Financ Mark Portf Manag (2010) 24: 193–213
DOI 10.1007/s11408-010-0130-1

Pair-copulas modeling in finance

Beatriz Vaz de Melo Mendes · Mariângela Mendes Semeraro · Ricardo P. Câmara Leal

Published online: 15 April 2010

Abstract This paper concerns itself with applications of pair-copulas in finance, and bridges the gap between theory and application. We provide a broad view of the problem of modeling multivariate financial log-returns using pair-copulas, gathering together for this purpose theoretical and computational results from the literature on canonical vines. From the practitioner's viewpoint, the paper shows the advantages of modeling through pair-copulas and makes clear that it is possible to implement this methodology on a daily basis. All the necessary steps (model selection, estimation, validation, simulations, and applications) are discussed at a level easily understood by all data analysts.

Keywords Pair-copulas · Multivariate modeling · Markowitz mean variance model

JEL Classification C16 · C51 · G11

1 Introduction

The Basel II international capital framework has encouraged, at least to some extent, the development of more sophisticated statistical tools for finance. Underlying each tool is a probabilistic model assumption. For a long time, finance modeling only considered the multivariate normal distribution. This situation was partially due to the fact that most of the important theoretical results in this area were based on the normality assumption, and also due to the lack of suitable alternative multivariate

B. Vaz de Melo Mendes (✉) · M. Mendes Semeraro
IM/COPPEAD, Federal University at Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: beatriz@im.ufrj.br

R. P. Câmara Leal
COPPEAD, Federal University at Rio de Janeiro, Rio de Janeiro, Brazil

distributions and software limitations. However, a simple exploratory analysis carried out on any collection of log-returns on indexes, stocks, portfolios, or bonds reveals significant departures from normality.

Data on log-returns present some well-known stylized facts and are characterized by two special features: (I) each series typically shows its own degree of asymmetry and high kurtosis, as well as some specific pattern of temporal dynamics, and (II) the dependence structure among pairs of variables will vary substantially, ranging from independence to complex forms of nonlinear dependence. No natural family of multivariate distribution (e.g., the elliptical family) includes all these features. This is true for the case of unconditional modeling, and also applies to the errors distribution of sophisticated dynamic models.

A solution is the use of copulas, which was introduced by Sklar (1959) and applied to the field of finance by Embrechts et al. (1999). Other works on the topic include Kassberger and Kiesel (2006), who offer as an alternative to copulas when modeling dependency structures in hedge funds a multivariate extension of the normal inverse Gaussian distribution and argue that it captures well the return characteristics of such funds. Gatzert et al. (2008) employ copulas to assess risk concentration and joint default probabilities given different compositions of financial groups and experiment with different types of copulas. They find that the joint default probabilities of the members of a financial group can vary substantially even when financial groups present similar risk concentration factors.

Using copulas simplifies both model specification and estimation. Initially, marginal distributions are fitted using the vast range of univariate models available. In a second step, the dependence between variables is modeled using a copula. However, this approach has its limitations. Although we are able to find very good (conditional and unconditional) univariate fits tailored for each margin, when it comes to copula fitting, there are significant obstacles to solving the required optimization problem over many dimensions, the so-called "curse of dimensionality" (Scott 1992). Most of the available software deal with the bivariate case only. Even if we are able to fit a d-dimensional copula, $d > 2$, parametric copula families usually restrict all pairs to having the same type or strength of dependence. For example, in the case of the t-copula, in addition to the correlation coefficients, a single parameter, the number of degrees of freedom, is used to compute the coefficient of tail dependence for *all pairs*, thus the t-copula fails when modeling (II).

Pair-copulas, being a collection of potentially different bivariate copulas, are flexible and very appealing. The method of construction is hierarchical, where variables are sequentially incorporated into the conditioning sets as one moves from level 1 (tree 1) to tree $d - 1$. The composing bivariate copulas may vary freely, from the parametric family to the parameter values. Therefore, all types and strengths of dependence can be covered. Pair-copulas are easy to estimate and simulate, making them very appropriate for modeling in finance.

Most of extant work on pair-copulas is theoretical in nature and by no means conclusive in that regard as of yet. A few papers provide applications in finance, but most of them simply fit a pair-copula to the data (see, e.g., Min and Czado 2010; Aas et al. 2007; Berg and Aas 2008; Fischer et al. 2008).

In this paper, we go beyond inference and provide practical applications, such as the pair-copula construction of efficient frontiers and risk computation. We consider both conditional and unconditional models for the univariate fits. The conditional models are the well-known combinations of ARFIMA and FIGARCH models (Sect. 4). As unconditional univariate models, we use the very flexible skew-t family (Sect. 3). Estimation of the models is based on the maximum likelihood method. Applications will follow the fits, and are intend to illustrate the utility of the pair-copulas approach since there are many applications in finance that rely on a good joint fit for the data. For example, computing risk measure estimates, finding optimal allocations for portfolios, pricing derivatives, and so on.

The main objective of this paper is to demonstrate to practitioners how pair- copulas modeling can be useful in finance. The paper also collects important results on this topic from the literature, thus the large number of references provided. In summary, the contributions of this paper are: (1) to show how a good multivariate (conditional or unconditional) fit for log-return data may be obtained with pair-copulas approach; (2) to propose the use of the skew-t distribution as the unconditional model for the margins; (3) to show how pair-copulas may be used on a daily basis in finance, particularly for constructing efficient frontiers and computing the value-at-risk; and (4) to show how parametric replications of the data may be obtained and used to assess variability and construct confidence intervals. In the case of the efficient frontier, pair-copulas make it possible to check portfolio quality, find out if a portfolio rebalance is needed, or discover if the inclusion of some other component would significantly improve the expected return for the same risk level.

The remainder of the paper is organized as follows. In Sect. 2, we briefly review definitions of copulas and pair-copulas. In Sect. 3, we consider the unconditional approach for the marginal fits combined with the pair-copulas fit, providing an application in Sect. 3.1, where we obtain optimal portfolios and show how to construct pair-copulas based replications of the efficient frontier. In Sect. 4, we take the conditional approach for the marginal fits; in Sect. 4.1 duplicating the application carried out in Sect. 3.1 but this time under the conditional approach. Section 5 concludes.

2 Copulas and pair-copulas: a brief review

2.1 Copulas

Consider a stationary d-variate process $(X_{1,t}, X_{2,t}, \dots, X_{d,t})_{t\in\mathcal{Z}}$, $\mathcal{Z}$ a set of indices. In the case the joint law of $(X_{1,t}, X_{2,t}, \dots, X_{d,t})$ is independent of t, the dependence structure of $\mathbf{X} = (X_1, X_2, \dots, X_d)$ is given by its (constant) copula C. If $\mathbf{X}$ is a continuous random vector with joint cumulative distribution function (c.d.f.) F with density function f, and marginal c.d.f.s F_i with density functions f_i, $i = 1, 2, \dots, d$, then there exists a unique copula C pertaining to F, defined on $[0, 1]^d$ such that

$$C\big(F_1(x_1), F_2(x_2), \dots, F_d(x_d)\big) = F(x_1, x_2, \dots, x_d) \tag{1}$$

holds for any $(x_1, x_2, \dots, x_d) \in \Re^d$ (Sklar's theorem, Sklar 1959). Let $F_i(X_i) \equiv U_i$, $i = 1, \dots, d$. From the assumptions made, U_i follows an uniform(0, 1) distribution. Therefore a copula is a multivariate distribution with standard uniform margins.

Multivariate modeling through copulas allows for factoring the joint distribution into its marginal univariate distributions and a dependence structure, its copula. By taking partial derivatives of (1) one obtains

$$f(x_1,\dots,x_d) = c_{1\cdots d}(F_1(x_1),\dots,F_d(x_d))\prod_{i=1}^{d} f_i(x_i) \tag{2}$$

for some d-dimensional copula density $c_{1\cdots d}$. This decomposition allows estimating the marginal distributions f_i separated from the dependence structure given by the d-variate copula. In practice, this aspect simplifies both *specification of the multivariate distribution* and its *estimation*.

The copula C provides all information about the dependence structure of F, independently of the specification of the marginal distributions. It is invariant under monotone increasing transformations of $\mathbf{X}$, making copula-based dependence measures interesting scale-free tools for studying dependence. For example, to measure monotone dependence (not necessarily linear), one may use Spearman's rank correlation (r)

$$r(X_1,X_2) = 12\int_0^1\int_0^1 u_1u_2\,dC(u_1,u_2) - 3. \tag{3}$$

The rank correlation r is invariant under strictly increasing transformations. It always exists in the interval $[-1,1]$, does not depend on the marginal distributions; the values ± 1 occur when the variables are functionally dependent, that is, when they are modeled by one of the Fréchet limit copulas.

Until recently, Pearson's product moment (linear) correlation ρ was used to measure the association between financial products. Although ρ is the canonical measure in the Gaussian world, ρ is not a copula-based dependence measure since it also depends on the marginal distributions. In addition to the drawback of being able to measure only *linear* correlation, ρ has other weaknesses, a number of which are by now well known (see e.g., Embrechts et al. 1999). Note that

$$r(X_1,X_2) = \rho\big(F_1(X_1),F_2(X_2)\big),$$

so that in the copula environment, the rank and the linear correlations coincide.

Another important copula-based dependence concept is the *coefficient of upper tail dependence*, defined as

$$\lambda_U = \lim_{\alpha\to 0^+}\lambda_U(\alpha) = \lim_{\alpha\to 0^+}\Pr\big\{X_1 > F_1^{-1}(1-\alpha)|X_2 > F_2^{-1}(1-\alpha)\big\},$$

provided a limit $\lambda_U \in [0,1]$ exists. If $\lambda_U \in (0,1]$, then X_1 and X_2 are said to be *asymptotically dependent* in the upper tail. If $\lambda_U = 0$, they are *asymptotically independent*. Similarly, the *lower tail dependence coefficient* is given by

$$\lambda_L = \lim_{\alpha\to 0^+}\lambda_L(\alpha) = \lim_{\alpha\to 0^+}\Pr\big\{X_1 < F_1^{-1}(\alpha)|X_2 < F_2^{-1}(\alpha)\big\},$$

provided a limit $\lambda_L \in [0,1]$ exists. The coefficient of tail dependence measures the amount of dependence in the upper (lower) quadrant tail of a bivariate distribution.

In finance, it is related to the strength of association during extreme events. The copula derived from the multivariate normal distribution does not have tail dependence. Therefore, if the Gaussian copula is assumed for modeling log-returns, for many pairs of variables the joint risks would be underestimated.

Let C be the copula of (X_1, X_2). It follows that

$$\lambda_U = \lim_{u \uparrow 1} \frac{\overline{C}(u,u)}{1-u}, \quad \text{where } \overline{C}(u_1,u_2) = \Pr\{U_1 > u_1, U_2 > u_2\} \quad \text{and}$$

$$\lambda_L = \lim_{u \downarrow 0} \frac{C(u,u)}{u}.$$

Other concepts of tail dependence do exist, including the concept of multivariate tail dependence (Joe 1996, IMS volume).

Parametric estimation of copulas is usually accomplished in two steps, as suggested by (2). In the first step, conditional (or unconditional) models are fitted to each margin, and the standardized innovations distributions F_i, $i = 1, \ldots, d$, (which may as well be the empirical distribution) are estimated. Through the probability integral transformation based on the $\widehat{F_i}$, the pseudo uniform(0, 1) data are obtained and used in the second step to estimate the best parametric copula family.

Copula parameters are usually estimated by maximum likelihood (Joe 1997), but can be obtained via robust and minimum distance estimators (Tsukahara 2005; Mendes et al. 2007), or semi-parametrically (Vandenhende and Lambert 2005). Goodness of fit may be assessed visually by means of pp-plots or based on some formal goodness-of-fit (GOF) test, usually based on the minimization of some criterion. GOF tests are proposed in Breymann et al. (2003), Fermanian (2005), Genest et al. (2006), and in the PIT algorithm (Rosenblatt 1952). The most commonly employed approach appears to be to transform the data into a set of independent and standard uniform variables and then calculate some measure of distance, such as the Anderson–Darling or the Kolmogorov–Smirnov measure, between the transformed variables and the uniform distribution. For a discussion of goodness-of-fit tests, see Genest et al. (2007).

2.2 Pair-copulas

The decomposition of a multivariate distribution in a cascade of pair-copulas was originally proposed by Joe (1996), and later discussed in detail by Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) and Aas et al. (2007).

Consider again the joint distribution F with density f and with strictly continuous marginal c.d.f.s $F_1, \ldots, F_d$ with densities f_i. First note that any multivariate density function may be uniquely decomposed as

$$f(x_1, \ldots, x_d) = f_d(x_d) \cdot f(x_{d-1}|x_d) \cdot f(x_{d-2}|x_{d-1}, x_d) \cdots f(x_1|x_2, \ldots, x_d). \quad (4)$$

The conditional densities in (4) may be written as functions of the corresponding copula densities. That is, for every j,

$$f(x \mid v_1, v_2, \ldots, v_d) = c_{xv_j|\mathbf{v}_{-j}}\big(F(x \mid \mathbf{v}_{-j}), F(v_j \mid \mathbf{v}_{-j})\big) \cdot f(x \mid \mathbf{v}_{-j}), \quad (5)$$

where $\mathbf{v}_{-j}$ denotes the d-dimensional vector $\mathbf{v}$ excluding the jth component. Note that $c_{xv_j|\mathbf{v}_{-j}}(\cdot,\cdot)$ is a *bivariate* marginal copula density. For example, when $d=3$,

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

and

$$f(x_2|x_3) = c_{23}\big(F(x_2), F(x_3)\big)\cdot f(x_2).$$

Expressing all conditional densities in (4) by means of (5), we derive a decomposition for $f(x_1,\dots,x_d)$ that consists of only univariate marginal distributions and bivariate copulas. Thus we obtain the *pair-copula decomposition* for the d-dimensional copula $c_{1\cdots d}$, a factorization of a d-dimensional copula based only on bivariate copulas. This is a very flexible and natural way of constructing a higher dimensional copula. Note that, given a specific factorization, there are many possible reparameterizations.

The conditional c.d.f.s necessary for pair-copulas construction are given (Joe 1996) by

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

For the special case (unconditional) when v is univariate, and x and v are standard uniform, we have

$$F(x\mid v) = \frac{\partial C_{xv}(x,v,\Theta)}{\partial v}$$

where Θ is the set of copula parameters.

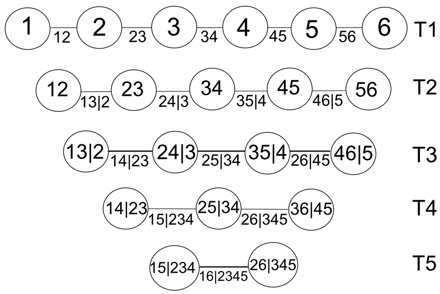
For large d, the number of possible pair-copula constructions is very large. As shown in Bedford and Cooke (2001), there are 240 different decompositions when $d=5$. These authors introduce a systematic way to obtain the decompositions, which involves graphical models that they call *regular vines*. They also aid in understanding the conditional specifications made for the joint distribution. Special cases are the hierarchical canonical vines (C-vines) and the D-vines. Each of these graphical models is a specific way of decomposing the density $f(x_1,\dots,x_d)$. For example, for a D-vine, $f()$ is equal to

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

In a D-vine, there are $d-1$ hierarchical *trees* with increasing conditioning sets, and there are $d(d-1)/2$ bivariate copulas. For a detailed description, see Aas et al. (2007). Figure 1 shows the D-vine decomposition for $d=6$. The D-vine consists of five nested trees, where tree T_j has $7-j$ nodes and $6-j$ edges corresponding to a pair-copula.

It is not essential that all the bivariate copulas involved belong to the same family. This flexibility is exactly what we are searching for, since our objective is to construct

Fig. 1 The D-vine graphical hierarchical representation of a six-dimensional pair-copula



(or estimate) a multivariate distribution that best represents the data at hand, which might be comprised of completely different margins (symmetric, asymmetric, with different dynamic structures, and so on) and, more importantly, could be pairwise joined by more complex dependence structures possessing linear and/or nonlinear forms of dependence, including tail dependence, or could be joined independently.

For example, one may combine the following types of (bivariate) copulas: Gaussian (no tail dependence, elliptical); t-student (equal lower and upper tail dependence, elliptical); Clayton (lower tail dependence, Archimedean); Gumbel (upper tail dependence, Archimedean); and BB7 (different lower and upper tail dependence, Archimedean). See Joe (1997) for a copula catalogue.

Simulations from both Canonical and D-vine pair-copulas can be easily implemented and take very little time to run. Maximum likelihood estimators depend on (i) the choice of factorization and (ii) the choice of pair-copula families. Algorithm implementation is straightforward. For smaller dimensions, we may compute the log-likelihood of all possible decompositions. For $d >= 5$, and for a D-vine, a specific decomposition may be chosen. One possibility is to look for the pairs of variables having the stronger tail dependence, and let those determine the decomposition to estimate. To this end, a t-copula may be fitted to all pairs and pairs would be ranked according to the smallest number of degrees of freedom.

3 Pair-copulas based unconditional modeling of log-returns

We present applications using a data set based on a global portfolio of an emerging market investor located in Brazil. We chose this perspective because of the higher volatility of Latin American stock markets and their greater potential for interdependence with the major markets. Thus we use six-dimensional contemporaneous daily log-returns comprised of: (1) a Brazilian composite hedge fund index (the ACI, Arsenal Composite Index); (2) a long-term inflation-indexed Brazilian treasury bonds index (the IMAC index, computed by the Brazilian Association of Financial Institutions, Andima); (3) a Brazilian stock index that includes the 100 largest capitalization companies (IBRX); (4) an index of large world stocks computed by MSCI (WLDLG); (5) an index of small capitalization world companies computed by MSCI (WLDSM); and (6) an index of total returns on US Treasury bonds computed by Lehman Brothers Barra (LBTBOND). All daily log-returns in US dollars are shown in Fig. 2. There are 1629 six-dimensional observations from January 2, 2002 to October 20, 2008.

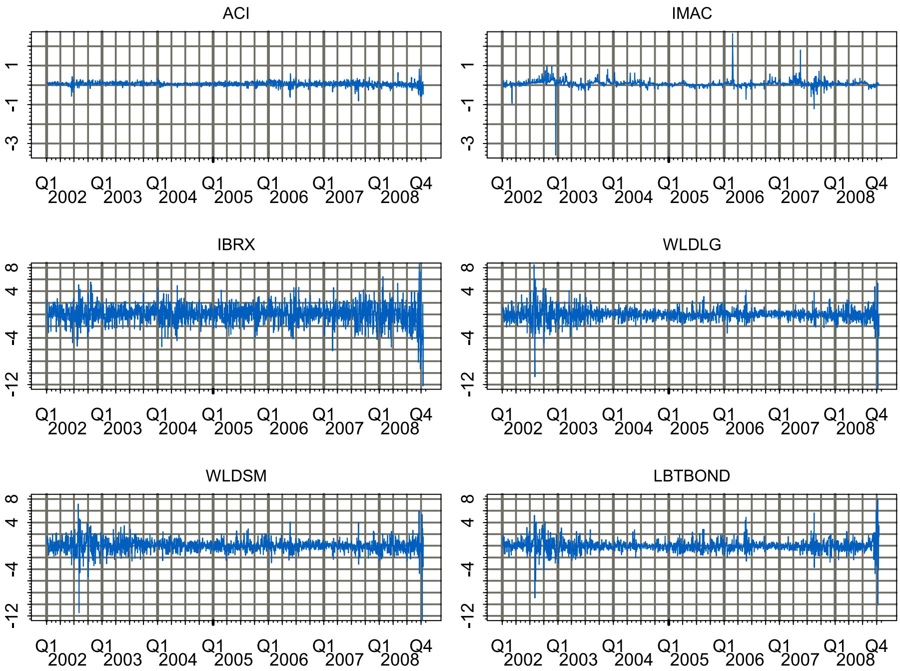


Fig. 2 Time series plot of daily log-returns used: ACI, IMAC, IBRX, WLDLQ, WLDSM, LBTBOND

According to Fig. 2 and Table 1, the hedge fund indexes ACI and IMC have less volatility compared to the long-run treasury bond index LBTBOND.

In this section, we take the unconditional approach to modeling the margins. The fits will be based on the flexible skew-t distribution (Hansen 1994), previously used by Patton (2006), Rockinger and Jondeau (2001), and Fantazzini (2006) in the context of dynamic modeling. This distribution generalizes the widely used normal distribution and its most common alternative, the t-student distribution, providing great flexibility since it covers left and right skewness and heavy tails.

The skew-t density has a closed form and implementation of the maximum likelihood method is feasible because there are only four parameters to estimate (μ, λ, ν, σ). Parameter μ equals the population mean and parameter σ equals the standard deviation (which exists when $\nu > 2$). When the skewness parameter λ is 0, the symmetric case is recovered. The mode of the skew-t distribution is smaller (larger) than μ in the case of right (left) skewness. Figure 3 shows the skew-t density for $\nu = 4$ and $\lambda = -0.6, 0.0, 0.6$.

The zero mean unit variance skew-t c.d.f. (see Fantazzini 2006) is given by

$$G(y;\nu,\lambda) = \begin{cases} (1-\lambda) G_T\Big(\sqrt{\frac{\nu}{\nu-2}}\big(\frac{by+a}{1-\lambda}\big);\nu\Big), & \text{for } y < -\frac{a}{b}, \\ (1-\lambda)/2, & \text{for } y = -\frac{a}{b}, \\ (1+\lambda) G_T\Big(\sqrt{\frac{\nu}{\nu-2}}\big(\frac{by+a}{1+\lambda}\big);\nu\Big) - \lambda, & \text{for } y > -\frac{a}{b}, \end{cases} \quad (6)$$

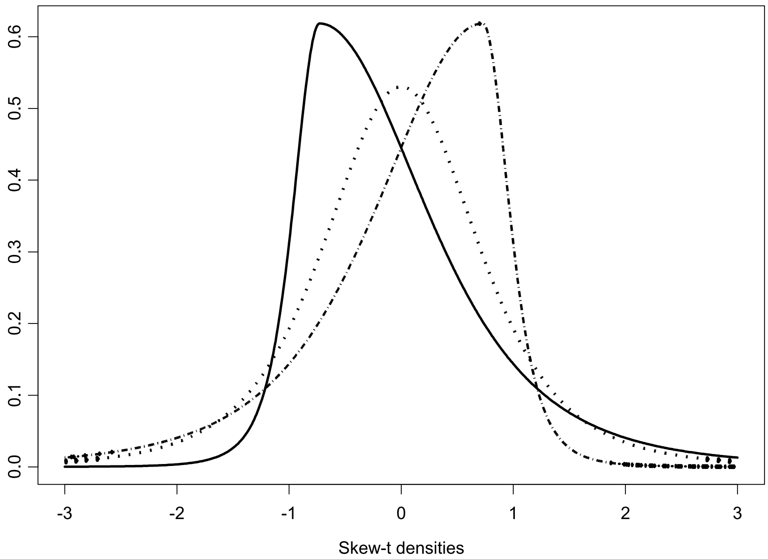


Fig. 3 Skew-t standard densities ($\mu = 0$, $\sigma = 1$) with $\nu = 4$, and $\lambda = -0.6, 0.0, 0.6$, shown respectively by: *solid*, *dotted*, and *dashed lines*

where $G_T(t;\nu)$ represents the c.d.f. of the symmetric t-student with ν degrees of freedom, and where

$$c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}},$$

$$b = \sqrt{1 + 3\lambda^2 - a^2},$$

$$a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right).$$

The maximum likelihood estimates of the skew-t distributions fitted to each variable are given in Table 1. The table also sets out the classic sample estimates of location and standard deviation, actually, maximum likelihood estimates under the univariate normal distribution.

Using the skew-t c.d.f. (6), the six transformed standard uniform series are obtained and used to estimate the pair-copulas. We estimate a D-vine, and to choose the variables in tree 1, we examined the scatterplots of all pairs and ranked the pairs according to the smallest number of degrees of freedom associated with a t-copula fit.

Having decided on the order of variables in tree 1, the D-vine decomposition follows. To estimate the pair-copulas (5 unconditional and 10 conditional), we considered as possible candidates four copula families: Normal, t-student, BB7, and the product copula, which models independence. These four copula families cover all the desired types of lower and upper tail dependence. Step-by-step instructions on how to perform the estimation are given in Aas et al. (2007).

Table 1 Skew-t parameters' estimates and sample estimates of location and scale

Estimates	ACI	IMAC	IBRX	WLDLG	WLDSM	LBTBOND
Sample mean	0.0645	0.0821	0.0738	−0.0048	0.0088	0.0186
Skew-t $\widehat{\mu}$	0.0655	0.0828	0.0745	−0.0065	0.0143	0.0249
Skew-t $\widehat{\lambda}$	−0.1522	0.1514	−0.1437	0.0445	0.0269	0.1163
Skew-t $\widehat{\nu}$	3.3170	2.1200	5.3971	2.9782	3.8503	2.9734
Skew-t $\widehat{\sigma}$	0.1159	0.2716	1.6920	1.2008	1.1228	1.1373
Sample st. deviation	0.1131	0.1853	1.6930	1.1638	1.1629	1.0891

To find the best copula fit, we compared the penalized log-likelihood (AIC), examined the pp-plots based on the estimated and the empirical copula, and computed a GOF test statistic. Any one of the GOF tests mentioned in Sect. 2.1 would be appropriate and, indeed, there is no general agreement as to the best GOF test for copulas. We used the one suggested by Genest and Rémillard (2005) and Genest et al. (2007). This test is based on the squared distance between the estimated and the empirical copulas. The limiting distribution of the test statistic depends on the parameter values and approximate p-values are obtained through bootstrap sampling.

The chosen pair-copula decomposition, along with best copula fits and parameter estimates, is shown in Fig. 4. In the case of the t-copula, the parameters are the correlation coefficient and the degrees of freedom, (ρ, υ). Even though for some t-copulas fitted, the number of degrees of freedom is large, we rejected the Gaussian copula based on the AIC values. The upper and lower tail dependence coefficients computed for each estimated bivariate copula are also shown in Fig. 4. Joe (1997) gives the formula for the tail dependence coefficient for several families. In the case of the t-copula, see Embrechts et al. (2001). We note that such accurate and tailored estimation of the data dependence structure, particularly its complex pattern of dependence in the tails, would not be possible using a d-dimensional copula. We observe in tree 1 that the stock indexes of large and small world companies show the strongest dependence during stressful times.

All copulas in tree 1 have positive upper and lower tail dependence coefficients and, according to Joe et al. (2010), this means that the D-vine modeling of the log-returns has multivariate upper and lower tail dependence.

As suggested by a referee, to illustrate the role of the t-copula when producing joint large values, we also compute the joint quantile exceedance probabilities (Demarta and McNeil 2005) for a selection of the bivariate copulas shown in Fig. 4. We compute $C(u, u)$, the probability of simultaneous quantile exceedances, for $u = (0.05, 0.01, 0.005, 0.001)$ using the fitted t-copulas and compare these values with the values obtained for the Gaussian copula. In Table 2 we report the values of C assuming C is Gaussian, and report the ratio between these exceedance probabilities computed under the t and the Gaussian assumptions. As pointed out in Demarta and McNeil (2005), these ratio values increase as the correlation parameter approaches -1 or as the number of degrees of freedom decreases.

For the applications that follow, we need to compute the unconditional rank correlation matrix. To obtain the rank correlation coefficients, we use (3) and the estimated

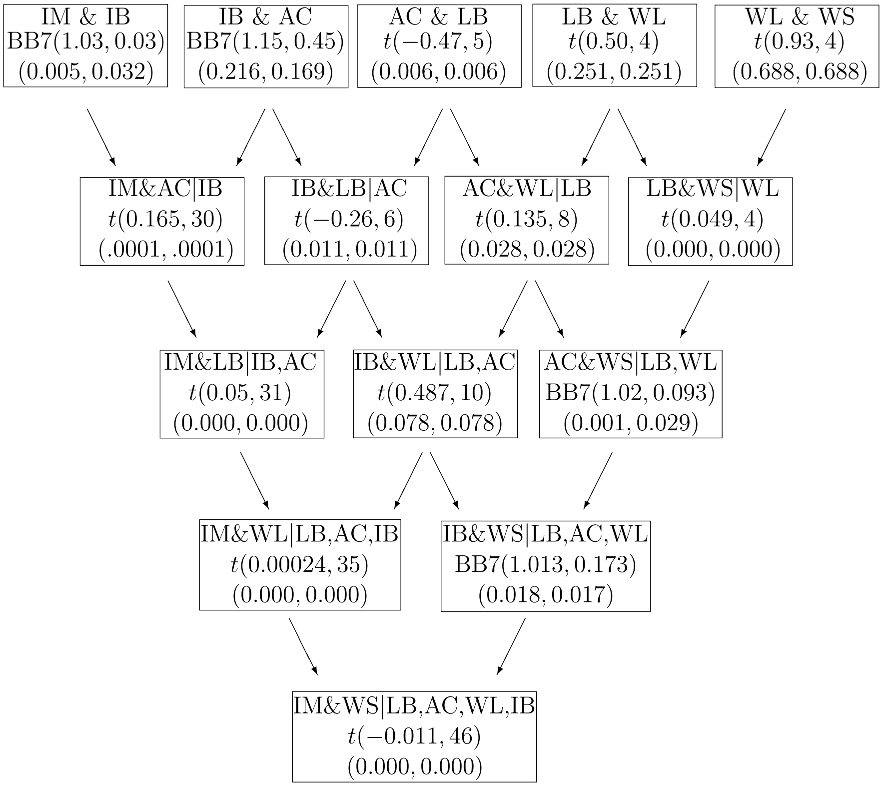


Fig. 4 D-vine decomposition. Best copula fits along with their parameters' estimates, which in the case of the t-copula are (ρ, υ). The *third row* inside the *box-table* gives the (λ_L, λ_U) estimates. Notation in figure: IM: IMAC; AC: ACI; IB: IBRX; LB: LBTBOND; WL: WLDLG; WS: WLDSM

pair-copulas. The fits in tree 1 result in five unconditional rank correlations. From fits in trees 2 through 5, we obtain conditional rank correlations. These conditional rank correlations are considered constant, they do not depend on the value of the conditioning variables, as proved in Kurowicka and Cooke (2001) for elliptical copulas and copulas in general. This leads to another important result in Misiewicz et al. (2001): for elliptical copulas, conditional linear and conditional rank correlations are equal, provided the conditional correlations are constant (recall that, by definition, for copulas, the unconditional linear and rank correlations are equal).

An important issue is the relation between conditional rank correlation and partial correlations. Partial rank correlations are defined in Yule and Kendall (1965). Their importance is stressed in Cooke and Bedford (1995), where the authors show that there is a one-to-one relation between partial correlations on a D-vine and correlation matrices. Kurowicka and Cooke (2001) show that the D-vine partial correlation matrix obtained from the fits uniquely determines the correlation matrix, and that every full rank correlation matrix may be decomposed in this way (Bedford and Cooke 2002).

Table 2 Joint quantile exceedance probabilities for bivariate Gaussian copula and a selection of the t copulas fitted. For Gaussian copula the probability of joint quantile exceedance is given; for the t copulas the factors by which the Gaussian probability must be multiplied are given

ρ	Copula	Fig. 4	Quantile			
			0.05	0.01	0.005	0.001
0.93	Gaussian	–	3.476×10^{-2}	0.612×10^{-2}	0.291×10^{-2}	0.052×10^{-2}
0.93	t_4	WL&WS	1.065	1.161	1.209	1.332
0.165	Gaussian	–	0.467×10^{-2}	0.028×10^{-2}	0.008×10^{-2}	0.001×10^{-2}
0.165	t_{30}	IM&AC|IB	1.128	1.447	1.676	2.551
−0.255	Gaussian	–	0.624×10^{-2}	0.045×10^{-2}	0.015×10^{-2}	0.001×10^{-2}
−0.255	t_6	IB&LB|AC	1.508	2.903	4.048	9.398
0.135	Gaussian	–	0.421×10^{-2}	0.024×10^{-2}	0.007×10^{-2}	0.000×10^{-2}
0.135	t_8	AC&WL|LB	1.525	3.123	4.544	12.020
0.00024	Gaussian	–	0.250×10^{-2}	0.010×10^{-2}	0.003×10^{-2}	0.000×10^{-2}
0.00024	t_{35}	IM&WL|LB,AC,IB	1.167	1.622	1.971	3.456

Table 3 Pair-copulas rank correlations and sample correlations

	ACI	IMAC	IBRX	WLDLG	WLDSM	LBTBOND
ACI	1.000	0.185	0.342	−0.117	−0.088	−0.453
IMAC	0.211	1.000	0.110	0.019	0.020	−0.054
IBRX	0.435	0.112	1.000	0.223	0.240	−0.360
WLDLG	−0.093	0.022	0.197	1.000	0.924	0.479
WLDSM	−0.080	0.024	0.197	0.938	1.000	0.458
LBTBOND	−0.465	−0.069	−0.373	0.571	0.596	1.000

Above the diagonal: pair-copulas rank correlations. Below the diagonal: sample correlations

Kurowicka and Cooke (2001) prove the equality of constant conditional correlations and partial correlations for elliptical and other copulas. They also study the relation between the conditional correlation and the conditional rank correlation.

All the above-cited results lay the foundation for (7), which links the partial and the unconditional correlations,

$$\rho_{12;3\ldots d} = \frac{\rho_{12;4\ldots d} - \rho_{13;4\ldots d}\rho_{23;4\ldots d}}{\sqrt{(1-\rho^2_{13;4\ldots d})(1-\rho^2_{23;4\ldots d})}}, \tag{7}$$

to inductively compute the unconditional (rank or linear) correlations. The final unconditional rank estimates are given in Table 2 above the diagonal.

Finally, to validate the fits, we simulated 2000 observations from the fitted D-vine using the algorithm given in Aas et al. (2007). We compute the sample rank correlations from the simulated data and compare the obtained values with those given above the diagonal in Table 2. The sum of the squares of the differences between the 15 rank correlations is 0.0110123, which validates the fit.

3.1 Application: efficient frontiers

Constructing the efficient frontier (EF) corresponding to optimal portfolios according to the Markowitz mean variance methodology (MV) only requires point estimates for the means, variances, and the linear correlation coefficients as inputs in the quadratic optimization problem. The most frequently used inputs are the classical sample mean and sample covariance matrix (hereafter referred to as the *classical* approach). The classical approach is very suitable if the data are not derived from a multivariate normal distribution, which, as is well known, is usually not the case for log-returns data. Outside the Gaussian world, the classical estimates lose efficiency and may become biased (Hampel et al. 1986), making the results somewhat doubtful.

Moreover, it would be interesting to measure dependence beyond correlations and capture all the linear and nonlinear associations among the portfolio components, and then incorporate them in the MV methodology. It would also be desirable to assess the variability of the efficient frontiers. All this can be accomplished by accurately estimating the multivariate distribution implied by the underlying assets, a task fulfilled in the previous subsection by using pair-copulas.

Scherer and Martin (2005) obtain robust versions of Markowitz mean variance optimal portfolios using some well-known robust estimates of covariance such as Rousseeuw's MCD (see Rousseeuw and Leroy 1987). Other robust alternatives are proposed, for example, by Mendes and Leal (2005). Ragea (2003) investigates whether highly volatile markets produce disruptions in the joint move of the risk factors. However, all the above-cited alternatives differ only with regard to how the inputs are estimated, particularly the linear correlation coefficient.

In this application, we use the location and standard deviation estimates provided by the skew-t model and the rank correlations provided by the pair-copula decomposition. We call these inputs "pair-copula-based estimates." Figure 5 shows the ellipsoids associated with the rank and sample correlations. For this particular data set there are no striking differences (e.g., a change of sign) between the correlation estimates even though, as we shall see, the resulting efficient frontiers will be quite different.

Using the two sets of inputs estimated across the entire period (approximately seven years), we run the long-only MV optimization algorithm and construct the classical and the pair-copula-based efficient frontiers (EF) containing 20 optimal linear combinations of the six series of log-returns. These are shown on the left-hand side of Fig. 6. We observe that the classical EF is below and to the right of the pair-copula-based efficient frontier.

An appealing feature of the pair-copula modeling strategy is that it allows for simulations of the data-fitted distribution, providing replications of any quantity of interest. Here, we compute parametric replications of the pair-copula-based efficient frontier. Let $\widehat{\mathbf{r}}$ represent the set of all rank correlations estimates r_{ij}, $i, j = 1, \ldots, 6$. Let θ represent the set of parameters from the pair-copula decomposition, and let $\widehat{\theta}$ represent their estimates. Let δ represent the set of parameters from the skew-t distribution, $\delta = (\mu, \lambda, \nu, \sigma)$, and let $\widehat{\delta}$ represent their estimates. For the generation, we assume that $\widehat{\theta}$ and $\widehat{\delta}$ are the true parameter values and implement the following parametric bootstrap algorithm:

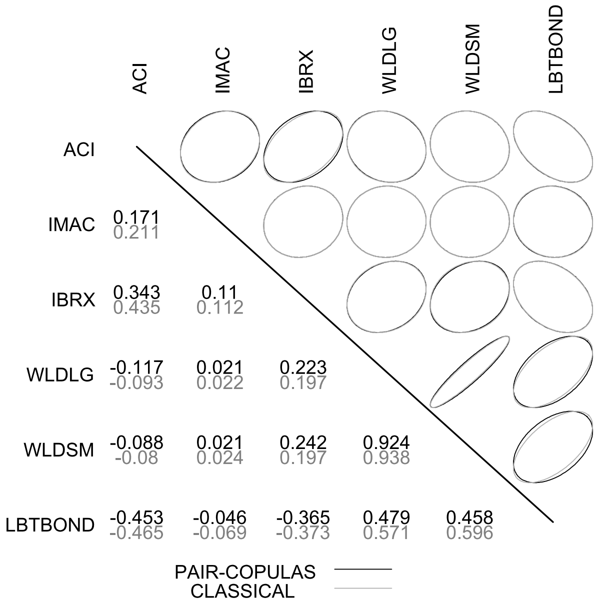


Fig. 5 Unconditional rank correlations and sample correlations

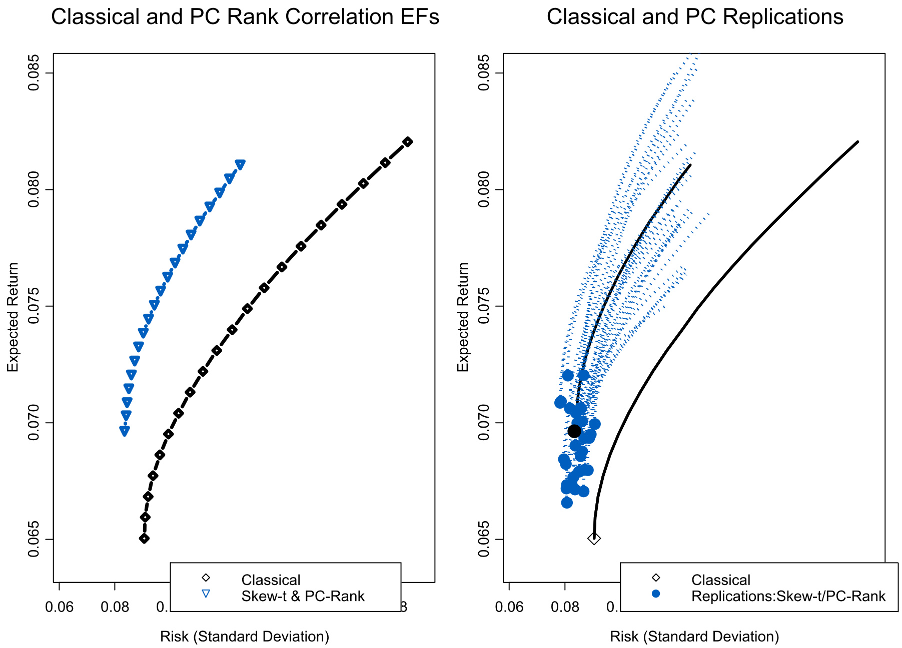


Fig. 6 (Color online) Long-only efficient frontiers. On the *left-hand side*, classical in *black*, and pair-copula-based (skew-t location estimates and pair-copula rank correlations) in *blue*. On the *right-hand side*, replications of the pair-copula-based EF

For $k = 1, \ldots, B$, with B large,

1. Using the algorithm given in Aas et al. (2007), simulate 1629×6 observations from the estimated pair-copula assuming $\widehat{\theta}$ as the true value.
2. Apply the corresponding inverses of the skew-t c.d.f.s to each margin, assuming $\widehat{\delta}$ as the true value, obtaining a six-dimensional sample $\mathbf{X}^{(k)}$, a replication of the original data.
3. Apply the entire estimation procedure (marginal and pair-copula fits) on $\mathbf{X}^{(k)}$, obtaining a new set of inputs $\widehat{\mu}^{(k)}$, $\widehat{\sigma}^{(k)}$, and $\widehat{\mathbf{r}}^{(k)}$, for construction of the efficient frontier $\mathrm{EF}^{(k)}$ composed of 20 long-only portfolios.

The right-hand side of Fig. 6 shows the original pair-copula-based EF and its parametric replications, along with the classical EF. The filled circles correspond to minimum risk portfolios. The set of all replications of some specific portfolio (in the figure, the number 1) may be used to define a $(1-\alpha)\%$ confidence level convex hull containing statistically equivalent portfolios. It is also possible to draw replications of the classical EF to verify if the corresponding convex hulls have an intersection. This would be useful for portfolio rebalancing and testing. Mendes and Leal (2009) propose a method for replicating the classical EF and use square distances to test equality of portfolios.

Figure 7 shows the boxplots of the weights from the replications for portfolios ranked 1 (P.1) and 7 (P.7), and for each variable. As expected, portfolios with less

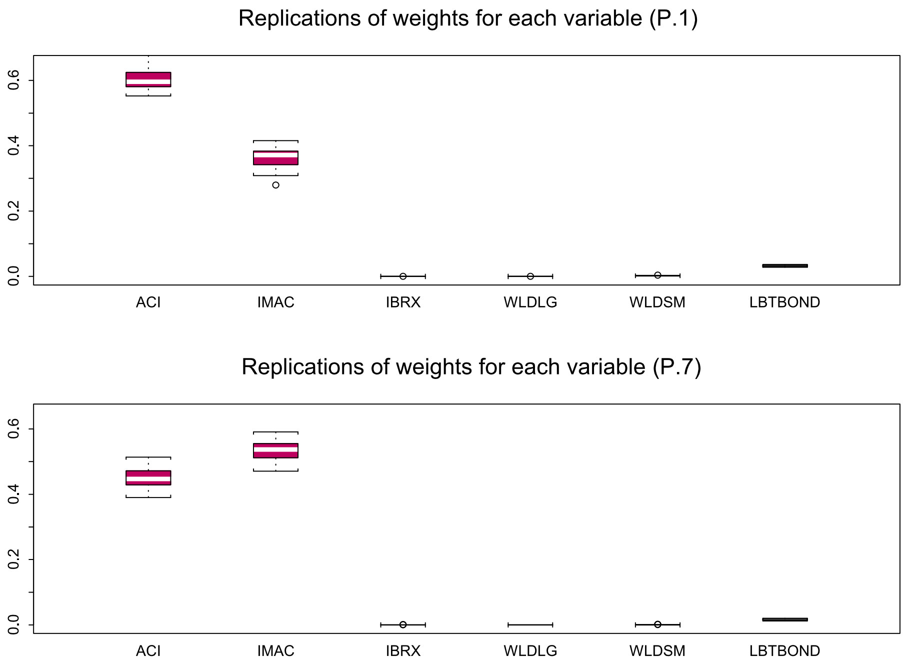


Fig. 7 Replications of weights for each variable and for portfolios ranked 1 and 7 based on the pair-copulas model

risk show less variability in the plane risk × return (Fig. 6), and are more stable in the d-dimensional space of the weights (Fig. 7). Actually, the stability of weights of a given rank portfolio, over the convex hull of replications, simply confirms that in general equivalent portfolios showing different return × risk values have similar weights composition. Yet, the utility of an EF construction is not the return/risk values of the portfolio, but their weight compositions.

4 Pair-copula-based conditional modeling of log-returns

Log-returns typically present temporal dependences in the mean and in the volatility. In this section, we first process the data using an ARFIMA-FIGARCH filter to obtain the standardized residuals, and then apply all the estimation steps of the previous section to the filtered data.

Let r_t represent the return at day t. The models specification is

$$\begin{aligned} r_t &= \mu_t + \sigma_t \epsilon_t, \\ \mu_t &= \phi_0 + \sum_{j=1}^{p} \phi_j r_{t-j} + \sum_{i=1}^{q} \theta_i \mu_{t-i}, \\ \sigma_t^2 &= \alpha_0 + \sum_{j=1}^{m} \alpha_j r_{t-j}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2, \\ E[\epsilon_t] &= 0, \qquad \text{var}(\epsilon_t) = 1. \end{aligned}$$

For each series of log-returns we fit the best ARMA-FIEGARCH model, considering as conditional distributions either the Normal or the t_ν, where ν is the number of degrees of freedom. Table 4 sets forth the estimates.

We compute the standardized residuals from the d univariate fits. The d filtered series are now free of temporal dependences in the first and second moments. To these i.i.d. series we apply the unconditional approach of the previous section.

Table 5 sets out the estimates of parameters of the skew-t distribution fitted to the marginal estimated innovations series. Of course, the μ estimates are close to 0 and the σ estimates are close to 1. Although the series still exhibits skewness and kurtosis, they are smaller than those estimated for the raw data.

The D-vine is fitted to the transformed standard uniform data obtained from the residuals from the skew-t fit. Recall that under the conditional approach, the pair-copula represents the dependence structure of the d-dimensional errors distribution, which is free of temporal dependences. The copula families found as best fits are practically the same as those shown in Fig. 4, the only difference being the conditional copula of IMAC & ACI given IBRX, which is now Gaussian, although all show smaller tail dependence. The unconditional copulas in tree 1, for example, have lower and upper tail dependence coefficients equal to $(0.104, 0.101)$ for the IBRX & ACI, $(0.149, 0.149)$ for LBTBOND & WLDLG, and $(0.505, 0.505)$ for WLDLG & WLDSM. Likewise, in the unconditional case, the multivariate distribution of the filtered data has tail dependence.

Table 4 Maximum likelihood estimates (standard errors) of the ARMA-FIEGARCH models fitted to the log-returns series

Parameter	ACI	IMAC	IBRX	WLDLG	WLDSM	LBTBOND
ϕ_0	0.0694 (0.002)	0.0648 (0.002)	0.0941 (0.041)	0.0011 (0.018)	0.0178 (0.022)	−0.0328 (0.019)
MA(1)	–	–	0.0784 (0.027)	0.0183 (0.006)	0.0808 (0.026)	0.1225 (0.027)
α_0 –	−0.1961 (0.025)	0.0026 (0.000)	−0.0797 (0.017)	0.1275 (0.019)	0.0308 (0.010)	0.0305 (0.007)
α_1 –	0.2237 (0.027)	0.7169 (0.077)	0.1155 (0.024)	0.8641 (0.019)	0.1338 (0.021)	0.1840 (0.026)
β_1	0.7695 (0.078)	0.2623 (0.036)	0.8120 (0.092)	–	0.8472 (0.021)	0.7721 (0.024)
Leverage term	−0.0267 (0.010)	–	−0.0634 (0.016)	–	–	0.2692 (0.064)
Fraction d	0.4533 (0.084)	–	0.3751 (0.125)	–	–	–
ϵ Distribution	Normal	t_4	Normal	t_7	t_8	t_{15}
AIC criterion	−3094.533	−2808.539	6049.564	4262.349	4450.753	3878.671

Table 5 Skew-*t* parameters' estimates and sample estimates of location and scale for the standardized residuals from the GARCH fits

Estimates	ACI	IMAC	IBRX	WLDLG	WLDSM	LBTBOND
Sample mean	−0.0007	0.0731	−0.0017	−0.0066	−0.0100	0.0309
Skew-*t* $\widehat{\mu}$	0.00003	0.0749	−0.0005	−0.0070	−0.0110	0.0300
Skew-*t* $\widehat{\lambda}$	−0.0995	0.0536	−0.1288	0.0442	−0.0033	0.0599
Skew-*t* $\widehat{\nu}$	6.8976	3.0000	13.2032	6.9746	8.3890	15.4386
Skew-*t* $\widehat{\sigma}$	0.9980	1.1352	0.9961	1.0005	0.9973	1.0029
Sample st. deviation	0.9970	1.4722	0.9970	1.0003	1.0009	1.0029

Table 6 Pair-copulas rank correlations (above the diagonal) and sample correlations (below the diagonal)

	ACI	IMAC	IBRX	WLDLG	WLDSM	LBTBOND
ACI	1.000	0.140	0.244	−0.134	−0.113	−0.434
IMAC	0.134	1.000	0.114	0.009	0.012	−0.062
IBRX	0.291	0.097	1.000	0.242	0.244	−0.337
WLDLG	−0.150	0.020	0.227	1.000	0.918	0.428
WLDSM	−0.118	0.025	0.239	0.923	1.000	0.411
LBTBOND	−0.451	−0.049	−0.364	0.449	0.435	1.000

Next, we compute the rank correlation matrix, which can be found in Table 6. We observe that, for most pairs, the values of the correlation coefficients (pair-copula-based) are slightly smaller, demonstrating that the volatility is in many cases responsible for the contagion and that it may (slightly) increase dependence.

4.1 Efficient frontier

We run the long-only MV algorithm and construct the 20 portfolios that comprise the efficient frontiers. As expected, the position on the risk × return plane of the efficient frontier associated with the filtered data is quite different from the one based on the original data. This is mostly due to changes in the means and standard deviation estimates. However, it would be interesting to examine the portfolio compositions to assess the effect of volatility on the weights.

Figure 8 shows the weights of the 20 portfolios in the pair-copula-based efficient frontiers for both the filtered and the original data. Tests using the weights and based on distances could be applied to test the equality of some compositions.

5 Conclusions

In this paper we explored the potential of pair-copulas modeling using dependent financial data. A fully flexible multivariate distribution was obtained by combining univariate fits and D-vines.

Our marginal specifications included the asymmetric and high kurtosis unconditional skew-*t* probabilistic models, as well as conditional models, combined with

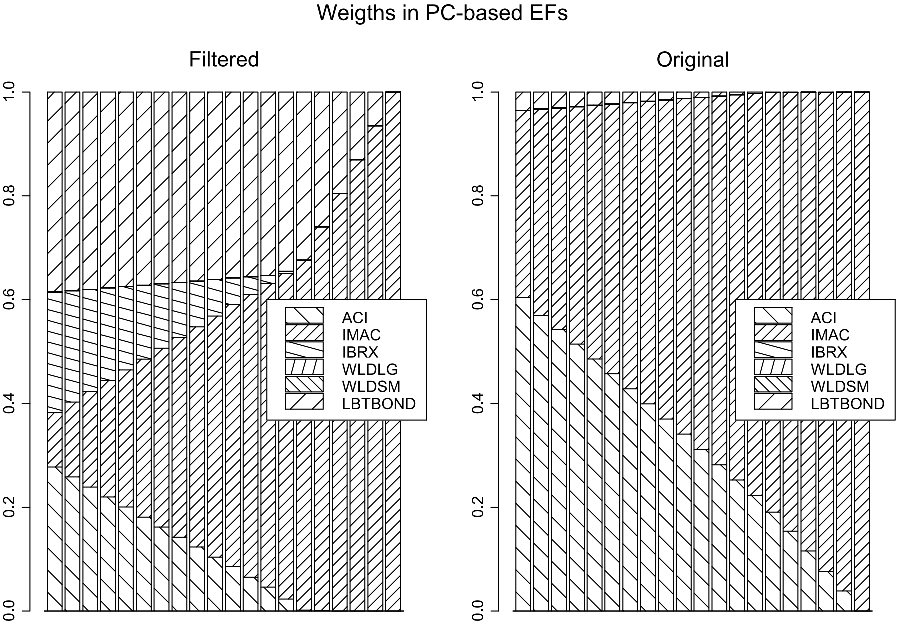


Fig. 8 Weights composing the 20 portfolios for the efficient frontiers based on the filtered and original data

pair-copulas having different upper and lower tail dependence. This results in a powerful model that can accurately estimate any quantity of interest, such as optimal portfolios and risk measures. An examination of the tail loss distributions shows that substantial differences result from the flexible pair-copula specification. Moreover, parametric replications of the fitted multivariate distribution may be used to assess variability of the estimates.

The pair-copulas field would benefit from research into how to choose among copula families and decompositions, and from the development of more powerful goodness-of-fit tests. Also of interest and practical use would be investigation into the potential of time-varying pair-copulas.

Acknowledgements The authors thank the Editor and an anonymous referee for their valuable comments and suggestions, which have greatly improved the paper. The authors gratefully acknowledge Brazilian financial support from CAPES, CNPq, and COPPEAD research grants.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insur.: Math. Econ. **2**, 1–25 (2007)

Bedford, T., Cooke, R.M.: Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. Intell. **32**, 245–268 (2001)

Bedford, T., Cooke, R.M.: Vines—a new graphical model for dependent random variables. Ann. Stat. **30**, 1031–1068 (2002)

Berg, D., Aas, K.: Models for construction of multivariate dependence: a comparison study. Eur. J. Finance, forthcoming (2008)

Breyman, W., Dias, A., Embrechts, P.: Dependence structures for multivariate high-frequency data in finance. Quant. Finance **3**, 1–16 (2003)
Cooke, R.M., Bedford, T.: Reliability Methods as Management Tools: Dependence Modeling and Partial Mission Success. Chameleon Press, London (1995)
Demarta, S., McNeil, A.: The t Copula and related Copulas. Int. Stat. Rev. **73**, 111–129 (2005)
Embrechts, P., McNeil, A., Straumann, D.: Correlation and dependence in risk management: properties and pitfalls. Preprint ETH Zurich (1999)
Embrechts, P., McNeil, A., Straumann, D.: Correlation and dependency in risk management: properties and pitfalls. In: Value at Risk and Beyond, pp. 176–223. Cambridge University Press, Cambridge (2001)
Fantazzini, D.: Dynamic copula modelling for value-at-risk. Frontiers in Finance and Economics. Available at SSRN: http://ssrn.com/abs=944172 (2006)
Fermanian, J.-D.: Goodness-of-fit tests for copulas. J. Multivar. Anal. Arch. **95**(1), 119–152 (2005)
Fischer, M., Köck, C., Schlüter, S., Weigert, F.: Multivariate copula models at work: outperforming the 'Desert Island Copula'? Universität Erlangen (2008)
Gatzert, N., Schmeiser, H., Schuckmann, S.: Enterprise risk management in financial group: analysis of risk concentration and default risk. Financ. Mark. Portf. Manag. **22**(3), 241–258 (2008)
Genest, C., Rémillard, B.: Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Technical Report G-2005-51, GERAD, Montreal, Canada (2005)
Genest, C., Quessy, J.F., Rémillard, B.: Goodness-of-fit procedures for copula models based on the probability integral transformation. Scand. J. Stat. **33**, 337–366 (2006)
Genest, C., Rémillard, B., Beaudoin, D.: Omnibus goodness-of-fit tests for copulas. A review and a power study. Working paper, Université Laval (2007)
Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J.: Robust Statistics: The Approach Based on Influence Functions. Willey, New York (1986)
Hansen, B.: Autoregressive conditional density estimation. Int. Econ. Rev. **35**, 705–730 (1994)
Joe, H.: Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In: Rüchendorf, L., Schweizer, B., Taylor, M.D. (eds.) Distributions with Fixed Marginals and Related Topics (1996)
Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
Joe, H., Li, H., Nikoloulopoulos, A.K.: Tail dependence functions and vine copulas. J. Multivar. Anal. **101**, 252–270 (2010)
Kassberger, S., Kiesel, R.: A fully parametric approach to return modeling and risk management of hedge funds. Financ. Mark. Portf. Manag. **20**, 472–491 (2006)
Kurowicka, D., Cooke, R.M.: Conditional, partial, and rank correlation for the elliptical copula: dependence modeling in uncertainty analysis. In: Proceedings ESREL (2001)
Kurowicka, D., Cooke, R.M.: Completion problem with partial correlation vines. Linear Algebra Its Appl. **418**, 188–200 (2006)
Mendes, B.V.M., Leal, R.P.C.: Robust multivariate modeling in finance. Int. J. Manag. Finance **1**, 95–106 (2005)
Mendes, B.V.M., Leal, R.P.C.: Portfolio management with semi-parametric bootstrapping. J. Risk Manag. Financ. Inst. **3**, 174–183 (2009)
Mendes, B.V.M., Melo, E.F.L., Nelsen, R.B.: Robust fits for copula models. Commun. Stat. **36**, 997–1017 (2007)
Min, A., Czado, C.: Bayesian inference for multivariate copulas using pair-copula constructions. J. Financ. Econ., forthcoming (2010)
Misiewicz, J., Kurowicka, D., Cooke, R.M.: Elliptical copulae. In: Schueller, Spanos (eds.) Monte Carlo Simulation. A.A. Balkema, Rotterdam (2001)
Patton, A.: Modeling asymmetric exchange rate dependence. Int. Econ. Rev. **47**, 527–556 (2006)
Ragea, V.: Testing correlation stability during hectic financial markets. Financ. Mark. Portf. Manag. **17**, 289–308 (2003)
Rockinger, M., Jondeau, E.: Conditional dependency of financial series: an application of copulas. In: HEC Paris DP 723
Rosenblatt, M.: Remarks on a multivariate transformation. Ann. Math. Stat. **23**, 470–472 (1952)
Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley, New York (1987)
Scott, D.W.: Multivariate Density Estimation. Wiley, New York (1992)
Scherer, B., Martin, R.D.: Introduction to Modern Optimization with Nuopt and SPlus. Springer, Berlin (2005)

Sklar, A.: Fonctions de répartition á n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris **8**, 229–231 (1959)
Tsukahara, H.: Semiparametric estimation in copula models. Can. J. Stat. **33**(3), 357–375 (2005)
Yule, G.U., Kendall, M.G.: An Introduction to the Theory of Statistics, 14th edn. Charles Griffin, London (1965)
Vandenhende, F., Lambert, P.: Local dependence estimation using semiparametric Archimedean copulas. Can. J. Stat. **33**(3), 377–388 (2005)

Beatriz Vaz de Melo Mendes Ph.D. in Statistics, Rutgers—The State University of New Jersey (1995), USA. Associate Professor at Institute of Mathematics/COPPEAD, Federal University at Rio de Janeiro, Brazil. She is author and co-author of 6 books in extremes, finance, risk and copulas, and has papers published in the Annals of Statistics, Journal of Derivatives, Computational Statistics, Journal of Statistical Computation and Simulation Journal of Risk, Applied Stochastic Models in Business and Industry, Journal of Actuarial Practice, among many others. Research areas include Financial Modeling and Risk Estimation, Robust Inference, Extremes, Copulas.

Mariângela Mendes Semeraro graduated in Applied Mathematics at the Federal University of Rio de Janeiro. During graduation studied in financial research projects. Currently, she is working as a consultant at the CPM Braxis in Rio de Janeiro, Brazil.

Ricardo P. Câmara Leal has taught at Georgetown University and at the University of Nevada. He was the dean of the Coppead Graduate School of Business where he is a Professor of Finance. His research and consulting cover emerging securities markets and corporate governance. He is a former president of the Brazilian Finance Society and of the Business Association of Latin American Studies. He is the editor of the Brazilian Finance Review and an editorial board member of many academic publications. He is a senior research fellow of the Brazilian Corporate Governance Institute.