Journal of Banking & Finance 54 (2015) 175–191

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Mixture pair-copula-constructions ☆

Gregor N.F. Weiß *, Marcus Scheffer [1]

Technische Universität Dortmund, Otto-Hahn-Str. 6, D-44227 Dortmund, Germany

ARTICLE INFO

Article history:
Received 7 May 2014
Accepted 15 January 2015
Available online 31 January 2015

JEL classification:
C52
C53
C58

Keywords:
Dependence structures
Vine copulas
Mixture copulas
Model selection

ABSTRACT

We propose the use of convex combinations of parametric copulas as pair-copulas in high-dimensional vine copula models. By doing so, we circumvent the error-prone need to choose and estimate a parametric copula for each pair-copula in a vine model. We show in simulations that our proposed model fits the dependence structure in a given data sample significantly better than a competing benchmark. In our empirical study on the models' accuracy for forecasting the Value-at-Risk of financial portfolios, we show that our proposed mixture pair-copula construction yields significantly better results in backtesting while the benchmark overestimates portfolio risk.

1. Introduction

It has become a stylized fact in both the finance as well as the risk management literature that elliptical models such as the multivariate Gaussian distribution cannot fully capture the dependence structures often found in financial asset returns. Starting with the work by Embrechts et al. (2002), several studies have criticized the inadequacy of correlation-based models for modeling the non-linear dependence in financial returns advocating the use of copulas instead. At the same time, elliptical copula models, which have become an industry standard in credit risk modeling following the influential study by Li (2000), have been found to perform just as poorly as their correlation-based counterparts due to their symmetric tail independence (see, e.g., Cherubini et al., 2004; Fischer et al., 2009). Especially in times of financial market turmoil, neglecting the tail dependence between financial time series can have disastrous effects on both banks and insurers as evidenced during the recent financial crisis.

While Archimedean copulas have been found to adequately model the lower tail dependence in bivariate financial portfolios (see, e.g., Weiß, 2011), simple parametric copula models are often not flexible enough to model the complex dependence structures of multivariate data. Consequently, studies by Joe (1996, 1997), Bedford and Cooke (2001, 2002) and Whelan (2004) have tried to construct high-dimensional copula models which are flexible enough to model complex multivariate data sets and yet at the same time tractable. The most prominent example of these high-dimensional copula models are the so-called vine copulas (also called pair-copula constructions, PCC).[2] Vine copulas are hierarchical in nature and only require the specification of bivariate copulas conditional on certain sets of variables (so called pair-copulas). The first application of vines in a risk management setting is due to Aas et al. (2009) and since then, vine copulas have emerged as the method of choice for modeling high-dimensional dependence structures due to their enormous flexibility.[3]

☆ We received helpful comments from Carol Alexander (the editor) and two anonymous referees. Janet Gabrysch, Sandra Gabrysch, Felix Irresberger, Janina Mühlnickel, Anne-Christine Schmidt, and Sara Schmidt provided outstanding research assistance. Support by the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823, project A7) of the German Research Foundation (DFG) is gratefully acknowledged.

* Corresponding author. Tel.: +49 231 755 4608.
E-mail addresses: gregor.weiss@tu-dortmund.de (G.N.F. Weiß), marcus.scheffer@tu-dortmund.de (M. Scheffer).
[1] Tel.: +49 231 755 4231.

[2] Overviews of different nested and hierarchical copula models in high dimensions are given in Aas and Berg (2009) and Fischer et al. (2009). A recent application of hierarchical Marshall–Olkin copula models to the estimation of the systemic risk of different countries is due to Baglioni and Cherubini (2013).

[3] For different applications of vines in asset pricing and risk management see Chollete and Heinen (2009), Heinen and Valdesogo (2009) and Weiß and Supper (2013).

http://dx.doi.org/10.1016/j.jbankfin.2015.01.008
0378-4266/

Although vine copulas are extremely flexible tools for modeling multivariate data, the increase in flexibility comes at the price of a greatly increased model risk. To fully specify a d-dimensional vine copula, one needs to choose $d(d-1)/2$ different pair-copulas from a set of candidate bivariate parametric copula families. Similar to the case of non-nested copulas, the question of how to select the optimal parametric copula family for the pair-copulas remains unanswered.[4] Methods for selecting the pair-copulas in vine copulas include the selection based on graphical data inspection and goodness-of-fit tests (Aas et al., 2009), sequential heuristics based on Akaike's Information Criterion (Brechmann et al., 2012) as well as tests based on empirical pair-copulas (Hobæk-Haff and Segers, 2015). A different approach is taken by Weiß and Scheffer (2012), who substitute the parametric pair-copulas in vines by smooth nonparametric Bernstein copulas to circumvent the otherwise necessary selection of parametric copula families for the pair-copulas. Finally, Kurowicka (2010) and Brechmann et al. (2012) propose strategies for simplifying vines by replacing certain pair-copulas by the independence copula (yielding a *truncated* vine copula) or the Gaussian copula (yielding a *simplified vine*).

In this article, we propose a new approach to solve the problem of selecting the pair-copulas in a vine model. As pair-copulas, we employ mixture copulas (i.e., convex combinations of parametric copulas) yielding so-called *mixture pair-copula-constructions* (*Mixture-PCCs* in short). Our modeling strategy is related to the work of Kim et al. (2013) and their idea of using mixtures of D-vines but tackles the problem of constructing a vine model from a different perspective. While their study is concerned with the use of mixtures of D-vines (yielding multivariate models of even higher flexibility), we construct C- and D-vines with mixture pair-copulas to minimize the possibility of misspecifying a vine model. The use of bivariate mixture copulas as pair-copulas is beneficial for two reasons: First, the use of mixture copulas instead of simple bivariate parametric copulas increases the flexibility of a vine model even further. As all mixture copulas are bivariate and thus tractable in estimation, and as pruning strategies like simplifying or truncating the vine can also be applied to Mixture-PCCs, the increase in flexibility is not canceled out by an increase in computational complexity. Second, mixture-pair-copula constructions completely obviate the need for the error-prone selection of pair-copulas from pre-specified sets of parametric copulas. Thus, this paper contributes significantly to the current state of the art by proposing an extremely flexible yet still tractable model for high-dimensional dependence structures that lacks the model risk of current vine model specifications.

We illustrate the superiority of our proposed model by performing both a simulation and an empirical study on the in-sample and out-of-sample Value-at-Risk forecasting accuracy of Mixture-PCCs and heuristically calibrated vine copulas. The results from our simulation study show that our proposed Mixture-PCCs produce Value-at-Risk estimates that possess a comparable and especially in higher dimensions a better in-sample fit than the heuristic benchmark from the recent literature. In the empirical study of a four-dimensional financial portfolio, our proposed Mixture-PCC is characterized by a significantly better out-of-sample fit than the benchmark which overestimates portfolio risk. Our model thus helps risk managers to reduce regulatory risk capital. At the same time, portfolio losses are satisfactorily forecasted.

The remainder of this article is structured as follows. Section 2 introduces the basic properties of vine and mixture copulas and outlines the idea to combine both of them into Mixture-PCCs. Section 3 presents the results of our simulation study on the in-sample fit of Mixture-PCCs for Value-at-Risk forecasting. In Section 4, we discuss the results of our empirical study in which we compare the out-of-sample fit of a Mixture-PCC model with that of a heuristic benchmark from the literature based on the Value-at-Risk forecasts for a four-dimensional financial portfolio. Section 5 concludes.

2. Combining mixture and vine copulas

The purpose of this section is to shortly review the basic properties of vine copulas, mixture copulas, and the combination of the two, respectively.

2.1. Pair-copula constructions

In essence, copulas can be used to separate a multivariate distribution into its marginals and the dependence structure which is fully captured by the copula. This idea is formalized in Sklar's theorem (1959) which states that a d-dimensional cumulative distribution function (cdf) F can be split in two parts, the marginal distribution functions F_i and a copula C which is a d-variate cdf on $[0;1]^d$ with uniform marginals and which fully describes the dependence structure inherent in F:[5]

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)), \tag{1}$$

with $\mathbf{x} = (x_1, x_2, \ldots, x_d)$. Similarly, the joint multivariate density f can be represented by

$$f(\mathbf{x}) = c(F_1(x_1), \ldots, F_d(x_d)) \prod_{i=1}^{d} f_i(x_i), \tag{2}$$

where $c(u_1, \ldots, u_d)$ is the d-variate copula density given by $\frac{\partial C(u_1,\ldots,u_d)}{\partial u_1 \ldots \partial u_d}$ and f_i $(i = 1, \ldots, d)$ are the marginal densities.

Building on this basic result, pair-copula constructions present an extremely flexible way of decomposing the multivariate density f into a cascade of bivariate copula densities.[6] We start our discussion of PCCs by observing that a joint probability density function of dimension d can be decomposed into its unconditional and conditional marginal densities via

$$f(x_1, \ldots, x_d) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_1, x_2) \cdots f(x_d|x_1, \ldots, x_{d-1}). \tag{3}$$

Each conditional marginal density in this product can then be decomposed further using copula densities, e.g., via the relation

$$f(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2), \tag{4}$$

with $F_i(\cdot)$ being the cumulative distribution function (cdf) of x_i $(i = 1, \ldots, d)$ and $c_{12}(\cdot)$ being the unconditional copula density of (x_1, x_2). Repeating this representation for $f(x_3|x_1, x_2)$, we get the decomposition

$$f(x_3|x_1, x_2) = c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3) \tag{5}$$

with $c_{23|1}(\cdot)$ being the conditional copula of (x_2, x_3) given x_1. The factorisation of conditional densities can be repeated in an iterative manner yielding a decomposition of the d-dimensional unconditional density in Eq. (3). For dimension $d = 3$, this decomposition is explicitly given by

$$\begin{aligned} f(x_1, x_2, x_3) = {} & c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot c_{12}(F_1(x_1), F_2(x_2)) \\ & \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \end{aligned} \tag{6}$$

[4] This well known problem in copula modeling has been addressed, e.g., by Breymann et al. (2003), Kole et al. (2007), Fischer et al. (2009) and Weiß (2011, 2013).

[5] See Rueschendorf (2013) for a complete formulation of Sklar's theorem.

[6] For a rigorous examination of the statistical properties of vine copulas, see Joe (1996, 1997) and Bedford and Cooke (2001, 2002).

with c_{12}, c_{13} and $c_{23|1}$ as *pair-copulas*. It is important to stress that this representation of the density is only one of several possible ways of decomposing $f(x_1, x_2, x_3)$ using marginal densities and pair-copulas depending on the variables one chooses to condition on.

The different possible decompositions of a d-dimensional joint density can be represented as nested sets of trees where two nodes are joined by an edge in tree $j+1, j = 1, \ldots, d-1$, only if the corresponding edges in tree j share a common node (see Bedford and Cooke, 2001, 2002). Consequently, there are $d-1$ trees, where tree j has $d+1-j$ nodes and $d-j$ edges with each edge corresponding to a pair-copula density, i.e., a density of a conditional bivariate parametric copula. Two well-known classes of such representations are called C- and D-vines with the number of distinct C- and D-vines being $d!/2$. In a Canonical or C-vine, each tree has a unique node (without loss of generality this is node 1) that is connected to all other nodes yielding the representation

$$f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}(F(x_i|x_{i+1}, \ldots, x_{i+j-1}), \times F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})), \tag{7}$$

where the subscript j identifies the tree, while i runs over all edges in each tree. In contrast to a C-vine, no node in any tree T_j is connected to more than two edges in a D-vine yielding the decomposition

$$f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\ldots,j-1}(F(x_j|x_1, \ldots, x_{j-1}), \times F(x_{j+i}|x_1, \ldots, x_{j-1})). \tag{8}$$

In Fig. 1, we illustrate the hierarchical nature of vine copulas by plotting the tree structure of both a C- and a D-vine model.

For the purpose of fitting a vine copula model in Section 2.2, note that the marginal conditional distributions included in the decompositions (7) and (8) can be represented via

$$F(x|\mathbf{v}) = \frac{\partial C_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})}, \tag{9}$$

where $\mathbf{v}$ denotes a d-dimensional vector and $C_{ij|\mathbf{k}}$ is a bivariate copula distribution function. In this context, v_j is an arbitrarily chosen component of the vector v and $\mathbf{v}_{-j}$ describes the vector that excludes this component. If the vector v is univariate, Eq. (9) simplifies to

$$F(x|v) = \frac{\partial C_{xv}(F_x(x), F_v(v))}{\partial F_v(v)}. \tag{10}$$

Later, the estimation of a vine copula model will require the use of so-called h-function which is defined for two uniform variables x and v (i.e., $F_x(x) = x$ and $F_v(v) = v$) as the conditional distribution function of x given v, i.e.,

$$h(x, v, \boldsymbol{\theta}) = F(x|v) = \frac{\partial C_{xv}(x, v)}{\partial v}, \tag{11}$$

with $\boldsymbol{\theta}$ being the set of parameters for the copula of the joint distribution function of x and v. The second parameter of $h(\cdot)$ always corresponds to the conditioning variable.

Pair-copula constructions, and C- and D-vines in particular, have become widely used in empirical applications due to the enormous flexibility of these models. This flexibility is due to the fact that the $d(d-1)/2$ different pair-copulas in Eqs. (7) and (8) can be selected from different parametric copula families. Furthermore, the sheer number of $d!/2$ possible C- and D-vine decompositions enable the statistician to choose from a wide range of flexible models for describing the dependence structure in high-dimensional distributions. Yet, the increased modeling flexibility of pair-copula constructions only comes at the cost of a greatly increased model risk as both the selection of pair-copulas and the choice of a vine decomposition are known to be error-prone. One possibility to alleviate the problem of selecting the best-fitting pair-copulas which we consider in this paper is the use of mixtures of parametric copulas as pair-copulas. In addition to C- and D-vines, recent studies on PCCs have proposed the use of so-called regular vines (R-vines in short) which put less restrictions on the vine trees and which include C- and D-vines as subsets (see Dissmann et al., 2013). Although R-vines are even more flexible than C- and D-vines and are thus becoming increasingly popular, their use is hindered by the same problem of selecting the right parametric pair-copulas one faces when using a C- or D-vine. In the following, we concentrate in our analysis on C- and D-vines to limit the computational complexity of our simulations but note that our results are readily applicable to R-vines as well.

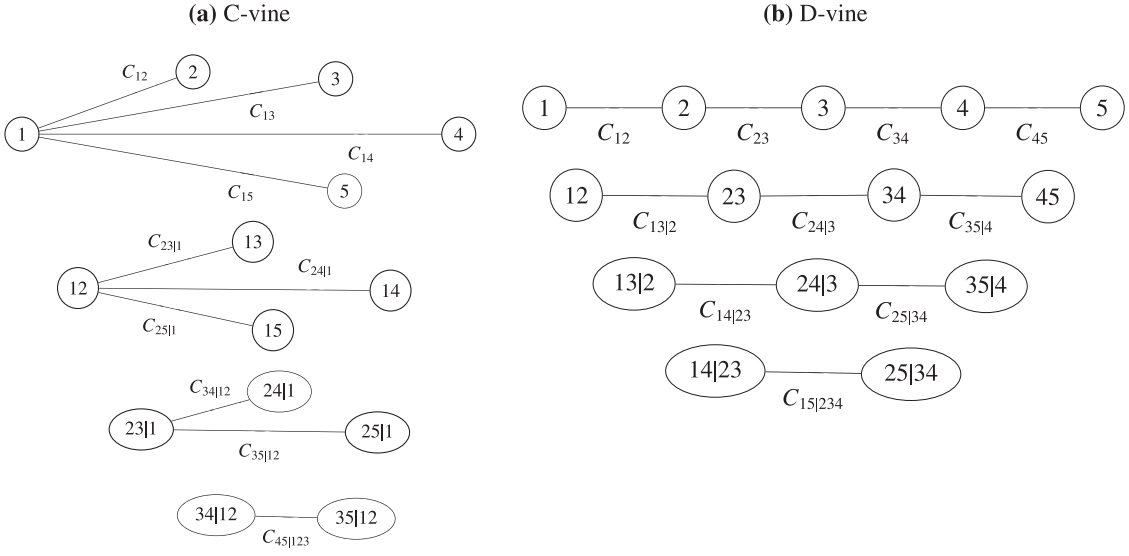

Fig. 1. Five-dimensional C- and D-vine copulas. The figure shows examples of a five-dimensional (a) C-vine and (b) D-vine copula with five random variables, four trees and ten edges. The nodes in the first tree correspond to the five random variables that are being modeled and each edge corresponds to a bivariate conditional or unconditional pair-copula. In both plots, C_{ij} is an unconditional bivariate copula while $C_{ij|x,y,\ldots}$ is the conditional bivariate copula of variables i and j given variables $x, y, \ldots$.

2.2. Mixture copulas

We now turn our attention briefly to the basic definition of a mixture copula, focusing on bivariate copula densities c_i $(i = 1, \ldots, g)$ with g being a fixed number of bivariate mixture copula components. Note that, in many applications of mixture models, not only the parametric forms of the components c_i but also the number g of components itself is unknown and has to be estimated from the available data of the underlying problem.

Let X_r and X_s be two univariate random variables (rvs) (we assume $r, s = 1, \ldots, d, r \neq s$) with continuous cdfs F_r and F_s, respectively. The probability integral transforms of X_r and X_s are given by $U_r = F_r(X_r)$ and $U_s = F_s(X_s)$, respectively. We intend to model the (unique) bivariate copula $C(u_r, u_s)$ of the joint distribution (X_r, X_s) by using a mixture (i.e., a convex combination) of g parametric copulas as components. With $c_1(u_r, u_s; \boldsymbol{\theta}_1), \ldots, c_g(u_r, u_s; \boldsymbol{\theta}_g)$ being the densities of the g parametric copulas, the density of the corresponding *mixture copula* is given by

$$c(u_r, u_s; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i c_i(u_r, u_s; \boldsymbol{\theta_i}), \tag{12}$$

where $\boldsymbol{\theta}_i$ is the vector of unknown parameters for the ith copula component of the mixture and $\boldsymbol{\Psi}$ denotes the vector including all unknown parameters, i.e., $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_g, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g)$.[7] The mixing proportions (or weights) π_i are nonnegative quantities that sum to one:

$$0 \leqslant \pi_i \leqslant 1 \qquad (i = 1, \ldots, g) \tag{13}$$

and

$$\sum_{i=1}^{g} \pi_i = 1. \tag{14}$$

As the components $c_1(u_r, u_s; \boldsymbol{\theta_1}), \ldots, c_g(u_r, u_s; \boldsymbol{\theta_g})$ are copula densities, it easily follows that (12) again defines a bivariate copula density (see Nelsen, 2006). We will refer to c_i as the ith *component copula density* of the mixture.

To illustrate the effect of mixing a second parametric copula to another one (with possibly different tail dependence), we plot simulated samples of size $T = 500$ from several different bivariate mixture copulas in Fig. 2.

The upper left plot in Fig. 2 shows a convex combination of a Clayton and a Student's t copula. Due to the large impact of the Student's t copula on the data, the simulated sample exhibits a significantly symmetric tail dependence structure although the Clayton copula with a weight of 0.2 introduces some lower tail dependence into the data. The upper right plot shows a sample of an equally weighted combination of a Frank copula and a Student's t copula. Here, the plot clearly reveals the symmetric tail dependence structure in the data caused by both mixture constituents. In the lower left plot, the upper tail dependence stemming from the dominant Gumbel copula in the mixture is quite evident while the lower tail dependence in the data caused by the introduction of the Clayton copula into the mixture is far less obvious. Finally, the lower right plot in Fig. 2 presents a mixture of a Student's t copula with a negative correlation parameter and an upper tail dependent Gumbel copula.[8]

The simulations from different bivariate mixture copulas given in Fig. 2 underline the importance to account for diverse tail dependence patterns in applied copula modeling. Conversely, the plots of all four mixture copulas highlight the flexibility with which these models can be used to describe the dependence structure in a given data set. Also, even in the bivariate case, simple parametric copula families could prove inadequate to model dependence structures in real life applications while mixture copulas produce a significantly better model fit. Modeling the pair-copulas in a PCC via convex combinations thus seems to be a sensible approach especially if one keeps in mind that the fit of the vine model critically depends on the fit of all $d(d-1)/2$ pair-copulas that need to be specified in a d-dimensional vine model.

2.3. Incomplete data problem and EM algorithm

When using finite mixture models, the parameter vector $\boldsymbol{\Psi}$ cannot be estimated via classical maximum likelihood estimation (MLE) as the likelihood function is unbounded due to the incomplete structure of the data (see, e.g., Dempster et al., 1977; McLachlan and Peel, 2000). Although some studies in the finance literature like, e.g., Rodriguez (2007) and Ruenzi and Weigert (2013) employ MLE, the identification problem of the mixture model should be taken into account by using the Expectation–Maximization (EM) algorithm for estimating the mixture parameters to avoid otherwise biased parameter estimates. In the following, we shortly restate some well-known facts on the estimation of mixture-copulas using the EM algorithm (for details, see, e.g., Hu, 2006; Li et al., 2011).

Assume that we are given a bivariate data sample $\mathbf{x} = ((x_r^1, x_s^1), \ldots, (x_r^T, x_s^T))$ of size T. A pseudo-sample of the underlying copula is given by

$$\mathbf{u} = (\mathbf{u_r}, \mathbf{u_s}) = ((u_r^1, u_s^1), \ldots, (u_r^T, u_s^T)), \tag{15}$$

where $u_r^t = \widehat{F}_r(x_r^t), u_s^t = \widehat{F}_s(x_s^t)$ $(t = 1, \ldots, T)$, and $\widehat{F}_r, \widehat{F}_s$ are the empirical cdfs of the marginal distributions.[9] The observed data vector $\mathbf{u}$ is treated as being incomplete and the tth observation (u_r^t, u_s^t) is assumed to be drawn from one of the g mixture elements.

Next, we define T component random variables Z_t of dimension g which specify for each observation in the sample the component-copula in (12) from which the observed tuple (u_r^t, u_s^t) $(r \neq s)$ are assumed to have been drawn. The ith element of $Z_t, (Z_t)_i := Z_{it}$ $(i = 1, \ldots, g)$ is an indicator variable that takes on the value 1 whenever the tth observation is drawn from the ith component copula, and 0 otherwise. For a random pseudo-sample $\mathbf{u}$ of a copula, the random vectors $Z_1, \ldots, Z_T$ are assumed to be unconditionally multinomially distributed with probabilities $\pi_1, \ldots, \pi_g$, i.e.,

$$Z_1, \ldots, Z_T \overset{i.i.d.}{\sim} \text{Multinomial}_g\ (1; \pi_1, \ldots, \pi_g), \tag{16}$$

with $z_1, \ldots, z_T$ denoting the associated vector realizations of $Z_1, \ldots, Z_T$. Consequently, the unobservable vector of component-indicator variables

$$\mathbf{z} = (z_1, \ldots, z_T) \tag{17}$$

complements the observed data vector $\mathbf{u}$ to yield the complete data vector

$$\mathbf{u_c} = (\mathbf{u}, \mathbf{z}). \tag{18}$$

Hence, the distribution of the incomplete-data vector $\mathbf{u}$ is included in the distribution of the complete-data vector $\mathbf{u_c}$. Under these conditions, the complete-data log-likelihood for $\boldsymbol{\Psi}$, $\log L_c(\boldsymbol{\Psi})$, can be factorised into the product of the marginal densities of Z_t

[7] Note that the mixture copula itself is given by $C(u_r, u_s; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i C_i(u_r, u_s; \boldsymbol{\theta}_i)$.

[8] Note that, strictly speaking, the parameter ρ of the Student's t copula is not exactly equal to the linear correlation of the data. However, we still refer to it as the "correlation parameter" of the t copula.

[9] The idea to transform the marginal series non-parametrically using the empirical cdfs goes back to Genest et al. (1995) and aims at solving the problem of misspecifying the marginal models. The effects of misspecified marginals on VaR estimates is illustrated by Fantazzini (2009) and Kim et al. (2007) show that estimating copula parameters using pseudo-observations is superior to the use of the Full-Maximum-Likelihood and Inference-for-Margins methods in finite samples.

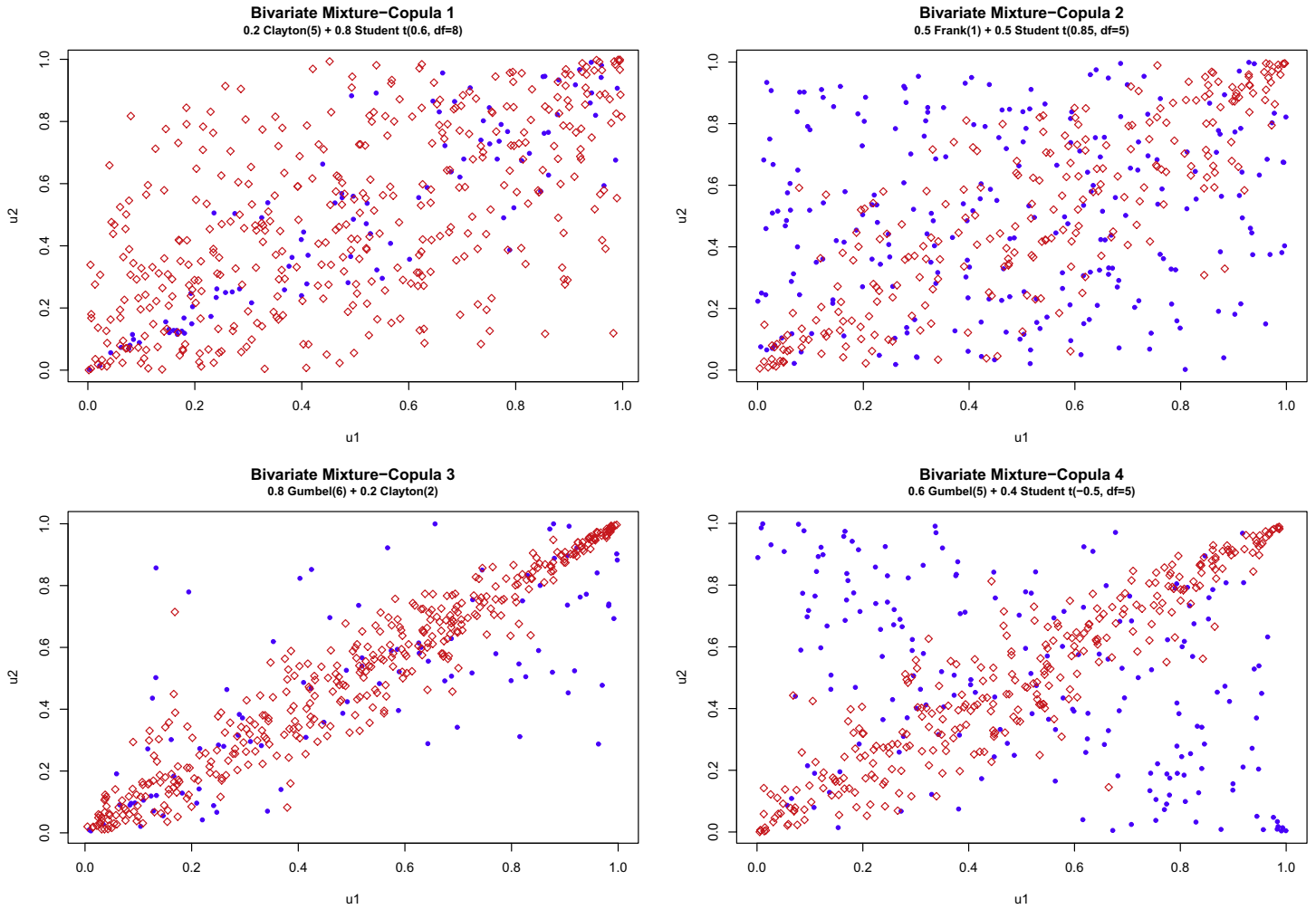

Fig. 2. Scatter plots of different mixture copulas. The figure shows plots of $T = 500$ observations simulated from four different mixture copulas. Panel (1) shows the plot of simulated observations from a mixture of a Clayton copula $(\theta = 5; \pi = 0.2)$ in blue dots and a Student's t copula $(\rho = 0.6; df = 8; 1 - \pi = 0.8)$ in red diamonds. Panel (2) shows the plot of simulated observations from a mixture of a Student's t copula $(\rho = 0.85; df = 5; 1 - \pi = 0.5)$ in red diamonds and a Frank copula $(\theta = 1; \pi = 0.5)$ in blue dots. Panel (3) shows the plot of simulated observations from a mixture of a Gumbel copula $(\theta = 6; \pi = 0.8)$ in red diamonds and a Clayton copula $(\theta = 2; 1 - \pi = 0.2)$ in blue dots. Panel (4) shows the plot of simulated observations from a mixture of a Gumbel copula $(\theta = 5; \pi = 0.6)$ in red diamonds and a Student's t copula $(\rho = -0.5; df = 5; 1 - \pi = 0.4)$ in blue dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and the conditional copula densities of (u_r^t, u_s^t) given the z_t, where the data vectors $(u_r^1, u_s^1), \ldots, (u_r^T, u_s^T)$ are assumed to be conditionally independent given $z_1, \ldots, z_T$:

$$\log L_c(\boldsymbol{\Psi}) := \sum_{i=1}^{g}\sum_{t=1}^{T} z_{it}(\log \pi_i + \log c_i(u_r^t, u_s^t; \boldsymbol{\theta_i})). \tag{19}$$

The EM algorithm is a generic method for computing the parameters in an incomplete-data problem by treating the z_{it} as missing data and has become the most commonly used method for fitting mixture distributions (Dempster et al., 1977). We shortly describe the two steps of each iteration of the EM algorithm, the expectation step (E-step) and the maximization step (M-step), which are successively repeated until convergence.

The E-step handles the addition of the unobservable data, i.e., computing the conditional expectation of the complete-data log-likelihood $L_c(\boldsymbol{\Psi})$, given the observed pair $(\mathbf{u_r}, \mathbf{u_s})$ and using the current estimate for $\boldsymbol{\Psi}$. Starting with an initial parameter vector $\boldsymbol{\Psi}^{(0)}$, the calculation of the conditional expectation of $L_c(\boldsymbol{\Psi})$ can be written as follows:

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(0)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(0)}}\{\log L_c(\boldsymbol{\Psi})|(\mathbf{u_r}, \mathbf{u_s})\}. \tag{20}$$

Consequently, on the $(m+1)$th iteration, the E-step requires the computation of $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(m)})$, where $\boldsymbol{\Psi}^{(m)}$ denotes the estimate of $\boldsymbol{\Psi}$ after the mth EM-iteration. Due to the fact that $\log L_c(\boldsymbol{\Psi})$ is linear in z_{it}, the computation of the conditional expectation of (19) given $\mathbf{u}$ on the $(m+1)$th iteration leads to the calculation of the conditional expectation of Z_{it}:

$$\mathbb{E}_{\boldsymbol{\Psi}^{(m)}}[Z_{it}|(\mathbf{u_r}, \mathbf{u_s})] = P_{\boldsymbol{\Psi}^{(m)}}[Z_{it} = 1|(\mathbf{u_r}, \mathbf{u_s})] \tag{21}$$

$$= \frac{\pi_i^{(m)} c_i(u_r^t, u_s^t; \boldsymbol{\theta_i}^{(m)})}{\sum_{h=1}^{g} \pi_h^{(m)} c_h(u_r^t, u_s^t; \boldsymbol{\theta_h}^{(m)})} \tag{22}$$

$$=: \tau_i((u_r^t, u_s^t); \boldsymbol{\Psi}^{(m)}). \tag{23}$$

While the mixing proportion π_i can be regarded as the prior probability that the tuple (u_r^t, u_s^t) corresponds to the ith mixture weight, the quantity $\tau_i((u_r^t, u_s^t)$ can be seen as the posterior probability that the observed data (u_r^t, u_s^t) belongs to the ith component of the mixture copula. Thus, using Eq. (22), the computation of the conditional expectation (19) is then equivalent to

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(m)}) = \sum_{i=1}^{g}\sum_{t=1}^{T} \tau_i((u_r^t, u_s^t); \boldsymbol{\Psi}^{(m)})(\log \pi_i + \log c_i(u_r^t, u_s^t; \boldsymbol{\theta_i})). \tag{24}$$

The M-step on the $(m+1)$th iteration calculates the updated estimate $\boldsymbol{\Psi}^{(m+1)}$. This is achieved by maximizing $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(m)})$ with respect to $\boldsymbol{\Psi}$. Note that the computation of $\boldsymbol{\Psi}^{(m+1)}$ is decomposed into two parts, because the updated mixing proportions $\pi_i^{(m+1)}$ are independent of the updated estimates $\boldsymbol{\theta}^{m+1} = (\boldsymbol{\theta_1}^{m+1}, \ldots, \boldsymbol{\theta_g}^{m+1})$. Using the current conditional expectation $\tau_i((u_r^t, u_s^t); \boldsymbol{\Psi}^{(m)})$, the updated estimate of π_i is given by

$$\pi_i^{(m+1)} = \sum_{t=1}^{T} \tau_i(u_r^t, u_s^t; \boldsymbol{\Psi}^{(m)})/T \qquad (i = 1, \ldots, g). \tag{25}$$

Observe that each observation u_r^t, u_s^t affects the ith mixture component $\pi_i^{(m+1)}$ via the current estimate of its posterior probability. The update of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta_g})$ of the $(m+1)$th iteration is directly given by (24), so that $\boldsymbol{\theta}^{(m+1)}$ must satisfy the following equation:

$$\sum_{i=1}^{g}\sum_{t=1}^{T} \tau_i(u_r^t, u_s^t; \boldsymbol{\Psi}^{(m)}) \frac{\partial \log c_i(u_r^t, u_s^t; \boldsymbol{\theta_i})}{\partial \boldsymbol{\theta}} = 0. \tag{26}$$

Dempster et al. (1977) show that the incomplete-data likelihood function $L(\boldsymbol{\Psi})$ does not decrease after an EM iteration, i.e.,

$$L(\boldsymbol{\Psi}^{(m+1)}) \geqslant L(\boldsymbol{\Psi}^{(m)}). \tag{27}$$

In addition to $L(\boldsymbol{\Psi})$ being non-decreasing in successive EM iterations, an upper bound for the likelihood assures convergence of the EM algorithm.

2.4. Mixture-pair-copula constructions

In this part of our analysis, we combine both concepts to yield so-called mixture-pair-copula constructions. The basic idea behind Mixture-PCCs is to use mixtures of parametric copulas as pair-copulas in a PCC. Obviously, the g components of the mixture-pair-copulas do not necessarily have to be identical throughout the vine, nor does the number of components g itself. In case both the number and the parametric form of the component copulas are fixed, we will refer to this as a *simple Mixture-PCC*. If both the number and the parametric forms are not prespecified but rather estimated prior to the EM algorithm, we will call such a model an *extended Mixture-PCC*. It is clear that an extended Mixture-PCC offers even more flexibility for modeling dependence structures than a simple Mixture-PCC as it imposes less restrictions on the parametric form of the mixture-pair-copulas. However, its specification requires the additional selection of both g and the component copulas separately for all $d(d-1)/2$ pair-copulas of the Mixture-PCC. Model selection for mixture models can be done by using a criterion for the fit of the model like, e.g., Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC) or the Consistent AIC (CAIC) (see Kim et al., 2013; McLachlan and Peel, 2000).

Due to the additional computational complexity of estimating an extended Mixture-PCC, we concentrate in the rest of our study on the analysis of simple Mixture-PCCs and their performance in Monte Carlo simulations and an empirical application. Nevertheless, we note that the results for extended Mixture-PCCs should be identical or even better than the results presented in the following based on simple Mixture-PCCs.[10]

Next, we demonstrate how to fit a Mixture-PCC model to a given multivariate data set as well as a known vine type with a given tree structure (i.e., the permutation of the data variables). Note that usually not only the bivariate copulas and the corresponding parameters but also the tree structure need to be estimated.[11] In our paper, we focus on techniques for fitting the building blocks (bivariate copulas) in the vine model. Thus, the vine type and the tree structure are assumed to be known.

When modeling the dependence structure of a multivariate data set by the use of a Mixture-PCC, two major problems have to be addressed. First, the true parametric families of the bivariate mixture copula densities describing the dependence structure are unknown. Hence, for a d-dimensional vine copula, $d(d-1)/2$ suitable bivariate mixture copulas must be selected. Second, we require a tool for estimating the weights as well as the parameters of the selected copula families for each mixture pair-copula. Since we are focusing on simple Mixture-PCCs, the number g and the parametric form of each bivariate component copula is prespecified, i.e., the first problem is not taken into further consideration in our study. The second task is handled by using the EM algorithm sequentially. This method fits each mixture pair-copula (density) separately by estimating $\boldsymbol{\Psi}$, the vector including the mixing weights $\Pi_1, \ldots, \Pi_g$ as well as the vectors of the unknown copula parameters $\theta_1, \ldots, \theta_g$, in Eq. (12). More precisely, we propose the following modified version of

[10] Selecting an extended Mixture-PCC will usually require the estimation of a model selection criterion like, e.g., Akaike's Information Criterion for several Mixture-PCC specifications. As such, the simple Mixture-PCC could be used as a benchmark which could be improved upon by estimating further extended Mixture-PCCs.

[11] See Brechmann and Schepsmeier (2013) for a discussion of the problem of selecting the optimal tree structure of a vine copula model.

the sequential estimation procedure outlined in Brechmann and Czado (2013) and Dissmann et al. (2013) to fully specify a d-dimensional Mixture-PCC:

Algorithm 1. (Fitting a d-dimensional Mixture-PCC sequentially)

For tree $j = 1, \dots, d-1$ in Eqs. (7) (C-vine) or (8) (D-vine), perform the following steps:

1. For each edge i $(i = 1, \dots, d-j)$ (bivariate mixture copula) of tree T_j of the vine, the parameter vector $\boldsymbol{\Psi}$ of each mixture pair-copula is estimated using the original data (tree T_1) corresponding to the variables i and $i+j$ for a C-vine or j and $j+i$ for a D-vine, or the transformed data (estimated in Step 2) from the previous tree T_{j-1} via the EM algorithm.
2. Transform the observations used in this iteration by using the estimated mixture copula from tree T_j and the h-function defined in Eq. (11) to compute the observations (i.e., conditional distribution functions) of $(F(x_i|x_{i+1},\dots,x_{i+j-1}), F(x_{i+j}|x_{i+1},\dots,x_{i+j-1}))$ (C-vine) or $(F(x_j|x_1,\dots,x_{j-1}), F(x_{j+i}|x_1,\dots,x_{j-1}))$ (D-vine) for the next tree. If $j=1$, use the original data to calculate the unconditional distribution functions $(F(x_i),F(x_{i+j}))$ (C-vine) or $(F(x_j),F(x_{j+i}))$ (D-vine), respectively.

The algorithm highlights the sequential manner of the procedure by exploiting the tree-by-tree structure of vine copulas. In the first tree, the parameter vector $\boldsymbol{\Psi}$ of each bivariate mixture copula is estimated using the original data via the EM algorithm. The transformed variables for the second tree are computed subsequently using the mixture's h-function (11) as shown in Section 2.1.[12] With these transformed data, each bivariate mixture copula of the second tree is estimated. This process is repeated until the last tree is reached and the final mixture pair-copula is estimated.

Due to the sequential estimation procedure, our mixture approach does not guarantee to find a global optimum with respect to the accuracy of the weights as well as the parameters of each mixture pair-copula. Performing a full estimation over all parameters simultaneously, might provide a better model fit. Nevertheless, our sequential mixture procedure is nevertheless a beneficial approach as it considerably limits the computational burden of the estimations.

Admittedly, modeling the dependence structure by using Mixture-PCCs goes hand in hand with additional cost of computation. In a simple Mixture-PCC with one-parameter copula components, $g-1$ mixture weights $\pi_1, \dots, \pi_{g-1}$ and g copula parameters $\theta_1, \dots, \theta_g$ have to be estimated for each building block in the vine model.[13] Thus, for a fully specified vine model, we have $d(d-1)/2 \cdot (2g-1)$ parameters in total that need to be estimated. However, all estimations are performed sequentially on bivariate data. Consequently, our model is still tractable even despite the increased number of parameters that need to be estimated and the gain in flexibility prevails the additional cost of computation.

3. Simulation study

In this section, we compare the performance of our proposed Mixture-PCC with that of the sequential heuristic of Brechmann et al. (2012) and Dissmann et al. (2013) for selecting the pair-copulas in a vine from prespecified parametric copula families. In our simulations, we mimic the frequent problem in financial applications of fitting a multivariate distribution to a time series of asset returns and subsequently estimating the Value-at-Risk of a portfolio consisting of these assets. More precisely, we are interested in assessing the quality of both model approaches by backtesting the vine copula models' VaR-estimates in the in-sample of our simulated data.

3.1. Simulation design

In the following, we shortly describe the data generating process (DGP) used in our simulation study. To simulate d asset price trajectories over a prespecified number of observations T, we employ standard GARCH(1,1) models to mimic frequently observed stylized facts in asset returns like volatility clustering, heavy tails, and serial correlation. Let z_{tj} $(t = 1, \dots, T; j = 1, \dots, d)$ denote an independent, identically distributed sample of a multivariate distribution with properly scaled Student's t marginals and a random vine copula describing the dependence structure.[14] Then, the log-return time series is modeled by $\epsilon_{tj} = \sigma_{tj} z_{tj}$. The dynamic variance is characterized by the equation $\sigma^2_{tj} = \alpha_{0j} + \alpha_{1j}\epsilon^2_{t-1j} + \beta_j \sigma^2_{t-1j}$, where all GARCH parameters are also chosen randomly within their respective domain of definition. In this context, α_{1j} denotes the first lag ARCH parameter (state memory factor) and β_j the first GARCH parameter (variance memory factor) for the log-return time series of asset j, respectively. Thus, $\alpha_{0j}/(1-\alpha_{1j}-\beta_j)$ is the unconditional variance for time-series j.

The randomly calibrated vine used as the DGP in each simulation run $l = 1, \dots, 100$ is referred to as $PCC^{(l)}_{DGP}$. The simulated sample from $PCC^{(l)}_{DGP}$ is then used to fit a benchmark vine model ($\widehat{PCC}^{(l)}_{AIC}$) via the sequential heuristic proposed by Brechmann and Czado (2013) and Dissmann et al. (2013) and a Mixture-PCC ($\widehat{PCC}^{(l)}_{Mix}$) using Algorithm 1. Based on the (known) DGP, the fit of both approximations is then evaluated by in-sample forecasting.

3.2. Estimation of PCCs

Next, we discuss the multivariate model that is fitted to the simulated data. As financial return data is frequently characterized by conditional heteroskedasticity and asymmetric dependence, we resort to GARCH-type models both as data generating processes as well as models that are fitted to the data. Therefore, we employ GARCH(1,1) models with Student's t-distributed innovations (see Jondeau and Rockinger, 2006; Fantazzini, 2009; Liu and Luger, 2009; Aas and Berg, 2009; Ausín and Lopes, 2010; Hafner and Reznikova, 2010) to describe the time series' marginal behavior.[15]

Then, let P_t $(t = 0, 1, \dots, T)$ denote the price of a financial asset at time t. The asset's log return R_t is defined by $R_t := log(P_t/P_{t-1})$ $(t \geqslant 1)$. In our simulation study, R_t is simulated and results from our data generating process whereas in the empirical study, real

[12] Note that the h-function of the mixture copula (distribution function) is given by the convex combination of the h-functions of the mixture components (i.e., the copula cdfs). The numerical computation of the mixture's h-function can then be achieved quite easily by (explicitly or numerically) solving the h-functions of the mixture components.

[13] Since the mixing weights sum to unity, one of them is redundant. Here, we arbitrarily omitted the gth mixing weight π_g.

[14] In this context, a random vine is given by a vine in which the vine type, the ordering of the variables, the parametric copula families for the pair-copulas, and the parameters are all chosen randomly in each simulation run. As candidate parametric copula families from which the pair-copulas of the true vine models are chosen, we use the Gaussian, Student's t, Clayton, Gumbel, Survival Clayton, Survival Gumbel, the rotated Clayton copula (90 degrees) and the rotated Gumbel copula (90 degrees). The parameters of the pair-copulas are then chosen randomly within the domain of the respective copula's parameters.

[15] Hansen and Lunde (2005) compare different versions of the original GARCH model and show that the majority of GARCH model variants are outperformed by the simple GARCH(1,1) specification. Consequently, the GARCH(1,1) time series filter with Student's t-distributed innovations is the model of choice for the marginal behavior of our data.

prices of financial assets are used. The return time series are then modeled using

$$R_{t,j} = \mu_j + \sigma_{t,j} Z_{t,j} \tag{28}$$

$$\sigma^2_{t,j} = \alpha_{0,j} + \alpha_{1,j}\left(R_{t-1,j} - \mu_j\right)^2 + \beta_j \sigma^2_{t-1,j}, \quad j = 1,\dots,d; t = 1,\dots,T \tag{29}$$

where Z_{tj} are independent and identically t-distributed innovations $Z_{t,j}$.

The vector $Z_t = (Z_{t,1},\dots,Z_{t,d})$ $(t = 1,\dots,T)$ of the GARCH innovations captures the dependence structure between the d financial assets and is assumed to be jointly distributed with distribution

$$F_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{v}_1,\dots,\boldsymbol{v}_d, \boldsymbol{\Psi} | \mathcal{F}_{t-1}) = C^{Mix}[F_1(z_1; \boldsymbol{v}_1|\mathcal{F}_{t-1}),\dots,F_d(z_d; \boldsymbol{v}_d|\mathcal{F}_{t-1}); \boldsymbol{\Psi}], \tag{30}$$

where $\boldsymbol{\Psi}$ denotes the parameter vector of a d-dimensional copula C and $v_1\dots,v_d$ are the parameter vectors of the innovations. The parameters of the univariate GARCH-models are estimated via quasi-maximum likelihood. Using these parameter estimates, we convert the original observations into standardized residuals which in turn are transformed into a pseudo-sample using Eq. (15). The pseudo-observations are then used to fit the Mixture-PCC and the heuristic benchmark-PCC. As a benchmark model, we calibrate a PCC using the sequential heuristic proposed by Brechmann and Czado (2013) and Dissmann et al. (2013). Here, for each of the $d(d-1)/2$ pair copulas, all candidate parametric copulas are fitted to the data using maximum likelihood estimation sequentially and the candidate that yields the optimal value of Akaike's Information Criterion (AIC) is chosen.

After both PCCs have been fitted, for each day in the in-sample, $K = 10,000$ observations $u^{(k)}_{T+1,1},\dots,u^{(k)}_{T+1,d}$ $(k = 1,\dots,K)$ from the fitted parametric vine copula are simulated. This sample is then converted into observations $z^{(k)}_{T+1,1},\dots,z^{(k)}_{T+1,d}$ from the distributions of the innovations via the inverse of the symmetric Student t-distributions. Then $\hat{\mu}_j + \hat{\sigma}_{T+1,j} z^{(k)}_{T+1,j}$ $(j = 1,\dots,d)$ transforms the simulated observations into simulated returns, where $\hat{\sigma}_{T+1,j}$ and $\hat{\mu}_j$ are the conditional volatility forecasts and mean values stemming from the fitted marginal GARCH(1,1) models. The portfolio return is then easily calculated via $R^{(k)}_{T+1,pf} = d^{-1}\sum_{j=1}^{d} R^{(k)}_{T+1,j}$ (assuming an equal-weighted portfolio) and an estimate for the portfolio's α%-VaR can be calculated by ranking the K simulated returns. Finally, VaR-violations can be identified by comparing the α%-VaR with the observed return of $R_{T+1,pf}$.

In the simulation study, we consider two different sample sizes $T = 300$, and $T = 500$ to assess the effect of different sample sizes on the fit of the pair-copula constructions. Furthermore, we repeat our simulations for two dimensions $d \in \{3;5\}$. We expect the overall fit of the PCCs to suffer at least to a small degree with increasing dimensionality. As the number of pair-copulas and trees in a vine model increases with the dimension of the data sample, rounding and estimation errors in the upper trees of a vine could be propagated thereby decreasing the fit of pair-copulas in the lower trees. Consequently, Aas et al. (2009), Brechmann et al. (2012) and Dissmann et al. (2013) propose to capture as much dependence of the joint distribution in the upper trees of a vine model so that the remaining pair-copulas in the lower trees can be truncated or simplified. Finally, we repeat all simulations for two different vine types (C- and D-vine) to test the robustness of our results to a change in vine type. See Section 3.1 for the candidate parametric copula families from which the pair-copulas of the true vine models are chosen.

For each sample size, dimension, and vine type, we simulate 100 random samples and fit to each sample a simple Mixture-PCC and a heuristically calibrated PCC following the procedure laid out in Brechmann et al. (2012). In the Mixture-PCCs, each pair-copula is modeled as a convex combination of a Clayton, a Frank, a Gumbel and a Student's t copula.[16] We chose these particular parametric copula families as constituents of the mixture pair-copulas because they account for both dependence and independence in the tails of the distribution. The Mixture-PCC is then fitted using the EM algorithm. For the benchmark approach, we use a widely extended set of candidate parametric copula families. On the one hand, the full range of parametric copula families used in the simulation design in Section 3.1 is taken into account. On the other hand, the Frank, Joe, BB1, BB6, BB7, BB8 and again rotated versions of these are provided additionally. Each candidate parametric pair-copula is fitted to the data via maximum likelihood estimation. Then, the candidate that yields the optimal AIC value is chosen. Using the fitted Mixture- and benchmark PCCs, we then estimate the 0.1%, 2.5%, and 5% Value-at-Risk of the simulated data and assess both models' fit by performing the two-tailed backtest of Christoffersen (1998) and the duration-based Weibull test of independent VaR-exceedances proposed by Christoffersen and Pelletier (2004) on the resulting sequences of VaR-violations. The outline of our simulation study can be summarized as follows:

1. For each sample size $(T = 300; T = 500)$, dimension $(d = 3, d = 5)$, and vine type (C- or D-vine) generate $l = 1,\dots,100$ random samples of size T from a randomly calibrated vine $PCC^{(l)}_{DGP}$.
2. Fit a benchmark vine model ($\widehat{PCC}^{(l)}_{AIC}$) via the sequential heuristic and a Mixture-PCC ($\widehat{PCC}^{(l)}_{Mix}$) to the data sample assuming the correct vine type to be known.
3. For each day in the in-sample of size T, simulate 10,000 observations both from $\widehat{PCC}^{(l)}_{AIC}$ and $\widehat{PCC}^{(l)}_{Mix}$.
4. Compute VaR-estimates for both vine models using the 10,000 simulated returns. Compare the estimated VaR of both models with the return in the original sample in Step 1.

The results of the simulations are presented in Table 1.

Table 1 presents the average number of expected and realized VaR-exceedances as well as the percentage of simulations in which the VaR-backtest was rejected at the 5% significance level for both our Mixture-PCC and the benchmark PCC.

Starting with the results for a data generating process of dimension $d = 3$ in Panel (a), several findings are noteworthy. First, both models yield similar results for the smallest VaR-level of 0.1% with almost no VaR-violations and a perfect 0% of simulation runs in which neither model is rejected by the backtest. This result holds for both sample sizes and both C- and D-vines. Second, results start to differ from model to model for VaR-levels of 2.5% and 5%. While both models appear to overestimate portfolio risk (as evidenced by the average number of violations which is always below the expected number of VaR-violations), our proposed Mixture-PCCs perform significantly better than the benchmark for a VaR-level of 5% as shown by the smaller percentages of simulations in which the backtest is rejected. For a VaR-level of 2.5%, the rejection percentages of the Mixture-PCCs are comparable or slightly worse than the corresponding averages of the benchmark. Third, the results for both models do not differ significantly across the vine types and sample sizes although the percentages of the backtest rejections increase slightly for the 2.5% VaR with an increase in sample size.

[16] To limit the computational cost of the parameter estimations, we restrict the mixture copulas in the Mixture-PCCs to four parametric copula constituents. Note that while the estimation of the parameters of the benchmark PCC required about ten minutes in total for all 100 simulations, the estimation of our Mixture-PCCs required about five times longer.

Table 1
Results of the simulation study – long position. The table presents the results of the simulation study on the in-sample fit of our proposed Mixture-PCC and a heuristically calibrated PCC. The data generating process is given by a multivariate GARCH model in which the marginal distributions follow GARCH(1,1) processes with Student's t distributed innovations and where the joint distribution of the innovations is characterized by a randomly calibrated vine copula (see Nikoloulopoulos et al., 2012, for details). The in-sample fit of both competing models is assessed by performing the two-tailed backtest of conditional coverage proposed in Christoffersen (1998) (CC) as well as the duration-based Weibull test of independence by Christoffersen and Pelletier (2004) (WB) on the in-sample estimates of the portfolio VaR. The table reports the expected and the realized number of VaR-exceedances as well as the rejection rates for a significance level for the backtest of 5%. The VaR-level α is set to 0.1%, 2.5%, and 5%, respectively. Results are presented separately for sample sizes $T = 300$ and $T = 500$ and for dimension $d = 3$ and $d = 5$, respectively.

Pair-copula construction	Sample size	VaR-level (%)	Exceedances (expected)	Sequential AIC			Mixture-PCC		
				P-value < 5% (No. of rejections)		∅-Exceedances (realized)	*P*-value < 5% (No. of rejections)		∅-Exceedances (realized)
				CC	WB		CC	WB	
Panel (a): Dimension d = 3									
C-Vine	$T = 300$	0.1	0.3	0.00	0.00	0.00	0.00	0.00	0.00
		2.5	7.5	0.36	0.18	3.23	0.39	0.18	2.97
		5	15	0.31	0.07	9.19	0.09	0.12	11.58
	$T = 500$	0.1	0.5	0.00	0.00	0.02	0.00	0.00	0.00
		2.5	12.5	0.47	0.09	5.96	0.59	0.07	5.41
		5	25	0.44	0.06	16.01	0.06	0.09	20.36
D-Vine	$T = 300$	0.1	0.3	0.00	0.00	0.02	0.00	0.00	0.00
		2.5	7.5	0.40	0.15	3.44	0.41	0.16	3.46
		5	15	0.23	0.05	9.53	0.06	0.03	13.45
	$T = 500$	0.1	0.5	0.00	0.00	0.06	0.00	0.00	0.00
		2.5	12.5	0.48	0.02	5.83	0.57	0.02	5.59
		5	25	0.45	0.03	15.35	0.05	0.11	21.44
Panel (b): Dimension d = 5									
C-Vine	$T = 300$	0.1	0.3	0.00	0.00	0.02	0.00	0.00	0.00
		2.5	7.5	0.36	0.22	3.32	0.28	0.17	3.54
		5	15	0.23	0.07	9.66	0.01	0.08	14.30
	$T = 500$	0.1	0.5	0.00	0.00	0.04	0.00	0.00	0.00
		2.5	12.5	0.52	0.07	5.69	0.52	0.05	5.97
		5	25	0.46	0.10	15.32	0.05	0.09	22.37
D-Vine	$T = 300$	0.1	0.3	0.00	0.00	0.03	0.00	0.00	0.00
		2.5	7.5	0.26	0.15	3.58	0.17	0.11	4.17
		5	15	0.35	0.06	9.13	0.04	0.09	14.61
	$T = 500$	0.1	0.5	0.00	0.00	0.05	0.00	0.00	0.00
		2.5	12.5	0.51	0.06	5.65	0.31	0.04	6.59
		5	25	0.44	0.09	15.36	0.03	0.10	24.45

Panel (b) presents the corresponding results for the simulations of data samples of dimension $d = 5$. Regarding the 0.1%-VaR-level, both approaches yield acceptable model fits with both models having a rejection percentage of zero in backtesting. Again, we find both approaches to be conservative and to overestimate portfolio risk to a certain degree as shown by the average number of VaR-violations. In contrast to the simulations for dimension $d = 3$, however, the results of the simulations for $d = 5$ clearly show that the benchmark PCC is outperformed by our Mixture-PCC. For all three VaR-levels, the Mixture-PCC is rejected in approximately the same or a smaller number of simulation runs than the benchmark PCC. For example, the in-sample VaR-estimates of the Mixture-PCC are only rejected in 5% (D-vine) and 3% (C-vine) of all simulations for a sample size of $T = 500$. In contrast, the benchmark PCC is rejected in 46% and 44% of all simulation runs, respectively. Moreover, the average number of violations produced by the Mixture-PCC is significantly closer to the expected number of violations. Turning to the analysis of the duration-based Weibull backtest of independent VaR-exceedances, we find similar results as for the test of conditional coverage with the exception that all estimated models (benchmark and Mixture-PCC) are rejected in far less simulations by the Weibull test than by the CC test. Table 2 shows corresponding results for the VaR-levels of 95%, 97.5%, and 99.9% for a short position in the portfolio.

Again, based on the results of the average number of exceedances and the results of the CC test, we find our Mixture-PCC to outperform the benchmark not only for a long but also for a short position in the portfolio.[17]

The results emphasize that both models yield acceptable Value-at-Risk forecasts. However, performing the two-tailed backtest of Christoffersen (1998) indicates a more appropriate model fit of our mixture approach for both long and short positions in the portfolio. This result is remarkable for the following reason. Since no mixtures are used to simulate the data, the Benchmark-PCC should in principle be able to perfectly recover the complete tree structure. However, due to the finite sample size, the Benchmark-PCC selects several pair-copulas incorrectly on average in our simulations. In contrast, our Mixture-PCCs appear to be significantly more flexible and to provide a better model fit even in this setting. As a consequence, our simulations provide us with ample evidence that especially for higher dimensions, the Mixture-PCC model yields significantly better in-sample VaR-estimates than the benchmark.

4. Empirical study

The results of our simulation study show that our proposed Mixture-PCCs perform exceptionally well in-sample especially for higher dimensions. In this section, we support our main finding by performing an empirical study on the out-of-sample forecasting accuracy of our Mixture-PCC and the heuristic benchmark.

[17] The results of the Weibull backtest for the short position in the portfolio are mixed and do not allow for a clear interpretation.

Table 2
Results of the simulation study – short position. The table presents the results of the simulation study on the in-sample fit of our proposed Mixture-PCC and a heuristically calibrated PCC. The data generating process is given by a multivariate GARCH model in which the marginal distributions follow GARCH(1,1) processes with Student's t distributed innovations and where the joint distribution of the innovations is characterized by a randomly calibrated vine copula (see Nikoloulopoulos et al., 2012, for details). The in-sample fit of both competing models is assessed by performing the two-tailed backtest of conditional coverage proposed in Christoffersen (1998) (CC) as well as the duration-based Weibull test of independence by Christoffersen and Pelletier (2004) (WB) on the in-sample estimates of the portfolio VaR. The table reports the expected and the realized number of VaR-exceedances as well as the rejection rates for a significance level for the backtest of 5%. The VaR-level α is set to 95%, 97.5% and 99.9%, respectively. Exceedances are given under the assumption of a short position in the portfolio. Results are presented separately for sample sizes $T = 300$ and $T = 500$ and for dimension $d = 3$ and $d = 5$, respectively.

Pair-copula construction	Sample size	VaR-level (%)	Exceedances (expected)	Sequential AIC			Mixture-PCC		
				P-value < 5% (No. of rejections)		∅-Exceedances (realized)	P-value < 5% (No. of rejections)		∅-Exceedances (realized)
				CC	WB		CC	WB	
Panel (a): Dimension d = 3									
C-Vine	$T = 300$	99.9	0.3	0.00	0.00	0.02	0.00	0.00	0.01
		97.5	7.5	0.42	0.16	3.18	0.32	0.15	3.63
		95	15	0.44	0.07	8.29	0.11	0.06	11.27
	$T = 500$	99.9	0.5	0.00	0.01	0.09	0.00	0.00	0.02
		97.5	12.5	0.40	0.15	6.04	0.33	0.11	6.67
		95	25	0.52	0.08	14.80	0.13	0.11	19.55
D-Vine	$T = 300$	99.9	0.3	0.01	0.00	0.07	0.00	0.00	0.07
		97.5	7.5	0.47	0.17	3.21	0.56	0.18	2.95
		95	15	0.45	0.13	8.44	0.29	0.07	9.41
	$T = 500$	99.9	0.5	0.00	0.00	0.04	0.01	0.01	0.16
		97.5	12.5	0.54	0.08	5.38	0.67	0.18	4.63
		95	25	0.74	0.03	13.71	0.47	0.06	15.53
Panel (b): Dimension d = 5									
C-Vine	$T = 300$	99.9	0.3	0.00	0.00	0.05	0.00	0.00	0.02
		97.5	7.5	0.38	0.16	3.17	0.16	0.11	4.46
		95	15	0.45	0.07	7.96	0.04	0.11	13.39
	$T = 500$	99.9	0.5	0.00	0.00	0.08	0.00	0.00	0.03
		97.5	12.5	0.62	0.07	5.13	0.19	0.09	7.29
		95	25	0.66	0.10	13.75	0.05	0.09	23.03
D-Vine	$T = 300$	99.9	0.3	0.00	0.00	0.04	0.01	0.00	0.21
		97.5	7.5	0.39	0.11	3.00	0.15	0.09	4.45
		95	15	0.41	0.09	8.11	0.07	0.06	13.10
	$T = 500$	99.9	0.5	0.00	0.00	0.07	0.01	0.02	0.24
		97.5	12.5	0.61	0.07	5.18	0.27	0.08	7.28
		95	25	0.69	0.10	13.78	0.10	0.06	21.71

4.1. Methodology and data

In our empirical study, we assess the accuracy of both the Mixture-PCC and the heuristically calibrated benchmark-PCC in out-of-sample portfolio-VaR forecasting. To this end, we analyze a four-dimensional portfolio consisting of three selected stocks and the price of Gold bullions. To be precise, we study the time series of daily log-returns $R_{t,pf} = 4^{-1}\sum_{j=1}^{4} R_{t,j}$ of an equal-weighted four-dimensional portfolio and forecast the Value-at-Risk of this portfolio. The multivariate GARCH model that we fit to the data resembles the one proposed by Nikoloulopoulos et al. (2012) and which we also employ in the simulation study (see Section 3). In contrast to our simulation study, however, we are now interested in forecasting the portfolio's Value-at-Risk out-of-sample.[18]

Our example portfolio consists of the stocks of Citigroup, General Electric, and Deutsche Bank, as well as the price for gold bullion LBM. We selected these four investments for our example portfolio as to cover a diverse set of asset classes (gold and stocks), industries (banks and industrial conglomerate) as well as regions (Europe and United States). While we expect the stocks of the two banks in our portfolio to experience strong comovements in stock returns, the addition of the stock of General Electric and gold bullion should allow for ample diversification effects. Moreover, we opted for a four-dimensional portfolio which suffices to exemplify the advantages of our proposed modeling approach while at the same time limiting the computational workload. Time-series data for the four assets are retrieved from *Thomson Reuters Financial Datastream*. To account for known data errors in *Datastream*, we follow Ince and Porter (2006) and apply the following data filters to our sample. First, we control for extreme log returns above 300% that are reversed within one month and exclude these returns from our data. Second, we exclude returns of prices below $1 to avoid incorrect log returns that could otherwise arise from *Datastream*'s practice of rounding prices. Neither of our four price time series exhibits any of these data errors.

The dependence is strongest between gold bullion and the individual stocks. Thus, we believe it is appropriate to model the empirical data using a canonical vine model with gold bullion as the so-called *root node*. Consequently, the pair-copulas in the first tree of the vine model the bivariate distributions of gold bullion-Citigroup, gold bullion-Deutsche Bank, and gold bullion-General Electric. In the benchmark model, we use the same set of candidate parametric copula families as in our simulation study presented in Section 3.1. Again, in our mixture approach, each pair-copula is modeled as a convex combination of a Clayton, a Frank, a Gumbel, and a Student's t copula.[19] Thus, both models allow for a wide range of possible dependence structures.

The data for our empirical study covers the time period from January 1, 2009 to October 31, 2012. Our sample period thus starts

[18] See Weiß (2013) for details of the extension of the model of Nikoloulopoulos et al. (2012) to out-of-sample forecasting.

[19] Note that the pre-specified set of parametric copulas from which the mixture pair-copulas are built can easily be extended to more than four candidate parametric copulas. Preliminary tests in our study, however, showed that the four parametric copulas we use suffice to model the dependence structure in the data.

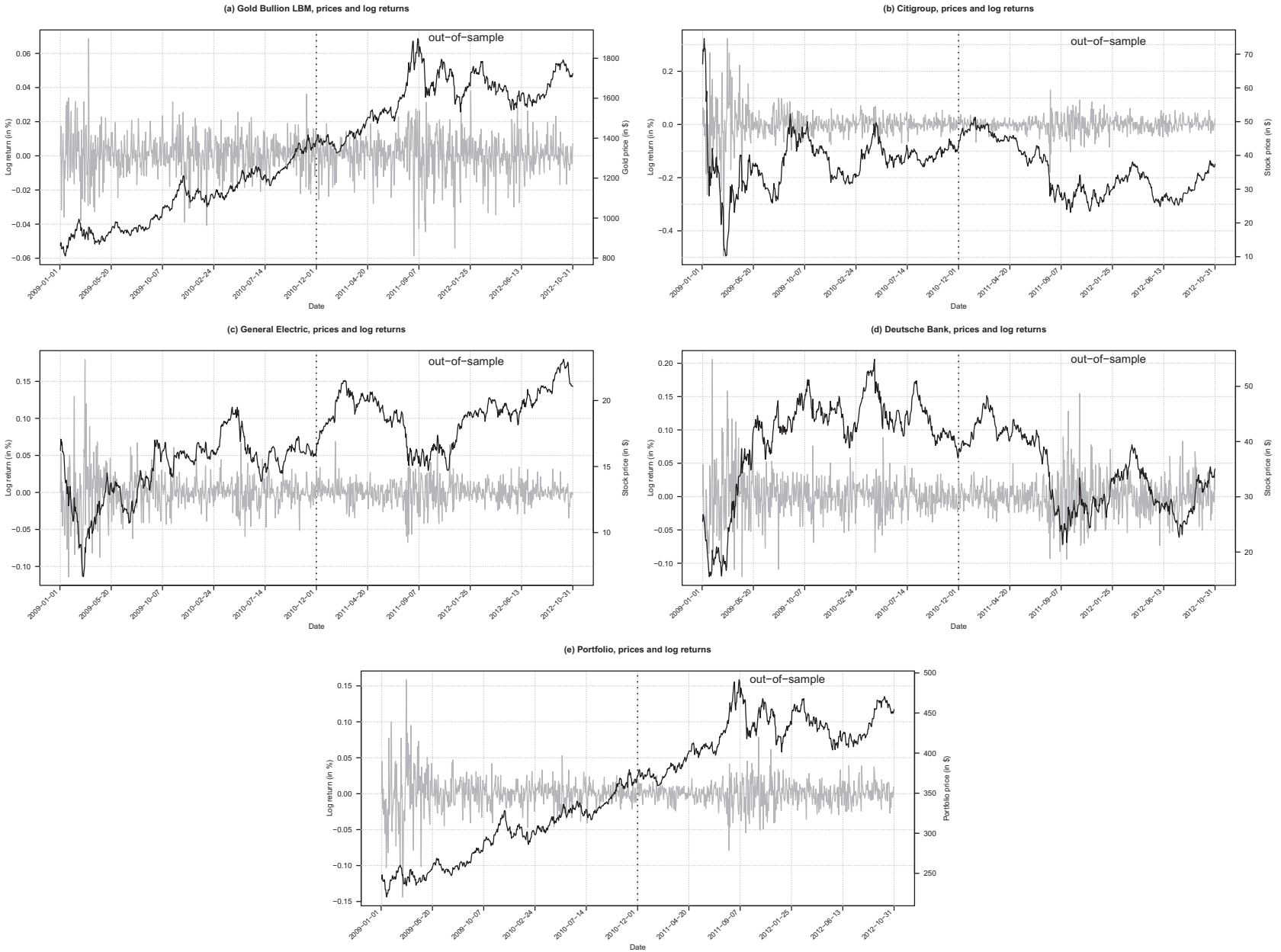

Fig. 3. Time series plots of quotes and log returns of individual assets and the portfolio used in the empirical study. The figure shows plots of the log returns (grey lines) and the quotes (black lines) on the financial assets considered in the empirical study. Panel (a) shows returns and quotes for Gold Bullion LBM ($/Troy Ounce), Panel (b) for the stock of Citigroup, Panel (c) for the stock of General Electric, Panel (d) for the stock of Deutsche Bank, and Panel (e) shows the quotes and log returns on an equal-weighted portfolio consisting of the four individual assets. The sample covers the period from January 1, 2009 to October 31, 2012 (1000 trading days). Each plot includes a vertical line to separate the initial in-sample and the out-of-sample. The data are taken from *Thomson Reuters Financial Datastream*.

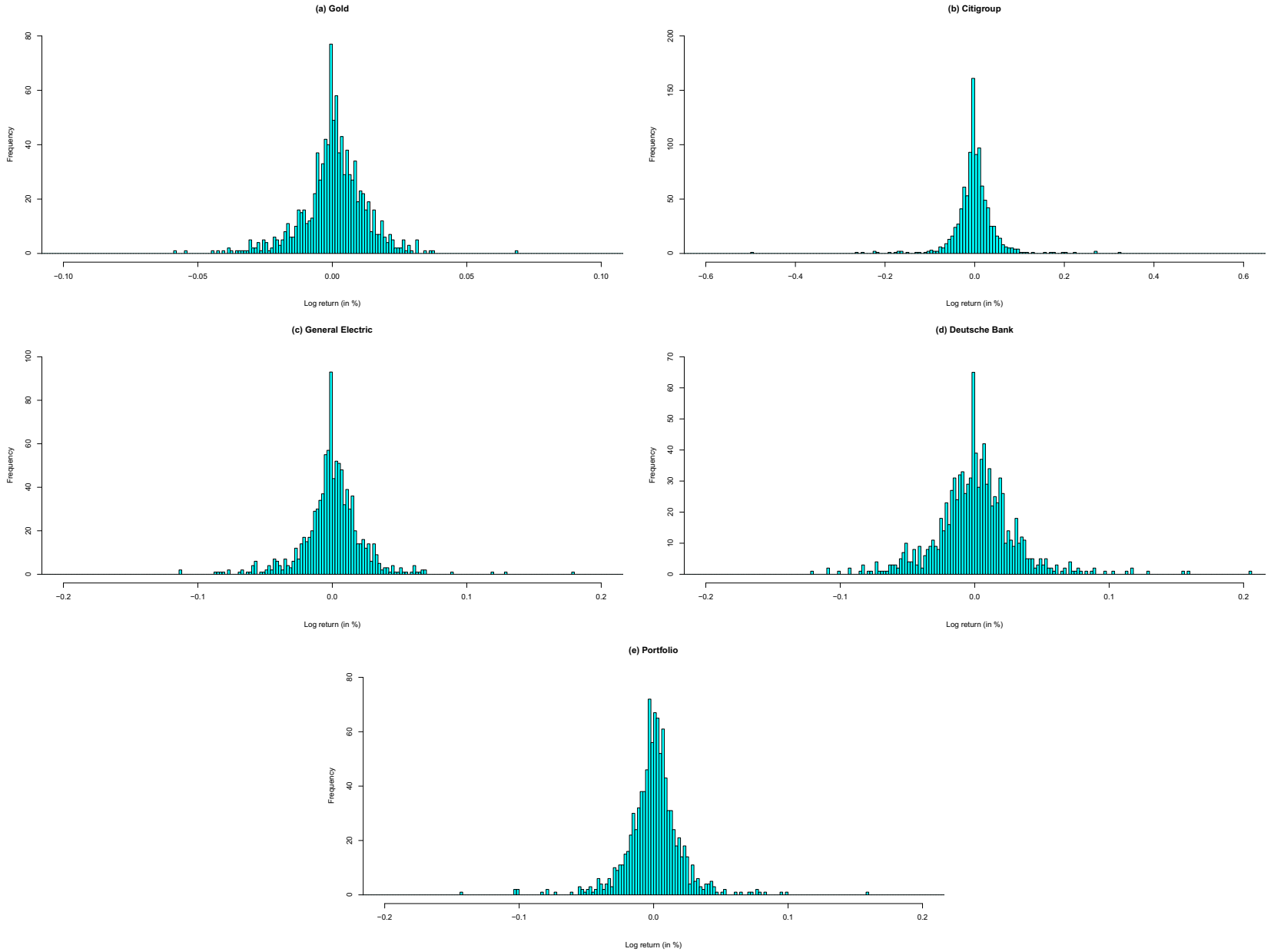

Fig. 4. Histograms of the log returns of individual assets and the portfolio used in the empirical study. The figure shows histograms of the log returns on the financial assets considered in the empirical study. Panel (a) shows returns for Gold Bullion LBM ($/Troy Ounce), Panel (b) for the stock of Citigroup, Panel (c) for the stock of General Electric, Panel (d) for the stock of Deutsche Bank, and Panel (e) shows the log returns on an equal-weighted portfolio consisting of the four individual assets. The sample covers the period from January 1, 2009 to October 31, 2012 (1000 trading days). The data are taken from *Thomson Reuters Financial Datastream*.

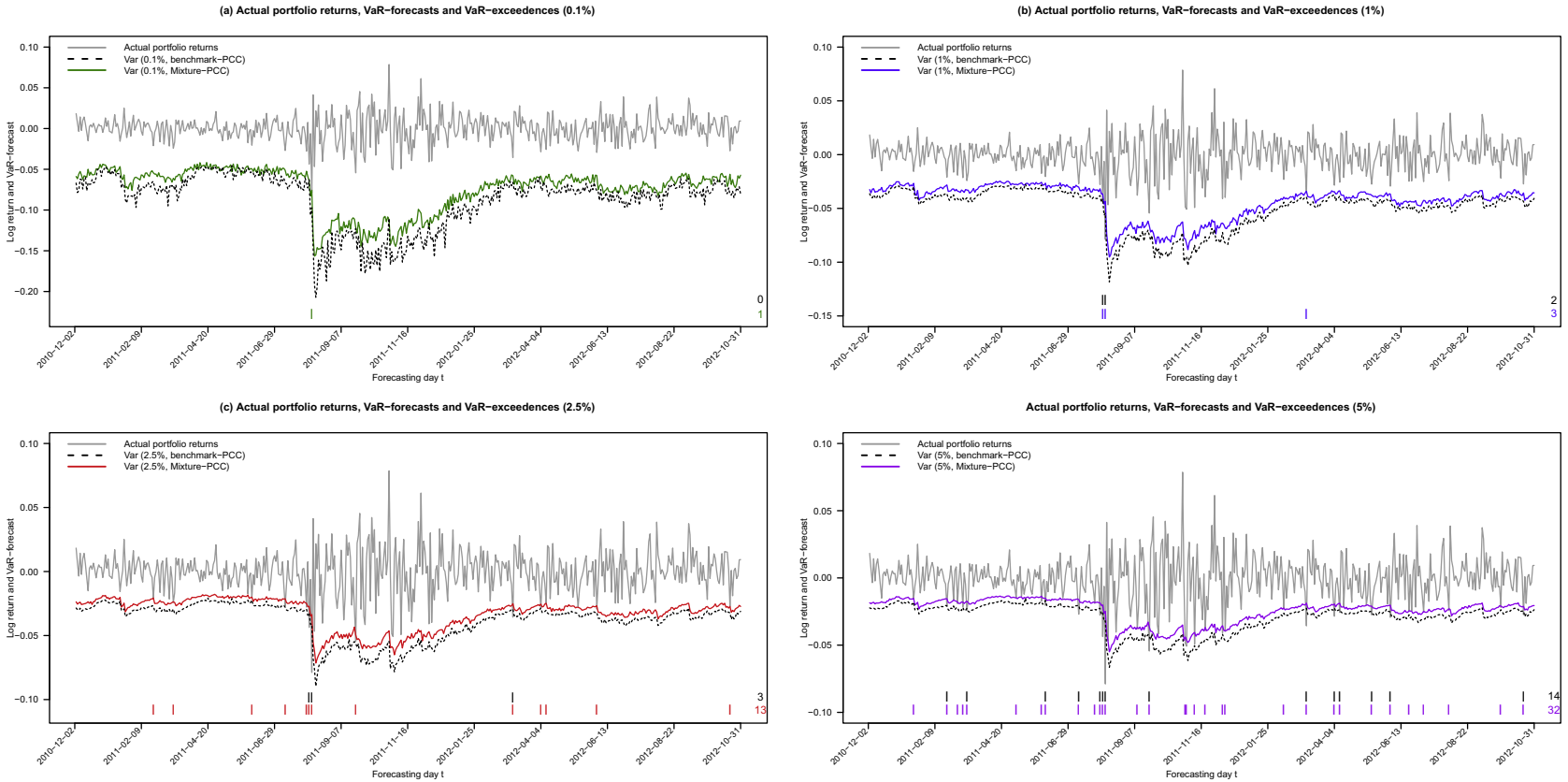

Fig. 5. Comparison of 0.1%-, 1%-, 2.5% and 5%-VaR forecasts and realized portfolio returns. The figure shows plots of the log returns on the four-dimensional portfolio considered in the empirical study and the out-of-sample VaR forecasts. Panel (a) presents the realized portfolio returns (grey line) and the VaR forecasts at the 0.1% significance level computed by the use of the benchmark pair-copula construction calibrated following the procedure laid out in Brechmann et al. (2012) (black line) and our mixture pair-copula-construction (colored line). Panels (b), (c) and (d) show similar comparisons for the 1%-, 2.5% and 5%-VaR. VaR-exceedances are indicated by vertical strokes at the bottom of each plot with the total number of exceedances given in the lower right corner of each plot. The size of the out-of-samples is $T = 500$. The equal-weight portfolio consists of the returns on Gold Bullion LBM and the stocks of Citigroup, General Electric, and Deutsche Bank. The data are taken from *Thomson Reuters Financial Datastream*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

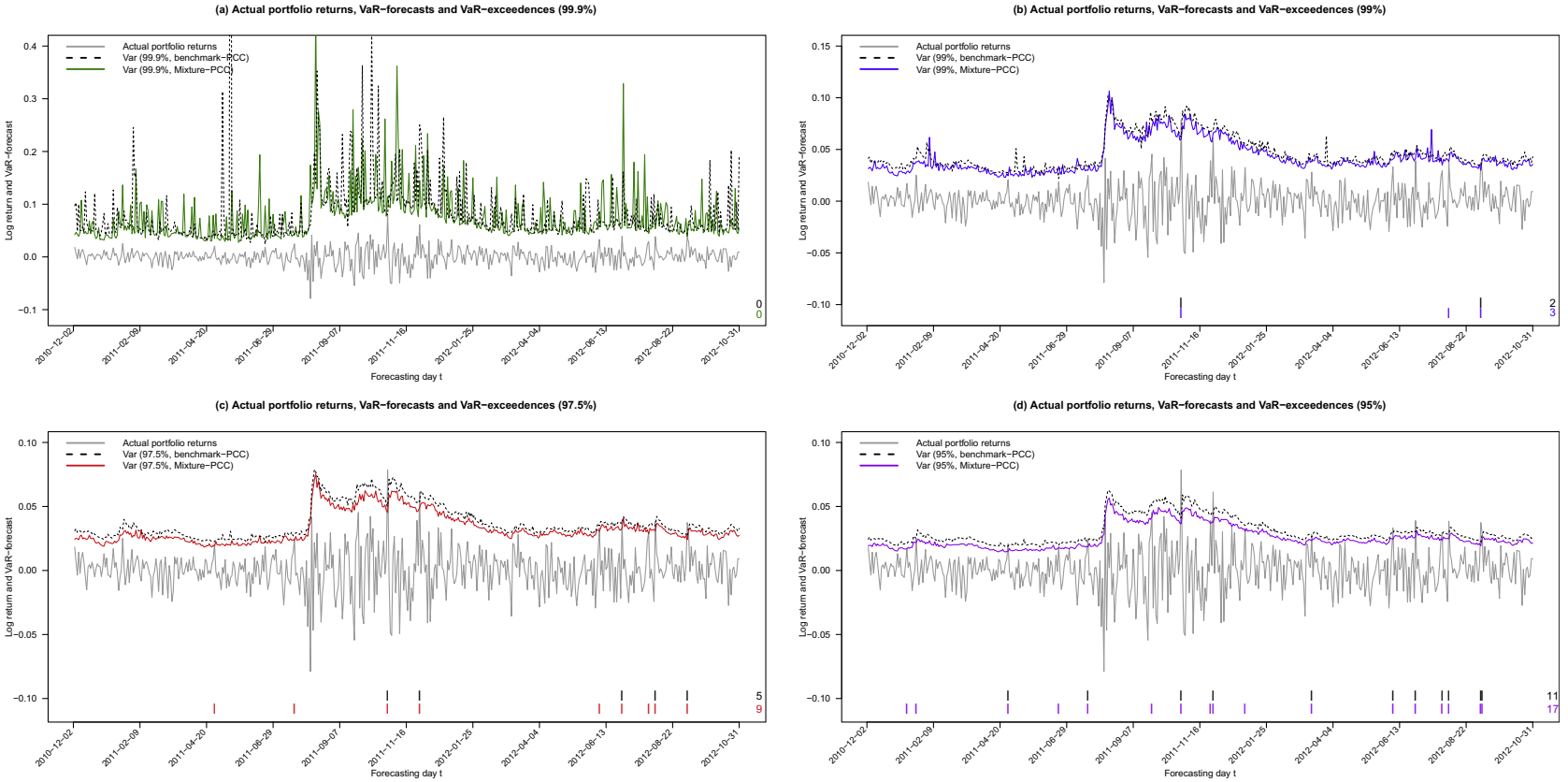

Fig. 6. Comparison of 99.9%-, 99%-, 97.5% and 95%-VaR forecasts and realized portfolio returns. The figure shows plots of the log returns on the four-dimensional portfolio considered in the empirical study and the out-of-sample VaR forecasts. Panel (a) presents the realized portfolio returns (grey line) and the VaR forecasts at the 99.9% significance level computed by the use of the benchmark pair-copula construction calibrated following the procedure laid out in Brechmann et al. (2012) (black line) and our mixture pair-copula-construction (colored line). Panels (b), (c) and (d) show similar comparisons for the 99%-, 97.5% and 95%-VaR. VaR-exceedances are given under the assumption of a short position in the portfolio and are indicated by vertical strokes at the bottom of each plot with the total number of exceedances given in the lower right corner of each plot. The size of the out-of-samples is $T = 500$. The equal-weight portfolio consists of the returns on Gold Bullion LBM and the stocks of Citigroup, General Electric, and Deutsche Bank. The data are taken from *Thomson Reuters Financial Datastream*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Backtesting results – Empirical study. The table presents the results of the two-tailed conditional coverage backtest proposed in Christoffersen (1998) performed on the out-of-sample forecasts for the portfolio-VaR estimated from the vine copula models calibrated by the sequential heuristic of Brechmann and Czado (2013) and Dissmann et al. (2013) and the Mixture-PCCs, respectively. The table reports the expected and the realized number of VaR-exceedances as well as the *p*-values of the two-tailed backtest of conditional coverage proposed in Christoffersen (1998) (CC) and the duration-based Weibull test of independence by Christoffersen and Pelletier (2004) (WB). For both models, the backtesting results are reported for the $(1-\alpha)$-VaR for significance level $\alpha \in \{0.1\%; 1\%; 2.5\%; 5\%\}$. For the 95%-, 97.5%-, 99%- and 99.9%-VaR, exceedances are given under the assumption of a short position in the portfolio. The out-of-sample consists of $T = 500$ trading days. The equal-weight portfolio consists of the returns on Gold Bullion LBM and the stocks of Citigroup, General Electric, and Deutsche Bank. The full sample covers the period from January 1, 2009 to October 31, 2012 (1000 trading days) with the data being taken from *Thomson Reuters Financial Datastream*.

	Exceedances (expected)	Sequential AIC			Mixture-PCC		
		CC test (*P*-value)	WB test (P-value)	Exceedances (realized)	CC test (*P*-value)	WB test (*P*-value)	Exceedances (realized)
VaR $\alpha = 0.1\%$ Conditional coverage	0.5	0.2285	1.0000	0	1.0000	1.0000	1
VaR $\alpha = 1\%$ Conditional coverage	5	0.2231	0.0402	2	0.4250	1.0000	3
VaR $\alpha = 2.5\%$ Conditional coverage	12.5	0.0032	1.0000	3	0.8211	0.7887	13
VaR $\alpha = 5\%$ Conditional coverage	25	0.0274	0.7695	14	0.3113	0.1764	32
VaR $\alpha = 95\%$ Conditional coverage	25	0.0016	0.9389	11	0.2048	0.3362	17
VaR $\alpha = 97.5\%$ Conditional coverage	12.5	0.0348	0.7567	5	0.4279	0.7065	9
VaR $\alpha = 99\%$ Conditional coverage	5	0.2264	0.0385	2	0.4126	0.9116	3
VaR $\alpha = 99.9\%$ Conditional coverage	0.5	0.2343	1.0000	1	0.2322	1.0000	0

in the post Lehman Bros. era and covers the climax of the European Sovereign Debt crisis. In total, our sample includes 1000 trading days. We forecast the portfolio's Value-at-Risk by employing rolling windows of 500 trading days. In turn, the out-of-sample period consists of 500 trading days. Fig. 3 shows plots of the log returns and the quotes on the four financial assets (Panel (a) through (d)) as well as the portfolio (Panel (e)), respectively. Each plot includes a vertical line to separate the initial in-sample and the out-of-sample.

The plots of the four portfolio constituents highlight several challenging features with respect to VaR-forecasting. First, all four assets possess different evolutions of their time series' volatility during the in- and the out-of-sample. While the returns of General Electric and Citigroup seem to calm down after a high volatility phase at the beginning of the in-sample (and thus, during the financial crisis), the time series of the stock of Deutsche Bank exhibits high (and clustered) volatility in the out-sample. Most probably, the increase in stock return volatility for Deutsche Bank was caused by the renewal of the European Sovereign Debt Crisis in the summer of 2011. Throughout our sample, the gold price increases steadily with volatility spiking in the out-of-sample. Complementing the plots of the individual assets, Panel (e) in Fig. 3 plots the time series prices and log returns of an equal-weighted portfolio of the four assets. While the time evolution of the price of the portfolio is almost identical to that of the gold bullion, the log returns on the portfolio exhibit both calm and extremely volatile phases. Moreover, the log returns also exhibit extreme positive and negative spikes at the start of the in-sample during the financial crisis.

To further illustrate the properties of our time series, we plot histograms of the log returns of the four individual assets and the portfolio in the full sample in Fig. 4.

The plots given in Fig. 4 underline the heavy left tail of the portfolio returns as well as the existence of several extreme returns in our data.

While this evidence prompts the use of copula models for modeling the joint distributions' behavior in the tails, symmetric marginals might not suffice to model asymmetries in both the distribution of the residuals as well as in conditional volatilities. We therefore employ both the Jarque–Bera and the Kolmogorov–Smirnov test to check the hypothesis of normally distributed residuals. Both tests are rejected. Furthermore, performing the D'Agostino normality test on the residuals indicates strong deviations from normality due to kurtosis. Finally, we check the adequacy of our assumption of Student-t distributed residuals using Quantile–Quantile-plots. The results underline the adequacy of our marginal models and the fact that asymmetric marginal models are not required for our particular sample.

In order to check for asymmetries in conditional volatility, we jointly conduct the Sign Bias Test, the Negative Size Bias Test as well as the Positive Size Bias Test as proposed by Engle and Ng (1993). More precisely, we test the null that the squared residuals of the return series cannot be predicted by the sign and the magnitude of return shocks. The resulting p-values of the tests are quite large for all series, indicating that there is no predictive power in the shocks with regard to the squared residuals. We thus find no evidence of asymmetric conditional volatility in any of the return series.

4.2. Results

We use the methodology presented in Section 4.1 to compute the one-step-ahead forecasts of the portfolio-VaR for each day in the out-of-sample using rolling windows of 500 trading days. To analyze the differential effect of different confidence levels for the VaR on our models' forecasting accuracy, we forecast the 0.1%-, 1%-, 2.5%- and 5%-VaR for a long position in the portfolio and the 95%-, 97.5%-, 99%-, and 99.9%-VaR for a short position, respectively. The selection of these VaR-levels is in part motivated by the results of the simulation study presented in the previous section. On the one hand, it has been highlighted that both models, the heuristically calibrated parametric benchmark as well as our Mixture-PCC, seem to be well-suited for producing accurate 0.1%-VaR-forecasts. This result is quite remarkable given the fact that the 0.1%-VaR-level is of high importance for the calculation of regulatory capital requirements implemented in the Basel III

framework. On the other hand, especially the heuristic benchmark was significantly outperformed in in-sample backtesting by the Mixture-PCC for higher significance levels.[20]

First, we present a comparison of the estimated pair-copula families and parameters for the benchmark model as well as the estimated weights and parameters in the mixture model for the dates for which the difference in the VaR-forecasts of both models is largest.[21] For the 0.1%-VaR-level, the largest difference between the forecasts of the two models is nearly 6% in portfolio-VaR (date: 2011–08-23). For the benchmark model, the Student's t (0.02; 13.11 degrees of freedom), the Survival Clayton (0.11) and the Rotated Joe-Clayton (1.06) copula are chosen in the first tree. In the second tree the Rotated Clayton-Gumbel copula is twice selected (1.71 and 1.53, respectively). The only copula in the third tree is the Gaussian (0.38).

The largest differences for the 1%- and 2.5%-VaR-levels are given on the same date (2011–08-11) and amount to 2.31% and 1.8% in portfolio-VaRs, respectively. The fitted benchmark model differs only marginally in the first two trees from the previous model. More precisely, the same copula families are chosen and the estimated parameters show only little variation. In the third tree, the Survival Clayton-Gumbel copula is chosen (2.19; 0.86). For the 5%-VaR-level the largest difference in forecasted VaRs is 1.34% (date: 2011–11-02). Here, in the first tree of the benchmark vine model, three Student's t pair-copulas ($-0.05, 0.01$ and -0.02; 6.59, 13.55 and 6.70 degrees of freedom) are used. Again, in the second tree the Rotated Clayton-Gumbel copula is selected twice (1.81 and 1.61 respectively). In the third tree the Survival Clayton-Gumbel copula is chosen (1.63; 0.96).

For all fitted Mixture-PCCs, the key component is the Student's t copula ($[-0.19, 0.25]$; degrees of freedom $\in [1.67, 14.37]$) with a percentage share of at least 31.36% and not more than 46.38% in all mixtures. The Frank copula ($[-89.04, 0.99]$) is the second main component and models the outcomes with strong negative dependence. However, the dependence in the tails of the Frank copula itself tends to be relatively weak. The weights are in the range of 23.27% to 29.50%. Furthermore, the Mixture-PCCs are significantly influenced by the Clayton copula $[1.33, 4.70]$ thus inducing significant lower tail dependence in the pair-copulas. The percentage share of this copula is in the range of 17.14% to 22.39%. Only a minor impact is given by the Gumbel Copula ($[2.33, 5.64]$), which captures the upper tail dependence in the model. Here, the weights are in the range between 10.62% and 21.64%.

Fig. 5 shows the out-of-sample VaR-forecasts as well as the realized portfolio returns for all significance levels of a long position in the portfolio. Each of the Panels (a), (b), (c) and (d) presents the realized portfolio returns and the VaR-forecasts for the corresponding confidence levels highlighting the results of our proposed Mixture-PCC and the simple PCC calibrated via the heuristic based on Akaike's Information Criterion, respectively. Corresponding results for the short position in the portfolio are illustrated in Fig. 6.

The plots in Fig. 5 highlight our previous finding that both the Mixture-PCC and benchmark PCC adequately forecast portfolio losses. Both models seem to adapt well to the specific evolution of the realized portfolio returns for all VaR-levels. Additionally, note that exceedances of the VaR-forecasts occur only in periods of high volatility combined with large losses in the long position of the portfolio investment. We would expect 0.5, 5, 12.5 and 25 VaR-exceedances for the four significance levels. While the benchmark-PCC yields 0, 2, 3 and 14 violations, our Mixture-PCC forecasts portfolio losses more accurately with 1, 3, 13 and 32 violations.

Panel (a) in Fig. 5 underlines our first impression from the simulation study that both models forecast the 0.1%-VaR of the portfolio quite accurately. In this case 0 and 1 realized exceedances for the benchmark model and our proposed Mixture-PCC, respectively, indicate an accurate forecast of the expected number of exceedances (0.5). Panel (b) shows similar VaR-forecasts for both models at the 1% significance level. Furthermore, both models again yield similar numbers of VaR-violations. Additionally, Panel (c) provides further interesting insights. On the one hand, the benchmark model again appears to forecast losses accurately. On the other hand, the mixture model yields almost the expected number of VaR-exceedances while the benchmark is significantly more conservative. Also, note that the plots of the VaR-forecasts of our Mixture-PCC are always above the plot of the forecasts of the benchmark. These findings are underlined by the plots shown in Fig. 6 for the short position in the portfolio.

The plots in Figs. 5 and 6 underline the impression that both models forecast the VaR of the portfolio quite accurately and yield (approximately) correct numbers of VaR-exceedances for the first three confidence levels. The key difference, however, is that our proposed Mixture-PCC model does not overestimate portfolio risk to such an extent as the benchmark model does. Our model can thus help risk managers to reduce regulatory risk capital.

To further substantiate these findings, we perform the two-tailed conditional coverage backtest proposed in Christoffersen (1998) on the out-of-sample VaR-forecasts of both models. The results of the formal backtest are presented in Table 3.[22]

The results of the backtest given in Table 3 again show that both models yield accurate VaR-forecasts for a significance level of 0.1%. The p-values differ considerably between both models, but both are not rejected at the 5% significance level. This is in line with our previous findings of the simulation study for the 0.1%-VaR-level. For the 1% significance level of the VaR, again, both models are not rejected by the formal backtest of conditional coverage. However, the formal backtest supports the results presented earlier for the $\alpha = 2.5\%$-Value-at-Risk forecasts. While the benchmark PCC is clearly rejected by the backtest due to its overestimation of portfolio risk, our Mixture-PCC is not rejected. This result also holds for the $\alpha = 5\%$-Value-at-Risk forecasts even though the number of the realized exceedances of the Mixture-PCC is higher than the expected number of exceedances. The VaR-levels of the short portfolios (95%, 97.5%, 99%, and 99.9%, respectively) permit a similar interpretation. The benchmark PCC is clearly rejected whereas our Mixture-PCC is not rejected.[23] Consequently, the results of the formal backtest support our main finding that Mixture-PCCs forecast losses of financial portfolios significantly more accurately than the heuristic benchmark PCC.

A note of caution is, however, in order when interpreting our results. The use of mixture copulas as pair-copulas increases the (already large) number of parameters that need to be estimated to fully specify a pair-copula-constructions. As a result, the parameter uncertainty and the risk of overfitting the data (especially when the sample size is small) increase in parallel with the model's increased flexibility. As the Mixture-PCCs produce accurate and reliable results in both our simulations and our empirical study, we believe parameter uncertainty to be of lesser concern in the particular setting of our study. However, the potential bias due

[20] The estimation of the benchmark PCC required less than ten minutes, on average, while the simulation of the returns and the forecasting of the VaR required less than one hour, on average. We again find the estimation of our Mixture-PCC to take about three to five times longer than the heuristic benchmark.

[21] In the following, parameter estimates are presented in brackets.

[22] We also computed VaR-forecasts using historical simulations as a benchmark. In unreported results, this benchmark model produced six VaR-exceedances for the 0.1%, 1%, and 2.5% significance levels, and eleven exceedances for the 5% VaR-level (long position). Consequently, this naive benchmark produced VaR-forecasts that were regularly off target with the model significantly overestimating portfolio risk for most VaR-levels.

[23] Complementing the test of conditional coverage, the results of the Weibull backtest show that with one exception no model is rejected based on the assessment of the independence of the exceedances.

to a (too) small sample size and a (too) high number of mixture components could significantly impact results when using Mixture-PCCs in different settings.

5. Conclusion

In this study, we propose the use of mixture copulas in d-dimensional Pair-Copula-Constructions as a new strategy to circumvent the otherwise necessary (and error-prone) selection of parametric forms for the $d(d-1)/2$ bivariate pair-copulas. While previous studies in the literature have tried different approaches to select optimally fitting pair-copulas from parametric copula families characterized by different tail dependence (e.g., goodness-of-fit tests, graphical tools, etc.), we propose to use convex combinations of these bivariate copulas for each pair-copula in a vine model. Each mixture pair-copula is then estimated using the well-known EM-algorithm yielding a fully specified vine model in which no parametric copula needs to be selected as all parametric candidate copulas can be included in the mixture pair-copulas. After outlining our proposed Mixture-PCC, we test the performance of our new model in comparison with a benchmark PCC in which each pair-copula is chosen by computing the Akaike's Information Criterion for each candidate parametric copula and selecting the copula with the optimal AIC value. We perform both a simulation study on the in-sample fit of both models as well as an empirical study in which we assess both models' out-of-sample forecasting accuracy.

Our main result can be summarized as follows: in our simulations and in the empirical study, both models yield acceptable Value-at-Risk forecasts. However, we show that our proposed Mixture-PCC yields better results in backtesting for at least the 2.5% significance level while the benchmark overestimates portfolio risk. Especially for higher dimensions, our Mixture-PCC model seems to approximate portfolio losses better than the benchmark which is far too conservative in many cases. Consequently, our model can help risk managers to save on regulatory risk capital while at the same time satisfactorily bounding possible portfolio losses.

For future research, one could think of an analysis of extended Mixture-PCCs in which the number of mixture pair-copulas is not fixed as it is done in this study. Furthermore, Mixture-PCCs could be combined with pruning strategies to truncate or simplify some of the mixture pair-copulas in lower trees to limit the computational cost. We expect all these extensions to lead to further improvements on the forecasting accuracy of our Mixture-PCC in comparison to models from the related literature and intend to address them in a future study.

References

Aas, K., Berg, D., 2009. Models for construction of multivariate dependence – a comparison study. European Journal of Finance 15, 639–659.

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

Ausín, M.C., Lopes, H., 2010. Time-varying joint distributions through copulas. Computational Statistics and Data Analysis 54, 2383–2399.

Baglioni, A., Cherubini, U., 2013. Within and between systemic country risk. Theory and evidence from the sovereign crisis in Europe. Journal of Economic Dynamics and Control 37 (8), 1581–1597.

Bedford, T., Cooke, R., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence 32, 245–268.

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

Brechmann, E., Czado, C., 2013. Risk management with high-dimensional vine copulas: an analysis of the Euro Stoxx 50. Statistics and Risk Modeling 30 (4), 307–342.

Brechmann, E., Schepsmeier, U., 2013. Modeling dependence with C- and D-vine copulas. Journal of Statistical Software 52 (3), 1–27.

Brechmann, E., Czado, C., Aas, K., 2012. Truncated regular vines in high dimensions with applications to financial data. Canadian Journal of Statistics 40 (1), 68–85.

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

Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance, 1 ed. Wiley.

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

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

Christoffersen, P., Pelletier, D., 2004. Backtesting value-at-risk: a duration-based approach. Journal of Financial Econometrics 2 (1), 84–108.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete via the EM algorithm. Journal of the Royal Statistical Society: Series B 39 (1), 1–38.

Dissmann, J., Brechmann, E., Czado, C., Kurowicka, D., 2013. Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics and Data Analysis 59 (1), 52–69.

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

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Fantazzini, D., 2009. The effects of misspecified marginals and copulas on computing the value at risk: a Monte Carlo study. Computational Statistics and Data Analysis 53, 2168–2188.

Fischer, M., Klöck, C., Schlüter, S., Weigert, F., 2009. An empirical analysis of multivariate copula model. Quantitative Finance 7, 839–854.

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

Hafner, C.M., Reznikova, O., 2010. Efficient estimation of a semiparametric dynamic copula model. Computational Statistics and Data Analysis 54, 2609–2627.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20, 873–889.

Heinen, A., Valdesogo, A., 2009. Asymmetric CAPM Dependence for Large Dimensions: The Canonical Vine Autoregressive Model, CORE Discussion Papers 2009069.

Hobæk-Haff, I., Segers, J., 2015. Non-parametric Estimation of Pair-copula Constructions with the Empirical Pair-copula. Computational Statistics & Data Analysis 84, 1–13. http://dx.doi.org/10.1016/j.csda.2014.10.020.

Hu, L., 2006. Dependence patterns across financial markets: a mixed copula approach. Applied Financial Economics 16, 717–729.

Ince, O., Porter, R., 2006. Individual equity return data from thomson datastream: handle with care! Journal of Financial Research 29, 463–479.

Joe, H., 1996. Families of m-variate distributions with given margins and $m(m-1)/2$ bi-variate dependence parameters. In: Rüschendorf, L., Schweizer, B., Taylor, M. (Eds.), Distributions with Fixed Marginals and Related Topics. Institute of Mathematical Statistics, pp. 120–141.

Joe, H., 1997. Multivariate Models and Dependence Concepts, 1 ed. Chapman & Hall.

Jondeau, E., Rockinger, M., 2006. The copula-GARCH model of conditional dependencies: an international stock market application. Journal of International Money and Finance 25, 827–853.

Kim, G., Silvapulle, M., Silvapulle, P., 2007. Comparison of semiparametric and parametric methods for estimating copulas. Computational Statistics and Data Analysis 51, 2836–2850.

Kim, D., Kim, J.-M., Liao, S.-M., Jung, Y.-S., 2013. Mixture of D-vine copulas for modeling dependence. Computational Statistics and Data Analysis 64, 1–19.

Kole, E., Koedijk, K., Verbeek, M., 2007. Selecting copulas for risk management. Journal of Banking and Finance 31, 405–423.

Kurowicka, D., 2010. Optimal truncation of vines. In: Kurowicka, D., Joe, H. (Eds.), Dependence Modeling: Handbook on Vine Copulae. World Scientific Publishing Co.

Li, D.X., 2000. On default correlation: a copula function approach. Journal of Fixed Income 9 (4), 43–54.

Li, Q., Brown, J., Huang, H., Bickel, P., 2011. Measuring reproducibility of high-throughput experiments. The Annals of Applied Statistics 5 (3), 1752–1779.

Liu, Y., Luger, R., 2009. Efficient estimation of copula-GARCH models. Computational Statistics and Data Analysis 53, 2284–2297.

McLachlan, G., Peel, D., 2000. Finite Mixture Models, 1 ed. Wiley.

Nelsen, R.B., 2006. An Introduction to Copulas, 2 ed. Springer.

Nikoloulopoulos, A.K., Joe, H., Li, H., 2012. Vine copulas with asymmetric tail dependence and applications to financial return data. Computational Statistics and Data Analysis 56 (11), 3659–3673.

Rodriguez, J., 2007. Measuring financial contagion: a copula approach. Journal of Empirical Finance 14 (3), 401–423.

Ruenzi, S., Weigert, F., 2013. Crash Sensitivity and the Cross-Section of Expected Stock Returns, Working Paper.

Rueschendorf, L., 2013. Mathematical Risk Analysis, 1 ed. Springer.

Weiß, G., 2011. Are copula-GoF-tests of any practical use? empirical evidence for stocks, commodities and FX futures. The Quarterly Review of Economics and Finance 51, 173–188.

Weiß, G., 2013. Copula-GARCH vs. dynamic conditional correlation – an empirical study on VaR and ES forecasting accuracy. Review of Quantitative Finance and Accounting 41 (2), 179–202.

Weiß, G., Scheffer, M., 2012. Smooth Nonparametric Bernstein Vine Copulas, Working Paper, September 2012.

Weiß, G., Supper, H., 2013. Forecasting liquidity-adjusted intraday value-at-risk with vine copulas. Journal of Banking and Finance 37 (9), 3334–3350.

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