Computational Statistics and Data Analysis 54 (2010) 2562–2579

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Three-stage semi-parametric estimation of T-copulas: Asymptotics, finite-sample properties and computational aspects

Dean Fantazzini *

Moscow School of Economics, Moscow State University, Russia

ARTICLE INFO

Article history:
Available online 12 February 2009

ABSTRACT

A two-stage semi-parametric estimation procedure for a broad class of copulas satisfying minimal regularity conditions has been recently proposed. In addition, a three-stage semi-parametric estimation method based on Kendall's tau in order to estimate the Student's t copula has also been designed. Its major advantage is to allow for greater computational tractability when dealing with high dimensional issues, where two-stage procedures are no more a viable choice. The asymptotic properties of this methodology are developed and its finite-sample behavior are examined via simulations. The advantages and disadvantages of this methodology are analyzed in terms of numerical convergence and positive definiteness of the estimated T-copula correlation matrix.

1. Introduction

The theory of copulas dates back to Sklar (1959), but its application in financial modelling is far more recent and dates back to the late 90s, instead. A copula is a function that embodies all the information about the dependence structure between the components of a random vector. When it is applied to marginal distributions which do not necessarily belong to the same distribution family, it results in a proper multivariate distribution. As a consequence, this theory enables us to incorporate a flexible modelling of the dependence structure between different variables, while allowing them to be modelled by different marginal distributions.

A semi-parametric estimation procedure for a broad class of copulas satisfying minimal regularity conditions was initially proposed by Genest et al. (1995), while a method-of-moment estimation procedure based on Kendall's tau was discussed by Genest and Rivest (1993). Genest et al. (1995) credit Oakes (1994) for the idea, although they were the first to examine its asymptotic properties. Shih and Louis (1995) also contributed to this field of the literature. However, recent empirical financial literature combined these two procedures to estimate the elliptical Student's t copula, too (see Bouyé et al. (2001), Marshall and Zeevi (2002), Cherubini et al. (2004), McNeil et al. (2005)). This new methodology entails three stages: a first stage where a non-parametric estimation of the marginal empirical distribution functions is performed; a second stage where a method-of-moment estimator based on Kendall's tau is used for the T-copula correlation matrix, and a third stage which considers Maximum Likelihood methods for the degrees of freedom ν.

This has become a common procedure among financial practitioners when dealing with high-dimensional portfolios and standard Maximum Likelihood methods cannot be used. Nevertheless, neither the asymptotic properties of such a multi-step procedure have been studied, nor its finite-sample properties. The former are important in order to build statistical tests for the structural parameters, while the latter are of interest to compare the small-sample efficiency and bias of this estimator with other estimators, such as the classical one-stage ML estimator where the margins are known. Although the

* Corresponding address: Moscow School of Economics, Moscow State University, 1, Building 61, Leninskie Gory, 119992, M.V. Lomonosov MSU, Moscow 119992, Russia. Tel.: +7 9857758816.
E-mail address: fantazzini@mse-msu.ru.

0167-9473/$ – see front matter
doi:10.1016/j.csda.2009.02.004

marginal distributions are unknown in practice, this hypothetical scenario wherein the marginals are assumed to be known, represents the ideal situation that can be used as a benchmark for comparative purposes, see, e.g., Kim et al. (2007).

What we do in this paper is to provide the asymptotic distribution for this recent semi-parametric method, and use simulations with different Data Generating Processes to examine the behavior of this estimator in small samples. The Monte Carlo study shows that this semi-parametric estimator is more efficient and less biased than the one-stage ML estimator when small samples and copulas with low degrees of freedom are of concern. We then analyze the pros and cons of this methodology in terms of numerical convergence and positive definiteness of the estimated T-copula correlation matrix. When small samples are of concern and ν is high, the number of times when the numerical maximization of the log-likelihood fails to converge is much higher for the ML method than for the three-stage Kendall's tau Moment Estimator–Canonical Maximum Likelihood (KME–CML) method. Yet, while the coverage rates at the 95% level for the ML estimates for ν do not show any particular bias or trend, the KME–CML estimates show very low coverage rates when ν becomes close to 30 and the correlations are not too strong. However, this drop in the coverage rates is large with bivariate T-copulas, only, while it is much lower when dealing with higher dimensional T-copulas, which is the usual case for real managed financial portfolios. Besides, both the ML and the KME–CML methods show high mean and median biases for the estimated correlations when the true ones are close to zero. Nevertheless, the effects on the coverage rates for the correlations are rather limited in this case.

Finally, we show that the eigenvalue method by Rousseeuw and Molenberghs (1993) has to be used to obtain a positive definite correlation matrix only when dealing with very small samples ($n < 100$) and when the true underlying process has the lowest eigenvalue close to zero. This fix induces a positive bias in the estimate of ν, but the effects on the coverage rates are rather limited. Besides, the number of times when this method has to be used quickly decreases when ν increases.

The rest of the paper is organized as follows. We introduce a recent semi-parametric estimation method of Student's t copulas based on Kendall's tau in Section 2 and we provide its asymptotics in Section 3. We present in Section 4 the results of a Monte Carlo study of the small-sample properties of this estimator, while we conclude in Section 5.

2. The three-stage KME–CML method

The study of copulas has originated with the seminal papers by Höffding (1940) and Sklar (1959) and has seen various applications in statistics and financial literature. Examples include Clayton (1978), Genest and MacKay (1986a, 1986b), Genest and Rivest (1993), Rosenberg (1998, 2003), Bouyé et al. (2001), Patton (2004, 2005), Dobric and Schmid (2006), Huarda et al. (2006), Granger et al. (2006), Fantazzini (2009, in press), Dalla Valle et al. (2008) and Karlis and Nikoloulopoulos (2008). For more details, we refer the interested reader to the recent methodological overviews by Joe (1997) and Nelsen (1999), while Cherubini et al. (2004) and McNeil et al. (2005) provide a comprehensive and detailed discussion of copula techniques for financial applications.

Genest et al. (1995) were the first to analyze a semi-parametric estimation of a bivariate copula with i.i.d. observations and to develop its asymptotic properties. Their Canonical Maximum Likelihood (CML) method differs from full Maximum Likelihood methods because no assumptions are made about the parametric form of the marginal distributions. We now briefly review this semi-parametric method since it constitutes the building block for our following analysis.

Let us consider a multivariate random sample represented by $X_i = (X_{i1}, \ldots, X_{id})$, and $i = 1, \ldots, n$, where d stands for the number of variables included and n represents the number of observations available. Let f_h be the density of the joint distribution of X. Then, by using Sklar's theorem (1959) and the relationship between the distribution and the density function we have:

$$f_h(X_i; \alpha_1, \ldots, \alpha_d, \gamma) = c(F_1(X_{1;}\alpha_1), \ldots, F_d(X_{d;}\alpha_d); \gamma) \cdot \prod_{j=1}^{d} f_j(X_j; \alpha_j) \tag{1}$$

where f_j is the univariate density of the marginal distribution F_j, c is the copula density, $\alpha_j, j = 1, \ldots, d$ is the vector of parameters of the marginal distribution F_j, while γ is the vector of the copula parameters. The CML estimation process is performed in two steps:

Definition 2.1 (*CML Copula Estimation*).

1. Transform the dataset $(X_{i1}, X_{i2}, \ldots, X_{id}), i = 1, \ldots, n$ into normalized ranks $(\hat{U}_{i1}, \hat{U}_{i2}, \ldots, \hat{U}_{id})$ using the empirical distributions $\hat{U}_{ij} = F_{nj}(X_{ij})$ defined as follows:

$$F_{nj}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(X_{ij} \leq x)} \tag{2}$$

   where $\mathbb{1}_{(\cdot)}$ represents the indicator function.
2. Estimate the copula parameters by maximizing the log-likelihood:

$$\hat{\gamma}_{CML} = \arg\max \sum_{i=1}^{n} \log(c(F_{n1}(X_{i1}), \ldots, F_{nd}(X_{id})); \gamma). \tag{3}$$

Genest et al. (1995) show in Proposition A.1 that under certain regularity conditions, the semi-parametric estimator $\hat{\gamma}_{CML}$ has the following asymptotic distribution (we consider the bivariate case $d = 2$ for the sake of simplicity, see Genest et al. (1995), Section 4, for the multivariate case):

$$\sqrt{T}(\hat{\gamma}_{CML} - \gamma_0) \xrightarrow{d} \mathrm{N}\left(0, \frac{\sigma^2}{h^2}\right) \quad \text{where,} \tag{4}$$

$$\sigma^2 = \mathrm{var}[l_\gamma\,(F_1(X_1), F_2(X_2); \gamma) + W_1(X_1) + W_2(X_2)], \tag{5}$$

$$W_j(x_j) = \int \mathbb{1}_{F_j(X_j) \le u_j}\; l_{\gamma,j}\,(u_1, u_2; \gamma)\, c(u_1, u_2; \gamma)\, \mathrm{d}u_1\, \mathrm{d}u_2 \quad j = 1, 2 \quad \text{and} \tag{6}$$

$$h = -E[l_{\gamma,\gamma}\,(F_1(X_1), F_2(X_2); \gamma)] \tag{7}$$

and where $l(u_1, u_2, \gamma) = \log c(u_1, u_2; \gamma)$ and the indices $1, 2, \gamma$ denote the partial derivatives of l with respect to u_1, u_2, γ, respectively. Besides, $W_j(x_j)$ can have this alternative expression too, upon integrating by parts with respect to u_j ($j = 1, 2$):

$$W_j(x_j) = -\int \mathbb{1}_{F_j(X_j) \le u_j}\; l_\gamma\,(u_1, u_2; \gamma)\, l_j\,(u_1, u_2; \gamma)\, c(u_1, u_2; \gamma)\, \mathrm{d}u_1\, \mathrm{d}u_2. \tag{8}$$

Since the seminal work by Genest et al. (1995), it has become common practice to use semi-parametric methods with high-dimensional elliptical Student's t copulas too; see Cherubini et al. (2004) and McNeil et al. (2005) for a detailed discussion about their financial applications. Particularly, after the marginal empirical distribution functions are computed in a first stage, the correlation matrix is estimated in a second stage using a method-of-moment estimator based on Kendall's tau, while the degrees of freedom are estimated in a third stage using Maximum Likelihood methods. We remark that the fact that the margins are estimated in the first stage is not necessary to estimate the correlation matrix, since Kendall's tau is based on the number of concordances, which is unaffected by the rank transformation. However, the estimation of the margins is necessary in the last stage to estimate the degrees of freedom.

Despite the widespread use of this procedure, its asymptotic properties have not been developed yet. Before doing that and properly define this estimation method, we have to introduce Kendall's tau, its relation with linear correlation for elliptical copulas and the method-of-moment estimator based on it. A method for estimating copula parameters based on Kendall's tau has been suggested in Genest and Rivest (1993), Lindskog (2000), Lindskog et al. (2002), Cherubini et al. (2004) and McNeil et al. (2005). This dependence measure can be defined as follows:

Definition 2.2 (*Kendall's Tau*)**.** If we have $(X_1; X_2)$ and $(\tilde{X}_1; \tilde{X}_2)$ two independent and identically distributed random vectors, the population version of Kendall's tau $\tau(X_1; X_2)$ is (see Kruskal (1958)):

$$\tau(X_1, X_2) = \mathrm{E}\left[\mathrm{sign}\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2)\right)\right]. \tag{9}$$

The generalization of Kendall's tau to $n > 2$ dimensions can be analogous to the procedure for linear correlation, where we have a $n \times n$ matrix of pairwise correlations. However, other n-variate generalization of Kendall's tau are possible, see Clemen and Jouini (1996) and Barbe et al. (1996) for a discussion about this issue. The main properties of this dependence measure and relative proofs are reported in Embrechts et al. (2002). Besides, Kendall's tau can be expressed in terms of copulas, thus simplifying calculus, see e.g. Nelsen (1999, p. 127).

$$\tau(X_1, X_2) = 4\int_0^1 \int_0^1 C(u_1, u_2)\, \mathrm{d}C(u_1, u_2) - 1. \tag{10}$$

Evaluating Kendall's tau requires the evaluation of a double integral and for elliptical copulas like the T-copula this is not an easy task: this problem was solved by Lindskog et al. (2002) who proved that Kendall's tau for elliptical distributions is given by

$$\tau(X_1, X_2) = \frac{2}{\pi} \arcsin \rho_{X_1 X_2} \tag{11}$$

where $\rho_{X_1 X_2}$ is the copula correlation parameter. It follows easily from the previous equation that the correlation can be retrieved by the inversion of Kendall's tau: in the case of the T-copula, the estimated correlation matrix $\hat{\Sigma}$ has off-diagonal elements given by $\hat{\rho}_{ij} = \sin(\pi \hat{\tau}_{ij}/2)$, see, e.g. Wang and Wells (2000), Kowalczyka and Niewiadomska-Bugaj (2002) and Genest et al. (2006).

More formally, Kendall's tau moment estimator (KME) can be defined as follows: let us consider the population version of Kendall's tau (9) and its relationship with copula parameters (10) to build a moment function of the type

$$\mathrm{E}\,[\psi\,(X_1, X_2; \gamma_0)] = 0. \tag{12}$$

Then, we can construct an empirical estimate of Kendall's tau pairwise correlation matrix and use relationship (10) to infer an estimate of the relevant parameters of the copula. This is a method-of-moments estimate because the true moment (9) is replaced by its empirical analogue,

$$\binom{n}{2}^{-1} \sum_{1\le i<s\le n} \operatorname{sign}\left((x_{i,1}-\tilde{x}_{s,1})(x_{i,2}-\tilde{x}_{s,2})\right) \tag{13}$$

and (10) is then used to estimate the copula parameters. For example, when using elliptical copulas and Eq. (11), this moment function becomes

$$\mathrm{E}\left[\psi\left(X_1, X_2; \rho_{X_1,X_2}\right)\right] = \mathrm{E}\left[\rho_{X_1,X_2} - \sin(\pi\tau(X_1,X_2)/2)\right] = 0 \tag{14}$$

where $\rho_{X_1X_2}$ is the copula correlation parameter.

There are cases when the copula parameter vector has different kinds of parameters and only some of them can be expressed as a function of Kendall's tau. This is the case for the T-copula which is parameterized by the correlation matrix $\boldsymbol{\Sigma}$ and the degrees of freedom ν, but only the former has a direct relationship with Kendall's tau. This is the copula of the multivariate Student's t-distribution and we can derive its density function by using Sklar's theorem (1959) and the relationship between the distribution and the density function:

$$\begin{aligned} c(t_\nu(x_1),\ldots,t_\nu(x_d)) &= \frac{f^{Student}(x_1,\ldots,x_d)}{\prod_{j=1}^{d} f_j^{Student}(x_i)} \\ &= |\boldsymbol{\Sigma}|^{-1/2}\,\frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right]^{d}\frac{\left(1+\frac{\zeta'\boldsymbol{\Sigma}^{-1}\zeta}{\nu}\right)^{-\frac{\nu+d}{2}}}{\prod_{i=1}^{d}\left(1+\frac{\zeta_i^2}{2}\right)^{-\frac{\nu+1}{2}}} \end{aligned} \tag{15}$$

where $\zeta = (t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_d))^\top$ is the vector of univariate Student's t inverse distribution functions, ν is the degrees of freedom, $u_j = t_\nu(x_j)$ is the univariate Student's t cumulative distribution function, while $\boldsymbol{\Sigma}$ is the correlation matrix. In this situation, Bouyé et al. (2001) and McNeil et al. (2005) have suggested the following estimation procedure:

Definition 2.3 (*Three-stage KME–CML Copula Estimation*)**.**

(1) Transform the dataset $(X_{i1}, X_{i2}, \ldots, X_{id})$ into normalized ranks $(F_{n1}(X_{i1}), F_{n2}(X_{i2}), \ldots, F_{nd}(X_{id}))$, using the empirical distribution function, and which are approximately uniform variates.

(2) Collect all pairwise estimates of the sample Kendall's tau given by (13) in an empirical Kendall's tau matrix $\hat{R}^\tau$ defined by $\hat{R}^\tau_{jk} = \tau(F_{nj}(X_j), F_{nk}(X_k))$, and then construct the correlation matrix using this relationship $\hat{\Sigma}_{j,k} = \sin(\frac{\pi}{2}\hat{R}^\tau_{j,k})$, where the estimated parameters are the $q = d\cdot(d-1)/2$ correlations $[\hat{\rho}_1,\ldots\hat{\rho}_q]'$. Since there is no guarantee that this componentwise transformation of the empirical Kendall's tau matrix is positive definite, when needed, $\hat{\boldsymbol{\Sigma}}$ can be adjusted to obtain a positive definite matrix using a procedure such as the eigenvalue method of Rousseeuw and Molenberghs (1993) or other methods.

(3) Look for the CML estimator of the degrees of freedom $\hat{\nu}_{CML}$ by maximizing the log-likelihood function of the T-copula density:

$$\hat{\nu}_{CML} = \arg\max \sum_{i=1}^{n} \log c_{T\text{-}copula}(F_{n1}(X_{i1}),\ldots,F_{nd}(X_{id}); \hat{\Sigma}, \nu). \tag{16}$$

3. Asymptotic properties of the three-stage method

The second step in the previous Definition 2.3 corresponds to a method-of-moments estimation based on q moments and Kendall tau rank correlations estimated with empirical distribution functions: we stress again that the estimation of the margins is not necessary to estimate the correlation matrix, since Kendall's tau is unaffected by the rank transformation. We can therefore generalize Eq. (14) and build a $q\times 1$ moments vector ψ for the parameter vector $\boldsymbol{\theta}_0 = [\rho_1,\ldots,\rho_q]'$ as reported below:

$$\psi\left(F_1(X_1),\ldots,F_n(X_n);\boldsymbol{\theta}_0\right) = \begin{pmatrix} \mathrm{E}\left[\psi_1\left(F_1(X_1),F_2(X_2);\ \rho_1\right)\right] \\ \vdots \\ \mathrm{E}\left[\psi_q\left(F_{d-1}(X_{d-1}),F_d(X_d);\ \rho_q\right)\right] \end{pmatrix} = 0. \tag{17}$$

Then these theorems follow (the proofs are reported in Appendix A in the technical report by Fantazzini (2009)):

Theorem 3.1 (*Consistency of* $\hat{\theta}$). *Let us assume that* $(X_{i1}, \ldots, X_{id})$ *are i.i.d random variables with dependence structure given by* $c(u_{i,1}, \ldots, u_{i,d}; \Sigma_0, \nu_0)$. *Suppose that*

(i) *the parameter space* Θ *is a compact subset of* $\mathbb{R}^q$,
(ii) *the q-variate moment vector* $\psi\,(F_1(X_1), \ldots, F_d(X_d); \boldsymbol{\theta}_0)$ *is continuous in* $\boldsymbol{\theta}_0$ *for all* X_j,
(iii) $\psi\,(F_1(X_1), \ldots, F_d(X_d); \boldsymbol{\theta})$ *is measurable in* X_j *for all* $\boldsymbol{\theta}$ *in* Θ,
(iv) $\mathrm{E}\left[\psi\,(F_1(X_1), \ldots, F_d(X_d); \boldsymbol{\theta})\right] \neq \mathbf{0}$ *for all* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *in* Θ,
(v) $\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\Theta} \|\psi\,(F_1(X_1), \ldots, F_d(X_d); \boldsymbol{\theta})\,\|\right] < \infty$,

then $\hat{\boldsymbol{\theta}} \xrightarrow{d} \boldsymbol{\theta}_0$ *as* $n \to \infty$.

Theorem 3.2 (*Consistency of* $\hat{\nu}_{CML}$). *Let the assumptions of the previous theorem hold, as well as the regularity conditions reported in Proposition A.1 in Genest et al. (1995). Then* $\hat{\nu}_{CML} \xrightarrow{d} \nu_0$ *as* $n \to \infty$.

The asymptotic normality is not straightforward, since we use a three-step procedure where we perform a different kind of estimation at the second and third stage. A possible solution is to consider the CML used in the third stage as a special method-of-moment estimator. Just note that the CML estimator is defined by the derivative of the log-likelihood function with respect to the degrees of freedom:

$$\frac{\partial l(\cdot;\nu)}{\partial \nu} = \sum_{i=1}^{n} l_\nu\left(F_{n1}(X_{i1}), \ldots, F_{nd}(X_{id}); \hat{\boldsymbol{\Sigma}}, \hat{\nu}\right) = 0. \tag{18}$$

Dividing both sides by n yields the definition of the method of moments estimator:

$$\frac{1}{n}\sum_{i=1}^{n} l_\nu\left(F_{n1}(X_{i1}), \ldots, F_{nd}(X_{id}); \hat{\boldsymbol{\Sigma}}, \hat{\nu}\right) = \frac{1}{n}\sum_{i=1}^{n} \psi_\nu(F_{n1}(X_{i1}), \ldots, F_{nd}(X_{id}); \hat{\boldsymbol{\Sigma}}, \hat{\nu}) = 0.$$

Thus, the CML estimator is a simple method-of-moments (MM) estimator with the score as its moment function, where the sample mean of the score is equal to the population mean of the score. We remind that the MM estimator $\hat{\theta}$ of θ_0 based on the k moment restrictions $E[\psi(Z_i, \theta_0)] = 0$ and where $Z_i = (Y_i, X_i)$ is a vector of endogenous and explanatory variables for observation $i = 1, \ldots, n$, is the solution to the problem

$$\frac{1}{n}\sum_{i=1}^{n} \psi(Z_i, \hat{\theta}) = E[\psi(Z_i, \theta_0)] = 0.$$

The method-of-moments estimator is simply the value of θ which sets the sample moments (sample means) equal to the true population moments. Since the number of parameters is equal to k, the definition of the simple MM estimator is the solution (the root) of the k (possibly nonlinear) equation system with respect to $\hat{\theta}$, see, e.g., Greene (2008) for more details. Therefore, we can now apply the method-of-moments asymptotics together with multivariate rank statistics.

Let us define the sample moments vector $\Psi_{KME\text{-}CML}$ for the parameter vector $\hat{\Xi} = [\hat{\rho}_1, \ldots \hat{\rho}_q, \hat{\nu}]'$ as follows:

$$\Psi_{KME-CML}\left(F_{n1}(X_{i1}), \ldots, F_{nd}(X_{id}); \hat{\boldsymbol{\Xi}}\right) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \psi_1\left(F_{n1}(X_{i1}), F_{n2}(X_{i2});\ \ \hat{\rho}_1\right) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \psi_q\left(F_{n(d-1)}(X_{i(d-1)}), F_{nd}(X_{id});\ \ \hat{\rho}_q\right) \\ \frac{1}{n}\sum_{i=1}^{n} \psi_\nu\left(F_{n1}(X_{i1}), \ldots, F_{nd}(X_{id}); \hat{\boldsymbol{\Sigma}}, \hat{\nu}\right) \end{pmatrix} = 0.$$

Let us also define the population moments vector with a correction to take the non-parametric estimation of the marginals into account, together with its variance (see Genest et al. (1995), § 4):

$$\Delta_0 = \begin{pmatrix} \psi_1\,(F_1(X_1), F_2(X_2);\ \ \rho_1) \\ \vdots \\ \psi_q\left(F_{d-1}(X_{d-1}), F_d(X_d);\ \ \rho_q\right) \\ \psi_\nu\,(F_1(X_1), \ldots, F_d(X_d); \boldsymbol{\Sigma_0}, \nu_0) + \sum_{j=1}^{d} W_{j,\nu}(X_j) \end{pmatrix} = 0, \tag{19}$$

$$\boldsymbol{\Upsilon}_0 \equiv \mathrm{var}[\boldsymbol{\Delta}_0] = \mathrm{E}[\boldsymbol{\Delta}_{KME-CML}\,\boldsymbol{\Delta}_{KME-CML}{}'] \tag{20}$$

where

$$W_{j,\nu}(X_j) = \int \mathbb{1}_{F_j(X_j)\leq u_j} \frac{\partial^2}{\partial \nu \partial u_j} \log c(u_1, \ldots, u_d)\, \mathrm{d}C(u_1, \ldots, u_d). \tag{21}$$

Note that the population moments used to estimate the correlations are not affected by the marginals empirical distribution functions, since Kendall's tau is invariant under strictly increasing marginal transformations. Then this theorem follows:

Theorem 3.3 (*Asymptotic Distribution Three-stages KME–CML Method*)**.** *Let the assumptions of the previous theorems hold. Assume further that* $\frac{\partial \Psi_{KME--CML}(\cdot;\Xi)}{\partial \Xi'}$ *is* $O(1)$ *and uniformly negative definite, while* $\boldsymbol{\Upsilon}_0$ *is* $O(1)$ *and uniformly positive definite. Then, the three-stage KME–CML estimator verifies the properties of asymptotic normality:*

$$\sqrt{T}(\hat{\Xi} - \Xi_0) \xrightarrow{d} \mathrm{N}\left(0, \mathrm{E}\left[\frac{\partial \Psi_{KME-CML}}{\partial \Xi'}\right]^{-1} \boldsymbol{\Upsilon}_0\, \mathrm{E}\left[\frac{\partial \Psi_{KME-CML}}{\partial \Xi'}\right]^{-1\prime}\right). \tag{22}$$

The previous asymptotic properties are still valid when dealing with multivariate heteroscedastic time series models, where one first obtains consistent estimates of the parameters of each univariate marginal time-series, and computes the corresponding residuals. These are then used to estimate the joint distribution of the multivariate error terms, which is specified using a copula. Such a result is a straightforward application of Theorems 1 and 2 in Kim et al. (2008), who make use of, and build upon, recent elegant results of Koul and Ling (2006) and Koul (2002) for these models. Besides, we remark that essentially the same result of Kim et al. (2008) is also used in Chen and Fan (2006) but without proofs.

Theorem 3.4 (*Asymptotic Distribution Three-stages KME–CML Method for Multivariate Heteroscedastic Time Series Models*)**.** *Let the regularity conditions* (i)–(v) *reported in Theorem* 3.1 *hold, together with conditions* (A.1)–(A.9) *in Kim et al. (2008). Then, the three-stage KME–CML estimator verifies the properties of asymptotic normality defined in (22).*

Conditions (A.1)–(A.4) do not involve the time-series aspects of the model and they were also used in Kim et al. (2005) for the linear regression case with iid errors. Particularly, conditions (A.1)–(A.2) require that the copula density has continuous partial derivatives up to third order, that they are finite and their second moment is finite. Condition (A.3) is a technical condition on the partial derivatives, while condition (A.4) requires that the conditions of Proposition A.1 in Genest et al. (1995) are satisfied. The conditions (A.5)–(A.8) are technical conditions taken from an earlier work by Koul (2002) and Koul and Ling (2006), while condition (A.9) is a mild one: for example, if the conditional mean is a function of past values of the time series which is strictly stationary and ergodic, then the summand forms a strictly stationary and ergodic process with mean zero, and hence (A.9) would be satisfied (see Taniguchi and Kakizawa (2000), Theorems 1.3.3–1.3.5, and Kim et al. (2008)).

Since the main results on the properties of the proposed semi-parametric method are asymptotic, a large-scale simulation study was carried out to compare its properties with other competing ones.

4. Finite-sample properties and computational aspects

In this section we present the results of a Monte Carlo study of the small-sample properties of the estimator discussed above for a representative collection of Data Generating Processes (DGPs). Furthermore, we analyze the pros and cons of this methodology in terms of numerical convergence and positive definiteness of the estimated T-copula correlation matrix.

We consider the following possible DGPs:

(1) We examine the case that two variables have a bivariate Student's t copula, with the copula linear correlation ρ ranging between -0.9 and 0.9 (step 0.1). We examine different values for the degree of freedom ν, too, ranging between 3 and 30 (step 1). The former corresponds to a case of strong tail dependence, that is there is a high probability of observing an extremely large observation on one variable, given that the other variable has yielded an extremely large observation. The latter exhibits low tail dependence, and corresponds to the case where the T-copula becomes almost indistinguishable from a Normal copula. We remark that the Student's t copula generates positive tail dependence, while the normal copula generates zero tail dependence, instead. See Cherubini et al. (2004) for more details. We consider two possible data situations: $n = 50$ and $n = 500$.

(2) We examine the case where ten variables have a multivariate Student's t copula, with the copula correlation matrix $\boldsymbol{\Sigma}$ equal to:

We choose this correlation matrix because its lowest eigenvalue is very close to zero (0.0786) and it allows us to study the effect that the eigenvalue method by Rousseeuw and Molenberghs (1993) has on the limiting distribution of

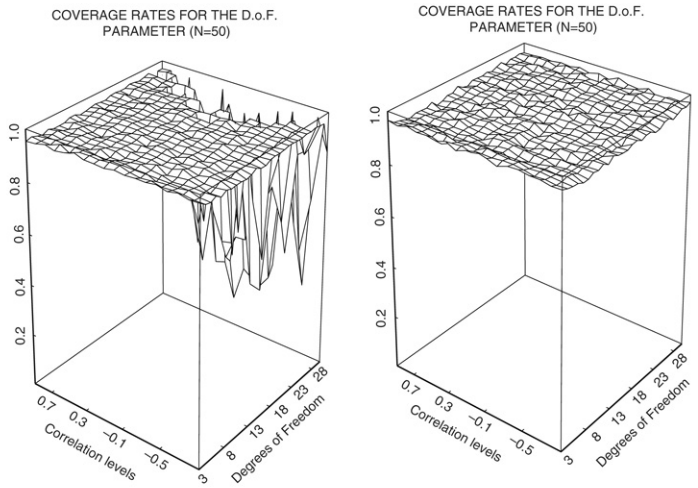

Fig. 1. Coverage rates for the 95% confidence intervals based on a normal approximation of the bivariate T-copula degrees of freedom ν, for $n = 50$. The first plot refer to the KME–CML method, while the second one to the ML method.

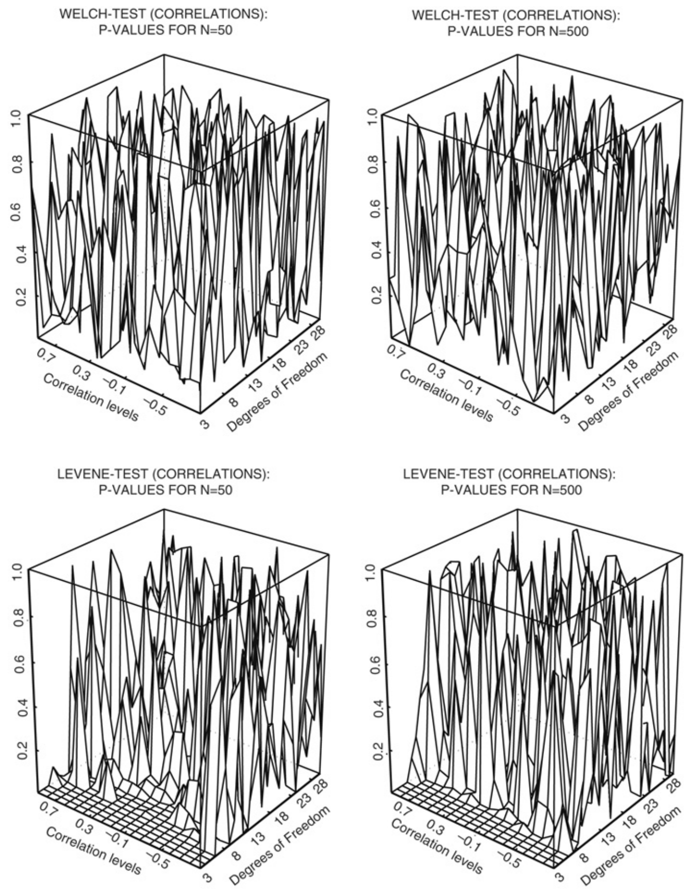

Fig. 2. The four plots report the p-values for the Welch tests (first row) and Levene tests (second row). The two samples compared in the tests are the *correlation parameters* of the bivariate T-copulas, estimated with the ML and KME–CML methods.

the KME–CML estimator. Furthermore, we examine different values for the degree of freedom ν, too, ranging between 3 and 30 (step 1), as well as two possible data situations: $n = 50$ and $n = 500$.

(3) We examine the case that ten variables have a multivariate Student's t copula, with the copula correlation matrix $\mathbf{\Sigma}$ equal to:

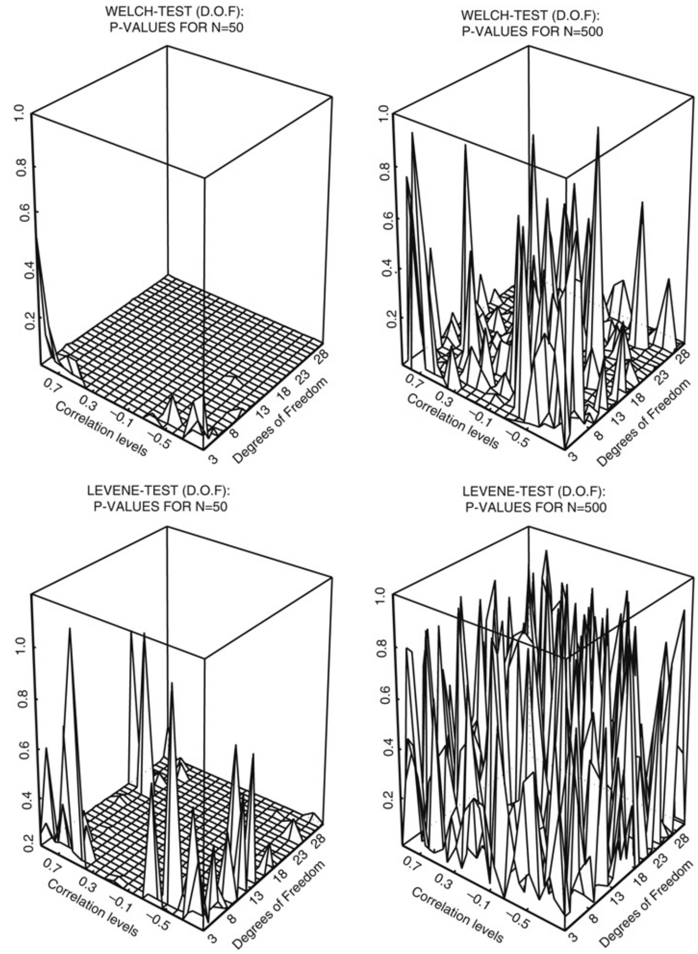

Fig. 3. The four plots report the *p*-values for the Welch tests (first row) and Levene tests (second row). The two samples compared in the tests are the *degrees of freedom parameters* of the bivariate *T*-copulas, estimated with the ML and KME–CML methods.

Table 1
Correlation matrix *T*-copula with lowest eigenvalue equal to 0.0768.

1	−0.15	−0.15	−0.15	−0.15	−0.14	−0.09	−0.03	0.05	0.13
−0.15	1	−0.15	−0.15	−0.15	−0.13	−0.08	−0.02	0.06	0.14
−0.15	−0.15	1	−0.15	−0.15	−0.12	−0.07	−0.01	0.07	0.15
−0.15	−0.15	−0.15	1	−0.15	−0.11	−0.06	0.01	0.08	0.15
−0.15	−0.15	−0.15	−0.15	1	−0.10	−0.05	0.02	0.09	0.15
−0.14	−0.13	−0.12	−0.11	−0.10	1	−0.04	0.03	0.10	0.15
−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	1	0.04	0.11	0.15
−0.03	−0.02	−0.01	0.01	0.02	0.03	0.04	1	0.12	0.15
0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	1	0.15
0.13	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	1

This is the correlation matrix of the returns of the first 10 stocks belonging to the Dow Jones Industrial Index, observed between 18/11/1988 and 20/11/2003. Furthermore, we examine different values for the degree of freedom ν, too, ranging between 3 and 30, as well as two possible data situations: $n = 50$ and $n = 500$.

The estimators considered are the three-stage KME–CML method described in Section 2, and the Maximum Likelihood estimator computed with given marginals, in order to assess the loss in efficiency associated with absence of knowledge of the marginals. We also considered the two-stage CML method (where $\mathbf{\Sigma}$ and ν are estimated jointly by ML), which delivered results in-between the KME–CML and ML methods, as expected. Therefore, we do not report its results for the sake of interest and space.

We generated 1000 Monte Carlo samples for each copula specification previously described and we estimated the T-copula parameters using both the KME–CML method and the ML method. Then, we computed the following evaluation statistics:

Table 2
Correlation matrix T-copula — returns Dow Jones Industrial Index.

1	0.21	0.33	0.22	0.36	0.30	0.37	0.34	0.31	0.47
0.21	1	0.20	0.15	0.27	0.18	0.18	0.31	0.20	0.21
0.33	0.20	1	0.16	0.32	0.28	0.40	0.33	0.17	0.42
0.22	0.15	0.16	1	0.20	0.16	0.18	0.20	0.27	0.20
0.36	0.27	0.32	0.20	1	0.32	0.33	0.55	0.33	0.35
0.30	0.18	0.28	0.16	0.32	1	0.28	0.32	0.26	0.31
0.37	0.18	0.40	0.18	0.33	0.28	1	0.35	0.23	0.40
0.34	0.31	0.33	0.20	0.55	0.32	0.35	1	0.31	0.35
0.31	0.20	0.17	0.27	0.33	0.26	0.23	0.31	1	0.30
0.47	0.21	0.42	0.20	0.35	0.31	0.40	0.35	0.30	1

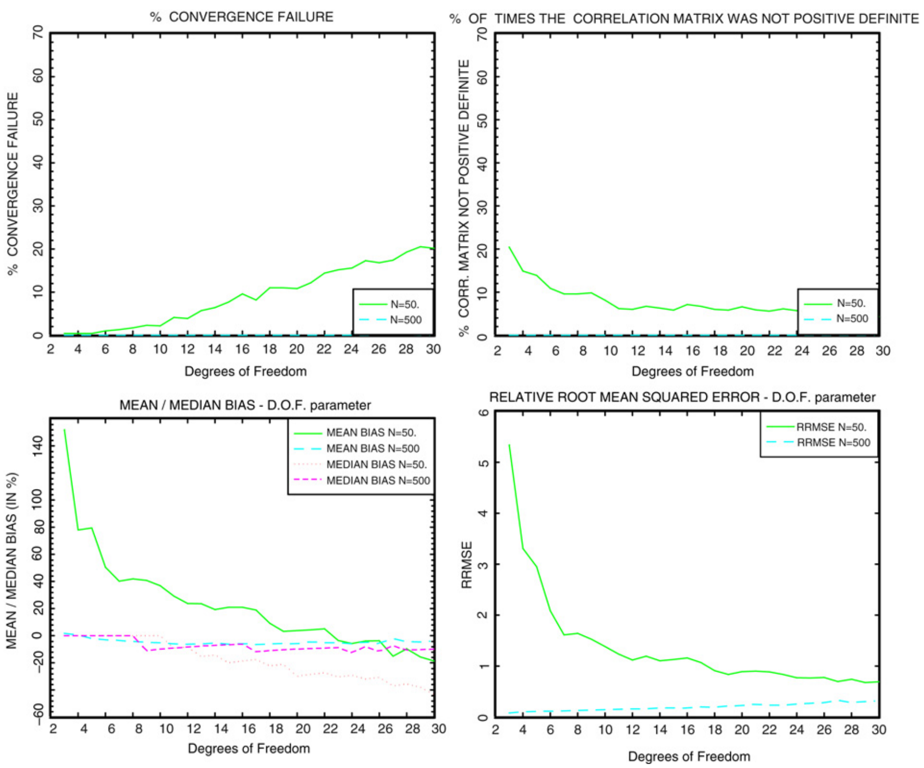

Fig. 4. The first plot reports the % of convergence failures when maximizing the log-likelihood. The second reports the % of times when the correlation matrix was not positive definite and the eigenvalue method was used. The third and the fourth plots report the Mean bias (in %), the Median bias (in %) and the Relative RMSE of the d.o.f. parameter, for the KME–CML method for the 10-variate T-copula with correlation matrix reported in Table 1.

- Mean bias (in %);
- Median bias (in %);
- Relative RMSE of the correlation and degrees of freedom parameters with respect to the true parameters;
- % of convergence failures when maximizing the log-likelihood. We used the `maxlik` library of the GAUSS software and a convergence tolerance for the gradient of the estimated parameters equal to 1e−5;
- % of times when the correlation matrix $\boldsymbol{\Sigma}$ was not positive definite and the eigenvalue method by Rousseeuw and Molenberghs (1993) was used;
- The coverage rate for the 95% confidence intervals of the T-copula parameters based on a normal approximation.
- t-tests under the null hypothesis that the sample mean across estimated parameters is equal to the true value.

Furthermore, we employed ANOVA based tests to test equality of the means and variances across the sample estimates delivered by the two considered methodologies. Particularly, we considered the following tests:

- The standard *F-test for the equality of variances of two groups*;
- The Levene test (1960) to assess the equality of variances in different samples. Levene test and the previous F-test are often used before a comparison of means: when Levene test is significant, modified procedures are used that do not

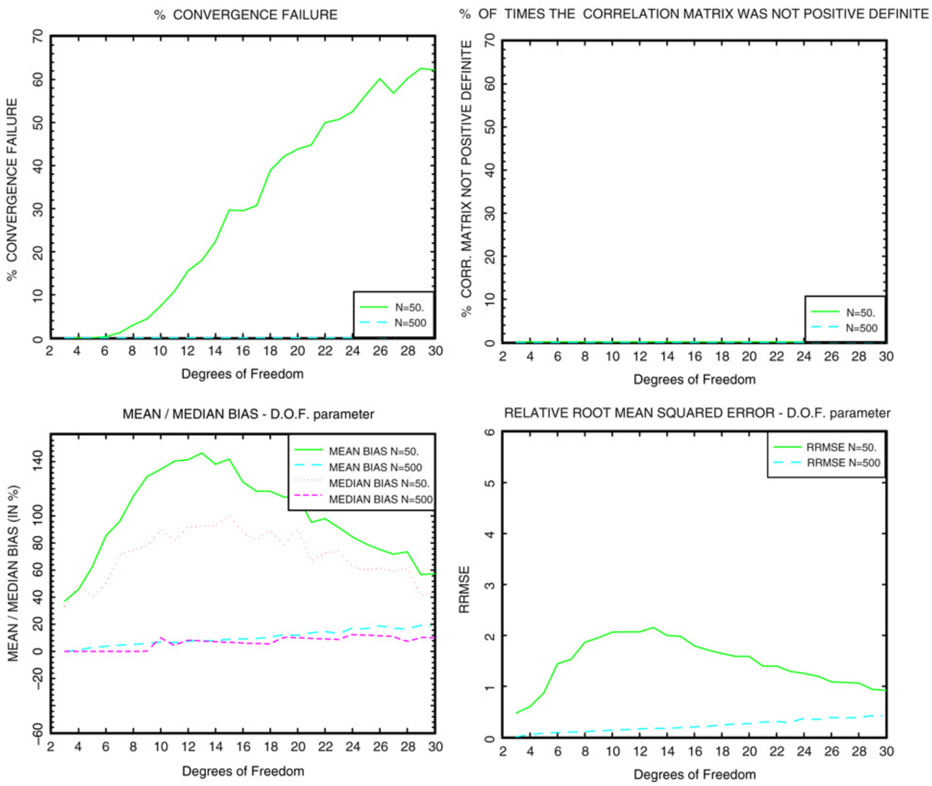

Fig. 5. The first plot reports the % of convergence failures when maximizing the log-likelihood. The second reports the % of times when the correlation matrix was not positive definite and the eigenvalue method was used. The third and the fourth plots report the Mean bias (in %), the Median bias (in %) and the Relative RMSE of the d.o.f. parameter, for the ML method for the 10-variate T-copula with correlation matrix reported in Table 1.

assume equality of variance:

$$F_{Levene} = \frac{N-G}{G-1} \cdot \frac{\sum_{g=1}^{G} n_g(\bar{Z}_g - \bar{Z})^2}{(G-1)\sum_{g=1}^{G}\sum_{j=1}^{n_g}(Z_{gj} - \bar{Z}_g)^2}$$

where $Z_{gj} = |x_{gj} - \bar{x}_g|$, $g = 1, \ldots, G$, $\bar{Z}$ is the overall mean of all Z_{gj}, $\bar{Z}_g$ is the mean of the Z_{gj} for group g, while n_g is the number of observations in group g. The F-statistic for the Levene test has an approximate F-distribution with $G = 1$ numerator degrees of freedom and $N - G$ denominator degrees of freedom under the null hypothesis of equal variances in each subgroup (Levene, 1960).

- The standard *F-test for the equality of means in two groups*;
- If the two groups' variances are heterogeneous, it is advisable to use the Welch (1951) *version of the previous test statistic*. The basic idea is to form a modified F-statistic that accounts for the unequal variances. Using the Cochran (1937) weight function,

$$w_g = n_g / s_g^2$$

where s_g^2 is the sample variance in group g, $g = 1, \ldots, G$, we may form the modified F-statistic as follows:

$$F^* = \frac{\sum_{g=1}^{G} w_g(\bar{x}_g - \bar{x}^*)^2/(G-1)}{1 + \frac{2(G-2)}{G^2-1}\sum_{g=1}^{G}\frac{(1-h_g)^2}{n_g-1}}$$

where $\bar{x}_g$ is the sample mean within group g, h_g is a normalized weight and $\bar{x}^*$ is the weighted grand mean,

$$h_g = w_g \Big/ \left(\sum_{g=1}^{G} w_k\right),$$

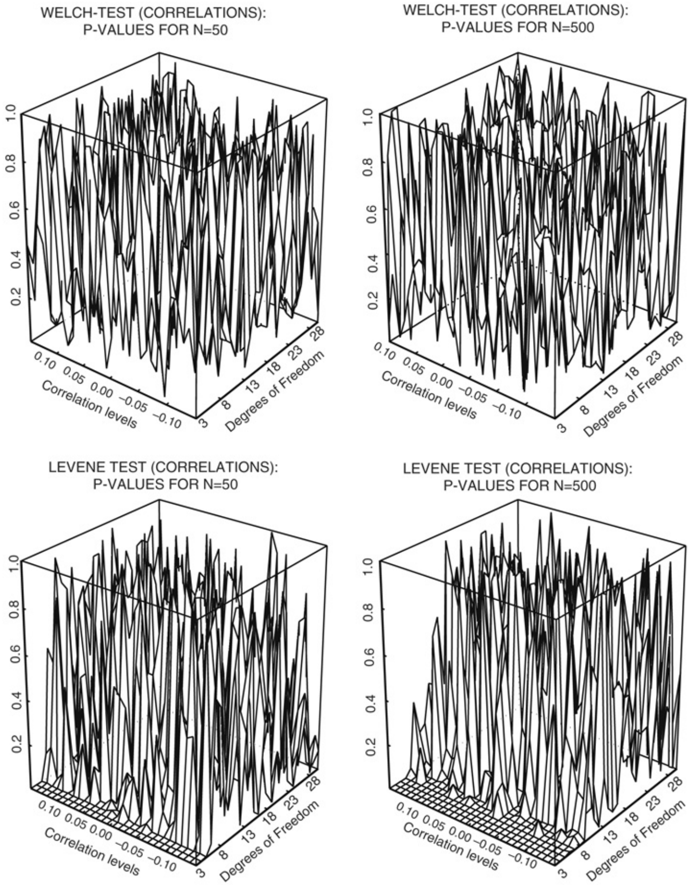

Fig. 6. The four plots report the p-values for the Welch tests (first row) and Levene tests (second row). The two samples compared in the tests are the *correlation parameters* of the T-copula with correlation matrix reported in Table 1, estimated with the ML and KME–CML methods.

$$\bar{x}^* = \sum_{g=1}^{G} h_g \bar{x}_g.$$

The numerator of the adjusted statistic is the weighted between-group mean squares and the denominator is the weighted within-group mean squares. Under the null hypothesis of equal means but possibly unequal variances, F^* has an approximate F-distribution with $(G-1, DF^*)$ degrees of freedom, where

$$DF^* = \frac{(G^2-1)}{3\sum_{g=1}^{G}\frac{(1-h_g)^2}{n_g-1}}.$$

Note that for $G = 2$ groups, this test reduces to the Satterthwaite (1946) test.

The standard F-tests for the equality of means and variance of two groups are not reported in the paper, but are available from the author upon request.

4.1. Simulation results: Bivariate T-copula

For the sake of interest and space, we report in the main manuscript only a limited set of results, while most of them are reported in the Technical Report by Fantazzini (2009). Particularly, Fig. 1 reports the coverage rates for the 95% confidence intervals based on a normal approximation of the bivariate T-copula degrees of freedom, for $n = 50$, while Figs. 2 and 3 report the p-values for the Welch tests and Levene tests, under the null hypothesis that the sample means and variances of the KME–CML and ML methods are equal, respectively.

The remaining set of results are reported in the Technical report by Fantazzini (2009): particularly, Figures 12–13 there report the Mean bias (in %), Median bias (in %) and Relative RMSE of the correlation and degrees of freedom parameters,

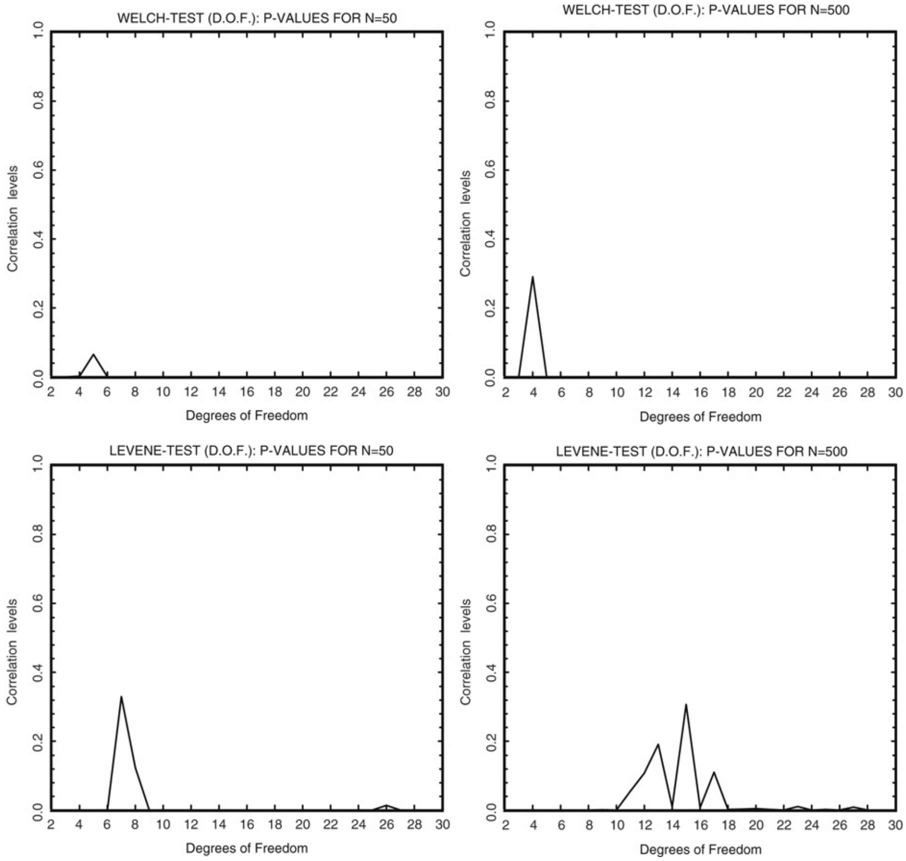

Fig. 7. The four plots report the p-values for the Welch tests (first row) and Levene tests (second row). The two samples compared in the tests are the *degrees of freedom parameters* of the T-copula with correlation matrix reported in Table 1, estimated with the ML and KME–CML methods.

respectively, for the KME–CML method, across different correlation levels and degrees of freedom parameters, as well as different data samples. Figure 14 in Fantazzini (2009) reports the coverage rate (in %) for a large sample 95% confidence interval based on a normal approximation for ν and ρ, together with the % of convergence failures when maximizing the log-likelihood. Figures 15–17 in Fantazzini (2009) report the same simulation statistics but for the ML method. Finally, Figures 26 (first column)–27 (first column) report the t-tests under the null hypothesis that the sample means across estimated correlations and degrees of freedom are equal to the true values, respectively.

The simulation studies show some interesting results:

With regard to the ***degrees of freedom*** "ν", when considering a small sample with $n = 50$ observations, both the ML and the KME–CML methods show poor results, with high mean and median biases, as well as high percentage of numerical convergence failures. However, if a low value for ν is used, the KME–CML method shows better results than the ML estimator, while the reverse is true if ν is high, that is if the T-copula becomes no more distinguishable from a Normal copula (the T-copula tends to the Normal copula when $\nu \rightarrow \infty$). The same results are confirmed by the t-tests in Fig. 27 in Fantazzini (2009) (first column).

When $n = 50$ and ν is high, the % of time when the numerical maximization of the log-likelihood failed to converge is much higher for the ML method than for the KME–CML method (over 50% vs 30%, respectively). However, while the coverage rates at the 95% level for the ML estimates do not show any particular bias or trend, the KME–CML estimates for ν show very low coverage rates when ν becomes close to 30 and the correlation is not too strong (between -0.5 and 0.5). Therefore, the lower number of convergence failures for the KME–CML method comes at a cost: the estimates show much stronger negative mean (and median) biases than the ML method, around the 70% of the true value vs the 50%, respectively. As a consequence, the confidence intervals based on a normal approximation are very poor. When a larger sample with $n = 500$ observations is considered, the two methods perform rather well and the ML estimator shows better properties, as expected. Moreover, the problem with the coverage rates for the KME–CML method disappears.

As for the ***correlation*** "ρ", the two methods produce consistent estimates already with very small samples and the analysis reveals no major difference between the two. However, if the variables are close to be uncorrelated, both methods present small biases and RRMSE that decrease when n increases (see Figures 12 and 15 in the Technical Appendix in Fantazzini (2009)).

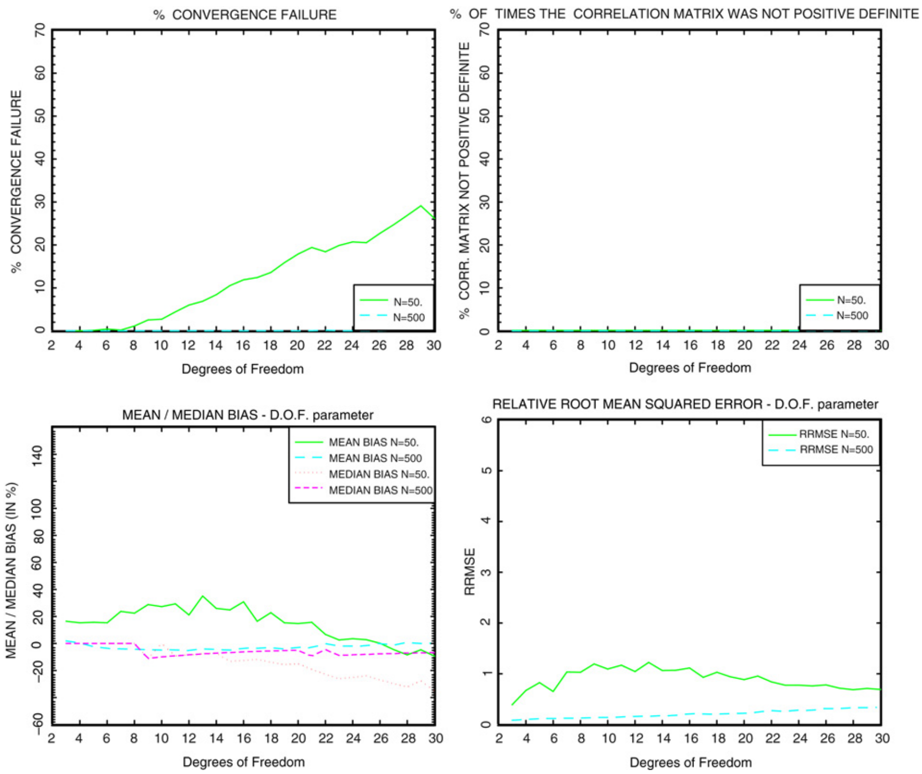

Fig. 8. The first plot reports the % of convergence failures when maximizing the log-likelihood. The second reports the % of times when the correlation matrix was not positive definite and the eigenvalue method was used. The third and the fourth plots report the Mean bias (in %), the Median bias (in %) and the Relative RMSE of the d.o.f. parameter, for the KME–CML method for the 10-variate T-copula with correlation matrix reported in Table 2.

With regard to ***ANOVA tests***, when the correlation parameters are of concern, the Welch tests in Fig. 2 show that the sample estimates delivered by the two methods (KME–CML and ML) have means which are not statistically different. A similar evidence is delivered by the Levene tests for the variances which, however, are statistically different when the true degrees of freedom ν are low. As for the degrees of freedom parameter, the Welch and Levene tests in Fig. 3 point out that for $n = 50$ the two methods deliver statistically different estimates, particularly with high values for ν and low correlations, thus confirming the previous evidence with coverage rates.

As regards the ***computational aspects***, as anticipated in the previous points, the ML method shows much higher convergence failures than the the KME–CML method when $n = 50$ and ν is higher than 10, while when $n = 500$ the numerical performances of the two methods are quite close. Furthermore, the analysis shows that when ν is high and the KME–CML is employed, the % of convergence failures is higher when T-copulas with *weaker* correlations are considered, reaching the maximum when the variables are uncorrelated (see the lower plots in Figure 14 in Fantazzini (2009)). This pattern is not present for the ML method, instead. This result together with the stronger negative biases of $\hat{\nu}_{KME-CML}$, helps us to explain the U-shape drop in coverage rates observed in Fig. 1 for the KME–CML method.

4.2. Simulation results: 10-variate T-copula — Ill-specified correlation matrix

For the sake of interest and space, we report in Figs. 4 and 5 the % of convergence failures when maximizing the log-likelihood, the % of times when the correlation matrix was not positive definite and the eigenvalue method was used, as well as the Mean bias (in %), the Median bias (in %) and the Relative RMSE of the d.o.f. parameter, for both the KME–CML and ML methods. Figs. 6 and 7 report the p-values for the Welch tests and Levene tests.

The remaining set of results are reported in the Technical Appendix in Fantazzini (2009): particularly, Figure 18 reports the Mean bias (in %), Median bias (in %) and Relative RMSE of the correlation parameters, for the KME–CML method, across different correlation levels and degrees of freedom as well as different data samples. The coverage rates for the 95% confidence intervals of the T-copula parameters based on a normal approximation are reported in Figure 20. Figures 19 and 21 in Fantazzini (2009) report the same simulation statistics but for the ML method. Finally, Figures 26 (second column)–27 (second column) in Fantazzini (2009) report the t-tests under the null hypothesis that the sample means across estimated correlations and degrees of freedom are equal to the true values, respectively.

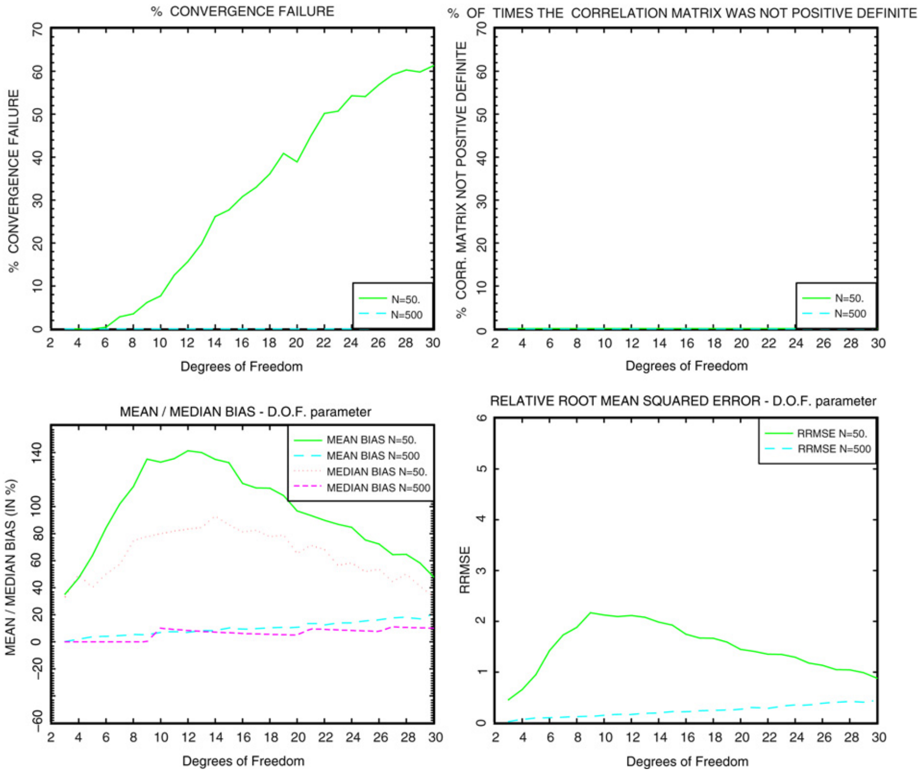

Fig. 9. The first plot reports the % of convergence failures when maximizing the log-likelihood. The second reports the % of times when the correlation matrix was not positive definite and the eigenvalue method was used. The third and the fourth plots report the Mean bias (in %), the Median bias (in %) and the Relative RMSE of the d.o.f. parameter, for the ML method for the 10-variate T-copula with correlation matrix reported in Table 2.

With regard to the ***degrees of freedom*** "ν", when $n = 50$, ν *is low* and the KME–CML method is employed, the analysis shows a strong positive mean bias and high RRMSE (see Fig. 4). However, the median bias is close to zero and the coverage rates are similar to the case with the ML method. This difference is due to the high % rates when the correlation matrix is not positive definite (between 10% and 20%) and the eigenvalue method by Rousseeuw and Molenberghs (1993) is used, see the second plot in Fig. 4. The analysis shows that this particular ad hoc fix has the effect to introduce a positive mean bias in $\hat{\nu}$, but the effect on the median is rather limited as well on the coverage rates. Besides, this bias quickly disappears when ν increases. Similar results can be retrieved by looking at the t-tests in Figure 27 in Fantazzini (2009) (second column).

Instead, when ν *is high* and the KME–CML method is employed, the simulation results show a negative median bias that has the effect to decrease the coverage rates below the 95% level. However, we remark that the drop in the coverage rates is much lower than those observed with bivariate T-copulas.

The ML method shows low positive mean and median biases as well as low RRMSE when ν is low. Then, these biases increase when ν increases, reaching the maximum around $\nu = 15$, after which they finally decrease (see Fig. 5). The effect on the 95% coverage rate is just specular: it is high when ν is low, then it decreases and finally it converges to the true value when ν is higher than 20.

When a larger sample with $n = 500$ observations is considered, the two methods perform rather well and the properties of the two estimators are quite similar. Interestingly, when $n = 500$ the correlation matrix is always positive definite for all ν, and the eigenvalue method by Rousseeuw and Molenberghs (1993) is not needed.

As for the ***correlation matrix***, when $n = 50$, both the ML and the KME–CML methods show mean and median biases as well as high RRMSE when the correlations are close to zero, see Figures 18 and 19 in Fantazzini (2009). However, the effects on the coverage rates are rather limited (see upper plots in Figures 20 and 21 in Fantazzini (2009)). Besides, the KME–CML method shows slightly better results than the ML method.

When $n = 500$ the previous biases decrease but they still remain quite high for the ML method when ν is low and the variables are close to be uncorrelated.

With regard to ***ANOVA tests***, when the correlation parameters are of concern, the Welch tests in Fig. 6 show again that the sample estimates delivered by the two methods (KME–CML and ML) have means which are not statistically significant different. A similar evidence is delivered by the Levene tests for the variances which are statistically different only when the true degrees of freedom ν are low.

As for the degrees of freedom parameter, the Welch and Levene tests in Fig. 7 point out that the two methods deliver statistically different estimates.

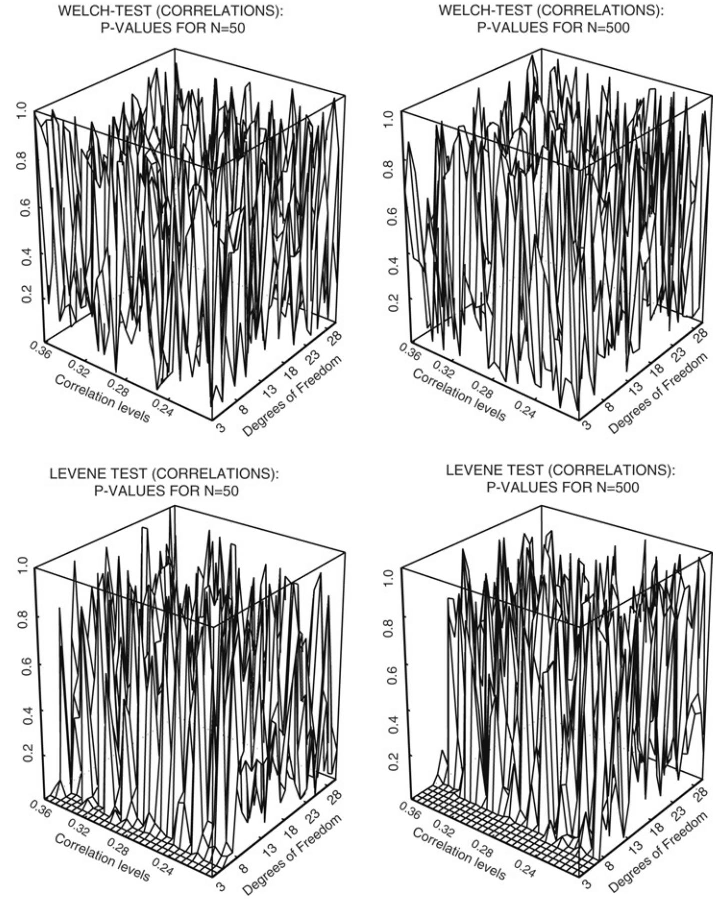

Fig. 10. The four plots report the p-values for the Welch tests (first row) and Levene tests (second row). The two samples compared in the tests are the *correlation parameters* of the T-copula with correlation matrix reported in Table 2, estimated with the ML and KME–CML methods.

Regarding the ***computational aspects***, similarly to the bivariate case, the ML method shows higher convergence failures than the KME–CML method when $n = 50$ and ν is higher than 10, while when $n = 500$ the numerical performances of the two methods are equal and there are no convergence failures. Furthermore, as previously discussed, when $n = 50$, ν is low and the KME–CML method is employed, the correlation matrix is not positive definite in 20% of cases and the eigenvalue method by Rousseeuw and Molenberghs (1993) has to be used. This fix induces a positive bias in the estimate of ν, but the effects on the coverage rates are rather limited. Besides, the number of times when this method has to be used quickly diminish when ν increases. Interestingly, the correlation matrix is always positive definite already with $n = 500$.

4.3. Simulation results: 10-variate T-copula — Dow Jones returns

For the sake of interest and space, we report in Figs. 8 and 9 the % of convergence failures when maximizing the log-likelihood, the % of times when the correlation matrix was not positive definite and the eigenvalue method was used, as well as the Mean bias (in %), the Median bias (in %) and the Relative RMSE of the d.o.f. parameter, for both the KME–CML and ML methods. Figs. 10 and 11 report the p-values for the Welch tests and Levene tests.

The remaining set of results are reported in the Technical Appendix in Fantazzini (2009): particularly, Figure 22 there reports the Mean bias (in %), Median bias (in %) and Relative RMSE of the correlation parameters, for the KME–CML method, across different correlation levels and degrees of freedom as well as different data samples. The coverage rate for the 95% confidence intervals of the T-copula parameters based on a normal approximation are reported in Figure 24 in Fantazzini (2009). Figures 23 and 25 report the same simulation statistics but for the ML method. Finally, Figures 26 (third column)–27 (third column) in Fantazzini (2009) report the t-tests under the null hypothesis that the sample means across estimated correlations and degrees of freedom are equal to the true values, respectively.

With regard to the ***degrees of freedom*** "ν", when $n = 50$, both the KME–CML and the ML methods show an inverted U-shape for the positive mean and median biases as well as for the RRMSE, as reported in Figs. 8 and 9. However, the biases for the KME–CML method are much lower than the ML method (20%–120% vs 0%–30%, respectively). Furthermore, while the biases for the ML method are very close to those observed for the ill-specified T-copula reported in Fig. 5, the mean

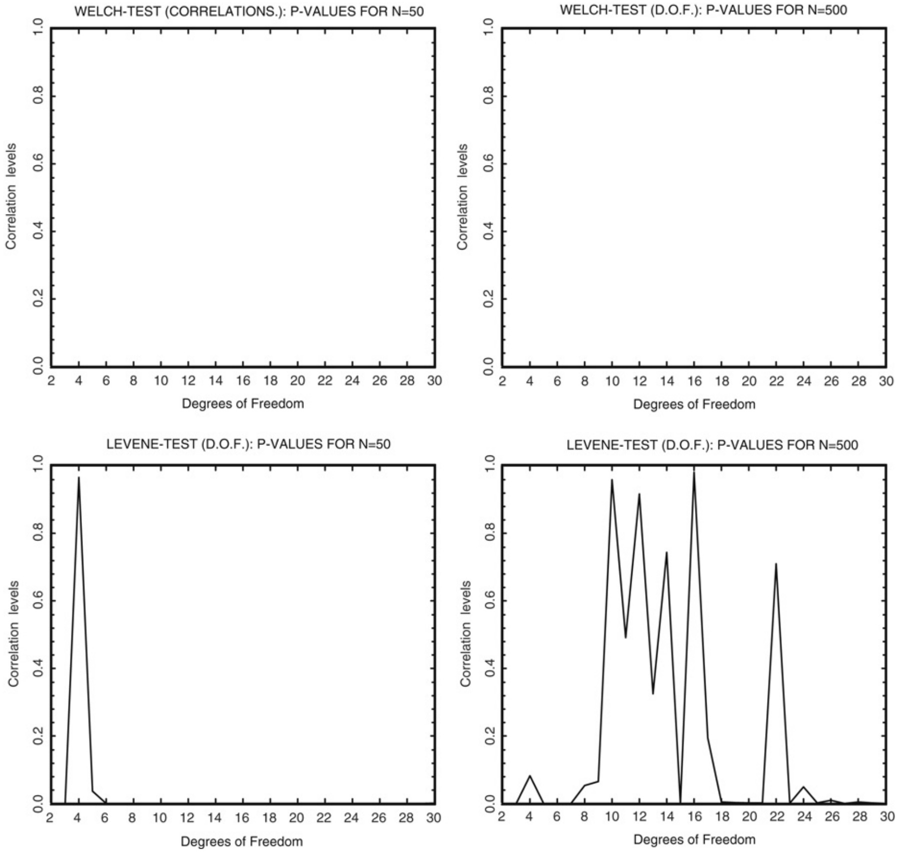

Fig. 11. The four plots report the p-values for the Welch tests (first row) and Levene tests (second row). The two samples compared in the tests are the *degrees of freedom parameters* of the T-copula with correlation matrix reported in Table 2, estimated with the ML and KME–CML methods.

biases observed for the KME–CML method are different from those observed for the same technique when dealing with the ill-specified T-copula, and reported in Fig. 4. This difference is due to the complete lack of negative definite correlation matrices (see Fig. 8), so that the eigenvalue method by Rousseeuw and Molenberghs (1993) is not needed.

Instead, both methods show median biases that are very close to those observed for the ill-specified T-copula, and similarly the 95% coverage rates reported in Figures 24 and 25 (lower plots) in Fantazzini (2009) are very close to those observed in Figures 20 and 21 (lower plots) in Fantazzini (2009), respectively: that is a U-shape that converges to the true rate when ν increases for the ML method, while a decreasing trend for the KME–CML method. However, we remark that the drop in the coverage rates for the KME–CML method is again much lower than that observed with bivariate T-copulas.

When a larger sample with $n = 500$ observations is considered, the two methods perform rather well and the properties of the two estimators are quite similar.

As for the ***correlation matrix***, the main results are close to those observed for the 10-variate T-copula with ill-specified correlation matrix: when $n = 50$, both the ML and the KME–CML methods show mean and median biases as well as high RRMSE when the correlations are close to zero, as shown in Figures 22 and 23 in Fantazzini (2009). However, the effects on the coverage rates are rather limited (see the upper plots in Figures 24 and 25 in Fantazzini (2009)). Besides, the KME–CML method shows slightly better results than the ML method. Differently from the ill-specified T-copula, the biases are much more lower, ranging between $-10/+10\%$ instead of $-100\%/+100\%$.

When $n = 500$ the previous biases decrease but they still remain quite high for the ML method when ν is low and the variables are close to be uncorrelated.

With regard to ***ANOVA tests***, when the correlation parameters are of concern, the Welch tests in Fig. 10 show again that the sample estimates delivered by the two methods (KME–CML and ML) have means which are not statistically significant different. A similar evidence is delivered by the Levene tests for the variances which are statistically different only when the true degrees of freedom ν are low.

As for the degrees of freedom parameter, the Welch and Levene tests in Fig. 11 point out that the two methods deliver statistically different estimates. However, when the dimension n increases, the Levene tests highlight that the variances of the sample estimates delivered by the two methods are no more statistically different.

Regarding the ***computational aspects***, similarly to the previous analysis, the ML method shows higher convergence failures than the the KME–CML method when $n = 50$ and ν is higher than 10, while when $n = 500$ the numerical performances of the two methods are equal and there are no convergence failures. However, as previously discussed, the eigenvalue method by Rousseeuw and Molenberghs (1993) is not needed already with $n = 50$: this simulation evidence confirms previous empirical evidence in Demarta and McNeil (2005)and McNeil et al. (2005, chapter 5) who claim that the componentwise transformation of the empirical Kendall's tau matrix is positive definite in most cases.

Therefore, the previous results suggest to use the KME–CML method when dealing with small samples and low degrees of freedom, while the ML method is a better choice otherwise. A possible strategy would be to first use the KME–CML method: if the estimated degrees of freedom are higher than 20, then one should try to use the ML method if it converges. Otherwise, an alternative solution would be to use the simple normal copula, given that the T-copula tends to the Normal copula when $\nu \to \infty$, and the two copulas are already quite close when $\nu > 20$. Besides, the drop in the 95% coverage rates highlighted by our analysis for the KME–CML method when ν is high, is quite low if the number of variables is high: this is the usual case for financial professionals, whose managed portfolios are rarely bivariate, but include a large number of assets to diversify financial risk.

5. Conclusions

We developed the asymptotics of a recent semi-parametric estimation method used in the financial literature with the multivariate Student's T-copula, which involves empirical distribution functions, method-of-moments and maximum likelihood methods. We examined the finite-sample properties of this estimator via a Monte Carlo study designed to replicate different Data Generating Processes, and we found that this estimator was more efficient and less biased than the one-stage ML estimator when small samples and T-copulas with low degrees of freedom were of concern.

We then analyzed the pros and cons of this methodology in terms of numerical convergence and positive definiteness of the estimated T-copula correlation matrix. When small samples were of concern and ν was high, the number of times when the numerical maximization of the log-likelihood failed to converge was much higher for the ML method than for the KME–CML method. Yet, while the coverage rates at the 95% level for the ML estimates for ν did not show any particular bias or trend, the KME–CML estimates showed very low rates when ν became close to 30 and the correlations were not too strong. However, this drop in the coverage rates was large with bivariate T-copulas, only, while it was much lower when dealing with higher dimensional T-copulas, which is the usual case for real managed financial portfolios. Besides, both the ML and the KME–CML methods showed high mean and median biases for the estimated correlations when the true ones were close to zero. Nevertheless, the effects on the coverage rates for correlations were rather limited in this case.

Finally, we showed that the eigenvalue method by Rousseeuw and Molenberghs (1993) has to be used to obtain a positive definite correlation matrix only when dealing with very small samples ($n < 100$) and when the true underlying process has the lowest eigenvalue close to zero. This fix induces a positive bias in the estimate of ν, but the effects on the coverage rates are rather limited. Besides, the number of times when this method has to be used quickly decreases when ν increases.

Acknowledgements

For helpful comments and suggestions I would like to thank Paolo Giudici, Antonio Lijoi, Mario Maggi, Pietro Rigo, Eduardo Rossi. Special thanks go to the participants of the 63rd European Meeting of the Econometric Society, 27–31 August 2008, Milan, for helpful comments.

References

Barbe, P., Genest, C., Ghoudi, K., Remillard, B., 1996. On Kendalls process. Journal of Multivariate Analysis 58, 197–229.
Bouye', E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T., 2001. Copulas for finance a reading guide and some applications. Groupe de Recherche Operationnelle, Credit Lyonnais, Working Paper.
Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. Wiley.
Clayton, D.G., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65 (1), 141–151.
Clemen, M.N., Jouini, R.T., 1996. Copula models for aggregating expert opinions. Operations Research 44, 444–457.
Chen, X., Fan, Y., 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. Journal of Econometrics 135, 125–154.
Cochran, W.G., 1937. Problems arising in the analysis of a series of similar experiments. Journal of the Royal Statistical Society 4 (Suppl. 1), 102–118.
Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: Properties and pitfalls. In: Dempster, M.A.H. (Ed.), Risk Management: Value at Risk and Beyond. Cambridge University Press, pp. 176–223.
Dalla Valle, L., Fantazzini, D., Giudici, P., 2008. Copulae and operational risks. International Journal of Risk Assessment and Management 9 (3), 238–257.
Demarta, S., McNeil, A.J., 2005. The t copula and related copulas. International Statistical Review 73, 111–129.
Dobric, J., Schmid, F., 2006. A goodness of fit test for copulas next term based on Rosenblatt's transformation. Computational Statistics and Data Analysis 51 (9), 4633–4642.
Fantazzini, D., 2009. Three-stage semi-parametric estimation of T-copulas: Asymptotics. Finite-Sample Properties and Computational Aspects. Moscow School of Economics Technical Report, Ferbuary 2009.
Fantazzini, D., 2009. The effects of misspecified marginals and copulas on computing the value at risk: A Monte Carlo study. Computational Statistics and Data Analysis, in press, doi:10.1016/j.csda.2008.02.002.
Genest, C., MacKay, R.J., 1986. Copules Archimédiennes et Familles de Lois Bidimensionnelles dont les Marges sont Donnes. Canadian Journal of Statistics 14, 145–159.
Genest, C., MacKay, R.J., 1986. The joy of copulas: Bivariate distributions with uniform marginals. The American Statistician 40, 280–285.
Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88, 1034–1043.

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

Genest, C., Quessy, J.F., Remillard, B., 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. Scandinavian Journal of Statistics 33, 337–366.

Granger, C.W.J., Patton, A.J., Tërasvirta, T., 2006. Common factors in conditional distributions for bivariate time series. Journal of Econometrics 132, 43–57.

Greene, W., 2008. Econometric Analysis, 6th ed. Prentice Hall.

Höffding, D., 1940. Masstabinvariante Korrelationstheorie. Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität 5, 181–233.

Huarda, D., Evin, G., Favre, A.C., 2006. Bayesian copula selection. Computational Statistics and Data Analysis 51 (2), 809–822.

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

Karlis, D., Nikoloulopoulos, A.K., 2008. Copula model evaluation based on parametric bootstrap. Computational Statistics and Data Analysis 52 (7), 3342–3353.

Kim, G., Silvapulle, M.J., Silvapulle, P., 2005. Semiparametric estimation of the joint distribution of two variables when each variable satisfies a regression model. Joint Statistical Meetings, August 2005, Minneapolis [Available as Working Paper Number 1/07, Department of Econometrics and Business Statistics, Monash University].

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

Kim, G., Silvapulle, M.J., Silvapulle, P., 2008. Estimating the error distribution in multivariate heteroscedastic time series models. Journal of Statistical Planning and Inference 138 (5), 1442–1458.

Koul, H.L., Ling, S., 2006. Fitting an error distribution in some heteroscedastic time series model. Annals of Statistics 34, 994–1012.

Koul, H.L., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models. In: Lecture Notes in Statistics, vol. 166. Springer Verlag.

Kowalczyka, T., Niewiadomska-Bugaj, M., 2002. An algorithm for maximizing Kendall's τ. Computational Statistics and Data Analysis 37 (2), 181–193.

Kruskal, W., 1958. Ordinal measures of association. Journal of the American Statistical Association 53, 814–861.

Levene, H., 1960. Robust tests for the equality of variances. In: Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G., Mann, H.B. (Eds.), Contribution to Probability and Statistics. Stanford University Press, Palo Alto, CA.

Lindskog, F., 2000. Modelling dependence with copulas. RiskLab Report, ETH Zurich.

Lindskog, F., McNeil, A.J., Schmock, U., 2002. Kendall's tau for elliptical distributions. In: Credit Risk Measurement Evaluation and Management. Physica-Verlag.

Marshall, R., Zeevi, A., 2002. Beyond correlation: Extreme co-movements between financial assets. Columbia University, working paper.

McNeil, A.J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton University Press.

Oakes, D., 1994. Multivariate survival distributions. Journal of Nonparametric Statistics 3 (3–4), 343–354.

Patton, A.J., 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. Journal of Financial Econometrics 2 (1), 130–168.

Patton, A.J., 2005. Estimation of copula models for time series of possibly different lengths. Journal of Applied Econometrics 21 (2), 147–173.

Nelsen, R.B., 1999. An Introduction to Copulas. In: Lecture Notes in Statistics, vol. 139. Springer.

Rosenberg, J.V., 1998. Pricing multivariate contingent claims using estimated risk-neutral density functions. Journal of International Money and Finance 17, 229–247.

Rosenberg, J.V., 2003. Nonparametric pricing of multivariate contingent claims. Journal of Derivatives 10 (3), 9–26.

Rousseeuw, P., Molenberghs, G., 1993. Transformation of non positive semidefinite correlation matrices. Communications in Statistics. Theory and Methods 22, 965–984.

Satterthwaite, F.E., 1946. An approximate distribution of estimates of variance components. Biometrics Bulletin 2 (6), 110–114.

Shih, J.H., Louis, T.A., 1995. Inferences on the association parameter in copula models for bivariate survival data. Biometrics 51, 1384–1399.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Université de Paris 8, 229–231.

Taniguchi, M., Kakizawa, Y., 2000. Asymptotic Theory of Statistical Inference for Time Series. Springer, NewYork.

Wang, W., Wells, M.A., 2000. Model selection and semiparametric inference for bivariate failure-time data. Journal of the American Statistical Association 95, 62–72.

Welch, B.L., 1951. On the comparison of several mean values: An alternative approach. Biometrika 38, 330–336.