Efficient maximum likelihood estimation of copula based meta t-distributions

Claudia Czado[a], Ran Zhang[a,*], Aleksey Min[a]

[a]*Zentrum Mathematik, Technische Universität München, Germany*

Abstract

Recently an efficient fixed point algorithm for finding maximum likelihood estimates has found its application in models based on Gaussian copulas. It requires a decomposition of a likelihood function into two parts and their iterative maximization. Therefore, this algorithm is called maximization by parts (MBP). For copula-based models, the algorithm MBP improves the efficiency of a two-step estimation approach called inference for margins (IFM) and is an promising alternative method to direct maximization of the likelihood function (DIR). For the first time, the MBP algorithm is derived and applied to Student t-copula based models. A superiority of the proposed algorithm over IFM and DIR methods is illustrated in a simulation study for data with small sample sizes. This makes the proposed algorithm an excellent candidate for estimation in a rolling window set up, which is able to account for time varying dependency structures. This approach is followed by the analysis of swap rates demonstrating the necessity of time varying copula effects.

Key words: Copula, inference for margins, maximum likelihood estimation, maximization by parts, meta-t distribution, rolling windows

1. Introduction

Due to Sklar's theorem (see Sklar, 1959) copulas describe an invariant dependence structure of multivariate distributions. Thus, using any monotone transformations, marginal variables of a given multivariate distribution can arbitrarily be transformed but the corresponding copula stays the same and does not change at all. Since Frees and Valdez (1998) and Embrechts et al. (1999) copulas have been started to be widely used in economics, finance and risk management. On the other hand, copulas have been used for the construction of flexible and more general multivariate distributions (see e.g. Johnson and Kotz 1972; Dall'Aglio et al. 1991; Krzysztofowicz and Kelly 1997). In particular, Fang et al. (2002) construct a class of meta-elliptical distributions with specified univariate marginal distributions using copulas of multivariate elliptically contoured distributions. For the theory and applications of elliptically contoured distributions we refer readers to Cambanis et al. (1981), Fang et al. (1990) and Embrechts et al. (2003).

Recently Song et al. (2005) introduced a new fixed point algorithm for finding the maximum likelihood estimates (MLE's) iteratively. This method is called the maximization by parts (MBP) algorithm since it requires a decomposition of a log likelihood function into two parts. The first part should be easily optimized and is called the working model, while the second part is called an error model. It involves additional variable over the ones in the working model. The global maximization is then determined in an iterative way. The purpose of this algorithm is to overcome optimization problems which can be encountered by using other known algorithms such as directed maximization (DIR) or expectation maximization (EM) algorithms. However, it

*Corresponding author: Zentrum Mathematik, Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany

Email addresses: cczado@ma.tum.de (Claudia Czado), ran.zhang@ma.tum.de (Ran Zhang), aleksmin@ma.tum.de (Aleksey Min)

seems that the MBP method is most appropriate for copula based models, where log-likelihood functions are already naturally decomposed into two parts. For copula based models, the working model usually consists only of copula parameters, while the error model includes copula as well as marginal parameters of the copula-based model.

Since Song et al. (2005) carried out the theory of maximization by parts many researchers have scrutinized MBP algorithm in numerous applications. Liu and Luger (2009) applied the MBP algorithm to copula-GARCH model based on a Gaussian copula and examined the performance of their MBP algorithm compared to so-called inference for margins (IFM) method (see Shih and Louis 1995; Joe and Xu 1996). They showed that the IFM strategy for estimating copula-GARCH models can be improved using the MBP algorithm. Further, Song et al. (2007) investigated the linear mixed-effects models using multivariate t-distributions by using the MBP algorithm to maximize the corresponding likelihood. They found out that the MBP algorithm is recommended to both high and low dimensional linear mixed-effects models. Fan et al. (2007) derived the MBP algorithm of Song et al. (2005) for more general extremum estimates which include the generalized method of moment (GMM) estimates and least absolute deviate estimates as well as MLE's. Further they investigated asymptotic properties of extremum estimates obtained via MBP, illustrated efficiency of their algorithm over existing ones for MLE's in the Merton's credit risk model and gave applications of the MBP algorithm for GMM estimates.

It is well known that economy has cycles and that the economic environment can change rapidly due to unpredictable circumstances. For many practitioners and researchers estimation of a time varying dependence structure in financial time series is of major interest. One easy approach to identify the need for models with time varying structures is to employ a rolling window approach. For example, Aussenegg and Cech (2008) applied a rolling windows technique to Eurostoxx 50 and Dow Jones Industrial 30 stock indices in order to compare Gaussian and Student t-copulas. As it is pointed out in Zivot and Wang (2003), this technique allows us to evaluate the model's stability over a time interval. If parameters of interest are truly constant over the entire sample, then the estimates over the rolling windows should not differ from each other very much. If the parameters change at some point during the sample significantly, it identifies the instability of the model and therefore the need to choose a model with time varying dependency variables. They might be slowly varying or rapidly varying. For rapidly varying structure Markov switching models might be useful.

In this paper we first derive a MBP algorithm for multivariate meta t-distributions. They are constructed by coupling a multivariate t-copula with arbitrarily specified univariate margins. Here and in the sequel we fix margins of meta t-distributions as univariate t-distributions not necessarily with the same degrees of freedom. The resulting multivariate distributions are rich enough and cover multivariate t-distributions. The usefulness of meta t-distributions with t-margins is illustrated in McNeil et al. (2005). We present our algorithm for bivariate meta t-distributions. However, it can be easily generalized to the multivariate case $d > 2$. In an extensive simulation study we demonstrate that the MBP algorithm is the most stable and efficient method for small samples than DIR and IFM methods.

The rolling window technique is computational intensive and uses data samples which are usually much smaller than an entire one. Therefore it requires a numerically stable estimation method for small sample sizes. In this situation the MBP algorithm for meta t-distributions is most suited. In contrast, the DIR method may often fail to find MLE's since the log likelihood functions are less peaked and smooth due to the small sample size. The two step IFM approach gives generally estimates which are less efficient than MLE's since marginal and copula parameter estimates are determined separately (see Joe, 2005).

The rest of the paper is organized as follows. In Section 2 we discuss a multivariate t-copula and a meta t-distribution. Section 3 presents the generic DIR, IFM and MBP algorithms for copula based models. Section 4 describes the MBP algorithm for meta t-distributions. Section 5 contains the simulation study illustrating the superiority of our algorithm over the DIR and IFM approaches. An application of the proposed MBP algorithm to rolling window analysis is given in Section 6, including pointwise bootstrapped confidence limits to assess the significance of time varying effects. Section 7 summarizes our findings and gives an outlook for further research.

Derivatives of working and error models for our MBP algorithm are deferred to the Appendix.

2. Copulas and meta t-distributions

Meta t-distributions are flexible extensions of multivariate t-distributions (see Kotz and Nadarajah, 2004). In their simplest form they have univariate t-margins with arbitrary degrees of freedom (df) parameter for each margin, but other margins are feasible as well. A comprehensive study of meta t-distributions and more general meta-elliptical distributions can be found in Fang et al. (2002) or Embrechts et al. (2003). Their construction is based on the concept of copulas (see Sklar, 1959), which will be briefly be presented in the following.

Copulas are multivariate distributions with uniformly distributed marginals on [0,1]. Classical textbook references on copulas are Joe (1997) and Nelsen (2006). Consider a d-variate random variable $\boldsymbol{X} := (X_1, \ldots, X_d)$ with joint cumulative distribution function (cdf) $F(\cdot, ..., \cdot)$ and marginal cdf's $F_i(\cdot)$, $i = 1, ..., d$, respectively. Then Sklar's Theorem states that there exists a d-variate copula cdf $C(\cdot, ..., \cdot)$ on $[0,1]^d$ with

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

for all $\boldsymbol{x} \in \bar{\mathbb{R}}^d$. If $F_1(\cdot), \ldots, F_d(\cdot)$ are continuous, then $C(\cdot, ..., \cdot)$ is unique. Conversely, if $C(\cdot, ..., \cdot)$ is a d-dimensional copula and $F_1(\cdot), \ldots, F_d(\cdot)$ are distribution functions, then

$$C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)). \tag{2}$$

Here $F_i^{-1}(u_i)$ denotes the inverse of the cdf $F_i(\cdot)$ for $i = 1, \ldots, d$. Using (2) we can write the corresponding copula density as

$$\begin{aligned} c(u_1, \ldots, u_d) &= \frac{\partial^2 C(u_1, \ldots, u_d)}{\partial u_1 \cdots \partial u_d} \\ &= \frac{\partial F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\partial u_1 \cdots \partial u_d} = \frac{f(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}. \end{aligned} \tag{3}$$

The d-variate meta t-distribution is constructed by linking a d-variate t-copula with d arbitrary marginal cdf's. In this paper we focus on meta t-distributions with univariate t-distributed margins. Let t_ν denote the univariate t-distribution with df parameter $\nu > 1$ and $t_{\nu,R}$ the central d-variate t-distribution with df $\nu > 1$ and a $d \times d$ symmetric and positive-definite scatter matrix $R := (\rho_{ij})_{1 \leqslant i,j \leqslant d}$ with unit diagonal entries and $-1 < \rho_{ij} < 1$. The corresponding probability density functions (pdf's) are then given by

$$\begin{aligned} f(x\,; \nu) &:= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \\ f(\boldsymbol{x}\,; \nu, R) &:= \frac{\Gamma[\frac{\nu+d}{2}]}{\Gamma[\frac{\nu}{2}](\nu\pi)^{\frac{d}{2}}} |R|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu}\boldsymbol{x}^t R^{-1} \boldsymbol{x}\right]^{-\frac{\nu+d}{2}}, \end{aligned} \tag{4}$$

where $|\cdot|$ denotes the determinant of a matrix and $\Gamma(\cdot)$ is the Gamma function. Finally we use $F(x\,; \nu)$ and $F(\boldsymbol{x}\,; \nu, R)$ to denote the corresponding cdf's. From (3) it follows that the d-variate t-copula pdf $c(\cdot\,; \nu, R)$ is given by

$$c(\boldsymbol{u}\,; \nu, R) = \frac{f(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\,; \nu, R)}{\prod_{j=1}^d f(t_\nu^{-1}(u_j)\,; \nu)} = \frac{\Gamma(\frac{\nu+d}{2})\Gamma(\frac{\nu}{2})^{d-1}\left(1 + \frac{\boldsymbol{z}^t R^{-1}\boldsymbol{z}}{\nu}\right)^{-\frac{\nu+d}{2}}}{\Gamma(\frac{\nu+1}{2})^d |R|^{\frac{1}{2}} \prod_{j=1}^d (1 + \frac{z_j^2}{\nu})^{-\frac{\nu+1}{2}}}, \tag{5}$$

where $\boldsymbol{z} := (z_1, ..., z_d)^t$ with $z_j := t_\nu^{-1}(u_j)$ and t_ν^{-1} is the quantile function of univariate t-distribution. The d-variate t-copula distribution is abbreviated as $C_d(\nu, R)$. If $X_j \sim t_{\nu_j}$, $j =$

$1,...,d$, it follows that $\boldsymbol{U} := (U_1,\dots,U_d) \sim C_d(\nu,R)$, where $U_j := F(X_j\,;\nu_j)$. The reversal holds if $\boldsymbol{U} \sim C_d(\nu,R)$, then $\boldsymbol{Z} := (t_\nu^{-1}(U_1),\dots,t_\nu^{-1}(U_d))^t \sim t_{\nu,R}$.

Let $\boldsymbol{\Theta} := \{\nu_1,\dots,\nu_d,\nu,R\,|\,\nu_1,\dots,\nu_d,\,\nu \in [1,\infty), -1 < \rho_{ij} < 1, i = 1,...,d, j < i\}$ be the parameter space of the d-variate meta t-distribution with univariate t-margins, abbreviated as $MT_d(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The parameters of the marginal distributions are df's $\nu_1,\dots,\nu_d$ while ν and R are the parameters of the copula. Using (3) with (4) and (5) we obtain the pdf of a d-variate meta t-distributed random vector $\boldsymbol{X} := (X_1,\dots,X_d) \sim MT_d(\boldsymbol{\theta})$ as:

$$\begin{aligned} h(\boldsymbol{x}\,;\boldsymbol{\theta}) &= c(F(x_1\,;\nu_1),\dots,F(x_d\,;\nu_d)\,;\nu,R)\prod_{j=1}^{d} f(x_j\,;\nu_j) \\ &= \frac{\Gamma(\frac{\nu+d}{2})\Gamma(\frac{\nu}{2})^{d-1}\big(1+\frac{\boldsymbol{z}^t R^{-1}\boldsymbol{z}}{\nu}\big)^{-\frac{\nu+d}{2}}}{\Gamma(\frac{\nu+1}{2})^d |R|^{\frac{1}{2}}\prod_{j=1}^{d}(1+\frac{z_j^2}{\nu})^{-\frac{\nu+1}{2}}} \prod_{j=1}^{d} \frac{\Gamma(\frac{\nu_j+1}{2})}{\sqrt{\nu_j\pi}\,\Gamma(\frac{\nu_j}{2})}\left(1+\frac{x_j^2}{\nu_j}\right)^{-\frac{\nu_j+1}{2}}. \end{aligned} \tag{6}$$

Figure 1 illustrates contour plots of bivariate meta t pdf's with $\rho = 0.8$ and different df values for ν_1, ν_2 and ν.

For d-variate i.i.d. observations $\boldsymbol{x} := (\boldsymbol{x}_1,...,\boldsymbol{x}_n)^t \in \mathbb{R}^{n\times d}$ of size n with $\boldsymbol{x}_i := (x_{i1},...,x_{id})^t$ for $i = 1,\dots,n$ from $MT_d(\boldsymbol{\theta})$ the corresponding log likelihood function based on the observations $\boldsymbol{x}$ is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}\,;\boldsymbol{x}) &:= \sum_{i=1}^{n} \ln c(F(x_{i1}\,;\nu_1),\dots,F(x_{id}\,;\nu_d)\,;\nu,R) + \sum_{j=1}^{d}\sum_{i=1}^{n} \ln f(x_{ij}\,;\nu_j) \qquad (7) \\ &= n\ln\Gamma(\frac{\nu+d}{2}) + n(d-1)\ln(\frac{\nu}{2}) - n\ln\Gamma(\frac{\nu+1}{2}) - \frac{n}{2}\ln(|R|) \\ &\quad - \frac{\nu+d}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{z}_i^t R^{-1}\boldsymbol{z}_i}{\nu}\big) + \frac{\nu+1}{2}\sum_{i=1}^{n}\sum_{j=1}^{d}\ln\big(1+\frac{z_{ij}^2}{\nu}\big) \\ &\quad + \sum_{j=1}^{d}\left(n\ln\Gamma(\frac{\nu_j+1}{2}) - \frac{n}{2}\ln(\nu_j\pi) - n\ln\Gamma(\frac{\nu_j}{2}) - \frac{\nu_j+1}{2}\sum_{i=1}^{n}\ln\big(1+\frac{x_{ij}^2}{\nu_j}\big)\right). \end{aligned}$$

In addition for the parameter vectors $\boldsymbol{\theta}_1 := (\nu_1,\dots,\nu_d)^t$ and $\boldsymbol{\theta}_2 := (\nu,\boldsymbol{\rho})^t$ where $\boldsymbol{\rho} := (\rho_{ij} : i = 1,...,d,\ j < i)$ we define

$$\ell_c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\,;\boldsymbol{x}) := \sum_{i=1}^{n} \ln c(F(x_{i1}\,;\nu_1),\dots,F(x_{id}\,;\nu_d)\,;\nu,R), \tag{8}$$

$$\ell_m(\boldsymbol{\theta}_1\,;\boldsymbol{x}) := \sum_{j=1}^{d}\sum_{i=1}^{n} \ln f(x_{ij}\,;\nu_j), \tag{9}$$

so that (7) can be rewritten as $\ell(\boldsymbol{\theta}\,;\boldsymbol{x}) = \ell_c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\,;\boldsymbol{x}) + \ell_m(\boldsymbol{\theta}_1\,;\boldsymbol{x})$.

Lindskog et al. (2003) show that Kendall's τ transformation estimator for the linear correlation of a pair of random variables, i.e.

$$\hat{\rho} = \sin\big(\frac{\pi}{2}\hat{\tau}\big), \tag{10}$$

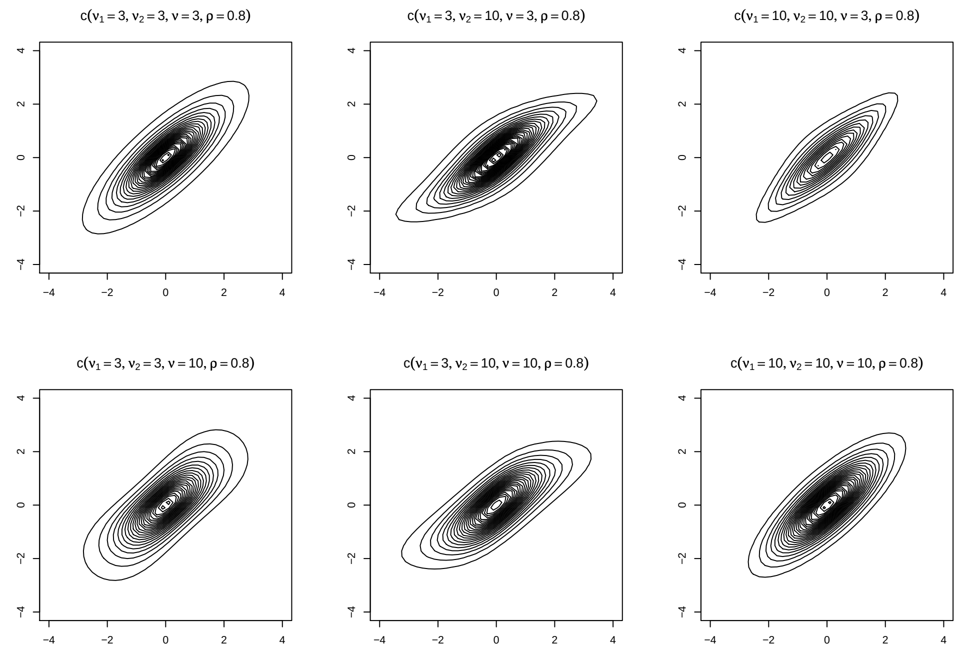
is more robust than the sample correlation estimator when estimating the bivariate correlation coefficient ρ. The association measure Kendall's τ for a bivariate random variable $\boldsymbol{X} = (X_1,X_2)$ is defined as

$$\tau := P((X_1-Y_1)(X_2-Y_2) > 0) - P((X_1-Y_1)(X_2-Y_2) < 0),$$

where $\boldsymbol{Y} = (Y_1,Y_2)$ are independent copies of $\boldsymbol{X}$. Kendall's τ is invariant under monotone transformations, thus it does not change for $\boldsymbol{X} = (X_1,X_2) \sim MT_2(\boldsymbol{\theta})$ (see Nelsen, 2006).

In many financial applications the tail dependence plays a crucial role when examining the im-

Figure 1: Contour plots of bivariate meta t pdf's with $\rho = 0.8$ and different degrees of freedom parameters ν_1, ν_2 and ν



pact of extremal events. According to Embrechts et al. (2003), the upper and lower tail dependence coefficients of symmetric copulas are equal. For the bivariate t-copula they showed that

$$\lambda(\nu, \rho) := 2t_{\nu+1}\left(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\right). \tag{11}$$

3. Likelihood based estimation methods

Without loss of generality we assume that we have a log likelihood $\ell(\boldsymbol{\theta}, \boldsymbol{x})$, given by (7), for parameter vector $\boldsymbol{\theta} \in \mathbb{R}^p$ and data $\boldsymbol{x} \in \mathbb{R}^{n \times d}$ available.

3.1. Direct maximization method (DIR)

A standard method to estimate the MLE is direct maximization (DIR). In the DIR method we conduct a p-dimensional maximization of $\ell(\boldsymbol{\theta}\,; \boldsymbol{x})$ in (7) with respect to $\boldsymbol{\theta}$ where $p = \frac{d(d+1)}{2} + 1$. We use numerical methods such as L-BFGS-B or BFGS to solve constrained or non-constrained nonlinear optimization problems by approximating the Hessian matrix. (see Lu et al., 1994). The algorithms are often not stable and computationally difficult for high dimensional parameter estimations or small sample size.

3.2. Inference for margins

The IFM method requires a partition of the log likelihood in (7) into two components - a marginal and a copula component, i.e. $\ell(\boldsymbol{\theta}\,; \boldsymbol{x}) = \ell_m(\boldsymbol{\theta}_1\,; \boldsymbol{x}) + \ell_c(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,; \boldsymbol{x})$. Here $\ell_m(\cdot\,;\cdot)$ and $\ell_c(\cdot\,;\cdot)$ are given in (8) and (9). The parameter vector $\boldsymbol{\theta}$ consists of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, i.e. $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t)^t$, where $\boldsymbol{\theta}_1 \in \mathbb{R}^d$ represents the parameters of the marginal distributions and $\boldsymbol{\theta}_2 \in \mathbb{R}^{(d^2-d)/2+1}$ includes the parameters of the copula. The IFM method proceeds as follows:

STEP 1. *Find* $\hat{\boldsymbol{\theta}}_1 = \arg\max$ *of* $\ell_m(\boldsymbol{\theta}_1\,; \boldsymbol{x}) = \mathbf{0}$.
STEP 2. *Find* $\hat{\boldsymbol{\theta}}_2 = \arg\max$ *of* $\ell_c(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\,; \boldsymbol{x}) = \mathbf{0}$.

Joe (2005) proved that under some regularity conditions the IFM estimators are consistent and asymptotically normal. However, the major disadvantage of this approach is the loss of efficiency in estimation since STEP 1 does not consider the dependence between the marginal distributions.

Since the high dimensional parameter estimation of a copula function in STEP 2 is often computationally challenging we calculate the IFM estimators following ideas in Lindskog (2000) and Demarta and McNeil (2005). The method uses the relationship (10) to estimate the correlation coefficients of bivariate margins. To obtain a positive definite matrix we can adapt the eigenvalue method suggested by Rousseeuw and Molenberghs (1993). Finally we estimate the remaining parameter ν by holding the correlation matrix fix. According to Mashal et al. (2003), this partial ML estimation approach gives very similar estimates to the full maximum likelihood procedure and has been widely used in practice.

3.3. MBP approach for maximum likelihood estimation

The multi-step fixed point algorithm MBP proposed by Song et al. (2005) requires a decomposition of the likelihood into two parts - a working model and an error model. In this paper we adapt a modified version of the maximization by parts (MBP) decomposition technique presented in Song et al. (2005). The modified version of MBP decomposition requires that the error model contains all parameters of the log likelihood while the working model only includes parts of them. In particular, consider a general likelihood function $\ell(\boldsymbol{\theta}\,;\boldsymbol{x})$ where $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector to be estimated and data $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^t$. We partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t)^t$ for $\boldsymbol{\theta}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\theta}_1 \in \mathbb{R}^{p_2}$. Here $\boldsymbol{\theta}_1$ denotes the parameter vector of the working model with log likelihood $\ell_w(\boldsymbol{\theta}_1\,;\boldsymbol{x})$ while the log likelihood of the error model is written as $\ell_e(\boldsymbol{\theta}\,;\boldsymbol{x})$. The decomposition of the full log likelihood is then $\ell(\boldsymbol{\theta}\,;\boldsymbol{x}) = \ell_w(\boldsymbol{\theta}_1\,;\boldsymbol{x}) + \ell_e(\boldsymbol{\theta}\,;\boldsymbol{x})$ and its corresponding score functions are therefore given by

$$\frac{\partial}{\partial \boldsymbol{\theta}}\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial}{\partial \boldsymbol{\theta}_1}\ell_w(\boldsymbol{\theta}_1\,;\boldsymbol{x}) + \dfrac{\partial}{\partial \boldsymbol{\theta}_1}\ell_e(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,;\boldsymbol{x}) \\ \dfrac{\partial}{\partial \boldsymbol{\theta}_2}\ell_e(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,;\boldsymbol{x}) \end{pmatrix} = \mathbf{0}.$$

The general MBP algorithm of Song et al. (2005) can now be formulated as follows:

STEP 0. *Find* $\boldsymbol{\theta}_1^0$*, which solves* $\dfrac{\partial}{\partial \boldsymbol{\theta}_1}\ell_w(\boldsymbol{\theta}_1\,;\boldsymbol{x}) = \mathbf{0}$.

Find $\boldsymbol{\theta}_2^0$*, which solves* $\dfrac{\partial}{\partial \boldsymbol{\theta}_2}\ell_e(\boldsymbol{\theta}_1^1, \boldsymbol{\theta}_2\,;\boldsymbol{x}) = \mathbf{0}$.

STEP k. *Find* $\boldsymbol{\theta}_1^k$*, which solves* $\dfrac{\partial}{\partial \boldsymbol{\theta}_1}\ell_w(\boldsymbol{\theta}_1\,;\boldsymbol{x}) = -\dfrac{\partial}{\partial \boldsymbol{\theta}_1}\ell_e(\boldsymbol{\theta}_1^{k-1}, \boldsymbol{\theta}_2^{k-1}\,;\boldsymbol{x})$.

k=1,2,... *Find* $\boldsymbol{\theta}_2^k$*, which solves* $\dfrac{\partial}{\partial \boldsymbol{\theta}_2}\ell_e(\boldsymbol{\theta}_1^{k-1}, \boldsymbol{\theta}_2\,;\boldsymbol{x}) = \mathbf{0}$.

For each STEP k with $k = 1, 2, ...,$ we use estimates from the previous step to update the estimates of the working and error models in the current step. Song et al. (2005) showed in Theorem 1, if $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$ is a consistent estimator of $\boldsymbol{\theta}$, then $\boldsymbol{\theta}^k = (\boldsymbol{\theta}_1^k, \boldsymbol{\theta}_2^k)$ for each k, $k = 1, 2, ...$ is consistent as well. Under regularity conditions, $\boldsymbol{\theta}^k$ converges to the MLE $\hat{\boldsymbol{\theta}}$ as $k \to \infty$ in probability. In Theorem 3 Song et al. (2005) showed the asymptotic normality of the estimator $\boldsymbol{\theta}^k$ arising from the MBP algorithm and derived an expression for the asymptotic variance of $\hat{\boldsymbol{\theta}}^k$.

4. MBP algorithm for meta t-distributions

For the meta t-distribution we start with the decomposition $\boldsymbol{\theta} := (\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t)^t$ where $\boldsymbol{\theta}_1 = (\nu_1, \ldots, \nu_d)^t \in \mathbb{R}^d$ are the parameters of the marginal distributions and $\boldsymbol{\theta}_2 = (\nu, \boldsymbol{\rho})^t \in \mathbb{R}^{(d^2-d)/2+1}$ with $\boldsymbol{\rho} := (\rho_{ij},\, i = 1, ..., d,\, j < i)$ are the parameters of the copula. As pointed out in Song et al. (2005), the convergence of the MBP algorithm strongly depends on the choice of the working and error models. A choice of a independence working model which only contains the univariate marginals would lead to a failure of the MBP algorithm, since it does not give no information about their dependence. In order to achieve the convergence we need to include additional information to the independence working model that accounts for some degree of correlation between the marginal distributions. Following Song et al. (2005), we decompose the log likelihood (7) of the meta t-distribution for data $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^t \in \mathbb{R}^{n\times d}$ and some known copula parameters $\boldsymbol{\theta}_{2,0} := (\nu_0, \boldsymbol{\rho}_0)^t$ as

$$\ell(\boldsymbol{\theta}\,; \boldsymbol{x}) = \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,; \boldsymbol{x}) = \ell_w(\boldsymbol{\theta}_1\,; \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}) + \ell_e(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,; \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}),$$

where

$$\ell_w(\boldsymbol{\theta}_1\,; \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}) := \ell_m(\boldsymbol{\theta}_1\,; \boldsymbol{x}) + \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}), \tag{12}$$

$$\ell_e(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,; \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}) := \ell_c(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,; \boldsymbol{x}) - \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}). \tag{13}$$

Here $\ell_c(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\,; \boldsymbol{x})$ and $\ell_m(\boldsymbol{\theta}_1\,; \boldsymbol{x})$ are as specified in (8) and (9). For the MBP algorithm specified in Section 3.3 we need to derive (12) and (13). We give the explicit form of these derivatives for $d = 2$ in the appendix.

In the algorithm we can set the known copula parameters $\boldsymbol{\theta}_{2,0} = (\nu_0, \boldsymbol{\rho}_0)^t$ in two ways - non-adaptive and adaptive. The non-adaptive method keeps $\boldsymbol{\theta}_{2,0}$ fixed for all MBP iterations, while the other one changes the value of $\boldsymbol{\theta}_{2,0}$ in each iteration, yielding two versions of the MBP algorithm. To determine the initial values in STEP 0 in the MBP algorithm we use the IFM method as described in Section 3.2.

Non-adaptive MBP (MBP1)

In the non-adaptive MBP method we hold $\boldsymbol{\theta}_{2,0}$ fixed during the entire MBP iterations.
STEP 0: Generate initial values $(\nu_1^0, \ldots, \nu_d^0, \nu^0, \boldsymbol{\rho}^0)$ using the IFM method
STEP k.1: Choose the fixed copula parameter vector $\boldsymbol{\theta}_{2,0} = (\nu^0, \boldsymbol{\rho}^0)^t$ and find $\boldsymbol{\theta}_1^k = (\nu_1^k, \ldots, \nu_d^k)^t$ which solves the following equation

$$\frac{\partial}{\partial \boldsymbol{\theta}_1}\ell_w(\boldsymbol{\theta}_1\,; \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}) + \frac{\partial}{\partial \boldsymbol{\theta}_1}\ell_e(\boldsymbol{\theta}_1^{k-1}, \boldsymbol{\theta}_2^{k-1}\,; \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}) = \mathbf{0}.$$

STEP k.2: Find the solution $\boldsymbol{\theta}_2^k = (\nu^k, \boldsymbol{\rho}^k)^t$ of the equation

$$\frac{\partial}{\partial \boldsymbol{\theta}_2}\ell_e(\boldsymbol{\theta}_1^{k-1}, \boldsymbol{\theta}_2\,; \boldsymbol{\theta}_{2,0}\,; \boldsymbol{x}) = \mathbf{0}.$$

Adaptive MBP (MBP2)

In the adaptive method we update the parameter $\boldsymbol{\theta}_{2,0} = (\nu^{k-1}, \boldsymbol{\rho}^{k-1})^t$ in each MBP steps with the estimated parameters from previous step.
STEP 0: Generate initial values $(\nu_1^0, \ldots, \nu_d^0, \nu^0, \boldsymbol{\rho}^0)$ using the IFM method
STEP k.1: Choose new copula parameter vector in each step k, i.e. $\boldsymbol{\theta}_{2,0} = \boldsymbol{\theta}_2^{k-1} = (\nu^{k-1}, \boldsymbol{\rho}^{k-1})^t$ and find $\boldsymbol{\theta}_1^k = (\nu_1^k, \ldots, \nu_2^k)^t$ which solves the following equation

$$\frac{\partial}{\partial \boldsymbol{\theta}_1}\ell_w(\boldsymbol{\theta}_1\,; \boldsymbol{\theta}_2^{k-1}\,; \boldsymbol{x}) - \frac{\partial}{\partial \boldsymbol{\theta}_1}\ell_m(\boldsymbol{\theta}_1^{k-1}\,; \boldsymbol{x}) = \mathbf{0}.$$

STEP k.2: Find the solution $\boldsymbol{\theta}_2^k = (\nu^k, \boldsymbol{\rho}^k)^t$ of the equation

$$\frac{\partial}{\partial \boldsymbol{\theta}_2}\ell_e(\boldsymbol{\theta}_1^{k-1}, \boldsymbol{\theta}_2\,; \boldsymbol{\theta}_2^{k-1}\,; \boldsymbol{x}) = \mathbf{0}.$$

We also investigated a mixed strategy for the choice of $\boldsymbol{\theta}_{2,0} = (\nu^{k-1}, \boldsymbol{\rho}^0)^t$ and $\boldsymbol{\theta}_{2,0} = (\nu^0, \boldsymbol{\rho}^{k-1})^t$. However, these algorithms fail to converge after few iteration steps.

In the implementation of the algorithm we apply the R function `optim` using Quasi-Newton with box constraints (L-BFGS-B) or Quasi-Newton without box constraints (BFGS). For the L-

BFGS-B method the parameter space for degrees of freedom is limited to [1,100]. If the `optim` function using L-BFGS-B fails to converge we apply the BFGS method which requires the Fisher's z-transformation for the parameters. The parameter space for degrees of freedom is then unconstrained. For more details we refer to Lu et al. (1994).

5. Small sample properties of MBP estimators

Here we demonstrate our findings in bivariate case. To evaluate the performance of the MBP estimators and its competitors DIR and IFM estimators we use robust estimates of bias and mean squared error (MSE). More precisely, let $\hat{\theta}_1, ..., \hat{\theta}_r$ be independent estimates of the true parameter θ_{tr}, then the median absolute deviation (mad) of estimator $\hat{\theta}$ is given by

$$mad(\hat{\theta}) := median(|\hat{\theta}_i - median(\hat{\theta}_1, ..., \hat{\theta}_r)|), \quad i = 1...r, \tag{14}$$

and the MSE of $\hat{\theta}$ is robustly estimated by

$$\widehat{rmse}(\hat{\theta}) := (median(\hat{\theta}_1, ..., \hat{\theta}_r) - \theta_{tr})^2 + \left(\frac{mad(\hat{\theta}_1, ..., \hat{\theta}_r)}{\Phi^{-1}(\frac{3}{4})}\right)^2 \tag{15}$$

(see Brown, 1982; Ripley and Venables, 2002).

We are interested in investigating the performance in small samples, since we want to apply our estimation method in a rolling window setup to capture time effects in financial data. Therefore we use a sample size of $n = 100$. For this small sample size it is more appropriate to use the robust performance measures introduced above than the classical estimates based on empirical means and variances. This allows to reduce the effects of outliers. We chose 12 parameter combinations for ν_1, ν_2, ν and ρ specified in column 2 to 5 in Table 1. For each scenario we generate an i.i.d. sample of size $n = 100$ from $MT_2(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\nu_1, \nu_2, \nu, \rho)^t$ is specified by the scenario. Finally we estimated $\boldsymbol{\theta}$ using MBP1, MBP2 as well as DIR and IFM methods.

Table 1 gives the estimated performance measures for each parameter component of $\boldsymbol{\theta}$ and for each scenario based on converged replications. From this we see that most of the times the lowest values for $\widehat{mad}(\hat{\theta})$ agrees with the lowest value of $\widehat{rmse}(\hat{\theta})$ for all components of $\boldsymbol{\theta}$, thus indicating no bias-variance trade-off between the different estimation methods. MBP1, MBP2 and DIR behave quite similar, while IFM tends to differ. Pattern differ from parameter component to parameter component. For the marginal parameter ν_1 and the copula component ρ IFM performs better than MBP1, MBP2 or DIR, while the opposite is true for parameter components ν_2 and ν.

To get a clearer overall picture we apply the average rank method implemented in R by the function `rank(,ties.method=average)`. This means that for each parameter we rank the results of the four estimation methods. Then we add up the ranks over all components and groups of scenarios involved for each estimation methods. These four sums corresponding to different estimation methods are finally ranked again. Table 2 gives these overall rank for different copula parameter combinations. From this we see that MBP1 outperforms IFM for a small copula df ($\nu = 3$) regardless of a smaller or larger value of the association parameter ρ. For large df ($\nu = 10$), MBP1 is still a good performer. Overall MBP1 is the best performer followed by IFM.

As we mentioned above the estimated *mad* and *rmse* values are only based on those replications out of 100, which converged. For scenarios with $\nu = 3$ ($\nu = 10$) the percentage of non-converged estimates ranges between 2% to 22% (24% to 42%) for the IFM method, while the DIR method converged always. For both MBP1 and MBP2 most scenarios had no convergence problems. At most 1% of non-converged estimates was observed for MBP1 and MBP2.

Our overall conclusion is that the non-adaptive maximization by parts methods is the best performing estimation method with respect to bias and MSE as well as with respect to lowest occurrence of convergence problems. The price for this good performance is a substantial increase in computing time. MBP requires often 10 to 50 times more than IFM and DIR.

Table 1: Small sample performance in terms of $\widehat{mad}(\hat{\theta})$ and $\widehat{rmse}(\hat{\theta})$ for each parameter component of $\boldsymbol{\theta}$ over 12 scenarios based on 100 data sets of size 100 from $MT_2(\boldsymbol{\theta})$

#	ν_1	ν_2	ν	ρ	Method	$\widehat{mad}$				$\widehat{rmse}$			
						$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\nu}$	$\hat{\rho}$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\nu}$	$\hat{\rho}$
01	3	3	3	0.3	MBP1	0.868	0.700	1.436	0.119	1.656	1.077	4.572	0.031
					MBP2	0.868	0.700	1.436	0.119	1.656	1.078	4.572	0.031
					DIR	0.868	0.700	1.437	0.119	1.657	1.078	4.574	0.031
					IFM	0.733	0.714	1.541	0.109	1.181	1.123	5.258	0.026
02	3	3	3	0.8	MBP1	0.805	0.694	1.653	0.037	1.426	1.068	6.011	0.003
					MBP2	0.806	0.694	1.653	0.037	1.428	1.069	6.009	0.003
					DIR	0.808	0.693	1.645	0.037	1.434	1.065	5.955	0.003
					IFM	0.800	0.701	1.749	0.041	1.414	1.094	6.792	0.004
03	10	10	10	0.3	MBP1	5.918	6.453	6.757	0.080	76.984	92.362	100.791	0.014
					MBP2	5.964	6.473	6.725	0.081	78.185	92.975	99.883	0.015
					DIR	6.029	6.525	6.899	0.079	79.904	94.514	105.040	0.014
					IFM	5.615	7.132	5.397	0.064	69.349	113.274	67.870	0.010
04	10	10	10	0.8	MBP1	6.888	7.163	8.297	0.036	104.746	118.531	151.356	0.003
					MBP2	6.888	7.413	8.263	0.036	104.732	126.678	150.120	0.003
					DIR	6.353	7.768	8.668	0.036	89.336	137.242	165.255	0.003
					IFM	6.928	6.960	7.541	0.026	105.939	107.582	125.048	0.002
05	3	3	10	0.3	MBP1	0.877	0.808	6.286	0.073	1.691	1.442	88.262	0.012
					MBP2	0.877	0.808	6.286	0.073	1.691	1.442	88.267	0.012
					DIR	0.880	0.806	6.342	0.073	1.703	1.434	89.490	0.012
					IFM	0.878	0.856	4.917	0.062	1.696	1.614	58.533	0.009
06	3	3	10	0.8	MBP1	0.867	0.761	7.451	0.035	1.680	1.344	122.682	0.003
					MBP2	0.865	0.762	7.454	0.033	1.675	1.347	122.785	0.002
					DIR	0.890	0.795	7.452	0.035	1.747	1.423	122.211	0.003
					IFM	0.752	0.776	6.658	0.026	1.249	1.323	97.673	0.002
07	3	10	3	0.3	MBP1	0.832	4.885	1.318	0.118	1.524	52.459	3.846	0.031
					MBP2	0.832	4.872	1.322	0.119	1.523	52.180	3.870	0.031
					DIR	0.828	4.910	1.300	0.119	1.508	52.996	3.741	0.031
					IFM	0.733	5.548	1.375	0.109	1.181	68.136	4.170	0.026
08	3	10	3	0.8	MBP1	0.746	5.819	1.655	0.041	1.224	75.192	6.051	0.004
					MBP2	0.762	5.823	1.646	0.041	1.279	75.273	5.989	0.004
					DIR	0.752	4.607	1.513	0.036	1.247	46.720	5.036	0.003
					IFM	0.800	5.676	1.607	0.041	1.414	72.347	5.718	0.004
09	3	10	10	0.3	MBP1	0.863	7.473	5.889	0.075	1.639	125.897	78.255	0.013
					MBP2	0.861	7.466	5.889	0.077	1.630	125.639	78.242	0.013
					DIR	0.886	7.454	5.985	0.078	1.726	125.235	80.759	0.013
					IFM	0.866	6.724	5.044	0.061	1.648	100.138	60.416	0.009
10	3	10	10	0.8	MBP1	0.933	6.811	7.010	0.035	1.923	106.115	109.422	0.003
					MBP2	0.900	6.829	7.128	0.036	1.789	105.692	113.109	0.003
					DIR	0.858	6.975	6.450	0.034	1.623	109.896	93.406	0.002
					IFM	0.827	5.578	6.099	0.027	1.504	68.416	84.020	0.002
11	10	10	3	0.3	MBP1	6.158	4.849	1.352	0.120	83.937	51.684	4.044	0.032
					MBP2	6.160	4.847	1.353	0.120	84.000	51.634	4.050	0.032
					DIR	6.320	4.927	1.362	0.120	88.558	53.358	4.114	0.032
					IFM	5.862	5.016	1.409	0.110	75.543	55.333	4.387	0.027
12	10	10	3	0.8	MBP1	7.140	5.076	1.621	0.040	113.923	56.917	5.778	0.004
					MBP2	7.107	5.053	1.625	0.040	112.860	56.314	5.806	0.003
					DIR	7.020	5.438	1.619	0.039	110.269	65.631	5.763	0.003
					IFM	6.441	5.484	1.664	0.041	91.587	67.329	6.146	0.004

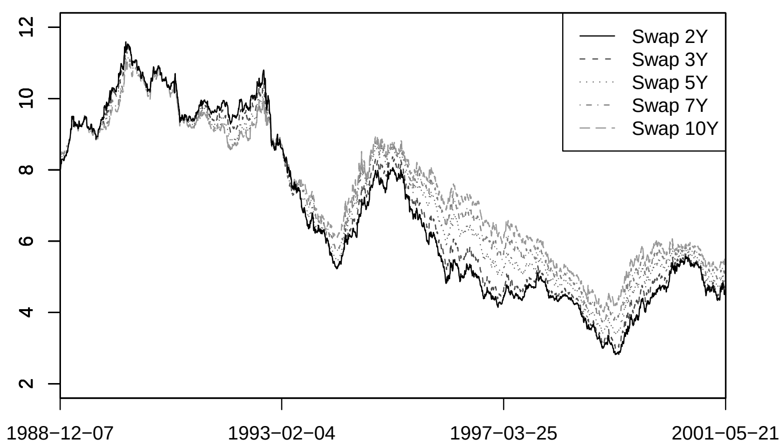
Table 2: Average ranks of lowest $\widehat{rmse}(\hat{\theta})$ over different copula parameter combinations for each estimation method

true copula parameters	averaged over scenario	MBP1	MBP2	DIR	IFM
$\nu = 3, \quad \rho = 0.3$	1, 7, 11	1.0	2.0	4.0	3.0
$\nu = 3, \quad \rho = 0.8$	2, 8, 12	2.0	3.0	1.0	4.0
$\nu = 10, \quad \rho = 0.3$	3, 5, 9	2.5	2.5	4.0	1.0
$\nu = 10, \quad \rho = 0.8$	4, 6, 10	2.0	3.0	4.0	1.0
	total $\sum$	7.5	10.5	13.0	9.0
	total average rank	1	3	4	2

6. Application to financial swap rates

We apply our estimation methods to annually compounded zero-coupon Euro swap rates from 12/07/1988 to 05/21/2001 with maturities of 2, 3, 5, 7 and 10 years, respectively (see Figure 2). A *zero-coupon swap* is a special kind of interest rate swap. In an interest rate swap one party agrees to pay a fixed rate (swap rate) and at the same time receives a floating rate which is usually linked to a reference rate such as the London Interbank Offered Rate (LIBOR). In the case of a zero-coupon swap, instead of paying the fixed rate periodically, one makes one large payment only at the end of the maturity.

Figure 2: swap rates for different maturities over time



6.1. Parameter estimation and hypothesis testing

Let X_{it} denote the zero-coupon Euro swap rate at time t for maturity i with $i = 2, 3, 5, 7$ and 10 and $t = 1, ..., 3150$. Czado et al. (2009) showed that an ARMA(1,1)-GARCH(1,1) model for $\{X_{it}, t = 1, ..., 3150\}$ is sufficient for each i to capture the time dependence. After forming standardized residuals $\hat{Z}_{it}$ we can therefore consider $\{\hat{Z}_{it}, t = 1, ..., 3150\}$ as an i.i.d. sample for each fixed maturity i. We now model pairs of $\{(\hat{Z}_{it}, \hat{Z}_{jt}), t = 1, ..., 3150\}$ as observations from $MT_2(\boldsymbol{\theta})$. For brevity we present results for the pairs with $i = 2$ years and $j = 3, 5, 7$ and 10 years. The remaining pairs are investigated in Zhang (2008).

In this section we regard four bivariate data sets of short term maturity versus medium and long term maturities: 2 years and 3 years ($\boldsymbol{Z}_2$,$\boldsymbol{Z}_3$), 2 years and 5 years ($\boldsymbol{Z}_2$,$\boldsymbol{Z}_5$), 2 years and 7 years ($\boldsymbol{Z}_2$,$\boldsymbol{Z}_7$), and 2 years and 10 years ($\boldsymbol{Z}_2$,$\boldsymbol{Z}_{10}$). We compare the bivariate t- and meta t-models regarding the data sets. The bivariate t-model is a special case of the bivariate meta t-model where we assume that the degrees of freedom of the marginal distributions and the joint distribution are equal. For the meta t-model we estimate the parameter vectors using MBP1, MBP2, DIR and IFM methods while for the bivariate t-model the parameter estimation only proceeds with MBP1 and DIR methods. The results are given in Table 3.

The estimation results suggest that most data pairs have marginal distributions with parameters ν_1 and ν_2 estimated between $[6.2, 6.9]$, while their joint parameter ν is considerably smaller than 6.2 and the correlation estimates are high. It also demonstrates that MBP1 and MBP2 yield similar estimates as the DIR method. Furthermore, the similar size of the log likelihood value for the MBP1, MBP2, DIR affirm that the MBP1 and MBP2 methods determine the maximum likelihood estimators as the DIR method while the log likelihood value results from the IFM method is not at its maximum.

In order to justify the choice of the meta t-model we use the likelihood ratio test. We construct a test with the null hypothesis $H_0 : \nu_1 = \nu_2 = \nu$ versus the alternative H_1 : not H_0. For $\hat{\boldsymbol{\theta}}^{H_0} := (\hat{\nu}_t, \hat{\nu}_t, \hat{\nu}_t, \hat{\rho}_t)$ and $\hat{\boldsymbol{\theta}}^{H_1} := (\hat{\nu}_1, \hat{\nu}_2, \hat{\nu}, \hat{\rho})$ we consider the test statistic

$$LR(\hat{\boldsymbol{\theta}}^{H_0}, \hat{\boldsymbol{\theta}}^{H_1}; \boldsymbol{x}) := -2(\ell(\hat{\boldsymbol{\theta}}^{H_0}; \boldsymbol{x}) - \ell(\hat{\boldsymbol{\theta}}^{H_1}; \boldsymbol{x})).$$

An α-level asymptotic test rejects $H_0 \iff LR(\hat{\boldsymbol{\theta}}^{H_0}, \hat{\boldsymbol{\theta}}^{H_1}; \boldsymbol{x}) \geqslant \chi_{1-\alpha,2}$, where $\ell(\hat{\boldsymbol{\theta}}; \boldsymbol{x})$ is the log likelihood function of the bivariate meta t-distribution defined in (7) and $\chi_{1-\alpha,2}$ is the $(1-\alpha)100\%$ quantile of the Chi-square distribution. Applying the estimates using MBP1 method for the likelihood ratio test we decide based on the p-values in the Table 3 that the null hypothesis is rejected for bivariate data sets $(\boldsymbol{Z}_2, \boldsymbol{Z}_3)$, $(\boldsymbol{Z}_2, \boldsymbol{Z}_5)$, $(\boldsymbol{Z}_2, \boldsymbol{Z}_7)$ and accepted for the data set $(\boldsymbol{Z}_2, \boldsymbol{Z}_{10})$. This means for the most pairs a bivariate t-model is not sufficient.

Table 3: Parameter estimates of bivariate t- and meta t-models and the likelihood ratio test for $H_0 : \nu_1 = \nu_2 = \nu$ versus H_1 : not H_0.

bivariate meta t-model $MT_2(\nu_1, \nu_2, \nu, \rho)$					
	method	$(\boldsymbol{Z}_2, \boldsymbol{Z}_3)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_5)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_7)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_{10})$
$\hat{\nu}_1$	MBP1	6.366	6.884	6.731	6.680
	MBP2	6.366	6.881	6.728	6.680
	DIR	6.362	6.877	6.725	6.677
	IFM	10.993	10.993	10.993	10.993
$\hat{\nu}_2$	MBP1	6.241	6.846	6.488	6.228
	MBP2	6.241	6.843	6.485	6.229
	DIR	6.234	6.839	6.482	6.225
	IFM	11.278	12.465	12.462	11.828
$\hat{\nu}$	MBP1	3.167	4.063	4.465	5.093
	MBP2	3.166	4.054	4.455	5.102
	DIR	3.165	4.048	4.450	5.090
	IFM	2.436	3.466	4.293	5.176
$\hat{\rho}$	MBP1	0.958	0.915	0.865	0.841
	MBP2	0.958	0.915	0.865	0.841
	DIR	0.958	0.915	0.865	0.841
	IFM	0.938	0.880	0.816	0.785
log likelihood	MBP1	−5374.532	−6438.849	−7121.341	−7352.185
	MBP2	−5374.532	−6438.848	−7121.341	−7352.185
	DIR	−5374.532	−6438.848	−7121.341	−7352.185
	IFM	−5426.637	−6483.500	−7163.900	−7396.200
bivariate t-model $MT_2(\nu_t, \nu_t, \nu_t, \rho)$					
	method	$(\boldsymbol{Z}_2, \boldsymbol{Z}_3)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_5)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_7)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_{10})$
$\hat{\nu}_t$	MBP1	5.399	6.171	6.111	6.210
	DIR	5.369	6.144	6.085	6.188
$\hat{\rho}_t$	MBP1	0.962	0.919	0.869	0.844
	DIR	0.962	0.919	0.870	0.844
log likelihood	MBP1	−5392.700	−6448.136	−7126.642	−7354.132
	DIR	−5392.696	−6448.134	−7126.640	−7354.131
likelihood ratio test: $H_0 : \nu_1 = \nu_2 = \nu$ versus H_1 : not H_0					
		$(\boldsymbol{Z}_2, \boldsymbol{Z}_3)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_5)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_7)$	$(\boldsymbol{Z}_2, \boldsymbol{Z}_{10})$
statistic		36.337	18.575	10.601	3.895
p-value		> 0.000	> 0.000	0.005	0.143

6.2. Rolling windows estimation

In the previous section we apply the bivariate t- and bivariate meta t-models to the swap data to estimate their parameters. We have obtained a single value for each of the parameters of the bivariate data series. However the economic environment often changes rapidly and it may not be reasonable to assume constant values of the model parameter. For this reason it would be interesting to see how the parameters of the model vary with respect to the time. To catch the changes of the models we apply a rolling window analysis to the observed time series.

We place a sequence of rolling windows on the time axis and estimate the parameters for each of these windows. Two sequential windows overlap by 50 data points. The data size of a window amounts up to 250 data points which implies the number of trading days in a year. The time span goes from December 07, 1988 to Jan 21st, 2001, which includes 3150 data points. We investigate the variability of parameter estimates over 59 rolling windows.

The sequence of estimated values shows the evolution of the parameters in our models over time. The 59 rolling windows estimates for each parameter $\nu_{1,h}$, $\nu_{2,h}$, ν_h and ρ_h, $h = 1, ..., 59$, are illustrated in the 1st, 2nd, 3rd and the 4th row of Figures 3, respectively. We can observe that the parameter estimates vary considerably over the rolling windows. They occasionally fluctuate

Figure 3: Rolling window estimates $\hat{\nu}_{1,h}$, $\hat{\nu}_{2,h}$, $\hat{\nu}_h$ and $\hat{\rho}_h$, $h = 1, ..., 59$, using MBP1, MBP2, Direct and IFM methods for meta t specification

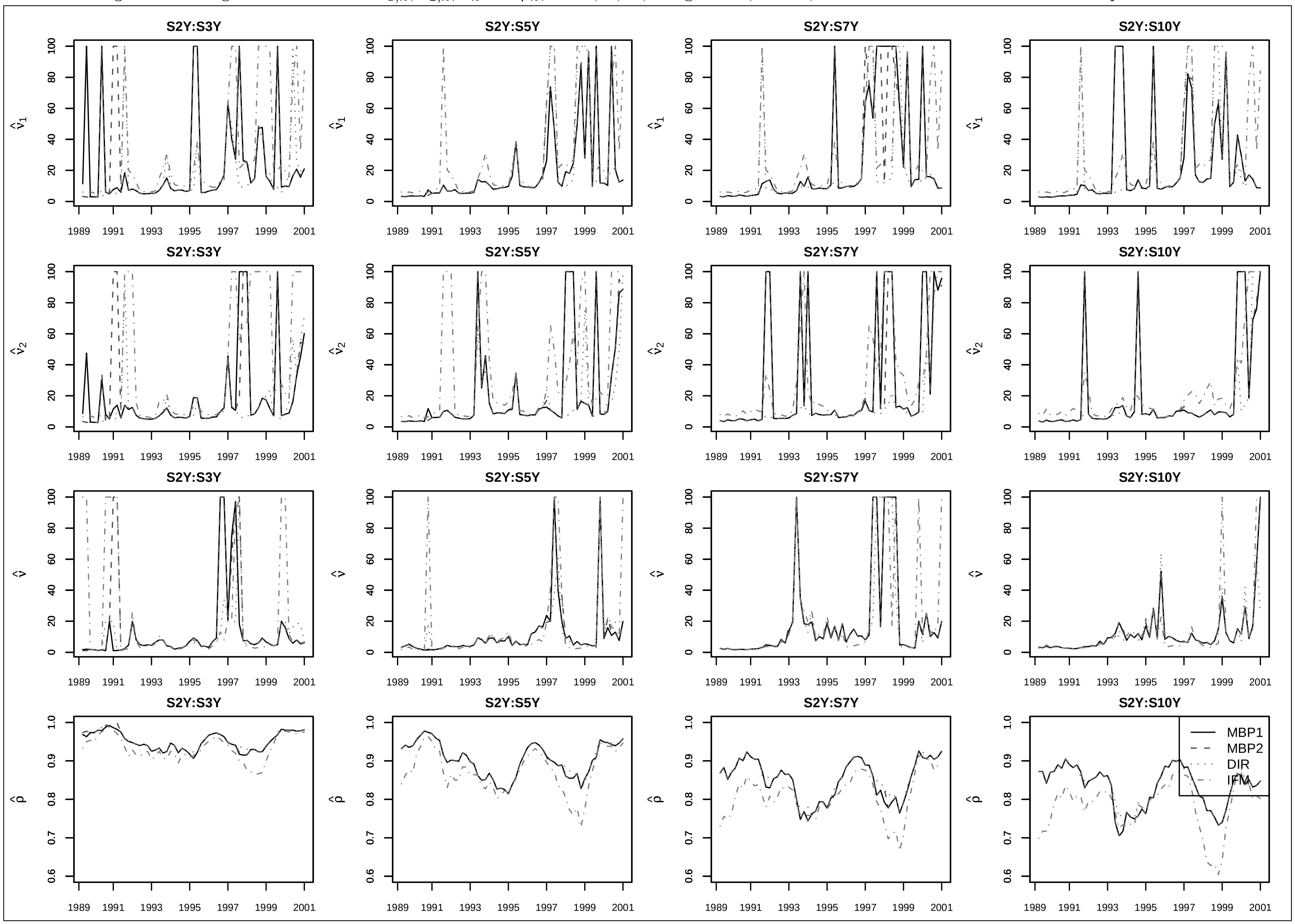
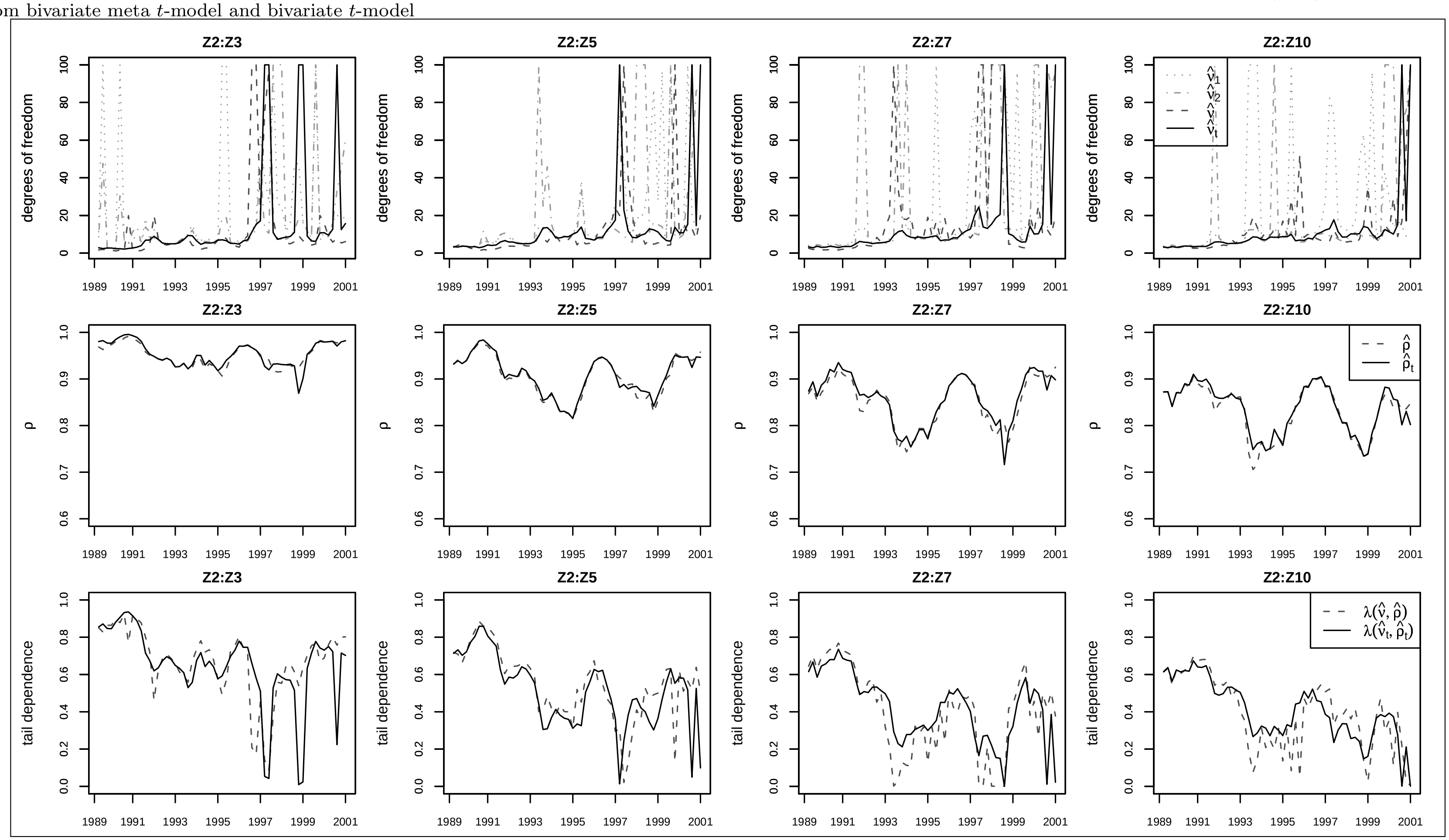


Figure 4: 1st Row: comparison of estimated rolling window df parameter estimates $\hat{\nu}_{1,h}$, $\hat{\nu}_{2,h}$, $\hat{\nu}_h$ estimated using bivariate meta t-model based on MBP1 method and estimated rolling parameter $\hat{\nu}_{t,h}$ using bivariate t-model based on MBP1 method; 2nd Row: comparison of estimated rolling correlation $\hat{\rho}_h$ using bivariate meta t-model based on MBP1 method and rolling estimates $\hat{\rho}_{t,h}$ using bivariate t-model based on MBP1 method; 3rd Row: estimated tail dependence coefficients $\lambda(\hat{\nu}_h, \hat{\rho}_h)$ and $\lambda(\hat{\nu}_{t,h}, \hat{\rho}_{t,h})$ using estimates from bivariate meta t-model and bivariate t-model



to 100 which indicate that data at those specific time intervals is almost normal distributed. Compared to the rolling estimated marginal degrees of freedom the rolling estimated joint degrees of freedom $\hat{\nu}_h$, $h = 1, ..., 59$, show much less extreme fluctuations over time. For $\hat{\rho}_h$ we notice that MBP1, MBP2 and DIR give similar values while the IFM estimates differ considerably.

Furthermore, the 1st row of Figure 4 illustrates the comparison of estimated parameters $\hat{\nu}_{1,h}$, $\hat{\nu}_{2,h}$, $\hat{\nu}_h$, $h = 1, ..., 59$, in bivariate meta t-model and the estimated degrees of freedom $\hat{\nu}_{t,h}$ in bivariate t-model. It shows that the rolling degrees of freedom $\hat{\nu}_t$ estimated with bivariate t-model and the estimated values of parameters ν_1, ν_2 and ν from the bivariate meta t-model differ from each other by a considerable amount over the entire time interval. This also means that any significant deviation from the solid line (rolling estimates of $\hat{\nu}_{t,h}$ from bivariate t-model, $h = 1, ..., 59$) affirms that a meta t-model is needed. Furthermore, we can observe that the dotted lines in the graphs for degrees of freedom parameters are often fluctuate to 100 while the solid black line, which represents the rolling estimates of MBP algorithm are more stable. In the 2nd row of Figure 4 we can see the rolling correlations $\hat{\rho}_h$ and $\hat{\rho}_{t,h}$, estimated by both bivariate meta t-model and bivariate t-model, respectively. The differences of the two estimated series are marginal but can be followed particularly in the periods between 1993-1995 and between 1997-1999. The 3rd row of Figure 4 illustrates the estimated tail dependence defined in (11) of the data using the estimated copula parameters of both models. The plots disclose a significant difference of the estimated tail dependence by applying the parameters of the bivariate meta t- and bivariate t-models.

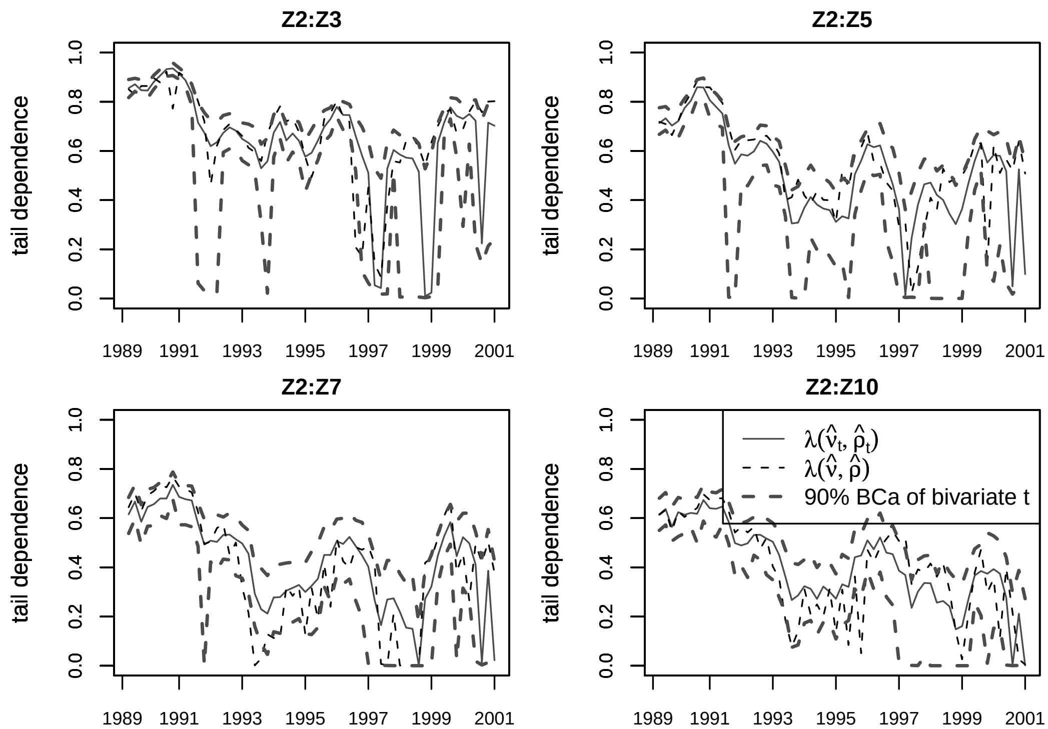
In order to assess overfitting of bivariate meta t-distributions we computed 90% bias corrected and accelerated (BCa) pointwise confidence intervals for the tail dependence coefficient of bivariate t-distributions. The BCa confidence interval is a modified bootstrap confidence interval which has good theoretical coverage probability and computational feasibility (see DiCiccio and Efron, 1996). For illustration we consider only four pairs as above. Thus, at each of the 59 rolling windows a bivariate t-distributions is fitted for 250 bootstrap samples each. Using (11) we obtain 250 bootstrap estimates for the tail dependence coefficients, on which the corresponding BCa interval is based. Figure 5 presents pointwise 90% BCa confidence intervals for $\lambda(\hat{\nu}_{t,h}, \hat{\rho}_{t,h})$ as well as rolling window estimates $\lambda(\hat{\nu}_{t,h}, \hat{\rho}_{t,h})$ and $\lambda(\hat{\nu}_h, \hat{\rho}_h)$, $h = 1, ..., 59$. The 90% BCa confidence intervals (bold dashed lines) mostly cover the rolling estimates of tail dependence for the bivariate meta t-model (dashed lines) but not always. In particular, for the pair $(\boldsymbol{Z}_2, \boldsymbol{Z}_3)$, more than a quarter of the rolling estimates of tail dependence $\lambda(\hat{\nu}_h, \hat{\rho}_h)$ based on meta t-model are not in the BCa confidence interval. For the pair $(\boldsymbol{Z}_2, \boldsymbol{Z}_5)$, $\lambda(\hat{\nu}_h, \hat{\rho}_h)$ are outside the BCa interval for $h = 9, 12, 23, 30, 43, 48$. For $(\boldsymbol{Z}_2, \boldsymbol{Z}_7)$, most of estimates between the 19th and 29th windows and some of the estimates in the end of the considered time horizon are not inside of the BCa interval. And for $h = 21, 31, 33, 40, 55$, the estimates $\lambda(\hat{\nu}_h, \hat{\rho}_h)$ of the pair $(\boldsymbol{Z}_2, \boldsymbol{Z}_{10})$ are outside the interval. These results indicate that both models have a significant difference in estimating the copula parameters and the bivariate t-distribution is not flexible enough to capture bivariate tail dependencies detected by a meta t-distribution. Finally, our bootstrap study reveals similar funding to one for stock indices in Aussenegg and Cech (2008) in sense that there are time periods when tail dependency cannot be captured well by the bivariate t-distribution.

7. Summary and outlook

Multivariate financial data exhibits heavy-tailedness, i.e. marginal extreme events occur often jointly. Therefore, a multivariate t-distribution and its copula are preferred over Gaussian ones for modeling in economics, finance and risk management. Meta t-distributions generalize multivariate t-distributions and allow for arbitrary marginal distributions. Estimation of meta t-distributions is numerical unstable and complex. Therefore there are different stepwise methods proposed for finding MLE's in literature (see Bouye et al. 2000; Embrechts et al. 2003; Demarta and McNeil 2005).

In this paper we adapt the general MBP estimation algorithm developed by Song et al. (2005) to the case of the multivariate meta t-distribution constructed by a t-copula. We implement and investigate the behavior of these MBP algorithms in two dimensions. This algorithm overcomes

Figure 5: 90% Bootstrap BCa interval for $\lambda(\hat{\nu}_{t,h}, \hat{\rho}_{t,h})$ based on bivariate t-model



many practical difficulties one may encounter using direct maximization or the IFM methods. The MBP algorithm calculates the maximum likelihoods estimators (MLE) in multi-step optimizations. It requires a decomposition of the log likelihood function into a working model and an error model and updates the parameters through an iterative approach that solve the sore equations of both models iteratively. The convergence of the MBP algorithm to a fixed point under regularity conditions is guaranteed by the existence of the asymptotic contraction mappings which requires the Hessian matrix of the working model to be more informative about the true value than the one of the error model. To ensure the convergence, we developed two strategies in the MBP algorithm $-$ a non-adaptive MBP method and an adaptive MBP method. In the non-adaptive MBP approach we keep some of the parameters in the working model fixed while those parameters are iteratively updated in the adaptive approach.

In our simulation study we found out that the non-adaptive MBP algorithm is a computationally cost intensive approach. However, it can handle small meta t-distributed samples very well compared to the adaptive MBP, direct maximization and IFM method, especially if the degrees of freedom of the copula parameter are low. In particular the non-adaptive MBP method results in minimal robustly estimated bias and MSE. Additionally it has a very low failure rate associated with algorithm convergence.

The rolling windows technique is a simple way to assess variability of model parameters in time. It relies on a much smaller part of an entire data set. Therefore, our non-adaptive MBP algorithm for meta t-distributions is especially tailored to this situation. Thus we applied both MBP algorithms along with the DIR and the IFM algorithms to Euro swap rates. For this we consider models based on bivariate t and meta t-distributions. Due to the small sample properties of the non-adaptive MBP algorithm, we apply the algorithm to rolling window analysis on the swap rates data. Pointwise bootstrap BCa confidence intervals for tail dependency coefficients indicate that a meta t-distribution is more suitable to the swap rates data than a simple bivariate t-model with a single degree of freedom.

Construction of flexible multivariate copulas ($d > 2$) is a fast growing and active research area.

According to recent empirical studies by Fischer et al. (2009) and Berg and Aas (2007), pair copula constructions (PCC's) are, at the moment, most successful in fitting multivariate financial data. In these constructions a multivariate copula is represented as a product of $d(d-1)/2$ bivariate unconditional and conditional copulas (for more details see Bedford and Cook, 2001, 2002 and Aas et al. 2009). A ML estimation of PCC's for high dimensional data is very time consuming and may be numerically unstable. Since parameters of unconditional copulas are involved in arguments of conditional pair-copulas, log likelihood functions of PCC's can be naturally split into working and error models and therefore we envision here an applicability if the MBP algorithm. This paper provides the ground work to proceed in this attractive future research area.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance Mathematics and Economics 44 (2), 182–198.

Aussenegg, W., Cech, C., 2008. Simple time-varying copula estimation. Tech. Rep. 53, University of Applied Science of bfi Vienna.

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

Bedford, T., Cook, R. M., 2002. Vines - a new graphical model for dependent random variables. Annals of Statistics 30 (4), 1031–1068.

Berg, D., Aas, K., 2007. Models for construction of multivariate dependence. Discussion Paper SAMBA/23/07.

Bouye, Durrleman, Nikeghbali, Riboulet, Roncalli, 2000. Copula for finance: A reading guide and some applications.

Brown, M., 1982. Robust line estimation with errors in both variables. Journal of the American Statistical Association 77 (377), 71–79.

Cambanis, S., Huang, S., Simons, G., 1981. On the theory of elliptically contoured distributions. Journal of Multivariate Analysis 11, 368–385.

Czado, C., Min, A., Baumann, T., Dakovic, R., 2009. Pair-copula constructions for modeling exchange rate dependence. Tech. rep., Technische Universtät München, http://www-m4.ma.tum.de/Papers/Czado/cc-pubs.html.

Dall'Aglio, G., Kotz, S., Salinetti, G. E., 1991. Advances in Probability Distributions with Given Marginals Beyond the Copulas. Vol. 67 of Mathematics and Its Applications. Springer.

Demarta, S., McNeil, A., 2005. The t copula and related copulas. International Statistical Review 73 (1), 111–129.

DiCiccio, J., Efron, B., 1996. Bootstrap confidence intervals. Statistical Science 11 (3), 189–212.

Embrechts, P., Lindskog, F., McNeil, A. J., 2003. Modelling dependence with copulas and applications to risk management. In: Handbook of heavy tailed distributions in finance. Elsevier/North-Holland, Amsterdam, pp. 329–384.

Embrechts, P., McNeil, A., Straumann, D., 1999. Correlation: Pitfalls and alternatives. RISK 12 (5), 69–71.

Fan, Y., Pastorello, S., Renault, E., 2007. Maximization py parts in extemum estimation. (unpublished manuscript).

Fang, H.-B., Fang, K.-T., Kotz, S., 2002. The meta-elliptical distributions with given marginals. Journal of Multivariate Analysis 82 (1), 1–16.

Fang, K. T., Kotz, S., Ng, K. W., 1990. Symmetric Multivariate and Related Distributions. Vol. 36 of Monographs on Statistics and Applied Probability. Chapman and Hall Ltd., London.

Fischer, M., Köck, C., Schlüter, S., Weigert, F., 2009. An empirical analysis of multivariate copula models. Quantitative Finance, To be published, DOI: 10.1080/14697680802595650.

Frees, E. W., Valdez, E. A., 1998. Understanding relationships using copulas. North American Actuarial Journal 2 (1), 1–25.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Vol. 73 of Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis 94 (2), 401–419.

Joe, H., Xu, J. J., 1996. The estimation method of inference functions for margins for multivariate models. Tech. Rep. 166, University of British Columbia.

Johnson, N. L., Kotz, S., 1972. Distributions in Statistics: Continuous Multivariate Distributions. John Wiley & Sons Inc., New York, Wiley Series in Probability and Mathematical Statistics.

Kotz, S., Nadarajah, S., 2004. Multivariate t Distributions and Their Applications. Cambridage University Press.

Krzysztofowicz, R., Kelly, K. S., 1997. A bivariate meta-gaussian density for use in hydrology. Stochastic Hydrology and Hydraulics 11 (1), 17–31.

Lindskog, F., 2000. Modelling dependence with copulas and applications to risk management. Master's thesis, Eidgösische Technische Hochschule Zürich.

Lindskog, F., A., M., Schmock, U., 2003. Kendall's tau for elliptical distributions. In: Credit Risk - Measurement, Evaluation and Management. Physica-Verlag Heidelberg, pp. 176–223.

Liu, Y., Luger, R., 2009. Efficient estimation of copula-GARCH models. Computational Statistics and Data Analysis 53 (6), 2284–2297.

Lu, P., Nocedal, J., Zhu, C., Richard Byrd, H., Byrd, R. H., 1994. A limited-memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16, 1190–1208.

Mashal, R., Zeevi, A., Naldi, M., 2003. Beyond correlation: Extreme events and multi-name credit derivatives. Tech. rep., Unpublished, Columbia University.

McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press.

Nelsen, R. B., 2006. An Introduction to Copulas (Springer Series in Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Ripley, B. D., Venables, W. N., 2002. Modern Applied Statistics with S, 4th Edition. Springer-Verlag.

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

Shih, J. H., Louis, T. A., 1995. Inferences on the association parameter in copula models for bivariate survival data. Biometrics 51 (4), 1384–1399.

Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231.

Song, P. X.-K., Fan, Y., Kalbfleisch, J. D., 2005. Maximization by parts in likelihood inference. Journal of the American Statistical Association 100 (472), 1145–1167, with comments and a rejoinder by the authors.

Song, P. X.-K., Zhang, P., Qu, A., 2007. Maximum likelihood inference in robust linear mixed-effects models using multivariate t-distributions. Statistica Sinica 17 (3), 929–943.

Zhang, R., 2008. Maximization by parts for bivariate t-and meta t-distributions. Master's thesis, Technische Universität München, `http://www-m4.ma.tum.de/Diplarb/da_txt.html`.

Zivot, E., Wang, J., 2003. Modeling Financial Time Series With S-Plus. Springer.

Appendix: Derivatives of the modified working and error models in bivariate case

To simplify the expressions of the derivatives of the modified working and error models, we begin with some definitions: Let $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^t \in \mathbb{R}^{n\times 2}$ with $\boldsymbol{x}_i := (x_{i1}, x_{i2})^t$, $i = 1, ..., n$. In the bivariate case, we have four parameters, i.e. ν_1, ν_2, ν, ρ. For clarity, we spare the notation $\boldsymbol{\theta}$ for the parameters. Now we define $\boldsymbol{z}_i := (z_{i1}, z_{i2})^t = (z_{i1}(\nu_1, \nu\,; x_{i1}), z_{i2}(\nu_2, \nu\,; x_{i2}))^t := \left(t_\nu^{-1}(F(x_{i1}, \nu_1)), t_\nu^{-1}(F(x_{i2}, \nu_2))\right)^t$,

$$\begin{aligned} A(\nu_1, \nu_2, \nu, \rho\,; \boldsymbol{x}_i) &:= \boldsymbol{z}_i^t R^{-1} \boldsymbol{z}_i = \frac{1}{1-\rho^2}\left(z_{i1}^2 + z_{i2}^2 - 2z_{i1}z_{i2}\rho\right), \\ B(\nu_j, \nu\,; x_{ij}) &:= z_{ij}^2, \quad j = 1, 2. \end{aligned} \tag{.1}$$

Their derivatives with respect to ν_1, ν_2, and ν are calculated as follows:

$$\begin{aligned} \frac{\partial A(\nu_1, \nu_2, \nu, \rho\,; \boldsymbol{x}_i)}{\partial \nu_j} &= \frac{1}{1-\rho^2}\left(2z_{ij}\frac{\partial z_{ij}}{\partial \nu_j} - 2\rho z_{ik}\frac{\partial z_{ij}}{\partial \nu_j}\right), \\ \frac{\partial B(\nu_j, \nu\,; x_{ij})}{\partial \nu_j} &= 2z_{ij}\frac{\partial z_{ij}}{\partial \nu_j}, \end{aligned}$$

where $k \neq j$, $k, j = 1, 2$, and $\dfrac{\partial z_{ij}}{\partial \nu_j} := \dfrac{\partial t_\nu^{-1}(F(x_{ij}\,; \nu_j))}{\partial \nu_j}$, $j = 1, 2$ Furthermore, the derivatives of functions A and B with respect to ν and ρ are given by:

$$\begin{aligned} \frac{\partial A(\nu_1, \nu_2, \nu, \rho\,; \boldsymbol{x}_i)}{\partial \nu} &= \frac{1}{1-\rho^2}\left(2\sum_{j=1}^{2} z_{ij}\frac{\partial z_{ij}}{\partial \nu} - 2\rho \sum_{j\neq k; j,k=1}^{2} z_{ik}\frac{\partial z_{ij}}{\partial \nu}\right), \\ \frac{\partial B(\nu_j, \nu\,; x_{ij})}{\partial \nu} &= 2z_{ij}\frac{\partial z_{ij}}{\partial \nu}, \end{aligned}$$

where $\dfrac{\partial z_{ij}}{\partial \nu} := \dfrac{\partial t_\nu^{-1}(u_j)}{\partial \nu}$ with $u_j := F(x_{ij}\,; \nu_j)$, $j = 1, 2$,

$$\frac{\partial}{\partial \rho} A(\nu_1, \nu_2, \nu, \rho\,; \boldsymbol{x}_i) = \frac{-2z_{i1}z_{i2}(1+\rho^2) + 2\rho(z_{i1}^2 + z_{i2}^2)}{(1-\rho^2)^2}.$$

The marginal model and its derivatives

For $a \in \mathbb{R}$ let $\ln\Gamma(a)$ and $\psi(a)$ be the digamma and trigamma functions of the gamma function $\Gamma(a)$. With the density function of univariate t-Distribution defined in (4) we have the marginal model and its derivatives with respect to ν_j, $j=1,2$, as follows:

$$\ell_m(\nu_1,\nu_2\,;\boldsymbol{x}) := \sum_{j=1}^{2}\sum_{i=1}^{n}\ln f(x_{ij}\,;\nu_j) = \sum_{i=1}^{n}\ln\left(\frac{\Gamma(\frac{\nu_j+1}{2})}{\sqrt{\nu_j\pi}\Gamma(\frac{\nu_j}{2})}\left(1+\frac{x^2}{\nu_j}\right)^{-\frac{\nu_j+1}{2}}\right)$$

$$= \sum_{j=1}^{2} n\ln\Gamma(\frac{\nu_j+1}{2}) - \frac{n}{2}\ln(\nu_j\pi) - n\ln\Gamma(\frac{\nu_j}{2}) - \frac{\nu_j+1}{2}\sum_{i=1}^{n}\ln\big(1+\frac{x_{ij}^2}{\nu_j}\big),$$

$$\frac{\partial}{\partial\nu_j}\ell_m(\nu_1,\nu_2\,;\boldsymbol{x}) = \frac{n}{2}\psi(\frac{\nu_j+1}{2}) - \frac{n}{2\nu_j} - \frac{n}{2}\psi(\frac{\nu_j}{2}) - \frac{1}{2}\sum_{i=1}^{n}\ln(1+\frac{x_{ij}^2}{\nu_j}) + \frac{\nu_j+1}{2\nu_j}\sum_{i=1}^{n}\frac{x_{ij}^2}{\nu_j+x_{ij}^2}.$$

The error model and its derivatives
With the density function of the bivariate t-Copula defined in (3) the error model is given by:

$$\ell_e(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}) := \sum_{i=1}^{n}\ln c(F(x_{i1}\,;\nu_1),F(x_{i2}\,;\nu_2)\,;\nu,\rho)$$

$$= \sum_{i=1}^{n}\ln\left(\frac{\frac{\nu}{2}\cdot\Gamma(\frac{\nu}{2})^2\big(1+\frac{z_1^2+z_2^2-2z_1z_2\rho}{\nu(1-\rho^2)}\big)^{-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu+1}{2})^2\sqrt{1-\rho^2}(1+\frac{z_1^2}{\nu})^{-\frac{\nu+1}{2}}(1+\frac{z_2^2}{\nu})^{-\frac{\nu+1}{2}}}\right)$$

$$= 2n\ln\Gamma(\frac{\nu}{2}) + n\ln(\frac{\nu}{2}) - 2n\ln\Gamma(\frac{\nu+1}{2}) - \frac{n}{2}\ln(1-\rho^2)$$

$$-\frac{\nu+2}{2}\sum_{i=1}^{n}\ln\big(1+\frac{A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)}{\nu}\big) + \frac{\nu+1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}\ln\big(1+\frac{B(\nu_j,\nu,x_{ij})}{\nu}\big),$$

with $A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)$ and $B(\nu_j,\nu,x_{ij})$, $j=1,2$, defined in (.1). The first order derivatives of the error model with respect to the parameters ν_1, ν_2, ν and ρ are calculated as follows:

$$\frac{\partial}{\partial\nu_j}\ell_e(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}) = -\frac{\nu+2}{2}\sum_{i=1}^{n}\frac{1}{1+\dfrac{A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)}{\nu}}\cdot\frac{\partial}{\partial\nu_j}A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)\frac{1}{\nu}$$

$$+\frac{\nu+1}{2}\sum_{i=1}^{n}\frac{1}{1+\dfrac{B(\nu_1,\nu\,;x_{ij})}{\nu}}\cdot\frac{\partial}{\partial\nu_j}B(\nu_1,\nu\,;x_{ij})\frac{1}{\nu},$$

$$\frac{\partial}{\partial\nu}\ell_e(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}) = \frac{n}{\nu} + n\psi(\frac{\nu}{2}) - n\psi(\frac{\nu+1}{2})$$

$$-\left[\frac{1}{2}(\sum_{i=1}^{n}\ln(1+\frac{A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)}{\nu}) + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\nu\frac{\partial A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)}{\partial\nu} - A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)}{\nu+A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)})\right]$$

$$+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}(\ln(1+\frac{B(\nu_j,\nu\,;x_{ij})}{\nu})) + \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\sum_{j=1}^{2}\frac{\frac{\partial}{\partial\nu}B(\nu_j,\nu\,;x_{ij})\cdot\nu - B(\nu_j,\nu\,;x_{ij})}{\nu+B(\nu_j,\nu\,;x_{ij})},$$

and

$$\frac{\partial}{\partial\rho}\ell_e(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}) = \frac{n\rho}{1-\rho^2} - \frac{\nu+2}{2}\sum_{i=1}^{n}\frac{\frac{\partial}{\partial\rho}A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)}{\nu+A(\nu_1,\nu_2,\nu,\rho\,;\boldsymbol{x}_i)}.$$