,1]$, set

$$\dot{C}_{1n}(u,v) = \begin{cases} \dfrac{\hat{C}_n(2\ell_n, v)}{2\ell_n} & \text{if } u \in [0, \ell_n), \\ \dfrac{\hat{C}_n(u+\ell_n, v) - \hat{C}_n(u-\ell_n, v)}{2\ell_n} & \text{if } u \in [\ell_n, 1-\ell_n], \\ \dfrac{\hat{C}_n(1, v) - \hat{C}_n(1-2\ell_n, v)}{2\ell_n} & \text{if } u \in (1-\ell_n, 1]. \end{cases}$$

Similarly, for arbitrary $u \in [0,1]$, set

$$\dot{C}_{2n}(u,v) = \begin{cases} \dfrac{\hat{C}_n(u, 2\ell_n)}{2\ell_n} & \text{if } v \in [0, \ell_n), \\ \dfrac{\hat{C}_n(u, v+\ell_n) - \hat{C}_n(u, v-\ell_n)}{2\ell_n} & \text{if } v \in [\ell_n, 1-\ell_n], \\ \dfrac{\hat{C}_n(u, 1) - \hat{C}_n(u, 1-2\ell_n)}{2\ell_n} & \text{if } v \in (1-\ell_n, 1]. \end{cases}$$

Step 2b. Draw a vector $\xi^{(h)} = (\xi_1^{(h)}, \ldots, \xi_n^{(h)})$ of independent nonnegative random variables with unit mean and unit variance; the standard exponential distribution is typically used to this end. Set

$$\bar{\xi}_n^{(h)} = \frac{1}{n}\,(\xi_1^{(h)} + \cdots + \xi_n^{(h)}) \quad \text{and} \quad \Xi_n^{(h)} = \left(\frac{\xi_1^{(h)}}{\bar{\xi}_n^{(h)}} - 1, \ldots, \frac{\xi_n^{(h)}}{\bar{\xi}_n^{(h)}} - 1 \right).$$

Step 2c. Define the bootstrap replicate $\hat{\mathbb{D}}_n^{(h)}$ of $\hat{\mathbb{D}}$ at any $u, v \in [0,1]$ by

$$\hat{\mathbb{D}}_n^{(h)}(u,v) = \frac{1}{\sqrt{n}}\,\Xi_n^{(h)}\{P_n(u,v) - \dot{C}_{1n}(u,v)\,P_n(u,1) - \dot{C}_{2n}(u,v)\,P_n(1,v)\}.$$

Step 2d. Compute the bootstrap replicate of the appropriate test statistic, viz.

$$R_n^{(h)} = \frac{1}{n}\int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n^{(h)}(u,v)\}^2 \mathrm{d}v\mathrm{d}u,$$

$$S_n^{(h)} = \frac{1}{n}\int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n^{(h)}(u,v)\}^2 \mathrm{d}\hat{C}_n(u,v),$$

$$T_n^{(h)} = \frac{1}{\sqrt{n}} \sup_{(u,v)\in[0,1]^2} |\hat{\mathbb{D}}_n^{(h)}(u,v)|.$$

Step 3. Compute the approximate p-value, viz.

$$\frac{1}{M}\sum_{h=1}^{M}\mathbf{1}(R_n^{(h)} > R_n), \quad \frac{1}{M}\sum_{h=1}^{M}\mathbf{1}(S_n^{(h)} > S_n), \quad \frac{1}{M}\sum_{h=1}^{M}\mathbf{1}(T_n^{(h)} > T_n).$$

It is easy to compute $S_n^{(h)}$ because $\hat{C}_n$ is a discrete distribution function. For the other two statistics, the computational burden can be reduced by resorting to a numerical approximation involving an $N \times N$ grid, viz.

$$R_n^{(h)} \approx \frac{1}{nN^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\left\{\hat{\mathbb{D}}_n^{(h)}\left(\frac{k}{N}, \frac{\ell}{N}\right)\right\}^2, \tag{5.4a}$$

$$T_n^{(h)} \approx \frac{1}{\sqrt{n}}\max_{k,\ell\in\{1,\ldots,N\}}\left|\hat{\mathbb{D}}_n^{(h)}\left(\frac{k}{N}, \frac{\ell}{N}\right)\right|. \tag{5.4b}$$

The results of an extensive Monte Carlo simulation study comparing the power of these three tests were reported in [23]. The test based on the Cramér—von Mises statistic S_n was found to be generally more powerful than its competitors. It is also the quickest to perform and hence can be recommended on that account too.

5.3.2 *Illustration*

Consider once again the monthly average flow rate of the Inn and Danube rivers. The tests based on R_n, S_n, and T_n were applied to both the raw and de-trended data using $M = 1,000$ multiplier replicates, a bandwidth $\ell_n = 0.04 \approx 1/\sqrt{660}$ and $N = 100$ grid points in the approximation (5.4). The hypothesis $\mathscr{H}_0$ of symmetry was rejected in all cases. The p-values were essentially zero except when the statistic T_n was used on the de-trended data, where it was found that $p \approx 1.5\,\%$.

5.4 Asymmetric Copula Families

When the null hypothesis of symmetry is rejected, dependence models based on Archimedean and meta-elliptical copula families are ruled out straightaway, because they cannot account for asymmetry; popular alternatives are reviewed below.

5.4.1 *Extreme-Value Copulas*

A copula C is said to be of the extreme-value type if and only if there exists a function $A : [0, 1] \to [1/2, 1]$ such that, for all $u, v \in [0, 1]$,

$$C(u,v) = \exp\left[\ln(uv)A\left\{\frac{\ln(v)}{\ln(uv)}\right\}\right]. \tag{5.5}$$

For C to be a copula, the so-called Pickands dependence function A must be convex and such that, for all $t \in [0,1]$, $\max(t, 1-t) \le A(t) \le 1$. The bounds $A(t) \equiv 1$ and $A(t) = \max(t, 1-t)$ correspond to the independence copula and the Fréchet–Hoeffding upper bound, respectively.

An extreme-value copula is asymmetric if and only if its Pickands dependence function A is asymmetric with respect to $1/2$, i.e., if there exists $t \in [0,1]$ such that $A(t) \neq A(1-t)$. Klement and Mesiar [28] state that if C is of the form (5.5), then

$$\mu_\infty(C) \le 4^4/5^5 \approx 0.082.$$

A detailed proof of this result is given by Durante and Mesiar [9], who show that the upper bound is reached for two members of the Marshall–Olkin extreme-value copula family whose Pickands dependence functions are given, for all $t \in [0,1]$, by

$$A_1(t) = \max\left(1-t, \frac{t+1}{2}\right), \quad A_2(t) = \max\left(t, \frac{2-t}{2}\right). \tag{5.6}$$

Typical examples of symmetric extreme-value copulas include the Galambos, Gumbel, Hüsler–Reiß, Tawn, and t-EV families; see [19] for their definitions and further details. Each of these families can be made asymmetric using Khoudraji's device [21, 27]. The latter is based on the observation that if C is an extreme-value copula with Pickands dependence function A and $\lambda, \kappa \in (0,1)$, then the copula given, for all $u, v \in [0,1]$, by

$$C_{\lambda,\kappa}(u,v) = u^{1-\lambda} v^{1-\kappa} C(u^\lambda, v^\kappa) \tag{5.7}$$

is again extreme-value with Pickands dependence function of the form

$$A_{\lambda,\kappa}(t) = (1-\kappa)t + (1-\lambda)(1-t) + \{\kappa t + \lambda(1-t)\}A\left\{\frac{\kappa t}{\kappa t + \lambda(1-t)}\right\}.$$

A random pair (U,V) from $C_{\lambda,\kappa}$ can be obtained as follows.

Step 1. Draw independent random variables W and Z, uniform on $[0,1]$.
Step 2. Draw a pair (X,Y) from the copula C and set

$$U = \max(W^{1/(1-\lambda)}, X^{1/\lambda}), \quad V = \max(Z^{1/(1-\kappa)}, Y^{1/\kappa}).$$

Clearly, $C_{\lambda,\kappa}$ is asymmetric when $\lambda \neq \kappa$. Figure 5.4 shows the effect of Khoudraji's device when C is a Gumbel copula with $\tau \in \{0.5, 0.75, 0.9\}$ and $\lambda = 0.5, \kappa = 0.7$; samples from the symmetric Gumbel copula and from $C_{\lambda,\kappa}$ are displayed in the top and middle row, respectively. The plots confirm that asymmetry restricts the range

of attainable positive association, as observed in Sect. 5.2.1. It is shown in [22] that Kendall's tau of $C_{\lambda,\kappa}$ satisfies

$$\tau(C_{\lambda,\kappa}) \leq \frac{\kappa\lambda}{\kappa+\lambda-\kappa\lambda} = \tau_{\max}(\lambda,\kappa). \tag{5.8}$$

Khoudraji's device can be generalized by choosing an integer $m \geq 1$, extreme-value copulas $C_1, \ldots, C_m$, and vectors $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m), \boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_m) \in [0,1]^m$ whose components sum up to 1. The copula defined, for all $u, v \in [0,1]$, by

$$C_{\boldsymbol{\lambda},\boldsymbol{\kappa}}(u,v) = \prod_{j=1}^{m} C_j(u^{\lambda_j}, v^{\kappa_j}) \tag{5.9}$$

is then an extreme-value copula with Pickands dependence function

$$A_{\boldsymbol{\lambda},\boldsymbol{\kappa}}(t) = \sum_{j=1}^{m} \{\kappa_j t + \lambda_j(1-t)\} A_j\left\{\frac{\kappa_j t}{\kappa_j t + \lambda_j(1-t)}\right\}.$$

This result is proved, e.g., by Bacigál et al. [2], who mention other constructions.

The availability of statistical tools for model fitting and validation makes extreme-value copulas particularly convenient. However, these dependence structures may not always be appropriate. In particular, copulas of the form (5.5) are stochastically increasing [17] and exhibit upper-tail but no lower-tail dependence.

5.4.2 *Asymmetric Generalizations of Archimedean Copulas*

A copula C is called Archimedean if it can be written, for all $u, v \in [0,1]$, in the form

$$C(u,v) = \varphi^{-1}\{\varphi(u) + \varphi(v)\}$$

in terms of a strictly decreasing, convex map $\varphi : (0,1] \to [0,\infty)$ such that $\varphi(1) = 0$. By convention, $\varphi(0) = \lim_{u \downarrow 0} \varphi(u)$ and $\varphi^{-1}(s) = 0$ when $s \geq \varphi(0)$. The function φ is referred to as an Archimedean generator; see Chap. 4 in [36] for examples.

Clearly, any Archimedean copula is symmetric. However, Khoudraji's device can again be used to generate an asymmetric copula from an Archimedean copula C with generator φ through (5.7). The resulting copula $C_{\lambda,\kappa}$ is no longer Archimedean; see [34] for further details and illustrations.

Although Khoudraji's device induces asymmetry, it does not provide much freedom in modeling association between the variables. To see this, observe that for given λ, κ, the Fréchet–Hoeffding inequality implies that, for all $u, v \in [0,1]$,

$$\max(0, uv^{1-\kappa} + u^{1-\lambda}v - u^{1-\lambda}v^{1-\lambda}) \leq C_{\lambda,\kappa}(u,v) \leq \min(uv^{1-\kappa}, u^{1-\lambda}v).$$

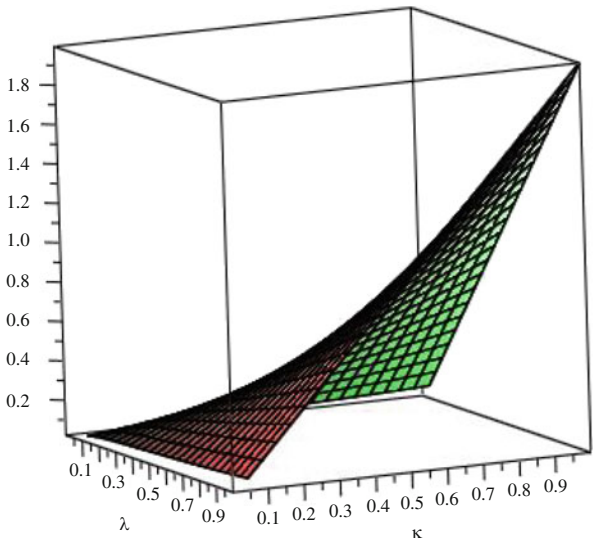

Fig. 5.3 Plot of the difference $\tau_{\max}(\lambda, \kappa) - \tau_{\min}(\lambda, \kappa)$ as a function of λ and κ

The attainable range of values of $\tau(C_{\lambda,\kappa})$ is thus $[\tau_{\min}(\lambda, \kappa), \tau_{\max}(\lambda, \kappa)]$ with $\tau_{\max}(C_{\lambda,\kappa})$ given in (5.8) and

$$\tau_{\min}(\lambda, \kappa) = \frac{\kappa\lambda}{\kappa + \lambda - \kappa\lambda} \mathrm{B}\left(\frac{\lambda}{2} - 1, \frac{\kappa}{2} - 1\right) - \frac{2\kappa\lambda}{\lambda + \kappa - \lambda\kappa},$$

where $\mathrm{B}(x, y)$, $x, y > 0$, denotes the Beta function. The length of this interval, $\tau_{\max}(\lambda, \kappa) - \tau_{\min}(\lambda, \kappa)$, is shown in Fig. 5.3. For example, it can be seen that $\tau(C_{\lambda,\kappa})$ is small when $|\lambda - \kappa|$ is large. Although the precise interplay between the parameters λ, κ and the asymmetry measures presented in Sect. 5.2 is not known at present, asymmetry appears to increase with $|\lambda - \kappa|$.

An alternative asymmetric generalization of Archimedean copulas has recently been proposed by McNeil and Nešlehová [33]. Using the fact that Archimedean copulas are survival copulas of simplex distributions [32], one can consider survival copulas of the more general class of Liouville distributions. A random pair (X, Y) is said to follow a Liouville distribution if $(X, Y) \stackrel{\mathrm{d}}{=} R \times (D_1, D_2)$, where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution, R is a strictly positive random variable independent of the random pair (D_1, D_2) having Dirichlet(α, β) distribution with $\alpha, \beta > 0$. When $\alpha = \beta = 1$, (X, Y) follows a simplex distribution and its survival copula is Archimedean. The interesting case arises when $\alpha \neq \beta$, as the survival copula is then asymmetric.

Although Liouville copulas do not have a closed form in general, the expression for their density as well as random number generation is comparatively tractable when α, β are integer-valued and the distribution of R is suitably chosen.

When the inverse generator $\psi = \varphi^{-1}$ is completely monotone, one option is to set $R \overset{\mathrm{d}}{=} \varphi(W_1) + \cdots + \varphi(W_{\alpha+\beta})$, where the distribution of $(W_1, \ldots, W_{\alpha+\beta})$ is an $(\alpha + \beta)$-dimensional Archimedean copula C_φ with generator φ. A random pair (U, V) from the corresponding Liouville copula with parameters α, β is then obtained as follows.

Step 1. Draw $(W_1, \ldots, W_{\alpha+\beta})$ from the multivariate Archimedean copula C_φ.
Step 2. Set $X = \varphi(W_1) + \cdots + \varphi(W_\alpha)$ and $Y = \varphi(W_{\alpha+1}) + \cdots + \varphi(W_{\alpha+\beta})$.
Step 3. For $j = 1, \ldots, \alpha \vee \beta - 1$, compute the jth derivative $\psi^{(j)}$ of ψ and set

$$U = \sum_{j=0}^{\alpha-1} (-1)^j \frac{X^j}{j!} \psi^{(j)}(X), \quad V = \sum_{j=0}^{\beta-1} (-1)^j \frac{Y^j}{j!} \psi^{(j)}(Y).$$

The bottom row of Fig. 5.4 shows samples of size $2,000$ from Liouville copulas $C_{\alpha,\beta}$ with $\alpha = 1$ and $\beta = 20$ when C_φ is Gumbel's copula with $\tau \in \{0.5, 0.75, 0.9\}$. The high-order derivatives of the Gumbel generator were computed using Theorem 2 in [25]. For lack of a theoretical expression, α and β were rigged so that the sample values of $\mu_2(C_{\alpha,\beta})$ and $\mu_2(C_{\lambda,\kappa})$ are close when $\tau = 0.5$. It transpires from the plots that in contrast to Khoudraji's device, the Liouville construction does not restrict the range of attainable association. This was demonstrated more formally in [33], where an explicit but cumbersome formula for Kendall's tau may be found.

5.4.3 *Archimax Copulas*

Archimax copulas provide yet another class of asymmetric bivariate copulas. Following Capéraà et al. [5], a copula C belongs to this class if, for all $u, v \in (0, 1)$,

$$C(u, v) = \varphi^{-1}\left[\{\varphi(u) + \varphi(v)\} \times A\left\{\frac{\varphi(v)}{\varphi(u) + \varphi(v)}\right\}\right], \tag{5.10}$$

where A is a Pickands dependence function and φ is a bivariate Archimedean generator. The class is called Archimax because:

(a) If $A \equiv 1$, then (5.10) reduces to an (exchangeable) Archimedean copula.
(b) If $\varphi(t) = -\ln(t)$ for all $t \in (0, 1]$, then (5.10) is an extreme-value copula.

An Archimax copula C with generator φ and Pickands dependence function A is symmetric when $A(t) = A(1 - t)$ for all $t \in [0, 1]$. To construct an asymmetric Archimax copula, one can thus resort to any of the techniques described in Sect. 5.4.1. Following [5, 24], a random pair (U, V) from C can be obtained as follows.

Step 1. Draw a pair (U_1, V_1) from an extreme-value copula with Pickands dependence function A and set $Z = \ln(U_1)/\ln(U_1 V_1)$.

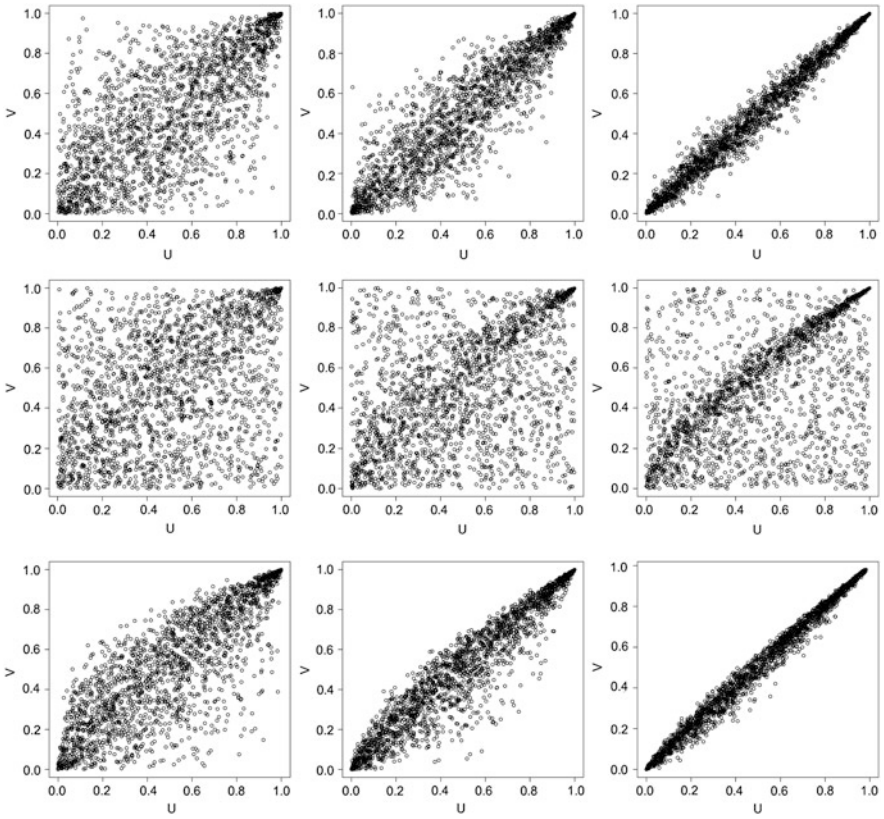

Fig. 5.4 *Top row*: samples of size 2,000 from the Gumbel copula with $\tau = 0.5$ (*left*), $\tau = 0.75$ (*middle*) and $\tau = 0.9$ (*right*). The *middle* and *bottom row* show the effect of asymmetrization of these copulas using Khoudraji's device with $\lambda = 0.5$ and $\kappa = 0.7$ and the Liouville copula construction with $\alpha = 1$ and $\beta = 20$, respectively

Step 2. Compute the first and second order derivatives A' and A'' of A and set

$$p(Z) = \frac{Z(1-Z)A''(Z)}{A(Z) + (1-2Z)A'(Z) + Z(1-Z)[A''(Z) - \{A'(Z)\}^2/A(Z)]}.$$

Step 3. Generate a pair (U_2, V_2) from an Archimedean copula C_φ with generator φ and a random variable U_3 uniformly distributed on $(0, 1)$.

Step 4. Set $W = U_2$ if $U_3 \le p(Z)$ and $W = C_\varphi(U_2, V_2)$ otherwise and compute

$$U = \varphi^{-1}\left\{\frac{Z\varphi(W)}{A(Z)}\right\}, \quad V = \varphi^{-1}\left\{\frac{(1-Z)\varphi(W)}{A(Z)}\right\}.$$

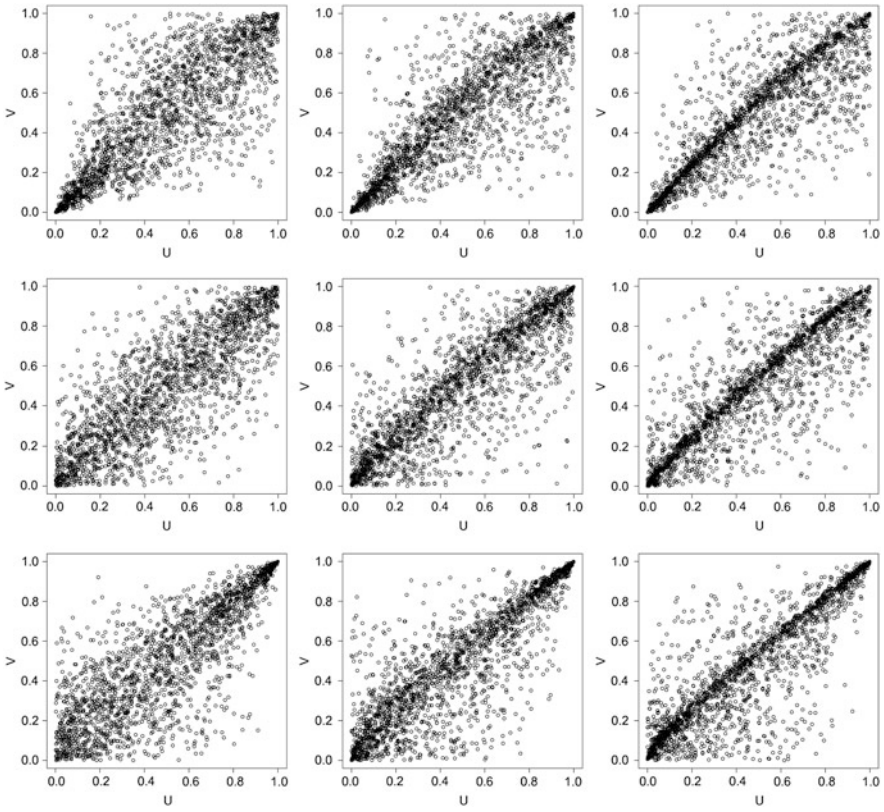

Fig. 5.5 Samples of size 2,000 from the Archimax copula with Gumbel's asymmetric Pickands dependence function with $\lambda = 0.5$, $\kappa = 0.7$ and $\theta = 2$ (*left*), $\theta = 4$ (*middle*) and $\theta = 10$ (*right*). The Archimedean generator φ is the Clayton (*top row*), Frank (*middle row*), and Joe (*bottom row*); its parameter is chosen so that $\tau(C_\varphi) = 0.5$

Typical samples from Archimax copulas are shown in Fig. 5.5. As can be seen, the degree of asymmetry is modest. This is consistent with the finding of Durante and Mesiar [9] who show that

$$\mu_\infty(C) \leq \sup_{t \in (0,\infty)} |\varphi^{-1}(t) - \varphi^{-1}(5t/4)|. \tag{5.11}$$

Once again, the upper bound is reached if either $A = A_1$ or $A = A_2$, as defined in (5.6). For the Clayton, Frank and Joe copulas used in Fig. 5.5, the function $\varphi^{-1}(t) - \varphi^{-1}(5t/4)$ is displayed in Fig. 5.6. The upper bound on μ_∞ in (5.11) is shown in the right panel of the same figure. One can see that the level of attainable asymmetry decreases rather quickly with increasing τ.

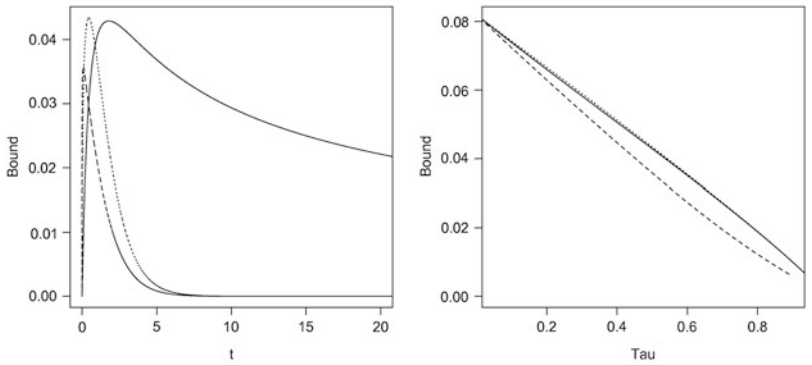

Fig. 5.6 *Left panel*: The function $\varphi^{-1}(t) - \varphi^{-1}(5t/4)$ for the Clayton (*full*), Frank (*dashed*), and Joe (*dotted*) copulas with $\tau = 0.5$. *Right panel*: The upper bound on μ_∞ from (5.11) as a function of τ for the Clayton (*full*), Frank (*dashed*) and Joe (*dotted*) copulas

5.4.4 Algebraic Constructions

In recent years, various other ways of constructing asymmetric copulas have been proposed. They are merely outlined below as in most cases, comparatively little is known about them, particularly from a practical perspective.

As was already apparent from [21], different extensions of Khoudraji's device are possible. Two of them have been investigated in detail. Given $m \geq 2$ copulas $C_1, \ldots, C_m$, Liebscher [30, 31] considers copulas defined, for all $u, v \in [0, 1]$, by

$$C(u, v) = \prod_{j=1}^{m} C_j\{f_j(u), g_j(v)\},$$

where $f_1, \ldots, f_m, g_1, \ldots, g_m$ are strictly increasing mappings from $[0, 1]$ to $[0, 1]$ such that, for all $t \in [0, 1]$, $f_1(t) \times \cdots \times f_m(t) = g_1(t) \times \cdots \times g_m(t) = t$. A simple procedure for generating observations from such copulas is given in [30, 31]. However, no practical guidance for the choice of the functions is provided.

Note that (5.9) corresponds to the case where $f_j(t) = t^{\lambda_j}$ and $g_j(t) = t^{\kappa_j}$ for all $j \in \{1, \ldots, m\}$. When $m = 2$, taking $f_2(t) = t/f_1(t)$ and $g_2(t) = t/g_1(t)$ for all $t \in (0, 1]$ leads to a construction proposed independently by Durante [7]. In his paper, this author also considers mappings defined, for all $u, v \in [0, 1]$, by

$$C(u, v) = C_3[C_1\{f_1(u), g_1(v)\}, C_2\{f_2(u), g_2(v)\}] \tag{5.12}$$

in terms of fixed copulas C_1, C_2, and C_3. Further suppose that C_3 is convex in each variable, so that if (U, V) has distribution C_3, U is stochastically decreasing

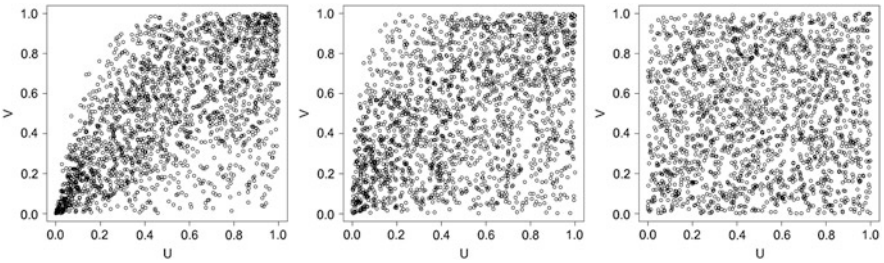

Fig. 5.7 Samples of size 2,000 from the Durante–Jaworski–Mesiar asymmetrization of the Gumbel copula with $\tau = 0.1$ (*left*), $\tau = 0.5$ (*middle*) and $\tau = 0.9$ (*right*)

in V and vice versa. Then the function C in (5.12) is a copula provided that for all $u \in [0, 1]$,

$$C_3\{f_1(u), f_2(u)\} = C_3\{g_1(u), g_2(u)\} = u.$$

Even greater generality is envisaged in [7] by relaxing the conditions on C_i, $i = 1, 2, 3$, but until rich examples, probabilistic interpretations, and simulation algorithms have been found, this approach remains somewhat of an empty vessel.

The work of Durante et al. [12] is of much more practical interest. Given a bivariate Archimedean generator φ, they define a copula C, for all $u, v \in (0, 1)$, by

$$C(u, v) = u\left[1 - \varphi^{-1}\{\varphi(1 - v)/u\}\right].$$

They prove that C is positive quadrant dependent and that it is symmetric if and only if it is the Fréchet–Hoeffding upper bound or a Clayton copula with parameter $\theta > 0$. To obtain the latter, Durante and Jaworski [8] show that one must take $\varphi(t) = \{(1 - t)^{-\theta} - 1\}^{-1/\theta}$ for all $t \in (0, 1)$. It is also easy to simulate from C as follows.

Step 1. Draw a pair (Z, V) from an Archimedean copula with generator φ.
Step 2. Return $(U, 1 - V)$, where $U = \varphi(V)/\{\varphi(Z) + \varphi(V)\}$.

An illustration of this asymmetrization technique is provided in Fig. 5.7 for three Gumbel copula generators corresponding to different degrees of dependence. The plots suggest that the degree of asymmetry decreases as τ increases. In [12], constructions leading to negative quadrant dependence are also discussed; they are obtained upon considering the copula given by $u - C(u, 1 - v) = u\varphi^{-1}\{\varphi(v)/u\}$.

Yet another approach is taken by Alfonsi and Brigo [1]. They consider absolutely continuous copulas whose density is expressed, at every $u, v \in [0, 1]$, by $c(u, v) = \ell(u - v)$ in terms of some function $\ell : [-1, 1] \to [0, \infty)$. For this construction to be valid, the authors must assume that ℓ is twice finitely integrable, that $\int_0^1 \ell(t)\mathrm{d}t = 1$,

and that $\ell(t) = \ell(t-1)$ for all $t \in [0,1]$. Letting $L(t) = \int_0^t \int_{-1}^s \ell(w)\mathrm{d}w\mathrm{d}s$, it can then be shown that C is indeed a copula and that, for all $u, v \in [0,1]$,

$$C_\ell(u,v) = L(u) + L(-v) - L(u-v).$$

This construction is of little interest, however, because if (U,V) has copula C_ℓ, the pair $(U, 1-V)$ then has a symmetric copula. Indeed, the conditions on ℓ imply that for all $t \in [0,1]$, $L(t) = L(1) + L(t-1) - (1-t)$ and hence, for all $u, v \in [0,1]$, $u - C_\ell(u, 1-v) = v - C_\ell(v, 1-u)$. Thus for C_ℓ to be a realistic copula of (X,Y), the dependence structure of $(X,-Y)$ must be symmetric. Given the wealth of such models, a statistician would be much better off analyzing the latter pair.

Finally, asymmetric copulas could also be constructed by gluing copulas or using ordinal sums; see, e.g., [35,42]. For asymmetric copulas with given diagonal section, see [6, 14, 15, 38].

5.5 Discussion

Returning to the hydrological data, it has already been seen that both the raw and de-trended data exhibit asymmetry. As it makes more sense to model pure dependence, this discussion concentrates on the analysis of the de-trended data.

While Bacigál et al. [2] consider Archimax models for these data, the hypothesis that the copula is of the simpler extreme-value form (5.5) cannot be rejected. For example, the rank-based tests developed in [3] and [29] yield approximate p-values of 96.8 % and 7.34 %, respectively.

In order to choose a suitable parametric copula family, asymmetric Galambos, Gumbel, Hüsler–Reiß and Tawn extreme-value copulas of the form (5.7) were fitted to the data by the maximum pseudo likelihood method of Genest et al. [20]. The four models returned very similar results; the highest likelihood was obtained for the Hüsler–Reiß with parameters $\hat{\theta} = 2.16$, $\hat{\lambda} = 0.938$ and $\hat{\kappa} = 1$. At 283.23, the corresponding log-likelihood is also the highest compared to the models in [2].

The parametric estimate of the Pickands dependence function of the asymmetric Hüsler–Reiß copula is shown in the left panel of Fig. 5.8. Rank-based versions of the nonparametric estimates of A due to Pickands (P) and Capéraà, Fougères, and Genest (CFG) are also displayed for comparison purposes. The right panel displays a random sample of size 659 from the fitted asymmetric Hüsler–Reiß model. Visual comparison with the rank plot of the original data suggests an adequate fit. This could be checked formally using the goodness-of-fit procedures introduced in [22].

In contrast, the hypothesis of extremeness is clearly rejected for the raw data. If one were to model the latter (i.e., deliberately ignoring the influence of time), one would need to resort to more elaborate dependence models, such as the Archimax, Liouville or Durante–Jaworski–Mesiar copulas. While these constructions clearly have potential, practical tools for their implementation remain to be developed.

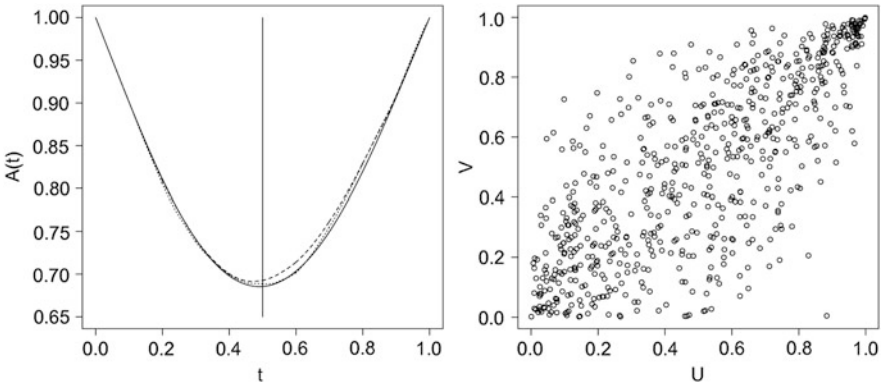

Fig. 5.8 *Left panel*: parametric Pickands dependence function estimated from the de-trended data using the asymmetric Hüsler–Reiß model (*solid line*), together with the rank-based nonparametric P (*dotted line*) and CFG (*dashed line*) estimators. *Right panel*: a sample of size 659 from the fitted asymmetric Hüsler–Reiß copula

Appendix

Proof of Proposition 5.3. To determine the limit of $\nu_p(\hat{C}_n)$, proceed as in the proof of Proposition 4 in Genest et al. [23]. Write

$$|\nu_p^p(\hat{C}_n) - \nu_p^p(C)| \le |\alpha_n| + |\beta_n|,$$

where

$$\alpha_n = \int_0^1 \int_0^1 \{\hat{C}_n(u,v) - \hat{C}_n(v,u)\}^p \, \mathrm{d}\hat{C}_n(u,v) - \int_0^1 \int_0^1 \{C(u,v) - C(v,u)\}^p \, \mathrm{d}\hat{C}_n(u,v)$$

and

$$\beta_n = \int_0^1 \int_0^1 \{C(u,v) - C(v,u)\}^p \, \mathrm{d}\hat{C}_n(u,v) - \int_0^1 \int_0^1 \{C(u,v) - C(v,u)\}^p \, \mathrm{d}C(u,v).$$

By the Mean Value Theorem, $|a^p - b^p| \le p\, 2^{p-1}|a-b|$ for all $a, b \in [0,2]$. Because $|C(u,v) - C(v,u)|$ and $|\hat{C}_n(u,v) - \hat{C}_n(v,u)|$ take values in $[0,2]$ for all $u, v \in [0,1]$,

$$\begin{aligned}
&\left| |\hat{C}_n(u,v) - \hat{C}_n(v,u)|^p - |C(u,v) - C(v,u)|^p \right| \\
&\le p\, 2^{p-1} \left| |\hat{C}_n(u,v) - \hat{C}_n(v,u)| - |C(u,v) - C(v,u)| \right| \\
&\le p\, 2^p \sup_{u,v \in [0,1]} |\hat{C}_n(u,v) - C(u,v)|,
\end{aligned}$$

where the last step follows from a twofold application of the triangular inequality. In particular, therefore,

$$|\alpha_n| \le p\, 2^p \sup_{u,v\in[0,1]} |\hat{C}_n(u,v) - C(u,v)|$$

and hence tends to 0 in probability, as $n \to \infty$. Turning to β_n, set $\psi = (C - C^\top)^p$ and apply Proposition A.1(i) in [20], taking $\delta = 1$ therein. It then follows that

$$\int_0^1 \int_0^1 \{C(u,v) - C(v,u)\}^p \, \mathrm{d}\hat{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^n \psi\left(\frac{R_i}{n}, \frac{S_i}{n}\right) \to \int_0^1 \int_0^1 \psi(u,v)\mathrm{d}C(u,v)$$

almost surely, whence $|\beta_n|$ converges to 0 in probability, as $n \to \infty$. □

Proof of Proposition 5.4. The fact that ν_p satisfies axioms (B1), (B3), and (B4) is easily seen. It is also immediate that $\nu_p(C) = 0$ if C is symmetric. To establish the converse, assume for simplicity that C has a density c. If $\nu_p(C) = 0$, then the supports of C and $C^\top$ are contained in $A = \{(u,v) : C(u,v) = C(v,u)\}$. Given that $c(u,v) = c(v,u)$ on A, it follows that

$$C(u,v) = \int_0^u \int_0^v \mathbf{1}\{(s,t) \in A\} c(s,t)\mathrm{d}t\,\mathrm{d}s = \int_0^u \int_0^v \mathbf{1}\{(s,t) \in A\} c(t,s)\mathrm{d}t\,\mathrm{d}s = C(v,u).$$

Regarding axiom (B5), write, in analogy with the proof of Proposition 5.3,

$$|\nu_p^p(C_n) - \nu_p^p(C)| \le |\alpha_n| + |\beta_n|,$$

where

$$\alpha_n = \int_0^1 \int_0^1 |C_n(u,v) - C_n(v,u)|^p \, \mathrm{d}C_n(u,v) - \int_0^1 \int_0^1 |C(u,v) - C(v,u)|^p \, \mathrm{d}C_n(u,v)$$

and

$$\beta_n = \int_0^1 \int_0^1 |C(u,v) - C(v,u)|^p \, \mathrm{d}C_n(u,v) - \int_0^1 \int_0^1 |C(u,v) - C(v,u)|^p \, \mathrm{d}C(u,v).$$

By the same argument as in the proof of Proposition 5.3,

$$|\alpha_n| \le p 2^p \sup_{u,v\in[0,1]} |C_n(u,v) - C(u,v)|$$

and hence $\alpha_n \to 0$ as $n \to \infty$. The fact that $\beta_n \to 0$ follows directly from the weak convergence of C_n to C; see, e.g., Lemma 2.2 in [45]. □

Acknowledgments Thanks are due to Tomáš Bacigál for providing the hydrological data set and his R code. The assistance of Marius Hofert and Ivan Kojadinovic with the R implementation of some of the simulation and estimation procedures used here is also gratefully acknowledged. Funding in partial support of this work was provided by the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council of Canada, and the Fonds de recherche du Québec—Nature et technologies.

References

1. Alfonsi, A., Brigo, D.: New families of copulas based on periodic functions. Comm. Stat. Theory Methods **34**, 1437–1447 (2005)
2. Bacigál, T., Jágr, V., Mesiar, R.: Non-exchangeable random variables, Archimax copulas and their fitting to real data. Kybernetika **47**, 519–531 (2011)
3. Ben Ghorbal, N., Genest, C., Nešlehová, J.: On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. Canad. J. Stat. **37**, 534–552 (2009)
4. Bücher, A., Dette, H.: A note on bootstrap approximations for the empirical copula process. Stat. Probab. Lett. **80**, 1925–1932 (2010)
5. Capéraà, P., Fougères, A.-L., Genest, C.: Bivariate distributions with given extreme value attractor. J. Multivar. Anal. **72**, 30–49 (2000)
6. De Baets, B., De Meyer, H., Mesiar, R.: Asymmetric semilinear copulas. Kybernetika **43**, 221–233 (2007)
7. Durante, F.: Construction of non-exchangeable bivariate distribution functions. Stat. Papers **50**, 383–391 (2009)
8. Durante, F., Jaworski, P.: Invariant dependence structure under univariate truncation. Statistics **46**, 263–277 (2012)
9. Durante, F., Mesiar, R.: L^∞-measure of non-exchangeability for bivariate extreme value and Archimax copulas. J. Math. Anal. Appl. **369**, 610–615 (2010)
10. Durante, F., Papini, P.L.: Componentwise concave copulas and their asymmetry. Kybernetika **45**, 1003–1011 (2009)
11. Durante, F., Papini, P.L.: Non-exchangeability of negatively dependent random variables. Metrika **71**, 139–149 (2010)
12. Durante, F., Jaworski, P., Mesiar, R.: Invariant dependence structures and Archimedean copulas. Stat. Probab. Lett. **81**, 1995–2003 (2011)
13. Durante, F., Klement, E.P., Sempi, C., Úbeda-Flores, M.: Measures of non-exchangeability for bivariate random vectors. Stat. Papers **51**, 687–699 (2010)
14. Durante, F., Kolesárová, A., Mesiar, R., Sempi, C.: Copulas with given diagonal sections: Novel constructions and applications. Int. J. Uncertain. Fuzziness Knowledge-Based Syst. **15**, 397–410 (2007)
15. Erdely, A., González-Barrios, J.M.: On the construction of families of absolutely continuous copulas with given restrictions. Comm. Stat. Theory Methods **35**, 649–659 (2006)
16. Fermanian, J.-D., Radulović, D., Wegkamp, M.H.: Weak convergence of empirical copula processes. Bernoulli **10**, 847–860 (2004)
17. Garralda-Guillem, A.I.: Structure de dépendance des lois de valeurs extrêmes bivariées. C. R. Acad. Sci. Paris Sér. I Math. **330**, 593–596 (2000)
18. Gänßler, P., Stute, W.: Seminar on Empirical Processes. Birkhäuser, Basel (1987)
19. Genest, C., Nešlehová, J.: Copula modeling for extremes. In: El-Shaarawi, A.H., Piegorsch, W.W. (eds.) Encyclopedia of Environmetrics, vol. 2, pp. 530–541, 2nd edn. Wiley, Chichester (2012)
20. Genest, C., Ghoudi, K., Rivest, L.-P.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika **82**, 543–552 (1995)

21. Genest, C., Ghoudi, K., Rivest, L.-P.: Comment on the paper by E.W. Frees and E.A. Valdez entitled Understanding relationships using copulas. N. Am. Act. J. **2**, 143–149 (1998)
22. Genest, C., Kojadinovic, I., Nešlehová, J., Yan, J.: A goodness-of-fit test for extreme-value copulas. Bernoulli **17**, 253–275 (2011)
23. Genest, C., Nešlehová, J., Quessy, J.-F.: Tests of symmetry for bivariate copulas. Ann. Inst. Stat. Math. **64**, 811–834 (2012)
24. Ghoudi, K., Khoudraji, A., Rivest, L.-P.: Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. Canad. J. Stat. **26**, 187–197 (1998)
25. Hofert, M., Mächler, M., McNeil, A.J.: Likelihood inference for Archimedean copulas in high dimensions under known margins. J. Multivar. Anal. **110**, 133–150 (2012)
26. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
27. Khoudraji, A.: Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivariées. Doctoral Dissertation, Université Laval, Québec, Canada (1995)
28. Klement, E.P., Mesiar, R.: How non-symmetric can a copula be? Comment. Math. Univ. Carolin. **47**, 141–148 (2006)
29. Kojadinovic, I., Yan, J.: Nonparametric rank-based tests of bivariate extreme-value dependence. J. Multivar. Anal. **101**, 2234–2249 (2010)
30. Liebscher, E.: Construction of asymmetric multivariate copulas. J. Multivar. Anal. **99**, 2234–2250 (2008)
31. Liebscher, E.: Erratum to Construction of asymmetric multivariate copulas. J. Multivar. Anal. **102**, 869–870 (2011) [J. Multivar. Anal. **99** 2234–2250 (2008)]
32. McNeil, A.J., Nešlehová, J.: Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. Ann. Stat. **37**, 3059–3097 (2009)
33. McNeil, A.J., Nešlehová, J.: From Archimedean to Liouville copulas. J. Multivar. Anal. **101**, 1772–1790 (2010)
34. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, Princeton (2005)
35. Mesiar, R., Jágr, V., Juráňová, M., Komorníková, M.: Univariate conditioning of copulas. Kybernetika **44**, 807–816 (2008)
36. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
37. Nelsen, R.B.: Extremes of nonexchangeability. Stat. Papers **48**, 329–336 (2007)
38. Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A., Úbeda-Flores, M.: On the construction of copulas and quasi-copulas with given diagonal sections. Insur. Math. Econom. **42**, 473–483 (2008)
39. Patton, A.J.: On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. J. Financ. Econom. **2**, 130–168 (2004)
40. Rüschendorf, L.: Asymptotic distributions of multivariate rank order statistics. Ann. Stat. **4**, 912–923 (1976)
41. Segers, J.: Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. Bernoulli **18**, 764–782 (2012)
42. Siburg, K.F., Stoimenov, P.A.: Gluing copulas. Comm. Stat. Theory Methods **37**, 3124–3134 (2008)
43. Siburg, K.F., Stoimenov, P.A.: Symmetry of functions and exchangeability of random variables. Stat. Papers **52**, 1–15 (2011)
44. Tsukahara, H.: Semiparametric estimation in copula models. Canad. J. Stat. **33**, 357–375 (2005)
45. van der Vaart, A.W.: Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (1998)
46. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes, with Applications to Statistics. Springer, New York (1996)
47. Vaz de Melo Mendez, B.: Asymmetric extreme interdependence in emerging equity markets. Appl. Stochas. Models Bus. Ind. **21**, 483–498 (2005)
48. Vorosmarty, C.J., Fekete, B.M., Tucker, B.A.: Global River Discharge, 1807–1991, V. 1.1 (RivDIS). Data set. Available on-line (http://www.daac.ornl.gov) from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN, USA (1998)

Chapter 6
Modeling Time-Varying Dependencies Between Positive-Valued High-Frequency Time Series

Nikolaus Hautsch, Ostap Okhrin, and Alexander Ristig

Abstract Multiplicative error models (MEM) became a standard tool for modeling conditional durations of intraday transactions, realized volatilities, and trading volumes. The parametric estimation of the corresponding multivariate model, the so-called vector MEM (VMEM), requires a specification of the joint error term distribution, which is due to the lack of multivariate distribution functions on $\mathbb{R}_+^d$ defined via a copula. Maximum likelihood estimation is based on the assumption of constant copula parameters and therefore leads to invalid inference if the dependence exhibits time variations or structural breaks. Hence, we suggest to test for time-varying dependence by calibrating a time-varying copula model and to re-estimate the VMEM based on identified intervals of homogenous dependence. This paper summarizes the important aspects of (V)MEM, its estimation, and a sequential test for changes in the dependence structure. The techniques are applied in an empirical example.

6.1 Multiplicative Error Models

Multiplicative Error Models (MEMs) are frequently applied to describe autocorrelated positive-valued processes. The multiplicative structure became popular in the context of (G)ARCH models, see [2, 8]. Engle and Russell [9] adopted this multiplicative approach to analyze the conditional duration of irregularly spaced financial transaction data under the assumption that the error term follows an

N. Hautsch · O. Okhrin · A. Ristig (✉)
School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: nikolaus.hautsch@wiwi.hu-berlin.de; ostap.okhrin@hu-berlin.de; ristigal@hu-berlin.de

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_6,

Exponential or Weibull distribution. This extends directly to MEMs, when other positive-valued random variables such as trading volumes, are considered. As stressed by [9], a joint model including volumes, transaction prices, and time variations in liquidity gives a better understanding of the fundamental mechanisms of stock markets than individual univariate analyses.

6.1.1 Univariate MEM

Let x_i be a nonnegative univariate time series, with time index $i = 1, \ldots, n$. The univariate MEM is defined as

$$\begin{aligned} x_i &= \mu_i \varepsilon_i \\ \mu_i &\stackrel{\text{def}}{=} \mathrm{E}\left(x_i | \mathscr{F}_{i-1}; \xi\right), \end{aligned} \tag{6.1}$$

where ξ denotes an m-dimensional vector of parameters. Furthermore, assume that ε_i follows an iid process with $\mathrm{E}(\varepsilon_i) = 1$ and density $f(\cdot)$. The conditional mean can be specified in several ways, e.g.,

$$\mu_i = \omega + \sum_{j=1}^{P} \alpha_j x_{i-j} + \sum_{j=1}^{Q} \beta_j \mu_{i-j}, \tag{6.2}$$

with $\omega \geq 0, \alpha_j \geq 0$ and $\beta_j \geq 0, \ \forall\, j, \xi = \left(\omega, \alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_Q\right)^{\top}$. Based on the filters $\phi(L) = \sum_{j=1}^{R} \phi_j L^j = \sum_{j=1}^{R} \left(\alpha_j + \beta_j\right) L^j$, $\beta(L) = \sum_{j=1}^{Q} \beta_j L^j$ and the martingale difference series $\eta_i = x_i - \mu_i$, (6.2) can be transformed to an ARMA(R, Q) model

$$x_i = \omega + \phi(L) x_i + \{1 - \beta(L)\} \eta_i, \tag{6.3}$$

where $R = \max(P, Q)$ and L denotes the lag operator with $L^j x_i = x_{i-j}$. According to standard time series arguments, (6.3) is guaranteed to be weakly stationary, if $\sum_{j=1}^{P} \alpha_j + \sum_{j=1}^{Q} \beta_j < 1$. Given the above set of assumptions, we implicitly assume an exponential decay of the autocorrelation function $\rho(\cdot)$, i.e., $\lim_{l \to \infty} \sum_{j=-l}^{l} |\rho(j)| < \infty$. However, in case of financial high-frequency data this assumption is often not fulfilled.

As such data typically reveal long memory, we provide a short review of the fractionally integrated MEM (FIMEM), which allows the autocorrelation function of the underlying random variable to decay hyperbolically. Formally, x_i exhibits long memory if $\lim_{l \to \infty} \sum_{j=-l}^{l} |\rho(j)| = \infty$. Following [1], Jasiak [14] specifies the FIMEM in the context of conditional durations by introducing the fractional difference operator $(1 - L)^{\delta}$ to (6.3), i.e.,

$$\{1-\phi(L)\}(1-L)^{\delta}x_i = \omega + \{1-\beta(L)\}\eta_i, \tag{6.4}$$

with $\delta \in [0,1]$ the fractional integration parameter. Hosking [13] defines the fractional difference operator by a binomial series:

$$(1-L)^{\delta} = \sum_{j=0}^{\infty}\binom{\delta}{j}(-1)^j L^j = \sum_{j=0}^{\infty}\pi_j L^j. \tag{6.5}$$

The FIMEM generalizes the integrated MEM, since it permits the degree of differencing to be a fractional value of the integrated case, for which $\delta = 1$.

Substituting the martingale difference series defined above in (6.4) leads to

$$\begin{aligned}\{1-\beta(L)\}\mu_i &= \omega + \left[1-\beta(L)-\{1-\phi(L)\}(1-L)^{\delta}\right]x_i \\ \mu_i &= \omega\{1-\beta(1)\}^{-1} + \lambda(L)x_i,\end{aligned} \tag{6.6}$$

where the linear filter $\lambda(L) = 1-\{1-\phi(L)\}\{1-\beta(L)\}^{-1}(1-L)^{\delta} = \sum_{j=1}^{\infty}\lambda_j L^j$ implies an infinite number of parameter restrictions to its coefficients to guarantee the nonnegativity of μ_i, i.e., $\lambda_j \geq 0$ for $j = 1,2,\ldots$. As a consequence, in practice the filter $\lambda(L)$ is truncated to a finite number of lags and one needs to apply Theorem 3 of [6] to verify that the combination of parameters of the FIMEM$(P;\delta;Q)$ is within the feasible parameter space. To emphasize this point, consider the following two extreme examples for which we assume that δ lies within the unit interval, such that $\pi_j < 0$, for $j > 0$. Then, (1) μ_i can become negative although all parameters are greater than zero and (2) μ_i can be positive almost surely for all i, even though all parameters except δ are negative. Note that these restrictions play a fundamental role for the validity of forecasts.

The first unconditional moment of x_i is not defined since the fractional difference operator evaluated at $L = 1$ equals zero. As a result, the FIMEM is not covariance stationary. If the parameters are nonnegative and $\sum_{j=1}^{P}\alpha_j + \sum_{j=1}^{Q}\beta_j < 1$, then the strict stationarity and ergodicity of the FIMEM can be deduced from the stationarity and ergodicity of the integrated MEM, since the infinite-order representation of (6.6) is dominated in an absolute value sense by the coefficients of the corresponding integrated MEM, cf. [1, 3]. Alternative covariance stationary long memory MEMs are discussed in [12].

In general, parametric ML estimation of univariate MEMs leads to asymptotically efficient and unbiased estimates if the distribution of the innovations ε_i is specified correctly. Typical candidates to describe ε_i are the standard Exponential or Weibull distribution, but flexible distributions as the generalized Gamma or F distribution can also be considered. In a standard ML framework for time series models, where $\ell_i(\xi), i = 1,\ldots,n$, denotes the ith contribution to the log likelihood $\ell(\xi) = \sum_{i=1}^{n}\ell_i(\xi)$, $H_n(\xi) = \sum_{i=1}^{n}\{\frac{\partial^2}{\partial\xi\partial\xi^{\top}}\ell_i(\xi)\}$ denotes the Hessian matrix

and $S_n(\xi) = \sum_{i=1}^{n} \{\frac{\partial}{\partial \xi} \ell_i(\xi) \frac{\partial}{\partial \xi^\top} \ell_i(\xi)\}$ the outer product of scores, the limiting distribution of the estimator $\hat{\xi}$ is given by

$$\left\{H_n(\xi)^{-1} S_n(\xi) H_n(\xi)^{-1}\right\}^{-1/2} \sqrt{n}(\hat{\xi} - \xi) \xrightarrow{\mathscr{L}} N\left(0_m, I_m\right), \tag{6.7}$$

with identity matrix I_m. Statistical inference is based on the finite sample approximation of (6.7), i.e., the Hessian matrix and the outer score product are replaced by the consistent estimates $H_n(\hat{\xi})$ and $S_n(\hat{\xi})$.

Furthermore, [9] adopts the asymptotic theory of [16] and proposes a quasi-ML setup which leads to consistent estimates for the linear and integrated MEM even if the true error term distribution does not correspond to the assumed standard Exponential distribution. In this case, $\hat{\xi}$ converges under weak regularity conditions to the asymptotic distribution of (6.7).

6.1.2 Vector MEM

Cipollini and Gallo [5] formalizes the VMEM as

$$x_i = \mu_i \odot \varepsilon_i, \tag{6.8}$$

where $\odot$ denotes the Hadamard product and $x_i = (x_{i1}, \ldots, x_{id})^\top$, $i = 1, \ldots, n$, is the vector of positive-valued processes. The multivariate scale factor $\mu_i \overset{\text{def}}{=} \mathrm{E}(x_i | \mathscr{F}_{i-1}; \xi)$ and the vector of error terms ε_i are $(d \times 1)$ vectors. The natural multivariate extension of (6.6) is given by

$$[I_d - B(L)]\, \mu_i = \omega + [I_d - B(L) - \{I_d - \Phi(L)\} D]\, x_i, \tag{6.9}$$

with $\Phi(L) = A(L) + B(L)$ and A, B being $(d \times d)$ matrices. Short-run effects enter equation (6.9) through the linear filters $A(L)$ and $B(L)$ and ω denotes the vector of constants. The univariate fractional difference operator from (6.6) extends to the diagonal matrix $\mathrm{diag}(D) = \{(1-L)^{\delta_1}, \ldots, (1-L)^{\delta_d}\}$, which contains the individual fractional difference operators, with $\delta_j \in [0, 1]$, $j = 1, \ldots, d$. By this restriction, we exclude deterministic low frequency patterns between the marginal time series. Note that the individual mean equations of (6.9) collapse to the univariate FIMEM (6.6), if A and B are diagonal and to the linearly parameterized MEM (6.2), if additionally $\delta_j = 0$, $j = 1, \ldots, d$. Based on the diagonality assumption for A and B the model can be estimated equation by equation and is stationary.

For the full parametric specification of the VMEM we need to define an innovation process ε_i, $i = 1, \ldots, n$, which must follow a distribution with only positive probabilities on $\mathbb{R}_+^d = [0, \infty)^d$ and $\mathrm{E}(\varepsilon_{ij}) = 1$, $j = 1, \ldots, d$. However,

the distribution function of a univariate error term process does not have a natural multivariate equivalent. Therefore, the d marginal distributions are coupled together with a copula splitting a multivariate distribution function into its margins and a pure dependence component—the copula. Copulae are introduced in [23] stating that if F is an arbitrary d-dimensional continuous distribution function of the random variables $X_1, \ldots, X_d$, then the associated copula is unique and defined as a continuous function $C : [0,1]^d \to [0,1]$ which satisfies the equality

$$C(u_1, \ldots, u_d) = F\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\}, \quad u_1, \ldots, u_d \in [0,1], \tag{6.10}$$

where $F_1^{-1}(\cdot), \ldots, F_d^{-1}(\cdot)$ are the quantile functions of the continuous marginal distribution functions $F_1(x_1), \ldots, F_d(x_d)$. Based on the copula density $c(\cdot, \ldots, \cdot; \theta)$ and the marginal densities $f_j(\cdot, \alpha_j)$ of ε_{ij}, $j = 1, \ldots, d$, the log likelihood of the VMEM can be written as

$$\begin{aligned} \ell(\theta, \xi, \alpha | \mathscr{F}_{i-1}) = & \sum_{i=1}^{n} \sum_{j=1}^{d} \left[\log \left[\varepsilon_{ij}(\xi) f_j \left\{ \varepsilon_{ij}(\xi); \alpha_j \right\} \right] - \log x_{ij} \right] \\ & + \sum_{i=1}^{n} \log c \left[F_1 \left\{ \varepsilon_{i1}(\xi), \alpha_1 \right\}, \ldots, F_1 \left\{ \varepsilon_{id}(\xi), \alpha_d \right\}; \theta \right], \end{aligned} \tag{6.11}$$

with $x_i / \mu_i(\xi) | \mathscr{F}_{i-1} = \varepsilon_i(\xi) | \mathscr{F}_{i-1} \sim C \left[F_1 \left\{ \varepsilon_{i1}(\xi), \alpha_1 \right\}, \ldots, F_1 \left\{ \varepsilon_{id}(\xi), \alpha_d \right\}; \theta \right]$ having expectation one, where θ denotes the copula-, α the marginal- and ξ the mean-parameters, cf. [5]. Conversely to the Hadamard product, x_i / μ_i denotes element-wise division. The efficient approach to obtain parameter estimates is given by full ML estimation, as the multivariate density function is assumed to be known, i.e., the product of the marginal densities multiplied with the copula density. On the other hand, full ML estimation is difficult to implement even if the induced dependence is non-elliptical. For example, if we assume a Vine- or hierarchical Archimedean copula (HAC), (see Sect. 6.2), the copula density varies with the structure of the underlying copula. Thus, the log likelihood must be optimized for each possible structure and the parameter vector generating the largest log likelihood value is selected as ML estimate. To avoid this computationally intensive method, a two-step procedure similar to [4] can be straightforwardly applied, since (6.11) can be decomposed into a marginal and a copula part as follows: First, the parameters of the mean equation are estimated to filter the residuals, for which only the information about the marginal distributions is used. Then, the copula is calibrated to the fitted values of the residuals' empirical distribution functions.

Similar to classical risk management applications, where several time-varying models for correlations and copulae are proposed, e.g., [7, 22], time-varying dependence cannot be excluded in our context and consequently, the copula estimated at the second step may contain time variations. Yet, the final target of VMEMs is not to predict, e.g., tail dependencies or risk measures, but to produce forecasts of μ_i, which crucially depend on precise parameter estimates and thus on the complete log

likelihood and the most recent data for which the dependence between the variables is constant. Thus, we suggest to reestimate the parameters of μ_i by maximizing (6.11) with fixed θ for time intervals at which the copula model calibrated at the second step supports constant dependence.

6.2 Hierarchical Archimedean Copulae

Among other important families, there exists the class of Archimedean copulae (AC), which (1) permits modeling non-elliptical dependencies, (2) can describe different types of tail dependencies and (3) has a closed form expression. Formally, AC are defined through the generator function $\phi_\theta \in \mathfrak{L} = \{\phi_\theta : [0;\infty) \to [0,1] \,|\, \phi_\theta(0) = 1, \phi_\theta(\infty) = 0; (-1)^i \phi_\theta^{(i)} \geq 0; i \in \mathbb{N}\}$ and $(-1)^i \phi_\theta^{(i)}(x)$ being non-decreasing and convex on $[0,\infty)$, for $x > 0$, which commonly depends on a single parameter θ, i.e.,

$$C(u_1,\ldots,u_d;\theta) = \phi_\theta\left\{\phi_\theta^{-1}(u_1) + \cdots + \phi_\theta^{-1}(u_d)\right\}, \quad u_1,\ldots,u_d \in [0,1]. \quad (6.12)$$

Properties of Archimedean copulae are reviewed and investigated in [15,18]. Nelsen [19] discusses generator families depending on two parameters. The restricted dependence structure induced by Archimedean generators is the major disadvantage of d-dimensional ACs since this assumption is mostly violated in practice.

To permit more flexibility, arguments of an AC can be replaced by further ACs leading to the concept of HAC, which can adopt arbitrary complicated structures denoted by s in the following. The generators of a single HAC, ϕ_j, can come from different generator families. However, if the ϕ_{θ_j}'s come from the same family, the required complete monotonicity of $\phi_{\theta_{j+1}}^{-1} \circ \phi_{\theta_j}$ imposes constraints on the parameters $\theta_1,\ldots,\theta_{d-1}$. The flexibility induced by the structure is accompanied by larger amounts of parameters, as each generator composition corresponds to one additional parameter. Sufficient conditions on the generator functions guaranteeing that C is a copula are stated in [17]. It holds that if $\phi_{\theta_j} \in \mathfrak{L}$, for $j = 1,\ldots,d-1$, and $\phi_{\theta_{j+1}}^{-1} \circ \phi_{\theta_j}$ have completely monotone derivatives, then C is a copula for $d \geq 2$. The major advantage of HACs compared to ACs is the non-exchangeability of the arguments beyond a single node, which is imposed by the structure of a HAC. Similar to the dependence parameters, s is generally unknown and can be regarded as an additional parameter to estimate.

A sequential estimation procedure for HACs is discussed by [20] providing statistical inference for parametric and nonparametric estimated margins. The procedure uses Proposition 1 of [21] stating that HACs can be uniquely reconstructed from marginal distributions and bivariate copula functions. The estimation procedure can be summarized in the following way: at the first step, estimate all binary copula parameters of a specified Archimedean family under the assumption of known marginal distribution functions. Select the largest parameter and fix the

binary copula as pseudo-variable. At next steps, assume the estimated margins and sub-copulae from lower levels are known and estimate all binary copula parameters by considering pairs of margins, pairs of pseudo variables, and pairs of margins and pseudo variables. Then, choose the largest parameter and fix the corresponding copula as a pseudo-variable. This procedure leads a binary approximation of an arbitrary HAC. Let $\varepsilon_i = \{\varepsilon_{i1}, \ldots, \varepsilon_{id}\}^\top$ be the sample, $i = 1, \ldots, n$, and $\theta = (\theta_1, \ldots, \theta_{d-1})^\top$ be the copula parameters ordered from the lowest to the highest hierarchical level. The multi-stage ML-estimator, $\hat{\theta}$, provides a solution for the following system of equations

$$\left(\frac{\partial \ell_1}{\partial \theta_1}, \ldots, \frac{\partial \ell_{d-1}}{\partial \theta_{d-1}} \right)^\top = \mathbf{0}, \tag{6.13}$$

$$\text{where} \quad \ell_j = \sum_{i=1}^{n} l_j(\varepsilon_i), \text{ for } j = 1, \ldots, d-1,$$

$$l_j(\varepsilon_i) = \log \left\{ c\left[\{\hat{F}_m(\varepsilon_{im})\}_{m \in s_j};\, s_j, \theta_j \right] \prod_{m \in s_j} \hat{f}_m(\varepsilon_{im}) \right\}$$

$$\text{for } i = 1, \ldots, n,$$

where s_j contains the indices, which are structured according to the fixed subcopulae (and margins) at lower hierarchical levels.

6.3 Change Point Detection

The time intervals for which the parameters of μ_i should be reestimated are identified by calibrating a time-varying copula. In this context, [11] proposes a framework, which incorporates time-varying HAC parameters θ_i and s_i and is closely related to the local change point (LCP) procedure applied in [24]. As a detailed discussion of this sophisticated method is beyond the scope of this paper, this section describes only the main ideas of the data driven LCP.

Let θ_i, s_i be the unknown time-varying parameters and structure of the HAC C. Let $I = [i_0 - m, i_0]$ denote an interval with reference point i_0, $m > 0$ and let $\Delta_I(\theta, s) = \sum_{i \in I} \mathscr{K}\{c(\cdot, \theta_i, s_i), c(\cdot; \theta, s)\}$ be a random quantity, which measures the quality of the approximation of the true copula with time-varying parameters $c(\cdot, \theta_i, s_i)$ by the (local) parametric copula $c(\cdot; \theta, s)$, where $\mathscr{K}(\cdot, \cdot)$ denotes the Kullback-Leibler divergence. Furthermore, let $\Delta_I(\theta, s) \leq \Delta$ be the small modeling bias (SMB) condition with $\Delta \geq 0$ and constant parameters θ, s. As $\mathscr{K}(\cdot, \cdot)$ measures the discrepancy between two densities, from the SMB condition follows that the data generating process can be well approximated by the local constant copula $C(\cdot; \theta, s)$ on I. In this sense [11] proposes testing whether a HAC

with time-varying parameters and structure can be locally described by a HAC with constant parameters and structure.

Under the null hypothesis assume that the SMB condition holds for interval I and parameters $\{\theta, s\}$ and define the set of possible change points $\mathscr{T}_I$ for interval I, which is tested for a single but unknown change point $\tau \in \mathscr{T}_I$. The test hypotheses are formalized as

$$H_0: \ \forall\, \tau \in \mathscr{T}_I, \theta_i = \theta, s_i = s, \forall\, i \in I = J \cup J^C = [\tau, i_0] \cup [i_0 - m, \tau) \tag{6.14}$$

$$\begin{aligned} H_1: \ & \exists\, \tau \in \mathscr{T}_I, \theta_i = \theta_1, s_i = s_1, \forall\, i \in J = [\tau, i_0], \\ & \text{and } \theta_i = \theta_2 \neq \theta_1 \text{ or } s_i = s_2 \neq s_1, \forall\, i \in J^C = [i_0 - m, \tau). \end{aligned}$$

The null hypothesis is rejected, if the likelihood ratio (LR) test statistic

$$T_I = \max_{\tau \in \mathscr{T}_I} \left[\max_{\theta_1, s_1} \{\ell_J(\theta_1, s_1)\} + \max_{\theta_2, s_2} \{\ell_{J^C}(\theta_2, s_2)\} - \max_{\theta, s} \{\ell_I(\theta, s)\} \right], \tag{6.15}$$

exceeds the critical value $\mathfrak{z}_I$. In practice, the length of the homogenous interval and the parameters of interest $\{\theta, s\}$ are estimated simultaneously due to their relation through the test statistic. For a well performing choice of the critical value, which is found via a Monte-Carlo simulation from the local parametric model and implicitly defines the significance level of the test, we refer to [24].

6.4 Empirical Analysis

The considered time span of NASDAQ trade data for Apple (AAPL) starts on January 2nd and ends on December 31st, 2009. Similar to the cleaning of TAQ data sets as, e.g., applied in [12], all non-executed trades, trades with a price smaller or equal to zero and outliers are removed from the tick-by-tick high-frequency data set. To overcome the phenomenon of simultaneous observations, trades with the same time stamp are merged and the corresponding values are aggregated by their median. A cleaned tick-by-tick data set provides information about (1) the price series p_j, (2) the trading volume v_j and (3) the time stamp of the trades t_j, $j = 1, \ldots, n^*$, where n^* is the number of daily observations. To investigate the relationships between these series, we construct the series of high-low ranges (HL), average volumes (Vol), and the number of trades (NT) on a sampling frequency of 10 min, i.e.,

$$\begin{aligned} \mathrm{HL}_i &= \max\left\{p_j | t_j \in (t_{i-1}, t_i]\right\} - \min\left\{p_j | t_j \in (t_{i-1}, t_i]\right\}, \\ \mathrm{NT}_i &= \#\left\{t_j | t_j \in (t_{i-1}, t_i]\right\}, \\ \mathrm{Vol}_i &= \mathrm{NT}_i^{-1} \sum_{t_j \in (t_{i-1}, t_i]} v_j, \end{aligned} \tag{6.16}$$

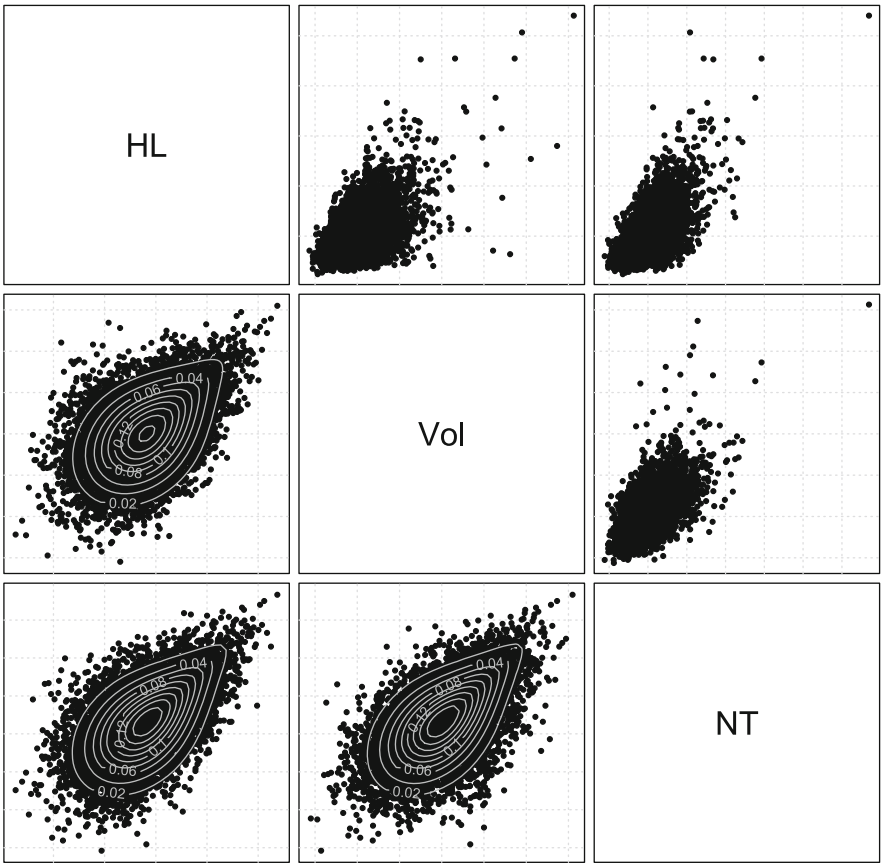

Fig. 6.1 The *upper diagonal elements* show the pairwise dependence between the filtered residuals. The *lower diagonal elements* present the values of the standard normal quantile applied to the values of the empirical distribution functions. Scales of the axes are not presented as they differ slightly. The origins of the coordinate planes of the upper diagonal elements correspond to zero

for $i = 1, \ldots, n$, where # counts the elements of the set $\{\cdot\}$. Note that other proxies for price variations, e.g., the 10 min realized volatility or the squared returns, can replace the high-low range.

To remove the U-shaped daily seasonal pattern provided by the variables defined above, the individual seasonal components are approximated by fitting cubic splines and each series is divided by the respective estimated seasonal factor. Then, model (6.8) with mean (6.9) is calibrated to the process, where $A(L)$ and $B(L)$ are restricted to be diagonal and to the first lag. The infinite sums of the mean equations of the FIMEMs are truncated to 400 lagged coefficients, i.e., $\sum_{l_j=0}^{400} \pi_{l_j} L^{l_j}$, since the parameters ξ_j are almost unaffected by including additional π_{l_j}'s, $j = 1, \ldots, d$. Despite these restrictions, the estimated models produce uncorrelated residuals. Figure 6.1 presents scatterplots of the filtered residuals. The lower diagonal elements

of Fig. 6.1 do not reveal elliptical dependencies, thus the Gaussian copula is inappropriate in this case. In the following, we prefer an approximation of the dependence structure by the hierarchical or simple Archimedean Gumbel copula, since the bivariate contour plots indicate almost the same dependencies as the underlying scatterplots.

The approach proposed in Sect. 6.3 considers only one single interval I, whose subintervals, defined through the set of possible change points $\mathscr{T}_I$, are tested for homogeneity. This method turns out to be time-varying, when it is applied as a sequential testing procedure. For this purpose, define the set $\mathscr{I}$, which contains the geometrically growing sequence of nested interval-candidates $I_0 \subset I_1 \subset \ldots \subset I_k \subset \ldots \subset I_K$, with $I_k = [i_0 - m_k, i_0]$, reference point i_0, geometric grid $m_k = [1.25^k m_0]$, and the sets of possible change points $\mathscr{T}_{I_k} = [i_0 - m_{k-1}, i_0 - m_{k-2}]$ for all $I_k \in \mathscr{I}$. $[x]$ means the integer part of x and $m_0 = 40$. If the null hypothesis of constant dependence is not rejected for interval I_k, the interval length is extended and interval I_{k+1} is tested for homogeneity. This procedure is continued until a change point is identified or the largest interval I_K is accepted as interval of homogeneity. If a change point is detected at $k + 1$, the local adaptive estimates are given by $\hat{\theta} = \tilde{\theta}_k$, $\hat{s} = \tilde{s}_k$, where $\tilde{\theta}_k, \tilde{s}_k$ denote the ML-estimates from Sect. 6.2. In practice, the HAC is tested for homogeneity at all points of the sample except the points of a "burn in" period, such that $i_0 = m_{K+1}, \ldots, n$. While other time-varying methods permit only the parameter(s) to vary over time, the structure of this time-varying HAC may change as well.

Based on the Gumbel family, we apply the LCP procedure to the filtered residuals, because an application of the LCP procedure to the full VMEM is cumbersome due to the large number of parameters. The first panel of Fig. 6.2 shows the changing HAC-structure estimated for an accepted interval of homogeneity, whose length is shown in the fourth panel. The two thick solid lines (gray and black) in the second panel present the time-varying parameters in terms of Kendall's $\hat{\tau}$. For the relationship between bivariate Archimedean generators and Kendall's τ, see [10]. Based on these results, we propose to reestimate the parameters of the VMEM's scale function μ_i for at least three intervals separated by the dashed vertical lines, using full ML with fixed copula parameters. The first interval ending in the middle of March can be clearly identified, as the structure is constant and the estimates of Kendall's τ exhibit a certain distance. Furthermore, the steadily increasing interval lengths support our choice of an homogenous HAC for this interval, which is given by $s^1_{\text{HAC}} = ((\text{NT Vol})_{1.92}\text{HL})_{1.50}$, where the subscript is related to $\hat{\theta}$. The second interval is characterized by an alternating structure, while the values of Kendall's $\hat{\tau}$ can almost be distinguished. This makes it, from our perspective, difficult, to decide, whether a HAC or a simple AC should be used for reestimating the VMEM. In general, the corresponding HAC, $s^2_{\text{HAC}} = ((\text{NT Vol})_{1.67}\text{HL})_{1.42}$, indicate a weaker dependence than the fitted HAC of the first interval. The simple AC is given by $s^2_{\text{AC}} = (\text{NT Vol HL})_{1.48}$. In the third selected interval beginning in June, the underlying copula corresponds with high probability to a simple AC, since the structure changes frequently and both parameters are very close to each other, such that the HAC can be aggregated to a simple AC at most

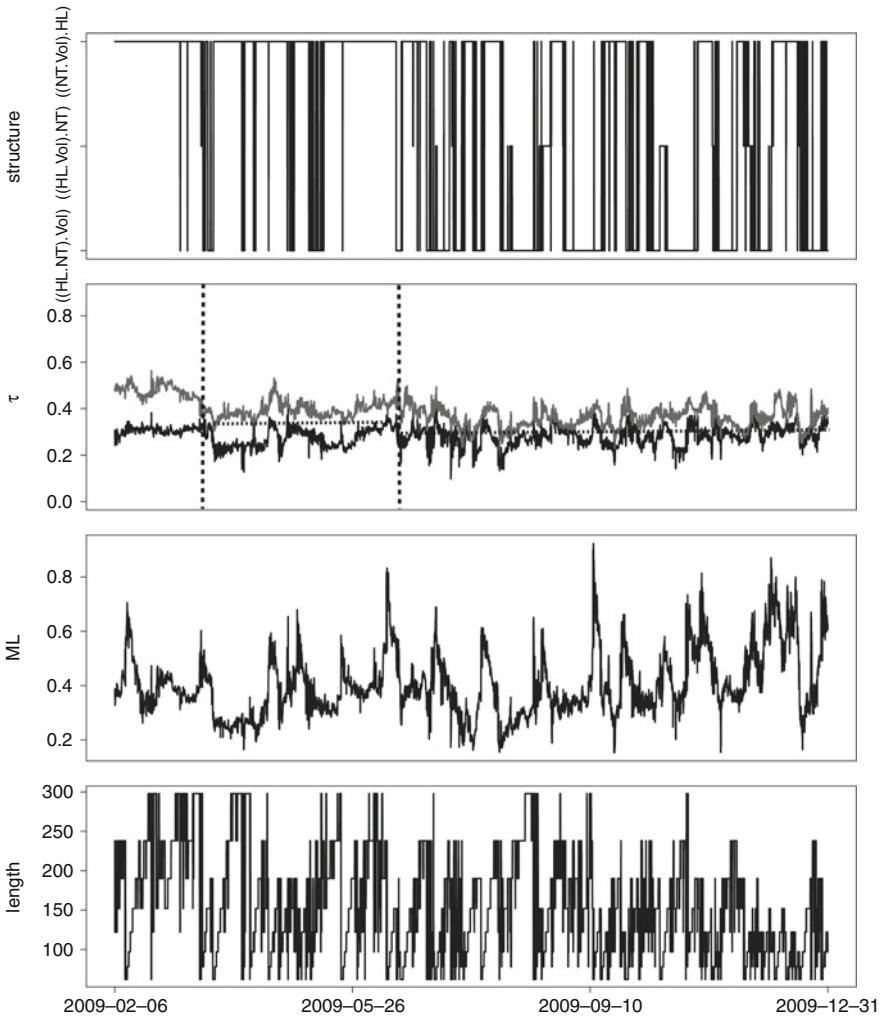

Fig. 6.2 Results of the LCP-procedure of AAPL. The *first panel* shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the ML process for the intervals of homogeneity, whose varying length is presented in the *lower panel*

points of the interval. The HAC of this interval, $s^3_{\text{HAC}} = ((\text{NT HL})_{1.52}\text{Vol})_{1.40}$, shows a different structure and the AC, $s^3_{\text{AC}} = (\text{NT Vol HL})_{1.42}$, a weaker dependence than the calibrated copulas of the first and second interval. We admit, at this point, that shorter interval specifications are possible, as the method provides a sensitive picture of the time-varying dependence. Note that shorter time intervals are accompanied with less data and therefore imply a loss in efficiency. The estimated HAC based on the entire sample is given by $s_{\text{HAC}} = ((\text{Vol NT})_{1.56}\text{HL})_{1.41}$ and the respective simple AC by $s_{\text{AC}} = (\text{NT HL Vol})_{1.45}$. We investigated the time-varying dependence for a few of other stocks and found similar results.

The third and fourth panels illustrate the performance of the LCP procedure. As proposed in Sect. 6.3, the LR test statistic measures the stability of the fitted model. Therefore, the length of the accepted intervals increase continuously in periods of a stable fit, whereas the interval length is typically short if the ML process is volatile. The dynamic of the ML process is presented in the third picture and allows to reproduce this relationship. The ML process exhibits a higher volatility in the last two months of the observed sample. This implies shorter intervals, for which the hypotheses of homogeneity are accepted, since the LR test statistics are smaller. Härdle et al. [11] illustrates in a simulation study that the procedure detects dependence changes with a short delay and [24] investigates the quality of the local adaptive estimators. A simple alternative approach is the rolling window method, which also allows for time-varying parameters but detects changes in the dependence with a larger delay.

Acknowledgments The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 Ökonomisches Risiko, Humboldt-Universität zu Berlin is gratefully acknowledged.

References

1. Baillie, R.T., Bollerslev, T., Mikkelsen, H.O.: Fractionally integrated generalized autoregressive conditional heteroskedasticity. J. Econometrics **74**, 3–30 (1996)
2. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econometrics **31**(3), 307–327 (1986)
3. Bougerol, P., Picard, N.: Stationarity of GARCH processes and of some nonnegative time series. J. Econometrics **52**(1–2), 115–127 (1992)
4. Chen, X., Fan, Y.: Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. J. Econometrics **135**(1–2), 125–154 (2006)
5. Cipollini, F., Gallo, G.M.: Automated variable selection in vector multiplicative error models. Comput. Stat. Data Anal. **54**(11), 2470–2486 (2010)
6. Conrad, C., Haag, B.R.: Inequality constraints in the fractionally integrated GARCH model. J. Financ. Econometrics **4**(3), 413–449 (2006)
7. Engle, R.: Dynamic conditional correlation: a simple class of multivariate GARCH models. J. Bus. Econ. Stat. **20**(3), 339–350 (2002)
8. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica **50**(4), 987–1007 (1982)
9. Engle, R.F., Russell, J.R.: Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica **66**(5), 1127–1162 (1998)
10. Genest, C., Rivest, L.P.: Statistical inference procedures for bivariate Archimedean copulas. J. Am. Stat. Assoc. **88**, 1034–1043 (1993)
11. Härdle, W.K., Okhrin, O., Okhrin, Y.: Time varying hierarchical Archimedean copulae. SFB 649 Discussion Paper 2010, 018, SFB 649, Economic Risk, Berlin (2010)
12. Hautsch, N.: Econometrics of Financial High-Frequency Data. Springer, Berlin (2012)
13. Hosking, J.R.M.: Fractional differencing. Biometrika **68**(1), 165–176 (1981)
14. Jasiak, J.: Persistence in intertrade durations. Finance **19**, 166–195 (1998)
15. Joe, H.: Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In: Rüschendorf, L., Schweizer, B., Taylor, M. (eds.) Distribution with Fixed Marginals and Related Topics. IMS Lecture Notes – Monograph Series. Institute of Mathematical Statistics, Hayward (1996)

16. Lee, S., Hansen, B.: Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. Economet. Theor. **10**, 29–52 (1994)
17. McNeil, A.J.: Sampling nested archimedean copulas. J. Stat. Comput. Simulat. **78**, 567–581 (2008)
18. McNeil, A.J., Nešlehová, J.: Multivariate Archimedean copulas, d-monotone functions and l_1 norm symmetric distributions. Ann. Stat. **37**(5b), 3059–3097 (2009)
19. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
20. Okhrin, O, Okhrin, Y, Schmid, W.: On the structure and estimation of hierarchical Archimedean copulas. J. Econom. **173**, 189–204 (2013)
21. Okhrin, O., Okhrin, Y., Schmid, W.: Properties of hierarchical Archimedean copulas. Stat. Risk Model. **30**(1), 21–54 (2013)
22. Patton, A.J.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**, 527–556 (2006)
23. Sklar, A.: Fonctions de répartition à n dimension et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 299–231 (1959)
24. Spokoiny, V.: Multiscale local change point detection with applications to value-at-risk. Ann. Stat. **37**(3), 1405–1436 (2009)

Chapter 7
The Limiting Properties of Copulas Under Univariate Conditioning

Piotr Jaworski

Abstract The dynamics of univariate conditioning of copulas with respect to the first variable is studied. Special attention is paid to the limiting properties when the first variable is attaining extreme values. We describe the copulas which are invariant with respect to the conditioning and study their sets of attraction. Furthermore we provide examples of the limit sets consisting of more than one element and discuss the chaotic nature of univariate conditioning.

7.1 Introduction

The interest in the construction of multivariate stochastic models describing the dependence among several variables has grown in the last years. In particular, the recent financial crisis emphasized the necessity of considering models that can serve to estimate better the occurrence of extremal events (see, for example, [3, 6, 45] or [19]).

In financial and actuarial risk management, the construction of appropriate models for dependence between risks is of obvious importance, due to the well recognized fact that neglecting dependence gives rise to a dramatic risk underestimation.

In risk management one often deals with a situation when there is one leading line of insurance or one leading asset in an investment portfolio. The crucial point is to determine what might happen with other assets/lines when a large loss caused by this leading one occurs, i.e. to what extent diversification could hedge the aggregated outcome.

P. Jaworski (✉)
Institute of Mathematics, Warsaw University, Warsaw, Poland
e-mail: P.Jaworski@mimuw.edu.pl

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_7,

Following the distributional approach proposed in [39, 40], the concept of *conditional copula*[1] has proved to be useful for the description of dependence among random variables, when we condition observations to lie above or below some thresholds. See also [1, 9, 10, 14, 26, 48, 52].

We recall that, given a vector $(X_1, X_2, \ldots, X_n)$ of continuous real-valued random variables, a conditional copula is the copula associated with the conditional distribution function of $(X_1, X_2, \ldots, X_n)$ knowing that $X_i, i = 1, \ldots, n$, are subject to some restrictions.

The goal of this paper is to study the behaviour of conditional copulas for extremal values of the first variable. Our main objectives are the limiting properties of families of copulas $C_{[\alpha]}$ satisfying the following property:

$C_{[1]}$ is the copula of a random vector $(X_1, X_2, \ldots, X_n)$.
$C_{[\alpha]}, \alpha \in (0, 1)$, is the copula of $(X_1, X_2, \ldots, X_n)$ supposing that $X_1 < q_{X_1}(\alpha)$.

The motivation for dealing with such families is closely connected with the ongoing research in the area of modelling in finance and reliability.

Market Contagion. Let X and Y model the returns from the indices of two stock markets. It is a well acknowledged stylized fact that the dependence of Y and X subject to the condition $X \leq q_X(\alpha)$ is greater than the dependence in the middle part of the distribution of Y and X. This phenomenon was studied, for example, in [8, 9, 16, 34].

Contagion in a Financial Network. We fix a time instance, say $t = 1$. Let C be a copula of joint distribution of welfares W_1^i of $n + 1$ banks/institutions at this time. We study the case when the first one is in trouble.

Reliability. Let us consider a complex system consisting of k components. Let T_i denote the lifetime of the ith component, T the lifetime of the system, and $T_{min} = \min(T_1, \ldots, T_k)$ the time of the first failure. Often we have to model the dependence between T and T_{min} subject to the condition $T_{min} \geq t^*$.

Credit Risk. Let us consider a pool of n homogeneous loans. Let T_i be the time of the default of ith debtor,

$$T_i = \inf\{t > 0 : X_{i,t} \leq 0\},$$

where $X_i = (X_{i,t})_{t \geq 0}$ is the wealth process of the ith debtor. If the wealth processes depend rather on systematic or systemic factors than on idiosyncratic (specific) ones, then one may expect that the dependence between $T_{k_1:n}$ and $T_{k_2:n}$, $k_1 < k_2 < n$, subject to the condition $T_{k_1:n} \geq t^*$ is increasing when t^* is increasing. Of course the most interesting case is when k_1/n and k_2/n are the attachment and detachment points of a CDO (collateralized debt obligation) tranche.

[1] Sometimes in some special cases the conditional copula is called tail-dependence or threshold copula.

The paper is organized as follows. Section 7.2 presents some facts that are necessary in order to fix the notation and make the paper self-contained. In the next two Sects. 7.3 and 7.4, the basic facts about univariate conditioning are provided. In Sect. 7.5 we discuss the existence of the limit for extremal conditionings. A characterization of all copulas that are invariant under univariate conditioning and some basic examples are presented in Sect. 7.6. Section 7.7 deals with the rates of convergence to the limit copula for selected families of copulas. Sections 7.8–7.10 illustrate the "strange" and "chaotic" behaviour of univariate conditioning. We start from the existence of periodic, toroidal and dense trajectories and finish with the butterfly effect. With a few exceptions, the proofs are collected in Sect. 7.11.

7.2 Preliminaries About Copulas

A *copula* is the restriction to the unit n-cube $[0,1]^n$ of a distribution function whose univariate margins are uniformly distributed on $[0,1]$. Specifically, a function $C : [0,1]^n \to [0,1]$ is a copula if it has the following properties: for every $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$ such that $0 \le u_i \le v_i \le 1$ for $i = 1, \ldots, n$,

(C1) $(\exists i \;\; u_i = 0) \Rightarrow C(\mathbf{u}) = 0$,
(C2) $\forall j \in \{1, \ldots, n\} \;\; (\forall i \neq j \;\; u_i = 1) \Rightarrow C(\mathbf{u}) = u_j$,
(C3) C is *n-non-decreasing*, that is, the C-volume $V_C(\mathbf{u}, \mathbf{v})$ of any n-rectangle with lower vertex $\mathbf{u}$ and upper vertex $\mathbf{v}$ is non-negative.

We recall that the C-volume is a signed sum of the values of C at the vertices of the n-rectangle,

$$V_C(\mathbf{u}, \mathbf{v}) = C(\mathbf{w})|_{w_1=u_1}^{v_1} \ldots |_{w_n=u_n}^{v_n} = \sum_{j_1=1}^{2} \cdots \sum_{j_n=1}^{2} (-1)^{j_1+\cdots+j_n} C(w_{1,j_1}, \ldots, w_{n,j_n}),$$

where $w_{i,1} = u_i$ and $w_{i,2} = v_i$.

We equip the set $\mathscr{C}^n$ of all n-variate copulas with the natural metric of uniform convergence,

$$d(C_1, C_2) = \sup\{|C_1(\mathbf{x}) - C_2(\mathbf{x})| : \mathbf{x} \in [0,1]^n\},$$

and obtain a complete, compact, separable, convex, metric space.

By the celebrated *Sklar's Theorem*, the joint distribution function F of any n-tuple $\mathbf{X} = (X_1, \ldots, X_n)$ of random variables defined on the probability space $(\Omega, \mathscr{F}, \mathbf{P})$ can be written as a composition of a copula C and the univariate marginals F_i, i.e. for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_n(x_n)).$$

Moreover, if X_i are continuous random variables, then the copula C is uniquely determined. For more details about copula theory and (some of) its applications, we refer to [4, 11, 17, 22, 32, 36, 37, 45, 49].

7.3 Univariate Conditioning of Copulas

We recall the basic facts concerning univariate conditioning (truncation, thresholding) of random variables and copulas. Without loss of generality we restrict ourselves to lower conditioning of the first variable. Indeed, formulas for conditioning of other variables may be obtained by a permutation of variables, while formulas for upper conditioning by changing the sign of the conditioned variable (compare [24]).

Special attention will be paid to conditioning of Archimedean copulas and ordinal sums of copulas. Furthermore we show that certain operations on copulas commute with conditioning.

7.3.1 Basics

To keep the notation consistent we introduce the following definition of the conditional copula $C_{[\alpha]}$.

Definition 7.1. Let an n-variate copula C be the joint distribution function of random variables $U_1, \ldots, U_n$ which are uniformly distributed on the unit interval. For every $\alpha \in (0, 1]$ we denote by $C_{[\alpha]}$ the copula of the conditional distribution of $U_1, \ldots, U_n$ with respect to the condition $U_1 \leq \alpha$.

By the Sklar's Theorem, $C_{[\alpha]}$ admits the following characterization (compare [31]):

Proposition 7.1. *Let C be an n-variate copula. For every $\alpha \in (0, 1]$ the conditional copula $C_{[\alpha]}$ is a unique solution of the equation*

$$C_{[\alpha]}\left(x_1, \frac{C_{1,2}(\alpha, x_2)}{\alpha}, \ldots, \frac{C_{1,n}(\alpha, x_n)}{\alpha}\right) = \frac{C(\alpha x_1, x_2, \ldots, x_n)}{\alpha}, \quad \mathbf{x} \in [0, 1]^n, \tag{7.1}$$

where $C_{1,i}$, $i = 2, \ldots, n$, are bivariate marginal copulas obtained by substituting $x_j = 1$ for $j \neq 1, i$,

$$C_{1,i}(x_1, x_i) = C(x_1, 1, \ldots, 1, x_i, 1, \ldots, 1).$$

In Definition 7.1 one can replace the uniformly distributed random variables $\mathbf{U}$ by any n-tuple of random variables having continuous distribution functions. Indeed:

Proposition 7.2 ([31]). *If the copula C describes the interdependencies between n random variables $X_1, \ldots, X_n$, then $C_{[\alpha]}$ is the copula of the conditional distribution of $X_1, \ldots, X_n$ with respect to the condition $X_1 \leq q$, where $\boldsymbol{P}(X_1 \leq q) = \alpha$.*

Note that even if the distribution function of one random variable X_i is discontinuous, the choice of the copula C is not unique. The same may occur for conditional copulas.

7.3.2 Conditional Archimedean Copulas

We recall that the n-variate copula C is called Archimedean if there exist generators ψ and φ such that

$$C(x_1, \ldots, x_n) = \psi(\varphi(x_1) + \cdots + \varphi(x_n)).$$

The generators are convex non-increasing functions

$$\psi : [0, \infty] \longrightarrow [0, 1], \quad \varphi : [0, 1] \longrightarrow [0, \infty],$$

such that

$$\psi(0) = 1, \quad \varphi(1) = 0 \quad \text{and} \quad \forall t \in [0, 1] \quad \psi(\varphi(t)) = t.$$

Note that for any constant $c > 0$ the pair of generators $\psi(ct)$ and $c^{-1}\varphi(t)$ is equivalent to the pair $\psi(t), \varphi(t)$. For more details, the reader is referred to [44, 49].

Univariate conditioning preserves the Archimedeanity (compare [48] Remark 2(iii)).

Theorem 7.1. *If C is an Archimedean copula with generators ψ and φ, then $C_{[\alpha]}$ is an Archimedean copula with generators $\psi_{[\alpha]}$ and $\varphi_{[\alpha]}$, where*

$$\varphi_{[\alpha]}(t) = \varphi(\alpha t) - \varphi(\alpha), \tag{7.2}$$

$$\psi_{[\alpha]}(t) = \frac{1}{\alpha}\psi(t + \varphi(\alpha)). \tag{7.3}$$

7.3.3 Conditional Ordinal Sums of Copulas

In the subsequent sections we denote by m the Lebesgue measure on the real line, and by F_U the cumulative distribution function of the random variable U uniformly distributed on the unit interval,

$$F_U(t) = \min(1, \max(0, t)) = \begin{cases} 0 & \text{for} \quad t < 0, \\ t & \text{for} \ 0 \le t \le 1, \\ 1 & \text{for} \quad 1 < t. \end{cases}$$

The ordinal sum of bivariate copulas is a classical concept (compare [49], 3.2.2) and was recently extended to n-variate copulas with $n \ge 2$ [7, 29, 35, 47].

Definition 7.2. The *ordinal sum* of n-variate copulas $\{C_k\}_{k=1}^N$ ($N \in \mathbb{N} \cup \{+\infty\}$) with respect to non-overlapping intervals $\{(a_k, b_k)\}_{k=1}^N$ ($0 \le a_k < b_k \le 1$) is the function $G : [0,1]^n \to [0,1]$ given by

$$G(x_1, \ldots, x_n) = \sum_{k=1}^{N} (b_k - a_k) C_k \left(F_U \left(\frac{x_1 - a_k}{b_k - a_k} \right), \ldots, F_U \left(\frac{x_n - a_k}{b_k - a_k} \right) \right) + m \left([0, \min(x_1, \ldots, x_n)] \setminus \bigcup_{k=1}^{N} [a_k, b_k] \right). \tag{7.4}$$

Note that for any copulas C_k and disjoint intervals the function G given by (7.4) is a copula.

We state two properties of ordinal sums which are crucial for the study of the dynamics of univariate conditioning; the proofs are in Sect. 7.11.

Theorem 7.2. *Let G be the ordinal sum of n-variate copulas $\{C_k\}_{k=1}^N$ with respect to non-overlapping intervals $\{(a_k, b_k)\}_{k=1}^N$. If $b_1 = 1$ and $a_1 > 0$, then:*

1. $d(G, C_1) \le (n+1)a_1$.
2. *$G_{[a_1]}$ is the ordinal sum of the copulas $\{C_{k+1}\}_{k=1}^{N-1}$ with respect to the intervals $\{(a_{k+1}/a_1, b_{k+1}/a_1)\}_{k=1}^{N-1}$.*

7.3.4 Operations on Copulas Which Commute with Univariate Conditioning

There are several operations on copulas which commute with conditioning of the first variable. For example the permutation or flipping of other variables, vertical gluing or generalized product of copulas.

Proposition 7.3. *Let C and D be two n-variate copulas such that*

$$D(x_1, \ldots, x_n) = C(x_1, x_{\sigma(2)}, \ldots x_{\sigma(n)}),$$

where $\sigma : \{2, \ldots, n\} \to \{2, \ldots, n\}$ is a permutation. Then for every $\alpha \in (0, 1]$,

$$D_{[\alpha]}(x_1, \ldots, x_n) = C_{[\alpha]}(x_1, x_{\sigma(2)}, \ldots x_{\sigma(n)}).$$

The change of signs of random variables, so-called flipping, induces a transformation of their copulas.

Definition 7.3. Let $\mathbf{v} = (v_1, \ldots, v_n)$, $v_i = 0, 1$, be a vertex of the unit cube. The associated dual copula

$$\widehat{C_{\mathbf{v}}} : [0,1]^n \longrightarrow [0,1]$$

is defined by

$$\widehat{C_{\mathbf{v}}}(\mathbf{x}) = V_C(\mathbf{u}, \mathbf{w}),$$

where the coordinates of $\mathbf{u}$ and $\mathbf{w}$ depend on the coordinates of $\mathbf{v}$ and $\mathbf{x}$ in the following way:

$$(u_i, w_i) = \begin{cases} (0, x_i) & \text{if } v_i = 0, \\ (1 - x_i, 1) & \text{if } v_i = 1. \end{cases}$$

Note that if C describes the joint distribution of random variables $X_1, \ldots, X_n$, then $\widehat{C_{\mathbf{v}}}$ does the same for $(-1)^{v_1} X_1, \ldots, (-1)^{v_n} X_n$.

Proposition 7.4. *Let C and D be two n-variate copulas such that*

$$D = \widehat{C_{\mathbf{v}}},$$

where $\mathbf{v} = (0, v_2, \ldots, v_n)$ is a vertex of the unit cube $[0,1]^n$ such that $v_1 = 0$. Then for every $\alpha \in (0, 1]$,

$$D_{[\alpha]} = \widehat{(C_{[\alpha]})_{\mathbf{v}}}. \tag{7.5}$$

Remark 7.1. If $v_1 = 1$, then in formula (7.5) one has to replace the conditional copula $C_{[\alpha]}$ by the conditional copula $C^{up}_{[\alpha]}$ obtained by upper conditioning of the first variable.[2]

The ordinal sums have been generalized in order to take into account possible asymmetries; for bivariate copulas the following construction method was presented in [48].

Definition 7.4. The *vertical gluing ordinal sum* of bivariate copulas $\{C_k\}_{k=1}^N$ ($N \in \mathbb{N} \cup \{+\infty\}$) with respect to non-overlapping intervals $\{(a_k, b_k)\}_{k=1}^N$ ($0 \le a_k < b_k \le 1$) is the function $G : [0,1]^2 \to [0,1]$ given by

[2] For basic facts concerning the univariate upper conditioning of random variables and copulas, see Section 3 of [24].

$$G(x_1, x_2) = \sum_{k=1}^{N} (b_k - a_k) C_k \left(x_1, F_U \left(\frac{x_2 - a_k}{b_k - a_k} \right) \right) + m \left([0, x_2] \setminus \bigcup_{k=1}^{N} [a_k, b_k] \right) x_1. \tag{7.6}$$

Note that the function G given by (7.6) is always a copula. Geometrically speaking, G is obtained by piecing together different copulas on slices of the unit multi-cube obtained by a partition of the vertical axis.

Proposition 7.5. *If G is the vertical gluing ordinal sum of copulas $(C_k)_{k\in\mathscr{I}}$ with respect to intervals $((a_k, b_k))_{k\in\mathscr{I}}$, then for every $\alpha \in (0, 1]$ the conditional copula $G_{[\alpha]}$ is the i-gluing ordinal sum of the conditional copulas $\left((C_k)_{[\alpha]}\right)_{k\in\mathscr{I}}$ with respect to the same collection $((a_k, b_k))_{k\in\mathscr{I}}$.*

This follows from [48, Th. 2].

Definition 7.5. Let C_i, $i = 1, \ldots, n$, be bivariate copulas and C_0 an n-variate one. The function $D : [0, 1]^{n+1} \to [0, 1]$ given by the formula

$$D(x, y_1, \ldots, y_n) = \int_0^x C_0 \left(\frac{\partial}{\partial \xi} C_1(\xi, y_1), \ldots, \frac{\partial}{\partial \xi} C_n(\xi, y_n) \right) d\xi \tag{7.7}$$

is called the *generalized product* of the copulas $C_1, \ldots, C_n$ induced by the copula C_0.

Note that copulas are Lipschitz functions (see [49] Theorem 2.10.7). Therefore, when we fix all but one variable we get functions which are absolutely continuous and differentiable almost everywhere on the unit interval. Furthermore the derivatives are measurable, positive and bounded by 1. Hence the integral in (7.7) is well defined.

Proposition 7.6. *For any n-variate copula C_0 and bivariate copulas C_i, $i = 1, \ldots, n$, the function D given by (7.7) is an $(n + 1)$-variate copula. Moreover the $(1, i + 1)$-marginal copula of D is equal to C_i*

$$D(x, 1, \ldots, 1, y_i, 1, \ldots, 1) = C_i(x, y_i).$$

For $n = 2$ formula (7.7) was widely used to construct copulas with given 2-margins (see, for example, [12, 13, 53]). Furthermore, if copulas C_0, C_1 and C_2 are absolutely continuous, then D is a vine 3-copula (compare [5]). For $n > 2$ we have the following link with vine copulas:

Proposition 7.7. *If C_0 is an n-variate vine copula and $C_1, \ldots, C_n$ are absolutely continuous bivariate copulas, then the copula D given by (7.7) is a vine copula. Moreover D is represented by a sequence $(T_i)_{i=1}^n$ of n trees with T_1 having the first node of degree n and n nodes of degree 1 and the other T_i's being the same as in the representation of C_0.*

Proposition 7.8. *If D is a generalized product of copulas $C_1, \ldots, C_n$ induced by a copula C_0, then for every $\alpha \in (0,1]$ the conditional copula $D_{[\alpha]}$ is the generalized product of the conditional copulas $(C_1)_{[\alpha]}, \ldots, (C_n)_{[\alpha]}$ induced by the same copula C_0.*

The proofs of Propositions 7.6 and 7.8 can be obtained as a slight generalization of the proof of Theorem 4.1 in [31].

7.4 Univariate Conditioning as a Dynamical System

The substitution

$$\alpha = \exp(-t), \quad t \in [1, \infty),$$

allows us to consider the conditioning as a topological dynamical system. Indeed:

Theorem 7.3. *The mapping*

$$Cond : \mathscr{C}^n \times (0,1] \longrightarrow \mathscr{C}^n, \quad Cond(C, \alpha) = C_{[\alpha]},$$

is continuous. Moreover, for any $\alpha, \beta \in (0,1]$ and $C \in \mathscr{C}$,

$$Cond(C,1) = C \quad \text{and} \quad Cond(Cond(C,\alpha),\beta) = Cond(C, \alpha\beta). \tag{7.8}$$

The proof is postponed to Sect. 7.11.

Hence the conditioning is a semi-flow, i.e. a continuous action of the multiplicative semigroup $(0,1]$ on the space of copulas (compare [20, Definitions 1.1.1 and 1.1.2]). In Sect. 7.10 we show that it is Devaney chaotic (compare [20, Definition 1.7.5]).

We adapt the definition of an invariant element, a periodic trajectory and a limit point from dynamical systems to conditioning of copulas (compare [20]).

Definition 7.6. An n-variate copula C is called invariant under the lower conditioning (truncation, thresholding) of the first variable if for every $\alpha \in (0,1]$,

$$C_{[\alpha]} = C.$$

Definition 7.7. An n-variate copula C is called log-periodic under the lower conditioning (truncation, thresholding) of the first variable if there exists a constant $0 < k < 1$, called a *period*, such that for every $\alpha \in (0,1]$,

$$C_{[k\alpha]} = C_{[\alpha]}.$$

As a consequence of the iteration formula (7.8) we get:

Proposition 7.9 ([26, Theorem 1]). *If*

$$\lim_{\alpha\to 0^+} C_{[\alpha]} = C_{[0]},$$

then the limit copula $C_{[0]}$ is invariant.

Proposition 7.10. *If for a copula C and some constant k, $0 < k < 1$,*

$$C_{[k]} = C,$$

then C is log-periodic with period k.

Proof. For any $\alpha \in (0, 1]$ we get

$$C_{[k\alpha]} = (C_{[k]})_{[\alpha]} = C_{[\alpha]}.$$

□

Definition 7.8. Let C and D be n-variate copulas. D is called a limit point of C under the lower conditioning of the first variable if there exists a sequence $(\alpha_k)_{k=1}^{\infty}$, $\alpha_k \in (0, 1]$, such that

$$\lim_{k\to\infty} \alpha_k = 0 \quad \text{and} \quad \lim_{k\to\infty} C_{[\alpha_k]} = D.$$

The set of all limit points of a copula C is called the limit set and denoted by $LimSet(C)$.

Due to the Ascoli Theorem limit sets are non-empty. Furthermore, since conditioning is continuous, the limit set of a copula is closed and connected. In the following sections we will discuss the simplest examples of limit sets:

- Consisting of one invariant copula,
- Consisting of the orbit of a log-periodic copula,
- Being the closure of a toroidal orbit,
- Coinciding with the whole space $\mathscr{C}^n$.

7.5 Existence of the Limit

As we will show in Sect. 7.10, the set of copulas for which the limit of the conditioning exists is a union of countably many nowhere dense subsets of $\mathscr{C}$. Below we show that it is also dense and discuss two important families of not necessarily invariant copulas for which the limit does exist.

7.5.1 Density

We start with the following observation:

Lemma 7.1. *The set of attraction of every n-variate invariant copula C is dense in $\mathscr{C}^n$, i.e.*

$$\forall \varepsilon > 0 \quad \forall B \in \mathscr{C}^n \quad \exists D \in \mathscr{C}^n \quad d(B, D) < \varepsilon, \quad \lim_{\alpha \to 0^+} D_{[\alpha]} = C.$$

Indeed, by Theorem 7.2, we can take as D the ordinal sum of the copulas B and C with respect to the intervals $(0, \frac{\varepsilon}{n+2})$ and $(\frac{\varepsilon}{n+2}, 1)$. Moreover, since sets of attraction of different invariant copulas are disjoint, we get:

Corollary 7.1. *The complement of the set of attraction of any n-variate invariant copula C is dense in $\mathscr{C}^n$.*

Thus every such set of attraction has an empty interior and should not be called a domain.

7.5.2 Conditionally Monotonic Copulas

Definition 7.9. A copula C is called non-decreasing (resp. non-increasing) under the lower conditioning of the first variable if for every $\alpha_1, \alpha_2 \in (0, 1]$, $\alpha_1 \geq \alpha_2$ (resp. $\alpha_1 \leq \alpha_2$)

$$C_{[\alpha_1]}(\mathbf{x}) \leq C_{[\alpha_2]}(\mathbf{x}).$$

Lemma 7.2. *If C is conditionally non-decreasing (resp. non-increasing) when $\alpha \to 0$, then there exists a copula $C_{[0]}$ such that*

$$\lim_{\alpha \to 0^+} C_{[\alpha]}(\mathbf{x}) = C_{[0]}(\mathbf{x}).$$

7.5.3 Copulas with Non-trivial Tail Expansions of Degree 1

We recall the definition of tail expansions (compare [27, 28, 30, 38, 50]).

Definition 7.10. We say that a copula C has a tail expansion of degree 1 at the vertex $\mathbf{v}$ of the unit cube if the limit

$$\lim_{t \to 0^+} \frac{\widehat{C_{\mathbf{v}}}(tx_1, \ldots, tx_n)}{t}$$

exists for all non-negative $x_1, \ldots, x_n$.

The function

$$L_{\mathbf{v}} : [0,\infty]^n \longrightarrow [0,\infty), \quad L_{\mathbf{v}}(\mathbf{x}) = \lim_{t\to 0} \frac{\widehat{C_{\mathbf{v}}}(t\mathbf{x})}{t},$$

is called the dependence function or the leading term of the expansion. Note that the expansion of a copula C at a vertex $\mathbf{v}$ is equal to the expansion of the dual copula $\widehat{C_{\mathbf{v}}}$ at the origin. For bivariate $(1, j)$-margins of a copula C we apply the following notation. When the expansions exist we put

$$L_{j,0}(x_1, x_j) = \lim_{t\to 0+} \frac{C_{1,j}(tx_1, tx_j)}{t},$$

$$L_{j,1}(x_1, x_j) = \lim_{t\to 0+} \frac{\widehat{(C_{1,j})_{(0,1)}}(tx_1, tx_j)}{t} = \lim_{t\to 0+} \frac{tx_1 - C_{1,j}(tx_1, 1-tx_j)}{t}.$$

We start with the case when the whole "tail" probability mass is concentrated close to one vertex $\mathbf{v}$, $\mathbf{v} = (0, v_2, \ldots, v_n)$.

Theorem 7.4. *Let $L_{\mathbf{v}}$ and L_{j,v_j} be the leading terms of the expansion of a copula C at a vertex $\mathbf{v}$ and its bivariate $(1, j)$-marginals at $(0, v_j)$. If*

$$\lim_{y\to\infty} L_{\mathbf{v}}(1, y, \ldots y) = 1,$$

then the limit of $C_{[\alpha]}$ when $\alpha \to 0^+$ exists and the limit copula $C_{[0]}$ is given by

$$\widehat{(C_{[0]})_{\mathbf{v}}}(x_1, L_{2,v_2}(1, x_2), \ldots, L_{n,v_n}(1, x_n)) = L_{\mathbf{v}}(x_1, x_2, \ldots, x_n).$$

The case when the whole "tail" probability mass is concentrated close to several vertices $\mathbf{v}$ with first coordinate 0 is a bit more complicated.

Theorem 7.5. *Let $L_{\mathbf{v}}$ and L_{j,e_j} be the leading terms of the expansions of a copula C at a vertex $\mathbf{v}$ and of its bivariate $(1, j)$-marginals at $(0, e_j)$, $e_j = 0, 1$. If for $j = 2, \ldots, n$,*

$$\lim_{y\to\infty} L_{j,0}(1, y) + \lim_{y\to\infty} L_{j,1}(1, y) = 1,$$

then the limit of $C_{[\alpha]}$ when $\alpha \to 0^+$ exists and the limit copula $C_{[0]}$ is determined by

$$\widehat{(C_{[0]})_{\mathbf{v}}}(x_1, L_{2,v_2}(1, x_2), \ldots, L_{n,v_n}(1, x_n)) = L_{\mathbf{v}}(x_1, x_2, \ldots, x_n),$$

where $\mathbf{v}$ ranges over the set of all vertices with $v_1 = 0$.

The proofs of both theorems are in Sect. 7.11.

7.6 Invariant Copulas

A characterization of invariant copulas was given in [10] (the bivariate case) and [31] (the higher dimensions). We recall the constructions.

7.6.1 Bivariate Case

Let $f : [0, +\infty] \to [0, 1]$ be a surjective, monotonic function and g its right inverse ($f(g(y)) = y$). We denote by C_f the function

$$C_f : [0,1]^2 \longrightarrow [0,1], \qquad C_f(x,y) = \begin{cases} 0 & \text{for } x = 0, \\ xf\left(\frac{g(y)}{x}\right) & \text{for } x > 0. \end{cases} \tag{7.9}$$

Proposition 7.11 ([10]). *A bivariate copula C is invariant if and only if either*

- $C(x, y) = \Pi(x, y) = xy$, *or*
- $C = C_f$, *where f is surjective, non-decreasing and concave, or*
- $C = C_f$, *where f is surjective, non-increasing and convex, or*
- C *is a vertical gluing ordinal sum of copulas of the above types.*

Note that the third case is a flipping of the second one.[3] Moreover these cases can be distinguished in terms of the concordance ordering of copulas.

Lemma 7.3 ([10]).

1. *If f is non-decreasing and concave, then the function C_f is a PQD copula, and moreover*
$$\forall (x, y) \in (0, 1)^2 \quad C_f(x, y) > \Pi(x, y) = xy.$$
2. *If f is non-increasing and convex, then the function C_f is a NQD copula, and moreover*
$$\forall (x, y) \in (0, 1)^2 \quad C_f(x, y) < \Pi(x, y) = xy.$$

7.6.1.1 Examples of Invariant Bivariate Copulas

Example 7.1. Copula of independent random variables: $\Pi(x, y) = xy$.

Example 7.2. Copula of comonotonic random variables:

$$M(x, y) = \min(x, y), \quad f(t) = \min(1, t).$$

[3] Copulas C_f are known as Durante–Jaworski–Mesiar copulas[18, Section 5.4.4].

Example 7.3. Copula of anticomonotonic random variables:

$$W(x,y) = \max(0, x+y-1), \quad f(t) = \max(0, 1-t).$$

Example 7.4. Clayton copula with positive parameter θ:

$$Cl_\theta(x,y) = (x^{-\theta} + y^{-\theta} - 1)^{-1/\theta}, \quad f(t) = (1+t^{-\theta})^{-1/\theta}.$$

Example 7.5. Clayton copula with negative parameter θ, $\theta \in (-1, 0)$:

$$Cl_\theta(x,y) = (\max(0, x^{-\theta} + y^{-\theta} - 1))^{-1/\theta}, \qquad f(t) = (\max(0, 1-t^{-\theta}))^{-1/\theta}.$$

Example 7.6. Marshall–Olkin copula with parameters $(\theta, 1)$, $\theta \in (0,1)$:

$$MO(x,y) = \min(1-x, yx^{1-\theta}), \qquad f(t) = \min(1, t^\theta).$$

Example 7.7. Let C be a copula associated with the distribution function

$$F(x,y) = \begin{cases} 0 & \text{for } x < 0 \ \vee \ y < 0, \\ L(x,y) & \text{for } 0 \le x \le 1 \ \wedge \ y \ge 0, \\ L(1,y) & \text{for } x > 1 \ \wedge \ y \ge 0, \end{cases}$$

where L is homogeneous:

$$\forall \alpha \in [0,1] \quad L(\alpha x, \alpha y) = \alpha L(x,y).$$

Then $C = C_f$ with $f(t) = L(1,t)$.

7.6.1.2 Similarities Between Invariant and Archimedean Bivariate Copulas

Let $A_f(u,v) = f(g(u) + g(v))$ be an Archimedean copula and $C_f(x,y) = xf(g(y)/x)$ an invariant copula with the same convex generator f. As was shown in [15], they are closely related to each other.

Theorem 7.6 ([15]).

(a) Let (U,V) be a pair of continuous random variables distributed according to A_f. Then C_f is the distribution function of the random pair (X,Y), where almost surely

$$X = \frac{g(V)}{g(U) + g(V)}, \qquad Y = V. \tag{7.10}$$

(b) Let (X, Y) be a pair of continuous random variables distributed according to C_f. Then A_f is the distribution function of the random pair (U, V), where almost surely

$$U = f\left(\frac{g(Y)}{X} - g(Y)\right), \qquad V = Y. \tag{7.11}$$

Transformation (7.10) composed with the flipping, $Y \to 1 - Y$, is known as Durante–Jaworski–Mesiar asymmetrization of Archimedean copulas (see [18] Figure 5.7).

7.6.2 Multivariate Case

The procedure of obtaining a characterization of the multivariate copulas that are invariant under univariate truncation is based on the generalized product (see Definition 7.5) and resembles Sklar's Theorem.

Theorem 7.7 ([31]). *If C is the generalized product of invariant copulas $C_1, \ldots, C_n$ induced by a copula C_0, then it is invariant. Moreover the $(1, i+1)$-marginal copula of C is equal to C_i, $i = 1, \ldots, n$.*

Theorem 7.8 ([31]). *For every invariant $(n+1)$-variate copula C, there exists an n-variate copula C_0 such that C is the generalized product of its bivariate marginal copulas $C_{1,2}, \ldots, C_{1,n+1}$ induced by C_0.*

7.6.2.1 Examples of Invariant Multivariate Copulas

Example 7.8. The comonotonic copula.
If an invariant copula C has a non-trivial singular component, then the copula C_0 need not be unique. Indeed, let C_0 be any n-variate copula. Then

$$\int_0^x C_0\left(\mathbb{I}_{\xi \le y_1}, \ldots, \mathbb{I}_{\xi \le y_n}\right) d\xi = M_{n+1}(x, y_1, \ldots, y_n),$$

where $M_{n+1}(x, y_1, \ldots, y_n) = \min(x, y_1, \ldots, y_n)$ is the copula of comonotonic random variables.

Example 7.9. Dual of the comonotonic copula.
Let C_0 be any n-variate copula. Then

$$\begin{aligned}\int_0^x C_0\left(\mathbb{I}_{\xi + y_1 \ge 1}, \ldots, \mathbb{I}_{\xi + y_n \ge 1}\right) d\xi &= M_n(\mathbf{y}) - M_{n+1}(1 - x, y_1, \ldots, y_n) \\ &= \widehat{M_{n+1}}_{(1,0,\ldots,0)}(x, y_1, \ldots, y_n).\end{aligned}$$

Example 7.10. The independence copula.
For the independence copula Π_{n+1} the copula C_0 is also the independence copula:

$$\int_0^x \Pi_n(y_1,\dots,y_n)d\xi = \int_0^x y_1\cdot\dots\cdot y_n d\xi = xy_1\cdot\dots\cdot y_n = \Pi_{n+1}(x,y_1,\dots,y_n).$$

A similar result holds for Clayton copulas.

Example 7.11. The family of Clayton copulas with positive parameter.
Let $C_{n,\gamma}$ denote the n-variate Clayton copula with parameter $\gamma > 0$:

$$C_{n,\gamma}(u) = \left(u_1^{-\gamma} + \dots + u_n^{-\gamma} - n + 1\right)^{-\frac{1}{\gamma}}.$$

Then

$$C_{n+1,\theta}(x,y_1,\dots,y_n) = \int_0^x C_{n,\frac{\theta}{1+\theta}}\left(\frac{\partial C_{2,\theta}}{\partial\xi}(\xi,y_1),\dots,\frac{\partial C_{2,\theta}}{\partial\xi}(\xi,y_n)\right)d\xi.$$

Example 7.10 can be generalized.

Example 7.12. If the first random variable is independent of the others, then

$$C(x,y_1,\dots,y_n) = xC_0(y_1,\dots,y_n) = \int_0^x C_0(y_1,\dots,y_n)d\xi.$$

The above example implies the following equivalence:

Lemma 7.4. *The first random variable X is independent of $\mathbf{Y} = (Y_1,\dots,Y_n)$ if and only if the copula C is invariant and for every i, $i = 1,\dots,n$, X is independent of Y_i.*

Example 7.13. The Gaussian copula.
If C_0 is a Gaussian copula and C_i are independence copulas, then C is also a Gaussian copula. Indeed, let G_Σ be the n-variate Gaussian copula with correlation matrix Σ. Then

$$\int_0^x G_\Sigma(y_1,\dots,y_n)d\xi = xG_\Sigma(y_1,\dots,y_n) = G_{\Sigma_1}(x,y_1,\dots,y_n),$$

where Σ_1 is the $(n+1)\times(n+1)$ correlation matrix

$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}.$$

Note that the copula G_{Σ_1} is in a sense unique in the family of elliptical copulas—see Corollary 7.3.

Even taking simple examples of C_0 one may get interesting examples of multivariate copulas. For example when $C_0 = M$ we obtain copulas describing

the dependence of the random variables $X, Y_1, \ldots, Y_n$, such that Y_i's conditioned with respect to X are comonotonic. Note that for $n = 2$ such copulas are maximal elements of Fréchet classes of 3-copulas with given two 2-margins [53].

Example 7.14. Invariant conditionally comonotonic copulas.
Let $C(x, y)$ be any invariant bivariate copula. Then

$$\int_0^x \min\left(\frac{\partial C}{\partial \xi}(\xi, y_1), \ldots, \frac{\partial C}{\partial \xi}(\xi, y_n)\right) d\xi = C(x, \min(y_1, \ldots, y_n)).$$

Example 7.15. Hierarchical copulas.
Let C be a hierarchical $n + m + 1$-variate copula given by

$$C(x, y_1, \ldots, y_n, z_1, \ldots, z_m) = C_2(C_1(x, y_1, \ldots, y_n), z_1, \ldots, z_m),$$

where C_2 and C_1 are $(n+1)$- and m-variate copulas. If C_2 and C_1 are invariant, then so is C. Indeed, for any $\alpha \in (0, 1]$ we get

$$\begin{aligned} &C_2\left(C_1\left(x, \frac{C_{1,1}(\alpha, y_1)}{\alpha}, \ldots, \frac{C_{1,n}(\alpha, y_n)}{\alpha}\right), \frac{C_{2,1}(\alpha, z_1)}{\alpha}, \ldots, \frac{C_{2,n}(\alpha, z_m)}{\alpha}\right) \\ &= C_2\left(\frac{C_1(\alpha x, y_1, \ldots, y_n)}{\alpha}, \frac{C_{2,1}(\alpha, z_1)}{\alpha}, \ldots, \frac{C_{2,n}(\alpha, z_m)}{\alpha}\right) \\ &= C_2(C_1(\alpha x, y_1, \ldots, y_n), z_1, \ldots, z_m). \end{aligned}$$

Note that the so-called hierarchical (or nested) Archimedean copulas (HAC for short), constructed by iterated compositions of Archimedean copulas, are widely studied in the literature—see, for example, [25, 43, 51]. Combining Examples 7.11 and 7.15 we see that hierarchical Clayton copulas are invariant.

Example 7.16. Outer product.
Basing on Examples 7.12 and 7.15 we deduce that the outer product of an invariant $n + 1$-variate copula $C(x, y)$ and any m-variate copula $C_1(z)$,

$$C_*(x, y, z) = C(x, y)C_1(z),$$

is invariant.

Example 7.17. Vine copulas.
For the definition and basic properties, the reader is referred to [5].

A vine copula $C(x_1, \ldots, x_n)$ such that

(i) The bivariate marginal copulas $C_{1,i}$, $i = 2, \ldots, n$, are invariant.
(ii) C is represented by a sequence $(T_i)_{i=1}^{n-1}$ of $n - 1$ trees with T_1 having the first node of degree $n - 1$ and $n - 1$ nodes of degree 1 is invariant (compare Proposition 7.7). Indeed, C admits the representation

$$C(x_1,\dots,x_n)=\int_0^{x_1} C_0\left(\frac{\partial}{\partial\xi}C_{1,2}(\xi,x_2),\dots,\frac{\partial}{\partial\xi}C_{1,n}(\xi,x_n)\right)d\xi,$$

where C_0 is the vine copula represented by the sequence $(T_{i+1})_{i=1}^{n-2}$.

Example 7.18. Homogeneous distributions.
Copulas C associated with the distribution functions,

$$F(x,y_1,\dots,y_n)=\begin{cases}0 & \text{for } x<0 \ \vee \ \min(y_1,\dots,y_n)<0,\\ L(x,y_1,\dots,y_n) & \text{for } 0\le x\le 1 \ \wedge \ \min(y_1,\dots,y_n)\ge 0,\\ L(1,y_1,\dots,y_n) & \text{for } x>1 \ \wedge \ \min(y_1,\dots,y_n)\ge 0,\end{cases}$$

where L is homogeneous,

$$\forall\alpha,\in[0,1]\quad L(\alpha x,\alpha y_1,\dots,\alpha y_n)=\alpha L(x,y_1,\dots,y_n),$$

are invariant (compare Theorem 7.4). The marginal copulas are as in Example 7.7:

$$C_{1,i+1}(x,y_i)=C_{f_i}(x,y_i),\quad f_i(y_i)=L(1,+\infty,\cdots+\infty,y_i,+\infty,\cdots+\infty)$$

and C is given by

$$C(x,y_1,\dots,y_n)=L(x,g_1(y_1),\dots,g_n(y_n)),$$

where g_i are the right inverses of f_i, $f_i(g_i(t))=t$. Moreover

$$C_0(f_1(z_1)-f_1'(z_1)z_1,\dots,f_n(z_n)-f_n'(z_n)z_n)=\frac{\partial L}{\partial x}(1,z_1,\dots z_n)\quad a.e.$$

7.7 Rates of Convergence to Invariant Copulas

In more subtle modelling of extreme events not only the existence of the limit copula is important but also the rate of convergence of the conditional copulas to the limit. In the following we discuss three cases of asymptotic behaviour of conditioning. We shall deal with the following families:

- Copulas obtained by distortion of the first variable (for $n=2$ see [46] and [33] Section 4.4),
- Archimedean copulas,
- Copulas of elliptical random vectors with radius belonging to the Gumbel max-domain of attraction.

The proofs of new results are given in Sect. 7.11.

7.7.1 Distortions

Proposition 7.12. *Let $D(x_1, x_2, \ldots, x_n)$ be an invariant copula and $h : [0,1] \longrightarrow [0,1]$ be a surjective function such that for $t > 0$ $h(t)$ is positive, both $h(t)$ and $t/h(t)$ are non-decreasing and for $r > 0$,*

$$\lim_{t \to 0} \frac{h(tr)}{h(t)} = r.$$

Then

$$C(x_1, x_2, \ldots, x_n) = \begin{cases} \frac{x_1}{h(x_1)} D(h(x_1), x_2, \ldots, x_n) & \text{for } x_1 \in (0,1], \\ 0 & \text{for } x_1 = 0 \end{cases}$$

is a copula. Furthermore

$$C_{[\alpha]}(x_1, x_2, \ldots, x_n) = \frac{x_1 h(\alpha)}{h(\alpha x_1)} D\left(\frac{h(\alpha x_1)}{h(\alpha)}, x_2, \ldots, x_n\right) \quad \text{and} \quad \lim_{\alpha \to 0^+} C_{[\alpha]} = D.$$

The speed of convergence is determined by the function

$$\eta : (0,1] \times [0,1] \longrightarrow [0,1], \qquad \eta(\alpha, x) = \frac{h(\alpha x)}{h(\alpha)} - x.$$

Note that $\eta(\alpha, x)$ converges uniformly to 0 when α tends to 0.

Proposition 7.13. *Under the same assumptions as in Proposition 7.12,*

$$C_{[\alpha]}(x_1, x_2, \ldots, x_n) = D(x_1, x_2, \ldots, x_n) + \eta(\alpha, x_1)\left(\partial_1 D(x_1^+, \ldots x_n) - \frac{1}{x_1} D(x_1, \ldots x_n) + o(\alpha)\right).$$

Note that since D is invariant, the right-side derivatives exist for all $x_1 \in [0,1)$. For $x_1 = 1$ we put $\partial_1 D(x_1^+, \ldots x_n) = 0$.

7.7.2 Archimedean Copulas

The set of n-variate Archimedean copulas is invariant with respect to lower conditioning. When we enlarge it by one point, the comonotonic copula M, we get a closed invariant set. Hence the limit set of any Archimedean copula may only consist of Archimedean copulas and the copula M.

The limiting behaviour of an Archimedean copula depends on the exponent ρ of regular variation of its generator φ at 0. For regularly varying strict generators this exponent is given by

$$\lim_{s\to 0^+} \frac{\varphi(st)}{\varphi(s)} = t^\rho, \quad \rho \in [-\infty, 0), \quad t > 0,$$

where $t^{-\infty} = \infty, 1, 0$ when, respectively, $t < 1$, $t = 1$ and $t > 1$. For non-strict generators we have

$$\lim_{s\to 0^+} \frac{\varphi(0) - \varphi(st)}{\varphi(0) - \varphi(s)} = t^\rho, \quad \rho \in [0, 1], \quad t > 0.$$

Extending the description of extremal behaviour of strict Archimedean copulas given in [2] we get:

Theorem 7.9. *Let C be an n-variate Archimedean copula with generator φ regularly varying at 0. If the exponent ρ is non-zero or the right-side derivative φ' is regularly varying at 0, then*

$$\lim_{\alpha\to 0^+} C_{[\alpha]} = \begin{cases} Cl_{n,-\rho} & \text{for } \rho \in (-\infty, 0) \cup (0, 1] \\ \Pi_n & \text{for } \rho = 0, \\ M_n & \text{for } \rho = -\infty. \end{cases}$$

Corollary 7.2. *If C is a non-strict n-variate Archimedean copula with generator regularly varying at 0, then*

$$\rho \le (n-1)^{-1}.$$

The speed of convergence is determined by the quotient of the generators of the given copula and of the limit copula. For a strict generator φ regularly varying at 0 with exponent $\rho \in (-\infty, 0)$ we put

$$\varphi(t) = -\frac{1}{\rho}(t^\rho - 1) l_\varphi(t).$$

l_φ ir regularly varying at 0 with exponent 0 (i.e., it is slowly varying). If furthermore its right-side derivative l'_φ is regularly varying at 0 with exponent γ, then γ is equal or greater than -1 and the following estimate is valid (compare [21]):

$$\frac{l_\varphi(\alpha t)}{l_\varphi(\alpha)} = 1 + \eta(\alpha)(K_\gamma(t) + o(\alpha)), \tag{7.12}$$

where $\eta(t)$ denotes the elasticity of $l_\varphi(t)$,

$$\eta(t) = \frac{t l'_\varphi(t)}{l_\varphi(t)},$$

and

$$K_\gamma(t) = \frac{t^{1+\gamma} - 1}{\gamma + 1} \quad \text{for } \gamma > -1, \qquad K_{-1}(t) = \ln(t).$$

Theorem 7.10. *Let C be an n-variate strict Archimedean copula with generator φ regularly varying at 0. If the exponent ρ of φ is finite and non-zero and the right-side derivative l'_φ is regularly varying at 0 with exponent γ, then for all $x \in [0,1]^n$*

$$C_{[\alpha]}(\mathbf{x}) = Cl_{n,-\rho}(\mathbf{x}) \times \left(1 + \frac{1}{\rho}\eta(\alpha)\left(-K_\gamma(Cl_{n,-\rho}(\mathbf{x})) + Cl_{n,-\rho}(\mathbf{x})^{-\rho}\sum_{i=1}^{n} x_i^\rho K_\gamma(x_i) + o(\alpha)\right)\right).$$

7.7.3 Elliptical Copulas

Let $X = (X_1, \ldots, X_{n+1})^T$ be an elliptical random vector with stochastic representation

$$X \overset{d}{=} RA^T U,$$

where $A = (a_{i,j})_{i,j=1}^n$ is a non-singular $(n+1) \times (n+1)$ upper triangular matrix,

$$A = \begin{pmatrix} 1 & a \\ 0 & A_* \end{pmatrix},$$

U is a random vector uniformly distributed on the unit sphere of $\mathbb{R}^{n+1}$, and R is a positive random variable independent of U with distribution function H belonging to the Gumbel max-domain of attraction, i.e. for every $x \in \mathbb{R}$,

$$\lim_{t \to \omega^-} \frac{1 - H(t + x/w(t))}{1 - H(t)} = \exp(-x), \tag{7.13}$$

where $\omega = \operatorname{ess\,supp}(R)$.

Since elliptical random vectors are radially symmetric, the results concerning the upper conditioning of the first variable from [23, 24] can be restated for the lower conditioning case (see Remark 7.1).

Theorem 7.11 ([23, 24]). *Let $C_{[\alpha]}$ be the copula of the conditional distribution of X. Then*

$$\lim_{\alpha \to 0} C_{[\alpha]} = \mathscr{G}$$

where $\mathscr{G}$ is a Gaussian copula with covariance matrix

$$\tilde{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & A_*^\top A_* \end{pmatrix}.$$

Corollary 7.3. *The only invariant Gaussian copulas are copulas of random vectors $(X_1, \ldots, X_{n+1})$ such that X_1 and $(X_2, \ldots, X_{n+1})$ are independent.*

In order to be able to describe the speed of convergence in a uniform way we add the following:

Additional Assumption. The random vector $\mathbf{X}$ is absolutely continuous and its pdf is given by

$$\frac{1}{|\Sigma|^{1/2}} g(\mathbf{x}^T \Sigma^{-1} \mathbf{x}/2).$$

For some constants $\theta \in \mathbb{R}$, $\beta_1 \in (-1, 0]$ and $\beta_2 \in [0, \infty)$ and some scaling function $w(t)$,

$$g\left(\frac{1}{2}t^2 + \frac{tz}{w(t)}\right) = g\left(\frac{1}{2}t^2\right) e^{-z}(1 + \lambda(z,t)\eta(t)),$$

where for all t in a left neighbourhood of $\omega \in (0, \infty]$ and $z > 0$,

$$|\lambda(t,z)| \leq \theta \max\left(z^{\beta_1}, z^{\beta_2}\right),$$

and $\eta(t)$ is some positive function such that $\lim_{t\to\omega} \eta(t) = 0$.

Theorem 7.12 ([24]). *Let $C_{[\alpha]}$ be the copula of the conditional df of X. If the additional assumption is valid, then for $P\{X_1 \leq -t\} = \alpha \to 0$ we have a uniform expansion in $(p, \mathbf{q}) \in (0,1] \times [0,1]^n$,*

$$C_{[\alpha]}(p, \mathbf{q}) = \mathscr{G}(p, \mathbf{q}) - \frac{p \ln p}{\sqrt{tw(t)}} \nabla_{\mathbf{a}} \Phi\left(\Phi_1^{-1}(q_1), \ldots, \Phi_n^{-1}(q_n)\right) + O\left(\eta^*(t)\right),$$

where $\mathscr{G}$ is the Gaussian copula from Theorem 7.11, Φ is a Gaussian df of a centered Gaussian random vector with covariance matrix $A_*^\top A_*$, Φ_i are its marginals and

$$\eta^*(t) = \max((tw(t))^{-1}, \eta(t)), \quad t \in (0, \omega).$$

Remark 7.2. For the Gaussian copula C the above formula simplifies. We have $w(t) = t$, $\eta(t) = 0$ and $\eta^*(t) = \frac{1}{t^2}$.

7.8 Conditionally Log-Periodic Copulas

In this section we show how to construct log-periodic copulas.

Proposition 7.14. *Let C be any n-variate copula and c a fixed constant, $c \in (0,1)$. Then the ordinal sum of the infinite set of copies of C with respect to the intervals $\{(c^k, c^{k-1})\}_{k=1}^{\infty}$,*

$$G(\mathbf{x}) = \sum_{k=1}^{\infty} (c^{k-1} - c^k) C\left(F\left(\frac{x_1 - c^k}{c^{k-1} - c^k}\right), \ldots, F\left(\frac{x_n - c^k}{c^{k-1} - c^k}\right)\right),$$

is conditionally log-periodic with period c.

Proof. From the second point of Theorem 7.2 we know that $G_{[c]} = G$. Hence from Proposition 7.10 we conclude that the copula G is conditionally log-periodic with period c. □

In the recent literature one can find examples of Archimedean copulas which turn out to be log-periodic. In [41], Example 3, the Archimedean copulas with generator

$$\psi_a(\mathbf{x}) = \frac{1 + a\sin(\ln(1+x))}{1+x}, \quad x \geq 0, \quad a \in \left(0, \frac{\sqrt{10}}{5}\right],$$

are studied.

Proposition 7.15. *The Archimedean copula with generator ψ_a is log-periodic with prime period* $\exp(-2\pi)$.

7.9 Toroidal Limit Sets

Other interesting examples of limit sets are homeomorphic images of finite or infinite (but countably) dimensional tori, i.e. Cartesian products of circles,

$$T^N = \underset{k=1}{\overset{N}{\mathrm{X}}} S^1.$$

The easiest way to construct them is to consider the gluing ordinal sum of bivariate log-periodic copulas with logarithms of periods linearly independent over the field $\mathbb{Q}$ of rational numbers. Namely:

Theorem 7.13. *Let $\{C_k\}_{k=1}^{N}$ ($N \in \mathbb{N} \cup \{+\infty\}$) be a sequence of log-periodic bivariate copulas with periods* $\exp(-\lambda_k)$*, where $\{\lambda_k\}_{k=1}^{N}$ are positive real numbers which are linearly independent over $\mathbb{Q}$, and let G be a vertical-gluing ordinal sum with respect to non-empty and non-overlapping intervals $\{(a_k, b_k)\}_{k=1}^{N}$. Then the set of conditional copulas*

$$Cond(G) = \{G_{[\alpha]} : \alpha \in (0,1]\}$$

is a dense subset of a set T (homeomorphic to the N-dimensional torus T^N) consisting of all vertical-gluing sums of the conditional copulas $\{(C_k)_{[\alpha_k]}\}_{k=1}^N$ with respect to the same intervals $\{(a_k, b_k)\}_{k=1}^N$, where $(\alpha_k)_{k=1}^N$ ranges over the set $\mathrm{X}_{k=1}^N(\exp(-\lambda_k), 1]$.

Toroidal orbits also occur in $\mathscr{C}^n$ for $n > 2$.

Remark 7.3. Let G be a bivariate copula such that $Cond(G)$ is a dense subset of a set homeomorphic to a torus. The same is valid for the n-variate copula

$$G_1(x_1, \ldots, x_n) = G(x_1, x_2)x_3 \ldots x_n.$$

7.10 Chaotic Nature of Conditioning

In this section we provide examples which illustrate the chaotic behaviour of conditioning.

Lemma 7.5. *For any n-variate copula C and any non-empty open subset U of $\mathscr{C}^n$ there exists $\alpha_* \in (0,1]$ such that for every $c \in (0, \alpha_*]$ there exists a copula G belonging to U such that $G_{[c]} = C$.*

Proof. Since U is non-empty and open, it contains an open ball B, say $B = B(C_1, r)$. For G take the ordinal sum of C_1 and C with respect to the intervals $(c, 1)$ and $(0, c)$ for any $0 < c \leq r/(n+2)$. Indeed, by Theorem 7.2, $d(G, C_1) \leq (n+1)c < r$ and $G_{[c]} = C$. □

Lemma 7.5 implies the *topological mixing*,

$$\forall U, V \overset{open}{\subset} \mathscr{C}^n, U \neq \emptyset \neq V \quad \exists \alpha_* \in (0,1] \quad \forall c \in (0, \alpha_*] \quad Cond(U, c) \cap V \neq \emptyset,$$

and the *topological transitivity*,

$$\exists C \in \mathscr{C}^n \quad LimSet(C) = \mathscr{C}^n,$$

of univariate conditioning. Furthermore (see, for example, [20] Proposition 1.2.5), the set of copulas for which the limit set is smaller than the whole space $\mathscr{C}^n$ is a countable union of nowhere dense subsets of $\mathscr{C}^n$.

Example 7.19. A conditionally transitive copula.
Let $\{C_k\}_{k=1}^\infty$ be a countable family of n-variate copulas which is dense in $\mathscr{C}^n$ and let $\kappa : \mathbb{N} \to \mathbb{N}$ be a function obtained by "gluing" arithmetic sequences of increasing length. More precisely, for $m \geq 0$ and $k_1 \in [0, 2^m)$ we put

$$\kappa(2^m + k_1) = 1 + k_1.$$

Then the copula G which is the ordinal sum of the copulas $\{C_{\kappa(k)}\}_{k=1}^{\infty}$ with respect to the intervals $\{(\exp(-2^k-1), \exp(-2^{k-1}-1)\}_{k=1}^{\infty}$ is conditionally transitive, i.e. the limit set of G is equal to the whole space $\mathscr{C}^n$. Indeed, by Theorem 7.2, for fixed k_1 we get

$$\lim_{m\to\infty} G_{[(2^m+k_1)^{-1}]} = C_{1+k_1}.$$

Since the limit set is closed it must be equal to the whole space.

Lemma 7.6. *The set of all conditionally log-periodic copulas is a dense subset of the space $\mathscr{C}^n$.*

Proof. It is enough to show that every open ball B contains a log-periodic copula. If B has radius r and is centred at C, then just take the ordinal sum G of infinitely many copies of C constructed in Proposition 7.14 for $c = r/(n+2)$. □

Since there exist invariant copulas, conditioning is a non-minimal dynamical system (not all orbits are dense). Therefore conditioning belongs to the vast family of *Devaney chaotic* dynamical systems (see Definition 1.7.5 in [20]). Such systems exhibit sensitive dependence on initial conditions: a small alteration of C may imply a major change of the orbit $Cond(C)$. This phenomenon is often referred to as the butterfly effect.

Example 7.20. The butterfly effect.
Let C, D_1 and D_2 be n-variate copulas. Let G_i, $i = 1, 2$, be the ordinal sum of C and D_i with respect to intervals $(c, 1)$ and $(0, c)$. Then by Theorem 7.2, $d(G_i, C) \leq (n+1)c$ and $(G_i)_{[c]} = D_i$.

7.11 Proofs

Proof (Theorem 7.2).
To prove point 1 we consider two cases: when the smallest coordinate of $\mathbf{x}$ is not greater than a_1 and when all coordinates of $\mathbf{x}$ are greater than a_1.
If $\min(x_1, \ldots, x_n) \leq a_1$, then

$$|G(\mathbf{x}) - C_1(\mathbf{x})| \leq G(\mathbf{x}) + C_1(\mathbf{x}) \leq 2\min(x_1, \ldots, x_n) \leq 2a_1.$$

If $\min(x_1, \ldots, x_n) > a_1$, then

$$G(\mathbf{x}) = a_1 + (1-a_1)C_1\left(\frac{x_1-a_1}{1-a_1}, \ldots, \frac{x_n-a_1}{1-a_1}\right).$$

Hence, since copulas are Lipschitz functions, we get

$$|G(\mathbf{x}) - C_1(\mathbf{x})| = \left| a_1 + (1-a_1) C_1\left(\frac{x_1-a_1}{1-a_1}, \ldots, \frac{x_n-a_1}{1-a_1}\right) - C_1(\mathbf{x}) \right|$$

$$\leq \left| a_1 \left(1 - C_1\left(\frac{x_1-a_1}{1-a_1}, \ldots, \frac{x_n-a_1}{1-a_1}\right)\right)\right| + \left| C_1\left(\frac{x_1-a_1}{1-a_1}, \ldots, \frac{x_n-a_1}{1-a_1}\right) - C_1(\mathbf{x}) \right|$$

$$\leq a_1 + \sum_{i=1}^{n} \left| \frac{x_i - a_1}{1-a_1} - x_i \right| \leq a_1 + \sum_{i=1}^{n} \left| \frac{a_1(x_1-1)}{1-a_1} \right| \leq a_1 + a_1 \sum_{i=1}^{n} \frac{1-x_1}{1-a_1} \leq (n+1)a_1$$

Therefore

$$d(G, C_1) = \sup\{G(\mathbf{x}) - C_1(\mathbf{x}) : x \in [0,1]^n\} \leq (n+1)a_1.$$

We start the proof of point 2 with the observation that the bivariate marginal copulas $G_{1,i}$ with first variable a_1 are equal to the smaller variable. Indeed, since the intervals (a_k, b_k), $k > 1$, do not overlap $(a_1, 1)$, we get

$$F_U\left(\frac{a_1 - a_k}{b_k - a_k}\right) = \begin{cases} 0 & \text{for } k = 1, \\ 1 & \text{for } k > 1. \end{cases}$$

Hence

$$G_{1,i}(a_1, y_i) = G(a_1, 1, \ldots, 1, y_i, 1, \ldots, 1) = \sum_{k=2}^{N} (b_k - a_k)$$

$$\times C\left(1, \ldots, 1, F_U\left(\frac{y_i - a_k}{b_k - a_k}\right), 1, \ldots, 1\right) + m\left([0, \min(a_1, y_i)] \setminus \bigcup_{k=1}^{N} [a_k, b_k]\right)$$

$$= \sum_{k=2}^{N} (b_k - a_k) F_U\left(\frac{y_i - a_k}{b_k - a_k}\right) + m\left([0, \min(a_1, y_i)] \setminus \bigcup_{k=1}^{N} [a_k, b_k]\right)$$

$$= \begin{cases} \sum_{k=2}^{N} (b_k - a_k) + m\left([0, a_1] \setminus \bigcup_{k=1}^{N} [a_k, b_k]\right) & \text{for } y_i \geq a_1 \\ \sum_{k: y_i \geq b_k} (b_k - a_k) + (y_i - a_{k^*}) & \\ \qquad + m\left([0, y_i] \setminus \bigcup_{k=1}^{N} [a_k, b_k]\right) & \text{for } y_i \in (a_{k^*}, b_{k^*}) \\ \sum_{k: y_i \geq b_k} (b_k - a_k) + m\left([0, y_i] \setminus \bigcup_{k=1}^{N} [a_k, b_k]\right) & \text{for } y_i \in [0, a_i) \setminus \bigcup_{k=1}^{N} (a_k, b_k) \end{cases}$$

$$= \begin{cases} a_1 & \text{for } y_i \geq a_1 \\ y_i & \text{for } y_i \in (a_{k^*}, b_{k^*}) \\ y_i & \text{for } y_i \in [0, a_i) \setminus \bigcup_{k=1}^{N} (a_k, b_k) \end{cases} = \min(a_1, y_i).$$

We check whether

$$G^*(\mathbf{x}) = \sum_{k=1}^{N} \frac{b_{k+1} - a_{k+1}}{a_1} \times C_{k+1}\left(F_U\left(\frac{x_1 - a_{k+1}/a_1}{b_{k+1}/a_1 - a_{k+1}/a_1}\right), \ldots, F_U\left(\frac{x_n - a_{k+1}/a_1}{b_{k+1}/a_1 - a_{k+1}/a_1}\right)\right) + m([0, \min(x_1, \ldots, x_n)] \setminus \bigcup_{k=1}^{N} [a_{k+1}/a_1, b_{k+1}/a_1])$$

fulfills (7.1) for $\alpha = a_1$. We have

$$G^*\left(x, \frac{G_1(a_1, y_1)}{a_1}, \ldots, \frac{G_n(a_1, y_n)}{a_1}\right) = G^*\left(\frac{a_1 x}{a_1}, \frac{\min(a_1, x_2)}{a_1}, \ldots, \frac{\min(a_1, x_n)}{a_1}\right)$$

$$= \sum_{k=2}^{N} \frac{b_k - a_k}{a_1} C_k\left(F_U\left(\frac{a_1 x_1 - a_k}{b_k - a_k}\right), F_U\left(\frac{\min(a_1, x_2) - a_k}{b_k - a_k}\right), \ldots,\right.$$

$$\left. F_U\left(\frac{\min(a_1, x_n) - a_k}{b_k - a_k}\right)\right) + m([0, \min(x_1 a_1, x_2 \ldots, x_n)/a_1] \setminus \bigcup_{k=2}^{N} [a_k/a_1, b_k/a_1]),$$

$$= \frac{1}{a_1} \sum_{k=2}^{N} (b_k - a_k) C_k\left(F_U\left(\frac{a_1 x_1 - a_k}{b_k - a_k}\right), F_U\left(\frac{x_2 - a_k}{b_k - a_k}\right) \ldots, F_U\left(\frac{x_n - a_k}{b_k - a_k}\right)\right)$$

$$+ \frac{1}{a_1} m([0, \min(x_1 a_1, x_2 \ldots, x_n)] \setminus \bigcup_{k=2}^{N} [a_k, b_k]) = *.$$

Since $F_U((a_1 x_1 - a_1)/(b_1 - a_1)) = 0$, we get

$$* = \frac{1}{a_1} \sum_{k=1}^{N} (b_k - a_k) C_k\left(F_U\left(\frac{a_1 x_1 - a_k}{b_k - a_k}\right), F_U\left(\frac{x_2 - a_k}{b_k - a_k}\right) \ldots, F_U\left(\frac{x_n - a_k}{b_k - a_k}\right)\right)$$

$$+ \frac{1}{a_1} m([0, \min(x_1 a_1, x_2 \ldots, x_n)] \setminus \bigcup_{k=2}^{N} [a_k, b_k]) = \frac{1}{a_1} G(a_1 x_1, x_2, \ldots, x_n).$$

□

Proof (Theorem 7.3).
To show the continuity of *Cond* we consider two convergent sequences (α_k) and (C_k) from $(0, 1]$ and $\mathscr{C}^{n+1}$, respectively. Let

$$\lim_{n \to \infty} \alpha_k = \alpha_\infty > 0, \quad \lim_{n \to \infty} C_k = C_\infty.$$

We fix $z \in [0,1]^n$. Let $y_k \in [0,1]^n$ be any solution of the equation

$$C_k(\alpha_k, y_k) = \alpha_k z.$$

We denote by y_∞^- and y_∞^+ the following vectors of limits

$$y_\infty^- = \liminf_{k\to\infty} y_k, \quad y_\infty^+ = \limsup_{k\to\infty} y_k.$$

Obviously

$$C_\infty(\alpha_\infty, y_\infty^-) = \alpha_\infty z = C_\infty(\alpha_\infty, y_\infty^+).$$

Therefore, since copulas are non-decreasing in the second variable, we get

$$\liminf_{k\to\infty}(C_k)_{[\alpha_k]}(x,z) = \liminf_{k\to\infty} \frac{1}{\alpha_k} C_k(\alpha_k x, y_k) \geq \frac{1}{\alpha_\infty} C_\infty(\alpha_\infty x, y_\infty^-) = (C_\infty)_{[\alpha_\infty]}(x,z)$$

$$= \frac{1}{\alpha_\infty} C_\infty(\alpha_\infty x, y_\infty^+) \geq \limsup_{k\to\infty} \frac{1}{\alpha_k} C_k(\alpha_k x, y_k) = \limsup_{k\to\infty}(C_k)_{[\alpha_k]}(x,z).$$

Hence the pointwise limit exists and equals $(C_\infty)_{[\alpha_\infty]}(x,z)$.

By the Ascoli Theorem (see, for example, Theorem 3.2.5 of [42]), we conclude that $(C_k)_{[\alpha_k]}$ converges to $(C_\infty)_{[\alpha_\infty]}$ as $k \to \infty$.

The proof of the second assertion follows along the same lines as the proof of the bivariate case (see [33] Proposition 2.2). □

Proof (Theorem 7.4).
Let y be greater than 1. For small t we get the inequality

$$t \geq \widehat{C_{\mathbf{v}}}(t, 1, \ldots, 1, ty, 1, \ldots 1) \geq \widehat{C_{\mathbf{v}}}(t, ty, \ldots, ty).$$

Dividing by t and passing to the limit we get

$$1 \geq L_{j,v_j}(1,y) \geq L_{\mathbf{v}}(1, y, \ldots, y).$$

Hence each $L_{j,v_j}(1,\cdot)$ is a continuous increasing function taking all values from $[0,1)$. We substitute x_j by αx_j in the equality given by Proposition 7.1 to obtain

$$(\widehat{C_{\mathbf{v}}})_{[\alpha]}\left(x_1, \frac{(\widehat{C_{\mathbf{v}}})_{1,2}(\alpha, \alpha x_2)}{\alpha}, \ldots, \frac{(\widehat{C_{\mathbf{v}}})_{1,n}(\alpha, \alpha x_n)}{\alpha}\right) = \frac{1}{\alpha}\widehat{C_{\mathbf{v}}}(\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

Let C_* be any limit point of $(\widehat{C_{\mathbf{v}}})_{[\alpha]}$. Passing to the limit we get

$$C_*(x_1, L_{2,v_2}(1,x_2), \ldots, L_{n,v_n}(1,x_n)) = L_{\mathbf{v}}(x_1, \ldots x_n).$$

Thus C_* is uniquely determined by L's. Hence there can be only one limit point, i.e. the limit $C_{[0]}$ exists. □

Proof (Theorem 7.5).
Let C_* be any limit point of $C_{[\alpha]}$. In the same way as in the proof of Theorem 7.4 we find that for every vertex $\mathbf{v}$,

$$\widehat{(C_*)}_{\mathbf{v}}(x_1, L_{2,v_2}(1, x_2), \ldots, L_{n,v_n}(1, x_n)) = L_{\mathbf{v}}(x_1, x_2, \ldots, x_n).$$

Since for every $j \in \{2, \ldots, n\}$,

$$\lim_{y\to\infty} L_{j,0}(1, y) + \lim_{y\to\infty} L_{j,1}(1, y) = 1,$$

C_* is unique. □

Proof (Proposition 7.12).
First we show that a distorted copula is a "true" copula. The boundary conditions (C1) and (C2) are obvious. We put

$$h_1(t) = \begin{cases} \frac{t}{h(t)} & \text{for } t \in (0, 1], \\ 0 & \text{for } t = 0. \end{cases}$$

Condition (C3) follows from the assumption that both $h(t)$ and $h_1(t)$ are non-decreasing. Indeed

$$\begin{aligned} V_C(\mathbf{u}, \mathbf{w}) &= V_C\left([u_1, w_1] \times \bigtimes_{i=2}^{n} [u_i, w_i]\right) \\ &= V_C\left([0, w_1] \times \bigtimes_{i=2}^{n} [u_i, w_i]\right) - V_C\left([0, u_1] \times \bigtimes_{i=2}^{n} [u_i, w_i]\right) \\ &= h_1(w_1) V_D\left([0, h(w_1)] \times \bigtimes_{i=2}^{n} [u_i, w_i]\right) \\ &\qquad - h_1(u_1) V_D\left([0, h(u_1)] \times \bigtimes_{i=2}^{n} [u_i, w_i]\right) \\ &= (h_1(w_1) - h_1(u_1))\ V_D\left([0, h(w_1)] \times \bigtimes_{i=2}^{n} [u_i, w_i]\right) \\ &\qquad + h_1(u_1) V_D\left([h(u_1), h(w_1)] \times \bigtimes_{i=2}^{n} [u_i, w_i]\right) \ \geq\ 0. \end{aligned}$$

Since the volumes V_D are non-negative, we get that the volume V_C is non-negative as well.

The fact that the conditional copula $C_{[\alpha]}$ is a distorted copula is implied by the invariance of the basic copula D. We apply Proposition 7.1.

$$\frac{x_1 h(\alpha)}{h(\alpha x_1)} D\left(\frac{h(\alpha x_1)}{h(\alpha)}, \frac{C_{1,2}(\alpha, x_2)}{\alpha}, \dots\right)$$
$$= \frac{x_1 h(\alpha)}{h(\alpha x_1)} D_{[h(\alpha)]}\left(\frac{h(\alpha x_1)}{h(\alpha)}, \frac{D_{1,2}(h(\alpha), x_2)}{h(\alpha)}, \dots\right)$$
$$= \frac{x_1}{h(\alpha x_1)} D\left(h(\alpha x_1), x_2, \dots\right) = C(\alpha x_1, x_2, \dots).$$

The convergence of $C_{[\alpha]}$ follows from the assumption that h is regularly varying with index 1. □

Proof (Proposition 7.13).
Since $t/h(t)$ is non-decreasing, $h(\alpha x) \geq x h(\alpha)$ and we may apply the following estimate:

$$C_{[\alpha]}(x_1, x_2, \dots, x_n) = \frac{x_1 h(\alpha)}{h(\alpha x_1)} D\left(\frac{h(\alpha x_1)}{h(\alpha)}, x_2, \dots, x_n\right)$$
$$= \frac{x_1}{x_1 + \eta(\alpha, x_1)} \left(D(x_1, \dots, x_n) + \eta(\alpha, x_1)(\partial_1 D(x_1^+, \dots x_n) + o(\alpha))\right)$$
$$= D(x_1, \dots, x_n) + \eta(\alpha, x_1)\left(-\frac{1}{x_1} D(x_1, \dots x_n) + \partial_1 D(x_1^+, \dots x_n) + o(\alpha)\right).$$

□

Proof (Theorem 7.9).
We will base on the formula

$$\varphi(\alpha z_\alpha) = \min\left(\varphi(0), \sum_{i=1}^{n} \varphi(\alpha x_i) - (n-1)\varphi(\alpha)\right), \tag{7.14}$$

where $z_\alpha = C_{[\alpha]}(x_1, \dots, x_n)$. For non-strict generators we get

$$\frac{\varphi(0) - \varphi(\alpha z_\alpha)}{\varphi(0) - \varphi(\alpha)} = \max\left(0, \sum_{i=1}^{n} \frac{\varphi(0) - \varphi(\alpha x_i)}{\varphi(0) - \varphi(\alpha)} - (n-1)\right). \tag{7.15}$$

Passing to the limit we obtain

$$\lim_{\alpha \to 0} z_\alpha^\rho = \max(0, \sum_{i=1}^{n} x_i^\rho - (n-1)),$$

which finishes the proof for $\rho \in (0, 1]$.

For strict generators we drop "min" (because $\varphi(0) = \infty$) and divide both sides of (7.14) by $\varphi(\alpha)$ to get

$$\frac{\varphi(\alpha z_\alpha)}{\varphi(\alpha)} = \sum_{i=1}^{n} \frac{\varphi(\alpha x_i)}{\varphi(\alpha)} - (n-1). \tag{7.16}$$

Passing to the limit we obtain

$$\lim_{\alpha \to 0} z_\alpha^\rho = \sum_{i=1}^{n} x_i^\rho - (n-1),$$

which finishes the proof for $\rho \in (-\infty, 0)$.

To prove the case $\rho = -\infty$ we divide both sides of (7.14) by $\varphi(\alpha z_\alpha)$ to get

$$1 = \sum_{i=1}^{n} \frac{\varphi(\alpha x_i)}{\varphi(\alpha z_\alpha)} - (n-1)\frac{\varphi(\alpha)}{\varphi(\alpha z_\alpha)}. \tag{7.17}$$

From Fréchet–Hoeffding bounds we have

$$z_\alpha \leq \min(x_1, \ldots x_n).$$

We will show that after passing to the limit the inequality "$\leq$" can be replaced by an equality. Indeed, if for any sequence $\alpha_k \to 0$,

$$\lim_{k \to \infty} z_{\alpha_k} < \min(x_1, \ldots x_n),$$

then the limit of the right side of (7.17) (for $\alpha = \alpha_k$) would be 0, a contradiction.

If $\rho = 0$ and φ' is regularly varying, then by de l'Hospital rule we get that φ' is regularly varying with exponent -1. We base on the approximations for, respectively, non-strict and strict generators

$$\frac{\varphi(0) - \varphi(\alpha x)}{\varphi(0) - \varphi(\alpha)} = 1 + \eta_1(\alpha)(\ln x + o(\alpha)),$$

$$\frac{\varphi(\alpha x)}{\varphi(\alpha)} = 1 + \eta_2(\alpha)(\ln x + o(\alpha)),$$

where η is, respectively, equal to the elasticity of $\varphi(0) - \varphi$ and φ,

$$\eta_1(\alpha) = -\frac{\varphi'(\alpha)\alpha}{\varphi(0) - \varphi(\alpha)}, \quad \eta_2(\alpha) = \frac{\varphi'(\alpha)\alpha}{\varphi(\alpha)}.$$

Inserting the above into (7.15) and (7.16) and passing to the limit we finish the proof. □

Proof (Theorem 7.10).
We will base on the formula

$$\frac{\alpha^\rho z_\alpha^\rho - 1}{-\rho} l_\varphi(\alpha z) = \sum_{i=1}^n \frac{\alpha^\rho x_i^\rho - 1}{-\rho} l_\varphi(\alpha x_i) - (n-1)\frac{\alpha^\rho - 1}{-\rho} l_\varphi(\alpha),$$

where $z_\alpha = C_{[\alpha]}(x_1, \ldots, x_n)$. We divide both sides by $\alpha^\rho l_\varphi(\alpha)$ and apply the estimate (7.12) to obtain

$$\begin{aligned}&(z_\alpha^\rho - \alpha^{-\rho})(1 + \eta(\alpha)(K_\gamma(z_\alpha) + o(\alpha)))\\ &= \textstyle\sum_{i=1}^n (x_i^\rho - \alpha^{-\rho})(1 + \eta(\alpha)(K_\gamma(x_i) + o(\alpha))) - (n-1)(1 - \alpha^{-\rho}).\end{aligned}$$

Since

$$z_\alpha = Cl_{n,-\rho}(\mathbf{x}) + o(\alpha),$$

we get the second order approximation,

$$z_\alpha^\rho = Cl_{n,-\rho}(\mathbf{x})^\rho \left(1 + \eta(\alpha)\left(K_\gamma(Cl_{n,-\rho}(\mathbf{x})) + Cl_{n,-\rho}(\mathbf{x})^{-\rho} \sum_{i=1}^n x_i^\rho K_\gamma(x_i) + o(\alpha)\right)\right).$$

□

Proof (Proposition 7.15).
We will show that the generator of the conditional copula with $\alpha = \exp(-2\pi)$ is equivalent to ψ_a. Note that

$$\psi_a(\exp(2\pi) - 1) = \frac{1 + a\sin(\ln(\exp(2\pi)))}{\exp(2\pi)} = \exp(-2\pi).$$

So $\varphi(\exp(-2\pi)) = \exp(2\pi) - 1$ and (see Theorem 7.1)

$$\psi_{[\exp(-2\pi)]}(t) = \exp(2\pi)\psi_a(t + \exp(2\pi))$$

$$= \exp(2\pi)\frac{1 + a\sin(\ln(e^{2\pi}(e^{-2\pi}t + 1))}{t + e^{2\pi}} = \psi_a(t\exp(-2\pi)).$$

□

Proof (Theorem 7.13).
We start with the following observation: since the copulas C_k are log-periodic, their orbits are homeomorphic to circles, and the gluing ordinal sum T of their orbits is homeomorphic to a torus.

As was shown in Proposition 7.5 gluing of bivariate copulas commutes with conditioning. Hence each $G_{[\alpha]}$ belongs to T. The density follows from the fact that any vector of real numbers $\beta = (\beta_k)_{k=1}^N$ can be approximated by a vector $\hat{\beta} = (\ln\alpha - n_k\lambda_k)_{k=1}^N$, where $\alpha \in (0, 1]$ and n_k are integers. □

Acknowledgments The author acknowledges the partial support by Polish Ministry of Science and Higher Education, via the grant N N201 547838.

References

1. Charpentier, A., Juri, A.: Limiting dependence structures for tail events, with applications to credit derivatives. J. Appl. Probab., **43**(2), 563–586 (2006)
2. Charpentier, A., Segers, J.: Tails of multivariate Archimedean copulas. J. Multivariate Anal. **100**, 1521–1537 (2009)
3. Chavez-Demoulin, V., Embrechts, P.: An EVT primer for credit risk. In: Lipton, A., Rennie, A. (eds.) Handbook of Credit Derivatives. Oxford University Press, Oxford (2010)
4. Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance. Wiley, Hoboken (2004)
5. Czado, C.: Pair copula constructions of multivariate copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Lecture Notes in Statistics - Proceedings, vol. 198, pp. 93–109. Springer, Berlin (2010)
6. Donnelly, C., Embrechts, P.: The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. Astin Bull. **40**, 1–33 (2010)
7. Durante, F., Fernández-Sánchez, J.: Multivariate shuffles and approximation of copulas. Stat. Probab. Lett. **80**, 1827–1834 (2010)
8. Durante, F., Foscolo, E.: An analysis of the dependence among financial markets by spatial contagion. Int. J. Intell. Syst. **28**(4), 319–331 (2013)
9. Durante, F., Jaworski, P.: Spatial contagion between financial markets: a copula-based approach. Appl. Stoch. Models Bus. Ind. **26**, 551–564 (2010)
10. Durante, F., Jaworski, P.: Invariant dependence structure under univariate truncation. Statistics **46**, 263–277 (2012)
11. Durante, F., Sempi, C.: Copula theory: an introduction. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Lecture Notes in Statistics - Proceedings, vol. 198, pp. 3–31. Springer, Berlin (2010)
12. Durante, F., Klement, E.P., Quesada-Molina, J.J., Sarkoci, P.: Remarks on two product-like construction for copulas. Kybernetika **43**, 235–244 (2007)
13. Durante, F., Klement, E.P., Quesada-Molina, J.J.: Bounds for trivariate copulas with given bivariate marginals. J. Inequal. Appl. **2008**, 1–9, Article ID 161537 (2008)
14. Durante, F., Foschi, R., Spizzichino, F.: Threshold copulas and positive dependence. Stat. Probab. Lett. **78**(17), 2902–2909 (2008)
15. Durante, F., Jaworski, P., Mesiar, R.: Invariant dependence structures and Archimedean copulas. Stat. Probab. Lett. **81**, 1995–2003 (2011)
16. Durante, F., Foscolo, E., Sabo, M.: A spatial contagion test for financial markets. In: Kruse, R. et al. (eds.) Synergies of Soft Computing and Statistics for Intelligent Data Analysis. Advances in Intelligent and Soft Computing, vol. 190, pp. 321–329. Springer, Berlin (2013)
17. Embrechts, P.: Copulas: a personal view. J. Risk Insur. **76**, 639–650 (2009)
18. Genest, C., Nešlehová, J.: Assessing and modeling asymmetry in bivariate continuous data. In: Jaworski, P., Durante, F., Härdle, W. (eds.) Copulae in Mathematical and Quantitative Finance, pp. 91–114. Springer, Berlin (2013)
19. Genest, C., Gendron, M., Bourdeau-Brien, M.: The advent of copulas in finance. Eur. J. Finance. **15**, 609–618 (2009)
20. Gidea, M., Niculescu, C.: Chaotic dynamical systems: An introduction. Universitaria Press, Craiova (2002)
21. de Haan, L.: Von Mises-type conditions in second order regular variation. J. Math. Anal. Appl. **197**, 400–410 (1996)
22. Härdle, W., Okhrin, O.: De copulis non est disputandum - Copulae: an overview. AStA Adv. Stat. Anal. **94**(1), 1–31 (2010)

23. Hashorva, E.: Gaussian approximation of conditional elliptical random vectors. Stoch. Model. **22**, 441–457 (2006)
24. Hashorva, E., Jaworski, P.: Gaussian approximation of conditional elliptical copulas. J. Multivariate Anal. **111**, 397–407 (2012)
25. Hofert, M.: Construction and sampling of nested Archimedean copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Lecture Notes in Statistics - Proceedings, vol. 198, pp. 147–160. Springer, Berlin (2010)
26. Jágr, V., Komorníková, M., Mesiar, R.: Conditioning stable copulas. Neural Netw. World **20**(1), 69–79 (2010)
27. Jaworski, P.: On uniform tail expansions of bivariate copulas. Appl. Math. **31**, 397–415 (2004)
28. Jaworski, P.: On uniform tail expansions of multivariate copulas and wide convergence of measures. Appl. Math. **33**, 159–184 (2006)
29. Jaworski, P.: On copulas and their diagonals. Inform. Sci. **179**, 2863–2871 (2009)
30. Jaworski, P.: Tail Behaviour of Copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Proceedings of the Workshop Held in Warsaw, 25–26 Sept 2009. Lecture Notes in Statistics - Proceedings, vol. 198, pp. 161–186. Springer, Berlin (2010)
31. Jaworski, P.: Invariant dependence structure under univariate truncation: the high-dimensional case. Statistics (2012) online. doi:10.1080/02331888.2012.664143
32. Jaworski, P.: On copulas and differential inclusions. In: Kruse, R. et al. (eds.) Synergies of Soft Computing and Statistics for Intelligent Data Analysis. Advances in Intelligent and Soft Computing, vol. 190, pp. 321–329. Springer, Berlin (2013)
33. Jaworski, P.: On the characterization of copulas by differential equations. In: Communications in Statistics - Theory and Methods (2013, in press)
34. Jaworski, P., Pitera, M.: On spatial contagion and multivariate GARCH models. Appl. Stoch. Models Bus. Ind. (2013, in press)
35. Jaworski, P., Rychlik, T.: On distributions of order statistics for absolutely continuous copulas, with applications to reliability. Kybernetika **44**, 757–776 (2008)
36. Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and its Applications. Lecture Notes in Statistics - Proceedings, vol. 198. Springer, Berlin (2010)
37. Joe, H.: Multivariate Models and Dependence Concepts. Monographs on Statistics and Applied Probability, vol. 73. Chapman & Hall, London (1997)
38. Joe, H., Li, H., Nikoloulopoulos, A.: Tail dependence functions and vine copulas. J. Multivariate Anal. **101**, 252–270 (2010)
39. Juri, A., Wüthrich, M.V.: Copula convergence theorems for tail events. Insur. Math. Econ. **30**(3), 405–420 (2002)
40. Juri, A., Wüthrich, M.V.: Tail dependence from a distributional point of view. Extremes **6**(3), 213–246 (2003)
41. Larsson, M., Nešlehová, J.: Extremal behavior of Archimedean copulas. Adv. Appl. Probab. **43**, 195–216 (2011)
42. Łojasiewicz, S.: An Introduction to the Theory of Real Functions, 3rd edn. Wiley, Chichester (1988)
43. McNeil, A.J.: Sampling nested Archimedean copulas. J. Stat. Comput. Simul. **78**, 567–581 (2008)
44. McNeil, A.J., Nešlehová, J.: Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. Ann. Stat. **37**(5b), 3059–3097 (2009)
45. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools. Princeton Series in Finance. Princeton University Press, Princeton (2005)
46. Mesiar, R., Pekárová, M.: DUCS copulas. Kybernetika **46**, 1069–1077 (2010)
47. Mesiar, R., Sempi, C.: Ordinal sums and idempotents of copulas. Aequationaes Math. **79**, 39–52 (2010)
48. Mesiar, R., Jágr, V., Juráňová, M., Komorníková, M.: Univariate conditioning of copulas. Kybernetika **44**, 807–816 (2008)

49. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer Series in Statistics. Springer, New York (2006)
50. Nikoloulopoulos, A., Joe, H., Li, H.: Extreme value properties of multivariate t copulas. Extremes **12**, 129–148 (2009)
51. Okhrin, O., Okhrin, Y., Schmid, W.: Properties of hierarchical Archimedean copulas. Stat. Risk Model. **30**(1), 21–54 (2013)
52. Patton, A.J.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**, 527–556 (2006)
53. Quesada-Molina, J.J., Rodriguez-Lallena, J.A.: Some advances in the study of the compatibility of three bivariate copulas. J. Ital. Stat. Soc. **3**, 397–417 (1994)

Chapter 8
Singular Mixture Copulas

Dominic Lauterbach and Dietmar Pfeifer

Abstract We present a new family of copulas—the *Singular Mixture Copulas*. We begin with the construction of singular copulas whose supports lie on the graphs of two given quantile functions. These copulas are then mixed with respect to a continuous distribution resulting in a nonsingular parametric copula.

The Singular Mixture Copulas we construct have a Lebesgue density and in special cases even a closed form representation. Moreover, they have positive lower and upper tail dependence. Because Singular Mixture Copulas are mixtures of "simple" singular copulas, they can be simulated easily.

8.1 Introduction

Copulas provide an effective and versatile tool for modeling multivariate stochastic dependence. Since their introduction by Sklar in 1959 (see [11]) there have been intense developments in both the copula theory and their applications, see, e.g., [1, 5–7, 9, 10, 12].

In [10] several geometric methods of constructing copulas are presented. One approach deals with the construction of singular copulas whose supports lie in a given set. Another approach mixes an infinite family of copulas with respect to a mixing distribution. We present a new family of copulas—the *Singular Mixture Copulas*. These copulas result from a combination of the above-mentioned methods. In Sect. 8.2 we construct singular copulas whose supports lie on the graphs of two given quantile functions. These copulas are then mixed with respect to a continuous distribution resulting in an absolutely continuous parametric copula (Sect. 8.3).

D. Lauterbach (✉) · D. Pfeifer
Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
e-mail: dominic.lauterbach@uni-oldenburg.de; dietmar.pfeifer@uni-oldenburg.de

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_8,

As mixing distribution we particularly use a generalized beta distribution, i.e., a linear transformation of a beta distribution. Section 8.4 summarizes the results and gives an outlook on some extensions of this approach.

8.2 Singular Copulas

Let F be a continuous distribution function on $[0, 1]$ and let α be some constant in $]0, 1[$. Then there exists a continuous function G such that

$$\alpha F(x) + (1 - \alpha)G(x) = x \tag{8.1}$$

for all $x \in [0, 1]$. The function G is given by

$$G(x) = \frac{x - \alpha F(x)}{1 - \alpha}. \tag{8.2}$$

In general, G is not necessarily a distribution function. However, we are interested in exactly this case.

Let us assume for a moment that G is also a distribution function. Let X be a random variable with a continuous uniform distribution on $[0, 1]$, and let I be a random variable, independent of X, with a binomial $B(1, \alpha)$-distribution. Define the random variable Y via

$$Y := I \cdot F^{-1}(X) + (1 - I) \cdot G^{-1}(X). \tag{8.3}$$

Easy calculations show that Y follows a continuous uniform distribution on $[0, 1]$. As a consequence the distribution function of (X, Y) is a certain singular copula. So with two distribution functions F and G satisfying (8.1) we can construct singular copulas. Those copulas are given by

$$\begin{aligned} C_{XY}(x, y) &= \mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x, I \cdot F^{-1}(X) + (1 - I) \cdot G^{-1}(X) \le y) \\ &= \mathbb{P}(I = 1)\mathbb{P}(X \le x, X \le F(y)) + \mathbb{P}(I = 0)\mathbb{P}(X \le x, X \le G(y)) \\ &= \alpha \min(x, F(y)) + (1 - \alpha) \min(x, G(y)). \end{aligned}$$

As mentioned above, G is not necessarily a distribution function, so we have to make assumptions on F to guarantee that G is also a distribution function.

Lemma 8.1. *Let F be a differentiable distribution function on $[0, 1]$. Then the function G given by (8.2) is a differentiable distribution function on $[0, 1]$ if and only if $F'(x) \le \frac{1}{\alpha}$ for all $x \in [0, 1]$.*

Proof. From $F(0) = 0$ and $F(1) = 1$ it follows immediately that $G(0) = 0$ and $G(1) = 1$. From (8.2) we have

$$G'(x) = \frac{1 - \alpha F'(x)}{1 - \alpha}, \tag{8.4}$$

so that $G'(x) \geq 0 \Leftrightarrow F'(x) \leq \frac{1}{\alpha}$, which completes the proof. □

In a more general approach we can formulate the following theorem which follows from Lemma 8.1 and the construction discussed above.

Theorem 8.1. *Let F be a differentiable function on $[0, 1]$ and let X be a random variable with a continuous uniform distribution on $[0, 1]$. Then the distribution function of (X, Y) with Y given by (8.3) and G given by (8.2) is a copula if and only if*

(i) $F(0) = 0$ *and* $F(1) = 1$,
(ii) $0 \leq F'(x) \leq \frac{1}{\alpha}$ *for all* $x \in [0, 1]$.

We denote the class of functions that fulfill the properties (i) and (ii) in Theorem 8.1 by $\mathscr{F}_\alpha$, i.e.,

$$\mathscr{F}_\alpha := \{F : [0, 1] \to [0, 1] \mid F(0) = 0, F(1) = 1, 0 \leq F'(x) \leq \tfrac{1}{\alpha}\}.$$

Lemma 8.2.

1. *Let F and G be two functions in $\mathscr{F}_\alpha$, then $F \cdot G$ is in $\mathscr{F}_{\frac{\alpha}{2}}$.*
2. *Let F and G be two functions in $\mathscr{F}_\alpha$ and $\theta \in [0, 1]$, then $\theta F + (1-\theta)G$ is in $\mathscr{F}_\alpha$.*
3. *Let $F_1, F_2, \ldots$ be functions in $\mathscr{F}_\alpha$ with $\lim_{n\to\infty} F_n = F$, where the convergence is uniform, then F is in $\mathscr{F}_\alpha$.*
4. *Let α and β be some constants in $[0, 1]$ with $\alpha \leq \beta$, then $\mathscr{F}_\beta \subseteq \mathscr{F}_\alpha$.*
5. *Let F be a function in $\mathscr{F}_\alpha$, then G given by (8.2) is an element of $\mathscr{F}_{1-\alpha}$.*

Proof. The proof is straightforward. □

Example 8.1. Let F be a rational function given by $F(x) = (ax+b)/(cx+d)$. For which coefficients is F an element of $\mathscr{F}_\alpha$? From $F(0) \overset{!}{=} 0$ it follows that $b = 0$ and from $F(1) \overset{!}{=} 1$ it follows that $a = c + d$. Consequently, without loss of generality F can be written as $F(x) = ((c + 1)x)/(cx + 1)$. From the conditions on F' it follows that F is in $\mathscr{F}_\alpha$ if and only if $c \in [\alpha - 1, \frac{1}{\alpha} - 1]$.

Example 8.2. Let F be a quadratic function given by $F(x) = a_2x^2 + a_1x + a_0$. For which coefficients is F an element of $\mathscr{F}_\alpha$? From $F(0) \overset{!}{=} 0$ it follows that $a_0 = 0$ and from $F(1) \overset{!}{=} 1$ it follows that $a_2 + a_1 = 1$. As a consequence we have $F'(x) = 2a_2x + 1 - a_2$. In order to satisfy $F'(x) \geq 0$ the coefficient a_2 has to be in $[-1, 1]$. To fulfill the condition $F'(x) \leq \frac{1}{\alpha}$ easy calculations show that a_2 has to be an element of $[1 - \frac{1}{\alpha}, \frac{1}{\alpha} - 1]$. Altogether we can conclude that the quadratic function F given by $F(x) = ax^2 + (1 - a)x$ is in $\mathscr{F}_\alpha$ if and only if $a \in [\max(-1, 1 - \frac{1}{\alpha}), \min(1, \frac{1}{\alpha} - 1)]$. Figure 8.1 shows the functions F given by $F(x) = x^2$ and G given by (8.2) for different values of α.

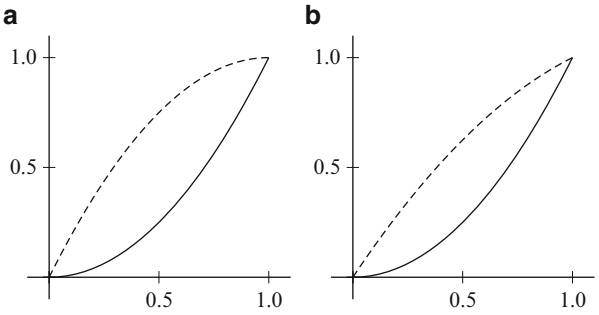

Fig. 8.1 F (*solid*), given by $F(x) = x^2$ and G, given by (8.2) for different values of α. (**a**) $\alpha = \frac{1}{2}$. (**b**) $\alpha = \frac{1}{3}$

Remark 8.1. The copula C_{XY} is a special case of the construction presented in [2] for the choice of $f_1 = f_2 = id_{[0,1]}$, $g_1 = F$, $g_2 = G$, $A(u, v) = B(u, v) = \min(u, v)$ and $H(x, y) = \alpha x + (1 - \alpha)y$. In this setting (8.1) can be obtained from Theorems 1 and 2 of [2].

In [3, 4] a copula $K_{\delta,\lambda}$ is presented that is given as follows

$$K_{\delta,\lambda}(x, y) = \min\{x, y, \lambda\delta(x) + (1 - \lambda)\delta(y)\}, \tag{8.5}$$

where δ is the diagonal section of a copula and λ is a constant that lies in an interval that is dependent on δ. They also show that $K_{\delta,\lambda}$ has a diagonal section equal to δ. Although the definitions of the copula $K_{\delta,\lambda}$ and C_{XY} might seem similar, they are not identical.

Remark 8.2. The copula $K_{\delta,\lambda}$ and the copula C_{XY} are essentially different.

Proof. Without loss of generality let $F(y) \leq y$ for all $y \in [0, 1]$, then $G(y) \geq y$ for all $y \in [0, 1]$. Choose x, y in a way that $y < x < G(y)$ holds. If $K_{\delta,\lambda}$ were equal to C_{XY}, then the diagonal δ of $K_{\delta,\lambda}$ would be given by $\delta(x) = C_{XY}(x, x) = \alpha F(x) + (1 - \alpha)x$. Consequently, the following equations would hold

$$C_{XY}(x, y) = \alpha F(y) + (1 - \alpha)x,$$
$$K_{\delta,\lambda}(x, y) = \min\{y, \alpha\lambda F(x) + \lambda(1 - \alpha)x + (1 - \lambda)\alpha F(y) + (1 - \alpha)(1 - \lambda)y\}.$$

Obviously, the equation $y = \alpha F(y) + (1 - \alpha)x$ does not hold for arbitrary x, y with $y < x < G(y)$, so it must hold

$$\alpha F(x) + (1 - \alpha)x = \alpha\lambda F(x) + \lambda(1 - \alpha)x + (1 - \lambda)\alpha F(y) + (1 - \alpha)(1 - \lambda)y \tag{8.6}$$

in order to satisfy $K_{\delta,\lambda} = C_{XY}$. Equation 8.6 is equivalent to

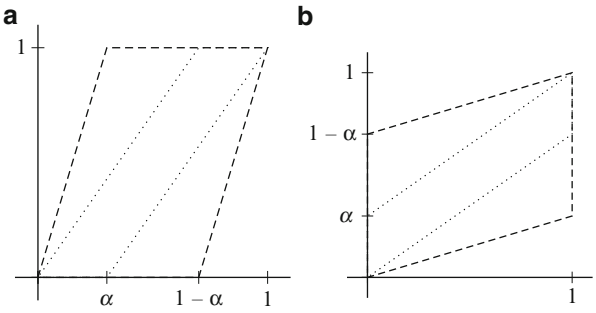

Fig. 8.2 (**a**) Borders for F and G. (**b**) Borders for the copula

$$\alpha\lambda(F(y) - F(x)) + (1-\lambda)(1-\alpha)(x-y) = 0, \tag{8.7}$$

which can be written as

$$\lambda\left(1 + \frac{(1-\alpha)(x-y)}{\alpha(F(x)-F(y))}\right) = \frac{(1-\alpha)(x-y)}{\alpha(F(x)-F(y))}, \tag{8.8}$$

since $F(x) \neq F(y)$ due to (8.7) and $y < x$. From (8.8) we can conclude[1] that

$$\lambda = \frac{1}{1 + \frac{\alpha(F(x)-F(y))}{(1-\alpha)(x-y)}}. \tag{8.9}$$

Since the right-hand side of the last equation is not constant for any nonlinear function F the statement follows. □

The support of the constructed copula C_{XY} always lies on the graphs of the functions F^{-1} and G^{-1}. Given a fixed α due to the restrictions on F (and G) there are points in $[0,1]^2$ which cannot be part of the support of the copula, regardless of which function $F \in \mathscr{F}_\alpha$ is chosen. Part (a) of Fig. 8.2 shows the borders in which the graphs of F (dashed line) and G (dotted line) have to lie. Having the borders of F and G it is easy to calculate the borders in which the support of the copula has to lie (see part (b) of Fig. 8.2). The function F has to fulfill the condition $F'(x) \leq \frac{1}{\alpha}$ for all $x \in [0,1]$, as a consequence points in the triangle $(1-\alpha, 0)(1,1)(1,0)$ or the triangle $(0,0)(0,1)(\alpha,1)$ cannot lie on the graph of F. Analogously, the borders for G can be obtained.

[1]The term in brackets on the left-hand side of (8.8) is unequal to zero because otherwise it would follow that $0 = -1$.

8.3 Singular Mixture Copulas

In this section we construct the convex sums (see [10]) of the singular copulas presented in Sect. 8.2. We start with a description of the general construction and subsequently consider specific mixing distributions.

8.3.1 General Construction

Consider a family $\{F_\omega\} \subset \mathscr{F}_\alpha$ of distribution functions, then for a fixed ω we can construct a singular copula $\check{C}_\omega$ using F_ω and G_ω given by

$$G_\omega(y) = \frac{y - \alpha F_\omega(y)}{1-\alpha}.$$

The copula $\check{C}_\omega$ is the distribution function of the random vector (X, Y) where X is uniformly distributed on $[0, 1]$ and Y is given by

$$Y := I \cdot F_\omega^{-1}(X) + (1 - I) \cdot G_\omega^{-1}(X),$$

with $I \sim \mathscr{B}(1, \alpha)$. If Ω is a real-valued random variable and $F_\omega \in \mathscr{F}_\alpha$ for all observations ω of Ω, then the convex sum of $\{\check{C}_\omega\}$ is given by

$$\begin{aligned}\dot{C}(x, y) &= \int \check{C}_\omega(x, y)\mathbb{P}^\Omega(d\omega) \\ &= \alpha \int \min(x, F_\omega(y))\mathbb{P}^\Omega(d\omega) + (1-\alpha)\int \min(x, G_\omega(y))\mathbb{P}^\Omega(d\omega).\end{aligned}$$

Especially, consider the family of distribution functions F_ω given by $F_\omega(y) = \omega y^2 + (1-\omega)y$ with $\omega \in [-1, 1]$. Let $0 < \alpha \leq \frac{1}{2}$, then F_ω is an element of $\mathscr{F}_\alpha$ for all $\omega \in [-1, 1]$ (see Example 8.2). Let Ω be a random variable with values in $[-1, 1]$, then the Singular Mixture Copula resulting from the family $\{F_\omega\}_{\omega\in[-1,1]}$ is given by

$$C_\alpha(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

$$= \begin{cases} x & , (x, y) \in A_1, \\ x + \alpha\left((x-y)\left(F_\Omega(\beta) - 1\right) + (y^2 - y)\int_\beta^1 \omega\mathbb{P}^\Omega(d\omega)\right) & , (x, y) \in A_2, \\ \alpha\left((x-y)F_\Omega(\beta) + y + (y^2-y)\int_\beta^1 \omega\mathbb{P}^\Omega(d\omega)\right) & \\ \quad +(1-\alpha)\left(x + (y-x)F_\Omega(b)\right) + \alpha(y - y^2)\int_{-1}^b \omega\mathbb{P}^\Omega(d\omega) & , (x, y) \in A_3, \\ \alpha(x-y)F_\Omega(\beta) + y + \alpha(y-y^2)\int_{-1}^\beta \omega\mathbb{P}^\Omega(d\omega) & , (x, y) \in A_4, \\ y & , (x, y) \in A_5, \end{cases}$$

where $\beta = \frac{x-y}{y^2-y}$, $b = \beta\frac{\alpha-1}{\alpha}$ and

$$A_1 = \left\{(x,y) \in [0,1]^2 \,|x < y^2\right\},$$
$$A_2 = \left\{(x,y) \in [0,1]^2 \,|y^2 \leq x < \frac{-\alpha}{1-\alpha}(y-y^2) + y\right\},$$
$$A_3 = \left\{(x,y) \in [0,1]^2 \,|\frac{-\alpha}{1-\alpha}(y-y^2) + y \leq x < \frac{\alpha}{1-\alpha}(y-y^2) + y\right\},$$
$$A_4 = \left\{(x,y) \in [0,1]^2 \,|\frac{\alpha}{1-\alpha}(y-y^2) + y \leq x < 2y - y^2\right\},$$
$$A_5 = \left\{(x,y) \in [0,1]^2 \,|2y - y^2 \leq x\right\}.$$

The density of the copula is given by

$$c_\alpha(x,y) = \begin{cases} 0 & ,(x,y) \in A_1, \\ \alpha f_\Omega(\beta)\frac{y^2-2xy+x}{(y^2-y)^2} & ,(x,y) \in A_2, \\ \frac{y^2-2xy+x}{(y^2-y)^2}\left(\alpha f_\Omega(\beta) + \frac{(1-\alpha)^2}{\alpha} f_\Omega(b)\right) & ,(x,y) \in A_3, \\ \alpha f_\Omega(\beta)\frac{y^2-2xy+x}{(y^2-y)^2} & ,(x,y) \in A_4, \\ 0 & ,(x,y) \in A_5. \end{cases}$$

Remark 8.3. For $\alpha > \frac{1}{2}$ it is possible to change the distribution of Ω in such a way that one receives the same copulas as for $\alpha < \frac{1}{2}$, so we restrict our investigation to the case $\alpha \leq \frac{1}{2}$.

Theorem 8.2. *The copula C_α has upper and lower tail dependence given by*

$$\lambda_U = 1 - \alpha\left(\int_0^1 \omega \mathbb{P}^\Omega(d\omega) - \int_{-1}^0 \omega \mathbb{P}^\Omega(d\omega)\right) = \lambda_L.$$

Proof. The proof is straightforward. □

Since Singular Mixture Copulas are convex sums of the singular copulas mentioned in Sect. 8.2 the borders described in part (b) of Fig. 8.2 are also valid for Singular Mixture Copulas. Moreover, because Singular Mixture Copulas are absolutely continuous we are able to compare the area of the copula's support with the area of the unit square. From the discussion in Sect. 8.2 we know that the support cannot lie in the triangles $(0,0)(0,1)(\alpha,1)$ and $(1-\alpha,0)(1,0)(1,1)$. Consequently, the parallelogram in which the support of the Singular Mixture Copula can lie has an area of $\max(\alpha, 1-\alpha)$.

In the special case where F_ω is a quadratic function for every ω the support of the Singular Mixture Copula is bounded by the inverses of the functions $F_{-1}(x) = 2x - x^2$ and $F_1(x) = x^2$, respectively. Here the support has an area of $\frac{1}{3}$.

8.3.2 Special Cases

In the above-mentioned construction the mixing distribution has to be concentrated on a finite interval. Therefore a generalized beta distribution, viz. a linear transformation of a beta distribution, provides a reasonable choice as a mixing distribution. Moreover, the beta distribution is very flexible so the resulting Singular Mixture Copulas should also show this flexibility.

Figures 8.3 and 8.4 show scatter plots of simulated Singular Mixture Copulas with a generalized beta distribution as mixing distribution.

Theorem 8.3. *Let $C_{\alpha,p,q}$ denote a Singular Mixture Copula with a Beta$(-1, 1, p, q)$ mixing distribution. Then the survival copula of $C_{\alpha,p,q}$ is given by $\hat{C}_{\alpha,p,q} = C_{\alpha,q,p}$.*

Proof. The proof is straightforward. □

Another possible mixing distribution is a uniform distribution on the interval $[-\theta, \theta]$ with $\theta \leq 1$. The choice $\theta = 1$ would be a special case of the above-mentioned generalized beta distribution. Here, the copula, which we will denote with $C_{\alpha,\theta}$, and its density have a closed form representation and the upper and lower tail dependence coefficients can be determined. See [8] for the proofs of this subsection. The copula $C_{\alpha,\theta}$ is given by

$$
\begin{aligned}
C_{\alpha,\theta}(x, y) &= \mathbb{P}(X \leq x, Y \leq y) \\
&= \begin{cases}
x & , (x, y) \in A_1, \\
\frac{\alpha}{4\theta}\left(\frac{(x-y)^2}{y^2-y} + 2\theta(x+y) + \theta^2(y^2-y)\right) + (1-\alpha)x & , (x, y) \in A_2, \\
\frac{1}{2}\left(\frac{(x-y)^2}{2\theta(y^2-y)}\left(\frac{(1-\alpha)^2}{\alpha} + \alpha\right) + x + (1-\alpha\theta)y + \alpha\theta y^2\right) & , (x, y) \in A_3, \\
\frac{\alpha}{4\theta}\left(\frac{(x-y)^2}{y^2-y} + 2\theta(x+y) + \theta^2(y^2-y)\right) + (1-\alpha)y & , (x, y) \in A_4, \\
y & , (x, y) \in A_5,
\end{cases}
\end{aligned}
$$

where

$$
\begin{aligned}
A_1 &= \left\{(x, y) \in [0, 1]^2 \,|x \leq -\theta(y - y^2) + y\right\}, \\
A_2 &= \left\{(x, y) \in [0, 1]^2 \,| -\theta(y - y^2) + y < x < -\theta\frac{\alpha}{1-\alpha}(y - y^2) + y\right\}, \\
A_3 &= \left\{(x, y) \in [0, 1]^2 \,| -\theta\frac{\alpha}{1-\alpha}(y - y^2) + y < x < \theta\frac{\alpha}{1-\alpha}(y - y^2) + y\right\}, \\
A_4 &= \left\{(x, y) \in [0, 1]^2 \,|\theta\frac{\alpha}{1-\alpha}(y - y^2) + y < x < \theta(y - y^2) + y\right\}, \\
A_5 &= \left\{(x, y) \in [0, 1]^2 \,|x \geq \theta(y - y^2) + y\right\}.
\end{aligned}
$$

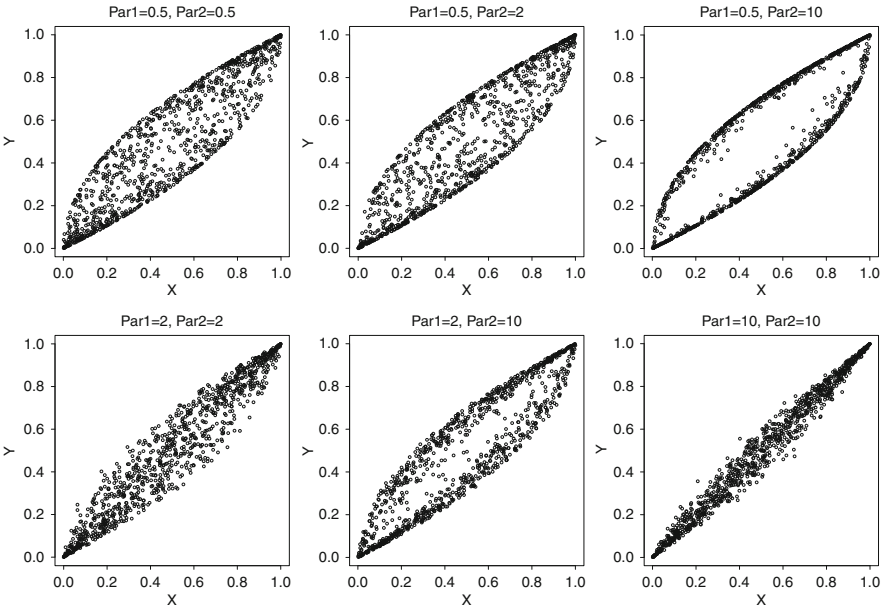

Fig. 8.3 Scatter plots of simulated points from a Singular Mixture Copula with generalized beta mixing distribution for $\alpha = 0.5$ and different shape parameters

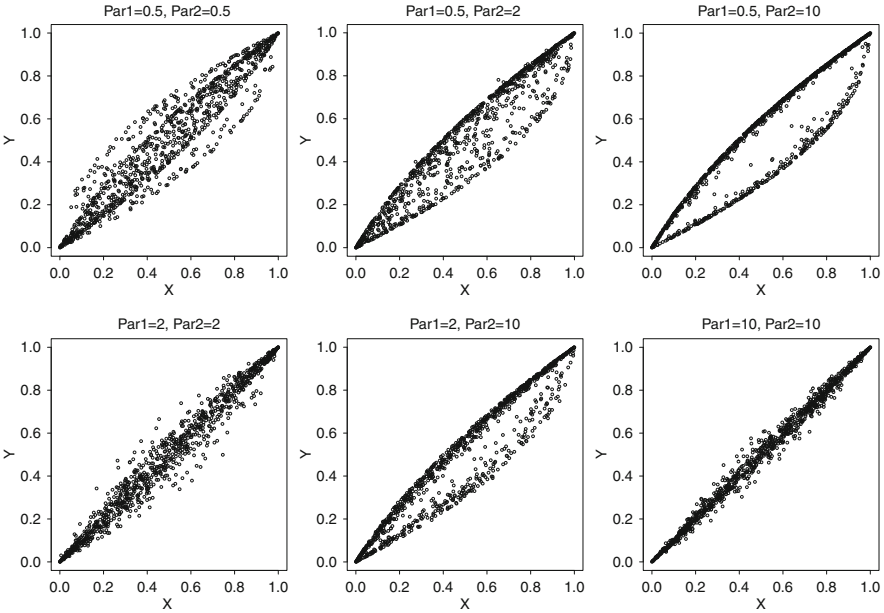

Fig. 8.4 Scatter plots of simulated points from a Singular Mixture Copula with generalized beta mixing distribution for $\alpha = 0.3$ and different shape parameters

The density of this copula is given by

$$c_{\alpha,\theta}(x,y) = \begin{cases} \frac{\alpha}{2\theta}\frac{y^2-2yx+x}{(y^2-y)^2} & ,(x,y) \in A_2, \\ \frac{1}{2\theta}\left(\frac{(1-\alpha)^2}{\alpha}+\alpha\right)\frac{y^2-2yx+x}{(y^2-y)^2} & ,(x,y) \in A_3, \\ \frac{\alpha}{2\theta}\frac{y^2-2yx+x}{(y^2-y)^2} & ,(x,y) \in A_4, \\ 0 & , \text{ otherwise.} \end{cases}$$

Theorem 8.4. *The copula $C_{\alpha,\theta}$ has upper and lower tail dependence given by*

$$\lambda_U = 1 - \frac{\alpha\theta}{2} = \lambda_L.$$

Theorem 8.5. *The copula $C_{\alpha,\theta}$ is radially symmetric, i.e., $C_{\alpha,\theta} = \hat{C}_{\alpha,\theta}$.*

Theorem 8.6. *The concordance measures Kendall's tau and Spearman's rho for the copula $C_{\alpha,\theta}$ are given by*

$$\tau_{\alpha,\theta} = 1 - \alpha\theta\frac{1+4(\alpha-1)^2}{9(1-\alpha)} \text{ and } \rho_{\alpha,\theta} = 1 - \frac{\alpha\theta^2}{15(1-\alpha)},$$

respectively.

Corollary 8.1. *Kendall's tau for the copula $C_{\alpha,\theta}$ lies in the interval $[\frac{7}{9}, 1]$, Spearman's rho for the copula $C_{\alpha,\theta}$ lies in the interval $[\frac{14}{15}, 1]$.*

8.4 Concluding Remarks

In this paper we presented a method for the construction of nonsingular copulas by mixing a family of singular copulas. We also showed how the constructed singular copulas differ from similar constructions in the literature. These copulas can be used to model strongly dependent random variables (see [8]).

In the future we want to investigate generalizations of the presented method, e.g., one could replace the quadratic functions in the definition of the singular copulas with other functions or use other mixing distributions.

References

1. Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance. Wiley, Chichester (2004)
2. Durante, F.: Construction of non-exchangeable bivariate distribution functions. Stat. Papers **50**, 383–391 (2009)

3. Durante, F., Jaworski, P.: Absolutely Continuous Copulas with Given Diagonal Sections. Comm. Statist. Theory Methods **37**, 2924–2942 (2008)
4. Durante, F., Kolesárová, A., Mesiar, R., Sempi, C.: Copulas with given diagonal sections: novel constructions and applications. Int. J. Uncertain. Fuzz. Knowledge-Based Syst. **15**, 397–410 (2007)
5. Embrechts, P., Lindskog, F., McNeil, A.J.: Modelling dependence with copulas and applications to risk management. In: Rachev, S.T. (ed.) Handbook of Heavy Tailed Distributions in Finance, pp. 329–384. Elsevier, Amsterdam (2003)
6. Genest, C., MacKay, R.: The joy of copulas: Bivariate distributions with uniform marginals. Am. Stat. **40**(4), 280–283 (1986)
7. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
8. Lauterbach, D.: Singular mixture copulas: a geometric method of constructing copulas. ZVersWiss **101**(5), 605–619 (2012)
9. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, Princeton (2005)
10. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (2006)
11. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
12. Wolff, E.F.: Measures of dependence derived from copulas. Ph.D. thesis, University of Massachusetts, Amherst (1977)

Chapter 9
Toward a Copula Theory for Multivariate Regular Variation

Haijun Li

Abstract Multivariate regular variation describes the relative decay rates of joint tail probabilities of a random vector with respect to tail probabilities of a norm of this random vector, and it is often used in studying heavy-tail phenomena observed in data analysis in various fields, such as finance and insurance. Multivariate regular variation can be analyzed in terms of the intensity measure or spectral measure but can also be studied by using the copula approach. In this paper, the basic ingredients of a measure-theoretic copula theory for multivariate regular variation are presented, and the method is based on extraction of scale-invariant extremal dependence from the intensity measure by standardizing its margins. Various examples as well as the advantages and disadvantages of the copula approach are also discussed.

9.1 Introduction

Multivariate regular variation has been widely used in analysis of multivariate extremes (see, e.g., [12, 59, 61]) and has also been used in analyzing tail risk [1, 3, 17, 18, 34, 38, 64, 66, 67] where risk (or failure) regions are usually non-orthant sets. In this paper we present the basic ingredients of a copula approach for multivariate regular variation. Our method is based on the upper tail-orthant limit representation of the Pickands dependence function ([15], also see Section 7.5.3 of [53] and [24]) that leads to vague convergence of measures induced by copulas. The copula approach yields rich distribution families for multivariate regular variation and also provides a tool for analyzing hidden and higher order regular variation [27, 28, 32].

H. Li (✉)
Department of Mathematics, Washington State University, Pullman, WA, USA
e-mail: lih@math.wsu.edu

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_9,

We begin with the standard setup of heavy-tail analysis on multivariate extremes. Let $X_n = (X_{1,n}, \cdots, X_{d,n})$, $n = 1, 2, \cdots$, be independent and identically distributed (iid) random vectors with common distribution function (df) F. Define component-wise maxima $M_{i,n} := \vee_{j=1}^{n} X_{i,j}$, $1 \leq i \leq d$. Here and hereafter $\vee$ ($\wedge$) denotes the maximum (minimum). This paper focuses on the limiting distributions, if any, of properly normalized vectors of component-wise maxima $M_n := (M_{1,n}, \ldots, M_{d,n})$, as $n \to \infty$, as well as on how structural properties of the limiting distributions can be described in terms of the asymptotic properties emerged from upper tails of F of the underlying sample $(X_n, n \geq 1)$.

For any two vectors $a, b \in \mathbb{R}^d$, the sum $a + b$, product ab, quotient a/b and vector power and vector inequalities such as $a \leq b$ are all operated component-wise. Let G be a df defined on $\mathbb{R}^d$ with non-degenerate margins. A df F is said to be in the domain of attraction of G for the maxima, denoted as $F \in \mathrm{DA}_{\vee}(G)$, if there exist $\mathbb{R}^d$-valued sequences $a_n = (a_{1,n}, \cdots, a_{d,n})$ with $a_{i,n} > 0$, $1 \leq i \leq d$, and $b_n = (b_{1,n}, \cdots, b_{d,n})$, $n = 1, 2, \cdots$, such that for any $x = (x_1, \cdots, x_d)$, as $n \to \infty$,

$$\mathbb{P}\left(\frac{M_{1,n} - b_{1,n}}{a_{1,n}} \leq x_1, \cdots, \frac{M_{d,n} - b_{d,n}}{a_{d,n}} \leq x_d\right) = F^n(a_n x + b_n) \to G(x), \tag{9.1}$$

and in this case, G is called a max *multivariate extreme value* (MEV) distribution. Similar definitions for min MEV distributions and their domain of attraction can be made. One needs only to study the case of maxima as the theory for minima is similar. A key property of an MEV distribution G is that all positive powers of G are also distributions, and max MEV distributions coincide with the max-stable distributions, which form a sub-class of max infinitely divisible distributions.

Let $X = (X_1, \ldots, X_d)$ denote a generic random vector with distribution F and continuous, univariate margins $F_1, \ldots, F_d$. If $F \in \mathrm{DA}_{\vee}(G)$, then G is closely related to the upper tail distribution of X. Without loss of generality, we may assume that X is nonnegative component-wise. Consider the standard case in which the survival functions $\overline{F}_i(x) := 1 - F_i(x)$, $1 \leq i \leq d$ of the margins are right tail equivalent; that is,

$$\frac{\overline{F}_i(x)}{\overline{F}_1(x)} = \frac{1 - F_i(x)}{1 - F_1(x)} \to 1, \text{ as } x \to \infty, \ 1 \leq i \leq d. \tag{9.2}$$

The distribution F or random vector X is said to be *multivariate regularly varying* (MRV) at ∞ with intensity measure ν if there exists a scaling function $b(t) \to \infty$ and a non-zero Radon measure $\nu(\cdot)$ such that as $t \to \infty$,

$$t\,\mathbb{P}\left(\frac{X}{b(t)} \in B\right) \to \nu(B), \ \forall \text{ relatively compact sets } B \subset \overline{\mathbb{R}}_+^d \backslash \{0\}, \text{ with } \nu(\partial B) = 0, \tag{9.3}$$

where $\overline{\mathbb{R}}_+^d := [0,\infty]^d$ is a compact set. By means of the one-point uncompactification (see pages 170–172 of [61]), the punctured version $\overline{\mathbb{R}}_+^d \backslash \{0\}$ provides a space in which any relatively compact neighborhood (e.g., any open neighborhood) of ∞ that is bounded away from the origin is bounded. The extremal dependence information of X is encoded in the intensity measure ν that satisfies the *scaling property* of order $-\alpha$; that is, $\nu(tB) = t^{-\alpha}\nu(B)$, for all relatively compact subsets B that are bounded away from the origin, where $\alpha > 0$ is known as the *tail index*. The examples of MRV dfs include (truncated) multivariate t distribution, multivariate Pareto distributions, and various members of the elliptical distribution family.

The law (9.3) of rare events can be rephrased in terms of relative decay rates of joint tail probabilities of X with respect to tail probabilities of its margin under the assumption (9.2). Since the set $B_1 = \{x \in \mathbb{R}_+^d : x_1 > 1\}$ is relatively compact within the cone $\overline{\mathbb{R}}_+^d \backslash \{0\}$ and $\nu(B_1) > 0$ under (9.2) for the nonzero measure $\nu(\cdot)$ with scaling property, it follows from (9.3) that the scaling function $b(t)$ can be chosen to satisfy that $\overline{F}_1(b(t)) = t^{-1}$, $t > 0$, after appropriately normalizing the intensity measure by $\nu(B_1)$. That is, $b(t)$ can be chosen as $b(t) = \overline{F}^{-1}(t^{-1}) = F_1^{-1}(1-t^{-1})$ under the condition (9.2), and thus (9.3) can be expressed equivalently as

$$\lim_{t\to\infty} \frac{\mathbb{P}(X \in tB)}{\mathbb{P}(X_1 > t)} = \nu(B), \ \forall \text{ relatively compact sets } B \subset \overline{\mathbb{R}}_+^d \backslash \{0\}, \tag{9.4}$$

satisfying that $\mu(\partial B) = 0$. It follows from (9.4) and (9.2) that for $1 \le i \le d$,

$$\lim_{t\to\infty} \frac{\mathbb{P}(X_i > ts)}{\mathbb{P}(X_i > t)} = \nu((s,\infty] \times \overline{\mathbb{R}}_+^{d-1}) = s^{-\alpha}\nu((1,\infty] \times \overline{\mathbb{R}}_+^{d-1}), \ \forall\, s > 0.$$

That is, univariate margins have regularly varying right tails. In general, a Borel-measurable function $g : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying with exponent $\rho \in \mathbb{R}$, denoted as $g \in \mathrm{RV}_\rho$, if and only if

$$g(t) = t^\rho \ell(t), \text{ with } \ell(\cdot) \ge 0 \text{ satisfying that } \lim_{t\to\infty} \frac{\ell(ts)}{\ell(t)} = 1, \text{ for } s > 0. \tag{9.5}$$

The function $\ell(\cdot)$ is known as a slowly varying function and denoted as $\ell \in \mathrm{RV}_0$. Since $\overline{F}_1 \in \mathrm{RV}_{-\alpha}$, $1/\overline{F}_1 \in \mathrm{RV}_\alpha$, and thus, by Proposition 2.6(v) of [61], the scaling function $b \in \mathrm{RV}_{\alpha^{-1}}$.

Since all the margins are tail equivalent as assumed in (9.2), one has

$$\overline{F}_i(t) = t^{-\alpha}\ell_i(t), \text{ where } \ell_i \in \mathrm{RV}_0, \text{ and } \ell_i(t)/\ell_j(t) \to 1 \text{ as } t \to \infty, \text{ for any } i \ne j, \tag{9.6}$$

which, together with $\overline{F}_1(b(t)) = t^{-1}$, imply that

$$\lim_{t\to\infty} t\,\mathbb{P}(X_i > b(t)s) = \lim_{t\to\infty} \frac{\mathbb{P}(X_i > b(t)s)}{\overline{F}_i(b(t))} \frac{\overline{F}_i(b(t))}{\overline{F}_1(b(t))} = s^{-\alpha},\ s > 0,\ 1 \le i \le d. \tag{9.7}$$

The multivariate regular variation and MEV distributions with Fréchet margins are related as described as follows.

Theorem 9.1 (Marshall and Olkin [51]). *Assume that (9.2) holds. Then there exist normalization vectors $a_n > 0$ and b_n such that*

$$\mathbb{P}\Big(\frac{M_n - b_n}{a_n} \le x\Big) \to G(x),\ as\ n \to \infty,\ \forall x \in \mathbb{R}_+^d,$$

where G is a d-dimensional distribution with Fréchet margins $G_i(s) = \exp\{-s^{-\alpha}\}$, $1 \le i \le d$, if and only if F is MRV with intensity measure $\nu([0,x]^c) := -\log G(x)$.

In other words, $F \in \mathrm{DA}_\vee(G)$ where G has Fréchet margins with tail index α if and only if F is MRV with intensity measure $\nu([0,x]^c) = -\log G(x)$.

Remark 9.1.

1. The normalization vectors $a_n > 0$ and b_n in Theorem 9.1 can be made precise so that $b_n = 0$ and $a_n = (\overline{F}_1^{-1}(1/n), \ldots, \overline{F}_d^{-1}(1/n))$ that depend only on the margins of F. Note that the Fréchet margins are not crucial but simplify the normalizing vectors. If $(X_1, \ldots, X_d)$ has a df $F \in \mathrm{DA}_\vee(G)$, then there exist monotone transformations $b^{(i)}(t)$, $1 \le i \le d$, such that $(b^{(1)}(X_1), \ldots, b^{(d)}(X_d))$ is MRV in the sense of (9.4).
2. If (9.2) does not hold, Theorem 9.1 can still be established but the non-standard global regular variation with different scaling functions among various margins needs to be used in place of (9.4), which uses the same scaling function among different margins.
3. The one-dimensional version of Theorem 9.1 is due to Gnedenko [23]. The univariate scaling property of $\nu(\cdot)$ on $\mathbb{R}_+$ leads to the explicit parametric expression for univariate MEV dfs, but the parametric feature enjoyed by univariate extremes is lost in the multivariate context.
4. The multivariate scaling property $\nu(\cdot)$ on $\mathbb{R}_+^d$ does allow a semi-parametric representation for G. Let $\mathbb{S}_+^{d-1} = \{a : a = (a_1, \ldots, a_d) \in \mathbb{R}_+^d, ||a|| = 1\}$, where $||\cdot||$ is a norm defined on $\mathbb{R}^d$. Using the polar coordinates, G can be expressed as follows:

$$G(x) = \exp\Big\{-c \int_{\mathbb{S}_+^{d-1}} \max_{1\le i\le d}\{(a_i/x_i)^\alpha\}\mathbb{Q}(da)\Big\}, \tag{9.8}$$

where $c > 0$ and $\mathbb{Q}$ is a probability measure defined on $\mathbb{S}_+^{d-1}$ such that

$$c \int_{\mathbb{S}_+^{d-1}} a_i^\alpha \mathbb{Q}(da) = 1, \quad 1 \le i \le d.$$

This is known as the Pickands representation [13, 58], and $c\mathbb{Q}(\cdot)$ is known as the spectral or angular measure.

5. Note that the spectral measure in (9.8) is a finite measure that can be approximated by a sequence of discrete measures. Using this idea, Marshall and Olkin [51] showed that the MEV distribution G is positively associated. This implies that as n is sufficiently large, one has asymptotically,

$$\mathbb{E}\Big(f\big(M_n\big)g\big(M_n\big)\Big) \ge \mathbb{E}\Big(f\big(M_n\big)\Big)\mathbb{E}\Big(g\big(M_n\big)\Big)$$

for all non-decreasing functions $f, g : \mathbb{R}^d \mapsto \mathbb{R}$. Observe that the sample vector X_n could have any dependence structure, but the strong positive dependence emerges among multivariate extremes.

6. Since G is max-infinitely divisible, all bivariate margins of G are TP_2, a positive dependence property that is even stronger than the positive association of bivariate margins (see Theorem 2.6 in [33]).

The detailed discussions on univariate and MEV theories can be found in [12, 21, 41, 59]. In contrast to the multivariate method, a geometric approach for extreme value analysis in high-dimensional spaces is discussed in [2]. Statistical methods for extreme value analysis are described in detail in [7]. The extreme value theory has found applications in various fields and, in particular, the applications of the univariate extreme value theory to modeling extremal events in insurance and finance can be found in [16].

Two basic ingredients of extreme value analysis are the regular variation and vague convergence of measures. The standard references for the theory of regularly varying functions include [8, 59, 63, 65]. The extensions of regular variation to linear operators and Borel measures can be found in [54]. A detailed account of the interplay between regular variation and vague convergence of Radon measures in heavy-tail analysis can be found in [61].

The discussions (9.2)–(9.7) for multivariate regular variation have been focused on $\mathbb{R}_+^d$. The extension of MRV beyond the nonnegative orthant can be done by using the tail probability of $||X||$, where $||\cdot||$ denotes a norm on $\mathbb{R}^d$, in place of the marginal tail probability in (9.4) (see [61], Section 6.5.5). The case that the limit in (9.2) is any nonzero constant can be easily converted into the standard tail equivalent case by properly rescaling margins. If the limit in (9.2) is zero or infinity, then some margins have heavier tails than others. One way to overcome this problem and to reveal the scaling property is to standardize the margins via marginal monotone transforms (see Theorem 6.5 in [61]), or to use the copula method [45], which is precisely the goal of this paper.

The rest of this paper is organized as follows. Section 9.2 introduces the basic elements of a copula theory for multivariate regular variation. The method is

based on rescaling the intensity measure in the copula setup that extracts the scale-invariant extremal dependence from the intensity measure by standardizing its margins. In contrast to the standardization used in [37], the standardization via the copula approach yields the limit measure with scaling property of order 1 that facilitates two representations:

1. The polar coordinate representation of Pickands type and
2. The Euler homogeneous representation.

Similar to any Pickands-type representation, the polar coordinate representation of the limit measure allows discretization of the angular measure, leading to various tractable, fully parametrized copula models that provide good approximations to a general MEV distribution [see Remark 9.1(5)]. The Euler representation, on the other hand, explores the extremal dependence decomposition of joint multivariate extremes that are driven by various univariate extremes and provides a natural limiting version of the total probability law for joint extremes. The power of the Euler representation is illustrated in Sect. 9.2 by explicitly deriving the intensity measure of random samples with multivariate t copulas. Section 9.3 discusses the tail densities of copulas that describe local extremal dependence. The local characterization of extremal dependence is especially useful for the distributions that are specified only by densities and provides a geometric approach for extremal dependence analysis [2]. Finally some remarks in Sect. 9.4 conclude the paper.

9.2 Copula Method for Multivariate Regular Variation

A copula C is a multivariate distribution with standard uniformly distributed margins on $[0, 1]$. Sklar's theorem (see Section 2.3 in [55], or Section 1.6 in [33]) states that every multivariate distribution F with margins $F_1, \ldots, F_d$ can be written as $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$ for some d-dimensional copula C. In fact, in the case of continuous margins, C is unique and

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

where $F_i^{-1}(u_i)$ is the quantile functions of the ith margin, $1 \leq i \leq d$. Let $(U_1, \ldots, U_d)$ denote a random vector with df C and $U_i, 1 \leq i \leq d$, being uniformly distributed on $[0, 1]$. The survival copula $\hat{C}$ is defined as follows:

$$\hat{C}(u_1, \ldots, u_n) = \mathbb{P}(1 - U_1 \leq u_1, \ldots, 1 - U_n \leq u_n) = \overline{C}(1 - u_1, \ldots, 1 - u_n) \tag{9.9}$$

where $\overline{C}$ is the joint survival function of C. The survival copula $\hat{C}$ can be used to transform lower tail properties of $(U_1, \ldots, U_d)$ into the corresponding upper tail properties of $(1-U_1, \ldots, 1-U_d)$ and thus only upper tail behaviors are studied and presented in this paper.

Consider the MEV df G described in Theorem 9.1 with Fréchet margins $G_i(s) = \exp\{-s^{-\alpha}\}$, $1 \le i \le d$, and clearly,

$$G_i^{-1}(u) = (-\ln u)^{-1/\alpha},\ 1 \le i \le d.$$

Using the Pickands representation (9.8), the copula of G, known as the extreme value copula (Section 7.5.1 of [53] and [24]), is given by

$$C_{\mathrm{EV}}(u_1, \ldots, u_d) = \exp\Big\{ - c \int_{\mathbb{S}_+^d} \max_{1 \le i \le d} \{(-\ln u_i) a_i^{\alpha}\} \mathbb{Q}(da) \Big\},\ (u_1, \ldots, u_d) \in [0,1]^d. \tag{9.10}$$

Note, however, that the spectral measure $\mathbb{Q}(\cdot)$ depends on margins of underlying samples (e.g., tail index α), and it would be desirable to express the extreme value copula in terms of scale-invariant information only.

Let C denote the copula of iid random vectors X_n, and $(U_1, \ldots, U_d)$ have the df C. Define the upper exponent function:

$$a(w; C) := \lim_{u \to 0+} \frac{\mathbb{P}\Big(\bigcup_{i=1}^d \{U_i > 1 - u w_i\} \Big)}{u},\ \forall w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d \backslash \{0\} \tag{9.11}$$

provided that the limit exists. This function is called the stable tail dependence function in [7, 15] and also known as the Pickands dependence function (see Section 6.1.5 of [12]). See [35, 56] for additional properties of this function.

Note that $a(\cdot; C)$ depends only on copula C. If the exponent function $a(\cdot; C)$ exists for a d-dimensional copula C, then the exponent function of any multivariate margin $C_I(u_i, i \in I)$ of C, $\emptyset \ne I \subset \{1, \ldots, d\}$,

$$a(w_I; C_I) = a((w_I, 0_{I^c}); C),\ w_I = (w_i, i \in I), \tag{9.12}$$

also exists, where 0_{I^c} is the $|I^c|$-dimensional sub-vector of zeros. As noted in [15] (also see Section 7.5.3 of [53] and [24]), the existence of $a(\cdot; C)$, as defined by (9.11), is equivalent to the existence of the EV copula C_{EV} of C, which in this case can be expressed as

$$C_{\mathrm{EV}}(u_1, \ldots, u_d) := \lim_{n \to \infty} C^n(u_1^{1/n}, \ldots, u_d^{1/n}) = \exp\{-a(-\log u_1, \ldots, -\log u_d; C)\}. \tag{9.13}$$

Compared to the condition (9.13) of copula domain of attraction, the limit (9.11) is more tractable. Furthermore, the limiting expression (9.11) can be rephrased in terms of vague convergence of Radon measures on $\overline{\mathbb{R}}_+^d \backslash \{0\}$ or on $\overline{\mathbb{R}}_+^d \backslash \{\infty\}$. Define the *exponent measure* $\mu(\cdot)$ on $\overline{\mathbb{R}}_+^d \backslash \{0\}$ generated by

$$\mu([0,w]^c) := a((w_1^{-1},\ldots,w_d^{-1});C),\ w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d\backslash\{0\}. \tag{9.14}$$

See, e.g., Section 6.1.3 of [12]. Since any relatively compact subset $K\subset\overline{\mathbb{R}}_+^d\backslash\{0\}$ that is bounded away from 0 is contained in $[0,w]^c$ for some $w\in\mathbb{R}_+^d\backslash\{0\}$, $\mu(\cdot)$ is a Radon measure on $\overline{\mathbb{R}}_+^d\backslash\{0\}$ and clearly homogeneous of order -1. Define the finite measures $\mu_n(\cdot)$, $n\geq 1$, generated by

$$\begin{aligned}\mu_n([0,w]^c) &:= n\,\mathbb{P}\left(\bigcup_{i=1}^d\left\{\frac{n^{-1}}{1-U_i}>w_i\right\}\right)\\ &= n\,\mathbb{P}\left((1-U_1,\ldots,1-U_d)\in n^{-1}\Big(\prod_{i=1}^d[w_i^{-1},\infty]\Big)^c\right)\end{aligned}$$

and then (9.11) can be rephrased via (9.14) as follows, as $n\to\infty$,

$$\mu_n([0,w]^c)\to\mu([0,w]^c),\ \forall w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d\backslash\{0\}.$$

It follows from Lemma 6.1 of [61] that (9.11) is equivalent to the vague convergence

$$\mu_n(K)\to\mu(K),\ \text{as } n\to\infty, \tag{9.15}$$

for all relatively compacts $K\subset\overline{\mathbb{R}}_+^d\backslash\{0\}$ that is bounded away from 0. The vague convergence (9.15), called the standard form of MRV, is the central object in multivariate regular variation [12, 61]. It can be easily shown that the standard multivariate regular variation with properly normalizing marginal transforms is equivalent to the multivariate domain of attraction (9.1) [see, e.g., [37] and Remark 9.1(1)]. Note that the standard MRV uses the standard Pareto margins, and such a standardization is necessary to reveal the underlying scaling property (of order -1).

In contrast to the standard MRV (9.14) and (9.15), the upper exponent function (9.11) induces directly a vague convergence of measures generated by copula C. Define a measure $\mathbb{U}(\cdot)$ on $\overline{\mathbb{R}}_+^d\backslash\{\infty\}$ as follows

$$\mathbb{U}\Big(\big(\prod_{i=1}^d[w_i,\infty]\big)^c\Big) := a((w_1,\ldots,w_d);C),\ (w_1,\ldots,w_d)\in\mathbb{R}_+^d. \tag{9.16}$$

Since any relatively compact subset $K\subset\overline{\mathbb{R}}_+^d\backslash\{\infty\}$ that is bounded away from ∞ is contained in $\big(\prod_{i=1}^d[w_i,\infty]\big)^c$ for some $w\in\mathbb{R}_+^d\backslash\{0\}$, the measure $\mathbb{U}(\cdot)$ is a Radon measure on $\overline{\mathbb{R}}_+^d\backslash\{\infty\}$ and clearly homogeneous of order 1. Hence (9.11) is equivalent to

$$n\,\mathbb{P}\left((1-U_1,\ldots,1-U_d)\in n^{-1}K\right)\to\mathbb{U}(K),\ \text{as } n\to\infty, \tag{9.17}$$

for all relatively compacts $K\subset\overline{\mathbb{R}}_+^d\backslash\{\infty\}$ that is bounded away from ∞. The law (9.17) of rare events resembles (9.3) and can be used to estimate upper joint tail probabilities of $(U_1,\ldots,U_d)$ with copula C. Note that the measure $\mathbb{U}(\cdot)$ is a rescaled exponent measure but is better suited to the copula framework.

For example, by taking $K=\prod_{i=1}^d[0,w_i]$, (9.17) reduces to the limiting expression that leads to the upper *tail dependence function*,

$$\begin{aligned} b(w;C) &:= \mathbb{U}\Big(\prod_{i=1}^d[0,w_i]\Big)\\ &= \lim_{u\to0^+}\frac{\overline{C}(1-uw_i,1\le i\le d)}{u},\ \forall w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d. \end{aligned} \tag{9.18}$$

The tail dependence function (9.18) was introduced and studied in [30, 31, 36] and studied further in [35, 56]. Various tail dependence parameters used in the copula literature (see, e.g., [44]) can actually be written in terms of the tail dependence and exponent functions. Note that exponent and tail dependence functions are related through inclusion–exclusion relations. But it is worth mentioning that instead of upper orthants used in (9.18), it is often more convenient to work with the complements of lower orthants used in the exponent function (9.11). For example, as indicated in (9.12)–(9.17), the existence of the exponent function $a(\cdot;C)$ implies that the exponent function $a(\cdot;C_I)$, and thus the upper tail dependence function $b(\cdot;C_I)$ of any multivariate margin $C_I(u_i,i\in I)$ of C exist.

After copula transformation, the limit measure $\mathbb{U}(\cdot)$ has scaling property of order 1 in Cartesian coordinates. Such a scaling in Cartesian coordinates allows transformation to polar coordinates to yield a product measure, leading to the Pickands representation for the upper exponent function $a(\cdot;C)$.

Theorem 9.2. *Let* $\mathbb{S}_+^{d-1}=\{a:a=(a_1,\ldots,a_d)\in\mathbb{R}_+^d,||a||=1\}$, *where* $||\cdot||$ *denotes any norm on* $\mathbb{R}_+^d$. *For any copula* C *for which* $a(\cdot;C)$, *as defined by* (9.11), *exists, one has*

$$a(w;C)=c\int_{\mathbb{S}_+^{d-1}}\max_{1\le i\le d}\{a_i^{-1}w_i\}\mathbb{Q}(da),\ \forall\ w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d\backslash\{0\}, \tag{9.19}$$

where $c=\mathbb{U}(\{x:||x||<1\})$ *and* $\mathbb{Q}$ *is a probability measure defined on* $\mathbb{S}_+^{d-1}$ *such that* $c\int_{\mathbb{S}_+^{d-1}}a_i\mathbb{Q}(da)=1$, $1\le i\le d$. *The measure* $c\mathbb{Q}(\cdot)$, *called the spectral measure, depends only on copula* C.

Remark 9.2.

1. The Pickands representation for the exponent function is not unique. If, instead of $\mathbb{U}(\cdot)$, the measure $\mu(\cdot)$ is used in the polar coordinate transformation, then one can obtain the following representation:

$$a(w;C) = \mu\Big(\big(\prod_{i=1}^{d}[0,w_i^{-1}]\big)^c\Big),\ \forall\ w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d\backslash\{0\},$$
$$= \mu(\{w:||w||>1\})\int_{\mathbb{S}_+^{d-1}}\max_{1\le i\le d}\{a_i w_i\}\mathbb{Q}'(da).$$

   The $\mathbb{Q}'$-measure in this representation is related to the $\mathbb{Q}$-measure in (9.19) via the inverse transform.
2. The Pickands representation for the tail dependence function $b(\cdot;C)$ can be also obtained similarly or through inclusion–exclusion relations. For example,

$$b(w;C) = c\int_{\mathbb{S}_+^{d-1}}\min_{1\le i\le d}\{a_i^{-1}w_i\}\mathbb{Q}(da),\ \forall\ w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d\backslash\{0\},$$

   where $c=\mathbb{U}(\{x:||x||<1\})$.
3. Combining (9.13) and Theorem 9.2, the EV copula C_{EV} can be rewritten as

$$C_{\mathrm{EV}}(u_1,\ldots,u_d) = \exp\left\{\int_{\mathbb{S}_+^{d-1}}\ln\big(\wedge_{i=1}^d u_i^{c/a_i}\big)\,\mathbb{Q}(da)\right\}, \tag{9.20}$$

   where $c=\mathbb{U}(\{x:||x||<1\})$. Note that the spectral measure $\mathbb{Q}(\cdot)$ can be approximated by a sequence of discrete measures, and by discretizing the finite measure $\mathbb{Q}(\cdot)$, various extreme value copulas can be obtained from (9.20). For example, the Marshall–Olkin copula [43, 47, 50] can arise from a discretization of the $\mathbb{Q}$-measure in (9.20).
4. The key to establishing the representation of Pickands type is Fubini's theorem and homogeneous (scaling) properties of the Radon measure $\mathbb{U}(\cdot)$ and norm $||\cdot||$ (both with order 1). The idea can be applied to any Radon measure and function that are homogeneous of certain, possibly different orders. The norm function is used here for an obvious geometric interpretation.

The exponent function (tail dependence function) and intensity measure are related via marginal homeomorphisms.

Theorem 9.3 (Li and Sun [45]). *Let $X=(X_1,\ldots,X_d)$ be a random vector with distribution F and copula C, satisfying (9.2).*

1. *If F is MRV as defined in (9.3) with intensity measure ν, then for all $w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d\backslash\{0\}$,*

$$b(w;C)=\nu\Big(\prod_{i=1}^{d}(w_i^{-1/\alpha},\infty]\Big),\ \text{and}\ a(w;C)=\nu\Big(\big(\prod_{i=1}^{d}[0,w_i^{-1/\alpha}]\big)^c\Big).$$

2. *If the limit* (9.11) *exists and marginal distributions* $F_1,\dots,F_d$ *are regularly varying with tail index* α*, then* $F(x_1,\dots,x_d)=C(F_1(x_1),\dots,F_d(x_d))$ *is MRV with intensity measure generated by* $\nu([0,w]^c)=a((w_1^{-\alpha},\dots,w_d^{-\alpha});C)$*, for all* $w=(w_1,\dots,w_d)\in\mathbb{R}_+^d\backslash\{0\}$.

Remark 9.3.

1. Note that the measure $\mu(\cdot)$ defined by (9.14) satisfies that

$$\mu\Big(\big(\prod_{i=1}^{d}[0,w_i]\big)^c\Big)=\nu\Big(\big(\prod_{i=1}^{d}[0,w_i^{1/\alpha}]\big)^c\Big),\ \forall\ w=(w_1,\dots,w_d)\in\mathbb{R}_+^d\backslash\{0\},$$

   which implies that

$$\mu(K)=\nu(K^{1/\alpha}),\ \text{where}\ K^{1/\alpha}:=\{x^{1/\alpha}:x\in K\},$$

   for all relatively compacts $K\subset\overline{\mathbb{R}}_+^d\backslash\{0\}$ that is bounded away from 0. Both measures $\mu(\cdot)$ and $\mathbb{U}(\cdot)$ are the standardization of the intensity measure $\nu(\cdot)$ induced by marginal monotone transforms. An advantage of standardization, as discussed in [37], is its ability to handle the situations where marginal tails are not necessarily tail equivalent, in order to recover homogeneous scaling properties among different margins at multiple scales. Scaling properties are crucial in establishing semi-parametric representations (e.g., the Pickands representation) for MEV dfs and their copulas.
2. If the limit (9.11) exists and the marginal distribution F_i is regularly varying with tail index α_i, $1\le i\le d$, then $F(x_1,\dots,x_d)=C(F_1(x_1),\dots,F_d(x_d))$ is a nonstandard MRV (see page 204 of [61]) with intensity measure $\nu(\cdot)$ generated by

$$\nu([0,w]^c)=a((w_1^{-\alpha_1},\dots,w_d^{-\alpha_d});C),\ \text{for all}\ w=(w_1,\dots,w_d)\in\mathbb{R}_+^d\backslash\{0\}.$$

   That is, the intensity measure is obtained from the standardized measure $\mathbb{U}(\cdot)$ via marginal homeomorphisms that are monotone in the same direction.
3. Note that (9.12) also follows from Theorem 9.3 immediately, but in general, for a multivariate margin $C_I(u_i,i\in I)$ of C, $\emptyset\ne I\subset\{1,\dots,d\}$,

$$b(w_I;C_I)\ge\lim_{w_j\to\infty,j\notin I}b(w;C),\ \text{for}\ w_I=(w_i,i\in I).$$

   It was shown in [35] that if copula C has continuous second-order partial derivatives with nonzero tail dependence function $b(\cdot;C)$, then

$$b(w_I; C_I) = \lim_{w_j \to \infty, j \notin I} b(w; C), \text{ for } w_I = (w_i, i \in I) \tag{9.21}$$

if and only if the conditional tail dependence functions are proper distribution functions. In contrast, it is easier to work with the exponent function $a(\cdot; C)$ when multivariate margins are involved.

4. Various tail dependence parameters used in the literature can be rephrased in terms of exponent and tail dependence functions, which, in turn, can be written in terms of the intensity or spectral measure. For example, the upper orthant tail dependence parameter is defined as follows [42–44, 62]: for some subset $\emptyset \neq J \subset \{1, \cdots, d\}$,

$$\tau_J := \lim_{u \uparrow 1} \mathbb{P}(F_j(X_j) > u, \forall j \notin J \mid F_i(X_i) > u, \forall i \in J),$$

provided that the limit exists. Clearly,

$$\tau_J = \frac{b((1, \ldots, 1); C)}{b((1, \ldots, 1); C_J)} = \frac{\nu\Big(\prod_{i=1}^d (1, \infty]\Big)}{\nu\Big(\prod_{i \in J}(1, \infty] \times \mathbb{R}_+^{d-|J|}\Big)}, \emptyset \neq J \subset \{1, \cdots, d\}.$$

The expressions of exponent and tail dependence functions of Archimedean copulas were explicitly derived in [1, 3, 10, 22] (also see Propositions 2.5 and 3.3 in [35]).

Proposition 9.1. *Let $C(u; \phi) = \phi(\sum_{i=1}^d \phi^{-1}(u_i))$ be an Archimedean copula where the generator ϕ^{-1} satisfies that $\phi^{-1}(1 - 1/t)$ is regularly varying at ∞ with tail index $\beta > 1$. The upper exponent function of C are given by $a(w; C) = \left(\sum_{j=1}^d w_j^\beta\right)^{1/\beta}$.*

Proposition 9.2. *Let $C(u; \phi) = \phi(\sum_{i=1}^d \phi^{-1}(u_i))$ be an Archimedean copula with strict generator ϕ^{-1}, where ϕ is regularly varying at ∞ with tail index $\theta > 0$. The upper tail dependence function of the survival copula $\hat{C}$ are given by $b(w; \hat{C}) = \left(\sum_{j=1}^d w_j^{-1/\theta}\right)^{-\theta}$, which obeys (9.21).*

If ϕ is completely monotone, then the Archimedean copulas are the survival copulas of scale mixtures of iid standard exponentially distributed random variables with strictly positive scale mixing. In general, consider a random vector $X = (RT_1, \ldots, RT_d)$ with df F and continuous margins $F_1, \ldots, F_d$, in which the mixing variable $R > 0$ is independent of $(T_1, \ldots, T_d)$. The exponent and tail dependence functions of X can be obtained via Theorem 9.3 and Proposition A.1 of [4].

Proposition 9.3. *Let $T_{i+} = \max\{T_i, 0\}$, $1 \leq i \leq d$. Assume that the survival function of R is regular varying with tail index $\alpha > 0$, and $0 < \mathbb{E}(T_{i+}^{\alpha+\epsilon}) < \infty$, $1 \leq i \leq d$, for some $\epsilon > 0$, then the upper tail dependence and exponent functions are given by*

$$b(w;C)=\mathbb{E}\left(\bigwedge_{i=1}^{d}\frac{w_i T_{i+}^{\alpha}}{\mathbb{E}(T_{i+}^{\alpha})}\right)\ \textit{and}\ a(w;C)=\mathbb{E}\left(\bigvee_{i=1}^{d}\frac{w_i T_{i+}^{\alpha}}{\mathbb{E}(T_{i+}^{\alpha})}\right).$$

The upper exponent and tail dependence functions of the elliptical distributions with regularly varying mixing variable can be obtained via Proposition 9.3. In particular, the tail dependence function of the t distribution with degrees of freedom ν can be written as the νth moment of the minimum of a weighted, truncated, marginally standardized version of the underlying normal random vector [9]. The t tail dependence function can be derived more explicitly using the Euler representation.

Consider the d-dimensional copula C of a random vector $(U_1,\ldots,U_d)$ with standard uniform margins and continuous second-order partial derivatives. Since both $a(\cdot;C)$ and $b(\cdot;C)$ are homogeneous of order 1, the well-known Euler's homogeneous theorem implies that

$$a(w;C)=\sum_{j=1}^{d}\frac{\partial a}{\partial w_j}w_j,\ \ b(w;C)=\sum_{j=1}^{d}\frac{\partial b}{\partial w_j}w_j,\ \ \forall w=(w_1,\ldots,w_d)\in\mathbb{R}_+^d, \tag{9.22}$$

where the partial derivatives can be interpreted as conditional limiting distributions of the underlying copula C [35,56]. For example,

$$\frac{\partial b}{\partial w_j}=\lim_{u\downarrow 0}\mathbb{P}(U_i>1-uw_i,\forall i\neq j\mid U_j=1-uw_j)=\lim_{u\downarrow 0}\overline{C}_j(1-uw_i,i\neq j\mid 1-uw_j),$$

where the notation $\overline{C}_j$ refers to the conditional survival function of $(U_i:i\neq j)$ given U_j. The Euler representations of $a(\cdot;C)$ and $b(\cdot;C)$ are especially useful for the distributions whose conditional distributions are tractable.

Let $X=(X_1,\ldots,X_d)$ have the t distribution $T_{d,\nu,\Sigma}$ with ν degrees of freedom and dispersion matrix Σ. Since the increasing location-scale marginal transforms convert X to the t distribution with identical margins, assume, without loss of generality, that the margins $F_i=T_\nu$ for all $1\le i\le d$, where T_ν is the t distribution function with ν degrees of freedom, and that $\Sigma=(\rho_{ij})$ satisfies $\rho_{ii}=1$ for all $1\le i\le d$. For each fixed j, define the $(d-1)\times(d-1)$ partial correlation matrix:

$$R_j=\begin{pmatrix}
1 & \cdots & \rho_{1,j-1;j} & \rho_{1,j+1;j} & \cdots & \rho_{1,d;j}\\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots\\
\rho_{1,j-1;j} & \cdots & 1 & \rho_{j-1,j+1;j} & \cdots & \rho_{j-1,d;j}\\
\rho_{1,j+1;j} & \cdots & \rho_{j-1,j+1;j} & 1 & \cdots & \rho_{j+1,d;j}\\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\
\rho_{1,d;j} & \cdots & \rho_{j-1,d;j} & \rho_{j+1,d;j} & \cdots & 1
\end{pmatrix}$$

with $\rho_{i,k;j}=\frac{\rho_{ik}-\rho_{ij}\rho_{kj}}{\sqrt{1-\rho_{ij}^2}\sqrt{1-\rho_{kj}^2}}$, $i\neq j,k\neq j$, being the partial correlations.

Theorem 9.4 (Nikoloulopoulos et al. [56]). *Let C be the copula of a multivariate t distribution with ν d.f. and dispersion matrix $\Sigma = (\rho_{ij})$ with $\rho_{ii} = 1$ for $1 \le i \le d$.*

1. *The tail dependence function of C is given by*

$$b(w; C) = \sum_{j=1}^{d} w_j T_{d-1,\nu+1,R_j} \left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho_{ij}^2}} \left[-\left(\frac{w_i}{w_j}\right)^{-1/\nu} + \rho_{ij} \right], i \neq j \right),$$

   for all $w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d$.
2. *The t-EV copula, obtained by the EV limit of the t copula, is given by*

$$C_{EV}(u_1, \ldots, u_d; \nu) = \exp\{-a(w_1, \ldots, w_d; C)\}, \quad w_j = -\log u_j, \ j = 1, \ldots, d,$$

   with exponent

$$a(w; C) = \sum_{j=1}^{d} w_j T_{d-1,\nu+1,R_j} \left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho_{ij}^2}} \left[\left(\frac{w_i}{w_j}\right)^{-1/\nu} - \rho_{ij} \right], i \neq j \right). \tag{9.23}$$

Remark 9.4.

1. Since T_ν has a regularly right tail with tail index ν, it follows from (9.23) and Theorem 9.3 that the truncated multivariate t distribution on $\mathbb{R}_+^d$ is regularly varying with intensity measure

$$\nu\left(\left(\prod_{i=1}^{d} [0, w_i] \right)^c \right) := \sum_{j=1}^{d} w_j^{-\alpha} T_{d-1,\nu+1,R_j} \left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho_{ij}^2}} \left[\left(\frac{w_i}{w_j}\right) - \rho_{ij} \right], i \neq j \right).$$

   That is, the limiting measure $\nu(\cdot)$ is driven by $(d-1)$-dimensional t distributions along all the axes. Such a recursive scheme holds precisely due to scaling properties of the limiting measures and the fact that the conditional distributions of a t distribution are also t distributions that belong to the same distribution family.
2. The explicit expressions obtained in Theorem 9.4 can be used to derive two well-known distributions as special cases [56].
   (a) Under some scaling conditions, $C_{\mathrm{EV}}(\cdot; \nu)$ converges weakly to the Hüsler-Reiss copula (see [29]) as $\nu \to \infty$.
   (b) As $\nu \to 0$, $C_{\mathrm{EV}}(\cdot; \nu)$ converges weakly to a Marshall–Olkin distribution (see [50]) that satisfies some linear constraints.

The limit to the Hüsler-Reiss copula as $\nu \to \infty$ is expected because the Hüsler-Reiss copula is the extreme value copula of normal random samples. On the other hand, the limit to a Marshall–Olkin distribution as $\nu \to 0$ is somewhat surprising and shows that regularity inherited from the t distribution breaks down as $\nu \to 0$ and some common cause dependence pattern emerges.

3. The unique feature of the copula method for multivariate regular variation is the rescaled exponent measure $\mathbb{U}(\cdot)$ with scaling property of order 1. Such a scaling property leads to the Euler representation for the intensity measure that serves as a total limiting probability law for multivariate extremes.

Using the Euler representation, the explicit expression of the exponent and tail dependence functions for skew-t distributions have been obtained in [57].

9.3 Tail Density of Multivariate Regular Variation

The tail dependence (or exponent) function and intensity measure are equivalent in extremal dependence analysis in the sense that the Radon measure generated by the tail dependence function is a marginally rescaled version of the intensity measure. Note, however, that the tail dependence function and intensity measure are cumulative in nature. A notion that describes extremal dependence locally is the tail density of multivariate regular variation studied in [14]. Consider again a distribution F with tail equivalent margins (9.2) and a norm $||\cdot||$ on $\mathbb{R}^d$.

Theorem 9.5 (de Haan and Resnick [14]). *Assume the density f of F exists and the margins F_i, $1 \le i \le d$, are regularly varying with tail index $\alpha > 0$. If $\frac{f(tx)}{t^{-d}\overline{F}_1(t)} \to \lambda(x) > 0$, as $t \to \infty$, on $\overline{\mathbb{R}}_+^d \backslash \{0\}$ and uniformly on $\mathbb{S}_+^{d-1} = \{x \ge 0 : ||x|| = 1\}$ where $\lambda(\cdot)$ is bounded, then, for any $x \in \mathbb{R}_+^d \backslash \{0\}$,*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{\overline{F}_1(t)} = \nu([0, x]^c) = \int_{[0,x]^c} \lambda(y)dy,$$

with homogeneous property that $\lambda(tx) = t^{-\alpha-d}\lambda(x)$ for $t > 0$.

The tail density $\lambda(\cdot)$ in Theorem 9.5 is especially tractable for the distributions that are specified by densities. Correspondingly, the tail densities for copulas, introduced in [46], can be applied to analyzing extremal dependence of the copulas that are specified only explicitly by densities, such as the t copula, and vine copulas that are built from bivariate linking copulas using local dependence properties [5, 6, 39, 40]. For the copula C of F, the *upper tail density function*, denoted by $\lambda(\cdot; C)$, is defined as follows:

$$\lambda(w; C) := \lim_{u\to 0} \frac{D_w \overline{C}(1 - uw_i; 1 \le i \le d)}{u}, \; w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d \backslash \{0\}, \tag{9.24}$$

provided that the limit exists, where $D_w = \frac{\partial^d}{\partial w_1 \cdots \partial w_d}$ denotes the d-order partial differentiation operator with respect to $w_1, \ldots, w_d$. It follows from (9.24) that the tail density function is homogeneous of order $1-d$; that is, $\lambda(tw; C) = t^{1-d}\lambda(w; C)$ for any $t > 0$ and $w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d$.

Assume throughout this section that the density $c(\cdot)$ of copula C exists. Assume furthermore that as $u \to 0$,

$$u^{d-1}c(1 - uw_i, 1 \le i \le d) \text{ converges for any } w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d \backslash \{0\}$$
$$\text{and converges uniformly on } \mathbb{S}_+^{d-1} = \{w \in \mathbb{R}_+^d : ||w|| = 1\}. \tag{9.25}$$

Let $(U_1, \ldots, U_d)$ be distributed with df C. Clearly (9.25) is equivalent to the partial derivative of order d for the ratio

$$\frac{\mathbb{P}(\cup_{i=1}^d \{U_i > 1 - uw_i\})}{u} \text{ converges for any } w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d \backslash \{0\}$$
$$\text{and converges uniformly on } \mathbb{S}_+^{d-1} = \{w \in \mathbb{R}_+^d : ||w|| = 1\}, \tag{9.26}$$

as $u \to 0$. Most copulas that are specified by densities satisfy this technical condition on uniform convergence. The uniform convergence condition (9.25) or (9.26) clearly ensures the exchange of limits on $\mathbb{S}_+^{d-1}$ that leads to that

$$\lambda(w; C) = \lim_{u \to 0} u^{d-1}c(1 - uw_i, 1 \le i \le d) = \frac{\partial^d b(w; C)}{\partial w_1 \cdots \partial w_d} = (-1)^{d-1}\frac{\partial^d a(w; C)}{\partial w_1 \cdots \partial w_d}, \tag{9.27}$$

for all $w \in \mathbb{S}_+^{d-1}$. The set on which (9.27) remains equal can be extended to $\mathbb{R}_+^d \backslash \{0\}$ from $\mathbb{S}_+^{d-1}$ by using the homogeneous property. Since $b(\cdot; C)$ is homogeneous of order 1, $\frac{\partial^d b(w;C)}{\partial w_1 \cdots \partial w_d}$ must be homogeneous of order $1-d$. On the other hand, the tail density function $\lambda(\cdot; C)$ is also homogeneous of order $1-d$. Thus, (9.27) holds for any $w \in \mathbb{R}_+^d \backslash \{0\}$ because of the polar coordinate representation $w = r\theta$, where $r = ||w|| > 0$ and $\theta = w/||w|| \in \mathbb{S}_+^{d-1}$.

Remark 9.5. It can be shown (see [14]) that under the condition specified in Theorem 9.5, $\frac{f(tx)}{t^{-d}\overline{F}_1(t)} \to \lambda(x)$, as $t \to \infty$, uniformly on $\{x \ge 0 : ||x|| = 1\}$ if and only if for any $\delta > 0$,

$$\frac{f(tx)}{t^{-d}\overline{F}_1(t)} \to \lambda(x), \text{ uniformly on } \{x \ge 0 : ||x|| > \delta\}.$$

Similarly, due to the homogeneous property, the uniform convergence in (9.25) is equivalent to that $u^{d-1}c(1 - uw_i, 1 \le i \le d)$ converges uniformly on $\{w \in \mathbb{R}_+^d : ||w|| > \delta\}$, for any $\delta > 0$.

Theorem 9.6. *Let F denote a distribution with tail equivalent, continuous margins F_i, $1 \le i \le d$. Suppose that F admits the density $f(\cdot)$ and its copula C admits the density $c(\cdot)$.*

1. *If the marginal density f_i of F_i, $1 \le i \le d$, is regularly varying with tail Index $\alpha + 1$, $\alpha > 0$, and the copula C of F satisfies the condition (9.26), then F is MRV with tail density $\lambda(\cdot)$ that is related to the upper tail density $\lambda(\cdot; C)$ of C as follows:*

$$\begin{aligned}\lambda(w_1, \ldots, w_d) &= \alpha^d (w_1 \cdots w_d)^{-\alpha-1} \lambda(w_1^{-\alpha}, \ldots, w_d^{-\alpha}; C) \\ &= \lambda^U(w_1^{-\alpha}, \ldots, w_d^{-\alpha}; C) |J(w_1^{-\alpha}, \ldots, w_d^{-\alpha})|, \end{aligned} \tag{9.28}$$

   where $J(w_1^{-\alpha}, \ldots, w_d^{-\alpha})$ is the Jacobian determinant of the homeomorphic transform $y_i = w_i^{-\alpha}$, $1 \le i \le d$.
2. *If F is MRV with tail density $\lambda(\cdot)$ as specified in Theorem 9.5, then the copula C of F admits the tail density as defined in (9.24).*

Proof. Consider

$$f(tx) = c(F_1(tx_1), \ldots, F_d(tx_d)) \prod_{i=1}^{d} f_i(tx_i),\ t > 0,\ x = (x_1, \ldots, x_d) \in \mathbb{R}_+^d. \tag{9.29}$$

Because of the regularly varying property of the tail equivalent margins, for sufficiently large $t > 0$, we have

$$f_i(tx_i) = t^{-\alpha-1}(x_i^{-\alpha-1} L_i(tx_i)) \approx t^{-\alpha-1} L_1(t) x_i^{-\alpha-1},\ x_i > 0,\ 1 \le i \le d,$$

where L_i, $1 \le i \le d$, are all slowly varying and $L_i(t)/L_1(t) \to 1$ as $t \to \infty$. Due to Karamata's theorem (see Theorem 2.1 in [61]), the margin F_i, $1 \le i \le d$, is regularly varying with tail index α and

$$F_i(tx_i) \approx 1 - \alpha^{-1}(tx_i) f_i(tx_i) \approx 1 - \alpha^{-1} t^{-\alpha} L_1(t) x_i^{-\alpha},\ x_i > 0,\ 1 \le i \le d.$$

Plug these tail estimates into (9.29) with $u := \alpha^{-1} t^{-\alpha} L_1(t) \to 0$ as $t \to \infty$, and one has

$$\begin{aligned}\frac{f(tx)}{t^{-d}\overline{F}_1(t)} &= \frac{\alpha^d t^{-d} u^d (\prod_{i=1}^d x_i^{-\alpha-1}) c(1 - u x_1^{-\alpha}, \ldots, 1 - u x_d^{-\alpha})}{t^{-d} u} \\ &= \alpha^d \Big(\prod_{i=1}^{d} x_i^{-\alpha-1} \Big) u^{d-1} c(1 - u x_1^{-\alpha}, \ldots, 1 - u x_d^{-\alpha}). \end{aligned} \tag{9.30}$$

(1) If (9.26) holds, then it follows from Remark 9.5 that

$$\frac{f(tw^{-1/\alpha})}{t^{-d}\overline{F}_1(t)} = \alpha^d \Big(\prod_{i=1}^{d} w_i^{1+1/\alpha}\Big) u^{d-1} c(1-uw_1, \ldots, 1-uw_d)$$

converges uniformly on $\{w \in \mathbb{R}_+^d : ||w|| > \delta\}$ as $t \to \infty$ or equivalently $u \to 0$. By selecting $0 < \delta < 1$, this shows that $\frac{f(tx)}{t^{-d}\overline{F}_1(t)}$ converges uniformly on $\{x \geq 0 : ||x|| = 1\}$. By Theorem 9.5, F is regularly varying with intensity measure ν and tail density λ, and for any $x \in \overline{\mathbb{R}}_+^d \backslash \{0\}$,

$$\lim_{t\to\infty} \frac{1-F(tx)}{\overline{F}_1(t)} = \nu([0,x]^c) = \int_{[0,x]^c} \lambda(y)dy.$$

Since ν is a Radon measure, one has that $\nu((x,\infty]) = \int_{(x,\infty]} \lambda(y)dy$. It follows from Theorem 9.3 that for any $w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d$,

$$b(w_1^{-\alpha}, \ldots, w_d^{-\alpha}; C) = \int_{(w,\infty]} \lambda(y)dy.$$

By taking the derivatives on both sides with respect to $w_1, \ldots, w_d$, (9.28) follows.

(2) It follows from (9.30) that

$$u^{d-1} c(1-ux_1^{-\alpha}, \ldots, 1-ux_d^{-\alpha}) = \alpha^{-d} \Big(\prod_{i=1}^{d} x_i^{\alpha+1}\Big) \frac{f(tx)}{t^{-d}\overline{F}_1(t)}$$

converges uniformly on $\{x \in \mathbb{R}_+^d : ||x|| > \delta\}$ as $t \to \infty$ or equivalently $u \to 0$. Again, by selecting $0 < \delta < 1$, this shows that $u^{d-1}c(1-uw_1, \ldots, 1-uw_d)$ converges uniformly on $\{w \geq 0 : ||w|| = 1\}$. Thus the tail density $\lambda(\cdot; C)$ exists and is given by (9.24) or (9.27). □

Remark 9.6.

1. Theorem 9.6(1) was obtained in [46] under a slightly stronger condition on uniform convergence. The current Theorem 9.6 shows the tail density of copula C is a rescaled version of the tail density of the multivariate regular varying distribution F with copula C and two tail densities are equivalent in analyzing local extremal dependence properties.
2. Since the tail dependence function is equal to zero if some variables take zero. then $b(w; C) = \int_0^{w_1} \cdots \int_0^{w_d} \lambda(x; C)dx$, for any $w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d \backslash \{0\}$. Note that the tail density of multivariate regular variation is the Radon-Nikodym derivative of the intensity measure with respect to the Lebesgue measure, whereas the tail density of a copula is the Radon-Nikodym derivative of the

rescaled exponent measure $\mathbb{U}(\cdot)$ defined in (9.16) with respect to the Lebesgue measure. That is, the intensity measure and rescaled exponent measure on any Borel measurable set $B \subset \mathbb{R}_+^d \backslash \{0\}$ can be written as integrals of respective tail densities over B. In risk analysis, remote critical sets (failure, overflow, or out-of-compliance regions) are usually the sets B that are neither upper nor lower orthant sets [1, 3, 17, 18, 34, 38, 48, 49, 64, 66, 67].

Example 9.1.

1. Let $C(u;\phi) = \phi(\sum_{i=1}^d \phi^{-1}(u_i))$ be an Archimedean copula where the generator ϕ^{-1} satisfies that $\phi^{-1}(1-1/t)$ is regularly varying at ∞ with tail index $\beta > 1$. It follows from Proposition 9.1 that the upper tail density is given by

$$\lambda(w;C) = \prod_{i=2}^{d}((i-1)\beta - 1)\Big(\prod_{i=1}^{d} w_i\Big)^{\beta-1}\Big(\sum_{i=1}^{d} w_i^{\beta}\Big)^{-d+1/\beta}.$$

2. Let $C(u;\phi) = \phi(\sum_{i=1}^d \phi^{-1}(u_i))$ be an Archimedean copula with strict generator ϕ^{-1}, where ϕ is regularly varying at ∞ with tail index $\theta > 0$. It follows from Proposition 9.2 that the upper tail density of the survival copula $\hat{C}$ is given by

$$\lambda(w;\hat{C}) = \prod_{i=2}^{d}\Big(1+\frac{i-1}{\theta}\Big)\Big(\prod_{i=1}^{d} w_i\Big)^{-1-1/\theta}\Big(\sum_{i=1}^{d} w_i^{-1/\theta}\Big)^{-\theta-d}.$$

   Examples 9.1 (1) and (2) are known at least since [11].
3. Consider a d-dimensional symmetric t distribution $T_{d,\nu,\Sigma}$ with mean 0 and its density function:

$$f_t(x;\nu,\Sigma) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{d/2}}|\Sigma|^{-\frac{1}{2}}\Big[1+\frac{1}{\nu}(x^\top\Sigma^{-1}x)\Big]^{-\frac{\nu+d}{2}}$$

   where $x = (x_1,\cdots,x_d) \in \mathbb{R}^d$, $\nu > 0$ is the degrees of freedom, and $\Sigma = (\rho_{ij})$ is a $d \times d$ symmetric dispersion matrix. If a random vector X has the t distribution $T_{d,\nu,\Sigma}$, then $X \overset{d}{=} \sqrt{R}(Z_1,\ldots,Z_d)$, where $(Z_1,\ldots,Z_d)$ has a multivariate normal distribution $N(0,\Sigma)$, and the scale variable R, independent of $(Z_1,\ldots,Z_d)$, has an inverse Gamma distribution, which is known to be regularly varying with tail index $\nu/2$ [53]. The upper tail density function of a multivariate t copula is given below [46]:

$$\lambda(w;C) = |\Sigma|^{-\frac{1}{2}}\nu^{1-d}\frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu+1}{2})\pi^{(d-1)/2}}\frac{[(w^{-\frac{1}{\nu}})^\top\Sigma^{-1}w^{-\frac{1}{\nu}}]^{-\frac{\nu+d}{2}}}{\prod_{i=1}^{d} w_i^{\frac{\nu+1}{\nu}}}.$$

   In contrast to the Euler representations presented in Theorem 9.4, the tail density of t copula is more explicit and shows geometrically the elliptical feature inherited from the t density function.

Let $X_k = (X_{1,k}, \ldots, X_{d,k})$, $1 \le k \le n$, be iid samples that form a sample cloud in $\mathbb{R}^d$ with df F and its copula C. The tail density of df F (or copula C) is a limiting conditional density of X_n given that one component (e.g., $X_{1,n}$) exceeds large thresholds. In general, one can introduce the tail density of F (or its copula C) given that a homogeneous function (e.g., linear function) of components $X_{1,n}, \ldots, X_{d,n}$ exceeds large thresholds [2]. Geometrically, such a tail density of the sample cloud $(X_k, 1 \le k \le n)$ describes distributional stability patterns of F (or its copula C) when a hypersurface of $(X_{1,n}, \ldots, X_{d,n})$ moves away from the center of the sample cloud. That is, the tail density describes distributional patterns of extreme data points near the boundary of the sample cloud under the stability condition of regular variation near the boundary. Tail risk, as measured by a functional of boundary measures of sample clouds, can then be estimated using the tail density approach.

9.4 Concluding Remarks

Multivariate regular variation on cone $\overline{\mathbb{R}}_+^d \backslash \{0\}$ yields laws of rare events that provide equivalent, tractable limit representations for the multivariate maximum domain of attraction (9.1). In this paper, we discussed a copula approach for multivariate regular variation that is based on the upper exponent function (9.11) and related vague convergences. As illustrated in this paper, standardizing univariate margins is necessary to reveal scaling properties that facilitate to establishing semi-parametric representations for multivariate extremes. The advantages of our copula method include the Euler homogeneous representation. We have not discussed the estimation and inference issues. References on estimation of rare events in MRV models and other models featuring EV copulas include, among many others, [19, 20, 25].

We have also not discussed multivariate regular variations on subcones of $\overline{\mathbb{R}}_+^d \backslash \{0\}$, which usually lead to hidden regular variation and conditioned limit laws [26, 52, 60, 61]. Whether or not the copula approach can be effectively applied to studying extremal dependence among multivariate margins (such as hidden regular variation) is still an open issue and currently under active investigation [27, 28, 32].

Acknowledgments The author would like to sincerely thank the reviewer for his/her detailed comments, which led to an improvement of the presentation of this paper and also for pointing out a few references related to this work. The author also likes to thank Professor Piotr Jaworski for the invitation to contribute to this Volume.
Furthermore, he acknowledges that his research is supported by NSF grants CMMI 0825960 and DMS 1007556.

References

1. Alink, S., Löwe, M., Wüthrich, M.V.: Diversification of aggregate dependent risks. Insur. Math. Econ. **35**, 77–95 (2004)
2. Balkema, G., Embrechts, P.: High Risk Scenarios and Extremes: A Geometric Approach. European Mathematical Society, Zürich (2007)
3. Barbe, P., Fougères, A.L., Genest, C.: On the tail behavior of sums of dependent risks. ASTIN Bull. **36**(2), 361–374 (2006)
4. Basrak, B., Davis, R.A., Mikosch, T.: Regular variation of GARCH processes. Stoch. Proc. Appl. **99**, 95–116 (2002)
5. Bedford, T., Cooke, R.M.: Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. Intell. **32**, 245–268 (2001)
6. Bedford, T., Cooke, R.M.: Vines: a new graphical model for dependent random variables. Ann. Stat. **30**, 1031–1068 (2002)
7. Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J.: Statistics of Extremes: Theory and Applications. Wiley, New York (2004)
8. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation. Cambridge University Press, Cambridge (1987)
9. Chan, Y., Li, H.: Tail dependence for multivariate t-copulas and its monotonicity. Insur. Math. Econ. **42**, 763–770 (2008)
10. Charpentier, A., Segers, J.: Tails of multivariate Archimedean copulas. J. Multivariate Anal. **100**, 1521–1537 (2009)
11. Coles, S.G., Tawn, J.A.: Modelling extreme multivariate events. J. R. Stat. Soc., Ser. B **53**(2), 377–392 (1991)
12. de Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer, New York (2006)
13. de Haan, L., Resnick, S.I.: Limit theory for multivariate sample extremes. Z. Wahrscheinlichkeitstheor. Verw. Gebiete **40**, 317–337 (1977)
14. de Haan, L., Resnick, S.I.: On regular variation of probability densities. Stoch. Process. Appl. **25**, 83–95 (1987)
15. Drees, H., Huang, X.: Best attainable rates of convergence for estimators of the stable tail dependence function. J. Multivariate Anal. **64**, 25–47 (1998)
16. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modeling Extremal Events for Insurance and Finance. Springer, New York (1997)
17. Embrechts, P., Lambrigger, D.D., Wüthrich, M.V.: Multivariate extremes and the aggregation of dependent risks: examples and counterexamples. Extremes **12**(2), 107–127 (2009)
18. Embrechts, P., Nešlehová, J., Wüthrich, M.V.: Additivity properties for value-at-risk under Archimedean dependence and heavy-tailedness. Insur. Math. Econ. **44**(2), 164–169 (2009)
19. Einmahl, J.H.J., de Haan, L., Huang, X.: Estimating a multidimensional extreme-value distribution. J. Multivariate Anal. **47**(1), 35–47 (1993)
20. Einmahl, J.H.J., de Haan, L., Piterbarg, V.I.: Nonparametric estimation of the spectral measure of an extreme value distribution. Ann. Stat. **29**(5), 1401–1423 (2001)
21. Falk, M., Hüsler, J., Reiss, R.-D.: Laws of Small Numbers: Extremes and Rare Events. Birkhäuser, Basel (2011)
22. Genest, C., Rivest, L.-P.: A characterization of Gumbel's family of extreme value distributions. Stat. Probab. Lett. **8**, 207–211 (1989)
23. Gnedenko, B.V.: Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. **44**, 423–453 (1943)
24. Gudendorf, G., Segers, J.: Extreme-value copulas. In: Jaworski, P., Durante, F., Karl Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Lecture Notes in Statistics, vol. 198. Springer, Berlin (2010)
25. Gudendorf, G., Segers, J.: Nonparametric estimation of an extreme value copula in arbitrary dimensions. J. Multivariate Anal. **102**(1), 37–47 (2011)

26. Heffernan, J.E., Resnick, S.I.: Hidden regular variation and the rank transform. Adv. Appl. Prob. **37**(2), 393–414 (2005)
27. Hua, L., Joe, H.: Tail order and intermediate tail dependence of multivariate copulas. J. Multivariate Anal. **102**, 1454–1471 (2011)
28. Hua, L., Joe, H., Li, H.: Relations between hidden regular variation and tail order of copulas (under review). Technical Report, Department of Mathematics, Washington State University (2012)
29. Hüsler, J., Reiss, R.-D.: Maxima of normal random vectors: between independence and complete dependence. Stat. Probab. Lett. **7**(4), 283–286 (1989)
30. Jaworski, P.: On uniform tail expansions of bivariate copulas. Appl. Math. **31**(4), 397–415 (2004)
31. Jaworski, P.: On uniform tail expansions of multivariate copulas and wide convergence of measures. Appl. Math. **33**(2), 159–184 (2006)
32. Jaworski, P.: Tail behaviour of copulas, chapter 8. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Lecture Notes in Statistics, vol. 198. Springer, New York (2010)
33. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
34. Joe, H., Li, H.: Tail risk of multivariate regular variation. Meth. Comput. Appl. Probab. **13**, 671–693 (2011)
35. Joe, H., Li, H., Nikoloulopoulos, A.K.: Tail dependence functions and vine copulas. J. Multivariate Anal. **101**, 252–270 (2010)
36. Klüppelberg, C., Kuhn, G., Peng, L.: Semi-parametric models for the multivariate tail dependence function – the asymptotically dependent. Scand. J. Stat. **35**(4), 701–718 (2008)
37. Klüppelberg, C., Resnick, S.: The Pareto copula, aggregation of risks, and the emperor's socks. J. Appl. Probab. **45**(1), 67–84 (2008)
38. Kortschak, D., Albrecher, H.: Asymptotic results for the sum of dependent non-identically distributed random variables. Meth. Comput. Appl. Probab. **11**, 279–306 (2009)
39. Kurowicka, D., Cooke, R.: Uncertainty Analysis with High Dimensional Dependence Modelling. Wiley, New York (2006)
40. Kurowicka, D., Joe, H.: Dependence Modeling: Vine Copula Handbook. World Scientific, Singapore (2011)
41. Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and Related Properties of Random Sequences and Processes. Springer, New York (1983)
42. Li, H.: Duality of the multivariate distributions of Marshall-Olkin type and tail dependence. Commun. Stat. Theory Meth. **37**(11), 1721–1733 (2008)
43. Li, H.: Tail dependence comparison of survival Marshall-Olkin copulas. Meth. Comput. Appl. Probab. **10**, 39–54 (2008)
44. Li, H.: Orthant tail dependence of multivariate extreme value distributions. J. Multivariate Anal. **100**, 243–256 (2009)
45. Li, H., Sun, Y.: Tail dependence for heavy-tailed scale mixtures of multivariate distributions. J. Appl. Prob. **46**(4), 925–937 (2009)
46. Li, H., Wu, P.: Extremal dependence of copulas: a tail density approach. J. Multivariate Anal. **114**, 99–111 (2012)
47. Mai, J.-F., Scherer, M.: Reparameterizing Marshall-Olkin copulas with applications to sampling. J. Stat. Comput. Simul. **81**, 1–23 (2009)
48. Mainik, G., Rüschendorf, L: On optimal portfolio diversification with respect to extreme risks. Finance Stoch. **14**(4), 593–623 (2010)
49. Mainik, G., Rüschendorf, L.: Ordering of multivariate risk models with respect to extreme portfolio losses. Stat. Risk Model. **29**(1), 73–106 (2012)
50. Marshall, A.W., Olkin, I.: A multivariate exponential distribution. J. Am. Stat. Assoc. **2**, 84–98 (1967)
51. Marshall, A.W., Olkin, I.: Domains of attraction of multivariate extreme value distributions. Ann. Probab. **11**, 168–177 (1983)

52. Maulik, K., Resnick, S.: Characterizations and examples of hidden regular variation. Extremes **7**, 31–67 (2004)
53. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, Princeton (2005)
54. Meerschaert, M.M., Scheffler, H.-P.: Limit Distributions for Sums of Independent Random Vectors. Wiley, New York (2001)
55. Nelsen, R.B.: An introduction to Copulas. Springer, New York (2006)
56. Nikoloulopoulos, A.K., Joe, H., Li, H.: Extreme value properties of multivariate t copulas. Extremes **12**, 129–148 (2009)
57. Padoan, S.: Multivariate extreme models based on underlying skew-t and skew-normal distributions. J. Multivariate Anal. **102**, 977–991 (2011)
58. Pickands, J.: III: Multivariate negative exponential and extreme value distributions. Unpublished manuscript, University of Pennsylvania (1980)
59. Resnick, S.: Extreme Values, Regularly Variation, and Point Processes. Springer, New York (1987)
60. Resnick, S.: Hidden regular variation, second order regular variation and asymptotic independence. Extremes **5**, 303–336 (2002)
61. Resnick, S.: Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer, New York (2007)
62. Schmidt, R.: Tail dependence for elliptically contoured distributions. Math. Meth. Oper. Res. **55**, 301–327 (2002)
63. Seneta, E.: Regularly Varying Functions. Springer, New York (1976)
64. Wüthrich, M.V.: Asymptotic Value-at-Risk estimates for sums of dependent random variables. ASTIN Bull. **33**(1), 75–92 (2003)
65. Yakimiv, A.L.: Probabilistic Applications of Tauberian Theorems. Kroninklijke Brill NV, Leiden (2005)
66. Zhou, C: Dependence structure of risk factors and diversification effects. Insur. Math. Econ. **46**(3), 531–540 (2010)
67. Zhu, L., Li, H.: Asymptotic analysis of conditional tail expectations. North American Actuarial J. **16**(3), 350–363 (2012)

Chapter 10
CIID Frailty Models and Implied Copulas

Jan-Frederik Mai, Matthias Scherer, and Rudi Zagst

Abstract A unified stochastic framework for all portfolio default models with conditionally independent and identically distributed (CIID) default times is presented. Desirable statistical properties of dependent default times are introduced in an axiomatic manner and related to the unified framework. It is shown how commonly used models, stemming from quite different mathematical and economic motivations, can be translated into a multivariate frailty model. After a discussion of popular specifications, two new models are introduced. The vector of default times in the first approach has an Archimax survival copula. The second innovation is capable of producing default pattern with interesting statistical properties. The motivation for the latter approach is to add an additional source of jump frailty to a classical intensity-based approach. An approximation of the portfolio-loss distribution is available in all cases.

10.1 Introduction

Following the seminal work of [57], various related portfolio default models have recently been proposed, see, e.g., [4, 8, 26, 29, 32, 35, 38, 44, 53] to provide some examples. Even though these papers use diverse economic motivations, rely

J.-F. Mai
XAIA Investment GmbH, Sonnenstraße 19, 80331 München, Germany
e-mail: jan-frederik.mai@xaia.com

M. Scherer (✉) · R. Zagst
Chair of Mathematical Finance, Technische Universität München, Munich, Germany
e-mail: scherer@tum.de; zagst@tum.de

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_10,

on alternative mathematical techniques,[1] and focus on different applications,[2] all models share as common ground a large homogeneous portfolio approximation, providing a convenient tool for applications that require the loss distribution of some large portfolio.[3] In this paper, a unified stochastic framework for all models in this spirit is constructed. Such a unified treatment provides several advantages:

- The mathematical structure behind this class of models becomes transparent. Instead of relying on specific distributional assumptions and related mathematical concepts, we provide as a generic framework a multivariate frailty model that uses the classical theorems of de Finetti and Glivenko–Cantelli as tools to obtain the portfolio-loss distribution. In contrast to several of the aforementioned examples, the present construction is consistent with respect to time and does not rely on some fix maturity (or a discrete number of maturities).
- Various statistical properties have been investigated with respect to the implied dependence structure as well as with respect to the implied portfolio-loss distribution. Having a unified framework at hand, an objective comparison of alternative model specifications is facilitated. We axiomatically define a list of (desirable) statistical properties and investigate the unified framework in this regard.
- Based on a generic framework, it is often easier to understand the mathematical concept behind a generalization of some model, e.g. to hierarchical dependence structures. Hence, there is a fair chance that one can transfer the idea of a generalization from one class of models to some other. Moreover, for two concrete cases, we show how given models can be combined to a framework that inherits all desirable statistical features of the building blocks. We show that it is even possible to combine alternative models over time, using them as some sort of local correlation model. Finally, we obtain a deeper understanding of how far we can stretch the limits of conditionally independent and identically distributed (CIID)-models and, related, what model generalizations come at the price of losing the mathematical viability.

Throughout we consider a portfolio of d defaultable assets and let $(\tau_1, \ldots, \tau_d)$ denote the vector of their default times. Both applications, the pricing of portfolio credit derivatives as well as risk management of credit portfolios, require the distribution of the accumulated loss within the reference portfolio up to time t. Currently, one of the most prominent applications in the context of portfolio credit derivatives is the pricing of collateralized debt obligations (CDOs). A CDO can be seen as an insurance contract for certain loss pieces of a credit portfolio. A convenient, and for sufficiently large portfolios widely used, assumption is a homogeneous portfolio

[1] The starting point might be a multivariate structural-default model, a certain dependence structure/copula for the vector of default times, a frailty model, or some latent-factor construction.

[2] The most important ones being risk-management and the pricing of portfolio derivatives.

[3] We focus on the pricing of insurance premia for tranches of a credit portfolio. Note, however, that applications to other insurance portfolios can be treated similarly.

structure with respect to recovery rates and portfolio weights. This allows to express the premium and default leg of the CDO's tranches as options on the (relative) portfolio-loss process $\{L_t\}_{t\geq 0}$, defined as $L_t := \frac{1}{d}\sum_{i=1}^{d} \mathbb{1}_{\{\tau_i \leq t\}}$, for $t \geq 0$. From a mathematical perspective, it is required to compute expectations of the form:

$$\mathbb{E}[f(L_t)] = \int_{[0,1]} f(x)\, \mathbb{P}(L_t \in dx), \quad f \text{ non-linear (collar type)},$$

where f depends on the considered tranche and the recovery rate. Hence, it is important to construct the vector of default times in such a way that the distribution of L_t can be identified or, at least, efficiently approximated. Due to the large dimensionality of the problem,[4] one has to accept simplifying assumptions to circumvent time-consuming Monte-Carlo techniques. In this regard, a popular class of models is based on the following ansatz: there is a market factor M, conditioned on which all default times are iid with distribution function $t \mapsto F_t := \text{function}(M, t)$. The core motivation for these models is to approximate the distribution of L_t by the (more tractable) distribution of the market factor M. The seminal model in this spirit is [38, 57], specifying M as a normal random variable, which results in a Gaussian dependence structure. Since this copula has several drawbacks, e.g., zero tail dependence, symmetric dependence pattern, and an insufficient fit to quoted CDO spreads, several authors extended the approach to other market factors.[5] More dynamic models are obtained when the market factor $M = \{M_t\}_{t\geq 0}$ is a nontrivial stochastic process and $F_t := \text{function}(M_t)$. Such a model is proposed by [44] with $\{M_t\}_{t\geq 0}$ being a Lévy subordinator.[6]

From a practical perspective, a calibration of the model typically relies on market quotes of (a) portfolio CDS and single-name CDS and (b) CDO tranche spreads. Considering (a), these are not affected by the dependence structure between the default times but do depend on the respective univariate default probabilities. Hence, the required term-structures of univariate default probabilities can be extracted. Considering (b), after having fixed the marginal default probabilities, spreads of

[4] A typical convention for credit derivatives is $d = 125$, insurance portfolios are often larger.

[5] For instance, [32] use a Student t-distribution, [29, 35] a Normal Inverse Gaussian (NIG) distribution, and [4] a general infinitely divisible distribution. In a related fashion, [53] assumes a positive random variable as market factor and constructs the model in such a way that the default times have an Archimedean survival copula. However, M is a single random variable in all aforementioned models, which equals the random parameter of a parametric family of distribution functions.

[6] The resulting survival copula of $(\tau_1, \ldots, \tau_d)$ is of Marshall–Olkin kind, see [43]. The Marshall–Olkin distribution is well studied and has several desirable properties for dependent defaults: an interpretation as a frailty model, asymmetric tail dependencies, and a singular component, i.e. positive joint default probabilities. Hence, Marshall–Olkin distributions have already been proposed for credit- and insurance-risk applications by [28, 40]. However, it is well known that the Marshall–Olkin distribution is characterized by the lack-of-memory property, see, e.g., [7, 27, 47]. This implies a somewhat unrealistic assumption for dependent defaults, since it excludes direct contagion effects.

the different tranches of a CDO can be used to calibrate the dependence parameters of the model. For this, it is very convenient if the model's dependence parameters do not affect the (already fixed) marginal default probabilities, i.e. the model allows for a separation of dependence structure from default probabilities. Such a separation naturally invokes a copula model. When the model is to be estimated to observed losses, it is crucial to explicitly know the model's dependence structure. Thus, we are especially interested in models whose copula can be identified explicitly.

Besides the generic frailty model and the investigation of its statistical properties, we present two new multivariate default models with interesting statistical properties. Both models allow for a convenient approximation of L_t by the distribution of the market factor and can thus be implemented without Monte-Carlo simulation. The first ansatz is based on a scale mixture of Lévy processes. The resulting survival copula of $(\tau_1, \ldots, \tau_d)$ is a scale mixture of Marshall–Olkin copulas, constituting a proper subclass of Archimax copulas, see, e.g., [39, p. 253]. The second extension is based on processes of CGMY-type, see [15]. This model incorporates stylized facts such as default clusters and excess clustering. It can be considered as an extension of a classical intensity-based ansatz in the spirit of [23], when an additional source of frailty—a latent Lévy subordinator—is present.

The remaining paper is organized as follows: a general probabilistic framework for latent one-factor models and a review of commonly used examples (reformulated as frailty models) is given in Sect. 10.2. Two new models are introduced and discussed in Sects. 10.3 and 10.4. Possible generalizations of the models are presented in Sect. 10.5. Besides technical proofs, the Appendix recalls, for the readers' convenience, the required notion of Lévy subordinators.

10.2 A General CIID-Framework

We consider a vector of default times $(\tau_1, \ldots, \tau_d) \in [0, \infty)^d$, defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The first aim of this article is to present a generic representation that contains all aforementioned models and, in fact, all possible models relying on the assumption of CIID default times. Assume that $(\tau_1, \ldots, \tau_d)$ is constructed on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ by the following generic two-step method.

Definition 10.1 (The canonical CIID-frailty model).

1. Let $\{F_t\}_{t \geq 0}$ be a non-decreasing, right-continuous stochastic process with left limits, such that $F_0 = 0$ and $\lim_{t \to \infty} F_t = 1$ hold almost surely. For fixed $\omega \in \Omega$, we consider $t \mapsto F_t(\omega)$ as the path of a distribution function of some random variable on $(0, \infty)$.
2. Conditioned on $\{F_t\}_{t \geq 0}$, let $(\tau_1, \ldots, \tau_d)$ be iid with distribution function $t \mapsto F_t$.

A canonical construction of such a model on $(\Omega, \mathscr{F}, \mathbb{P})$ is given by

$$\tau_k := \inf\{t \geq 0 : U_k \leq F_t\}, \quad k = 1, \ldots, d, \tag{10.1}$$

where $U_1, \ldots, U_d$ are iid with $U_k \sim \text{Uni}[0,1]$ and $\{F_t\}_{t\geq 0}$ is independent of the vector $(U_1, \ldots, U_d)$. Such a multivariate default model is called CIID-model in the following; CIID being the acronym of *conditionally independent and identically distributed*. On the one hand, this CIID-construction is a restrictive assumption. For instance, it implies that the law of the default times is invariant under permutations of the components of $(\tau_1, \ldots, \tau_d)$. In particular, each τ_k is distributed according to the distribution function $p(t) := \mathbb{E}[F_t]$, $t \geq 0$. Furthermore, it implicitly inherits a *large homogeneous portfolio assumption*, since the construction above is independent of the dimension d in the sense that one can consider [as an immediate extension of (10.1)] an infinite sequence $\{\tau_k\}_{k\in\mathbb{N}}$ of default times. On the other hand, a seminal theorem of De Finetti, see [19], guarantees that *all* infinite exchangeable sequences of random variables can be constructed as above. This implies that the approach is more general than it might have appeared at first. It is shown below how several popular models are embedded into the general CIID-framework of Definition 10.1 by identifying the respective specification of $\{F_t\}_{t\geq 0}$.

10.2.1 The Portfolio Loss Distribution

The key advantage of CIID-models is that the distribution of the portfolio-loss process $L_t := \big(\mathbb{1}_{\{\tau_1\leq t\}} + \ldots + \mathbb{1}_{\{\tau_d\leq t\}}\big)/d$, $t \geq 0$, is available. More precisely:

$$\mathbb{P}\Big(L_t = \frac{k}{d}\Big) = \binom{d}{k}\mathbb{E}\big[F_t^k\,(1-F_t)^{d-k}\big], \quad k = 0, 1, \ldots, d.$$

For $d \gg 2$ the complexity of the above expectation value as well as the size of the binomial coefficient prevent this formula from being of practical value. Since CIID-models are typically applied in large dimensions, the numerical difficulties are avoided by working with an infinite portfolio size (letting $d \to \infty$) which allows to approximate $\mathbb{P}(L_t \in dx)$ by $\mathbb{P}(F_t \in dx)$. For instance, it is not difficult to verify the following lemma, a proof of which is provided in the Appendix.

Lemma 10.1 (Approximation of the portfolio loss). *Consider the canonical probability space $(\Omega, \mathscr{F}, \mathbb{P})$ of a CIID-model as above. Then*

$$\mathbb{P}\Big(\lim_{d\to\infty}\sup_{t\geq 0}\big|F_t - L_t\big| = 0\Big) = 1.$$

Alternatively, for each $T > 0$ it holds true that

$$\{L_t\}_{t\in[0,T]} \to \{F_t\}_{t\in[0,T]}, \quad d \to \infty,$$

in the space $L^2(\Omega \times [0,T])$ of square-integrable stochastic processes on $[0,T]$.

As an application, the above result is used to justify approximations (for sufficiently large d) such as

$$\mathbb{E}\big[f(L_t)\big] = \int_{[0,1]} f(x)\,\mathbb{P}(L_t \in dx) \approx \int_{[0,1]} f(x)\,\mathbb{P}(F_t \in dx).$$

Approximation results such as Lemma 10.1 are called *large homogeneous portfolio approximation*. In our framework, it is possible to obtain this result as an application of the Theorem of Glivenko–Cantelli. Unlike most of the aforementioned references, we do not have to fix a certain time $t > 0$. This is due to the new formulation as a frailty model, which reveals the underlying structure of (time consistent) CIID-models.

10.2.2 Properties of CIID-Models

CIID-models are appreciated for their mathematical viability. However, for the selection of an appropriate model it is crucial to understand the different dependence structures that are implied by the possible specifications. In an axiomatic way, a list of properties of the resulting vector of default times is specified below.

(Sep): The *separation* of dependence structure from marginals is extremely convenient for practical applications (e.g., the calibration or estimation of the model in two steps) and is also required for the derivation of the model's implied copula. Given the term structure of default probabilities, i.e. $t \mapsto p(t)$, as model input, the separation condition (Sep) is valid if the parametric model for $\{F_t\}_{t\geq 0}$ is specified in such a way that $\mathbb{E}[F_t] = p(t)$ for all $t > 0$. This means that the parameters of $\{F_t\}_{t\geq 0}$ only affect the dependence structure but not the marginal default probabilities.

(Cop): The joint distribution function of $(\tau_1, \ldots, \tau_d)$ in a CIID-model is a priori implicit. More clearly, it is given by

$$\begin{aligned}\mathbb{P}(\tau_1 \leq t_1, \ldots, \tau_d \leq t_d) &= \mathbb{E}\big[\mathbb{E}\big[\prod_{k=1}^{d} \mathbb{1}_{\{\tau_k \leq t_k\}} \big| \{F_t\}_{t\geq 0}\big]\big] = \mathbb{E}\big[\prod_{k=1}^{d} \mathbb{E}\big[\mathbb{1}_{\{\tau_k \leq t_k\}} \big| \{F_t\}_{t\geq 0}\big]\big] \\ &= \mathbb{E}\big[F_{t_1} \cdots F_{t_d}\big], \quad t_1, \ldots, t_d \geq 0.\end{aligned}$$

In some specifications, the latter expectation value can be computed explicitly. In such a case, one can conveniently rely on known statistical properties of the model to judge on its realism. Some distributions even allow for an intuitive economic interpretation. If, in addition, the model satisfies the separation property (Sep), then the marginal distributions of the default times are given a priori. In this case, the dependence structure can be studied from the *implied copula* or, if more convenient, from the *implied survival copula*. If the prespecified term structure of default

probabilities $t \mapsto p(t)$ is continuous, then the implied copula C and the survival copula $\hat{C}$ of the default times are given by

$$C(u_1, \ldots, u_d) = \mathbb{E}\big[F_{p^{-1}(u_1)} \cdots F_{p^{-1}(u_d)}\big], \tag{10.2}$$

$$\hat{C}(u_1, \ldots, u_d) = \mathbb{E}\big[(1 - F_{p^{-1}(1-u_1)}) \cdots (1 - F_{p^{-1}(1-u_d)})\big], \tag{10.3}$$

where $p^{-1}(\cdot)$ denotes the generalized inverse of $p(\cdot)$ and $u_1, \ldots, u_d \in [0, 1]$.

(Exc): Time series of realized corporate defaults or insurance claims often exhibit points in time with accumulations of defaults. This property is termed *excess clustering*. It might even be reasonable to support *multiple defaults* at the same time. In the general CIID-framework, this corresponds to possible jumps in the paths of $\{F_t\}_{t\geq 0}$. In terms of multivariate distribution functions, this corresponds to a singular component of the implied copula of $(\tau_1, \ldots, \tau_d)$.

(Fs): The *qualitative structure* of the underlying *frailty distribution* $\{F_t\}_{t\geq 0}$ is important to understand the dynamics of the model. Three cases are distinguished:

(Fs$_\ominus$): The source of frailty is static, i.e. for each $t > 0$, F_t is measurable with respect to the σ-algebra $\bigcap_{u>0} \sigma(F_s : 0 \leq s \leq u)$. This situation is typical for models that define $\{F_t\}_{t\geq 0}$ as a member of a parametric family of distribution functions with randomly drawn parameter. In most cases, $\{F_t\}_{t\geq 0}$ is monotonically affected by this parameter. This prevents the model from supporting changing market conditions, since the market frailty process $\{F_t\}_{t\geq 0}$ cannot change randomly.

(Fs$_\odot$): The source of frailty is dynamic, but the innovations of the process $\{F_t\}_{t\geq 0}$ are driven by a time-homogeneous stochastic process. Interpreted from an economic perspective, this implies that the market uncertainty is affected by random changes, but these changes occur in a time-homogeneous pattern.

(Fs$_\oplus$): The source of frailty is dynamic and the innovations of the process $\{F_t\}_{t\geq 0}$ are driven by a time-inhomogeneous stochastic process. From an economic point of view, this allows for realizations with randomly varying default environments. In particular, a typical realization of $\{F_t\}_{t\geq 0}$ inherits time periods with different local default probabilities and dependence structures.

(Tdc): A measure of dependence for the likelihood of joint early defaults is *tail dependence*. In the context of default risk, a positive lower tail dependence coefficient of the default times corresponds to a positive limit (as time goes to zero) of pairwise default correlations, see [54, Chap. 10]. Hence, this property is of specific interest for models with small default probabilities or small time horizon. Additionally, empirical studies suggest that models supporting positive lower tail dependence of the default times are more successful in explaining CDO quotes. In mathematical terms, the lower tail dependence coefficient λ_l of a pair (τ_i, τ_j) of default times in a CIID-setup is given by[7]

[7]Thus, this important measure of dependence is related to the specification of F_t in a rather simple way. The derivation of (10.4) is straightforward and therefore omitted.

$$\lambda_l := \lim_{t \downarrow 0} \mathbb{P}(\tau_i \leq t \mid \tau_j \leq t) = \lim_{t \downarrow 0} \frac{\mathbb{E}\big[F_t^2\big]}{\mathbb{E}[F_t]}. \tag{10.4}$$

(Den): When implementing the model, it is convenient if the distribution of F_t is tractable for all $t > 0$. Most convenient is the case when the *density* of F_t is available. Some specifications allow for a closed-form expression without special functions. In some models, the density is available through Laplace-inversion techniques.

It is shown below how various popular models are comprised in the CIID-framework. These models are discussed with regard to the aforementioned properties.

10.2.3 The Gaussian Copula Model and Extensions

Li and Vasicek [38, 57] generalize the univariate structural default model of [50] to d identical firms. Dependence is introduced through a single-factor structure: the idiosyncratic factors $\epsilon_1, \ldots, \epsilon_d$ and the market factor M are iid standard normally distributed random variables. Given the prespecified term structure of default probabilities $t \mapsto p(t)$ as model input, the default time of firm k is defined as

$$\begin{aligned} \tau_k &:= \inf\big\{t \geq 0 : \sqrt{\rho}\, M + \sqrt{1-\rho}\, \epsilon_k \leq \Phi^{-1}\big(p(t)\big)\big\} \\ &= \inf\Big\{t \geq 0 : U_k \leq \Phi\Big(\frac{\Phi^{-1}\big(p(t)\big) - \sqrt{\rho}\, M}{\sqrt{1-\rho}}\Big)\Big\}, \quad k = 1, \ldots, d, \end{aligned} \tag{10.5}$$

where $\rho \in (0,1)$ adjusts the dependence, Φ denotes the distribution function of the standard normal law, and $U_1, \ldots, U_d$ are iid and obtained by $U_k := \Phi(\epsilon_k) \sim \text{Uni}[0,1]$. By construction, the vector $(\tau_1, \ldots, \tau_d)$ has a Gaussian copula with identical pairwise correlation ρ as dependence structure and marginal distributions $\mathbb{P}(\tau_k \leq t) = p(t)$. Reformulating the model in the general CIID-setting, one obtains $F_t := \Phi\big(\big(\Phi^{-1}\big(p(t)\big) - \sqrt{\rho}\, M\big)/\sqrt{1-\rho}\big)$, for $t \geq 0$. Summing up, the distribution of L_t can be approximated via the standard normal distribution of M. Generalizing this approach to distributions other than the normal, [32] propose to replace it by the (heavier tailed) Student t-distribution. In a similar spirit, [4] consider a Lévy process $X = \{X_t\}_{t \in [0,1]}$, satisfying $\mathbb{E}[X_1] = 0$ and $\text{Var}[X_1] = 1$. Letting $X^{(0)}, \ldots, X^{(d)}$ be $d+1$ independent copies of X and $\rho \in (0,1)$, they define the individual factors $\epsilon_k := X^{(k)}_{1-\rho}$, $k = 1, \ldots, d$, and the market factor $M := X^{(0)}_{\rho}$. Then, construction (10.5) is replaced by

$$\begin{aligned} \tau_k &:= \inf\big\{t \geq 0 : M + \epsilon_k \leq H^{-1}_{[1]}\big(p(t)\big)\big\} \\ &= \inf\big\{t \geq 0 : U_k \leq H_{[1-\rho]}\big(H^{-1}_{[1]}\big(p(t)\big) - M\big)\big\}, \quad k = 1, \ldots, d, \end{aligned}$$

where $H_{[t]}$ denotes the distribution function of X_t and $U_1, \ldots, U_d$ are iid and obtained by $U_k := H_{[1-\rho]}(\epsilon_k) \sim \text{Uni}[0,1]$. This obviously corresponds to the choice $F_t := H_{[1-\rho]}\big(H_{[1]}^{-1}\big(p(t)\big) - M\big)$, for $t \geq 0$, in the general CIID-setup. When X is a Brownian motion, this approach is equivalent to (10.5). Considering other specifications, [51] uses a Variance-Gamma process, [29, 35] a NIG process, and [8] the sum of a Brownian motion and a Variance-Gamma process. Again, the distribution of L_t can be approximated via the distribution of the single random variable M, which is easy to handle.

Properties of the Model

(Sep): In this specification, one can take $t \mapsto p(t)$ as model input for the univariate marginal laws and obtains $\mathbb{E}[F_t] = p(t)$, $t \geq 0$.

(Cop): The copula behind the model is identified in the Gaussian specification and for Student t-factors as the respective distribution's copula. For the general Lévy framework, the implicitly defined copula in (10.2) and (10.3) is not well studied.

(Exc): In all specifications with continuous $t \mapsto H_{[\rho]}(t)$, the resulting copula (explicitly or implicitly given) does not have a singular component, so multiple defaults at the same time are impossible. Equivalently, $t \mapsto F_t$ is almost surely continuous (for continuous $t \mapsto p(t)$).

($\text{Fs}_\ominus$): The randomness in this class of models is induced by the randomness of the parameter of an otherwise deterministic distribution function. This makes these models static and also difficult to interpret.

(Tdc): One major disadvantage of a Gaussian dependence structure is zero tail dependence, see [49, p. 211]. This means that joint early defaults are very unlikely. For other model specifications with a heavier tailed common factor, positive tail dependence is possible. For instance, it is shown in [4] that for the Lévy construction, the lower tail dependence coefficient of any pair of default times is

$$\lambda_l = \lim_{x \to -\infty} \int_{\mathbb{R}} \frac{H_{[1-\rho]}(x-y)^2}{H_{[1]}(x)} dH_{[\rho]}(y).$$

(Den): The density of F_t, $t > 0$, is known for various choices of the common factor, making this class of models quite viable.

When calibrating the Gaussian model to the CDO market it is often the case that for matching spreads of senior tranches, extremely high correlation parameters are required. When distributions with heavier tails are used, e.g. the NIG model, the model seems to be better suited for a calibration to the CDO market.

10.2.4 A Model Based on Mixtures of Exponential Distributions

Marshall–Olkin [48] show that the dependence structure behind iid exponential random variables with randomly drawn parameter is Archimedean. Denoting the Laplace transform of the positive random variable M by $\varphi(x) := \mathbb{E}[\exp(-x\,M)]$, $x \geq 0$, it follows that $\mathbb{P}\big(\epsilon_1/M > t_1, \ldots, \epsilon_d/M > t_d\big) = C_\varphi\big(\varphi(t_1), \ldots, \varphi(t_d)\big)$, $t_1, \ldots, t_d \geq 0$, where $\epsilon_1, \ldots, \epsilon_d$ are iid unit exponentially distributed and independent of M. The function $C_\varphi(u_1, \ldots, u_d) := \varphi\big(\varphi^{-1}(u_1) + \ldots + \varphi^{-1}(u_d)\big)$ is called *Archimedean copula* with *generator* φ. Transforming the components to standard uniform marginals, it follows that

$$(V_1, \ldots, V_d) := \Big(\varphi\Big(\frac{\epsilon_1}{M}\Big), \ldots, \varphi\Big(\frac{\epsilon_d}{M}\Big)\Big) \sim C_\varphi.$$

Schönbucher [53] uses this probabilistic model for portfolio credit risk and derives a large homogeneous portfolio approximation. Formulated as a frailty model and using the notations from above, given the prespecified term structure of default probabilities $t \mapsto p(t)$, one defines[8]

$$\begin{aligned}\tau_k &:= \inf\big\{t \geq 0 : 1 - p(t) \leq V_k\big\},\\ &= \inf\big\{t \geq 0 : U_k \leq 1 - \exp\big(-M\varphi^{-1}(1 - p(t))\big)\big\}, \quad k = 1, \ldots, d,\end{aligned}$$

where $U_1, \ldots, U_d$ are iid obtained by $U_k := 1 - \exp(-\epsilon_k) \sim \mathrm{Uni}[0, 1]$. Translating this construction into the present CIID-setup yields $F_t := 1 - \exp\big(-M\,\varphi^{-1}\big(1 - p(t)\big)\big)$, for $t \geq 0$. Summarizing, this implies that the distribution of L_t can be approximated using the distribution of M and default times defined in this way have an Archimedean survival copula C_φ.[9]

[8] The first line indicates the idea of [53]: starting from the canonical construction of a default time with distribution function $t \mapsto p(t)$, see [54, p. 122], dependent trigger variables $(V_1, \ldots, V_d)$ are used as the source of dependence.

[9] If one wishes to define default times in such a way that they have C_φ as copula instead of survival copula, one must use $F_t := \exp\big(-M\,\varphi^{-1}\big(p(t)\big)\big)$, $t \geq 0$. This can be deduced by replacing $(V_1, \ldots, V_d)$ in the above derivation by $(1 - V_1, \ldots, 1 - V_d)$. This alternative ansatz can be used to switch tail dependencies: the lower tail dependence of C_φ equals the upper tail dependence of its survival copula, and vice versa. Since lower tail dependence between default times is desirable, one should use the latter approach when C_φ exhibits lower tail dependence and the first approach when C_φ exhibits upper tail dependence.

Properties of the Model

(Sep): The separation property holds, i.e. the term structure of default probabilities $t \mapsto p(t)$ can be prespecified and $\mathbb{E}[F_t] = 1 - \varphi\big(\varphi^{-1}(1 - p(t))\big) = p(t)$, for $t \geq 0$.

(Cop): By construction, the copula behind the model is of Archimedean kind. Such copulas, being parameterized by a function φ, are quite flexible. On the other side, they do not provide a firm economic interpretation.

(Exc): Multiple defaults at the same time are not possible, since Archimedean copulas do not assign positive mass to the diagonal of the unit cube. Stated differently, the process $\{F_t\}_{t\geq 0}$ is almost surely continuous (for continuous $t \mapsto p(t)$).

(Fs$_\ominus$): As outlined above, the model is based on an exponential distribution with randomly chosen parameter. Hence, the model is static and difficult to interpret from an economic perspective.

(Tdc): The upper and lower tail dependence parameters of the Archimedean copula C_φ are given by

$$\lambda_u = \begin{Bmatrix} 0, & \varphi'(0) < \infty \\ 2 - 2 \lim\limits_{t \searrow 0} \frac{\varphi'(2t)}{\varphi'(t)}, & \text{else} \end{Bmatrix}, \quad \lambda_l = 2 \lim_{t \nearrow \infty} \frac{\varphi'(2t)}{\varphi'(t)},$$

see [34, p. 103ff]. For several classes, these parameters are positive and might even be unequal.

(Den): The density of F_t, $t > 0$, is known for various choices of the common factor, rendering this class of models quite viable. A list of popular generator functions φ and their associated random variables M is provided in, e.g., [16].

10.2.5 An Intensity-Based Approach

On a univariate level, intensity-based models are introduced and developed further in, e.g., [21, 33, 37, 42]. On a multivariate level, the so-called *doubly-stochastic approaches* and extensions thereof are considered in, e.g., [18, 23, 24, 58]. A single-factor specification, which fits into the setup of general CIID-models, can be constructed as a special case of the model in [23]. On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, consider a positive stochastic process $\{\lambda_t\}_{t\geq 0}$, which is $\mathbb{P}$-almost surely integrable on $[0, t]$ for all $t > 0$ and satisfies $\int_{(0,\infty)} \lambda_s \, ds = \infty$. Independently of this *market intensity* process, let $\epsilon_1, \ldots, \epsilon_d$ be iid unit exponential random variables. The vector $(\tau_1, \ldots, \tau_d)$ of default times is defined by setting

$$\tau_k := \inf\big\{t > 0 : M_t \geq \epsilon_k\big\}, \quad M_t := \int_0^t \lambda_s \, ds, \quad k = 1, \ldots, d.$$

Translated into the setup of CIID-models, this is equivalent to modeling $\{F_t\}_{t\geq 0}$ as $F_t := 1 - \exp(-M_t)$, for $t \geq 0$. A prominent choice for $\{\lambda_t\}_{t\geq 0}$ is a *basic affine process*. This means that $\{\lambda_t\}_{t\geq 0}$ has parameters $(\kappa, \theta, \sigma, \mu, l)$ and is defined as the (unique) solution of the stochastic differential equation (SDE)

$$d\lambda_t = \kappa\,(\theta - \lambda_t)\,dt + \sigma\,\sqrt{\lambda_t}\,dB_t + dZ_t, \quad \lambda_0 > 0, \tag{10.6}$$

where $\{B_t\}_{t\geq 0}$ is a standard Brownian motion and $\{Z_t\}_{t\geq 0}$ is an independent compound Poisson process with intensity l and exponential jump sizes with mean $1/\mu$. Besides the immediate interpretation of the SDE for λ, one important advantage of using basic affine processes is that the Laplace transform of M_t is available from general results in [22]. More clearly, it is known that

$$\mathbb{E}\big[e^{-x\,M_t}\big] = e^{\alpha(x,t)+\beta(x,t)\,\lambda_0}, \quad x \geq 0, \tag{10.7}$$

where the functions α and β are given by

$$\beta(x,t) = \frac{1 - e^{b(x)\,t}}{c(x) + d(x)\,e^{b(x)\,t}}, \tag{10.8}$$

$$\begin{aligned}\alpha(x,t) = {} & 2\kappa\theta\Big(-\sigma^{-2}\log\Big(\frac{c(x)+d(x)e^{b(x)t}}{c(x)+d(x)}\Big) + \frac{t}{2c(x)}\Big) \\ & + l\Big(\frac{2\mu\log\big(1 - e^{b(x)t} - \mu(c(x)+d(x)e^{b(x)t})\big)}{2\mu\kappa + 2x - \sigma^2\mu^2}\Big) \\ & + l\Big(\frac{-t}{1-\mu c(x)} - \frac{2\mu\log\big(-\mu(c(x)+d(x))\big)}{2\mu\kappa + 2x - \sigma^2\mu^2}\Big),\end{aligned} \tag{10.9}$$

with $b(x)$, $c(x)$, and $d(x)$ defined by

$$b(x) = -\sqrt{\kappa^2 + 2\sigma^2 x}, \quad c(x) = \frac{\kappa + \sqrt{\kappa^2 + 2\sigma^2 x}}{-2x}, \quad d(x) = \frac{-\kappa + \sqrt{\kappa^2 + 2\sigma^2 x}}{-2x}.$$

This allows to compute $p(t) = \mathbb{E}[F_t]$ in closed form, as $p(t) = 1 - \mathbb{E}\big[\exp(-M_t)\big] = 1 - \exp\big(\alpha(1,t) + \beta(1,t)\,\lambda_0\big)$.

Properties of the Model

(Sep): All parameters $(\kappa, \theta, \sigma, \mu, l)$ of a basic affine process enter the formulas for the marginal and the joint default probabilities. There is no parameter that solely affects the dependence structure; a separation is not possible. This complicates the

calibration of the model and the interpretation of the parameters. However, the five parameters $(\kappa, \theta, \sigma, \mu, l)$ provide a good fit of the function $t \mapsto p(t) = \mathbb{E}[F_t]$ to market quotes of single-name CDS or portfolio CDS.

(Cop): The copula behind the model is not well studied. This makes it difficult to study the underlying dependence structure of default times.

(Exc): [18] find evidence that intensity-based approaches fail to explain excess clustering as observed in the markets, e.g. during the recent credit crisis. This is due to the fact that the integrated intensity process is continuous, and, hence, the random distribution function $\{F_t\}_{t\geq 0}$ is continuous, too. In Sect. 10.4 we propose an extension of this intensity-based approach to incorporate excess clustering. This is achieved by incorporating jumps into $\{F_t\}_{t\geq 0}$.

($\text{Fs}_{\oplus}$): The intensity process $\{\lambda_t\}_{t\geq 0}$ is interpreted as an instantaneous default rate, making the model quite intuitive. The larger the intensity λ_t, the larger is the default probability over $[t, t + dt]$. Consequently, a typical realization of $\{F_t\}_{t\geq 0}$ inherits time periods with different local default probabilities, resulting from periods with high or low λ_t.

(Tdc): For a specification of the model using a basic affine process $\{\lambda_t\}_{t\geq 0}$, the resulting bivariate lower tail dependence coefficient (10.4) is zero. The required computation is very tedious and postponed to the Appendix.

(Den): For the approximation $\mathbb{P}(L_t \in dx) \approx \mathbb{P}(F_t \in dx)$ the density of M_t is required, which in the basic affine case can be obtained from the known Laplace transform via Laplace inversion. Even though this makes the implementation of the model more involved, it is still more efficient compared to a Monte-Carlo simulation. References for Laplace inversion algorithms, based on different theoretical inversion formulas, include [1–3, 56] and especially for the integrated CIR process [36]. We performed several numerical experiments and identified Talbot-type algorithms to be best suited for the present problem.

10.2.6 A Model Based on Lévy Subordinators

The model of [44] is related to the intensity-based approach, the key difference being that the (continuous) integrated intensity process is replaced by a (discontinuous) jump process. To set up the model, let $\epsilon_1, \ldots, \epsilon_d$ be iid random variables with unit exponential distribution and let $\Lambda = \{\Lambda_t\}_{t\geq 0}$ be an independent (killed) Lévy subordinator with Laplace exponent Ψ. We further assume Λ to be nondegenerate, i.e. $0 \neq \Lambda$. A brief introduction to Lévy subordinators is provided in the Appendix. Given the prespecified continuous and strictly increasing term structure of default probabilities $t \mapsto p(t)$, one defines the cumulative hazard function $t \mapsto h(t) := -\log\big(1 - p(t)\big)$ and sets the market frailty process as $M_t := \Lambda_{h(t)/\Psi(1)}$, $t \geq 0$. The default times are defined as

$$\tau_k := \inf\big\{t \geq 0 : M_t \geq \epsilon_k\big\} = \inf\big\{t \geq 0 : U_k \leq 1 - e^{-M_t}\big\}, \quad k = 1, \ldots, d,$$

where $U_1, \ldots, U_d$ are iid obtained by $U_k := 1-\exp(-\epsilon_k) \sim \text{Uni}[0,1]$. Translating it into the framework of a general CIID-model, this approach corresponds to defining $F_t := 1 - \exp(-M_t)$, where $M_t := \Lambda_{h(t)/\Psi(1)}$ for $t \geq 0$.

Properties of the Model

(Sep): A separation of marginals from dependence structure is valid: independently of the choice of subordinator Λ and with prespecified term structure $t \mapsto p(t)$, it holds that $\mathbb{E}[F_t] = 1 - \exp\big(-h(t)\big) = p(t)$, $t \geq 0$.

(Cop): The survival copula of the default times is known to be of Marshall–Olkin kind, see [43]. The Marshall–Olkin distribution has already been proposed for credit-risk modeling by [28, 40], since it provides an intuitive interpretation as an exogenous shock model.

(Exc): Multiple defaults at the same time are possible, since the subordinator Λ can jump across more than one trigger variable at a time. Hence, the survival copula behind the multivariate default model (which is an exchangeable Marshall–Olkin copula) has a singular component on the diagonal and the model supports joint defaults. This property distinguishes the model from all aforementioned model specifications.

($\text{Fs}_{\odot}$): Since the common factor is a stochastic process instead of a single random variable, one obtains a dynamic structure of $\{F_t\}_{t\geq 0}$. Unfortunately, the dependence structure behind the default times exhibits the so-called *multivariate lack-of memory property*, see, e.g., [27, 47]. Heuristically, the Lévy properties of Λ, corresponding to the lack-of memory property of the Marshall–Olkin distribution, force jumps of the market frailty to occur in time-homogeneous pattern.

(Tdc): The lower tail dependence coefficient of any pair of default times equals $\lambda_l = 2 - \Psi(2)/\Psi(1)$, see [44], which is always positive unless $\Lambda_t = t$, $t \geq 0$.

(Den): Several classes of Lévy subordinators have known densities. Examples include the Inverse Gaussian and the Gamma subordinator. Other examples have semi-explicit densities, e.g., the stable subordinator and several compound Poisson subordinators.[10] Efficiently obtaining the density via Laplace-inversion is treated in [10].

In the following sections, two extensions are presented. These aim at combining the desirable properties of the aforementioned models while preserving their viability. The first generalization combines the models of [44, 53]. The implied dependence structure is of Archimax kind. The second generalization combines the intensity-based approach with the model of [44]. The result is the so-called *triply-stochastic model* which supports default clustering.

[10] This construction contains the simple Marshall–Olkin model with one armageddon shock of [13] as a special case. Rewriting this example in this framework, the subordinator Λ must be linearly increasing until a single jump to infinity simultaneously destroys all components.

10.3 A Model Based on Scale Mixtures of Marshall–Olkin Copulas

It is possible to unify the approaches of [44, 53]. In the language of copula theory, this corresponds to combining Archimedean with Marshall–Olkin copulas, which represents a family of copulas termed *scale mixtures of Marshall–Olkin copulas*, see [39]. It constitutes a proper subclass of the so-called Archimax copulas, as introduced for the bivariate case in [14]. Let $\bar{M} > 0$ be a random variable with Laplace transform $\varphi(x) := \mathbb{E}[\exp(-x\bar{M})]$, $x \geq 0$, and let $\Lambda \neq 0$ be an independent (killed) Lévy subordinator with Laplace exponent Ψ. Independently of $(\bar{M}, \Lambda)$, let $\epsilon_1, \ldots, \epsilon_d$ be iid unit exponentially distributed. Given the continuous and strictly increasing prespecified term structure of default probabilities $t \mapsto p(t)$, one defines the market frailty process $M_t := \Lambda_{\bar{M}\varphi^{-1}(1-p(t))/\Psi(1)}$. The default times are defined by

$$\tau_k := \inf\{t \geq 0 : M_t \geq \epsilon_k\} = \inf\{t \geq 0 : U_k \leq 1 - e^{-M_t}\}, \quad k = 1, \ldots, d,$$

where $U_1, \ldots, U_d$ are iid obtained by $U_k := 1 - \exp(-\epsilon_k) \sim \text{Uni}[0, 1]$. In particular, choosing the calendar time $\Lambda_t = t$, $t \geq 0$, the model is equivalent to that of [53]. Similarly, the choice $\bar{M} \equiv 1$ implies the model of [44]. Rewriting the model in the general CIID-setup, this means that

$$F_t := 1 - e^{-M_t}, \quad M_t := \Lambda_{\bar{M}\varphi^{-1}(1-p(t))/\Psi(1)}, \quad t \geq 0.$$

Suitable choices of $(\bar{M}, \Lambda)$ render the distribution of F_t tractable enough to be useful for efficient pricing. It is shown below that this class admits several desirable properties for the modeling of joint defaults. More precisely, it contains the full flexibility of the Archimedean class, inherits the singular component of the Marshall–Olkin class, combines the positive dependence coefficients of both classes of copulas, and improves the dynamic aspects of the original Lévy model. An important property of this model is that the resulting dependence structure can be identified. The specific form of the survival copula is provided in Theorem 10.1 below, a proof is given in the Appendix.

Theorem 10.1 (The survival copula of the vector of default times). *The survival copula of the vector* $(\tau_1, \ldots, \tau_d)$ *is*

$$\hat{C}(u_1, \ldots, u_d) = \varphi\Big(\frac{1}{\Psi(1)} \sum_{i=1}^{d} \varphi^{-1}(u_{(i)}) \big(\Psi(i) - \Psi(i-1)\big)\Big), \tag{10.10}$$

where $u_{(1)} \leq \ldots \leq u_{(d)}$ *denotes the ordered list of* $u_1, \ldots, u_d \in [0, 1]$.

Properties of the Model

(Sep): A separation of marginals from dependence structure is model inherent, since

$$\begin{aligned}\mathbb{E}[F_t] &= 1 - \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(-\Lambda_{\bar{M}\,\varphi^{-1}\left(1-p(t)\right)/\Psi(1)}\Big)\Big|\bar{M}\Big]\Big] \\ &= 1 - \mathbb{E}\Big[\exp\Big(-\bar{M}\,\varphi^{-1}\big(1-p(t)\big)\Big)\Big] = p(t), \quad t \geq 0.\end{aligned}$$

(Cop): The default times have a scale mixture of Marshall–Olkin copulas as survival copula. Statistical properties of this class are investigated, e.g., in [9, 39]. The specific form is computed in Theorem 10.1.

(Exc): Multiple defaults are possible, since $\{F_t\}_{t\geq 0}$ might have jumps. In the language of copula theory, the survival copula (10.10) has a singular component on the diagonal. More precisely, the events $\{\tau_1 = \ldots = \tau_k\}$, $k = 2, \ldots, d$, are independent of the realization of $\bar{M}$. Hence, their probabilities solely depend on the Lévy subordinator and can be extracted from a computation in [45]:

$$\mathbb{P}(\tau_1 = \ldots = \tau_k) = \frac{\sum_{i=0}^{k}\binom{k}{i}(-1)^{i+1}\Psi(i)}{\Psi(k)}, \quad k = 2, \ldots, d.$$

($\mathrm{Fs}_{\odot}$): The process $\{F_t\}_{t\geq 0}$ is driven by a mixture of Lévy subordinators. Since $\bar{M}$ is independent of the time t, no time-inhomogeneity is introduced to $\{F_t\}_{t\geq 0}$.

(Tdc): It is not difficult to compute the tail dependence of a pair (τ_i, τ_j) of default times. Using L'Hospital's rule, it is given by[11] (whenever this limit exists)

$$\lambda_l = \lim_{t\downarrow 0}\frac{\mathbb{E}\big[F_t^2\big]}{\mathbb{E}[F_t]} = 2 - \frac{\Psi(2)}{\Psi(1)}\lim_{t\downarrow 0}\frac{\varphi'\big(t\,\Psi(2)/\Psi(1)\big)}{\varphi'(t)}.$$

(Den): Assume that Λ_t admits a density $f_t^{(\Lambda)}$ for $t > 0$. It follows from Fubini's Theorem that the random variable $\Lambda_{\bar{M}\,t}$, $t > 0$, has the density f_t, given by

$$f_t(x) = \mathbb{E}\Big[f_{\bar{M}\,t}^{(\Lambda)}(x)\Big] = \int_0^\infty f_{y\,t}^{(\Lambda)}(x)\,\mathbb{P}(\bar{M} \in dy), \quad x > 0.$$

The latter integral can efficiently be computed when $\bar{M}$ admits a density.

We close this section by giving two specifications of $(\bar{M}, \Lambda)$, which imply viable formulas for all required quantities.

---

[11] This result can be validated for the subclasses of Archimedean and Marshall–Olkin survival copulas: the case $\bar{M} \equiv 1$ gives $\lambda_l = 2 - \Psi(2)/\Psi(1)$, which agrees with the result obtained in the model of [44]. Similarly, the case $\Lambda_t = t$, $t \geq 0$, leads to $\lambda_l = 2 - 2\lim_{t\downarrow 0}\varphi'(2\,t)/\varphi'(t)$, which agrees with the tail dependence parameter in the model of [53]. Higher-dimensional dependence measures can be retrieved from results in [39].

Example 10.1 (An Archimedean model with Armageddon-scenario). A new parametric family is obtained when an Archimedean model is combined with a Lévy subordinator that increases linearly with drift $\alpha \in [0, 1)$ and might jump to infinity, i.e. its Lévy measure is determined by $\nu(\{\infty\}) = (1-\alpha)$, $\nu((0,\infty)) = 0$. Put differently,

$$\Lambda_t := \alpha\, t + \infty \cdot \mathbb{1}_{\{t>E\}}, \quad t \geq 0,$$

where E is an exponential random variable with mean $1/(1-\alpha)$. Interpreted from an economic point of view, this corresponds to an Archimedean-type dependence structure combined with the positive probability of an Armageddon-scenario killing all remaining components. The resulting survival copula of default times interpolates between the co-monotonicity copula and the chosen Archimedean copula and is given by

$$\hat{C}(u_1,\ldots,u_d) = \varphi\Big((1-\alpha)\,\varphi^{-1}(u_{(1)}) + \alpha \sum_{i=1}^{d} \varphi^{-1}(u_{(i)})\Big), \quad u_1,\ldots,u_d \in [0,1].$$

The required distributions of $\Lambda_{\bar{M}\,t}$, $t > 0$, are found to be:

$$\begin{aligned}
\mathbb{P}(\Lambda_{\bar{M}\,t} = \infty) &= \mathbb{P}(\bar{M}\,t > E) = 1 - \varphi\big(t\,(1-\alpha)\big),\\
\mathbb{P}(\Lambda_{\bar{M}\,t} \leq x) &= \mathbb{E}\Big[e^{-(1-\alpha)\,\bar{M}\,t}\,\mathbb{1}_{\{\bar{M}\leq \frac{x}{\alpha\,t}\}}\Big], \quad x \in [0,\infty).
\end{aligned}$$

Example 10.2 (Gamma scale mixture of Cuadras–Augé copulas). A model specification with explicit distribution of F_t is obtained as follows: let Λ be a Poisson process with intensity $\beta > 0$ and $\bar{M}$ be a $\Gamma(1, 1/\theta)$-distributed random variable. In the language of Archimedean copulas, this correponds to C_φ being a Clayton-copula. It follows that

$$\begin{aligned}
\mathbb{P}(\Lambda_{\bar{M}\,t} = k) &= \frac{(t\,\beta)^k}{k!}\,\mathbb{E}\Big[\bar{M}^k\,e^{-\beta\,t\,\bar{M}}\Big] = \frac{(t\,\beta)^k}{\Gamma(1/\theta)\,k!}\int_0^\infty y^k\,e^{-\beta\,t\,y}\,y^{\frac{1}{\theta}-1}e^{-y}\,dy\\
&= \frac{(t\,\beta)^k}{\Gamma(1/\theta)\,k!}\Big(\frac{1}{1+\beta\,t}\Big)^{k+\frac{1}{\theta}}\,\Gamma\Big(k+\frac{1}{\theta}\Big), \quad k \in \mathbb{N}_0.
\end{aligned}$$

This choice corresponds to a Gamma scale mixture of exchangeable Cuadras–Augé copulas, see [17] for an introduction to the Cuadras–Augé family. Interpreted differently, it corresponds to a generalization of Clayton copulas. The specific form of the copula is obtained from Theorem 10.1 with $\varphi(t) = (1+t)^{1/\theta}$ and $\Psi(x) = \beta\,\big(1-e^{-x}\big)$.

10.4 A Model Based on CGMY-Type Processes

In this section, we propose a CIID-model which combines the intensity-based approach with the approach of [44]. Recall that one shortfall of the former class is the fact that $\{F_t\}_{t\geq 0}$ is continuous. Consequently, it does not support joint defaults. On the other side, the model of [44], although supporting jumps of $\{F_t\}_{t\geq 0}$, is based on the somewhat unrealistic lack-of-memory properties (inhereted from Λ being a Lévy process). The idea of this generalization is to combine both approaches to create a model that overcomes both shortcomings and produces realistic default pattern over time. Still, it remains tractable enough to allow for efficient pricing routines without Monte-Carlo techniques. When empirical corporate defaults are monitored over time, two stylized facts are observed: (a) There are time periods with few and time periods with many defaults. In between those periods, there is typically a gradual change from one regime to the other. (b) Occasionally, there are times with a sudden peak in the number of corporate defaults. This model is designed to mimic these properties. To formally define the model, consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ supporting the following (independent) objects:

- A basic affine process $\{\lambda_t\}_{t\geq 0}$ as given by the SDE (10.6).
- A Lévy subordinator $0 \neq \Lambda = \{\Lambda_t\}_{t\geq 0}$ with Laplace exponent Ψ.
- A list of iid unit exponential random variables $\epsilon_1, \ldots, \epsilon_d$.

With $M_t := \Lambda_{\int_0^t \lambda_s ds/\Psi(1)}$, $t \geq 0$, the individual default times are defined as

$$\tau_k := \inf\{t \geq 0 : M_t \geq \epsilon_k\}, \quad k = 1, \ldots, d.$$

Due to the definition of τ via three stochastic objects, we term this class of models *triply stochastic*. The according CIID-model stems from

$$F_t := 1 - e^{-M_t}, \quad M_t := \Lambda_{\int_0^t \lambda_s \, ds/\Psi(1)}, \quad t \geq 0.$$

The process $\{M_t\}_{t\geq 0}$ is a time-changed Lévy process in the spirit of [15]. The Lévy subordinator $\{\Lambda_t\}_{t\geq 0}$ incorporates jumps into $\{F_t\}_{t\geq 0}$, which corresponds to positive probabilities of joint defaults and excess clustering. This accounts for the occurrence of peaks in the number of defaults. The intensity process $\{\lambda_t\}_{t\geq 0}$ incorporates time-inhomogeneity: the larger λ_t, the larger the probability of defaults over the next instance of time. The stochastic nature of $\{\lambda_t\}_{t\geq 0}$ overcomes the lack-of-memory property of the model presented in [44]. An important argument for the use of CGMY-type processes is that their Laplace transform is known in closed form. Thus, numerical pricing routines are available using Laplace-inversion techniques, see [1–3, 56], which are much more efficient than Monte-Carlo pricing routines. More clearly, one computes

$$\mathbb{E}[F_t] = 1 - \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(-\Lambda_{\int_0^t \lambda_s\,ds/\Psi(1)}\Big)\Big|\int_0^t \lambda_s\,ds\Big]\Big]$$
$$= 1 - \mathbb{E}\Big[e^{-\int_0^t \lambda_s\,ds}\Big] = 1 - e^{\alpha(1,t)+\beta(1,t)\,\lambda_0}, \quad t \geq 0, \tag{10.11}$$

with functions β and α as given in (10.8) and (10.9).

This implies that the marginal default probabilities are equal to the ones in the intensity-based approach. Stated differently, the Lévy subordinator only affects the dependence structure. In this regard, $\{\Lambda_t\}_{t\geq 0}$ is an additional source of frailty, which accounts for excess clustering.

Properties of the Model

(Sep): The parameters of $\{\lambda_t\}_{t\geq 0}$ enter the formula for $p(t) = \mathbb{E}[F_t]$. However, the parameters of the jump process $\{\Lambda_t\}_{t\geq 0}$ do not affect $p(t)$, see (10.11). Consequently, the parameters of the intensity can be calibrated to correlation-insensitive market quotes in a first step, and the remaining parameters of the Lévy subordinator provide additional freedom to calibrate the dependence structure in a second step. Hence, even though (Sep) is not fully valid, one can still apply a two-step calibration routine.

(Cop): The multivariate distribution of $(\tau_1, \ldots, \tau_d)$ is not well studied. This complicates the investigation of the underlying dependence structure of the default times.

(Exc): Regarding joint default probabilities, the model inherits all desired properties from the approach of [44], since the events $\{\tau_1 = \ldots = \tau_k\}$, for $k = 2, \ldots, d$, are independent of the process $\{\lambda_t\}_{t\geq 0}$. In particular, it holds that

$$\mathbb{P}(\tau_1 = \ldots = \tau_k) = \frac{\sum_{i=0}^{k} \binom{k}{i}(-1)^{i+1}\Psi(i)}{\Psi(k)}, \quad k = 2, \ldots, d.$$

(Fs$_\oplus$): The process $\{F_t\}_{t\geq 0}$ is a transformation of a time-changed Lévy subordinator in the spirit of [15]. Intuitively, the Lévy subordinator $\{\Lambda_t\}_{t\geq 0}$ accounts for jumps of $\{F_t\}_{t\geq 0}$. Since it is affected by a random time-change, these jumps can occur in a time-inhomogeneous pattern: the larger the intensity λ_t, the more likely a jump of F_t is to occur. This property can be observed in Fig. 10.1, where one realization of the model is illustrated.

(Tdc): Starting from (10.4), a lengthy computation (related to the one in the Appendix for the model without subordinator Λ) involving the specific form of the Laplace transform identifies the lower tail-dependence coefficient of any two default times as $2 - \Psi(2)/\Psi(1)$. This result is intuitive: it agrees with the coefficient in the model of [44]. Compared with this model, the deterministic function $t \mapsto h(t)$ is replaced in this framework by a function of the integrated (random) intensity. However, the presented intensity-model alone does not generate tail dependence.

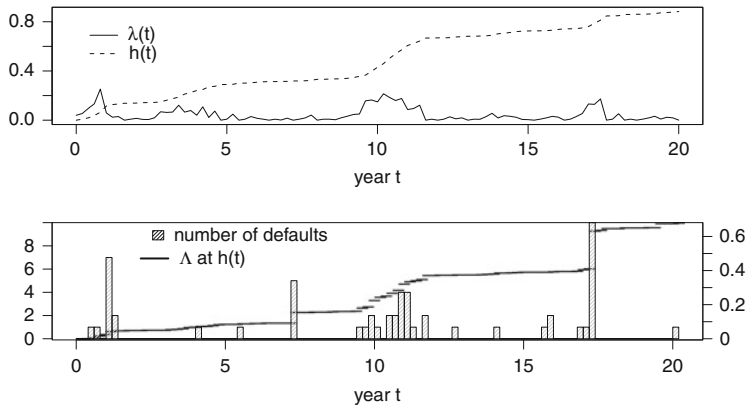

Fig. 10.1 One realization of the CGMY-based model with portfolio size $d = 125$ over 20 years. Specification: $\{\lambda_t\}_{t\geq 0}$ is a Cox–Ingersoll–Ross process (a basic affine process without jumps) with $\lambda_0 = 0.04$ and parameters $(\kappa, \theta, \sigma) = (1, 0.04, 0.25)$. The Lévy subordinator is specified by $\Psi(x) = x^{0.8}$, $x \geq 0$, i.e. it is a 0.8-stable subordinator. The upper plot illustrates the simulated paths of $\{\lambda_t\}_{t\geq 0}$ and $\{h(t)\}_{t\geq 0}$, where $h(t) := \int_0^t \lambda_s \, ds$. The lower plot illustrates the path of $\{M_t\}_{t\geq 0}$ as well as the observed defaults

(Den): The density of M_t, $t > 0$, is not known in closed form. However, it can be recovered from its known Laplace transform via numerical Laplace inversion. Using independence of $\{\lambda_t\}_{t\geq 0}$ and $\{\Lambda_t\}_{t\geq 0}$, the Laplace transform is given by

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(-x\,\Lambda_{\int_0^t \lambda_s\,ds\,/\Psi(1)}\Big)\Big] &= \mathbb{E}\Big[\exp\Big(-\frac{\Psi(x)}{\Psi(1)}\int_0^t \lambda_s\,ds\Big)\Big] \\ &= e^{\alpha\left(\Psi(x)/\Psi(1),t\right)+\beta\left(\Psi(x)/\Psi(1),t\right)\lambda_0}, \quad x \geq 0,\end{aligned}$$

where α and β are given as in (10.9) and (10.8). Since the Lévy subordinator is typically specified in such a way that Ψ has a simple form, the model is as convenient to work with as the classical intensity-based approach.

10.5 Extensions of the CIID Model

This section illustrates classical and new model extensions, formulated in the spirit of the unified stochastic framework of Sect. 10.2. This allows to easily adopt extensions from one model class to another. Note, however, that in most cases the convenient large homogeneous portfolio approximation, which is the major selling point of CIID models, is lost.

10.5.1 Hierarchical Dependence Structures

CIID implies exchangeability, an assumption that one might question from an economic point of view. For instance, it is reasonable to assume that companies in the same geographic region (or in the same industry sector) are affected by similar risk factors. Mathematically speaking, to construct a hierarchical model one partitions all firms in J groups—given some economic criterion. Then, all firms are affected by a global factor. In addition to that, specific factors affecting certain subgroups are introduced. Such a model can be translated into our framework, however, the CIID structure is given up in exchange for a hierarchical model structure. Formally, denote by $d_1, \dots, d_J$ the number of firms in subgroup $1, \dots, J$. The default time of company i from subgroup j is denoted τ_{ij} and defined by

$$\tau_{ij} := \inf\left\{t \geq 0 : U_{ij} \leq F_t^{(j)}\right\}, \quad j = 1, \dots, J,\ i = 1, \dots, d_j, \tag{10.12}$$

where $F^{(1)}, \dots, F^{(J)}$ are group specific frailty distributions and $U_{11}, \dots, U_{d_J J}$ is a list of iid Uni[0, 1]-distributed random variables. It is reasonable to assume

$$F^{(j)} = \text{function}_j\left(\{M_t\}_{t\geq 0}, \{M_t^{(j)}\}_{t\geq 0}\right),$$

where $\{M_t\}_{t\geq 0}$ is a global factor and $\{M_t^{(j)}\}_{t\geq 0}$ is specific for group j. Within each group, the resulting dependence structure is again CIID. However, the group specific dependencies might differ from one group to another, since the group specific factors need not be iid. Firms from different groups inherit their dependence structure from the global factor. However, a large homogeneous portfolio approximation is not available anymore. To work with the model, one must instead rely on Monte-Carlo simulations based on construction (10.12). For a structural discussion of such hierarchical structures and several examples, see [46].

Example 10.3 (Nested Archimedean copulas). Nested Archimedean copulas are succesfully applied in the context of CDO pricing in [31]. Formulating models based on nested Archimedean copulas in the present language requires a positive random variable M with Laplace transform φ as global factor and group-specific independent Lévy subordinators $\Lambda^{(j)} = \{\Lambda_t^{(j)}\}_{t\geq 0}$, $j = 1, \dots, J$, with Laplace exponent Ψ_j. Then, the frailty distribution for group j is defined as

$$F_t^{(j)} := 1 - e^{-\Lambda_M^{(j)} \Psi_j^{-1}\left(\varphi^{-1}(1-p(t))\right)}, \quad t \geq 0.$$

Consequently, two default times in group j are coupled by an Archimedean survival copula with generator $\Psi_j(\varphi(x))$, two firms from different groups are coupled by an Archimedean survival copula with generator $\varphi(x)$, see [30]. One can show that the dependence within each group is at least as large as the dependence between default times of different groups.

Example 10.4 (Hierarchical scale mixture of Marshall–Olkin copulas). A new example for a hierarchical extension is to start with a Lévy subordinator $\Lambda = \{\Lambda_t\}_{t\geq 0}$ as global factor. For each group j, an independent positive random variable M_j with Laplace transform φ_j is considered as additional group-specific factor. Finally, $F^{(j)}$ is defined as

$$F_t^{(j)} := 1 - e^{-\Lambda_{M_j \varphi_j^{-1}\left(1-p(t)\right)/\Psi(1)}}, \quad t \geq 0.$$

10.5.2 Inhomogeneous Marginal Distributions

Starting from a CIID model, one possible generalization is to assume conditionally independent (but not identically distributed) default times; in short, inhomogeneous marginal distributions. Assuming that the marginal default probabilities are model input, i.e., the market frailty is a function of the term structure of default probabilities $t \mapsto p_k(t)$ and some market factor $M = \{M_t\}_{t\geq 0}$, this is achieved by defining

$$\tau_k := \inf\left\{t \geq 0 : U_k \leq F_t^{(k)} := \text{function}\big(p_k(t), M\big)\right\}, \quad t \geq 0,$$

where the marginal distribution function $\mathbb{E}\big[F_t^{(k)}\big] = p_k(t)$ is now specific to obligor k. In this case, it is still possible to exploit the fact that the resulting default times are conditionally independent. For instance,

$$\mathbb{P}(\tau_1 \leq t_1, \ldots, \tau_d \leq t_d) = \mathbb{E}\Big[\mathbb{E}\Big[\prod_{k=1}^{d} \mathbb{1}_{\{\tau_k \leq t_k\}} \big| M\Big]\Big] = \mathbb{E}\Big[\prod_{k=1}^{d} F_{t_k}^{(k)}\Big], \quad t_1, \ldots, t_d \geq 0.$$

One way to obtain the loss distribution in this case is to use the classical recursion formula, see, e.g., [5, 20], adapted to this framework:

$$\Pi_k^{M,n+1}(t) := \big(1 - F_t^{(n+1)}\big)\Pi_k^{M,n}(t) + F_t^{(n+1)}\Pi_{k-1}^{M,n}, \quad t \geq 0,\ 0 \leq n \leq d-1,$$

where at the end of the iteration, $\Pi_k^{M,d}(t)$ denotes the conditional probability (given M) of having precisely k defaults until time t, where $0 \leq k \leq d$. The iteration must be initialized with $\Pi_0^{M,0} = 1$ and $\Pi_{-1}^{M,d} = 0$. The unconditional probability is obtained by integrating out the market factor. In order to obtain a viable model, this distribution must again be tractable. Moreover, this iterative approach becomes slow and prone to rounding errors for large portfolio sizes.

An alternative way to compute the loss distribution is available when the copula behind the default times is known. In this case, it is even possible to compute the probability of having $0 \leq k \leq d$ distinct defaults up to time t, see,

e.g. [54, Theorem 10.6]. Then, we have to sum over all d choose k subsets to obtain the probability of having k (not further specified) defaults from d names—which of course, is computationally only possible for small portfolios.

10.5.3 Model Fitting Across a Term Structure of Maturities

Standardized CDO contracts are traded with maturities 3, 5, 7, and 10 years, respectively, the most liquid ones being 5 and 10 years. To price contracts with non-standard maturities consistent to market data, one has to match model and market prices across all CDO tranches and traded maturities. Since the latter is especially demanding, most investors fix some maturity and fit their model to the tranches for this maturity. However, proceeding like this for each maturity leads to different model specifications—one for each maturity—and it is not clear how to obtain arbitrage-free prices for other maturities. The present setup is well suited to allow for a bootstrapping-like routine to fit the model across all maturities, starting with the shortest maturity, and ending with the longest. To describe it, assume a given tenor structure $0 < T_1 < T_2 < \ldots < T_K$ of maturities for which CDO quotes are available. The fundamental idea is to partition the frailty distribution into distinct pieces on the intervals $[0, T_1], (T_1, T_2], \ldots, (T_{K-1}, T_K]$ and to iteratively extend the fit of the CIID-model to the next maturity. Each piece might be interpreted as a *local frailty distribution*. Therefore, the model consists of K (stochastically independent) market frailties. Combining them to an overall stochastic process $\{F_t\}_{t\geq 0}$, such that the resulting model is maturity-consistent, is done in Lemma 10.2 below, the proof is given in the Appendix.

Lemma 10.2 (Bootstrapping CIID structures across maturities). *Given K stochastically independent market frailties*

$$\{F_t^{[0,T_1]}\}_{t\geq 0}, \{F_t^{(T_1,T_2]}\}_{t\geq 0}, \ldots, \{F_t^{(T_{K-1},T_K]}\}_{t\geq 0},$$

we iteratively define the stochastic process $\{F_t\}_{t\geq 0}$ as follows: on $t \in [0, T_1]$, we let $F_t^{(1)} := F_t^{[0,T_1]}$. For $k = 2, \ldots, K$, we then let

$$F_t^{(k)} := \mathbb{1}_{\{t\in[0,T_{k-1}]\}} F_t^{(k-1)} + \mathbb{1}_{\{t\in(T_{k-1},T_k]\}} F_{T_{k-1}}^{(k-1)} \left(1 + \frac{1 - F_{T_{k-1}}^{(k-1)}}{F_{T_{k-1}}^{(k-1)}} F_{t-T_{k-1}}^{(T_{k-1},T_k]}\right), \quad t \in [0, T_k].$$

Finally, $F_t := F_t^{(K)}$ for $t \geq 0$ is an admissible market frailty, i.e. a proper distribution function for each $\omega \in \Omega$.

Lemma 10.2 guarantees the validity of an iterative calibration of a CIID-model to CDO quotes referring to different maturities $T_1 < \ldots < T_K$. Recall that the

CIID-model implied pricing formulas corresponding to quotes for maturity T_k involve expectation values of the form $\mathbb{E}[f(F_t)]$ for time points $t \leq T_k$. With $\{F_t\}_{t\geq 0}$ being specified such as in Lemma 10.2, it follows that these expectation values only depend on the stochastic factors $\{F_t^{[0,T_1]}\}_{t\geq 0}$, $\{F_t^{(T_1,T_2]}\}_{t\geq 0}, \ldots, \{F_t^{(T_{k-1},T_k]}\}_{t\geq 0}$. By iteration, the parameters of the first $k-1$ factors

$$\{F_t^{[0,T_1]}\}_{t\geq 0}, \{F_t^{(T_1,T_2]}\}_{t\geq 0}, \ldots, \{F_t^{(T_{k-2},T_{k-1}]}\}_{t\geq 0}$$

are already determined. Therefore, it is straightforward to calibrate the parameters of the kth factor $\{F_t^{(T_{k-1},T_k]}\}_{t\geq 0}$ to market quotes of maturity T_k. Depending on the specific forms of the stochastic factors, the expectation values $\mathbb{E}[f(F_t)]$ can either be computed analytically or must be solved via Monte-Carlo simulations. As an example, consider three different maturities, say 5, 7, and 10 years, and specify each of the three factors $\{F_t^{[0,5]}\}_{t\geq 0}$, $\{F_t^{(5,7]}\}_{t\geq 0}$, and $\{F_t^{(7,10]}\}_{t\geq 0}$ like in the Archimedean model of Sect. 10.2.4. This means that we have three independent, absolutely continuous and positive random variables $M_{[0,5]}$, $M_{(5,7]}$, and $M_{(7,10]}$, with corresponding Laplace transforms $\varphi_{[0,5]}$, $\varphi_{(5,7]}$, and $\varphi_{(7,10]}$. To calibrate the model to CDO quotes corresponding to contracts maturing in 5 years, all involved expectation values are integrals w.r.t. the density of $M_{[0,5]}$. Proceeding with the next maturity of 7 years, all involved expectation values are double integrals w.r.t. the product of (independent) densities of $M_{[0,5]}$ and $M_{(5,7]}$. For the third maturity, we then need to evaluate triple integrals, which is of course much more computationally burdensome. Nevertheless, conceptually the routine is straightforward and the number of different maturities considered in real-life is typically amongst $2, 3$, rendering this effort acceptable, in particular because a simultaneous fit of only one market frailty across different maturities is typically not satisfying.

10.6 Conclusion

A unified approach for CIID portfolio default models was presented. Desirable stochastic properties of these models were introduced in an axiomatic manner and discussed from an economic perspective. State-of-the-art models that fit into this context were discussed and compared with respect to these properties. Two new models were introduced. The first one was shown to unify the approaches of [44,53]. The resulting implied copula is of Archimax type. The second ansatz combines a classical intensity approach with a Lévy-based approach to allow for excess clustering and time-inhomogeneity. In both cases, one could derive the Laplace transform of the required underlying frailty distribution in closed form. Finally, several model generalizations are discussed.

Appendix

Lévy Subordinators

A Lévy subordinator $\Lambda = \{\Lambda_t\}_{t\geq 0}$ is a non-decreasing stochastic process. It starts at zero, is stochastically continuous, and has stationary and independent increments. Standard textbooks on this theory comprise [6, 11, 12, 52, 55]. A Lévy subordinator is uniquely characterized by its Laplace transforms, which admit the form

$$\mathbb{E}\big[e^{-x\,\Lambda_t}\big] = e^{-t\,\Psi(x)}, \quad \forall\, x \geq 0,\, t \geq 0,$$

for a function $\Psi : [0,\infty) \to [0,\infty)$ which has a completely monotone derivative and satisfies $\Psi(0) = 0$, see [25, p. 450]. The function Ψ is called *Laplace exponent* of Λ and is strictly increasing unless $\Lambda_t \equiv 0$.

Proof of Lemma 10.1

The first statement follows immediately from the Theorem of Glivenko–Cantelli, see [41, p. 20]: conditioned on $\{F_t\}_{t\geq 0}$, $\{L_t\}_{t\geq 0}$ is precisely the empirical distribution function of the law $\{F_t\}_{t\geq 0}$ based on d samples. Hence,

$$\begin{aligned}\mathbb{P}\Big(\lim_{d\to\infty}\sup_{t\geq 0}\big|F_t - L_t\big| = 0\Big) &= \mathbb{E}\Big[\mathbb{P}\Big(\lim_{d\to\infty}\sup_{t\geq 0}\big|F_t - L_t\big| = 0\,\Big|\,\{F_t\}_{t\geq 0}\Big)\Big]\\ &= \mathbb{E}[1] = 1.\end{aligned}$$

For the second statement, immediate computations show that

$$\mathbb{E}\big[L_t^2\big] = \frac{1}{d}\,\mathbb{E}[F_t] + \frac{d-1}{d}\,\mathbb{E}[F_t^2], \quad \mathbb{E}[L_t\,F_t] = \mathbb{E}[F_t^2],$$

which implies that

$$\int_{[0,T]}\mathbb{E}\big[(L_t - F_t)^2\big]\,dt = \frac{1}{d}\int_{[0,T]}\big(\mathbb{E}[F_t] - \mathbb{E}[F_t^2]\big)\,dt \quad \overset{d\to\infty}{\longrightarrow} \quad 0.$$

The claim is established. □

Proof of Zero Tail Dependence in the Model of Sect. 10.2.5

The first step is to compute for β and α, given in (10.8) and (10.9), that

$$\alpha'(x,0) := \lim_{t\downarrow 0}\frac{d}{dt}\alpha(x,t) \overset{(*)}{=} 0, \quad \beta'(x,0) := \lim_{t\downarrow 0}\frac{d}{dt}\beta(x,t) \overset{(**)}{=} -x, \quad x > 0.$$

Both $(*)$ and $(**)$ are tedious computations that become simpler with the identities

$$c(x)+d(x)=-\frac{\sqrt{\kappa^2+2\sigma^2 x}}{x},\quad c(x)-d(x)=-\frac{\kappa}{x},$$
$$c(x)-d(x)=\frac{\sigma^2}{2x},\quad \frac{b(x)}{c(x)+d(x)}=x.$$

Then, using formula (10.7), one computes with L'Hospital's rule that[12]

$$\lambda_l=\lim_{t\downarrow 0}\frac{\mathbb{E}\big[F_t^2\big]}{\mathbb{E}[F_t]}=\lim_{t\downarrow 0}\left\{1+\frac{e^{\alpha(2,t)+\beta(2,t)\,\lambda_0}-e^{\alpha(1,t)+\beta(1,t)\,\lambda_0}}{1-e^{\alpha(1,t)+\beta(1,t)\,\lambda_0}}\right\}$$
$$=2-\frac{\alpha'(2,0)+\beta'(2,0)\,\lambda_0}{\alpha'(1,0)+\beta'(1,0)\,\lambda_0}=0.$$

Proof of Theorem 10.1

For $t_1,\ldots,t_d\in[0,\infty)$ with ordered list $t_{(1)}\le\ldots\le t_{(d)}$ and $t_{(0)}:=0$ one has

$$\sum_{i=1}^{d}(d+1-i)\left(\Lambda_{t_{(i)}}-\Lambda_{t_{(i-1)}}\right)=\sum_{i=1}^{d}(d+1-i)\,\Lambda_{t_{(i)}}-\sum_{i=0}^{d-1}(d-i)\,\Lambda_{t_{(i)}}=\sum_{i=1}^{d}\Lambda_{t_i}.$$

Since Λ is a Lévy process, the vector of increments $\left(\Lambda_{t_{(d)}}-\Lambda_{t_{(d-1)}},\ldots,\Lambda_{t_{(1)}}-\Lambda_{t_{(0)}}\right)$ has independent components and $\Lambda_{t_{(i)}}-\Lambda_{t_{(i-1)}}$ is equal in distribution to $\Lambda_{t_{(i)}-t_{(i-1)}}$. Consequently

$$\mathbb{E}\left[e^{-\sum_{i=1}^{d}\Lambda_{t_i}}\right]=\prod_{i=1}^{d}\mathbb{E}\left[e^{-(d+1-i)\,\Lambda_{(t_{(i)}-t_{(i-1)})}}\right]=\prod_{i=1}^{d}e^{-(t_{(i)}-t_{(i-1)})\,\Psi(d+1-i)}.$$

Secondly, compute the joint survival function (using the above identity)

[12]The above argument makes use of the explicit form of the Laplace transform of an integrated basic affine intensity. If the intensity $\{\lambda_t\}_{t\ge 0}$ is specified differently, one might end up with positive tail dependence. Giving one example, assume that $\lambda_t:=\bar{M}$ for a positive random variable $\bar{M}$ with Laplace transform φ, i.e. $M_t=\bar{M}\,t, t\ge 0$. The resulting dependence structure is of Archimedean kind, and there are choices for $\bar{M}$ that imply positive tail dependence. A related observation is that for $\lambda_t:=\bar{M}\,\frac{\partial}{\partial t}\big(\varphi^{-1}(1-p(t))\big)$, the model of [53] is a special case of the intensity approach.

$$
\begin{aligned}
G(t_1,\dots,t_d) &:= \mathbb{P}\big(\tau_1 > t_1,\dots,\tau_d > t_d\big)\\
&= \mathbb{P}\big(\epsilon_1 > \Lambda_{\bar{M}\,\varphi^{-1}(1-p(t_1))/\Psi(1)},\dots,\epsilon_d > \Lambda_{\bar{M}\,\varphi^{-1}(1-p(t_d))/\Psi(1)}\big)\\
&= \mathbb{E}\left[\exp\Big(-\sum_{i=1}^{d}\Lambda_{\bar{M}\,\varphi^{-1}(1-p(t_i))/\Psi(1)}\Big)\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\exp\Big(-\sum_{i=1}^{d}\Lambda_{\bar{M}\,\varphi^{-1}(1-p(t_i))/\Psi(1)}\Big)\Big|\bar{M}\right]\right]\\
&= \mathbb{E}\left[\exp\Big(-\frac{\bar{M}}{\Psi(1)}\sum_{i=1}^{d}\Psi(d+1-i)\big(\varphi^{-1}(1-p(t_{(i)}))-\varphi^{-1}(1-p(t_{(i-1)}))\big)\Big)\right]\\
&= \varphi\Big(\frac{1}{\Psi(1)}\sum_{i=1}^{d}\Psi(d+1-i)\big(\varphi^{-1}(1-p(t_{(i)}))-\varphi^{-1}(1-p(t_{(i-1)}))\big)\Big)\\
&= \varphi\Big(\frac{1}{\Psi(1)}\sum_{i=1}^{d}\varphi^{-1}(1-p(t_{(d+1-i)}))\big(\Psi(i)-\Psi(i-1)\big)\Big).
\end{aligned}
$$

The last step involves expanding the sum of differences to two sums and shifting the summation index in the second sum by one. The resulting sums can then be recombined using $\Psi(0) = 0$.

Thirdly, for the margins one obtains using similar arguments

$$\mathbb{P}\big(\tau_i > t\big) = \mathbb{P}\big(\epsilon_i > \Lambda_{\bar{M}\,\varphi^{-1}(1-p(t))/\Psi(1)}\big) = 1-p(t), \quad i=1,\dots,d, \quad t \geq 0.$$

Thus, τ_i is distributed according to $p(t)$. Finally, using the survival analogue of Sklar's Theorem, see [49, p. 195], there exists a unique copula $\hat{C}$, called the survival copula of $(\tau_1,\dots,\tau_d)$, which satisfies

$$G(t_1,\dots,t_d) = \hat{C}\big(1-p(t_1),\dots,1-p(t_d)\big).$$

Testing the copula claimed in (10.10), one finds

$$\hat{C}\big(1-p(t_1),\dots,1-p(t_d)\big) = \varphi\Big(\frac{1}{\Psi(1)}\sum_{i=1}^{d}\varphi^{-1}(1-p(t_{(d+1-i)}))\big(\Psi(i)-\Psi(i-1)\big)\Big).$$

Thus, the claim is established by the uniqueness of the survival copula. □

Proof of Lemma 10.2

The claimed composition of different distribution functions to a new one is based on an elementary decomposition of a distribution function. Considering only two intervals $[0, T_1]$ and $(T_1, T_2]$, it is easy to verify for $t \in [0, T_2]$ that

$$\mathbb{P}(\tau \leq t) = \mathbb{1}_{\{t \in [0,T_1]\}} \mathbb{P}(\tau \leq t) + \mathbb{1}_{\{t \in (T_1,T_2]\}} \mathbb{P}(\tau \leq T_1) \left(1 + \frac{\mathbb{P}(\tau > T_1)}{\mathbb{P}(\tau \leq T_1)} \mathbb{P}(\tau \leq t \mid \tau > T_1)\right).$$

The crucial observation from this elementary computation is that having determined the distribution $p_1(t) := \mathbb{P}(\tau \leq t)$ on $[0, T_1]$ already, to determine the distribution on $[0, T_2]$ it suffices to determine $p_2(u) := \mathbb{P}(\tau \leq u + T_1 \mid \tau > T_1)$ for $u \in (0, T_2 - T_1]$. However, the function p_2 is a proper distribution function on $[0, \infty)$. Hence, starting with two given distribution functions p_1, p_2, the claimed composition of those yields a proper distribution function with the interpretation that p_2 is the conditional distribution in case of survival until time T_1. The general case $K > 2$ is now easily obtained by iterating the above argument.

References

1. Abate, J., Choudhury, G.L., Whitt, W.: On the Laguerre method for numerically inverting Laplace transforms. Informs. J. Comput. **8**(4), 413–427 (1996)
2. Abate, J., Valkó, P.: Comparison of sequence accelerators for the Gaver Method of numerical Laplace transform inversion. Comput. Math. Appl. **48**, 629–636 (2002)
3. Abate, J., Valkó, P.: Multi-precision Laplace transform inversion. Int. J. Numer. Methods Eng. **60**, 979–993 (2004)
4. Albrecher, H., Ladoucette, S.A., Schoutens, W.: A generic one-factor Lévy model for pricing synthetic CDOs. Advances in Mathematical Finance (Applied and Numerical Harmonic Analysis) Birkhäuser Boston 259–278 (2007)
5. Andersen, L., Sidenius, J., Basu, S.: All your hedges in one basket. Risk Mag. 67–72 (2003)
6. Applebaum, D.: Lévy processes and Stochastic Calculus. Cambridge University Press, Cambridge (2004)
7. Barlow, R.E., Proschan, F.: Statistical Theory of Reliability and Life Testing: Probability Models. Holt Rinehart and Winston, New York (1975)
8. Baxter, M.: Dynamic modelling of single-name credits and CDO tranches. Working paper - Nomura Fixed Income Quant Group retrievable from http://www.nomura.com/resources/europe/pdfs/dynamiccdomodelling2.pdf (2006). Accessed January 14, 2012
9. Bernhart, G., Escobar Anel, M., Mai, J.-F., Scherer, M.: Default models based on scale mixtures of Marshall-Olkin Copulas: properties and applications. Metrika **76**, 179–203 (2013)
10. Bernhart, G., Mai, J.-F., Scherer, M., Schenk, S.: The density of distributions from the Bondesson class. J. Comput. Financ. (2012, Forthcoming)
11. Bertoin, J.: Lévy Processes. Cambridge University Press, Cambridge (1996)
12. Bertoin, J.: Subordinators: Examples and Applications. Springer École d'Été de Probabilités de Saint-Flour XXVII-1997. Lect. Notes Math. **1717**, 1–91 (1999)

13. Burtschell, X., Gregory, J., Laurent, J-P.: A comparative analysis of CDO pricing models. In: Meissner, G. (ed.) The Definitive Guide to CDOs, Chapter 15, pp. 389–427, Risk Books, Incisive Media, London (2009)
14. Capéraà, P., Fougères, A.-L., Genest, C.: Bivariate distributions with given extreme value attractor. J. Multivar. Anal. **72**, 30–49 (2000)
15. Carr, P., Geman, H., Madan, D., Yor, M.: Stochastic volatility for Lévy processes. Math. Financ. **13**, 345–382 (2003)
16. Charpentier, A., Segers, J.: Tails of multivariate Archimedean copulas. J. Multivar. Anal. **100**(7), 1521–1537 (2009)
17. Cuadras Avellanas, C.M., Augé, J.: A continuous general multivariate distribution and its properties. Comm. Stat. A Theory Method. **10**(4), 339–353 (1981)
18. Das, S.R., Duffie, D., Kapadia, N., Saita, L.: Common failings: How corporate defaults are correlated. J. Financ. **62**(1), 93–117 (2007)
19. De Finetti, B.: La prévision: Ses lois logiques ses sources subjectives. Annales de l'Institut Henri Poincaré **7**, 1–68 (1937)
20. Dobránszky, P., Schoutens, W.: Generic Lévy one-factor models for the joint modelling of prepayment and default: Modelling LCDX. Technical report 08-03. Section of Statistics K.U. Leuven (2009)
21. Duffie, D., Singleton, K.: Modeling term structures of defaultable bonds. Rev. Financ. Stud. **12**(4), 687–720 (1999)
22. Duffie, D., Pan, J., Singleton, K.: Transform analysis and asset pricing for affine jump-diffusions. Econometrica **68**, 1343–1376 (2000)
23. Duffie, D., Gârleanu, N.: Risk and valuation of collateralized debt obligations. Financ. Anal. J. **57**(1), 41–59 (2001)
24. Duffie, D., Saita, L., Wang, K.: Multi-period corporate default prediction with stochastic covariates. J. Financ. Econ. **83**, 635–665 (2007)
25. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. II, 2nd edn. Wiley, New York (1966)
26. Frey, R., McNeil, A.J.: Modelling Dependent Defaults. Working paper available at http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=273 (2001). Accessed January 14, 2012
27. Galambos, J., Kotz, S.: Characterizations of probability distributions. In: Lecture Notes in Mathematics, vol. 675. Springer, Heidelberg (1978)
28. Giesecke, K.: A simple exponential model for dependent defaults. J. Fixed Income **13**(3), 74–83 (2003)
29. Guégan, D., Houdain, J.: Collateralized Debt Obligations pricing and factor models: A new methodology using Normal Inverse Gaussian distributions. Research Report IDHE-MORA ENS Cachan 7 (2005)
30. Hering, C., Hofert, M., Mai, J.-F., Scherer, M.: Constructing hierarchical Archimedean copulas with Lévy subordinators. J. Multivar. Anal. **101**(6), 1428–1433 (2010)
31. Hofert, M., Scherer, M.: CDO pricing with nested Archimedean copulas. Quant. Financ. **11**(5), 775–787 (2011)
32. Hull, J., White, A.: Valuation of a CDO and an n-th to default CDS without monte carlo simulation. J. Derivatives **12**(2), 8–23 (2004)
33. Jarrow, R.A., Turnbull, S.: Pricing derivatives on financial securities subject to credit risk. J. Financ. **50**(1), 53–85 (1995)
34. Joe, H.: Multivariate Models and Dependence Concepts. Chapman and Hall/CRC, University of Bristish Colombia, Vancouver (1997)
35. Kalemanova, A., Schmid, B., Werner, R.: The normal inverse gaussian distribution for synthetic CDO pricing. J. Derivatives **14**(3), 80–93 (2007)
36. Kiesel, R., Lutz, M.: Efficient pricing of constant maturity swap spread options in a stochastic volatility LIBOR market model. J. Comput. Financ. **14**(4), 37–72 (2011)
37. Lando, D.: On cox processes and credit risky securities. Rev. Derivatives Res. **2**(2–3), 99–120 (1998)

38. Li, D.X.: On default correlation: A copula function approach. J. Fixed Income **9**(4), 43–54 (2000)
39. Li, H.: Orthant tail dependence of multivariate extreme value distributions. J. Multivar. Anal. **100**(1), 243–256 (2009)
40. Lindskog, F., McNeil, A.J.: Common poisson shock models: Applications to insurance and credit risk modelling. Astin Bull. **33**(2), 209–238 (2003)
41. Loève, M.: Probability Theory I, 4th edn. Springer, Heidelberg (1977)
42. Madan, D.B., Unal, H.: Pricing the risks of default. Rev. Derivatives Res. **2**(2–3), 121–160 (1998)
43. Mai, J.-F., Scherer, M.: Reparameterizing Marshall–Olkin copulas with applications to sampling. J. Stat. Comput. Simul. **81**(1), 59–78 (2011)
44. Mai, J.-F., Scherer, M.: A tractable multivariate default model based on a stochastic time-change. Int. J. Theo. Appl. Financ. **12**(2), 227–249 (2009)
45. Mai, J.-F., Scherer, M.: Lévy-frailty copulas. J. Multivar. Anal. **100**(7), 1567–1585 (2009)
46. Mai, J.-F., Scherer, M.: H-extendible copulas. J. Multivar. Anal. **11**, 151–160 (2012)
47. Marshall, A.W., Olkin, I.: A multivariate exponential distribution. J. Am. Stat. Assoc. **62**, 30–44 (1967)
48. Marshall, A.W., Olkin, I.: Families of multivariate distributions. J. Am. Stat. Assoc. **83**, 834–841 (1988)
49. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management. Princeton University Press, Princeton (2005)
50. Merton, R.: On the pricing of corporate debt: The risk structure of interest rates. J. Financ. **29**, 449–470 (1974)
51. Moosbrucker, T.: Pricing CDOs with Correlated Variance Gamma Distributions. Research Report Department of Banking University of Cologne, Albertus-Magnus-Platz (2006)
52. Sato, K.-I.: Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press, Cambridge (1999)
53. Schönbucher, P.J.: Taken to the limit: Simple and not-so-simple loan loss distributions. Working paper retrievable from http://www.gloriamundi.org/picsresources/pjs.pdf (2002). Accessed January 14, 2012
54. Schönbucher, P.J.: Credit Derivatives Pricing Models. Wiley, New York (2003)
55. Schoutens, W.: Lévy processes in Finance: Pricing financial derivatives. Wiley Series in Probability and Statistics. Wiley, Chichester, England (2003)
56. Talbot, A.: The accurate numerical inversion of Laplace transforms. J. Inst. Math. Appl. **23**, 97–120 (1979)
57. Vasicek, O.: Probability of loss on loan portfolio. Working paper KMV Corporation (1987)
58. Yu, F.: Correlated defaults in intensity-based models. Math. Financ. **17**(2), 155–173 (2007)

Chapter 11
Copula-Based Models for Multivariate Discrete Response Data

Aristidis K. Nikoloulopoulos

Abstract In this survey we review copula-based models and methods for multivariate discrete data modeling. Advantages and disadvantages of recent contributions are summarized and a general modeling procedure is suggested in this context.

11.1 Introduction

One goal in the theory of dependence modeling and multivariate copulas is to develop copula-based models and inferential procedures for multivariate discrete responses with covariates. Discrete response types include binary, ordinal categorical, and count data. Examples of data include, among others, familial data (measurements for each member of an extended multi-generation family) in medical genetics applications, repeated measurements in health studies, item response data in psychometrics applications, etc. These multivariate discrete data have different dependence structures including features such as negative dependence. To this end, the desiderata properties of multivariate copula families for modeling multivariate discrete data are given below (see also [19, 45, 48]):

P1: Wide range of dependence, allowing both positive and negative dependence.
P2: Flexible dependence, meaning that the number of bivariate marginals is (approximately) equal to the number of dependence parameters.
P3: Computationally feasible cumulative distribution function (cdf) for likelihood estimation.
P4: Closure property under marginalization, meaning that lower-order marginals belong to the same parametric family.

A.K. Nikoloulopoulos (✉)
School of Computing Sciences, University of East Anglia, Norwich, UK
e-mail: A.Nikoloulopoulos@uea.ac.uk

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*,
Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_11,

P5: No joint constraints for the dependence parameters, meaning that the use of covariate functions for the dependence parameters is straightforward.

In the existing literature, none of the existing parametric families of multivariate copulas satisfy all these conditions; hence there are many challenges for copula-based models for discrete response data.

Multivariate copulas for discrete response data have been around a long time, e.g., in [19], and earlier for some simple copula models. There are also papers with simple bivariate (multivariate) discrete distributions where actually the construction is more or less a copula, but the authors do not refer to copulas, e.g., [6, 27, 34]. Simple parametric families of copulas satisfy P3; hence the joint likelihood is straightforward to derive from the probability mass function (pmf) as a finite difference of the cdf, but they provide limited dependence; see, e.g., the contributions in [8, 29, 32, 44–46, 62].

The multivariate normal (MVN) copula generated by the MVN distribution inherits the useful properties of the latter, thus allowing a wide range for dependence (P1–P2) and overcomes the drawback of limited dependence inherent in simple parametric families of copulas [41]. The MVN copula with discrete margins has been in use for a considerable length of time, e.g. [19], and much earlier in the biostatistics [2], psychometrics [35], econometrics [14], and literature. It is usually known as a multivariate, or multinomial, probit model. The multivariate probit model is a simple example of the MVN copula with univariate probit regressions as the marginals. Implementation of the MVN copula for discrete data (discretized MVN) is possible, but not easy, because the MVN distribution as a latent model for discrete response requires rectangle probabilities based on high-dimensional integrations or their approximations [45].

Similarly, this is the case for other elliptical copulas which have also been applied to discrete data [10] and lead to a model with more probabilities in the joint upper and joint lower tails than expected with discretized MVN. Another interesting contribution and flexible modeling approach are the pair-copula constructions as developed in [48] which can also allow asymmetries, i.e., more probability in joint upper or lower tails.

The remainder of the survey proceeds as follows. Section 11.2 sets the notation and provides background material on copulas for multivariate discrete response data. In Sect. 11.3 the parametric families of copulas used so far in the literature for modeling-dependent discrete data are presented. Their properties, which inherit to the copula-based models advantages and disadvantages, are described. Section 11.4 discusses estimation methods and classifies them depending on the properties of the parametric family of copulas. In Sect. 11.5 the Kendall's tau for discrete response data is presented. We conclude this survey with some discussion.

11.2 Multivariate Discrete Distributions via Copulas

By definition, a d-variate copula $C(u_1, \ldots, u_d)$ is a multivariate cdf with uniform marginals on the unit interval; see, e.g., [19, 37]. From Sklar [51], in order to express a multivariate discrete distribution for the discrete (binary, count, etc.) vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ given a vector of covariates $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_d)$ with $\mathbf{x}_j \in \mathbb{R}^{p_j}$, $j = 1, \ldots, d$, one needs to combine discrete (Bernoulli, Poisson, etc.) marginal distribution functions $F_{Y_1}(y_1; \mathbf{x}_1), \ldots, F_{Y_d}(y_d; \mathbf{x}_d)$ with a d-variate copula such for all $\mathbf{y} = (y_1, \ldots, y_d)$,

$$H_{\mathbf{Y}}(\mathbf{y}; \mathbf{x}) = \Pr(Y_1 \leq y_1, \ldots, Y_d \leq y_d; \mathbf{x}) = C\Big(F_{Y_1}(y_1; \mathbf{x}_1), \ldots, F_{Y_d}(y_d; \mathbf{x}_d)\Big). \tag{11.1}$$

Because the margins are discrete, as emphasized in [9], there are many possible copulas, but all of these coincide on the closure of $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_d)$, where $\mathrm{Ran}(F_j)$ denotes the range of F_j.

For discrete random vectors, multivariate probabilities of the form $h_{\mathbf{Y}}(\mathbf{y}; \mathbf{x}) = \Pr(Y_1 = y_1, \ldots, Y_d = y_d; \mathbf{x})$ involve 2^d finite differences of $H_{\mathbf{Y}}(\mathbf{y}; \mathbf{x})$. Let $\mathbf{s} = (s_1, \ldots, s_d)$ be vertices where each s_j is equal to either y_j or $y_j - 1$, $j = 1, \ldots, d$. Then the joint pmf $h_{\mathbf{Y}}(\cdot)$ is given by,

$$h_{\mathbf{Y}}(\mathbf{y}; \mathbf{x}) = \sum \mathrm{sgn}(\mathbf{s}) C\Big(F_{Y_1}(s_1; \mathbf{x}_1), \ldots, F_{Y_d}(s_d; \mathbf{x}_d)\Big), \tag{11.2}$$

where the sum is taken over all vertices $\mathbf{s}$, and $\mathrm{sgn}(\mathbf{s})$ is given by,

$$\mathrm{sgn}(\mathbf{s}) = \begin{cases} 1, \text{ if } s_j = y_j - 1 \text{ for an even number of } j\text{'s.} \\ -1, \text{ if } s_j = y_j - 1 \text{ for an odd number of } j\text{'s.} \end{cases}$$

Therefore likelihood inference for discrete data is simpler for copulas with computationally feasible form of the cdf (P3). Essentially, the specification of the multivariate discrete distribution in (11.1), exploiting the use of copula functions, provides complete inference, i.e., maximum likelihood estimation, calculation of joint and conditional probabilities, and standard goodness of fit procedures.

11.3 Copula-Based Models for Discrete Response Data

In this section, we review several existed copula-based models for discrete data [8, 10, 29, 32, 44, 46, 48, 56, 57, 62]. The authors assumed that the copula C comes from a specific parametric family or class of copulas; hence, its properties are inherited to the model. Although C is not uniquely defined outside the Cartesian product of the ranges of the marginal distribution functions, there is no harm in assuming that it arises from a parametric class of copulas [10, 48].

If the same copula applies for all clusters and have covariates on board, in particular continuous covariates, the number of potential values is so high and the copula becomes unique in the limit (infinite clusters). However, generally speaking the copula is not unique (identifiable) in the discrete case except on the range of the marginals [9]. The non-identifiability is a separate theoretical issue and does not have any bearing on copula dependence modeling for discrete data [19, 55], which is the main focus of this survey.

11.3.1 Archimedean

Meester and Mackay [32] proposed a parametric model for cluster correlated categorical (binary and ordinal) data, based on the d-variate Frank copula. The Frank copula belongs to the large class of Archimedean copulas. Multivariate Archimedean copulas, see, e.g., [19], have the form,

$$C(u_1,\ldots,u_d\,;\,\theta)=\phi\left(\sum_{j=1}^{d}\phi^{-1}(u_j\,;\,\theta)\,;\,\theta\right), \tag{11.3}$$

where the generator $\phi(u\,;\,\theta)$ is the Laplace transform (LT) of a univariate family of distributions of positive random variables indexed by the parameter θ, such that $\phi(\cdot)$ and its inverse have closed forms. One may refer to [31] for a general definition of an Archimedean copula where the generator is more general than an LT but still needs to satisfy certain regularity conditions. Hence, one can relax the completely monotone condition to d times alternating in sign, then Archimedean copulas based on extensions of LTs are obtained, and some of these might have negative dependence.

The Frank copula in the Archimedean family [19, page 141] has Laplace transform $\phi_F(t)=-\theta^{-1}\log\left[1-(1-e^{-\theta})e^{-t}\right]$, $\theta>0$. This d-variate copula is permutation-symmetric in the d arguments, thus it is a distribution for exchangeable uniform random variables on the unit interval. The Frank copula interpolates from the independence ($\theta\to 0$) to the Fréchet upper (perfect positive dependence) bound ($\theta\to\infty$). For extension of $\phi_F(t)$ for $\theta<0$, the Frank family extends to countercomonotonicity ($\theta\to-\infty$) for $d=2$ and only a little into negative dependence for dimensions $d\geq 3$ [19, 31]. Joe [19, pages 158–159] shows how narrow is the range of negative dependence for trivariate Frank and beyond. Hence, for bivariate discrete data a model based on Frank copula is quite popular [3, 28, 30, 32]. For another application of d-variate Frank copula for familial binary data, see [57], and for applications of various Archimedean copula-based models for multivariate count data, see [46].

To sum up, d-variate ($d>2$) Archimedean copulas satisfy properties P3, P4, and P5, but not P1 and P2, because they allow only for exchangeable dependence, and its range becomes narrower as the dimension increases.

11.3.2 Partially Symmetric

Zimmer and Trivedi [62] and Nikoloulopoulos and Karlis [46] modeled dependent discrete response data using partially symmetric copulas. Joe [17] extended multivariate Archimedean copulas to a more flexible class of copulas using nested LTs, the so-called partially-symmetric d-variate copulas with $d-1$ dependence parameters. Note in passing that these copulas are also called Hierarchical or nested Archimedean copulas; see, e.g., [4, 15, 16]. The multivariate form has a complex notation, so we present the trivariate extension of (11.3) to help the exposition. The trivariate form is given by,

$$C(u_1,u_2,u_3)=\phi_1\left(\phi_1^{-1}\circ\phi_2\left(\phi_2^{-1}(u_1)+\phi_2^{-1}(u_2)\right)+\phi_1^{-1}(u_3)\right), \tag{11.4}$$

where ϕ_1,ϕ_2 are LTs and $\phi_1^{-1}\circ\phi_2\in \text{Ł}_\infty^\star=\{\omega:[0,\infty)\longrightarrow[0,\infty)|\omega(0)=0,\ \omega(\infty)=\infty,(-1)^{j-1}\omega^j\geq 0,j=1,\ldots,\infty\}$. From the above formula it is clear that (11.4) has (1,2) bivariate margin of the form (11.3) with LT $\phi_2(\cdot;\theta_2)$, and (1,3), (2,3) bivariate margins of the form (11.3) with LT $\phi_1(\cdot;\theta_1)$. As the dimension increases there are many possible LT nestings. Bivariate margins associated with LTs that are more nested are larger in concordance than those that are less nested. For example, for (11.4) the (1,2) bivariate margins is more dependent (concordant) than the remaining bivariate margins.

Although partially symmetric copulas have a closed form cdf, they do not provide flexible dependence due to moderate number of dependence parameters ($d-1$ distinct parameters) and do not allow for negative dependence by construction. To sum up, partially symmetric copulas satisfy properties P3, P4, and P5, but not P1 and P2.

11.3.3 Farlie–Gumbel–Morgenstern

Gauvreau and Pagano [8] considered a d-variate Farlie–Gumbel–Morgenstern (FGM) copula . The multivariate FGM copula is,

$$\begin{aligned}C(u_1,\ldots,u_d;\boldsymbol{\Theta})=\Big(1&+\sum_{1\leq j<k\leq d}^{d}\theta_{jk}(1-u_j)(1-u_k)\\&+\sum_{1\leq j<k<l\leq d}^{d}\theta_{jkl}(1-u_j)(1-u_k)(1-u_l)+\cdots\\&+\theta_{12\cdots d}(1-u_1)(1-u_2)\cdots(1-u_d)\Big)\prod_{j=1}^{d}u_j,\end{aligned} \tag{11.5}$$

where $\boldsymbol{\Theta}=\{\theta_{jk},\theta_{jkl},\cdots,\theta_{12\cdots d}\}$. For more details, see [25, 26].

However, the conditions on the parameters $\boldsymbol{\Theta}$ so that FGM is indeed a copula are not investigated in [8]. The conditions on the parameters so that FGM is indeed a copula can be obtained by considering the 2^d cases for $u_j = 0$ or 1, $j = 1 \ldots, d$, and verifying that the copula density is positive, i.e. $c(u_1, \ldots, u_d) \geq 0$.

To simplify the notation a simpler version of a d-variate FGM copula that does not include higher order terms is given below,

$$C(u_1, \ldots, u_d; \theta_{jk} : 1 \leq j < k \leq d) = \Big(1 + \sum_{1 \leq j < k \leq d} \theta_{jk}(1-u_j)(1-u_k)\Big) \prod_{j=1}^{d} u_j. \tag{11.6}$$

It has density function,

$$c(u_1, \ldots, u_d; \theta_{jk} : 1 \leq j < k \leq d) = 1 + \sum_{j<k}^{d} \theta_{jk}(1-2u_j)(1-2u_k).$$

The necessary and sufficient conditions on the parameters θ_{jk} so that (11.6) is a copula are straightforward. For $d = 3$, the conditions can be conveniently summarized as follows: $1 + \theta_{12} + \theta_{13} + \theta_{23} \geq 0, 1 + \theta_{12} \geq \theta_{13} + \theta_{23}, 1 + \theta_{13} \geq \theta_{12} + \theta_{23}, 1 + \theta_{23} \geq \theta_{12} + \theta_{13}$, or more succinctly $-1 + |\theta_{12} + \theta_{23}| \leq \theta_{13} \leq 1 - |\theta_{12} - \theta_{23}|, -1 \leq \theta_{12}, \theta_{13}, \theta_{23} \leq 1$. Similar conditions for higher dimension $d > 3$ can also be obtained by considering the 2^d cases for $u_j = 0$ or 1, $j = 1 \ldots, d$, and verifying that $c(u_1, \ldots, u_d) \geq 0$. For further details see [60].

In addition to the joint constraints limitation, the FGM copula has a limited range of dependence and is inappropriate for general modeling unless the responses are weakly dependent. Even for the bivariate case with no joint constraints between the parameters, it is easy to see that the range of dependence is limited. Gauvreau and Pagano [8] studied the range of the dependence parameter, say θ_{12}, in terms of Pearson's correlation parameter for binary data, say ρ_{12}, through the relation

$$\rho_{12} = \theta_{12}\sqrt{\pi_1\pi_2(1-\pi_1)(1-\pi_2)},$$

where $\pi_j = \Pr(Y_j = 1)$, $j = 1, 2$. However, since $-1 \leq \theta_{12} \leq 1$ the bounds of the Pearson's correlation are,

$$\pm\sqrt{\pi_1\pi_2(1-\pi_1)(1-\pi_2)}.$$

Li and Wong [29] used a similar parametric family of copulas with the FGM copula in [8],

$$C(u_1, \ldots, u_d; \theta_{jk} : 1 \leq j < k \leq d) = \prod_{j=1}^{d} u_j \prod_{1 \leq j < k \leq d}^{d} \Big(1 + \theta_{jk}(1-u_j)(1-u_k)\Big). \tag{11.7}$$

However, the conditions on the parameters θ_{jk} so that (11.7) is a copula are not investigated by the authors. For $d = 3$, the necessary conditions can be conveniently summarized as follows: $-1 \leq \theta_{12}, \theta_{13}, \theta_{23} \leq 1$ and $-(1 + \theta_{j\ell}) \leq \theta_{jk} + \theta_{k\ell} \leq \min(1, 1 + \theta_{j\ell} + \theta_{jk}\theta_{k\ell})$ for all different permutations of (j, k, ℓ) in $(1, 2, 3)$, see [39].

The sufficient conditions (nonnegativity of the entire density function in $[0.1]^d$) are hard to prove for $d > 2$ because the density of (11.7) is a higher order polynomial function (quadratic for $d = 3$, cubic for $d = 4$, etc.) of each u_j taken separately. However, considering the 2^d cases for $u_j = 0$ or 1, $j = 1, \ldots, d$, and verifying that the copula density of (11.7) is positive provides the necessary conditions on the parameters for the copula in (11.7); these are also sufficient for $d = 2$ since the bivariate density is a linear function of each u_j taken separately, see [25, Sect. 4, page 419]. In addition to the joint constraints limitation, the copula in (11.7) has a limited range of dependence as the FGM copula in (11.5) or (11.6) and it resembles FGM for the bivariate case.

To sum up, the FGM copulas satisfy properties P3 and P4, but not P1, P2, and P5. Because of the dependence range limitation, the FGM copulas are not very useful for general modeling unless the responses are weakly dependent.

11.3.4 Finite Normal Mixture

Nikoloulopoulos and Karlis [45] modeled multivariate count data proposing a copula generated by a mixture of two independent MVN distributions. The finite normal mixture copula cdf takes the form,

$$C(u_1, \ldots, u_d; \pi, \mu_1 = 1, \ldots, \mu_d) = \mathscr{F}_{1\ldots d}\left[\mathscr{F}_1^{-1}(u_1; \pi, 1), \ldots, \mathscr{F}_d^{-1}(u_d; \pi, \mu_d); \pi, \boldsymbol{\mu}\right],$$

where

$$\mathscr{F}_{1\ldots d}(\cdot; \pi, \boldsymbol{\mu}) = \pi\Phi_d(\cdot; \boldsymbol{\mu}, \mathbf{I}_d) + (1 - \pi)\Phi_d(\cdot; -\boldsymbol{\mu}, \mathbf{I}_d) \tag{11.8}$$

is the d-variate cdf of a mixture of two d-variate normal cdfs with mixing probability π, $\Phi_d(\cdot; \boldsymbol{\mu}, \mathbf{I}_d)$ denotes the cdf of the d-variate normal distribution function with mean $\boldsymbol{\mu} = (1, \mu_2, \ldots, \mu_d)$ and covariance matrix the d-variate diagonal identity matrix $\mathbf{I}_d$, and $\mathscr{F}_j(\cdot; \pi, \mu_j) = \pi\Phi(\cdot; \mu_j, 1) + (1 - \pi)\Phi(\cdot; -\mu_j, 1)$, $j = 1, \ldots, d$ is the univariate cdf of a mixture of two univariate normal cdfs. Essentially, since the variables are uncorrelated upon conditioning by the component, the d-variate normal cdfs in (11.8) can be easily calculated as the product of univariate normal cdfs.

In this construction the mixing operation introduces the dependence structure. The covariance matrix of the 2-finite normal mixture distribution is of the form,

$$\boldsymbol{\Delta} = \mathbf{I}_d + \boldsymbol{\mu}\boldsymbol{\mu}^\top \tag{11.9}$$

$$\boldsymbol{\Delta} = \begin{bmatrix} 2 & \mu_2 & \dots & \mu_{d-1} & \mu_d \\ \mu_2 & 1+\mu_2^2 & \dots & \mu_2\mu_{d-1} & \mu_2\mu_d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{m-1} & \mu_2\mu_{d-1} & \dots & 1+\mu_{d-1}^2 & \mu_{d-1}\mu_d \\ \mu_d & \mu_2\mu_d & \dots & \mu_{d-1}\mu_d & 1+\mu_d^2 \end{bmatrix}.$$

Clearly, the covariance matrix in (11.9) is identifiable and has $d-1$ dependence parameters. This dependence construction is similar to the partially symmetric copula of [17]; in the present case, however, the (j,k) marginal for $j \neq k \neq 1$ has two copula parameters, and thus more flexible association.

Mathematically, this family has nice features, a moderate number of parameters to model dependence (including negative dependence), and a rather simple computational form but does not provide such flexible or wide range of dependence. For example, it cannot model multivariate discrete data with strong or with negative dependence among many random variables or at least it cannot capture all the possible structures. To sum up, finite normal mixture copulas satisfy properties P3, P4, and P5, but not P1 and P2.

11.3.5 Mixtures of max-id

Joe [19] and Nikoloulopoulos and Karlis [44, 46] applied mixtures of max-id copulas to model multivariate discrete data. Joe and Hu [22] extended multivariate Archimedean copulas to a more flexible class of copulas using mixture of max-id copulas $C_{jk}^{(m)}$ of the form,

$$C(u_1,\dots,u_d;\theta,\theta_{jk}:1\le j<k\le d) = \tag{11.10}$$
$$\phi\left(-\sum_{1\le j<k\le d}\log C_{jk}^{(m)}\left(e^{-p_j\phi^{-1}(u_j;\theta)},e^{-p_k\phi^{-1}(u_k;\theta)};\theta_{jk}\right)+\sum_{j=1}^{d}\nu_j\,p_j\,\phi^{-1}(u_j;\theta)\,;\,\theta\right),$$

where $p_j=(\nu_j+d-1)^{-1}$, $j=1,\dots,d$. Since the mixing operation introduces dependence, this copula has a dependence structure that comes from the form of $C_{jk}^{(m)}(\cdot;\theta_{jk})$ and the form of the Laplace transform $\phi(\cdot;\theta)$. Another interesting interpretation is that the Laplace transform ϕ introduces the smallest dependence between random variables (exchangeable dependence), while the copulas $C_{jk}^{(m)}$ add some pairwise dependence. The parameters ν_j are included in order that the parametric family of multivariate copulas (11.10) is closed under margins. Regarding ν_j zero or fixed, the copula of the form (11.10) is a family with $1+d(d-1)/2$ parameters that allows only positive but flexible dependence structure.

One may simplify the form of the copula by assuming $C_{jk}^{(m)}(u_j, u_k) = u_j\, u_k$ (known as independence or product copula) together with $\nu_j = \nu_k = -1$, for some pairs. This implies that for those pairs of variables, the minimum level of dependence is introduced by ϕ.

This construction, on the one hand, does not impose any constraints between the dependence parameters θ_{jk}, but, on the other hand, does not allow for negative dependence [22]. The latter is the only drawback of this class of parametric families of copulas.

To sum up, d-variate mixtures of max-infinitely divisible copulas satisfy all properties except P1, since they don't allow for negative associations. Note in passing that using mixtures of max-id copulas the use of covariate functions for the copula dependence parameters is straightforward since they fulfill P5.

11.3.6 Elliptical

Two well-known members of elliptical copulas [1, 7], the MVN and Student t copulas, have been used in the literature for prediction and modeling of dependent discrete data.

Joe [19] and Song [55] modeled dependent discrete data using the MVN copula,

$$C(u_1, \ldots, u_d; \mathbf{R}) = \Phi_d\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d); \mathbf{R}\right), \tag{11.11}$$

where $\Phi_d(\cdot; \mathbf{R})$ denotes the standard MVN distribution function with correlation matrix $\mathbf{R} = (\rho_{jk} : 1 \le j < k \le d)$ and Φ is the cdf of the univariate standard normal. The MVN copula inherits the dependence structure of MVN distribution, and thus admits a wide range of flexible dependence allowing both positive and negative dependence (P1–P2). The drawback of the MVN copula is with relation to the computation of the rectangle probabilities. This computation involves repeated multidimensional integration since MVN lacks a closed form cdf. Consequently, likelihood inference might be difficult; see [45]. Note that for the special case of positive exchangeable correlation structures, the d-dimensional integrals conveniently reduce to 1-dimensional integrals [24, p. 48].

The pmf can be obtained by computing the following rectangle probability [40, 48],

$$\begin{aligned} h_{\mathbf{Y}}(\mathbf{y}; \mathbf{x}) &= \Pr(Y_1 = y_1, \ldots, Y_d = y_d; \mathbf{x}) \\ &= \Pr(y_1 - 1 < Y_1 \le y_1, \ldots, y_d - 1 < Y_d \le y_d; \mathbf{x}) \\ &= \int_{\Phi^{-1}[F_{Y_1}(y_1-1;\mathbf{x}_1)]}^{\Phi^{-1}[F_{Y_1}(y_1;\mathbf{x}_1)]} \cdots \int_{\Phi^{-1}[F_{Y_d}(y_d-1;\mathbf{x}_d)]}^{\Phi^{-1}[F_{Y_d}(y_d;\mathbf{x}_d)]} \phi_d(z_1, \ldots, z_d; \mathbf{R}) dz_1 \ldots dz_d, \end{aligned} \tag{11.12}$$

where ϕ_d denotes the standard MVN density with correlation matrix $\mathbf{R}$.

There are several papers in the literature that focus on the computation of the MVN rectangle probabilities for general correlation structures, and, conveniently, the implementation of the proposed algorithms is available in contributed R packages.[1] Schervish [50] proposed a locally adaptive numerical integration method but this method, while more accurate, is time consuming and restricted to a low dimension. Therefore, Genz and Bretz [11] proposed a randomized quasi Monte Carlo method with the use of antithetic variates and Joe [18] proposed two approximations to MVN probabilities. These advances in computation of MVN probabilities can be used to implement MVN copula models with discrete response data.

Genest et al. [10] modeled dependent binary data using the Student t copula,

$$C(u_1,\ldots,u_d;\mathbf{R},\nu) = T_d\left(T^{-1}(u_1;\nu),\ldots,T^{-1}(u_d;\nu);\mathbf{R},\nu\right), \tag{11.13}$$

where $T(\cdot;\nu)$ is the univariate Student t cdf with (non-integer) ν degrees of freedom, and $T_d(\cdot;\mathbf{R},\nu)$ is the cdf of a multivariate Student t distribution with ν degrees of freedom and correlation matrix $\mathbf{R}$. Student t copula share with the MVN copula the ability to accommodate any feasible pattern of association in a set of random variables. However, Student t copula can also account for tail dependence in multivariate continuous data [42], whereas MVN copula cannot. In the context of multivariate discrete data that means that more probabilities can be assigned in the joint upper and joint lower tails than with the MVN copula. Student t copula cannot also be expressed in closed form; however, the rectangle probabilities can also be computed using the methods in [11].

To sum up, elliptical copulas satisfy properties P1, P2, and P4, but not P3, and P5, since they lack a closed form cdf and a positive-definite matrix is required respectively.

11.3.7 Vine

In the literature, vine copulas are suitable for modeling multivariate continuous data with various features such as tail dependence [23]. Since the densities of multivariate vine copulas can be factorized in terms of bivariate linking copulas and lower-dimensional margins, they are computationally tractable for high-dimensional continuous variables. The cdf of d-dimensional vine copula lacks a closed form and requires $(d-1)$-dimensional integration [19]. Hence, in order to derive the pmf as finite difference of the cdf poses nonnegligible numerical challenges.

[1] Both approximations to MVN rectangle in [18], the 1-dimensional integral in the positive exchangeable case, and the method in [50], can be computed with the functions `mvnapp`, `exchmvn`, and `pmnorm`, respectively, in the R package `mprobit` [21]. The methods in [11] can be computed with the function `pmvnorm` in the R package `mvtnorm` [12].

Recently, Panagiotelis et al. [48] decomposed the pmf as follows,

$$\Pr(Y_1 = y_1, \ldots, Y_d = y_d) = \Pr(Y_1 = y_1 | Y_2 = y_2, \ldots, Y_d = y_d) \times \Pr(Y_2 = y_2 | Y_3 = y_3, \ldots, Y_d = y_d) \times \cdots \times \Pr(Y_d = y_d). \tag{11.14}$$

Letting V_h be any scalar element of $\mathbf{V}$ and $\mathbf{V}_{\setminus h}$ its complement, with Y_j not an element of $\mathbf{V}$, each term on the right-hand side of (11.14) has the form $Pr(Y_j = y_j | V = v)$ where y_j is a scalar element of y and $\mathbf{v}$ is a subset of $\mathbf{y}$,

$$\begin{aligned} Pr(Y_j = y_j | \mathbf{V} = \mathbf{v}) &= \frac{\Pr(Y_j = y_j, \mathbf{V}_h = \mathbf{v}_h | V_{\setminus h} = \mathbf{v}_{\setminus h})}{\Pr(\mathbf{V}_h = \mathbf{v}_h, \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h})} \\ &= \frac{\sum_{i_j=0,1} \sum_{i_h=0,1} (-1)^{i_j + i_h} \Pr(Y_j \leq y_j - i_j, \mathbf{V}_h \leq \mathbf{v}_h - i_h | \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h})}{\Pr(\mathbf{V}_h = \mathbf{v}_h, V_{\setminus h} = \mathbf{v}_{\setminus h})} \\ &= \frac{\sum_{i_j=0,1} \sum_{i_h=0,1} (-1)^{i_j + i_h} C_{Y_j, \mathbf{v}_h | \mathbf{v}_{\setminus h}} \left(F_{Y_j | \mathbf{v}_{\setminus h}}(y_j - i_j | \mathbf{v}_{\setminus h}), F_{\mathbf{v}_h | V_{\setminus h}}(\mathbf{v}_h - i_h | v_{\setminus h}) \right)}{\Pr(\mathbf{V}_h = \mathbf{v}_h, \mathbf{V}_{\setminus h} = \mathbf{v}_{\setminus h})}. \end{aligned}$$

The above can be applied recursively to (11.14) to decompose a multivariate pmf into bivariate copula families. More details and a three-dimensional illustration can be found in [48].

The computation of the pmf for a discrete vine only requires $2d(d-1)$ bivariate copula function evaluations, compared to 2^d multivariate copula evaluations for the finite difference approach (11.2), and Panagiotelis et al. [48] have developed a fast algorithm for computing the pmf of a vine copula with discrete margins.

A wide variety of dependence structures can be modeled by selecting different copula families as building blocks. Selecting different bivariate copula families in a discrete vine has a substantial impact on the joint probabilities of the multivariate distribution and can provide better fits when we have some discrete multivariate data where asymmetries can easily be seen.

To sum up, discrete vine copulas or pair-copula constructions satisfy all properties except P4. Note that although their cdf is not of closed form the pmf is successively decomposed and likelihood estimation is feasible even for high dimensions.

11.4 Methods of Estimation

For a sample of size n with data $\mathbf{y}_1, \ldots, \mathbf{y}_n$ the joint log-likelihood of the copula-based model is,

$$\ell = \sum_{i=1}^{n} \log h_{\mathbf{Y}}(y_{i1}, \ldots, y_{id}; \mathbf{x}_{i1}, \ldots, \mathbf{x}_{id}). \tag{11.15}$$

Estimation of the model parameters can be approached by the standard maximum likelihood method, by maximizing the joint log-likelihood in (11.15) over the univariate and copula parameters [19] or by a two-step approach called Inference Function of Margins (IFM) method in [19, 20]. In the first step, the univariate parameters are estimated assuming independence, and in the second step the joint log-likelihood in (11.15) is maximized over the copula parameters with the univariate parameters fixed at the estimated values from the first step. When the dependence is not too strong which is a realistic scenario for discrete response data, the IFM method can efficiently (in sense of computing time and asymptotic variance) estimate the model parameters. For parametric families of copulas with a closed form cdf and vine copulas, maximum likelihood or IFM estimation is straightforward.

For the elliptical copulas likelihood inference involves the computation of multidimensional rectangle probabilities of the form (11.12). The advances in computation of rectangle probabilities can be used to implement elliptical copula-based models with discrete response data. Using the the first-order (makes use of all of the univariate and bivariate marginal probabilities) or the second-order approximation (also makes use of trivariate and four-variate marginal probabilities) in [18] to compute the rectangle MVN probabilities in (11.15), the likelihood is successively approximated for weak to moderate correlation parameters. Computing the rectangle MVN/Student t probabilities in (11.15) via simulation based on the methods in [11], a simulated likelihood is implemented; see [40]. Since the estimation of the parameters of the copula-based models is obtained using a quasi-Newton routine [36] applied to the joint log-likelihood in (11.15), the use of quasi Monte Carlo simulation to four decimal place accuracy for evaluations of the rectangles works poorly, because numerical derivatives of the joint log-likelihood with respect to the parameters are not smooth. In order to achieve smoothness, the same set of uniform random variables should be used for every rectangle probability that comes up in the optimization of the simulated likelihood [40]. Asymptotic and small-sample efficiency calculations in [40] show that the simulated likelihood method, which is based on evaluating the multidimensional integrals of the joint likelihood with randomized quasi Monte Carlo methods developed in [11], is as good as maximum likelihood as shown for dimension 10 or lower. These findings are expected to hold in higher dimensions. Although there is an issue of computational burden as the dimension and the sample size increase, this will become marginal, as computing technology is advancing rapidly.

Zhao and Joe [61] proposed composite likelihood estimation methods to overcome the computational issues at the maximization routines for the MVN copula in a high-dimensional context. Composite likelihood is a surrogate likelihood which leads to unbiased estimating equations obtained by the derivatives of the composite log-likelihoods. Estimation of the model parameters can be approached by solving the estimating equations in [61] or equivalently by maximizing the sum of composite likelihoods. First consider the sum of univariate log-likelihoods,

$$\ell_1 = \sum_{i=1}^{n} \sum_{j=1}^{d} \log f_{Y_j}(y_{ij}; \mathbf{x}_{ij}),$$

where $f_{Y_j}(y_1; \mathbf{x}_1), \ldots, f_{Y_d}(y_d; \mathbf{x}_d)$ are the univariate marginal pmfs, and then the sum of bivariate log-likelihoods,

$$\ell_2 = \sum_{i=1}^{n} \sum_{j<k} \log h_{\mathbf{Y}_2}(y_{ij}, y_{ik}; \mathbf{x}_{ij}, \mathbf{x}_{ik}),$$

where $\mathbf{Y}_2 = (Y_j, Y_k)$. Composite likelihood estimates can be obtained using a two-stage method (CL1):

1. At the first step the ℓ_1 is maximized over the univariate marginal parameters.
2. At the second step the ℓ_2 is maximized over the copula parameters with univariate marginal parameters fixed as estimated at the first step of the method.

Alternatively, one can use the one stage composite likelihood estimation procedure (CL2), that is maximizing the ℓ_2 over the univariate and copula parameters at one step. The efficiency of composite likelihood estimates has been studied and shown in a series of a papers; see, e.g., [58, 59, 61]. If the interest is both to the univariate and dependence parameters, CL2 method should be performed since CL1 ignores the dependence at the estimation of the univariate marginal parameters.

Bayesian methods have also been used on the estimation of an elliptical-copula-based model. Pit et al. [49] proposed a general Bayesian approach for estimating a MVN copula-based model. Smith, Gan and Kohn [53] extend the work in [49] to other elliptical copula-based models. Very recently, Smith and Khaled [54] suggest efficient Bayesian data augmentation methodology for the estimation of copula-based models for multivariate discrete data. For a detailed exposition of Bayesian approaches on estimation of copula-based models for discrete response data we refer the interested reader to the excellent survey by Smith [52].

11.5 Dependence as Measured by Kendall's Tau

The copula parameters for different parametric families have different range; hence, they are not comparable. To compare strengths of dependence among different copula-based models and ease interpretation, it is useful to convert the estimated parameters to concordance measures such as Kendall's τ's.

For continuous random variables dependence as measured by Kendall's tau $\tau = P_c - P_d$, the difference between the probabilities of concordance (P_c) and discordance (P_d), is associated only with the copula parameters. However for discrete data the marginal distributions also play a role on dependence, and

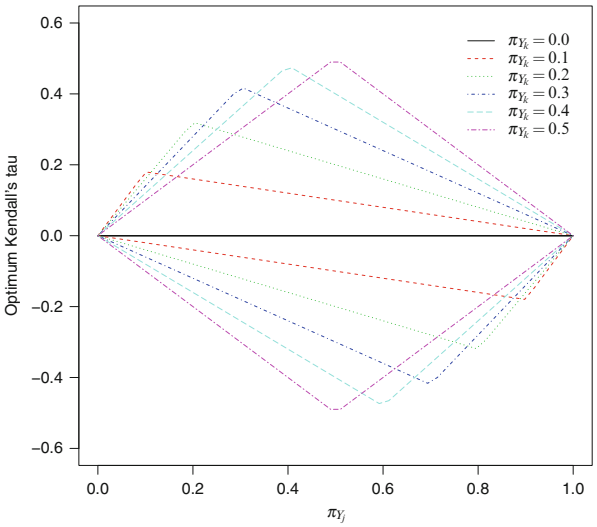

Fig. 11.1 Kendall's tau bounds when $Y_j \sim Bin(1, \pi_{Y_j})$ and $Y_k \sim Bin(1, \pi_{Y_k})$ with $(\pi_{Y_j}, \pi_{Y_k}) \in [0,1] \times \{0, 0.1, \ldots, 0.5\}$

τ does not attain the boundary values of ± 1, because the probability of ties $P_t = 1 - (P_c + P_d)$ is positive; see [5, 33, 38].

The Kendall's tau for each pair $\mathbf{Y}_2$ is given as below [47],

$$\tau(Y_j, Y_k) = \sum_{y_j=0}^{\infty}\sum_{y_k=0}^{\infty} h_{\mathbf{Y}_2}(y_j, y_k; \mathbf{x}_j, \mathbf{x}_k)\Big\{4C(F_{Y_j}(y_j - 1; \mathbf{x}_j), F_{Y_k}(y_k - 1; \mathbf{x}_k)) - h_{\mathbf{Y}_2}(y_j, y_k; \mathbf{x}_j, \mathbf{x}_k)\Big\} + \sum_{y_j=0}^{\infty} f^2_{Y_j}(y_j; \mathbf{x}_j) + \sum_{y_k=0}^{\infty} f^2_{Y_k}(y_k; \mathbf{x}_k) - 1. \tag{11.16}$$

This formula helps to see clearly that in the discrete case the marginals do affect Kendall's tau.

To visualize the effect of the marginal distributions/parameters, we computed the optimum Kendall's tau values using various discrete marginal distributions, i.e., Bernoulli, binomial and Poisson and the Fréchet bound copulas. In Figs. 11.1 and 11.2 optimum Kendall's tau values have been plotted for Bernoulli, i.e., $Y_j \sim Bin(1, \pi_{Y_j})$ and $Y_k \sim Bin(1, \pi_{Y_k})$ and binomial margins, i.e., $Y_j \sim Bin(5, \pi_{Y_j})$ and $Y_k \sim Bin(5, \pi_{Y_k})$ for a grid of (π_{Y_j}, π_{Y_k}) values in $\mathscr{P}_{Y_j} \times \mathscr{P}_{Y_k}$ where $\mathscr{P}_{Y_j} = [0,1]$ and $\mathscr{P}_{Y_k} = \{0, 0.1, \ldots, 0.5\}$, respectively. In Fig. 11.3 optimum Kendall's tau values have been plotted for Poisson marginal distributions with the same parameter λ up to 50.

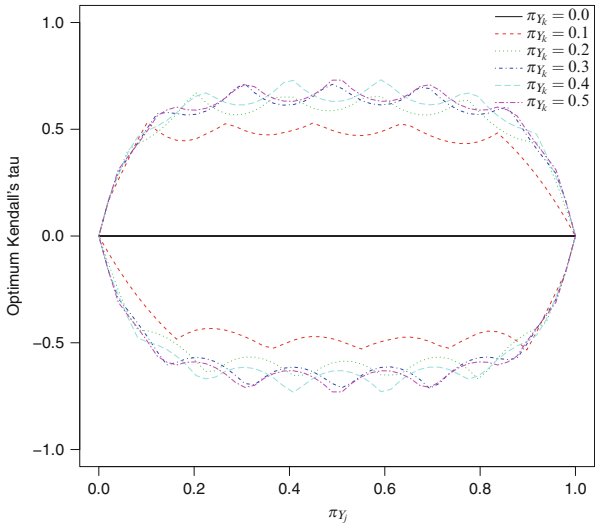

Fig. 11.2 Kendall's tau bounds when $Y_j \sim Bin(5, \pi_{Y_j})$ and $Y_k \sim Bin(5, \pi_{Y_k})$ with $(\pi_{Y_j}, \pi_{Y_k}) \in [0, 1] \times \{0, 0.1, \ldots, 0.5\}$

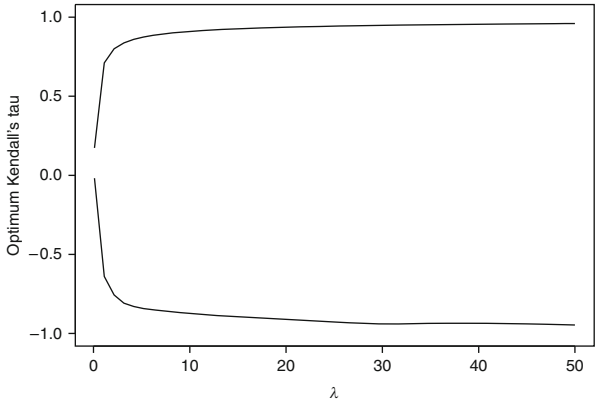

Fig. 11.3 Kendall's tau bounds when both random variables are Poisson with parameter λ

From the figures, one can see that Kendall's tau does not reach the bounds ± 1 for countermonotonic and comonotonic marginals. There is also a clear association between the optimum value of Kendall's tau and the marginal probabilities for binary and binomial data, while this association is negligible for count data with marginal parameters greater than 10. For normalized versions of Kendall's tau one can refer to [13, 38].

11.6 Discussion

This survey summarized copula-based models for discrete response data. We list several desirable properties such a model should have and introduce the models that have been used in copula dependence modeling for discrete data so far. For copula modeling with multivariate discrete data, we suggest models that admit a wide range of dependence, such as the MVN copula. Given the wide range of dependence, MVN copula provides often the best fit or nearly the best fit for discrete data [41]. However MVN copula is inadequate to model multivariate data with refection asymmetry or tail dependence [43]. Although tail dependence degenerate in the discrete case, reflection asymmetry is a realistic scenario. Vine copula constructions are suitable for modeling this kind of data since by using as bivariate blocks asymmetric bivariate copulas tail asymmetry can be accommodated, i.e., more probability in one or both joint tails can be obtained. Essentially, discrete vine copulas are highly flexible since any multivariate discrete distribution can be decomposed as a vine copula, under a set of conditions outlined in [48].

If the discrete responses are positively associated, then parametric families of copulas with a closed form cdf could be also used. Archimedean copulas could be used to model clustered data with exchangeable dependence, while mixtures of max-infinitely divisible copulas could be used for data with a more general positive dependence. Note in passing that, from copulas with positive dependence by construction, one could always get some negative dependence by applying decreasing transformations on some subset of the random variables, but this is restrictive in general, because this construction cannot model negative dependence among many random variables [46].

If the interest is to study the effect of explanatory variables on the dependence structure, Archimedean, partially symmetric, mixtures of max-id, and vine copulas are suitable since allow the use of covariate functions for the copula dependence parameters (see, e.g., [44, 47]); this is not the case for the FGM and elliptical copulas in (11.5)–(11.7), (11.11) and (11.13), because of the joint constraints for the dependence parameters.

Acknowledgments I would like to thank Professor Harry Joe, University of British Columbia, for helpful comments and suggestions on an earlier version of this survey.

References

1. Abdous, B., Genest, C., Rémillard, B.: Dependence properties of meta-elliptical distributions. In: Duchesne, P., Rémillard, B. (eds.) Statistical Modelling and Analysis for Complex Data Problems, pp. 1–15. Kluwer, Dordrecht (2005)
2. Ashford, J.R., Sowden, R.R.: Multivariate probit analysis. Biometrics **26**, 535–546 (1970)
3. Cameron, A.C., Li, T., Trivedi, P.K., Zimmer, D.M.: Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. Econom. J. **7**(2), 566–584 (2004)

4. Choros, B., Härdle, W., Okhrin, O.: CDO pricing with copulae. Discussion paper 2009-013, SFB 649, Humboldt Universität zu Berlin, Berlin (2009)
5. Denuit, M., Lambert, P.: Constraints on concordance measures in bivariate discrete data. J. Multivar. Anal. **93**(1), 40–57 (2005)
6. Famoye, F.: A new bivariate generalized Poisson distribution. Stat. Neerl. **64**, 112–124 (2010)
7. Fang, H.B., Fang, K.T., Kotz, S.: The meta-elliptical distributions with given marginals. J. Multivar. Anal. **82**, 1–16 (2002)
8. Gauvreau, K., Pagano, M.: The analysis of correlated binary outcomes using multivariate logistic regression. Biometrical J. **39**, 309–325 (1997)
9. Genest, C., Nešlehová, J.: A primer on copulas for count data. Astin Bull. **37**, 475–515 (2007)
10. Genest, C., Nikoloulopoulos, A.K., Rivest, L.P., Fortin, M.: Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas. Braz. J. Probab. Stat. (2013, in press)
11. Genz, A., Bretz, F.: Methods for the computation of the multivariate t-probabilities. J. Comput. Graph. Stat. **11**, 950–971 (2002)
12. Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T.: mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-9992. URL http://CRAN.R-project.org/package=mvtnorm (2012)
13. Goodman, L., Kruskal, W.: Measures of association for cross classifications. J. Am. Stat. Assoc. **49**, 732–764 (1954)
14. Hausman, J., Wise, D.: A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica **45**, 319–339 (1978)
15. Hofert, M.: Construction and sampling of nested archimedean copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications, Proceedings of the Workshop held in Warsaw, 25–26 September, pp. 147–160. Springer, Berlin (2009)
16. Hofert, M., Mächler, M.: Nested archimedean copulas meet R: The `nacopula` package. J. Stat. Softw. **39**, 1–20 (2011)
17. Joe, H.: Parametric families of multivariate distributions with given margins. J. Multivar. Anal. **46**, 262–282 (1993)
18. Joe, H.: Approximations to multivariate normal rectangle probabilities based on conditional expectations. J. Am. Stat. Assoc. **90**(431), 957–964 (1995)
19. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
20. Joe, H.: Asymptotic efficiency of the two-stage estimation method for copula-based models. J. Multivar. Anal. **94**(2), 401–419 (2005)
21. Joe, H., Chou, L.W., Zhang, H.: mprobit: Multivariate probit model for binary/ordinal response. R package version 0.9-3. URL http://CRAN.R-project.org/package=mprobit (2011)
22. Joe, H., Hu, T.: Multivariate distributions from mixtures of max-infinitely divisible distributions. J. Multivar. Anal. **57**(2), 240–265 (1996)
23. Joe, H., Li, H., Nikoloulopoulos, A.K.: Tail dependence functions and vine copulas. J. Multivar. Anal. **101**, 252–270 (2010)
24. Johnson, N.L., Kotz, S.: Continuous Multivariate Distributions. Wiley, New York (1972)
25. Johnson, N.L., Kotz, S.: On some generalized Farlie-Gumbel-Morgenstern distributions. Commun. Stat. **4**, 415–427 (1975)
26. Johnson, N.L., Kotz, S.: On some generalized Farlie-Gumbel-Morgenstern distributions – II Regression, correlation and further generalizations. Commun. Stat. **6**, 485–496 (1977)
27. Lakshminarayana, J., Pandit, S.N.N., Rao, K.S.: On a bivariate Poisson distribution. Commun. Stat.-Theory Method. **28**, 267–276 (1999)
28. Lee, A.: Modelling rugby league data via bivariate negative binomial regression. Aust. N. Z. J. Stat. **41**(2), 141–152 (1999)
29. Li, J., Wong, W.K.: Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness. Biostatistics **12**, 143–155 (2011)
30. McHale, I., Scarf, P.: Modelling soccer matches using bivariate discrete distributions with general dependence structure. Stat. Neerl. **61**(4), 432–445 (2007)

31. McNeil, A.J., Nešlehová, J.: Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. Ann. Stat. **37**, 3059–3097 (2009)
32. Meester, S., MacKay, J.: A parametric model for cluster correlated categorical data. Biometrics **50**, 954–963 (1994)
33. Mesfioui, M., Tajar, A.: On the properties of some nonparametric concordance measures in the discrete case. J. Nonparametr. Stat. **17**(5), 541–554 (2005)
34. Miravete, E.J.: Multivariate Sarmanov count data models. CEPR Discussion Paper No. DP7463, University of Texas at Austin; Centre for Economic Policy Research (2009)
35. Muthén, B.: Contributions to factor analysis of dichotomous variables. Psychometrika **43**, 551–560 (1978)
36. Nash, J.: Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd edn. Hilger, New York (1990)
37. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (2006)
38. Nešlehová, J.: On rank correlation measures for non-continuous random variables. J. Multivar. Anal. **98**, 544–567 (2007)
39. Nikoloulopoulos, A.K.: Letter to the editor. Biostatistics **13**, 1–3 (2012)
40. Nikoloulopoulos, A.K.: On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood. ArXiv e-prints, arXiv:1304.0905 (2013)
41. Nikoloulopoulos, A.K., Joe, H., Chaganty, N.R.: Weighted scores method for regression models with dependent data. Biostatistics **12**, 653–665 (2011)
42. Nikoloulopoulos, A.K., Joe, H., Li, H.: Extreme value properties of multivariate t copulas. Extremes **12**, 129–148 (2009)
43. Nikoloulopoulos, A.K., Joe, H., Li, H.: Vine copulas with asymmetric tail dependence and applications to financial return data. Comput. Stat. Data Anal. **56**, 659–3673 (2012)
44. Nikoloulopoulos, A.K., Karlis, D.: Multivariate logit copula model with an application to dental data. Stat. Med. **27**, 6393–6406 (2008)
45. Nikoloulopoulos, A.K., Karlis, D.: Finite normal mixture copulas for multivariate discrete data modeling. J. Stat. Plan. Inference **139**, 3878–3890 (2009)
46. Nikoloulopoulos, A.K., Karlis, D.: Modeling multivariate count data using copulas. Commun. Stat. Simul. Comput. **39**, 172–187 (2010)
47. Nikoloulopoulos, A.K., Karlis, D.: Regression in a copula model for bivariate count data. J. Appl. Stat. **37**, 1555–1568 (2010)
48. Panagiotelis, A., Czado, C., Joe, H.: Pair copula constructions for multivariate discrete data. J. Am. Stat. Assoc. **107**, 1063–1072 (2012)
49. Pitt, M., Chan, D., Kohn, R.: Efficient Bayesian inference for Gaussian copula regression models. Biometrika **93**, 537–554 (2006)
50. Schervish, M.: Algorithm AS 195. Multivariate normal probabilities with error bound. Appl. Stat. **33**, 81–94 (1984)
51. Sklar, M.: Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris **8**, 229–231 (1959)
52. Smith, M.S.: Bayesian approaches to copula modelling. In: Damien, P., Dellaportas, P., Polson, N., Stephens, D. (eds.) Bayesian Theory and Applications. Oxford University Press, USA (2013)
53. Smith, M.S., Gan, Q., Kohn, R.: Modeling dependence using skew t copulas: Bayesian inference and applications. J. Appl. Econom. **27**, 500–522 (2012)
54. Smith, M.S., Khaled, M.A.: Estimation of copula models with discrete margins via Bayesian data augmentation. J. Am. Stat. Assoc. **107**, 290–303 (2012)
55. Song, P.X.K.: Correlated Data Analysis: Modeling, Analytics, and Application. Springer, New Yory (2007)
56. Song, P.X.K., Li, M., Yuan, Y.: Joint regression analysis of correlated data using Gaussian copulas. Biometrics **65**, 60–68 (2009)
57. Trégouët, D.A., Ducimetière, P., Bocquet, V., Visvikis, S., Soubrier, F., Tiret, L.: A parametric copula model for analysis of familial binary data. Am. J. Hum. Genet. **64**(3), 886–93 (1999)
58. Varin, C.: On composite marginal likelihoods. Adv. Stat. Anal. **92**, 1–28 (2008)

59. Varin, C., Reid, N., Firth, D.: An overview of composite likelihood methods. Stat. Sin. **21**, 5–42 (2011)
60. Xu, J.: Statistical modelling and inference for multivariate and longitudinal discrete response. Ph.D. thesis, The University of British Columbia (1996)
61. Zhao, Y., Joe, H.: Composite likelihood estimation in multivariate data analysis. Can. J. Stat. **33**(3), 335–356 (2005)
62. Zimmer, D., Trivedi, P.: Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. J. Bus. Econ. Stat. **24**(1), 63–76 (2006)

Chapter 12
Vector Generalized Linear Models: A Gaussian Copula Approach

Peter X.-K. Song, Mingyao Li, and Peng Zhang

Abstract In this chapter we introduce a class of multi-dimensional regression models for vector outcomes, termed as the vector generalized linear models (VGLMs), which is a multivariate analogue of the univariate generalized linear models (GLMs). A unified framework of such regression models is established with the utility of Gaussian copula, accommodating discrete, continuous and mixed vector outcomes. Both full likelihood and composite likelihood estimations and inferences are discussed. A Gauss–Newton type algorithm is suggested to carry out the simultaneous estimation for all model parameters. Numerical illustrations are focused on VGLMs for correlated binary outcomes, correlated count outcomes, and mixed normal and binary outcomes. In the simulation studies, we compare the VGLM to the popular generalized estimating equations (GEEs) approach. The simulation results indicate that the VGLMs provide more efficient inference for the regression coefficients than the GEEs. The VGLM is also illustrated via real-data examples.

P.X.-K. Song (✉)
Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
e-mail: pxsong@umich.edu

M. Li
Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA
e-mail: mingyao@mail.med.upenn.edu

P. Zhang
Department of Mathematics, Zhejiang University, Hongzhou, China
e-mail: pengz@zju.edu.cn

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*, Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_12,

12.1 Introduction

Generalized linear models (GLMs) [20] have been playing an important role in the regression analysis of non-normal data. A GLM assumes that a univariate response y follows an exponential dispersion (ED) family distribution [14], denoted by $ED(\mu, \varphi)$, with mean μ and dispersion parameter φ, and the density function is given by

$$g(y; \mu, \varphi) = c\,(y; \varphi) \exp\left[\{\theta y - \kappa(\theta)\}/\varphi\right], y \in \mathscr{R}, \theta \in \Theta, \tag{12.1}$$

where $\kappa(\cdot)$ is the cumulant generating function, Θ is an open interval and φ varies in a subset of $(0, \infty)$. It is known that the mean and variance are, respectively, $\mu = \mathrm{E}(y) = \tau(\theta)$ and $\mathrm{var}(y) = \varphi v(\mu)$, where $v(\cdot)$ is the unit variance function, and $\tau(\cdot) = \dot{\kappa}(\cdot)$ and $v(\cdot) = \dot{\tau}\{\tau^{-1}(\cdot)\}$ are the respective first order derivatives of $\kappa(\cdot)$ and $\tau(\cdot)$.

A GLM postulates that the mean μ is related to p covariates $\mathbf{x} = (x_1, \ldots, x_p)^T$ by an equation

$$h(\mu) = \eta(\mathbf{x}) = \mathbf{x}^T \beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \tag{12.2}$$

where h is a known link function, and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a vector of regression coefficients. Statistical inference for β is one of the main tasks in the theory of the GLMs.

A key ingredient required for the extension of the above univariate GLM to a general multivariate framework for vector outcomes is a multivariate analogue of the ED family distributions in (12.1). Suppose that for each subject an m-element response vector $\mathbf{y} = (y_1, \ldots, y_m)^T$ and a p-element associated covariate vector $\mathbf{x}$ are observed. For example, the vector $\mathbf{y}$ is comprised of measurements from different response variables, such as blood pressure, heart rate, weight and temperature for a subject. Other examples of such data include clustered data of an equal cluster size, longitudinal data with a fixed number of repeated measurements, and spatial data collected from a fixed number of spatial locations. It is worth noting that such restriction of equal dimension for outcome vectors can be relaxed by the means of composite likelihood [28], in which only low dimensional likelihood objectives will be used to carry out statistical estimation and inference. For the ease of exposition, let us first focus on the scenario of fixed dimension at m.

To analyze such data by the GLM approach, vector GLMs (VGLMs) can be formulated as a model for which the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ takes the form

$$f(\mathbf{y}|\mathbf{x}; \beta, \varphi, \Gamma) = \delta(\mathbf{y}, \eta_1, \ldots, \eta_m; \varphi, \Gamma), \tag{12.3}$$

where in general the regression coefficients $\beta = (\beta_1^T, \ldots, \beta_K^T)^T$ and the linear predictors $\eta_j = \eta_j(\mathbf{x}) = \mathbf{x}^T \beta_j$, $j = 1, \ldots, m$, under a certain joint density function $\delta(\cdot; \varphi, \Gamma)$ that is parametrized by a vector of dispersion parameters

$\varphi = (\varphi_1, \ldots, \varphi_m)^T$ and a dependence matrix Γ. Here $\Gamma = (\gamma_{ij})$ characterizes dependencies among the components of $\mathbf{y}$. Note that K may not equal to m, the dimension of $\mathbf{y}$.

To complete the specification of a VGLM in (12.3), it is necessary to specify the $\delta(\cdot)$ function and the parameter set Γ. In our views, a desired density function $\delta(\cdot; \Gamma)$ should satisfy the following basic properties:

a. The regression parameters β_j's and the correlation parameters in Γ are ideally made 'orthogonal' by the chosen $\delta(\cdot)$, in a spirit similar to that in the multivariate normal. This property of orthogonality, if available, would give rise to much ease in the development of efficient statistical inference for the large number of model parameters in the VGLMs.
b. The VGLM resulted from a chosen $\delta(\cdot)$ should be reproducible or marginally closed, namely the low dimensional regression models retain the same error distribution type as the joint model. This is because in most practical problems the data types for the individual components of $\mathbf{y}$ are relatively easy to recognize, so that the corresponding marginal error distributions can be readily assumed, as is the practice for GLMs. In addition, in the setting of composite likelihood approach [28], low dimensional likelihood objectives should be ideally defined with the same distributional type as that of the full dimensional likelihood.
c. The correlation parameters in matrix Γ enable to characterize both positive and negative associations for the vector $\mathbf{y}$. Although positive association is often seen in practice, a model that allows negative association certainly provides flexibility to depict a broader range of data types.
d. The VGLM (12.3) based on differently chosen $\delta(\cdot)$ can easily accommodate discrete, continuous and mixed outcomes of various types within one unified theoretical framework.

Examples of the VGLM (12.3) with the common regression parameter include the log-linear representation [5] or the Bahadur representation [3] for correlated binary responses (see [10, 29]) and generalized linear mixed models (GLMMs) (see [6, 8, 19]). Examples of models with different β_j's include the bivariate logit model (see [18], Sect. 6.5.6) and the bivariate probit model [2] for correlated binary responses, among others.

In this chapter, we consider a new class of $\delta(\cdot)$ functions based on the multivariate distributions generated by parametric copulas (see [13], Chap. 5). In particular, the class of multivariate exponential dispersion (MED) distributions generated by the Gaussian copula [25] will be applied and examined in detail, due to the fact that such a class of MED distributions satisfies the four desirable properties above. This then results in a variety of VGLMs under a unified modeling framework, which is useful to conduct regression analysis of continuous vector outcomes, discrete vector outcomes and mixed vector outcomes. Refer to Sect. 12.2 for details concerning the different forms of the $\delta(\cdot)$ density for different types of outcomes. The class of VGLMs possesses several advantageous features compared to some of the existing models.

First, unlike the Bahadur's [3] representation of the joint distribution for binary vector outcomes, the VGLMs do not suffer the drawback that the correlations are constrained by the marginal probabilities. This constraint usually causes a substantial shrinkage on the range of correlation [7]. In addition, unlike the log-linear representation [10] for binary vector outcomes, the VGLMs are reproducible. This reproducibility leads to a better interpretation for the VGLMs than the log-linear representation model, as well as the valid basis of developing composite likelihood approach.

Second, the VGLMs share a great deal of similarity in spirit with Liang and Zeger's [16] marginal models. However, the difference between the two approaches is crucial: the VGLMs that are built upon joint probability distributions allow to perform a likelihood inference for all model parameters, but the marginal models based on GEEs are relied on a quasi-likelihood inference developed only for the regression coefficients β.

Third, the formulation of the VGLMs allows to develop a unified likelihood inference theory and model selection procedures for a broad range of data types such as those considered in the classical GLM theory. Compared to the existing *ad hoc* vector models, each suitable for one specific data type, this unification offers a useful multivariate analogue of the theory of the univariate GLMs, and particularly gives rise to great ease in developing a flexible statistical software package for the class of vector regression models, which is of practical importance.

Fourth, this unified framework sheds light on the analysis of outcomes of mixed types. Within this framework, one general statistical inference theory can be developed for a number of models useful to analyze mixed binomial, Poisson, normal and gamma outcomes. The VGLMs for mixed outcomes sustain the marginal distributions as desired, simply because the copula models are known to be marginally closed. This differs from Fitzmaurice and Laird's [9] conditional model approach and Sammel et al.'s [23] latent variable model approach in which the marginal distribution is specified by a mixture distribution.

Last, as pointed by Song [25], the dependence matrix Γ in the Gaussian copula is inherited from and hence behaves similarly to that of the multivariate normal. For example, the components of response $\mathbf{y}$ are independent if and only if Γ is the identity matrix. Therefore, as in the normal multivariate analysis, it is convenient to impose some structure on matrix Γ, such as exchangeable, AR(1) and m-dependence, to yield a parsimonious specification of the dependence structure and hence gain power in inference. In addition, the inverse matrix Γ^{-1} provides the interpretation of conditional independence, which is essential to describe relationships in graphic models.

Although the copula approach is getting increasingly popular in the past two decades or so, using such models to analyze discrete correlated data has not yet fully discussed in the literature. Thus, this chapter will focus on the VGLMs for correlated binary, correlated count and correlated mixed outcomes. Despite the cautious remarks given by researchers (e.g., [11]), our simulation studies and data analyses have shown good performances of vector logistic/probit, vector log-linear

models and vector GLMs for mixed outcomes. In all cases, comparisons to the GEEs are undertaken.

The organization of the chapter is as follows. We begin with a brief review of Gaussian copula multivariate ED distributions in Sect. 12.2, followed by the discussion of full likelihood estimation and inference as well as composite likelihood estimation and inference in Sect. 12.3, where some details are supplied for VGLMs for bivariate discrete data. Sections 12.4 and 12.5 concern vector logistic/probit models and vector log-linear models, where comparisons between estimators from the VGLMs and estimators from the GEEs are shown. Moreover, Sect. 12.6 presents the VGLMs for mixed outcomes. Several simulation and data analysis examples are scattered in Sects. 12.4, 12.5, and 12.6 to illustrate the VGLMs. Section 12.7 gives concluding remarks. Some technical details are listed in the appendices.

12.2 Multivariate ED Family Distributions

We now give a brief review of the multivariate ED (MED) distributions proposed by Song [25], which are useful in the specification of the random component of the VGLMs as in the theory of univariate GLMs, where the univariate ED family distributions [15] are assumed.

12.2.1 Definition

For component j, $j = 1, \ldots, m$, denote the marginal CDF of $\text{ED}(\mu_j, \varphi_j)$ by $G_j(y_j; \mu_j, \varphi_j)$ or simply $G_j(y_j)$. Following Sklar [24], we may construct a joint CDF with m ED margins by the Gaussian copula in the form

$$F(\mathbf{y}; \mu, \varphi, \Gamma) = C\left\{G_1(y_1; \mu_1, \varphi_1), \ldots, G_m(y_m; \mu_m, \varphi_m) | \Gamma\right\}, \tag{12.4}$$

where $\mu = (\mu_1, \ldots, \mu_m)^T$ is the vector of m means, $\varphi = (\varphi_1, \ldots, \varphi_m)^T$ is the vector of m dispersion parameters, and $C(\cdot)$ is the m-variate Gaussian copula with the CDF given by

$$C(\mathbf{u}|\Gamma) = \Phi_m\left\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m)|\Gamma\right\}, \ \mathbf{u} = (u_1, \ldots, u_m)^T \in (0,1)^m. \tag{12.5}$$

Here Φ_m (or ϕ_m) and Φ (or ϕ) are the respective CDFs (or densities) of m-variate normal $N_m(0, \Gamma)$ with a correlation matrix Γ and the standard univariate normal $N(0,1)$ marginal. Note that all marginal parameters are contained in the F, and the parameters for correlation are inherited from the correlation matrix Γ of the multivariate normal. It is known that the Gaussian copula in (12.5) is a joint CDF of m uniform random variables on $(0,1)$ with dependence matrix $\Gamma = (\gamma_{jj'})_{m\times m}$

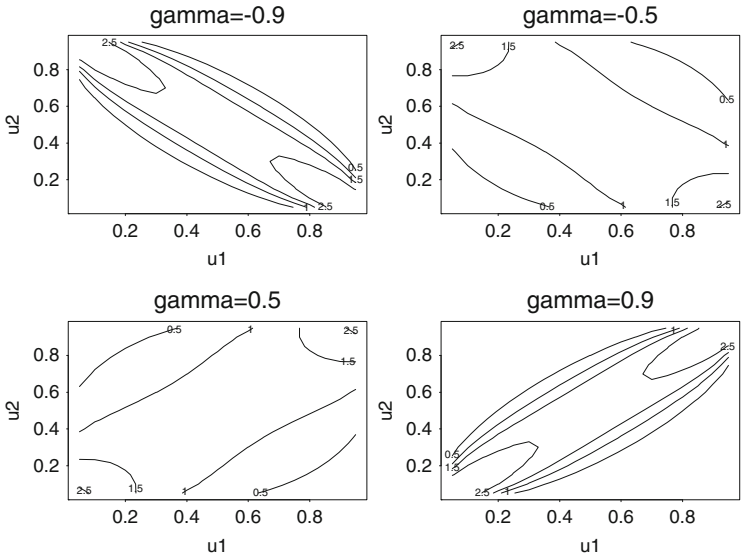

Fig. 12.1 Contours of the bivariate Gaussian copula distribution

with the diagonals $\gamma_{jj} = 1$ and off-diagonals $|\gamma_{jj'}| < 1$. Like the multivariate Gaussian distribution, an MED distribution is fully parametrized by the three sets of parameters, μ, φ and Γ.

Figure 12.1 displays contours of the bivariate Gaussian copula with different values of dependence parameter γ. Clearly, this copula accommodates both positive and negative dependence, indicated by the opposite directions of concentration in the contours. The degree of concentration representing the variation of the distribution increases as parameter γ tends to ± 1.

Clearly, the multivariate normal distribution is a special case of the MED when all margins are univariate Gaussian. In this case μ is the vector of mean parameters, φ is the vector of variance parameters, and Γ is the Pearson correlation matrix. With non-Gaussian margins, the (i, j)-the element of Γ becomes a pairwise nonlinear dependence defined by

$$\gamma_{ij} = \text{corr}\left[\Phi^{-1}\{G_i(y_i)\}, \Phi^{-1}\{G_j(y_j)\}\right]. \tag{12.6}$$

When both marginal CDFs $G_t(\cdot), t = i, j$ are continuous, γ_{ij} represents the linear correlation of two normal scores $\Phi^{-1}(G_t(y_t)), t = i, j$. When y_i and y_j are discrete, the equation (12.6) still holds, but the interpretation would be different with different data types. For example, when $y_t, t = i, j$ are both binary, the resulting bivariate binary model will have the same joint probability mass function as that induced from the threshold latent variable model via dichotomization. This implies that the dependence parameter γ_{ij} can be interpreted as the *polychoric correlation*

given by Anderson and Pemberton [1]. See Song [25] for more details in other distribution cases such as Poisson distribution.

12.2.2 Density Functions

When all m margins are continuous, the joint density of an MED in (12.5) is given by

$$f(\mathbf{y};\mu,\varphi,\Gamma) = c\{G_1(y_1),\ldots,G_m(y_m)|\Gamma\}\prod_{i=1}^{m} g_i(y_i;\mu_i,\varphi_i), \tag{12.7}$$

where $c(\cdot)$ is the density of the copula $C(\cdot)$ in (4) given by

$$c(\mathbf{u}|\Gamma) = |\Gamma|^{-1/2}\exp\left\{\frac{1}{2}\mathbf{q}^T(I_m-\Gamma^{-1})\mathbf{q}\right\},$$

with $\mathbf{q} = (q_1,\ldots,q_m)^T$ being a vector of normal scores $q_i = \Phi^{-1}(u_i), i = 1,\ldots,m$, and I_m being the m-dimensional identity matrix. Obviously, $\Gamma = I_m$ implies the independence of the m components, similar to the multivariate normal.

Consequently, when the function δ required in the VGLM (12.3) is chosen to be the density f specified by (12.7), the VGLM yields a large class of vector regression models for various continuous vector outcomes, including the vector normal linear model, the vector gamma GLM model, the vector inverse Gaussian GLM model, and the vector compound Poisson GLM model.

When all m margins are discrete, the joint probability function of a discrete MED distribution takes the form

$$f(\mathbf{y}) = \mathrm{P}(Y_1 = y_1,\ldots,Y_m = y_m) = \sum_{j_1=1}^{2}\cdots\sum_{j_m=1}^{2}(-1)^{j_1+\cdots+j_m}C(u_{1,j_1},\ldots,u_{m,j_m}|\Gamma) \tag{12.8}$$

where $u_{j,1} = G_j(y_j-)$ and $u_{j,2} = G_j(y_j)$. Here $G_j(y_j-)$ is the left-hand limit of G_j at y_j. For the VLGMs presented in this chapter, the δ function has been specified by the Gaussian copula, so the point mass probabilities defined in (12.8) are unique and identical to those generated by discretizing latent m-variate normal vectors [25].

Likewise, a large class of vector regression models for vector discrete outcomes is specified under a unified framework, by taking this probability mass function f in (12.8) as the δ for the VGLM in (12.3). In this paper, two special vector GLMs from this class, the vector logistic/probit model for correlated dichotomous data and the vector Poisson/negative binomial log-linear models for correlated count data, will be studied in Sects. 12.4 and 12.5, respectively.

When the m margins appear to be mixed outcomes, say, the first m_1 margins being continuous and the rest $m_2 = m - m_1$ margins being discrete, according to

Song et al. [27], the joint density function is given as follows. Let $\mathbf{u} = (\mathbf{u}_1^T, \mathbf{u}_2^T)^T$, with $\mathbf{u}_1 = (u_1, \ldots, u_{m_1})^T$ and $\mathbf{u}_2 = (u_{m_1+1}, \ldots, u_m)^T$. The same partition and notation are applied for vectors $\mathbf{x}$ and $\mathbf{q}$. Let

$$
\begin{aligned}
&C_1^{m_1}(\mathbf{u}_1, \mathbf{u}_2|\Gamma) = \frac{\partial^{m_1}}{\partial u_1 \cdots \partial u_{m_1}} C(u_1, \ldots, u_m|\Gamma) \\
&= (2\pi)^{-\frac{m_2}{2}} |\Gamma|^{-\frac{1}{2}} \\
&\times \int_{-\infty}^{\Phi^{-1}(u_{m_1+1})} \cdots \int_{-\infty}^{\Phi^{-1}(u_m)} \exp\left\{ \frac{1}{2}(\mathbf{q}_1^T, \mathbf{x}_2^T)\Gamma^{-1}(\mathbf{q}_1^T, \mathbf{x}_2^T)^T - \frac{1}{2}\mathbf{q}_1^T\mathbf{q}_1 \right\} d\,\mathbf{x}_2.
\end{aligned}
$$

Then, the joint density is given by

$$
\begin{aligned}
f(\mathbf{y}) = \prod_{j=1}^{m_1} g_j(y_j) \sum_{j_{m_1+1}=1}^{2} \cdots \sum_{j_m=1}^{2} (-1)^{j_{m_1+1}+\cdots+j_m} \\
\times C_1^{m_1}(G_1(y_1), \ldots, G_{m_1}(y_{m_1}), u_{m_1+1,j_{m_1+1}}, \ldots, u_{m,j_m}|\Gamma), \qquad (12.9)
\end{aligned}
$$

where u_{t,j_t}'s are defined in (12.8).

Section 12.6 will present one example of the VGLM with mixed binomial and normal outcomes. Other types of mixed outcomes, such as mixed gamma and Poisson outcomes as well as mixed binomial and Poisson outcomes, can be analyzed in a similar way. Note that the two latter data types cannot be modeled by the threshold latent model approach via discretization.

12.2.3 Conditional Density

An important feature of fitting a multivariate regression model is to predict an expected value of one outcome conditional on the other outcomes and covariates. This task essentially requires a conditional distribution derived from the joint MED distribution (12.4). For example, in the case of all continuous margins, the conditional density of y_m given $y_1, \ldots, y_{m-1}$ is

$$
f(y_m|y_1, \ldots, y_{m-1}) = \frac{1}{\omega} g_m(y_m; \mu_m, \varphi_m) \exp\left[-\frac{1}{2} \left\{ \frac{(q_m - \gamma^T \Gamma_{11}^{-1}\mathbf{q}_{-m})^2}{\omega^2} - q_m^2 \right\} \right],
$$

where $\mathbf{q}_{-m} = (q_1, \ldots, q_{m-1})^T$, $\omega = 1 - \gamma^T \Gamma_{11}^{-1}\gamma$, and both Γ_{11} and γ are elements given in the following partition:

$$
\Gamma = \begin{pmatrix} \Gamma_{11} & \gamma \\ \gamma^T & 1 \end{pmatrix}.
$$

12.3 Simultaneous Maximum Likelihood Estimation and Inference

12.3.1 General Theory: Full Likelihood

Suppose data $(\mathbf{y}_1, X_1), \ldots, (\mathbf{y}_n, X_n)$ follow an m-variate MED distribution,

$$\mathbf{y}_i | X_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im}) \sim \text{MED}_m(\mu_i, \varphi_i, \Gamma), \ i = 1, \ldots, n$$

where response vector $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})^T$ has mean $\mu_i = (\mu_{i1}(\mathbf{x}_{i1}) \ldots, \mu_{im}(\mathbf{x}_{im}))^T$ and dispersion $\varphi_i = (\varphi_{i1}, \ldots, \varphi_{im})^T$, in which the j-th component $\varphi_{ij} = \varphi_j / w_{ij}$ with a known positive weight w_{ij} and dispersion φ_j, $j = 1, \ldots, m, i = 1, \ldots, n$. Here $\mathbf{x}_{ij}$ is a p-element vector of covariates associated with subject i for component j, and $X_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im})$ is a $p \times m$ matrix of covariates. Moreover, the mean μ_{ij} follows a marginal GLM, $h_j(\mu_{ij}) = \eta_j(\mathbf{x}_{ij})$ with $\eta_{ij} = \eta_j(\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \beta_j$ and link function h_j, $j = 1, \ldots, m$. The primary objective of this section is to establish simultaneous maximum likelihood inference for all model parameters $\theta = (\beta, \varphi, \Gamma)$.

In many cases, the general model above may become more specific. For example, a VGLM (12.3) takes a common regression parameter vector β, which appears typically in longitudinal or clustered data analysis with a common link function. In addition, the dependence matrix Γ may be further parametrized by a parameter vector α, denoted by $\Gamma(\alpha)$, such as exchangeable, AR(1) or 1-dependence. In this case, we have $\theta = (\beta, \varphi, \alpha)$. Moreover, for the convenience, we set all weights $w_{ij} = 1$ in the rest of the chapter.

Let the log-likelihood function of parameters in a given model be

$$\ell(\theta; Y, X) = \sum_{i=1}^{n} \ell_i(\theta; \mathbf{y}_i, X_i). \tag{12.10}$$

Then, the MLE of θ is

$$\hat{\theta} = \text{argmax}_\theta \ell(\theta; Y, X).$$

To find the MLE $\hat{\theta}$ numerically, we implement a Gauss–Newton-type algorithm that allows us to search for the global maximizer with no need of second order derivatives of the log-likelihood function [22]. The second order derivatives of the log-likelihood usually appear to be very complicated, so that the explicit expressions of their analytic forms are difficult to derive and to be implemented in computation. This issue appears more challenging in the case of discrete data.

Under some mild regularity conditions, the MLE $\hat{\theta}$ is consistent and asymptotically normal. When the second order derivatives of the log-likelihood are

not available, we estimate the observed Fisher Information using the following sandwich form:

$$\hat{\mathbf{i}} = \mathbf{A}_n^{-1}(\hat{\theta})\mathbf{B}_n(\hat{\theta})\mathbf{A}_n^{-1}(\hat{\theta}), \tag{12.11}$$

where $\mathbf{A}_n(\theta)$ is the numerical Hessian approximating the observed Fisher information and an observed variability matrix $\mathbf{B}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_i(\theta; \mathbf{y}_i, X_i)\dot{\ell}_i(\theta; \mathbf{y}_i, X_i)^T$. This estimation is robust since it would consistently estimate the standard errors even if the model is misspecified.

The optimization procedure we adopt to obtain the MLE is a Gauss–Newton-type algorithm. The key step of this algorithm is to take step-halving that guarantees a steady increase in the likelihood from the previous iteration. Precisely, the $(k+1)^{th}$ iteration proceeds as

$$\theta^{k+1} = \theta^k + \epsilon\{\mathbf{B}_n(\theta^k)\}^{-1}\dot{\ell}(\theta^k),$$

where $\mathbf{B}_n$ is the observed variability matrix given above and ϵ is the step-halving term that is chosen as follows: starting at 1, it halves each time until $\ell(\theta^{k+1}) > \ell(\theta^k)$ holds in one iteration. Finally, the algorithm stops when the increase in the likelihood is no longer possible or the difference between two consecutive updates is smaller than a pre-specified precision level.

12.3.2 General Theory: Composite Likelihood

Computational difficulty in parameter estimation with the full likelihood approach may be alleviated by the means of composite likelihood (CL). The CL method has drawn much attention in recent years due to its computational convenience as it uses only low dimensional likelihood objectives (e.g., two-dimensional likelihoods). Readers may refer to Varin et al. [28] for a comprehensive overview of the CL methodology and more references therein. Now let us focus on the pairwise CL that takes low dimensional margins or submodels based on *all possible pairs* to form a pseudo-likelihood. This strategy of dimension reduction results in an inference function derived from the following pairwise CL:

$$L_c(\theta; \mathbf{y}) = \prod_{a<b} f(y_a, y_b; \theta)^{w_{ab}},$$

where $f(y_a, y_b; \theta)$ is a bivariate margin of (y_a, y_b) from the joint MED distribution (12.4) and w_{ab} is a weight defined on $[0, \infty)$. Theoretically the full likelihood is a special case of the CL when the composite set is chosen as the full set of m elements. Importantly the CL approach represents a trade-off between estimation efficiency and computational efficiency.

The fundamental argument for the validity of CL method lies on the theory of estimating functions [11, Chap. 3]. In short, as long as the bivariate marginal

distributions $f(y_a, y_b; \theta)$ (or the bivariate copula) are properly specified, the pairwise CL estimator of θ, obtained by maximizing the CL function L_c with respect to the θ can achieve desired performances. Let the log-CL function be given by

$$\ell_c(\theta; Y) = \sum_{i=1}^{n} \log L_{i,c}(\theta; Y) = \sum_{i=1}^{n} \sum_{a<b} w_{ab} \log f(y_{i,a}, y_{i,b}; \theta).$$

Thus, the CL estimator is obtained by

$$\hat{\theta}_c = \text{argmax}_\theta \ell_c(\theta; Y).$$

Clearly, the strategy of using only bivariate marginals in the above estimation and inference can naturally accommodate unbalanced longitudinal data. Searching the maximizer of ℓ_c may be similarly done using the Gauss–Newton type algorithm described above. Being a pseudo-likelihood approach, under mild regularity conditions, the root-n consistent CL estimator is asymptotically normal with the asymptotic covariance matrix given by $\left\{\text{E}(-\ddot{\ell}_c(\theta))\right\}^{-1} \text{var}(\dot{\ell}_c(\theta)) \left\{\text{E}(-\ddot{\ell}_c(\theta))\right\}^{-T}$.

12.3.3 VGLMs For Bivariate Discrete Data

In this section, the VGLMs for bivariate discrete data are discussed. The focus of two-dimensional data allows us to study performances of both full likelihood and pairwise CL under one common setting. Readers may find general results regarding the log-likelihood functions and their scores in the VGLMs for continuous, discrete, and mixed data types from Sect. 6.6 of Song [26].

For convenience, we always use notation ℓ to denote the likelihood for the case of $m = 2$, although it is indeed also ℓ_c. The log-likelihood is $\ell(\theta; Y, X) = \sum_{i=1}^{n} \ln f(\theta; \mathbf{y}_i)$, where the bivariate probability mass function f is obtained immediately from (12.8) as

$$f(\mathbf{y}_i; \theta) = C_\alpha(u_{i1}, u_{i2}) - C_\alpha(u_{i1}, v_{i2}) - C_\alpha(v_{i1}, u_{i2}) + C_\alpha(v_{i1}, v_{i2}),$$

with $u_{ij} = G_i(y_{ij})$ and $v_{ij} = G_i(y_{ij}-)$, $j = 1, 2$. Clearly the parameter vector is $\theta = (\beta, \alpha)$.

The scores $\dot{\ell}_{\theta_k}(\theta)$ for the k-th component of θ is $\sum_{i=1}^{n} \dot{f}_k(\mathbf{y}_i; \theta)/f(\mathbf{y}_i; \theta)$, where $\dot{f}_k(\cdot)$ denotes the first order derivative with respect to θ_k. By the chain rule, the scores with respect to β_j are given by

$$\begin{aligned}\frac{\partial f(\mathbf{y}_i; \theta)}{\partial \beta_j} &= \frac{\partial C_\alpha(u_{i1}, u_{i2})}{\partial \beta_j} - \frac{\partial C_\alpha(u_{i1}, v_{i2})}{\partial \beta_j} - \frac{\partial C_\alpha(v_{i1}, u_{i2})}{\partial \beta_j} + \frac{\partial C_\alpha(v_{i1}, v_{i2})}{\partial \beta_j} \\ &= \left[\frac{\partial C_\alpha(u_{i1}, u_{i2})}{\partial u_{i1}} - \frac{\partial C_\alpha(u_{i1}, v_{i2})}{\partial u_{i1}}\right] \frac{x_{i1j}}{\dot{h}_1(\mu_{i1})} \frac{\partial u_{i1}}{\partial \mu_{i1}}\end{aligned}$$

$$+\left[\frac{\partial C_\alpha(u_{i1},u_{i2})}{\partial u_{i2}}-\frac{\partial C_\alpha(v_{i1},u_{i2})}{\partial u_{i2}}\right]\frac{x_{i2j}}{\dot{h}_2(\mu_{i2})}\frac{\partial u_{i2}}{\partial \mu_{i2}}$$
$$-\left[\frac{\partial C_\alpha(u_{i1},v_{i2})}{\partial v_{i2}}-\frac{\partial C_\alpha(v_{i1},v_{i2})}{\partial v_{i2}}\right]\frac{x_{i2j}}{\dot{h}_2(\mu_{i2})}\frac{\partial v_{i2}}{\partial \mu_{i2}}$$
$$-\left[\frac{\partial C_\alpha(v_{i1},u_{i2})}{\partial v_{i1}}-\frac{\partial C_\alpha(v_{i1},v_{i2})}{\partial v_{i1}}\right]\frac{x_{i1j}}{\dot{h}_1(\mu_{i1})}\frac{\partial v_{i1}}{\partial \mu_{i1}}, \tag{12.12}$$

where, suppressing subscripts,

$$\frac{\partial C_\alpha(u,v)}{\partial u}=\Phi\left(\frac{\Phi^{-1}(v)-\alpha\Phi^{-1}(u)}{\sqrt{1-\alpha^2}}\right),$$
$$\frac{\partial C_\alpha(u,v)}{\partial v}=\Phi\left(\frac{\Phi^{-1}(u)-\alpha\Phi^{-1}(v)}{\sqrt{1-\alpha^2}}\right).$$

Note that derivatives $\partial u_{ij}/\partial\mu_{ij}$ and $\partial v_{ij}/\partial\mu_{ij}$ in (12.12) can have closed form expressions when certain marginal distributions are assumed. For example, the Bernoulli margin for binary data gives

$$\frac{\partial u_{ij}}{\partial\mu_{ij}}=-1[y_{ij}=0],\ \frac{\partial v_{ij}}{\partial\mu_{ij}}=-1[y_{ij}=1],$$

where $1[A]$ denotes the indicator function on set A, while the Poisson margin for count data gives

$$\frac{\partial u_{ij}}{\partial\mu_{ij}}=G_{ij}(y_{ij}-1)-G_{ij}(y_{ij}),\ \frac{\partial v_{ij}}{\partial\mu_{ij}}=G_{ij}(y_{ij}-2)-G_{ij}(y_{ij}-1),$$

where $G_{ij}(\cdot)$ is the Poisson CDF with mean μ_{ij}.

Similarly, the score with respect to the parameter α is

$$\frac{\partial f(\mathbf{y}_i;\theta)}{\partial\alpha}=\frac{\partial C_\alpha(u_{i1},u_{i2})}{\partial\alpha}-\frac{\partial C_\alpha(u_{i1},v_{i2})}{\partial\alpha}-\frac{\partial C_\alpha(v_{i1},u_{i2})}{\partial\alpha}+\frac{\partial C_\alpha(v_{i1},v_{i2})}{\partial\alpha},$$

where, suppressing the subscripts,

$$\frac{\partial C_\alpha(u,v)}{\partial\alpha}=\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{\partial}{\partial\alpha}\left\{\ln\phi_2(x_1,x_2;\alpha)\right\}\phi_2(x_1,x_2;\alpha)dx_1dx_2, \tag{12.13}$$

with

$$\frac{\partial\ln\phi_2(x_1,x_2;\alpha)}{\partial\alpha}=\left\{\alpha+(1-\alpha^2)^{-1}(\alpha x_1-x_2)(\alpha x_2-x_1)\right\}(1-\alpha^2)^{-1}.$$

To evaluate the double integral in (12.13), we adopt the Gaussian–Hermite quadrature method with details given in Appendix A.

12.4 Vector GLMs for Binary Data

In this section, we present a detailed investigation on the VGLMs for bivariate binary data, the simplest discrete data type from which some useful insights can be drawn to guide the use of such models to analyze higher dimensional vector data. For the ease of exposition, we present our results under the assumption of a common $\beta = \beta_j, j = 1, \ldots, m$. The related procedures can be extended to the general case with different β_j with little effort.

12.4.1 Models

It is easy to derive the bivariate probability mass function,

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}) = \begin{cases} C_\alpha(1-\mu_{i1}, 1-\mu_{i2}), & \text{if } y_{i1} = 0, y_{i2} = 0 \\ 1-\mu_{i1} - C_\alpha(1-\mu_{i1}, 1-\mu_{i2}), & \text{if } y_{i1} = 0, y_{i2} = 1 \\ 1-\mu_{i2} - C_\alpha(1-\mu_{i1}, 1-\mu_{i2}), & \text{if } y_{i1} = 1, y_{i2} = 0 \\ \mu_{i1} + \mu_{i2} + C_\alpha(1-\mu_{i1}, 1-\mu_{i2}) - 1, & \text{if } y_{i1} = 1, y_{i2} = 1, \end{cases} \tag{12.14}$$

where $C_\alpha(\cdot, \cdot)$ is given by (12.5) with $m = 2$.

Therefore, a bivariate logistic model is obtained by specifying the marginal probabilities $\mu_{ij} = P(Y_{ij} = 1)$ in (12.14) as $\text{logit}(\mu_{ij}) = \eta(\mathbf{x}_{ij})$, $j = 1, 2$. This gives $(\dot{h}(\mu_{ij}))^{-1} = \mu_{ij}(1-\mu_{ij})$.

Similarly, a bivariate probit model specifies the marginal probabilities in (12.14) as the form of $\Phi^{-1}(\mu_{ij}) = \eta(\mathbf{x}_{ij})$ for $j = 1, 2$, which implies that $\{\dot{h}(\mu_{ij})\}^{-1} = \phi\{\Phi(\mu_j)\}$. According to Song [25], this vector probit model is effectively identical to the threshold latent model induced from dichotomization. Therefore, in the case of the probit link the dependence parameter α has the same interpretation as the correlation parameter in the latent bivariate normal distribution. However, such a correspondence is not true under the logit link.

In the context of multivariate categorical data, odds ratios are usually used to measure pairwise association, which are equal to, suppressing index i,

$$\begin{aligned} \text{OR} &= \frac{P(Y_1 = 1, Y_2 = 1) P(Y_1 = 0, Y_2 = 0)}{P(Y_1 = 1, Y_2 = 0) P(Y_1 = 0, Y_2 = 1)} \\ &= \frac{\{\mu_1 + \mu_2 + C_\alpha(1-\mu_1, 1-\mu_2) - 1\}\{C_\alpha(1-\mu_1, 1-\mu_2)\}}{\{1-\mu_1 - C_\alpha(1-\mu_1, 1-\mu_2)\}\{1-\mu_2 - C_\alpha(1-\mu_1, 1-\mu_2)\}}, \\ &\quad \mu_1, \mu_2 \in (0, 1). \end{aligned} \tag{12.15}$$

Clearly, the odds ratio equals 1 if and only if $\alpha = 0$, the case of independence.

Figure 12.2 illustrates the relationship of the log-10 base odds ratio to the dependence parameter α with equal mean parameters $\mu_1 = \mu_2$. It can be seen that

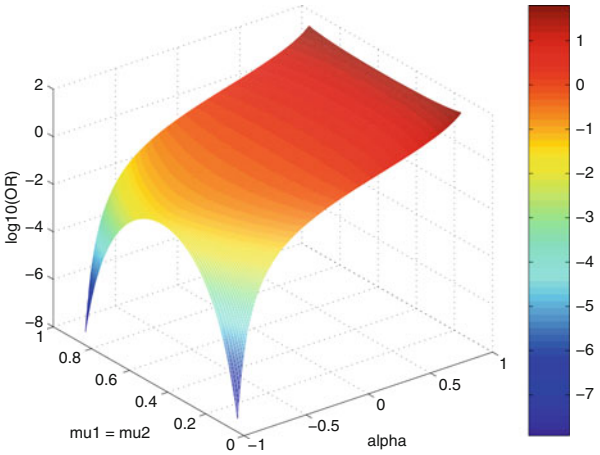

Fig. 12.2 log10 (OR) with equal mean parameters

the log-10 odds ratio is predominantly related to the dependence parameter α, but only slightly related to the mean parameters near boundaries. Also, both positive and negative associations are modeled in the VGLMs.

12.4.2 Comparison of Asymptotic Efficiencies

We now address the issue of the asymptotic efficiency of estimators from the VGLMs and GEE estimators of the marginal parameters in models for binary responses. In particular, we compare the asymptotic efficiency of maximum likelihood estimators, under the fully parametric VGLMs, to the GEE estimators that are obtained under a specified correlation structure. Our comparison will focus only on the regression parameters β, since the correlation parameter α is treated as a nuisance parameter in the GEE.

Let V_{vglm} be the asymptotic covariance of the ML estimator of β from the VGLM. We will compare it to the asymptotic covariance of the GEE estimator of β, $V_{gee} = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta}^T \text{cov}(\mathbf{y}_i)^{-1} \frac{\partial \mu_i}{\partial \beta}$, obtained under the working correlation being specified as the true correlation. The asymptotic relative efficiency (ARE) is defined as follows:

$$\text{ARE}(\beta) = \text{diag}\{V_{vglm}\}[\text{diag}\{V_{gee}\}]^{-1}. \tag{12.16}$$

Consider a hypothetical cross-over trial in which two repeated measurements are observed from one subject with $x_{i1} = 0$ for placebo and $x_{i2} = 1$ for active drug. The marginal probabilities of the bivariate binary response are specified as

$$h(\mu_{ij}) = \beta_0 + \beta_1 x_{ij}, \ j = 1, 2, \tag{12.17}$$

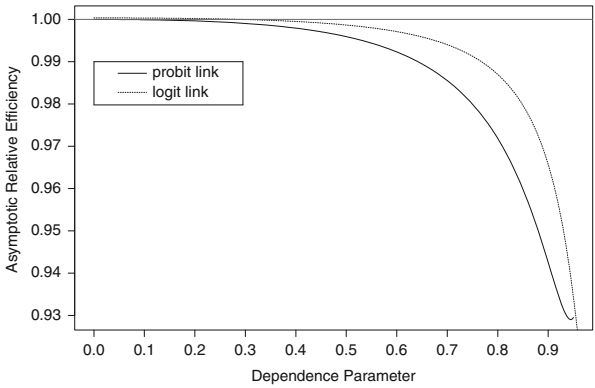

Fig. 12.3 Asymptotic efficiencies of the VGLM estimator of the slope parameter β_1 relative to the estimator from the GEE under the respective bivariate logistic and probit models

where $\beta_0 = \beta_1 = 0.5$. For such a simple model, we can effectively obtain the closed form expressions for both asymptotic covariance matrices V_{gee} and V_{vglm}, and related details are given in Appendix B. Therefore, in this case the ARE can be evaluated exactly with no need of simulated data.

Figure 12.3 displays the ARE values for the estimator of the slope parameter β_1 as a function of the within cluster association parameter $\alpha \in [0, 1)$ with both logit and probit link functions. Evidently, Fig. 12.3 suggests that the estimator from the VGLM is more efficient than the GEE estimator with either logit link or probit link, especially when the within cluster correlation is high (α close to 1). Note that, under the probit link, the VGLM is identical to the threshold latent model. This means that the estimator of the slope parameter from the threshold latent model appears to be more efficient than the GEE estimator. Also note that, under the logit link, the resulting GEEs are indeed coincident with the score equations derived from the log-linear model representation [8, Sect. 8.2]. Figure 12.3 implies that the VGLM is more appealing than the log-linear model representation to fit the correlated binary data.

12.5 Vector GLMs for Count Data

In this section, we focus on the VGLMs for correlated count data. Especially, we use the bivariate Poisson VGLM for simulation studies and for data analysis.

12.5.1 Models

The m-variate Poisson VGLM has the joint probability mass function given by (12.8), where the left-hand limit of Poisson CDF $G_{ij}(\cdot)$ at y_{ij} is $G_{ij}(y_{ij}-) = G_{ij}(y_{ij} - 1)$ with marginal mean μ_{ij} follows a log-linear model $\log(\mu_{ij}) = \mathbf{x}_{ij}^T \beta_j$, $j = 1, 2, \ldots, m$. As usual, we set $G_{ij}(y_{ij}) = 0$ if y_{ij} is negative or if its mean $\mu_{ij} = 0$. Clearly, $\{\dot{h}(\mu_{ij})\}^{-1} = \mu_{ij}$ if we take the log link function.

In the presence of overdispersion, negative binomial margins appear to be more appealing than Poisson margins in the specification of the VGLM for overdispersed count data. For a negative binomial distribution with mean μ and variance $\mu(1+\sigma^2\mu)$, the probability mass function is given by $P(Y=y)=\frac{\Gamma(a+y)}{y!\Gamma(a)}\frac{a^a\mu^y}{(\mu+a)^{a+y}}$, where $a=1/\sigma^2$. Clearly, if σ^2 is zero, it reduces to the Poisson variance, and a nonzero σ^2 leads to a model of overdispersion. To specify the VGLM, similar to the Poisson VGLM, we assume the marginal means μ_{ij} take the form of log-linear models, $\log(\mu_{ij})=\mathbf{x}_{ij}^T\beta_j$, $j=1,\ldots,m$. Let $u_{ij}=G_{ij}(y_{ij})$ and $v_{ij}=G_{ij}(y_{ij}-1)$. Then the derivatives needed in the calculation of the scores are immediately yielded as follows:

$$\frac{\partial u_{ij}}{\partial \mu_{ij}}=\sum_{k=0}^{y_{ij}}\frac{\Gamma(a_j+k)}{k!\Gamma(a_j)}\frac{a_j^{a_j}\mu_{ij}^k}{(\mu_{ij}+a_j)^{a_j+k}}\left[\frac{k}{\mu_{ij}}-\frac{a_j+k}{a_j+\mu_{ij}}\right]$$

$$\frac{\partial v_{ij}}{\partial \mu_{ij}}=\sum_{k=0}^{y_{ij}-1}\frac{\Gamma(a_j+k)}{k!\Gamma(a_j)}\frac{a_j^{a_j}\mu_{ij}^k}{(\mu_{ij}+a_j)^{a_j+k}}\left[\frac{k}{\mu_{ij}}-\frac{a_j+k}{a_j+\mu_{ij}}\right].$$

It is worth noting that according to Song [25], the dependence parameter α in the bivariate Poisson VGLM is approximately equal to the Pearson correlation in the bivariate Poisson distribution generated by the stochastic representation [13, Sect. 7.2]. The stochastic representation method constructs the bivariate Poisson distribution as follows,

$$(Y_1,Y_2)\stackrel{d}{=}(Z_1+Z_{12},Z_2+Z_{12}),\ \text{with } Z_j\stackrel{iid}{\sim}Po(\lambda_j),\ j=1,2,12, \qquad (12.18)$$

where both components share a common Poisson variable Z_{12}. Obviously the correlation given in this bivariate Poisson distribution is always non-negative with 0 corresponding to independence.

12.5.2 Comparison of Asymptotic Efficiencies

This section addresses the issue of the relative asymptotic efficiency between the estimators from the VGLMs and the estimators from the GEEs, when both methods are used to fit a common data. The comparison of asymptotic efficiencies will be focused only on the regression coefficients.

Here we consider a bivariate Poisson VGLM with a common $\beta=\beta_j$, $j=1,2$. Unlike the case of the bivariate binary VGLM, the closed form expression for V_{vglm} is difficult to derive analytically. So, a numerical evaluation on V_{vglm} is inevitable. In fact, we compare the estimated observed Fisher Information given in (12.11) to $V_{gee}=\sum_{i=1}^n\frac{\partial\mu_i}{\partial\beta}^T\text{cov}(\mathbf{y}_i)^{-1}\frac{\partial\mu_i}{\partial\beta}$ with the working correlation being specified as the true correlation in the simulation study. In order to achieve high precision for formula (12.11), we chose a large sample size $n=1000$.

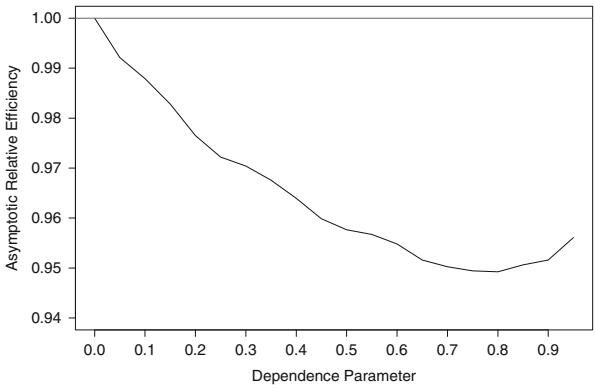

Fig. 12.4 Average asymptotic efficiency of the VGLM estimator relative to the estimator from the GEE under the bivariate Poisson model over 2,000 replications, in which each cluster contains 1,000 pairs

Data are generated from the following marginal log-linear models:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{ij}, \ j = 1, 2, i = 1, \ldots, n \tag{12.19}$$

where $\beta_0 = \beta_1 = 0.5$, covariate x_{i1} is generated randomly according to uniform $U(0, 1)$, and for simplicity, covariate $x_{i2} = x_{i1}$ is assigned with the underlying assumption that two subjects in each cluster have the same exposure level. $n = 1000$ clusters were simulated from the stochastic representation model (12.18), each containing a pair of counts (y_{i1}, y_{i2}), with means determined by (12.19). This can be done through an explicit one-to-one correspondence between two sets of parameters, $(\mu_{i1}, \mu_{i2}, \alpha)$ and $(\lambda_{i1}, \lambda_{i2}, \lambda_{i12})$. Note that the assignment of the same exposure level, i.e. $x_{i2} = x_{i1}$, in each cluster is just one way to ensure that the generated outcomes satisfy the positivity constraint for the marginal expectations of the Poisson model.

This simulation was carried out only for $\alpha \in [0, 1)$. Obviously, under the independence correlation, the GEEs appear to be the same as the score equations given by the VGLMs. The average ARE of the form (12.16) for the slope parameter β_1 is plotted in Fig. 12.4 over 2,000 replications at each of 20 grid points with 0.05 apart in $[0, 1)$.

Figure 12.4 indicates that high dependence leads to low ARE, which implies that the estimator from the proposed VGLM is more efficient than the estimator from the GEEs. Note that the above calculations were based on the data generated from the stochastic representation (12.18), a model that is different from the proposed VGLM, and the GEEs method is not dependent on a fully specified probability model but only on the correctly specified first two moments. We learned from this simulation that when both VGLM and GEE are used to analyze data from a third model, the VGLM outperforms the GEE method in terms of estimation efficiency. It is interesting to notice from Fig. 12.4 that the ARE stops dropping when two outcomes become highly dependent. This may be because when the correlation is high, the GEE becomes more sensitive to appreciate the correlation of the data.

Table 12.1 Estimated regression coefficients (standard errors) and Z-statistics from the VGLM and GEE

	VGLM		GEE	
Variable	$\hat{\beta}$ $(s.e.)$	Z	$\hat{\beta}(s.e.)$	Z
Intercept	1.91 (0.07)	26.99	1.90(0.11)	17.36
Treatment (x_1)	0.29 (0.07)	4.16	0.28(0.09)	3.26
Period (x_2)	−0.12 (0.07)	−1.77	−0.12(0.08)	−1.44

12.5.3 Data Examples

We illustrate the bivariate Poisson model by fitting data arising from a two-period cross-over trial for the treatment of enuresis, reported by Hills and Armitage [12]. Twenty nine children were treated with a new drug (A) or placebo (B) for 14 days. The number of dry nights for each child was recorded [21, Table 16.2] and was assumed to follow a Poisson distribution. Piantadosi [21] studied the period effect and treatment effect by assuming a normal distribution. Here we used Poisson distribution to model the counts of dry nights, and its marginal means are assumed to follow the log-linear model:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}, j = 1, 2, i = 1, \ldots, 29,$$

where $x_{1ij} = 1$ for new drug and 0 otherwise, and $x_{2ij} = 1$ for period 2 and 0 otherwise. Two methods, the Poisson VGLM and the GEEs were applied to analyze the data. Table 12.1 lists the results of the two methods. Both VGLM and GEE methods found statistical significance of the new drug to increase the mean number of dry nights for child, where the VGLM gives a smaller standard error and larger Z-statistic than the GEE, indicating a higher statistical power in the hypothesis test for the effect of treatment. It is in agreement with the findings from the simulation study; that is, the VGLM is more efficient than the GEE. In addition, both methods did not find the importance of covariate period.

12.6 Vector GLMs for Mixed Outcomes

As pointed out in Sect. 12.2, the copula approach allows us to jointly model response variables of mixed types. In this section we illustrate the use of such a model to handle a bivariate response vector of mixed normal and binary outcomes.

12.6.1 Models

Consider a bivariate vector $\mathbf{y} = (y_1, y_2)$ in that $y_1 \sim N(\mu_1, \varphi_1)$ and $y_2 \sim$ Bernoulli(μ_2). The VGLM for the data is specified by (12.5) with the marginal

means given by $\mu_1 = \mathbf{x}_1^T\beta_1$ and $h(\mu_2) = \mathbf{x}_2^T\beta_2$. Here β_1 and β_2 are usually different, but covariates $\mathbf{x}_j, j = 1, 2$ may be the same in some cases. It follows from (12.9) that the joint density of $\mathbf{y}$ is

$$f(y_1, y_2) = \begin{cases} \phi(y_1; \mu_1, \varphi_1)\{1 - C_1^*(\mu_2, z_1)\}, & \text{if } y_2 = 0, \\ \phi(y_1; \mu_1, \varphi_1)C_1^*(\mu_2, z_1), & \text{if } y_2 = 1, \end{cases} \tag{12.20}$$

where $\phi(\cdot; \mu_1, \varphi_1)$ is the density of $N(\mu_1, \varphi_1)$, $z_1 = (y_1 - \mu_1)/\sqrt{\varphi_1}$, and $C_1^*(a, b) = \Phi\left(\frac{\Phi^{-1}(a) + \alpha b}{\sqrt{1-\alpha^2}}\right)$.

We use the following conditional mass function of $y_2|y_1$ to generate correlated mixed outcomes in the simulation study,

$$f(y_2|y_1) = \begin{cases} 1 - C_1^*(\mu_2, z_1), & \text{if } y_2 = 0, \\ C_1^*(\mu_2, z_1), & \text{if } y_2 = 1. \end{cases}$$

With a given data $\{\mathbf{y}_i, (\mathbf{x}_{i1}, \mathbf{x}_{i2})\}, i = 1, \ldots, n$, the log-likelihood for parameter $\theta = (\beta_1, \beta_2, \varphi_1, \alpha)$ is expressed as follows:

$$\begin{aligned} \ell(\theta) &= \sum_{i \in I_0} \ln[\phi(y_{i1}; \mu_{i1}, \varphi_1)\{1 - C_1^*(\mu_{i2}, z_{i1})\}] + \sum_{i \in \bar{I}_0} \ln[\phi(y_{i1}; \mu_{i1}, \varphi_1)C_1^*(\mu_{i2}, z_{i1})] \\ &= \sum_{i=1}^{n} \ln \phi(y_{i1}; \mu_{i1}, \varphi_1) + \sum_{i \in I_0} \ln\{1 - C_1^*(\mu_{i2}, z_{i1})\} + \sum_{i \in \bar{I}_0} \ln C_1^*(\mu_{i2}, z_{i1}), \end{aligned}$$

where set $I_0 = \{i : y_{i2} = 0\}$ and $\bar{I}_0 = \{i : y_{i2} = 1\}$.

12.6.2 An Simulation Experiment

The marginal models used in the simulation study are specified as follows:

$$\begin{aligned} \mu_{i1} &= \beta_{01} + \beta_{11}x_{i1} \\ \Phi^{-1}(\mu_{i2}) &= \beta_{02} + \beta_{12}x_{i2} \end{aligned}$$

where x_{i1} and x_{i2} are generated randomly according to N(0, 1), and $\beta_{01} = \beta_{11} = \beta_{02} = \beta_{12} = 0.5$.

The correlated mixed outcomes are simulated as follows: First, simulate $y_{i1} \sim N(\mu_{i1}, \varphi_1)$ with $\varphi_1 = 1$, and then simulate $y_{i2} \sim$ Bernoulli$\{C_1^*(\mu_{i2}, z_{i1})\}$ with $\alpha = 0.5$. We generated 500 samples, each containing 300 pairs of observations. We analyzed the simulated data sets using both VGLM and the naive method that regards Y_1 and Y_2 as independent. In the naive method, Y_1 is analyzed using

Table 12.2 Average estimates, average standard errors (s.e.), and empirical standard deviations (e.s.d.) from the joint copula model and separate models for normal and binary mixed outcomes, respectively

		VGLM			Univariate models		
Model	β	$\hat{\beta}$	*s.e.*	*e.s.d.*	$\hat{\beta}$	*s.e.*	*e.s.d.*
Linear	$\beta_{01} = 0.5$	0.4976	0.0586	0.0560	0.4977	0.0576	0.0559
	$\beta_{11} = 0.5$	0.5032	0.0552	0.0567	0.5038	0.0577	0.0605
Probit	$\beta_{02} = 0.5$	0.5005	0.0799	0.0751	0.5011	0.0801	0.0755
	$\beta_{12} = 0.5$	0.5012	0.0812	0.0835	0.5019	0.0865	0.0895

Five hundred simulations were run in the study

the standard linear regression method, and Y_2 is analyzed using the probit model. Average estimates, standard errors, and empirical standard deviations over the 500 replications are listed in Table 12.2.

The estimation method in the VGLM appears to be numerically stable and is clearly more efficient than the naive method that uses univariate models to separately analyze the data, where correlated mixed outcomes were treated as if they were independent. The efficiency gain by the VGLM is expected to be more substantial when correlation between the mixed outcomes is high.

12.7 Concluding Remarks

This chapter presents a class of vector generalized linear models that can accommodate a variety of discrete, continuous and mixed vector outcomes. We developed a simultaneous maximum likelihood estimation and inference as well as composite likelihood estimation and inference, which were implemented by a Gauss–Newton type algorithm. Our focus of this paper is on the analysis of discrete and mixed outcomes through a joint model, since both have not been investigated thoroughly in the literature. An advantage of the presented theory is that all different types of data can be treated under one unified framework of modeling, estimation and inference.

Since few fully probability model-based methods are available for practitioners to analyze vector discrete and mixed data, the proposed models provide a powerful arsenal to conduct maximum likelihood or composite likelihood statistical inference in the line of the classical generalized linear models, which has been proved to be an appealing way in the statistical literature to study the relationship between the response and covariates of interest. With available full likelihood or composite likelihood, it is straightforward to define likelihood ratio type statistics in the context of hypothesis testing. Testing hypothesis based on likelihood ratio statistics for discrete outcomes presents a challenge to the GEEs approach because of the lack of likelihood function. An alternative to the composite likelihood approach is the simulated likelihood approach [17] that utilizes importance sampling algorithm

to overcome the numerical difficulty of evaluating Gaussian copula. This is a promising domain of research that is worth further exploration.

In the simulation studies and data analyses, we compared the VGLMs to GEEs. Our comparisons have clearly shown that estimators from the VGLMs appeared to be more efficient than estimators from GEEs for the regression coefficients. While the VGLMs enjoy better efficiency than the currently popular GEEs, the approach based on the full likelihood may be limited due to the computational burden involved in the evaluation of multivariate normal CDFs. However this limitation may be greatly reduced when the pairwise composite likelihood approach is used, in which only two-dimensional integrals need to be evaluated. Of course this computational gain may be paid by a certain potential loss of estimation efficiency and it is worth a further investigation about the amount of efficient loss in the context of VGLMs. In addition, as always when parametric models are applied for data analysis, model assumption diagnostics are necessary before the results are used to make final conclusions. Checking assumptions on the marginal model specifications can be done similarly as in the classical GLM theory. However, checking the dependence structure induced by Gaussian copula or even Gaussian copula itself is a challenging task, which has not been thoroughly investigated yet in the literature.

The VGLMs may be extended to handle multi-level data that often arise from many practical settings such as spatially clustered data. Bai [4] proposed a GeoCopula model to study the prevalence of malaria among village resident children in Gambia. Two thousand and thirty five children were randomly sampled from 65 villages along the Gambia river. Two levels of correlations, inter-village spatial correlation and intra-village correlation, arise in the data and need to be accounted for in the data analysis. To deal with such multi-level correlation, the dependence matrix may be written as of the form:

$$\Gamma = \Gamma_{village}(\alpha_v) \otimes \Gamma_{resident}(\alpha_r)$$

where $\Gamma_{village}(\alpha_f)$ and $\Gamma_{resident}(\alpha_t)$, respectively, characterize the spatial dependence among villages and the intra-village dependence among residents in a common village and $\otimes$ denotes the Kronecker product. Moreover, an explicit appearance of matrix Γ in the joint distribution enables us to specify desired variance structures to efficiently identify environmental risk factors associated with disease prevalence as well as to perform kriging. More details may be found in Bai's PhD thesis [4].

Appendix A: Gaussian–Hermite Quadrature

Let $\phi_2(x_1, x_2)$ be the bivariate normal density function with mean $\mathbf{0}$ and covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. We want to evaluate the following integral

$$\int_{-\infty}^{a}\int_{-\infty}^{b}\phi_2(x_1,x_2)dx_1dx_2$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}1_{\{x_1<a,x_2<b\}}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\{-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\}dx_1dx_2$$

where $\mathbf{x}^T=(x_1,x_2)$. Let

$$\mathbf{z}^T=(\frac{1}{2}\Sigma^{-\frac{1}{2}}\mathbf{x})^T=\frac{1}{\sqrt{2}}(x_1,x_2)\begin{pmatrix}\frac{1}{\sqrt{1-\rho^2}} & 0\\ -\frac{\rho}{\sqrt{1-\rho^2}} & 1\end{pmatrix}=(\frac{x_1-\rho x_2}{\sqrt{2(1-\rho^2)}},\frac{x_2}{\sqrt{2}}),$$

then

$$\int_{-\infty}^{a}\int_{-\infty}^{b}\phi_2(x_1,x_2)dx_1dx_2$$
$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}1_{\{\sqrt{2(1-\rho^2)}z_1+\sqrt{2}\rho z_2<a,\sqrt{2}z_2<b\}}\exp\{-(z_1^2+z_2^2)\}dz_1dz_2.$$

Let (z_{i1},z_{i2}) and (w_{i1},w_{i2}), $i=1,\ldots,Q$, be the abscissas and weights of Gaussian–Hermite quadrature, where Q is the number of quadrature points, then

$$\int_{-\infty}^{a}\int_{-\infty}^{b}\phi_2(x_1,x_2)dx_1dx_2$$
$$\approx\frac{1}{\pi}\sum_{i1=1}^{Q}\sum_{i2=1}^{Q}w_{i1}w_{i2}1[\sqrt{2(1-\rho^2)}z_{i1}+\sqrt{2}\rho z_{i2}<a,\sqrt{2}z_{i2}<b].$$

Similarly,

$$\int_{-\infty}^{a}\int_{-\infty}^{b}h(x_1,x_2)\phi_2(x_1,x_2)dx_1dx_2\approx\frac{1}{\pi}\sum_{i1=1}^{Q}\sum_{i2=1}^{Q}w_{i1}w_{i2}$$
$$\times h(\sqrt{2(1-\rho^2)}z_{i1}+\sqrt{2}\rho z_{i2},\sqrt{2}z_{i2})1[\sqrt{2(1-\rho^2)}z_{i1}+\sqrt{2}\rho z_{i2}<a,\sqrt{2}z_{i2}<b].$$

Appendix B: AREs of GEE and VGLM for Binary Data

Closed form expressions are available and used in the calculation of the asymptotic relative efficiency (ARE) for the bivariate binary data in Sect. 12.4.2. The following formulas in Appendices B.1 and B.2 are used to produce Fig. 12.3.

B.1: Efficiency of GEE Estimation

We consider the special case where the working correlation is equal to the true correlation, so the optimal asymptotic covariance matrix is

$$V_{gee}(\hat{\beta}) = \Big(\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta}^T \operatorname{cov}(\mathbf{Y}_i)^{-1} \frac{\partial \mu_i}{\partial \beta}\Big)^{-1}.$$

Clearly, the covariance matrix of $\mathbf{y}_i$ can be written in the form

$$\operatorname{cov}(\mathbf{Y}_i) = \begin{bmatrix} \operatorname{var}(Y_{i1}), & \operatorname{cov}(Y_{i1}, Y_{i2}) \\ \operatorname{cov}(Y_{i1}, Y_{i2}), & \operatorname{var}(Y_{i2}) \end{bmatrix} = \begin{bmatrix} \mu_{i1}(1-\mu_{i1}), & \operatorname{cov}(y_{i1}, y_{i2}) \\ \operatorname{cov}(y_{i1}, y_{i2}), & \mu_{i2}(1-\mu_{i2}) \end{bmatrix}$$

with

$$\operatorname{cov}(Y_{i1}, Y_{i2}) = \mu_{i1} + \mu_{i2} + \Phi_2(\Phi^{-1}(1-\mu_{i1}), \Phi^{-1}(1-\mu_{i2})|\alpha) - 1 - \mu_{i1}\mu_{i2}.$$

For the marginal logit model (12.17), it is easy to obtain

$$\frac{\partial \mu_i}{\partial \beta}^T = \begin{bmatrix} \mu_{i1}(1-\mu_{i1}), & \mu_{i2}(1-\mu_{i2}) \\ \mu_{i1}(1-\mu_{i1})x_{i1}, & \mu_{i2}(1-\mu_{i2})x_{i2} \end{bmatrix}.$$

Thus,

$$\begin{aligned} V_{gee} &= \left(\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta}^T \frac{\begin{bmatrix} \mu_{i2}(1-\mu_{i2}), & -\operatorname{cov}(Y_{i1}, Y_{i2}) \\ -\operatorname{cov}(Y_{i1}, Y_{i2}), & \mu_{i1}(1-\mu_{i1}) \end{bmatrix}}{\mu_{i1}(1-\mu_{i1})\mu_{i2}(1-\mu_{i2}) - \operatorname{cov}(Y_{i1}, Y_{i2})^2} \frac{\partial \mu_i}{\partial \beta}\right)^{-1} \\ &\stackrel{def.}{=} \begin{bmatrix} \sum_{i=1}^{n} v_{i,00}, & \sum_{i=1}^{n} v_{i,01} \\ \sum_{i=1}^{n} v_{i,10}, & \sum_{i=1}^{n} v_{i,11} \end{bmatrix}^{-1} \end{aligned}$$

Moreover, when $x_{i1} = x_{j1}$, and $x_{i2} = x_{j2}$, $i \neq j$, then $v_{i,lk} = v_{j,lk} = v_{lk}$, $l, k = 0, 1$, and

$$V_{gee} = \frac{1}{n(v_{00}v_{11} - v_{01}^2)} \begin{bmatrix} v_{11}, & -v_{01} \\ -v_{10}, & v_{00} \end{bmatrix}$$

So,

$$\begin{aligned} \widehat{\operatorname{var}}_{gee}(\hat{\beta}_0) &= \frac{v_{11}}{n(v_{00}v_{11} - {v_{01}}^2)}, \\ \widehat{\operatorname{var}}_{gee}(\hat{\beta}_1) &= \frac{v_{00}}{n(v_{00}v_{11} - {v_{01}}^2)}. \end{aligned}$$

For the probit model (12.17),

$$\frac{\partial \mu_i}{\partial \beta}^T = \begin{bmatrix} \phi(\Phi^{-1}(\mu_{i1})), & \phi(\Phi^{-1}(\mu_{i2})) \\ \phi(\Phi^{-1}(\mu_{i1}))x_{i1}, & \phi(\Phi^{-1}(\mu_{i2}))x_{i2} \end{bmatrix},$$

and the V_{gee} and $\widehat{\text{var}}_{gee}(\hat{\beta}_j)$, $j = 0, 1$ can be similarly calculated.

B.2: Efficiency of VGLM Estimation

Let $f_{lk} = P(Y_1 = l, Y_2 = k|\alpha), l, k = 1, 2$. It follows from Eq. (12.12) in Sect. 12.3.3 that for the logit model (12.17),

$$\begin{aligned}
\dot{f}_{11,\beta_0} &= \mu_1(1-\mu_1) + \mu_2(1-\mu_2) - \Phi\left(\frac{\Phi^{-1}(1-\mu_2) - \alpha\Phi^{-1}(1-\mu_1)}{\sqrt{1-\alpha^2}}\right)\mu_1(1-\mu_1) \\
&\quad -\Phi\left(\frac{\Phi^{-1}(1-\mu_1) - \alpha\Phi^{-1}(1-\mu_2)}{\sqrt{1-\alpha^2}}\right)\mu_2(1-\mu_2), \\
\dot{f}_{10,\beta_0} &= -\mu_2(1-\mu_2) - \Phi\left(\frac{\Phi^{-1}(1-\mu_1) - \alpha\Phi^{-1}(1-\mu_2)}{\sqrt{1-\alpha^2}}\right)\mu_1(1-\mu_1) \\
&\quad -\Phi\left(\frac{\Phi^{-1}(1-\mu_2) - \alpha\Phi^{-1}(1-\mu_1)}{\sqrt{1-\alpha^2}}\right)\mu_2(1-\mu_2), \\
\dot{f}_{01,\beta_0} &= -\mu_1(1-\mu_1) - \Phi\left(\frac{\Phi^{-1}(1-\mu_1) - \alpha\Phi^{-1}(1-\mu_2)}{\sqrt{1-\alpha^2}}\right)\mu_1(1-\mu_1) \\
&\quad -\Phi\left(\frac{\Phi^{-1}(1-\mu_2) - \alpha\Phi^{-1}(1-\mu_1)}{\sqrt{1-\alpha^2}}\right)\mu_2(1-\mu_2), \\
\dot{f}_{00,\beta_0} &= -\Phi\left(\frac{\Phi^{-1}(1-\mu_1) - \alpha\Phi^{-1}(1-\mu_2)}{\sqrt{1-\alpha^2}}\right)\mu_1(1-\mu_1) \\
&\quad -\Phi\left(\frac{\Phi^{-1}(1-\mu_2) - \alpha\Phi^{-1}(1-\mu_1)}{\sqrt{1-\alpha^2}}\right)\mu_2(1-\mu_2).
\end{aligned}$$

The Fisher Information matrix for the VGLM estimators is

$$I_{vglm} = \text{var}(\dot{l}_\beta) = \begin{bmatrix} \text{var}(\dot{l}_{\beta_0}), & \text{cov}(\dot{l}_{\beta_0}, \dot{l}_{\beta_1}) \\ \text{cov}(\dot{l}_{\beta_0}, \dot{l}_{\beta_1}), & \text{var}(\dot{l}_{\beta_1}) \end{bmatrix} = \begin{bmatrix} I_{00}, & I_{01} \\ I_{10}, & I_{11} \end{bmatrix},$$

where

$$
\begin{aligned}
I_{00} &= \mathrm{var}(\dot{l}_{\beta_0}) = \frac{\dot{f}_{11,\beta_0}^2}{f_{11}} + \frac{\dot{f}_{10,\beta_0}^2}{f_{10}} + \frac{\dot{f}_{01,\beta_0}^2}{f_{01}} + \frac{\dot{f}_{00,\beta_0}^2}{f_{00}} \\
I_{01} &= I_{10} = \mathrm{cov}(\dot{l}_{\beta_0}, \dot{l}_{\beta_1}) \\
&= \frac{\dot{f}_{11,\beta_0}\dot{f}_{11,\beta_1}}{f_{11}} + \frac{\dot{f}_{10,\beta_0}\dot{f}_{10,\beta_1}}{f_{10}} + \frac{\dot{f}_{01,\beta_0}\dot{f}_{01,\beta_1}}{f_{01}} + \frac{\dot{f}_{00,\beta_0}\dot{f}_{00,\beta_1}}{f_{00}} \\
I_{11} &= \mathrm{var}(\dot{l}_{\beta_0}) = \frac{\dot{f}_{11,\beta_1}^2}{f_{11}} + \frac{\dot{f}_{10,\beta_1}^2}{f_{10}} + \frac{\dot{f}_{01,\beta_1}^2}{f_{01}} + \frac{\dot{f}_{00,\beta_1}^2}{f_{00}}.
\end{aligned}
$$

The asymptotic covariance matrix of $\hat{\beta}$ is

$$
V_{vglm}(\hat{\beta}) = \frac{1}{n(I_{00}I_{11} - {I_{01}}^2)} \begin{bmatrix} I_{11}, & -I_{01} \\ -I_{10}, & I_{00} \end{bmatrix}.
$$

Therefore,

$$
\begin{aligned}
\widehat{\mathrm{var}}_{vglm}(\hat{\beta}_0) &= \frac{I_{11}}{n(I_{00}I_{11} - {I_{01}}^2)}, \\
\widehat{\mathrm{var}}_{vglm}(\hat{\beta}_1) &= \frac{I_{00}}{n(I_{00}I_{11} - {I_{01}}^2)}.
\end{aligned}
$$

For the probit model (12.17), the first derivatives $\dot{f}_{lk} = P(Y_1 = l, Y_2 = k|\alpha)$, $l, k = 1, 2$ can be similarly derived, and the only difference from the above logit model case is on the derivatives of $\frac{\partial \mu_j}{\partial \beta_k}, k = 0, 1$.

References

1. Anderson, J.A., Pemberton, J.D.: The grouped continuous model for multivariate ordered categorical variables and covariate adjustment. Biometrics **41**, 875–885 (1985)
2. Ashford, J.r., Sowden, R.R.: Multi-variate probit analysis. Biometrics **26**, 535–546 (1970)
3. Bahadur, R.R.: A Representation of the joint distribution of responses on *n* dichotomous items. In: Solomon, H. (ed.) Studies in Item Analysis and Prediction, Stanford Mathematical Studies in the Social Sciences VI, pp. 158–168. Stanford University Press, California (1961)
4. Bai, Y.: Joint composite estimating functions in spatial and spatio-temporal models. PhD Dissertation, Department of Biostatistics, University of Michigan (2011)
5. Bishop, Y.M.M., Fienberg, S.E., Holland, P.W.: Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge (1975)
6. Breslow, N.E., Clayton, D.G.: Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. **88**, 9–25 (1993)
7. Chaganty, N.R., Joe, H.: Efficiency of generalized estimating equations for binary responses. J. Roy. Stat. Soc. Ser. B **66**, 851–860 (2004)
8. Diggle, P.J., Liang, K.-Y., Heggerty, P., Zeger, S.L.: The Analysis of Longitudinal Data, 2nd edn. Oxford University Press, Oxford (2002)

9. Fitzmaurice, G.M., Laird, N.M.: Regression models for a bivariate discrete and continuous outcome with clustering. J. Am. Stat. Assoc. **90**, 845–852 (1995)
10. Fitzmaurice, G.M., Laird, N.M., Rotnitzky, A.G.: Regression models for discrete longitudinal responses. Stat. Sci. **8**, 284–309 (1993)
11. Genest, C., Nešlehová, J.: A primer on copulas for count data. Astin Bulletin **37**, 475–515 (2007)
12. Hills, A., Armitage, P.: The Two-period cross-over clinical trial. Brit. J. Clin. Pharmaco. **8**, 7–20 (1979)
13. Joe, H.: Multivariate Models and Dependence Concepts. Chapman and Hall, London (1997)
14. Jørgensen, B.: Exponential dispersion models (with discussion). J. Roy. Stat. Soc. Ser. B **49**, 127–162 (1987)
15. Jørgensen, B.: The Theory of Dispersion Models. Chapman and Hall, London (1997)
16. Liang, K.-Y., Zeger, S.L.: Longitudinal data analysis using generalized linear models. Biometrika **73**, 13–22 (1986)
17. Masarotto, G., Varin, C.: Gaussian copula marginal regression. Electron. J. Stat. **6**, 1517–1549 (2012)
18. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman and Hall, London (1989)
19. McCulloch, C.E., Searle, S.R.: Generalized, Linear, and Mixed Models. Wiley, New York (2001)
20. Nelder, J.A., Wedderburn, R.W.M.: Generalized linear models. J. Roy. Stat. Soc. Ser. A **135**, 370–384 (1972)
21. Piantadosi, S.: Clinical Trials: A Methodologic Perspective. Wiley, New York (1997)
22. Ruppert, D.: Discussion of maximization by parts in likelihood inference by song, fan, kalbfleisch. J. Am. Stat. Assoc. **100**, 1161–1163 (2005)
23. Sammel, M.D., Ryan, L.M., Legler, J.M.: Latent variable models for mixed discrete and continuous outcomes. J. Roy. Stat. Soc. Ser. B **59**, 667–678 (1997)
24. Sklar, A.: Fonctions de Répartition à n dimensions et leurs marges. Publications de L'Institut de Statistiques de l'Universite de Paris **8**, 229–231 (1959)
25. Song, P.X.-K.: Multivariate dispersion models generated from gaussian copula. Scand. J. Stat. **27**, 305–320 (2000)
26. Song, P.X.-K.: Correlated Data Analysis: Modeling Analytics and Applications. Springer, New York (2007)
27. Song, P.X.-K., Li, M., Yuan, Y.: Joint regression analysis of correlated data using Gaussian copulas. Biometrics **65**, 60–68 (2009)
28. Varin, C., Reid, N., Firth, D.: An overview of composite likelihood methods. Stat. Sinica **21**, 5–42 (2011)
29. Zhao, P.L., Prentice, R.L.: Correlated binary regression using a generalized quadratic model. Biometrika **77**, 642–648 (1990)

Chapter 13
Application of Bernstein Copulas to the Pricing of Multi-Asset Derivatives

Bertrand Tavin

Abstract This paper deals with the application of Bernstein copulas to the pricing of derivatives written on several underlying assets. We review the main characteristics of this particular family of copulas. We then analyze their properties in a context of multi-asset derivatives pricing, with a focus on the approximation property. We finally give details about implementation steps and provide numerical evidences to illustrate the reviewed properties.

13.1 Introduction

When facing a multivariate modeling problem, one needs a proper tool to model dependence. Copula functions are such a tool as they allow for the dependence to be modeled separately from the marginals, whenever these marginals are continuous. This paper focuses on a particular family of copulas, Bernstein copulas, and considers their application in finance. For further definitions, properties and references about copula functions, see [6] and the monograph [15].

Erdely and Díaz-Viera [8] and Hernández-Maldonado et al. [9] use Bernstein copulas in geology, for the modeling of dependence between petrophysical properties of oil reservoirs. In insurance, [5] uses Bernstein copulas to model the dependence between non-life insurance risks. In finance, [17] and [10] apply Bernstein copulas to the pricing of two-asset derivatives written on foreign exchange rates. Their approach focuses on the flexibility of the Bernstein copula that can be fitted to available market data, namely vanilla options on the cross exchange rate. In this paper we also work with Bernstein copulas for the pricing of multi-asset

B. Tavin (✉)
Department of Management and Business Economics, Université Paris 1 - Panthéon Sorbonne, Paris, France
e-mail: btavin@gmail.com

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*, Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6_13,

derivatives but our focus is different. It is on the approximation property of Bernstein copulas when a dependence model has already been chosen or fitted.

13.2 The Financial Framework

We consider a financial market with one period and $n+1$ primary assets, $t=0$ is the initial time and $t=T<+\infty$ is the final time. The final prices of the primary assets are modeled as positive random variables on $(\Omega,\mathscr{F},\mathbb{P})$ and are denoted by $\left(B_T,S_T^1,\ldots,S_T^n\right)$. The 0^{th} asset, B, is a maturity T and risk-free zero-coupon bond.

A European multi-asset derivative Z is a derivative that is written on up to n risky assets and that pays Z_T at maturity. Z_T is a positive random variable on $(\Omega,\mathscr{F},\mathbb{P})$ written $Z_T=z(S_T^1,\ldots,S_T^n)$ for a positive payoff function z on $[0,+\infty[^n$. In accordance with the First Fundamental Theorem of Asset Pricing, the time 0 price of Z can be obtained as the discounted expectation of its payoff under a risk-neutral probability measure $\mathbb{Q}$. The time 0 price of Z is denoted by Z_0 and writes $Z_0=B_0\mathbb{E}^{\mathbb{Q}}[Z_T]$.

A probability measure is said risk-neutral when it is equivalent to $\mathbb{P}$ and the discounted asset prices are $\mathbb{Q}$-martingales. There are many ways to build such a measure for pricing purposes. An approach that is particularly suitable to our context is to construct the joint distribution of log-returns of the primary asset prices in two steps. In the first step their marginal distributions are computed from, or fitted to, the available vanilla option prices. And in the second step a dependence structure is applied to the marginals by means of a copula function. This two-step approach allows for a separated modeling of risk factors and for flexibility in the choice of the risk-neutral marginals. Background and details on this approach can be found, among others, in [2, 16, 20] and in the monograph [3].

Gaussian and Student copulas are derived from the associated multivariate distributions and both are popular choices. The former is parametrized with a correlation matrix R. The latter is parametrized with a correlation matrix R and a degree of freedom ν, it is symmetric and have tail dependence. See [4] for details. The skew t copula is the copula derived from the multivariate skew t distribution built in [1]. It works with a correlation matrix R, a degree of freedom ν and a skew vector α. This copula is able to describe asymmetric dependence. See [1] and [12] for details.

Let $\left(Y_0^1,\ldots,Y_0^n\right)$ be the T-forward prices of the n risky assets and define the associated log-returns, for $k=1,\ldots,n$, as $X_T^k=\ln\left(S_T^k/Y_0^k\right)$. F denotes the joint distribution of $\left(X_T^1,\ldots,X_T^n\right)$ under $\mathbb{Q}$, with $F_1,\ldots,F_n$ its marginals. Z_T rewrites $Z_T=g(X_T^1,\ldots,X_T^n)$ with g a positive function on $]-\infty,+\infty[^n$. We denote by $\mathscr{C}^{(n)}$ the set of n-dimensional copulas. Let $C\in\mathscr{C}^{(n)}$ be the chosen risk-neutral copula to model the dependence structure of the asset price log-returns. We have

$$\mathbb{E}^{\mathbb{Q}}[Z_T] = \mathbb{E}^{\mathbb{Q}}\left[g\left(X_T^1, \ldots, X_T^n\right)\right] = \int_{]-\infty,+\infty[^n} g(x) dC(F_1(x_1), \ldots, F_n(x_n))$$
$$= \int_{u\in[0,1]^n} g(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) dC(u).$$

When C is absolutely continuous, and $c = \frac{\partial^n C}{\partial u_1 \ldots \partial u_n}$ is its density, the integral becomes

$$\mathbb{E}^{\mathbb{Q}}[Z_T] = \int_{u\in[0,1]^n} g(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) c(u) du \tag{13.1}$$

13.3 Bernstein Copulas and Their Properties

The family of Bernstein copulas was introduced in [13] and [14]. This family of copulas is built with Bernstein polynomials as building blocks. Bivariate Bernstein copulas are studied in [7] and their multivariate extension is considered in [19] and [18]. Janssen et al. [11] studies the asymptotic properties of the Bernstein copula estimator.

Definition 13.1 (Bernstein Polynomial).
$(B_{i,m})_{i=0}^m$ are the $m+1$ Bernstein polynomials of degree $m \in \mathbb{N}$, defined for $x \in [0,1]$ as

$$B_{i,m}(x) = \binom{m}{i} x^i (1-x)^{m-i}$$

Let $\mathscr{L}^{n,m}$ be a discretization of the n-dimensional unit hypercube $[0,1]^n$, with $m \in \mathbb{N}$ discretization steps in all dimensions, and written as

$$\mathscr{L}^{n,m} = \left\{ \left(\frac{\alpha_1}{m}, \ldots, \frac{\alpha_n}{m}\right) \middle| \alpha_j \in \mathbb{N} \text{ and } 0 \le \alpha_j \le m \text{ for } j = 1, \ldots, n \right\}$$

For ease of readability $u = (u_1, \ldots, u_n)$ will denote an element of $[0,1]^n$ and $v = (v_1, \ldots, v_n)$ will denote an element of $\mathscr{L}^{n,m}$, with $v_j = \frac{\alpha_j}{m}$ for some $\alpha_j \in \mathbb{N}$ and $0 \le \alpha_j \le m$ $(j = 1, \ldots, n)$ and so that v can also be written $\left(\frac{\alpha_1}{m}, \ldots, \frac{\alpha_n}{m}\right)$.

Definition 13.2 (Bernstein Copula).
For ξ a given real-valued function on $\mathscr{L}^{n,m}$, define $C_B^m : [0,1]^n \longrightarrow [0,1]$ as

$$C_B^m(u) = \sum_{v\in\mathscr{L}^{n,m}} \xi(v) \prod_{i=1}^n B_{\alpha_i,m}(u_i) = \sum_{\alpha_1=0}^m \cdots \sum_{\alpha_n=0}^m \left(\xi\left(\frac{\alpha_1}{m}, \ldots, \frac{\alpha_n}{m}\right) \prod_{i=1}^n B_{\alpha_i,m}(u_i) \right) \tag{13.2}$$

If ξ fulfills the two conditions stated below, then C_B^m is a proper copula, named Bernstein Copula with parameter function ξ.

1. For $0 \leq \alpha_j \leq m-1$ $(j = 1, \ldots, n)$ and with $\delta_n = 0$ or 1 whether n is even or odd, respectively.

$$\sum_{l_1=0}^{1} \cdots \sum_{l_n=0}^{1} (-1)^{\left(\delta_n + \sum_{j=1}^{n} l_j\right)} \xi\left(\frac{\alpha_1 + l_1}{m}, \ldots, \frac{\alpha_n + l_n}{m}\right) \geq 0$$

2. For $v \in \mathscr{L}^{n,m}$

$$\max\left(\sum_{j=1}^{n} v_j - n + 1, 0\right) \leq \xi(v_1, \ldots, v_n) \leq \min_{j=1,\ldots,n}(v_j)$$

In the sequel and unless stated differently, we consider C_B^m an order m Bernstein copula with parameter function ξ. Let $\Delta_n \xi$ be the n-dimensional volume operator applied to the function ξ and defined, for $0 \leq \alpha_j \leq m-1$ $(j = 1, \ldots, n)$, as

$$\Delta_n \xi\left(\frac{\alpha_1}{m}, \ldots, \frac{\alpha_n}{m}\right) = \sum_{l_1=0}^{1} \cdots \sum_{l_n=0}^{1} (-1)^{\left(\delta_n + \sum_{j=1}^{n} l_j\right)} \xi\left(\frac{\alpha_1 + l_1}{m}, \ldots, \frac{\alpha_n + l_n}{m}\right)$$

Definition 13.3 (Bernstein Copula Density).
The Bernstein copula C_B^m is absolutely continuous and has density c_B^m defined as

$$c_B^m(u) = \frac{\partial^n C_B^m}{\partial u_1 \ldots \partial u_n}(u) = \sum_{\alpha_1=0}^{m-1} \cdots \sum_{\alpha_n=0}^{m-1} \left(\Delta_n \xi\left(\frac{\alpha_1}{m}, \ldots, \frac{\alpha_n}{m}\right) \prod_{i=1}^{n} B_{\alpha_i, m-1}(u_i)\right) \tag{13.3}$$

As it is proven in [19], a Bernstein copula can always be decomposed as a sum of the product copula and a perturbation term. This decomposition is written

$$C_B^m(u) = \prod_{i=1}^{n} u_i + \sum_{v \in \mathscr{L}^{n,m}} \gamma(v) \prod_{i=1}^{n} B_{\alpha_i, m}(u_i)$$

$$\gamma(v) = \xi(v) - \prod_{i=1}^{n} v_i$$

If $C \in \mathscr{C}^{(n)}$, then $C_B^m(C)$ defined, for $u \in [0,1]^n$, as

$$C_B^m(C)(u) = \sum_{v \in \mathscr{L}^{n,m}} C(v) \prod_{i=1}^{n} B_{\alpha_i, m}(u_i) = \sum_{\alpha_1=0}^{m} \cdots \sum_{\alpha_n=0}^{m} \left(C\left(\frac{\alpha_1}{m}, \ldots, \frac{\alpha_n}{m}\right) \prod_{i=1}^{n} B_{\alpha_i, m}(u_i)\right) \tag{13.4}$$

is a proper copula named Bernstein copula approximation of C with order m. The associated parameter function is written $\xi(C)(v) = C\,(v_1, \ldots, v_n)$, for $v \in \mathscr{L}^{n,m}$. $C_B^m(C)$ uniformly converges to C as the order m grows, so that a given copula can be approximated to any precision level by a Bernstein copula.

The decomposition result is useful to understand how the Bernstein copula approximation behaves. If the copula to be approximated is the product copula, then the perturbation term is zero as well as the approximation error. If a given copula to be approximated is different from the product copula, then the perturbation term departs from zero. In probabilistic terms, the product copula represents independence. We could expect that, for a fixed order m, the approximation quality worsens as the given copula represents a dependence structure that departs from independence.

In order to investigate this behavior, we restrict ourselves to the bivariate case and we consider a measure of the approximation error that is defined, for $C \in \mathscr{C}^{(n)}$ and $m \in \mathbb{N}$, as

$$sup_{u\in[0,1]^2}|C_B^m(C)(u) - C(u)| \tag{13.5}$$

This kind of sup-norm distance-based measure of the approximation error is used in [7] in the same context. It takes only positive values and goes to zero as the order m grows.

In the bivariate case, there are, at least, three ways for a copula to depart from independence, namely association, tail dependence, and asymmetry. We now consider the approximation of different families of parametric copulas and we compute the approximation error (13.5) for different sets of parameters. In Figs. 13.1, 13.2, and 13.3 below, we plot the approximation error as a function of the Bernstein copula order m. Plotted values are obtained with Matlab minimization routine `fmincon`.

In Fig. 13.1 the approximated copulas are Gaussian copulas with different correlation parameters, corresponding to different levels of association. The three curves of approximation error are ordered according to the levels of association. As expected, the lower the association level is, the lower the approximation error is.

In Fig. 13.2 the approximated copulas are Student copulas with a fixed correlation parameter and different degrees of freedom, corresponding to different levels of tail dependence. Even though the three curves of approximation error are close to each other, they are ordered according to the levels of tail dependence. As expected, the approximation error worsens with the level of tail dependence.

In Fig. 13.3 the approximated copulas are skew t copulas with fixed correlation and degree of freedom and different skew parameter vectors, corresponding to different cases of asymmetry. The three curves of approximation error are ordered according to the corresponding levels of asymmetry. As expected, the approximation error worsens with the level of asymmetry. Curves corresponding to asymmetric cases decrease at a slower rate than the curve corresponding to the symmetric case.

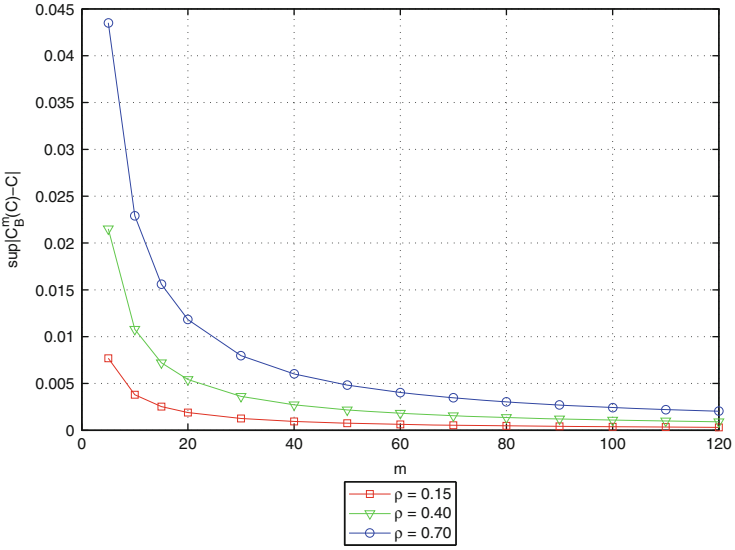

Fig. 13.1 Error measure associated with Bernstein copula approximations of Gaussian copulas, as a function of the order m and for different correlation parameter values

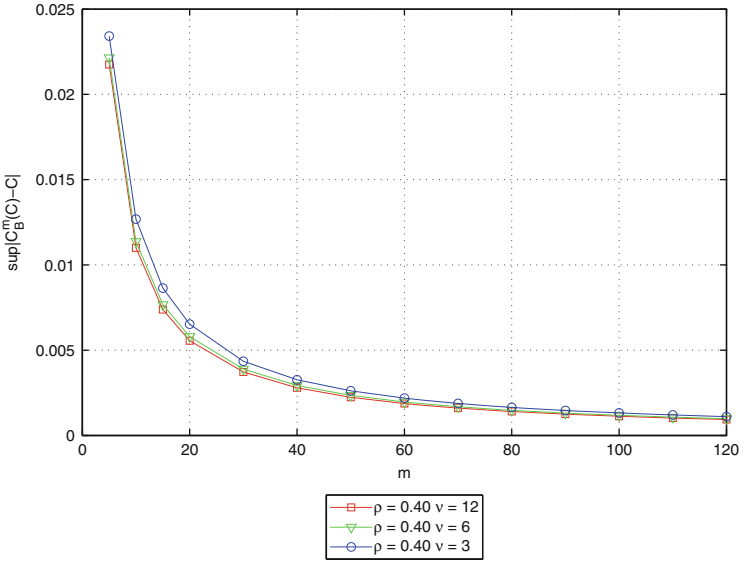

Fig. 13.2 Error measure associated with Bernstein copula approximations of Student copulas, as a function of the order m and for different degrees of freedom

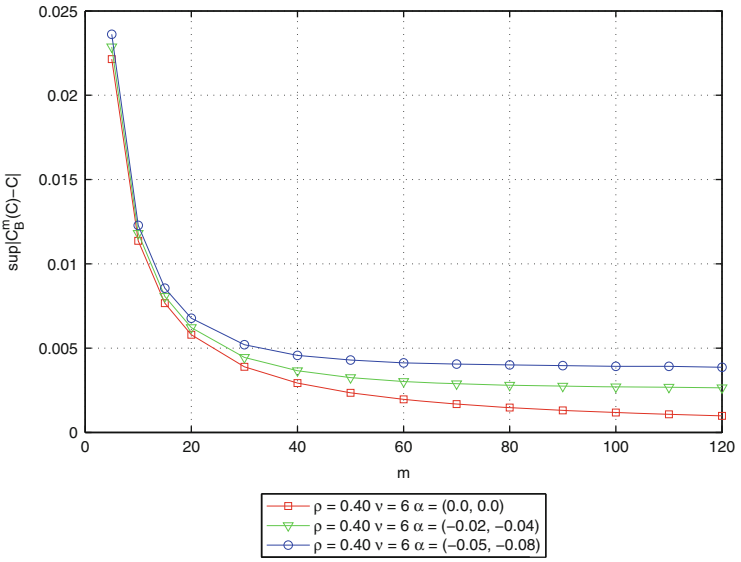

Fig. 13.3 Error measure associated with Bernstein copula approximations of skew t copulas, as a function of the order m and for different skew vectors

The behavior of the Bernstein copula approximation hence depends on the characteristics of the approximated copula. The initial intuition that this behavior usually depends on how the approximated copula departs from independence is thereby confirmed by these numerical investigations.

13.4 The Pricing of Multi-Asset Derivatives

Within the framework of Sect. 13.2 we consider the pricing of multi-asset derivatives by means of Bernstein copulas, the definition and properties of which were detailed in Sect. 13.3. Let $C \in \mathscr{C}^{(n)}$ be the chosen risk-neutral copula and $C_B^m(C)$ be its order m Bernstein copula approximation. An approximation of Z_0, the multi-asset derivative price, is obtained by replacing the chosen copula by its approximation. Bernstein copulas are absolutely continuous. The multiple integral to be computed then writes

$$\mathbb{E}^{\mathbb{Q}}[Z_T] \approx \int_{u\in[0,1]^n} g(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) c_B^m(C)(u) du \tag{13.6}$$

This representation is particularly suitable for the use of numerical quadrature methods or quasi Monte Carlo integration methods. The choice between the two methods to solve this multiple integral usually depends on its dimensionality.

In the sequel of this section we restrict ourselves to the two-asset case. We perform a numerical investigation for common multi-asset payoffs, namely calls on the equally weighted basket of S_1 and S_2 and puts on the maximum of the same pair of assets. We work with realistic market data conditions that could correspond to derivatives written on equity market indices such as the French and German euro-denominated market indices, CAC40 and DAX30. To compute the prices of basket options we use the Matlab quadrature routine `quad2d`. To compute the prices of puts on the maximum we use the one-dimensional Matlab quadrature routine `quadgk` because the multiple integrals (13.1) and (13.6) simplify to one dimensional integrals, see [3] for details about this simplification.

We consider different parametric copulas. G_1 and G_2 are Gaussian copulas with parameters $\rho = 0.55$ and $\rho = 0.75$, respectively. ST_1 is a Student copula with parameters $\rho = 0.55$ and $\nu = 6$. SKT_1 and SKT_2 are skew t copulas with respective parameters $(\rho = 0.55, \nu = 6, \alpha_1 = -0.02, \alpha_2 = -0.04)$ and $(\rho = 0.55, \nu = 6, \alpha_1 = -0.05, \alpha_2 = -0.08)$.

We consider the risk-neutral marginal distributions of X_T^1 and X_T^2 to be normal inverse gaussian (NIG) and Gaussian distributions. The NIG distributions for the three months log-returns have parameters $(a_1 = 22.8, b_1 = -16.0, m_1 = 0.10, d_1 = 0.11)$ and $(a_2 = 26.9, b_2 = -18.3, m_2 = 0.10, d_2 = 0.11)$ and the Gaussian distributions for the same log-returns are parametrized with the corresponding ATM volatilities, respectively $\sigma_1 = 0.208$ and $\sigma_2 = 0.188$. The NIG distributions for the six months log-returns have parameters $(a_1 = 16.4, b_1 = -12.5, m_1 = 0.15, d_1 = 0.14)$ and $(a_2 = 17.8, b_2 = -13.4, m_2 = 0.15, d_2 = 0.15)$. The other market parameters are as follows. Three and six months forward prices of S_1 and S_2 are both set equal to 100. The three and six months zero-coupon bond prices are 0.9975 and 0.9950, respectively.

In Tables 13.1, 13.2, and 13.3 below we have gathered the exact and approximated prices, as well as the differences between both, for basket options struck below, above and at the underlying forward value. Their strikes are denoted by K. In Table 13.4 we have gathered the exact and approximated prices, as well as the differences between both, for puts on the maximum struck below, above and close to the underlying forward value. Their strikes are also denoted by K.

For basket options, the considered Bernstein copula order is $m = 50$ and for puts on the maximum, this order is $m = 120$. These values are chosen for an illustrative purpose and correspond, at the same time, to an acceptable precision for the approximated price and to a reasonable computational load to handle. The considered order is larger for puts on the maximum than for basket options because the price of the former is more sensitive to the dependence structure.

All computations are done with Matlab routines on a personal laptop and the required computation times are indicated in the last column of the tables.

Tables 13.1 and 13.4 show that the use of Bernstein copulas can lead to acceptable levels of approximation while keeping computation time reasonable. It is particularly true for basket options, for which the pricing error magnitude is around two cents of the underlying forward value. Tables 13.2 and 13.3 confirm this feature as the magnitude of pricing errors is little affected by a change in marginal

Table 13.1 Prices and approximation errors for call options on the equally weighted basket of S_1 and S_2

Copula	$K = 95$	$K = 98$	$K = 100$	$K = 102$	$K = 105$	Comp. time
Gaussian G_1	6.747	4.658	3.462	2.451	1.309	5.7 s
Gaussian G_2	6.943	4.865	3.667	2.643	1.465	4.0 s
Student ST_1	6.728	4.627	3.428	2.420	1.290	6.9 s
Skew t SKT_1	6.762	4.675	3.486	2.488	1.373	197 s
Skew t SKT_2	6.797	4.724	3.547	2.561	1.464	196 s
Bernstein $\xi(G_1)$	6.725	4.637	3.442	2.433	1.292	9.1 s
Bernstein $\xi(G_2)$	6.915	4.839	3.642	2.620	1.444	5.0 s
Bernstein $\xi(ST_1)$	6.708	4.608	3.410	2.403	1.275	11.2 s
Bernstein $\xi(SKT_1)$	6.747	4.661	3.474	2.477	1.365	10.2 s
Bernstein $\xi(SKT_2)$	6.788	4.718	3.542	2.559	1.466	10.2 s
Price diff. (G_1)	0.022	0.021	0.02	0.019	0.017	
Price diff. (G_2)	0.028	0.026	0.025	0.023	0.020	
Price diff. (ST_1)	0.020	0.019	0.018	0.017	0.015	
Price diff. (SKT_1)	0.016	0.013	0.012	0.010	0.008	
Price diff. (SKT_2)	0.01	0.006	0.004	0.002	0.002	

Maturity is $T = 3$ months, marginals are NIG and the Bernstein copulas order is $m = 50$

Table 13.2 Prices and approximation errors for call options on the equally weighted basket of S_1 and S_2

Copula	$K = 95$	$K = 98$	$K = 100$	$K = 102$	$K = 105$	Comp. time
Gaussian G_1	6.444	4.523	3.469	2.596	1.604	0.1 s
Gaussian G_2	6.623	4.732	3.685	2.809	1.795	0.1 s
Student ST_1	6.423	4.490	3.433	2.563	1.583	0.2 s
Skew t SKT_1	6.450	4.531	3.485	2.626	1.660	173.1 s
Skew t SKT_2	6.478	4.575	3.540	2.694	1.746	173.8 s
Bernstein $\xi(G_1)$	6.425	4.503	3.448	2.575	1.583	7.7 s
Bernstein $\xi(G_2)$	6.599	4.706	3.659	2.783	1.769	5.6 s
Bernstein $\xi(ST_1)$	6.406	4.472	3.414	2.544	1.564	7.6 s
Bernstein $\xi(SKT_1)$	6.437	4.517	3.471	2.612	1.647	7.6 s
Bernstein $\xi(SKT_2)$	6.469	4.567	3.533	2.687	1.740	7.6 s
Price diff. (G_1)	0.019	0.021	0.021	0.021	0.021	
Price diff. (G_2)	0.024	0.025	0.026	0.026	0.026	
Price diff. (ST_1)	0.017	0.018	0.019	0.019	0.019	
Price diff. (SKT_1)	0.014	0.014	0.014	0.014	0.014	
Price diff. (SKT_2)	0.009	0.008	0.008	0.007	0.006	

Maturity is $T = 3$ months, marginals are Gaussian and the Bernstein copulas order is $m = 50$

distributions or in time to maturity. Across our numerical results, the main driving factors of the approximation behavior are the choice of the approximated copula, particularly its correlation parameter and the derivative payoff. This is in accordance with the remarks made in Sect. 13.2.

Table 13.3 Prices and approximation errors for call options on the equally weighted basket of S_1 and S_2

Copula	$K = 95$	$K = 98$	$K = 100$	$K = 102$	$K = 105$	Comp. time
Gaussian G_1	8.339	6.382	5.214	4.166	2.836	7.1 s
Gaussian G_2	8.652	6.698	5.525	4.464	3.104	4.7 s
Student ST_1	8.300	6.334	5.163	4.117	2.796	8.9 s
Skew t SKT_1	8.362	6.410	5.249	4.213	2.908	253.3 s
Skew t SKT_2	8.426	6.490	5.341	4.317	3.030	244.4 s
Bernstein $\xi(G_1)$	8.306	6.351	5.183	4.137	2.810	12.5 s
Bernstein $\xi(G_2)$	8.611	6.659	5.487	4.428	3.071	6.9 s
Bernstein $\xi(ST_1)$	8.270	6.306	5.137	4.091	2.772	12.6 s
Bernstein $\xi(SKT_1)$	8.341	6.391	5.233	4.198	2.896	12.8 s
Bernstein $\xi(SKT_2)$	8.415	6.483	5.337	4.315	3.032	12.4 s
Price diff. (G_1)	0.033	0.031	0.03	0.029	0.027	
Price diff. (G_2)	0.041	0.039	0.038	0.036	0.033	
Price diff. (ST_1)	0.030	0.028	0.027	0.026	0.024	
Price diff. (SKT_1)	0.021	0.018	0.016	0.015	0.012	
Price diff. (SKT_2)	0.010	0.006	0.003	0.001	0.003	

Maturity is $T = 6$ months, marginals are NIG and the Bernstein copulas order is $m = 50$

Table 13.4 Prices and approximation errors for put options written on the maximum of S_1 and S_2

Copula	$K = 105$	$K = 108$	$K = 110$	$K = 112$	$K = 115$	Comp. time
Gaussian G_1	3.799	5.599	7.062	8.702	11.402	0.7 s
Gaussian G_2	4.516	6.416	7.927	9.600	12.327	0.6 s
Student ST_1	3.888	5.695	7.164	8.812	11.522	0.6 s
Skew t SKT_1	3.960	5.782	7.261	8.915	11.631	12.6 s
Skew t SKT_2	4.037	5.877	7.366	9.028	11.752	12.6 s
Bernstein $\xi(G_1)$	3.763	5.560	7.020	8.658	11.356	2.2 s
Bernstein $\xi(G_2)$	4.445	6.339	7.846	9.516	12.239	2.2 s
Bernstein $\xi(ST_1)$	3.836	5.639	7.106	8.751	11.458	2.3 s
Bernstein $\xi(SKT_1)$	3.906	5.724	7.199	8.851	11.563	2.3 s
Bernstein $\xi(SKT_2)$	3.981	5.816	7.301	8.961	11.680	2.4 s
Price diff. (G_1)	0.036	0.040	0.042	0.044	0.046	
Price diff. (G_2)	0.071	0.077	0.081	0.084	0.089	
Price diff. (ST_1)	0.052	0.056	0.059	0.061	0.064	
Price diff. (SKT_1)	0.054	0.059	0.062	0.065	0.068	
Price diff. (SKT_2)	0.055	0.061	0.064	0.067	0.072	

Maturity is $T = 3$ months and marginals are NIG distributions. The Bernstein copulas order is $m = 120$

From Tables 13.1, 13.2, and 13.3, pricing with Bernstein copulas appears particularly suitable when the chosen copula is slow to compute because it has a complex expression, like the skew t copula. In such a case, the approach with Bernstein copulas offers a clear reduction of the computation time required to value

a portfolio of derivatives. The reduction of computational time, however smaller, is confirmed for rainbow options by Table 13.4. This smaller edge is explained by the possibility to compute the exact prices faster with one dimensional integrals and by the higher order of a Bernstein copula required to have an acceptable level of precision.

Acknowledgments I thank two anonymous referees and an editor for their comments on an earlier version.

References

1. Azzalini, A., Capitanio, A.: Distribution generated by perturbation of symmetry with emphasis on a multivariate skew *t* distribution. J. R. Stat. Soc. Ser. B Stat. Methodol. **65**, 367–389 (2003)
2. Cherubini, U., Luciano, E.: Bivariate option pricing with copulas. Appl. Math. Finance **9**, 69–82 (2002)
3. Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance, Wiley Finance. Wiley, Chichester (2004)
4. Demarta, S., McNeil, A.: The *t* Copula and related copulas. Int. Stat. Rev. **73**, 111–129 (2005)
5. Diers, D., Eling, M., Marek, S.: Dependence modeling in non-life insurance using the Bernstein copula. Insur. Math. Econ. **50**(3), 430–436 (2012)
6. Durante, F., Sempi, C.: Copula theory: an introduction. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications, Proceedings of the Workshop Held in Warsaw 25–26 September 2009, Lecture Notes in Statistics—Proceedings, vol. 198. Springer, Dordrecht (2010)
7. Durrleman, V., Nikeghbali, A., Roncalli, T.: Copulas approximations and new families. Working Paper, Groupe de Recherche Opérationnelle du Crédit Lyonnais (2000)
8. Erdely, A., Díaz-Viera, M.: Nonparametric and semiparametric bivariate modeling of petrophysical porosity-permeability dependence from well log data. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications, Proceedings of the Workshop Held in Warsaw 25–26 September 2009, Lecture Notes in Statistics—Proceedings, vol. 198. Springer, Dordrecht (2010)
9. Hernández-Maldonado, V., Díaz-Viera, M., Erdely, A.: A joint stochastic simulation method using the Bernstein copula as a flexible tool for modeling nonlinear dependence structures between petrophysical properties. J. Petrophys. Sci. Eng. **90–91**, 112–123 (2012)
10. Hurd, M., Salmon, M., Schleicher, C.: Using copulas to construct bivariate foreign exchange distributions with an application to the sterling exchange rate index. Bank of England, Working paper 334 (2007)
11. Janssen, P., Swanepoel, J., Veraverbeke, N.: Large sample behavior of the Bernstein copula estimator. J. Stat. Plan. Inference **142**(5), 1189–1197 (2012)
12. Kollo, T., Pettere, G.: Parameter estimation and application of the multivariate skew *t*-copula. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications, Proceedings of the Workshop Held in Warsaw 25–26 September 2009, Lecture Notes in Statistics—Proceedings, vol. 198. Springer, Dordrecht (2010)
13. Li, X., Mikusiński, P., Sherwood, H., Taylor, M.: On approximations of copulas. In: Beneš, V., Štěpán, J. (eds.) Distributions With Given Marginals and Moment Problems. Kluwer, Dordrecht (1997)
14. Li, X., Mikusiński, P., Taylor, M.: Strong approximation of copulas. J. Math. Anal. Appl. **225**(2), 608–623 (1998)

15. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistics, vol. 139. Springer, New York (1999)
16. Rosenberg, J.: Nonparametric pricing of multivariate contingent claims. J. Deriv. **10**(3), 9–26 (2003)
17. Salmon, M., Schleicher, C.: Pricing multivariate currency options with copulas. In: Rank, J. (ed.) Copulas: From Theory to Application in Finance. RISK Books, London (2006)
18. Sancetta, A.: Nonparametric estimation of distributions with given marginals via Bernstein-Kantorovitch Polynomilas: L_1 and pointwise convergence theory. J. Multivar. Anal. **98**(7), 1376–1390 (2007)
19. Sancetta, A., Satchell, S.: The Bernstein copula and its application to modeling and approximations of multivariate distributions. Econom. Theory **20**(3), 535–562 (2004)
20. van der Goorbergh, R., Genest, C., Werker, B.: Bivariate option pricing using dynamic copula models. Insur. Math. Econ. **37**(1), 101–114 (2005)

Index

P. Jaworski et al. (eds.), *Copulae in Mathematical and Quantitative Finance*, Lecture Notes in Statistics 213, DOI 10.1007/978-3-642-35407-6,