Pair-Copulas Modeling in Finance

Beatriz Vaz de Melo Mendes

IM/COPPEAD, Federal University at Rio de Janeiro, Brazil.

Mariângela Mendes Semeraro

IM/COPPEAD, Federal University at Rio de Janeiro, Brazil.

Ricardo P. Câmara Leal

COPPEAD, Federal University at Rio de Janeiro, Brazil.

Abstract

This paper is concerned with applications of pair-copulas in finance, and bridges the gap between theory and application. We give a broad view of the problem of modeling multivariate financial log-returns using pair-copulas, gathering theoretical and computational results scattered among many papers on canonical vines. We show to the practitioner the advantages of modeling through pair-copulas and send the message that this is a possible methodology to be implemented in a daily basis. All steps (model selection, estimation, validation, simulations and applications) are given in a level reached by all data analysts.

Pair-Copulas Modeling in Finance

1 Introduction

The Basel II international capital framework has been, in some way, promoting the development of more sophisticated statistical tools for finance. Underlying each tool there is always a probabilistic model assumption. For a long time, modeling in finance would just mean considering the multivariate normal distribution. This 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 distributions and the restrictions imposed by the softwares available. However, a simple exploratory analysis carried on any collection of log-returns on indexes, stocks, portfolios, or bonds, will reveal significant departures from normality.

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

A solution, by now popular, is the use of copulas, introduced by Sklar (1959) and in finance by Embrechts et al. (1999). 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 also 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 solve the required optimization problem over many dimensions, the so called "curse of dimensionality" (Scott, 1992). Most of the available softwares deal only with the bivariate case. Even if we are able to fit a d-dimensional copula, $d > 2$, parametric copula families usually restrict all pairs to possess the same type or strength of dependence. For example, in the case of the t-copula, besides 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 violating (II).

Pair-copulas, being a collection of potentially different bivariate copulas, is a flexible and very appealing concept. The method for 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 parameters values. Therefore, all types and strengths of dependence may be covered. Pair-copulas are easy to estimate and to simulate being very appropriate for modeling in finance.

Most existing papers on pair-copulas deal with theoretical details on their construction and there still are some open questions. A few papers provide applications in finance, but most of them just fit a pair-copula to the data, see Min and Czado (2008), Aas, Czado, Frigessi, and Bakken (2007), Berg and Aas (2008), Fischer, Köck, Schlüter and Weigert (2008), among others.

In this paper we go beyond inference and provide 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 (Section 4). As unconditional univariate models we propose to use the very flexible skew-t family (Section 3). Estimation of the models is based on the maximum likelihood method. Applications will follow the fits, and they intend to show how the pair-copulas approach may be useful, since there are many applications in finance which rely on a good joint fit for the data. For example, computing risk measures estimates, finding portfolios' optimal allocations, pricing derivatives, and so on.

The main objective of this paper is to show at a practitioner's level, how pair-copulas modeling may be useful in finance. The paper also gathers a collection of important results which are scattered in many papers, 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-returns data may be obtained with the help of the 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, in particular for constructing efficient frontiers and computing the Value-at-Risk; (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, they allow for testing the equality of efficient portfolios and for testing if a portfolio re-balance is needed, or if the inclusion of some other component would significantly improve the expected return for the same risk level.

The remainder of this paper is organized as follows. In Section 2, we briefly review copulas and pair-copulas definitions. In Section 3 we consider the e unconditional approach for the marginal fits combined with the pair-copulas fit, and provide an application in 3.1, where we obtain optimal portfolios and show how to construct pair-copulas based replications of the efficient frontier. In Section 4 we take the conditional approach for the marginal fits, and in 4.1 we provide the same application carried in 3.1. Section 5 contains some concluding remarks.

2 Copulas and Pair-Copulas: a brief review

2.1 Copulas

Consider a stationary d-variate process $(X_{1,t}, X_{2,t}, \cdots, X_{d,t})_{t\in\mathcal{Z}}$, $\mathcal{Z}$ a set of indices. In the case the joint law of $(X_{1,t}, X_{2,t}, \cdots, X_{d,t})$ is independent of t, the dependence structure of $\mathbf{X} = (X_1, X_2, \cdots, X_d)$ 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, $i = 1, 2, \cdots, d$, then there exists a unique copula C pertaining to F, defined on $[0, 1]^d$ such that

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

holds for any $(x_1, x_2, \cdots, x_d) \in \Re^d$ (Sklar's theorem, Sklar (1959)).

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

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

for some d-dimensional copula density $c_{1 \cdots d}$. This decomposition allows for estimating the marginal distributions f_i separated from the dependence structure given by the d-variate copula. In practice, this fact simplifies both the *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 the copula based dependence measures interesting scale-free tools for studying dependence. For example, to measure monotone dependence (not necessarily linear) one may use the Spearman's rank correlation (r)

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

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 on of the Fréchet limit copulas.

Until recently, the Pearson's product moment (linear) correlation ρ was the quantity used to measure 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. Besides the drawback of measuring only *linear* correlation, ρ presents other weaknesses. A number of fallacies related to this quantity are by now well known, see, for example, Embrechts, McNeil, and Straumann (1999). Note that

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

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

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

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

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

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 it is assumed for modeling log-returns, for many pairs of variables it will underestimate joint risks.

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

$$\lambda_U = \lim_{u \uparrow 1} \frac{\overline{C}(u,u)}{1-u}, \text{ where } \overline{C}(u_1,u_2) = Pr\{U_1 > u_1, U_2 > u_2\} \text{ and } \lambda_L = \lim_{u \downarrow 0} \frac{C(u,u)}{u} \ .$$

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

Parametric estimation of copulas are usually accomplished in two steps, suggested by decomposition (2). In the first step, conditional (or unconditional) models are fitted to each margin, and the standardized innovations distributions F_i, $i = 1, \cdots, d$, (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 may also be obtained through the robust and minimum distance estimators (Tsukahara (2005), Mendes, Melo and Nelsen (2007)), or semi parametrically (Vandenhende and Lambert (2005). Goodness of fits may be assessed visually through pp-plots or based on some formal goodness of fit (GOF) test, usually based on the minimization of some criterion. GOF tests have been proposed in Wang and Wells (2000), Breymann, Dias & Embrechts (2003), Chen, Fan & Patton (2004), Genest, Quessy, and Rémillard. (2006), the PIT algorithm (Rosenblatt, 1952), Berg & Bakken (2006). It seems that the most accepted idea is to transform the data into a set of independent and standard uniform variables, and to calculate some measure of distance, such as the Anderson-Darling or the Kolmogorov-Smirnov distance between the transformed variables and the uniform distribution. For a discussion on goodness-of-fit tests see Genest, Rémillard, and Beaudoin (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, Czado, Frigessi, and Bakken (2007).

Consider again the joint distribution F with density f and with strictly continuous marginal c.d.f.s $F_1, \cdots, F_d$ with densities f_i. First note that any multivariate density function may be uniquely (up to relabel of variables) decomposed as

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

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

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

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 $d = 3$,

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

and

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

Expressing all conditional densities in (4) by means of (5) we derive a decomposition for $f(x_1, \cdots, x_d)$ that only consists of 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 in bivariate copulas. Given a specific factorization there are many possible reparametrizations. This is a very flexible and natural way of constructing a higher dimensional copula.

The conditional c.d.f.s needed in the 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 Bedfort and Cooke (2001) and Kurowicka and Cooke (2004), there are 240 different decompositions when $d = 5$. In these papers the authors have introduced a systematic way for obtaining the decompositions, which are graphical models denominated *regular vines*. They help 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 gives a specific way of decomposing the density $f(x_1, \cdots, x_d)$. For example, for a D-vine, $f()$ is equal to

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

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, Czado, Frigessi, and Bakken (2007). Figure 1 shows the D-vine decomposition for $d = 6$. It consists of 5 nested trees, where tree T_j possess $7 - j$ nodes and $6 - j$ edges corresponding to a pair-copula.

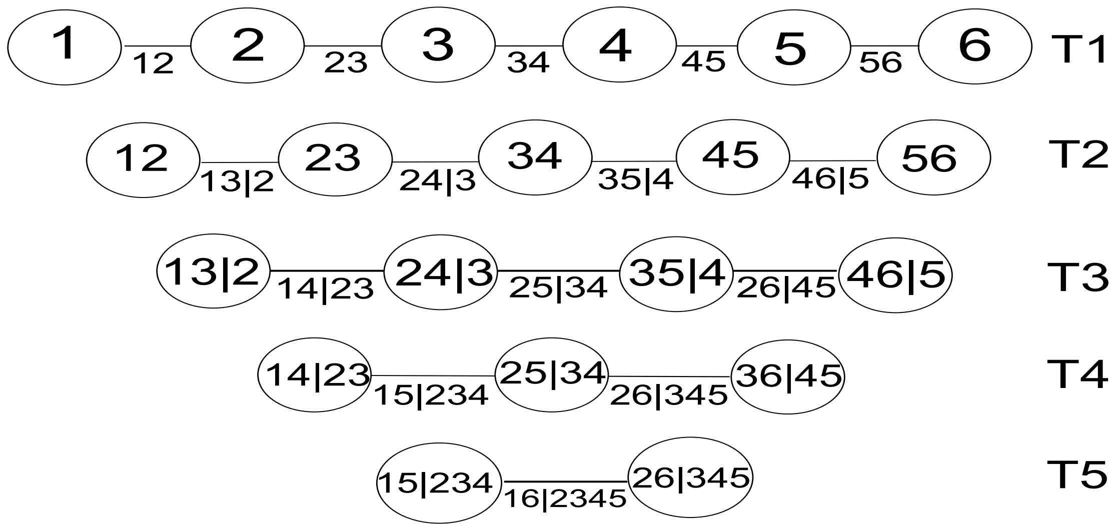

Figure 1: *Six-dimensional D-vine.*

It is not essential that all the bivariate copulas involved belong to the same family. This is exactly what we are searching for, since, recall, our objective is to construct (or estimate) a multivariate distribution which best represents the data at hand, which might be composed by completely different margins (symmetric, asymmetric, with different dynamic structures, and so on) and, more importantly, could be pair-wise joined by more complex dependence structures possessing linear and/or non-linear 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); 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 run fast. Maximum likelihood estimators depend on (i) the choice of factorization and (ii) on the choice of pair-copula families. Algorithms 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 smaller number of degrees of freedom.

3 Pair-Copulas based unconditional modeling of log-returns

We present applications using a data set of a global portfolio from the perspective 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 interdependence with the major markets. Thus we use a 6-dimensional contemporaneous daily log-returns composed by (1) a Brazilian composite hedge fund index (the ACI, Arsenal Composite Index); (2) a long-term inflation-indexed Brazilian treasury bonds index (the IMA-C index, computed by the Brazilian Association of Financial Institutions, Andima); (3) a Brazilian stock index with 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 depicted in Figure 2. There are 1629 6-dimensional observations from January 2, 2002 to October 20, 2008. We observe that the hedge fund index ACI presents a lower volatility than the long run treasury bond indexes (IMA-C and LBTBond).

In this section we take the unconditional approach for modelling the margins. The fits will be based on the flexible skew-t distribution (Hansen 1994), previously used by Patton (2006), Rockinger and Jondeau (2003), Fantazzini (2006), in the context of dynamic modeling. It 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 the implementation of the maximum likelihood method is still feasible because there are only 4 parameters to estimate

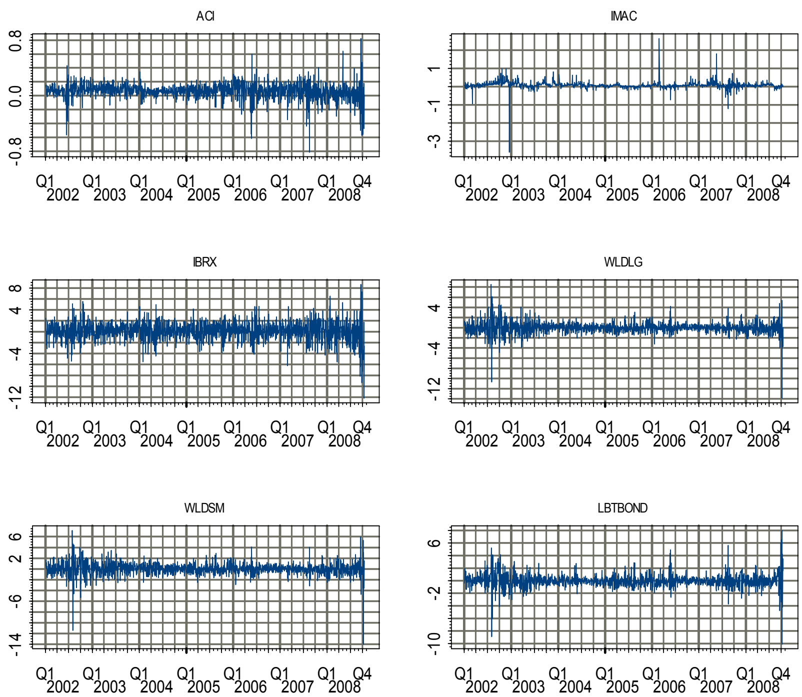

Figure 2: *Series of daily log-returns: ACI, IMAC, IBRX, WLDLQ, WLDSM, LBTBOND.*

(μ, σ, λ, ν). The parameter μ equals the population mean and the parameter σ equals the standard deviation (it exists if $\nu > 2$). When the skewness parameter λ is zero the symmetric case is recovered. Its mode 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

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

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

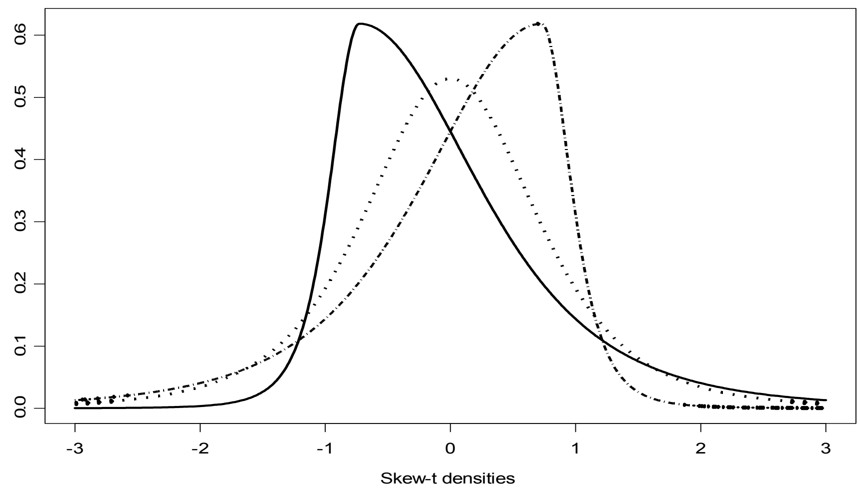

Figure 3: *Skew-t with $\nu = 4$, and $\lambda = -0.6, 0.0, 0.6$, respectively, solid, dotted, and dashed lines.*

freedom, and where

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

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

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

The maximum likelihood estimates of the skew-t distributions fitted to each variable are given in Table 1. In this table we also provide the classical 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 6 transformed standard uniform series are obtained and used to estimate the pair-copulas. We estimate a D-vine, and to help choosing the variables in tree 1 we examined the scatterplots of all pairs and ranked the pairs according the smaller number of degrees of freedom associated with a t-copula fit.

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	0.0811	0.0745	-0.0065	0.0143	0.0247
Skew-t $\widehat{\lambda}$	-0.1522	0.1537	-0.1437	0.0447	0.0269	0.1164
Skew-t $\widehat{\nu}$	3.3170	3.0000	5.3971	3.0000	3.8503	3.0000
Sample St. Deviation	0.1131	0.1853	1.6930	1.1638	1.1629	1.0891
Skew-t $\widehat{\sigma}$	0.1159	0.1251	1.6920	1.1942	1.1228	1.1295

Having decided about 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. Step by step instructions on how to perform the estimation is given in Aas, Czado, Frigessi, and Bakken (2006).

To help selecting the best copula fit we compared the penalized log-likelihood, examined the pp-plots based on the estimated and the empirical copula, and computed a GOF test statistic. Anyone of the GOF tests cited in subsection 2.1 could be applied. Actually, there is no general agreement on the best copulas GOF test. We used the one suggested by Genest and Rémillard (2005) and Genest, Rémillard and Beaudoin (2007). The test is based on the squared distance between the estimated and the empirical copula. The limiting distribution of the test statistic depends on the parameters values and approximate p-values are obtained through bootstrap.

The chosen pair-copula decomposition along with best copula fits are shown in Exhibit 1. The upper and lower tail dependence coefficients computed for each estimated bivariate copula are also shown in Exhibit 1. Joe (1997) gives the formula of 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, in particular of its complex pattern of dependence in the tails, would not be possible using a d-dimensional copula. The stock indexes of large and small world companies show the strongest dependence during stressful times.

All copulas in tree 1 possess positive upper and lower tail dependence coefficients and, according to Joe, Li, and Nikoloulopoulos (2008), in this case we say that the D-vine modeling the log-returns has multivariate upper and lower tail dependence.

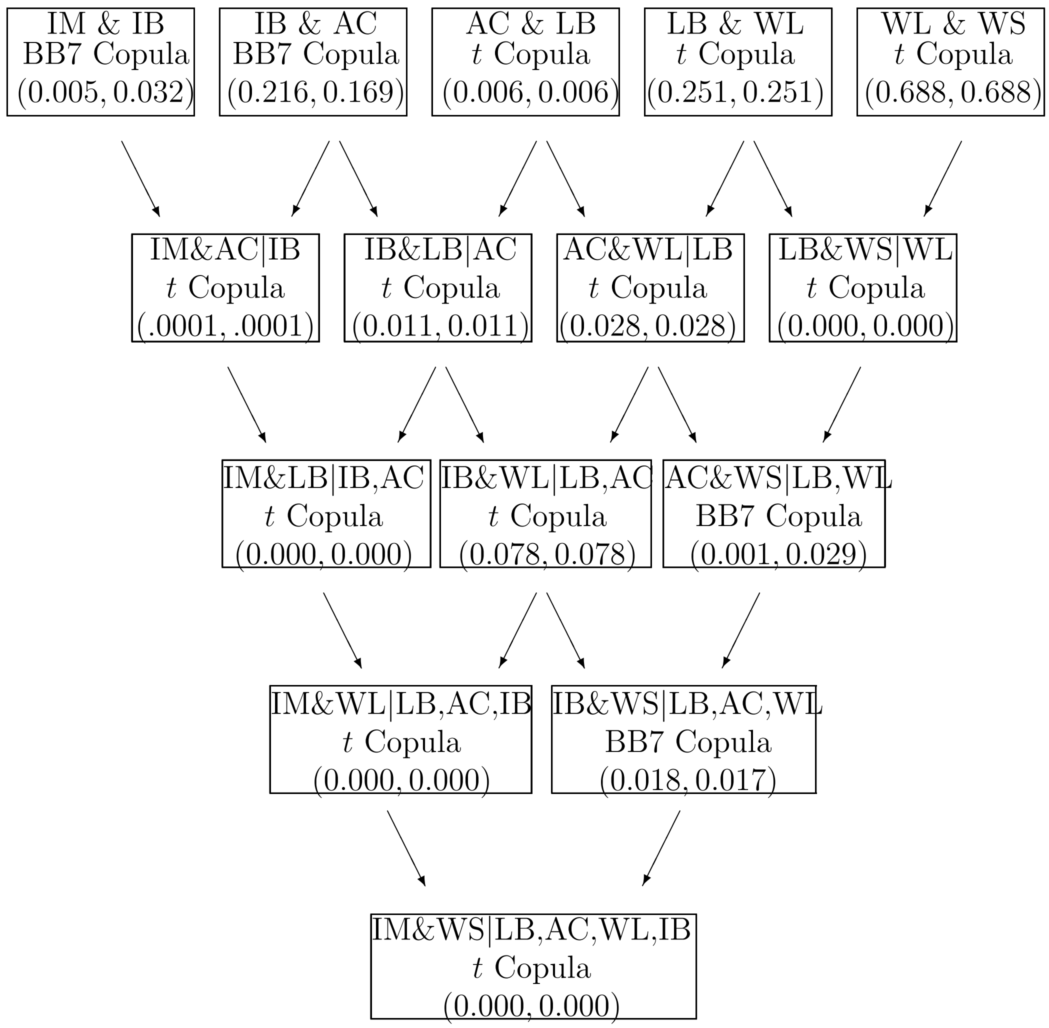

Exhibit 1: D-vine decomposition, best copula fits with λ_L and λ_U estimates. Notation in figure: IM: IMAC; AC: ACI; IB: IBRX; LB: LBTBOND; WL: WLDLG; WS: WLDSM.

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 pair-copulas. The fits in tree 1 result in 5 unconditional rank correlations. From fits in tree 2 through 5 we obtain conditional rank correlations. These conditional rank correlations are considered constant, not depending on the value of the conditioning variables, as proved in Kurowicka and Cooke (2001) for elliptical and copulas in general. This leads to another important result in Misiewicz, Kurowicka, and Cooke (2000): 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 have been stressed in Cooke and Bedford (1995), where the authors show that there exists 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 every full rank correlation matrix may be decomposed in this way (Bedford and Cooke, 1999).

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

Table 2: *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 diagonal: Pair-copulas rank correlations. Below diagonal: sample correlations.

All above cited results form the basis for using formula (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 above diagonal of Table 2.

To validate the fits we now simulate 2000 observations from the fitted D-vine using the algorithm given in Aas, Czado, Frigessi, and Bakken (2007). We compute the sample rank correlations from the simulated data and compare with those given above diagonal of Table 2. The sum of the squares of the differences between the 15 rank correlations is 0.0110123, validating the fit.

3.1 Application: Efficient Frontiers

Constructing the efficient frontier (EF) corresponding to optimal portfolios according to the Markowitz Mean Variance methodology (MV) requires just point estimates for the means, variances, and the linear correlation coefficients as inputs in the quadratic optimization problem. However it would be interesting to measure dependence beyond correlations and capture all possible different types of linear and non-linear associations among the portfolio components, and to incorporate this in the MV methodology. It would also be desirable to assess the variability of the efficient frontiers. All this translates to accurately estimating the multivariate distribution implied by the underlying assets, a task done in the previous subsection, through pair-copulas.

The widely used inputs are the classical sample mean and sample covariance matrix (from now on this approach will be referred to as *classical*). The classical approach possesses good properties if the data do come from a multivariate normal distribution, which, as it is well known, is usually not the case for log-returns data. Outside the Gaussian world the classical estimates loose efficiency and may become biased (Hampel et al., 1986) and this leads to questionable results.

Scherer and Martin (2007) obtain robust versions of Markowitz mean-variance optimal portfolios using some well known robust estimates of covariance such as the MCD of Rousseeaw (see Rousseeaw and Leroy, 1887), and other robust alternatives have been proposed, for example, Mendes and Leal (2005). However all cited alternatives differ just on the estimation method of the inputs, in particular of the linear correlation coefficient.

In this application we propose to use the location and standard deviation estimates provided by the skew-t model, and the rank correlations provided by the

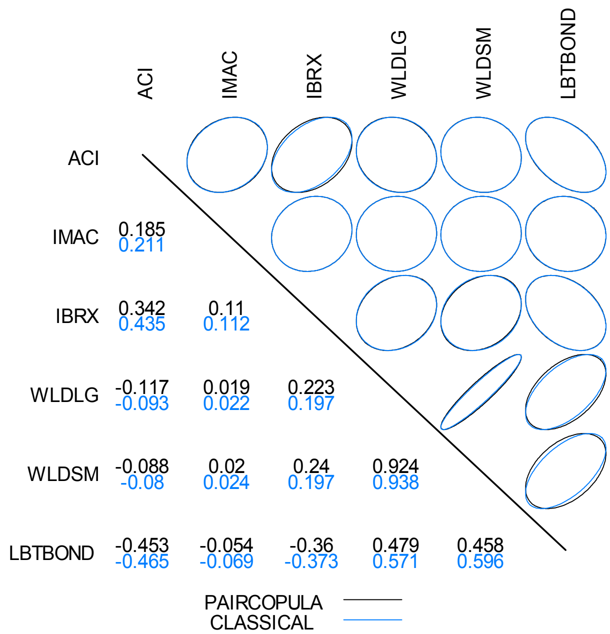

Figure 4: *Unconditional rank correlations and sample correlations.*

pair-copula decomposition. Figure 4 shows the ellipsoids associated with the rank and sample correlations. For this particular data set there are no striking differences (for example, a change of sign) between the correlation estimates. Even though, as we shall see, the resulting efficient frontiers will be quite different.

Using the MV algorithm and the two sets of inputs we construct the classical and the pair-copula based efficient frontiers containing 20 optimal linear combinations of the 6 series of log-returns. They are shown on the left hand side of Figure 5. We observe that the classical one 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 fitted data 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},$i,j=1,\cdots,6$. 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

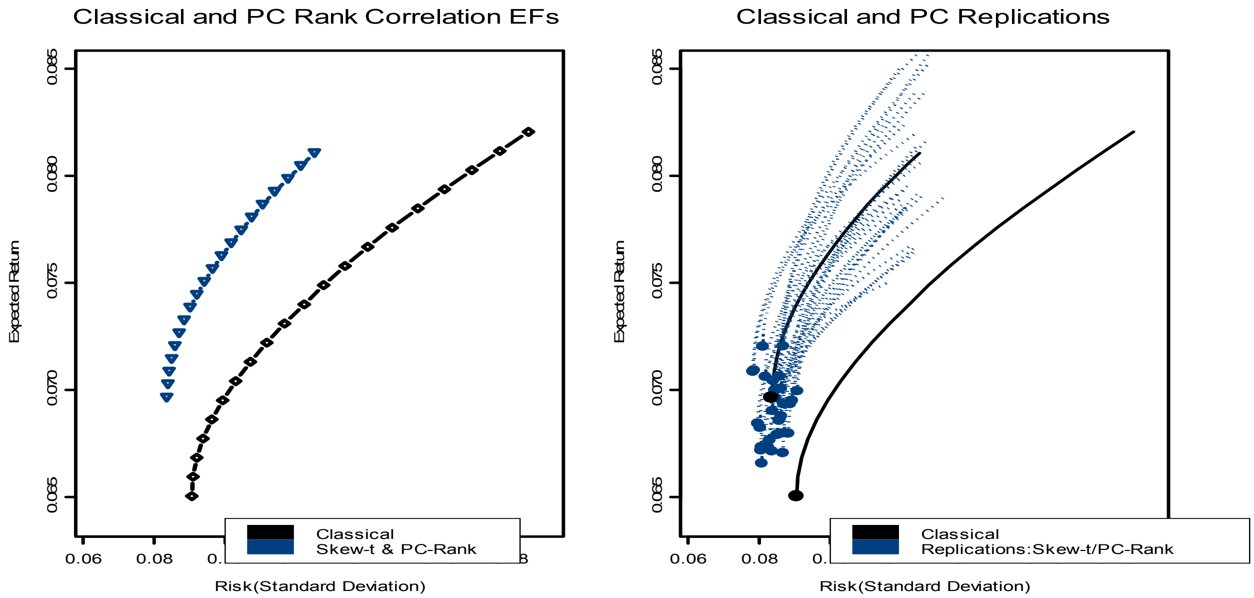

Figure 5: *Efficient frontiers: On the left hand side, classical in black, and skew-t-rank correlations based in blue. On the right hand side, replications of the PC-based EF.*

distribution, $\delta = (\lambda, \nu, \mu, \sigma)$, and let $\widehat{\delta}$ represent their estimates. For the generation, we assume that $\widehat{\theta}$ and $\widehat{\delta}$ are the true parameters values and implement the following parametric bootstrap algorithm:

For $k = 1, \cdots, B$, B large,

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

On its right hand side, Figure 5 shows the original pair-copula based EF and its parametric replications, along with the classical FE. The filled circles correspond to minimum risk portfolio. The set of all replications of some specific portfolio (in the figure, the number 1) give rise to a $(1-\alpha)\%$ confidence level convex hull containing statistically equivalent portfolios. It is also possible to draw the replications of the classical EF to verify if the corresponding convex hulls have an intersection. This would be useful for portfolio re-balancing and testing. See Mendes and Leal (2009) where the authors propose a method for replicating the classical FE and use square distances to test equality of portfolios.

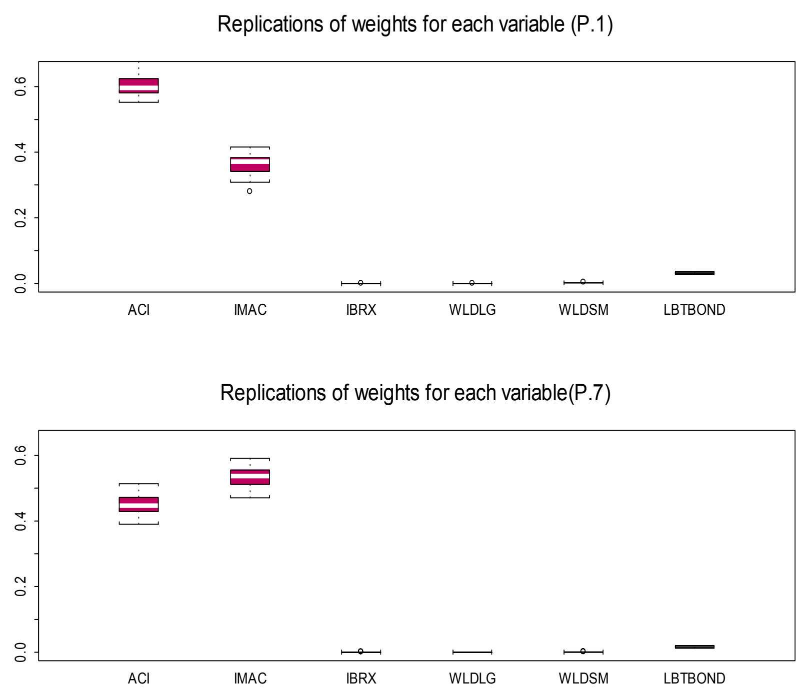

Figure 6: *Replications of weights for each variable and for portfolios ranked 1 and 7.*

We also provide in Figure 6 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 possessing smaller risks show less variability in the plane risk $\times$ return (Figure 5), and are more stable in the d-dimensional space of the weights (Figure 6). Actually, the stability of weights of a given rank portfolio, over the convex hull of replications, just confirms that equivalent portfolios showing different return × risk values in general have similar weights composition. Yet, the utility of a EF construction is not the return/risk values of the portfolio but rather their weights compositions.

4 Pair-Copulas 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 some ARFIMA-FIGARCH filter obtaining the standardized residuals, and then apply all estimation steps of previous section on the filtered data.

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

$$r_t = \mu_t + \sigma_t \epsilon_t$$

$$\mu_t = \phi_0 + \sum_{j=1}^{p} \phi_j r_{t-j} + \sum_{i=1}^{q} \theta_i \mu_{t-i}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{m} \alpha_j r_{t-j}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2$$

$$E[\epsilon_t] = 0 \quad var(\epsilon_t) = 1.$$

For each series of log-returns we fit the best ARMA-FIEGARCH model, considering as conditional distributions either the Normal or the t$_\nu$, where ν denote the numbers of degrees of freedom. Table 3 give the estimates.

Table 3: *Maximum likelihood estimates (standard errors) of the ARMA-FIEGARCH models fitted to the log-return 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

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 4 show estimates of parameters of the skew-t distribution fitted to the marginal estimated innovations series. Of course the μ estimates are close to zero and the σ estimates are close to one. Although the series still present skewness and kurtosis, they are smaller than those estimated for the raw data.

Table 4: *Skew-t parameters estimates and sample estimates of location and scale.*

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
Sample St. Deviation	0.9970	1.4722	0.9970	1.0003	1.0009	1.0029
Skew-t $\widehat{\sigma}$	0.9980	1.1352	0.9961	1.0005	0.9973	1.0029

Table 5: *Pair-copulas rank correlations (above diagonal) and sample correlations (below).*

	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

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 are free of temporal dependences. The copula families found as best fits are practically the same given in Figure 3, 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 the unconditional case, the multivariate distribution of the filtered data has tail dependence.

Next we compute the rank correlation matrix. They are given in Table 4. We observe that, for most pairs, the values of the correlation coefficients (pair-copulas based) are slightly smaller. This shows that the volatility is in many cases responsible for the contagion and that it may (slightly) increase dependence.

4.1 Efficient Frontier

As expected, the position on the risk $\times$ return plane of the efficient frontier associated with the filtered data is quite different from the previous one based on the original data. This is mostly due to changes in the means and standard deviations estimates. However, it would be interesting to examine the portfolios compositions to assess the effect of volatility on the weigths.

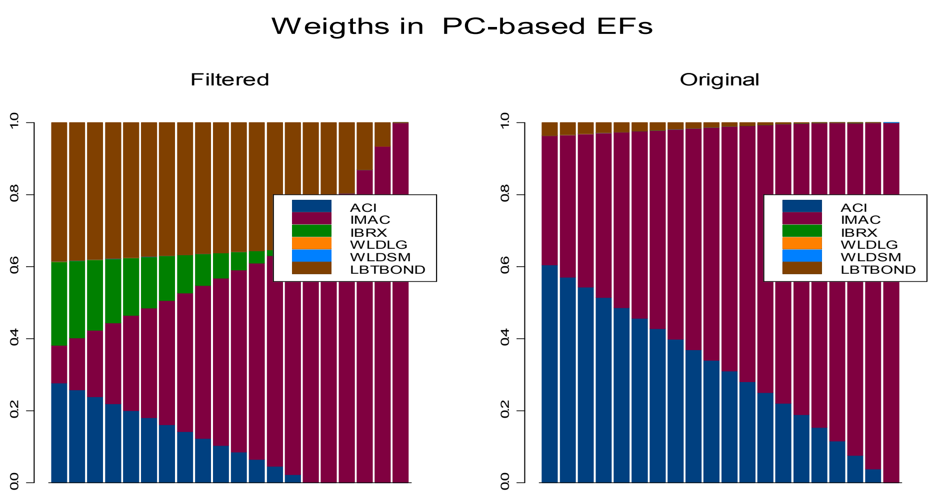

Figure 7: *Weights composing the 20 portfolios for the efficient frontiers based on the filtered and original data.*

Figure 7 shows the weights of the 20 portfolios in the efficient frontiers based on the filtered and original data. Tests using the weights and based on distances could be applied to test equality of some compositions.

5 Conclusions

In this paper we have explored the potentials 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 model, as well as conditional models, combined with pair-copulas possessing different strengths of upper and lower tail dependence. This results in a powerful model allowing for an accurate estimation of 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 variablity of the estimates.

The pair-copulas field still needs research on tests for choosing among copula families and among decompositions, and more powerful goodness-of-fit tests. Further research topics include time-varying pair-copulas.

References

Aas, K, Czado, C., Frigessi, A., and Bakken, H. 2007. "Pair-copula constructions of multiple dependence". *Insurance: Mathematics and Economics*, 2, 1, 1-25.

Bedfort, T.J., and Cooke, R.M. 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245-268.

Bedford, T. and Cooke, R.M. (2002). "Vines - a new graphical model for dependent random variables". *Annals of Statistics*, 30(4), 10311068.

Berg, D. and Aas, K. (2008). "Models for construction of multivariate dependence: A comparison study". Forthcoming ıEuropean Journal of Finance.

Berg, D. (2008). "Copula Goodness-of-fit testing: An overview and power comparison". Forthcoming in The European Journal of Finance.

Cooke, R.M., Bedford, T.J. (1995). *Reliability methods as management tools: dependence modeling and partial mission success*. In Chameleon Press, editor, ESREL95, 153160. London.

Embrechts P., McNeil A.J. and Straumann D. (1999). "Correlation and Dependence in Risk Management: Properties and Pitfalls", Preprint ETH Zurich, available from http://www.math. ethz.ch/#embrechts. http://citeseer.ist. psu.edu/article/ embrechts99correlation.html

Embrechts, McNeil, A., and Straumann, D. 2001. "Correlation and Dependency in Risk Management: Properties and Pitfalls". In *Value at Risk and Beyond*. Cambridge University Press.

Fantazzini, D. (2006). "Dynamic Copula Modelling for value-at-Risk". Frontiers in Finance and Economics, Forthcoming. Available at SSRN: http://ssrn.com/abs=944172.

Fischer, M., Köck, C., Schlüter, S., and Weigert, F. (2008). "Multivariate copula models at work: outperforming the "Desert Island Copula" ?" http://www.statistik.wiso.uni-erlangen.de/forschung/d0079.pdf.

Genest, C. and Rivest, L.P. (1993). Statistical inference procedures for bivariate Archimedian copulas, *Journal of Amer. Statist. Assoc.*, 88, 423, 1034-1043.

Genest, C. and Rémillard, B. (2005). "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models". Technical Report G-2005-51, GERAD, Montreal, Canada.

Genest, C., Quessy, J.F., and Rémillard, B. (2006). "Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation". *Scandinavian J. of Statistics*, 33, 337-366.

Genest, C., Rémillard, B., Beaudoin, D. (2007). "Omnibus goodness-of-fit tests for copulas. A review and a power study". Working paper, Université Laval.

Hampel, F.R., Ronchetti, E.M., Rousseeaw, P.J. (1986). *Robust Statistics: The Approach based on influence functions*. J. Willey and Sons, Inc.

Hansen, B. (1994). "Autoregressive Conditional Density Estimation". *International Economic Review*, 35,3.

Joe, H. (1996). "Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters". In L. Rüchendorf and B. Schweizer and M. D. Taylor (Ed.), *Distributions with fixed marginals and Related Topics*.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* London: Chapman & Hall.

Joe, H., Li, H. and Nikoloulopoulos, A.K. (2008) Tail dependence functions and vine copulas. Math technical report 2008-3, Washington State University.

Kurowicka, D. and Cooke, R. M. (2001). "Conditional, partial, and rank correlation for the elliptical copula; Dependence modeling in uncertainty analysis". Proceedings ESREL.

Kurowicka, D. and Cooke, R. M. (2000). "Conditional and partial for graphical uncertainty models". In *Recent Advances in Reliability Theory. Methodology, Practice, and Inference.* By Nikolaos Limnios, Mikhail Stepanovich Nikulin, Birkhäuser.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments.* J. Willey, N.Y.

Mendes, B. V. M. and Leal, R. P. C. 2005. "Robust Multivariate Modeling in Finance". *International Journal of Managerial Finance.* V.1, N. 2, pp. 95-106.

Mendes, B. V. M. and Leal, R. P. C. 2009. "On the Resampling of Efficient Frontiers". *Submitted.*.

Min, A. and C. Czado (2008). "Bayesian inference for multivariate copulas using pair-copula constructions". Preprint, available under http://www-m4.ma. tum.de/Papers /index.html.

Nelsen, R.B. (2007). *An introduction to copulas*, Lectures Notes in Statistics, Springer, N.York.

Misiewicz, J., Kurowicka, D., and Cooke, R.M. (2000). "Elliptical copulae". to appear.

Rockinger, M. and Jondeau, E. (2001). "Conditional dependency of financial series: An application of copulas". HEC Paris DP 723.

Rosenblatt, M. 1952."Remarks on a multivariate transformation". *Annals of Mathematical Statistics*, 23, 470-472.

Rousseeuw, P.J., and A.M. Leroy, 1987. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.

Scott, D.W. (1992). *Multivariate Density Estimation*, NY: Wiley

Patton, A. (2001). "On the out-of-sample importance of skewness and asymmetric dependence for asset allocation". *Journal of Financial Econometrics*, 2, 1, 130-168.

Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47, 2.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229-231.

Sklar, A. (1996). Random variables, distribution functions, and copulas (a personal look backward and forward), in *Distributions with Fixed Marginals and Related Topics*, ed. by L. Rüschendor, B. Schweizer, and M. Taylor,. 1-14. IMS, Hayward, CA.

Yule, G. U. and Kendall, M. G. (1965). *An Introduction to the Theory of Statistics*, 281-309, Chapter 12, Charles Griffin & Co., 14th Edition, London.