Technische Universität München

Zentrum Mathematik

Maximization by Parts for Bivariate t- and Meta t-Distributions

Diplomarbeit

von

Ran Zhang

Themensteller/in: Prof. Dr. Claudia Czado
Betreuer/in: Prof. Dr. Claudia Czado
Dr. Aleksey Min
Abgabetermin: 01. Dezember 2008

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

London, den 21. November 2008

Acknowledgments

First, I would like to express my sincere gratitude to Prof. Dr. Claudia Czado for making this diploma thesis possible. I am very thankful for the frequent meeting with many discussions and hints which gave me right directions for my research as well as the the numerous advices in academic writing. Furthermore, I want to address many thanks to Dr. Aleksey Min for his tremendous support whenever I had difficulties with the theory and implementations of the algorithms. Finally, I would like to thank Mr. Lanzendörfer for his continuous helps on the topic.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

This thesis presents and numerically implements the maximization by parts method (MbP) from Song et al. (2005) and applies it to bivariate t and bivariate meta t-distributions in order to analyze the dependence structure in the embedded bivariate t-copula model. Furthermore, other estimation methods such as direct maximization, inference for margins from Joe (1996), semi-parametric method using Kendall's τ and canonical maximum likelihood method for parameter estimation are presented, implemented and compared with the MbP method. The aim of this thesis is to demonstrate the benefits as well as the disadvantages of the MbP method in parameter estimation. Finally, we show an application using the MbP method involving swap rate data.

Chapter 2

Background Material

2.1 Introduction to Copulas

2.1.1 Definition of a copula

The reference for this section is the book Nelsen (1999) which provides a comprehensive introduction to copulas. The following definitions and Sklar's theorem apply to the multivariate case. In the interest of clarity, we first introduce some notations.

- Let $\mathbb{R}$ denote the ordinary real line $(-\infty, \infty)$, $\bar{\mathbb{R}}$ denote the extended real line $[-\infty, \infty]$ and $\bar{\mathbb{R}}^n$ denote the extended n-space $\bar{\mathbb{R}} \times \bar{\mathbb{R}} \times \cdots \times \bar{\mathbb{R}}$, where n is a positive integer.
- We will use vector notation for points in $\bar{\mathbb{R}}^n$, e.g. $\boldsymbol{a} = (a_1, ...a_n)^t$.
- Let $[\boldsymbol{a}, \boldsymbol{b}]$ denote the *n-dimensional set* $B = [a_1, b_1] \times \cdots \times [a_n, b_n]$, the Cartesian product of n closed intervals. The *vertices* of an n-dimensional set B are the points $\boldsymbol{c} = (c_1, ...c_n)^t$ where each c_k is equal to either a_k or b_k.
- The unit n-cube I^n is the product $I \times \cdots \times I$, where $I = [0, 1]$.
- An *n-place real function* H is a function whose domain, Dom$H$, is a subset of $\bar{\mathbb{R}}^n$ and whose range, Ran$H$, is a subset of $\mathbb{R}$.

Definition 2.1 *Let $S_1, \ldots, S_n$ be nonempty subset of $\bar{\mathbb{R}}$, and let H be an n-place real function such that Dom$H$ $= S_1 \times S_n$. Let $B = [\boldsymbol{a}, \boldsymbol{b}]$ be an n-box. The H volume of B is defined by*

$$\begin{aligned} V_H(B) &:= \sum \operatorname{sgn}(\boldsymbol{d}) H(\boldsymbol{d}) \\ &= \sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots i_n} H(d_{1i_1}, d_{2i_2}, \ldots, d_{ni_n}) \geqslant 0, \end{aligned}$$

Example 2.2 *Let H be a 2-place real function with domain $\bar{\mathbb{R}}^2$. The H-volumne of a rectangle $B = [a_1, b_1] \times [a_2, b_2]$ with all of whose vertices in Dom$H$ is given by*

$$V_H(B) = H(b_1, b_2) - H(b_1, a_2) - H(a_1, b_2) + H(a_1, a_2)$$

The definition of an n-dimensional copula in Nelsen (1999) is given by:

Definition 2.3 *An n-dimensional copula is a function C from the unit n-cube $I^n = [0,1]^n$ to $I = [0,1]$, denote $C : I^n \to I$, with the following properties:*

(i) $C(\boldsymbol{u}) = 0$ *for all* $\boldsymbol{u}$ *in* $[0,1]^n$ *for which at least one coordinate of* $\boldsymbol{u}$ *equals zero;*

(ii) $C(\boldsymbol{u}) = u_k$ *if all coordinates of u are 1 except the kth one;*

(iii) *C is n-increasing in the sense that for all $\boldsymbol{a} = (a_1, ..., a_n)^t$ and $\boldsymbol{b} = (b_1, ..., b_n)^t$ in $[0,1]^n$ with $a_i \leqslant b_i$ the volume assigned by C to the n-box $[\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times ... \times [a_n, b_n]$ is nonnegative $(V_C([\boldsymbol{a}, \boldsymbol{b}]) \geqslant 0)$.*

This definition shows that an n-dimensional copula is an n-dimensional distribution function restricted to $[0,1]^n$ with standard uniformly distributed marginals. The most relevant result in the copula framework is introduce by Sklar (1959).

Theorem 2.4 (Sklar's Theorem) *Let F be a d-dimensional cumulative distribution function (c.d.f.) with univariate margins F_1,...,F_d. Let R_i be the range of F_i, for $i = 1, 2, ..., d$. Then there exists an d-dimensional copula C such that for all $\boldsymbol{x} \in \bar{\mathbb{R}}^d$,*

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{2.1}$$

If $F_1, F_2, ..., F_d$ are all continuous, then C is unique; otherwise, C is uniquely determined on $RanF_1 \times \cdot \times RanF_d$. Conversely, if C is an d-dimensional copula and $F_1, ..., F_d$ are distribution functions, then the function F defined by (2.1) is an d-dimensional distribution function with univariate margins $F_1, ..., F_d$.

Proof: See Nelsen (1999) page 41.

Therefore, the joint distribution can be described as a function of copula and the univariate marginal distributions. Since the copula does not depend on univariate marginal distributions, we can say that it captures completely all the information about the dependence between the variables. For this reason, a copula is often interpreted as the *dependence structure*. Certainly, it makes sense when all the univariate marginals are continuous and the copula is unique; in the discrete case there will be more than one way to describe the dependence structure.
As pointed out in many literatures, Sklar's theorem has an important corollary. Before continuing with the corollary, we need to define the quantile function, probability-integral and quantile transformations since they will be repeatedly utilized in the thesis. The definition of quantile functions is given as follows:

Definition 2.5 (Quantile Function) *Consider the random variables X, X_1,...,X_d with distribution functions F, F_1,...,F_d respectively.*

(i) The quantile function of F is defined for all $u \in (0,1)$ by the generalized inverse of F:

$$F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geqslant u\}. \tag{2.2}$$

(ii) The probability-integral transformation is the mapping $T : \mathbb{R}^d \to [0,1]^d$ *with*

$$(x_1, x_2, ..., x_d) \mapsto (F_1(x_1), ..., F_d(x_d)). \tag{2.3}$$

(iii) The quantile transformation is the operation $T^{-1} : [0,1]^d \to \mathbb{R}^d$ *with* $(u_1, u_2, ..., u_d) \mapsto (F_1^{-1}(u_1), ..., F_d^{-1}(u_d))$.

If U is an uniform random variable on $(0,1)$, then $F^{-1}(U)$ has distribution function F. On the other hand, the distribution function F of a random variable X is continuous if and only if $F(X)$ is uniformly distributed on $(0,1)$.

Corollary 2.6 *Let F and C be, respectively, an d-dimensional distribution function with continuous univariate margins $F_1, ..., F_d$ and an d-dimensional copula function. Then, for any $\boldsymbol{u} \in [0,1]^d$, we have*

$$C(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)) \tag{2.4}$$

where $F_j^{-1}(u_j)$ denotes the inverse of the cumulative distribution function $F_j(\cdot)$.

By applying Sklar's theorem in (2.1) and by exploiting the relation between the distribution and density function in (2.4), the multivariate copula density $c(F_1(x_1), ..., F_d(x_d))$ can be derived associated with a copula function $C(F_1(x_1), ..., F_d(x_d))$ as the following. Let $f(\cdot)$ be the density function of a d-dimensional c.d.f. F,

$$\begin{aligned} f(x_1, ..., x_d) &= \frac{\partial F(x_1, ..., x_d)}{\partial x_1 ... \partial x_d} \\ &= \frac{\partial^d C(F_1(x_1), ..., F_d(x_d))}{\partial x_1 ... \partial x_d} \\ &= c(F_1(x_1), ..., F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d) \end{aligned}$$

Where we define the copula's density as

$$\begin{aligned} c(u_1, ..., u_d) &= \frac{\partial^d C(u_1, ..., u_d)}{\partial u_1 ... \partial u_d} \\ &= \frac{\partial F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))}{\partial u_1 ... \partial u_d} \\ &= f(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)) \cdot \frac{1}{f_1(F_1^{-1}(u_1))} \cdots \frac{1}{f_d(F_d^{-1}(u_d))} \end{aligned} \tag{2.5}$$

As we will see later that knowledge of the associated copula density will be particularly useful in order to calibrate its parameters to real market data.

2.1.2 Multivariate and Bivariate t-Copulas

To define the multivariate and the bivariate t-copulas, we need first to define the probability density functions of the multivariate, bivariate and univariate t-distributions.

Definition 2.7 (Univariate t-Distribution) *The random variable $X \in \mathbb{R}$ has the univariate Student's t distribution with degrees of freedom ν, if its probability density function (p.d.f.) is*

$$f(x;\nu) := \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{2.6}$$

It has mean 0 (for $\nu > 1$) and variance $\nu/(\nu-2)$ (for $\nu > 2$). In the thesis, we denote $F(x;\nu)$ as the univariate t cumulative distribution function (c.d.f.)

$$F(x;\nu) := \int_{-\infty}^{x} f(t;\nu)dt \tag{2.7}$$

and t_ν^{-1} as the quantile function of $F(x;\nu)$, which is given by

$$t_\nu^{-1}(u) := \inf\{x \in \mathbb{R} : F(x;\nu) \geqslant u\}. \tag{2.8}$$

According to Demarta and McNeil (2004) the multivariate Student's t joint density function is given as follows.

Definition 2.8 (Multivariate t-Distribution) $\boldsymbol{X} = (X_1, ..., X_d)^t$ *has multivariate (or d-variate) Student's t-distribution with $\nu > 1$ degrees of freedom (d.f.), mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and a $d \times d$ symmetric and positive-definite scatter matrix R, denoted by $t_d(\nu, \boldsymbol{\mu}, R)$, if its probability density function is*

$$f(\boldsymbol{x};\nu,\boldsymbol{\mu},R) := \frac{\Gamma[\frac{\nu+d}{2}]}{\Gamma[\frac{\nu}{2}](\nu\pi)^{\frac{d}{2}}}|R|^{-\frac{1}{2}}\left[1+\frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^t R^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-\frac{\nu+d}{2}} \tag{2.9}$$

$|R|$ represents the absolute value of the determinant of the matrix R.

If the mean vector $\boldsymbol{\mu} = \mathbf{0}$, we say that the Student's t-distribution is central, otherwise non-central. In this thesis, we only consider the central t-distribution.

Remark: Based on Prop 2.7 in the previous section, the copula remains invariant under strictly increasing transformations of the components of random vector $\boldsymbol{X}$, therefore the copulas of distributions $t_d(\nu, \boldsymbol{\mu}, R)$ and $t_d(\nu, \mathbf{0}, R)$ are identical. We can also transform any scatter matrix $\widetilde{R}$ to a positive definite matrix with unit diagonal elements and the copula does not change. In the thesis we consider only a scatter matrix R with *unit diagonal entries*. In addition, if $\nu > 2$ then the covariance matrix of the Student's t-distribution is given by

$$Cov(\boldsymbol{X}) = \frac{\nu}{\nu-2}R$$

otherwise i.e. $1 < \nu \leqslant 2$ it does not exist. Thus if $\nu > 2$ the scatter matrix R is equal to the correlation matrix.

Definition 2.9 (Bivariate t-Distribution) *Let $\boldsymbol{x} = (x_1, x_2)^t$. The bivariate t-density function $f(x_1, x_2; \nu, \rho)$ with parameters ν and ρ is given by*

$$\begin{aligned} f(x_1, x_2; \nu, \rho) &= \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)} |R|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu}\boldsymbol{x}^t R^{-1}\boldsymbol{x}\right]^{-\frac{\nu+2}{2}} \\ &= \frac{1}{2\pi}|R|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu}\boldsymbol{x}^t R^{-1}\boldsymbol{x}\right]^{-\frac{\nu+2}{2}} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \left[1 + \frac{1}{\nu(1-\rho^2)}(x_1^2 + x_2^2 - 2x_1x_2\rho)\right]^{-\frac{\nu+2}{2}} \end{aligned} \tag{2.10}$$

where we used the identities

$$\begin{aligned} \Gamma\left(\frac{\nu+2}{2}\right) &= \Gamma\left(\frac{\nu}{2}+1\right) = \frac{\nu}{2}\Gamma\left(\frac{\nu}{2}\right) \\ \textit{and} \quad \boldsymbol{x}^t R^{-1}\boldsymbol{x} &= \frac{1}{1-\rho^2}(x_1^2 + x_2^2 - 2x_1x_2\rho), \end{aligned}$$

with $R := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$

By applying the Sklar's theorem, the t-copula is defined as follows:

Definition 2.10 (Multivariate t-Copula) *Let $F(\cdot; \nu, R)$ be the cumulative distribution function (c.d.f.) of the standardized multivariate Student's t distribution with correlation matrix R and ν degrees of freedom. Then the multivariate Student's t-copula is defined as follows:*

$$\begin{aligned} C(u_1, ..., u_d; \nu, R) &= P[F(X_1; \nu) \le u_1, ..., F(X_d; \nu) \le u_d] \\ &= F(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d); \nu, R) \\ &= \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_d)} f(\boldsymbol{x}; \nu, R) d\boldsymbol{x}, \quad \boldsymbol{u} \in (0,1)^d \end{aligned}$$

where $t_\nu^{-1}(\cdot)$ denotes the quantile function of a standard univariate $F(\cdot; \nu)$ distribution and $f(\cdot; \nu, R)$ the joint density function of $F(\cdot; \nu, R)$.

The associated Student's t-copula density is obtained by applying the equations in (2.5):

$$\begin{aligned} c(\boldsymbol{u}; \nu, R) &= \frac{f(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d); \nu, R)}{\prod_{i=1}^d f(t_\nu^{-1}(u_i); \nu)} \\ &= \frac{\Gamma(\frac{\nu+d}{2})\Gamma(\frac{\nu}{2})^{d-1}\left(1 + \frac{\boldsymbol{z}^t R^{-1}\boldsymbol{z}}{\nu}\right)^{-\frac{\nu+d}{2}}}{\Gamma(\frac{\nu+1}{2})^d |R|^{\frac{1}{2}} \prod_{j=1}^d (1 + \frac{z_j^2}{\nu})^{-\frac{\nu+1}{2}}} \end{aligned} \tag{2.11}$$

with $\boldsymbol{z} := (z_1, ..., z_d)^t := (t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d))^t \in \mathbb{R}^d$.

Definition 2.11 (Bivariate t-Copula) *The density of a bivariate t copula is given by*

$$\begin{aligned} c(u_1, u_2; \nu, R) &= \frac{f(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2); \nu, R)}{f(t_\nu^{-1}(u_1); \nu) f(t_\nu^{-1}(u_2); \nu)} \\ &= \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})\big(1+\frac{\boldsymbol{z}^t R^{-1}\boldsymbol{z}}{\nu}\big)^{-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu+1}{2})^2 |R|^{\frac{1}{2}} (1+\frac{z_1^2}{\nu})^{-\frac{\nu+1}{2}} (1+\frac{z_2^2}{\nu})^{-\frac{\nu+1}{2}}} \\ &= \frac{\frac{\nu}{2}\cdot\Gamma(\frac{\nu}{2})^2\big(1+\frac{z_1^2+z_2^2-2z_1z_2\rho}{\nu(1-\rho^2)}\big)^{-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu+1}{2})^2\sqrt{1-\rho^2}(1+\frac{z_1^2}{\nu})^{-\frac{\nu+1}{2}}(1+\frac{z_2^2}{\nu})^{-\frac{\nu+1}{2}}} \end{aligned} \tag{2.12}$$

$(u_1, u_2)^t \in [0,1]^2$ and $\boldsymbol{z} := (z_1, z_2)^t := (t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))^t$ with quantile function t_ν^{-1} of univariate t distribution as defined in (2.8)*, $R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ and $|R| = 1-\rho^2$.*

In general, consider a bivariate t random vectors $\boldsymbol{X} := (X_1, X_2)^t \in \mathbb{R}^2$ and a bivariate t-copula $\boldsymbol{U} := (U_1, U_2)^t \in [0,1]^2$ with $U_1 = F(X_1, \nu_2)$ and $U_2 = F(X_2, \nu_2)$ where $F(;\nu_j)$ the univariate t-distribution function with degrees of freedom ν_j, $j = 1, 2$. It is important to note that the correlation coefficient of the bivariate t-copula is not identical to the correlation of a t-distribution, i.e.

$$\rho(U_1, U_2) \neq \rho(X_1, X_2)$$

In addition, according to Embrechts (2001), the upper and lower tail dependence coefficient for bivariate t-copula with parameters ν and ρ is defined as

$$\lambda(\nu, \rho) := 2t_{\nu+1}\left(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\right) \tag{2.13}$$

2.1.3 Invariance Property

One attractive feature of the copula representation of dependence is that the copula is *invariant* under increasing and continuous transformations of the marginals.

Proposition 2.12 *If $(X_1, ..., X_n)^t$ has copula C and $T_1, ..., T_n$ are increasing continuous functions, then $(T_1(X_1), ..., T_n(X_n))^t$ also has copula C.*

Proof. Let the random vector $(U_1, ..., U_n)^t$ have distribution function C with continuous marginals F_{X_i}, in particular $U_i = F_{X_i}(X_i)$. With P denoting the probability of any event, we may write

$$\begin{aligned} C(F_{T_1(X_1)}(x_1), ..., F_{T_n(X_n)}(x_n))) &= P[U_1 \leq F_{T_1(X_1)}(x_1), ..., U_n \leq F_{T_n(X_n)}(x_n)] \\ &= P[F^{-1}_{T_1(X_1)}(U_1) \leq x_1, ..., F^{-1}_{T_n(X_n)}(U_n) \leq x_n] \\ &= P[T_1 \circ F^{-1}_{X_1}(U_1) \leq x_1, ..., T_n \circ F^{-1}_{X_n}(U_n) \leq x_n] \\ &= P[T_1(X_1) \leq x_1, ..., T_n(X_n) \leq x_n]. \qquad \square \end{aligned}$$

The continuity of T_i, $i = 1, ..., n$, is a necessary condition, since the calculation above requires $F^{-1}_{T_i(X_i)}(U_1) = T_1 \circ F^{-1}_{X_i}(U_i)$. With this feature we can transform, for instance, a model of percentage returns on financial assets to a model of logarithmic returns, the copula of the data will not change, only the marginal distributions. In summary, in terms of copula modeling, it is indifferent to which of the marginal scales we work.

2.1.4 Linear Correlation and Kendall's τ Rank Correlation

This section will recall the definition of linear correlation and illustrate a brief description of the Kendall's τ rank correlation and its relation to the linear correlation. The presentation of this section follows mainly in Embrechts (2001).

Definition 2.13 (Linear Correlation) *Let X and Y be two real valued random variables with finite variances. The linear correlation coefficient between X and Y is*

$$\rho(X,Y) := \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}, \tag{2.14}$$

where $Cov(X,Y) := E(XY) - E(X)E(Y)$ is the covariance between X and Y, and $Var(X)$, $Var(Y)$ denotes the variances of X and Y.

In the case of perfect linear dependence, i.e. $Y = aX + b$ almost surely for $a \in \mathbb{R}\backslash\{0\}$ and $b \in R$, we have $|\rho(X,Y)| = 1$. Alternatively, $-1 < \rho(X,Y) < 1$. Moreover,

$$\rho(\alpha X + \beta, \gamma Y + \delta) = sign(\alpha\gamma)\rho(X,Y),$$

for $\alpha, \gamma \in \mathbb{R}\backslash\{0\}$ and $\beta, \delta \in \mathbb{R}$. The linear correlation is invariant under strictly increasing *linear* transformation, thus it can be easily calculated under linear operations. Let A, B be $m \times n$ matrices, $a, b \in \mathbb{R}^m$ and consider the n-dimensional vectors $\boldsymbol{X}, \boldsymbol{Y}$. Then

$$Cov(A\boldsymbol{X} + a, B\boldsymbol{Y} + b) = ACov(\boldsymbol{X}, \boldsymbol{Y})B^t.$$

It follows that for $\lambda \in \mathbb{R}^n$,

$$Var(\lambda^t \boldsymbol{X}) = \lambda^t Cov(\boldsymbol{X})\lambda$$

where $Cov(\boldsymbol{X}) := Cov(\boldsymbol{X}, \boldsymbol{X}) = Var(\boldsymbol{X})$.

In practice, the linear correlation is frequently used in measuring dependance structure between random variables. As mentioned before it is easily manipulatable under linear operations due to its invariance under strictly increasing *linear* transformations. For modeling multivariate data people often use joint multivariate normal or the joint multivariate Student's t-distribution in which linear correlation exposes its natural measure of dependence. However the linear correlation might prove inappropriate and misleading in certain situations. For instance, it does not make sense of using linear correlation for the jointly t-distributed random variables with degrees of freedom less than 3 since the linear correlation coefficient is not even defined because of infinite second moments. Another case of inappropriate use of linear correlation is where the random variables has different type of distributions and are actually not jointly normal or t-distributed. In such situations, copula-based measures of dependence demonstrate its benefit since it has the invariance property under strictly increasing transformations of the marginal distributions.

Kendall's τ rank correlation provides a distribution free test of independence and a measure of the strength of dependence between two variables.

Definition 2.14 *Let $(X,Y)^t$, $(\tilde{X},\tilde{Y})^t$ be two independent vectors of random variables with joint distributions F. The Kendall's τ for the random variables X,Y is defined as*

$$\rho_\tau(X,Y) := P[(X-\tilde{X})(Y-\tilde{Y})>0] - P[(X-\tilde{X})(Y-\tilde{Y})<0],$$

Theorem 2.15 *Let $(X,Y)^t$ be a vector of continuous random variables with copula C. The Kendall's τ for $(X,Y)^t$ is given by*

$$\tau(X,Y) = 4\int\int_{[0,1]^2} C(u,v)dC(u,v) - 1 \tag{2.15}$$

for $U,V \sim U(0,1)$.

The integral in (2.15) is the expected value of the random variable $C(U,V)$, i.e.

$$\tau(X,Y) = 4E[C(U,V)] - 1$$

Lindskog (2001) proves the relationship between the linear correlation coefficient and Kendall's τ of joint normal or joint t-distributed random vectors $(X,Y)^t$ can be expressed as

$$\tau(X,Y) = \frac{2}{\pi}\arcsin\rho(X,Y) \tag{2.16}$$

The above formula (2.16) provide a way of estimating Kendall's τ based on linear correlation coefficient of two random variables. Inversely, given the estimator of Kendall's τ we can estimate the linear correlation ρ, in particular,

$$\rho_\tau := \rho(X,Y) = \sin\left(\frac{\pi}{2}\tau\right) \tag{2.17}$$

with $\tau := \tau(X,Y)$.

2.2 Introduction to Meta-Elliptical Distributions

In this thesis we are specially interested in modeling data from t-distributions with specified marginal t-distributions and a given dependence structure. Thus an introduction to the class of elliptical distributions is essential. Furthermore we will obtain such a distribution by means of the copula construction. The theory in this section will follow the presentation in Embrechts (2001) and Fang et al. (2005).

2.2.1 Elliptical Distributions

According to Embrechts (2001) the elliptical distributions are stated as follows:

Definition 2.16 (Elliptical Distributions) *If* $\boldsymbol{X} = (X_1, ..., X_d)^t$ *is a* d*-dimensional random vector and for some* $\boldsymbol{\mu} \in \mathbb{R}^d$ *and some* $d \times d$ *nonnegative definite, symmetric matrix* Σ*, the characteristic function* $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\boldsymbol{t})$ *of* $\boldsymbol{X}-\boldsymbol{\mu}$ *is a function of the quadratic form* $\boldsymbol{t}^t\Sigma\boldsymbol{t}$*,* $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\boldsymbol{t}) = \phi(\boldsymbol{t}^t\Sigma\boldsymbol{t})$*, we say that* $\boldsymbol{X}$ *has an elliptical distribution with parameters* $\boldsymbol{\mu}$*,* Σ *and* ϕ*, and we write* $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \phi)$*.*

For $d = 1$ corresponds the class of elliptical distributions exactly to the class of one-dimensional symmetric distributions, which, for instance, the univariate normal or Student's t-distributions belong to. The function ϕ is called a *characteristic generator.*

Theorem 2.17 $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \phi)$ *with* $rank(\Sigma) = k$ *if and only if there exist a random variable* $\zeta \geqslant 0$ *independent of* $\boldsymbol{U}$*, an* k*-dimensional random vector uniformly distributed on the unit hypersphere* $\{\boldsymbol{z} \in \mathbb{R}^k | \boldsymbol{z}^t\boldsymbol{z} = 1\}$*, and an* $d \times k$ *matrix* A *with* $AA^t = \Sigma$*, such that*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \zeta A\boldsymbol{U} \tag{2.18}$$

The symbol "$\stackrel{d}{=}$" implicates that $\boldsymbol{X}$ has the same distribution as the right hand side of the equation (2.18).
Proof: see Fang and Ng (1987).

As pointed out in Embrechts (2001) as well as in Fang et al. (2005) an elliptical distributed random vector $\boldsymbol{X}$ does not necessarily possess a density. However, if $\boldsymbol{X}$ has a density, then the density of $\boldsymbol{X}$ is of the form

$$|\Sigma|^{-\frac{1}{2}} g((\boldsymbol{X}-\boldsymbol{\mu})^t\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})) \tag{2.19}$$

for some nonnegative function g of one scalar variable. $|\Sigma|$ represents the absolute value of the determinant of the matrix Σ. From equation (2.19) it follows that the function g is uniquely determined by the distribution of ζ since $\boldsymbol{X} - \boldsymbol{\mu} \stackrel{d}{=} \zeta A\boldsymbol{U}$ from the transformation of equation (2.18). Thus we can also use the notation $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g)$ for elliptical distributed vectors.

According to Kibria (2006) the characteristic function of a d-variate t-distributed random vector $\boldsymbol{Y}$ is given by

$$\phi_{\boldsymbol{Y}}(\boldsymbol{y}) = exp(i\boldsymbol{y}^t\boldsymbol{\mu}) \cdot \frac{\|(\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y}\|^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}-1}\Gamma(\frac{\nu}{2})} \cdot K_{\frac{\nu}{2}}(\|(\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y}\|)$$

where $\|\boldsymbol{t}\| = \sqrt{\boldsymbol{t}^t\boldsymbol{t}}$ and $K_{\frac{\nu}{2}}(\|(\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y}\|)$ is the Macdonald function with order $\frac{\nu}{2}$ and argument $\|(\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y}\|$. The notation $(\nu\Sigma)^{\frac{1}{2}}$ indicates that the upper triangle matrix of cholesky decomposition. The Macdonald function is defined as

$$K_\alpha(t) = \left(\frac{2}{t}\right)^\alpha \frac{\Gamma(\alpha+1)}{\sqrt{\pi}} \int_0^\infty (1+u^2)^{-(\alpha+\frac{1}{2})} \cos(tu)du$$

where $t > 0$ and $\alpha > -\frac{1}{2}$. We can rewrite the argument $\|(\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y}\|$ as

$$\begin{aligned}\|(\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y}\| &= \sqrt{((\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y})^t((\nu\Sigma)^{\frac{1}{2}}\boldsymbol{y})} \\ &= \sqrt{\nu(\Sigma^{\frac{1}{2}}\boldsymbol{y})^t(\Sigma^{\frac{1}{2}}\boldsymbol{y})} \\ &= \sqrt{\nu(\Sigma^{\frac{1}{2}}\boldsymbol{y})^t(\Sigma^{\frac{1}{2}}\boldsymbol{y})} \\ &= \sqrt{\nu\boldsymbol{y}^t\Sigma^{\frac{t}{2}}\Sigma^{\frac{1}{2}}\boldsymbol{y}} \\ &= \sqrt{\nu\boldsymbol{y}^t\Sigma\boldsymbol{y}},\end{aligned}$$

which is a function of $\boldsymbol{y}^t\Sigma\boldsymbol{y}$. Thus, the characteristic function $\phi_{\boldsymbol{Y}}(\boldsymbol{y})$ can be written as

$$\phi_{\boldsymbol{Y}}(\boldsymbol{y}) = \exp(i\boldsymbol{y}^t\mu) \cdot \psi(\boldsymbol{y}^t\Sigma\boldsymbol{y})$$

with

$$\psi(z) := \frac{\left(\nu z^{\frac{1}{2}}\right)^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}-1}\Gamma(\frac{\nu}{2})} \cdot K_{\frac{\nu}{2}}(\nu \cdot z^{\frac{1}{2}})$$

and therefore the multivariate t-distributions belong to the class of elliptical distributions.

Furthermore we use the representation (2.18) stated in Theorem 2.17 we obtain $Cov(\boldsymbol{X})$ for $\mathrm{E}[\zeta^2] < \infty$ as follows:

$$\begin{aligned}Cov(\boldsymbol{X}) &= Cov(\boldsymbol{\mu} + \zeta A\boldsymbol{U}) \\ &= A \cdot \mathrm{E}[\zeta^2] Cov(\boldsymbol{U}) \cdot A^t \\ &= AA^t \cdot \mathrm{E}[\zeta^2] \cdot \frac{1}{d}\end{aligned}$$

Quoted from Embrechts (2001), $Cov(\boldsymbol{U}) = \frac{1}{d}I_d$ can be shown as follows: Let $\boldsymbol{Y} \sim \mathcal{N}_d(\boldsymbol{0}, I_d)$, a d-variate standard normal distributed random vector. $\boldsymbol{U}$ is as defined in theorem 2.17. Then it follows that $\boldsymbol{Y} \overset{d}{=} \|\boldsymbol{Y}\|\boldsymbol{U}$, where $\|\boldsymbol{Y}\|$ is independent of $\boldsymbol{U}$. Moreover is $\|\boldsymbol{Y}\|^2 \sim \chi^2_d$ and therefore $E[\|\boldsymbol{Y}\|^2] = d$. Since $Cov(\boldsymbol{Y}) = I_d$ and $\boldsymbol{U}$ is uniformly distributed on the unit hypersphere in $\mathbb{R}^d$, then $Cov(\boldsymbol{U}) = \frac{1}{d}I_d$.

2.2.2 Meta-Elliptical Distributions

Following the presentations in Fang et al. (2005) let $\boldsymbol{Z} = (Z_1, ..., Z_d)^t \sim E_d(\boldsymbol{0}, R, g)$, where R is a symmetric positive definite matrix with entries

$$R := \{\rho_{ij} : -1 < \rho_{ij} < 1 \text{ for } i \neq j, \rho_{ij} = \rho_{ji} \text{ and } \rho_{ii} = 1, \text{ for } i, j = 1, ..., d\} \tag{2.20}$$

and some function $g(\cdot)$ satisfied (2.19). Then all the marginal distribution of $\boldsymbol{Z}$ are identical with the probability density function

$$q_g(x) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_{x^2}^{\infty} (y - x^2)^{\frac{d-1}{2}-1} g(y) dy, \tag{2.21}$$

and the cumulative distribution function

$$Q_g(x) = \frac{1}{2} + \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^x \int_{u^2}^{\infty} (y - u^2)^{\frac{d-1}{2}-1} g(y) dy du. \tag{2.22}$$

Let $\boldsymbol{X} = (X_1, ..., X_d)^t$ be a d-variate random vector with each component X_i having a continuous density $f_i(x_i)$ and cumulative distribution function $F_i(x_i)$ and let the random vector $\boldsymbol{Z} = (Z_1, ..., Z_d)^t \sim E_d(\mathbf{0}, R, g)$. Assume that

$$Z_i = Q_g^{-1}(F_i(X_i)), \quad i = 1, ..., d$$

where Q_g^{-1} is the inverse of Q_g. Then with the Jacobian of the transformation

$$\begin{aligned} J\{(z_1, ..., z_d)^t \to (x_1, ..., x_d)^t\} &= \prod_{i=1}^{d} \frac{dz_i}{dx_i} \\ &= \prod_{i=1}^{d} \frac{f_i(x_i)}{q_g(Q_g^{-1}(F_i(x_i)))}, \end{aligned}$$

and the density function of $\boldsymbol{X}$ is given by

$$h((x_1, ..., x_d)) = w(Q_g^{-1}(F_1(x_1)), ..., Q_g^{-1}(F_d(x_d))) \prod_{i=1}^{d} f_i(x_i), \tag{2.23}$$

where w is the d-variate density weighting function

$$w(z_1, ..., z_d) = |R|^{-\frac{1}{2}} g(\boldsymbol{z}^t \Sigma^{-1} \boldsymbol{z}) \cdot \frac{1}{\prod_{i=1}^{d} q_g(z_i)}. \tag{2.24}$$

The definition of meta-elliptical distributions is given by:

Definition 2.18 (Meta-Elliptical Distribution) *The d-dimensional random vector $\boldsymbol{X}$ is said to have a meta-elliptical distribution, if its density function is given by* (2.23), *i.e.*

$$h((x_1, ..., x_d)) = w(Q_g^{-1}(F_1(x_1)), ..., Q_g^{-1}(F_d(x_d))) \prod_{i=1}^{d} f_i(x_i),$$

As stated in Fang et al. (2005) the class of meta-elliptical distributions includes distributions such as elliptical distributions and diverse asymmetric distributions such as the asymmetric t-distributions (or we call them in this thesis the *meta t-distributions*) where the marginal distributions $F_i(\cdot)$ can be arbitrarily determined.

2.2.3 Bivariate Meta t-Distribution

As pointed out in Demarta and McNeil (2004) if a random vector $\boldsymbol{X}$ has the t-copula $C(\cdot; \nu, R)$ and univariate t-marginals with the same degree of freedom ν, then it has a multivariate t-distribution with ν degrees of freedom. If we use the Sklar's theorem to combine any other set of univariate distributions using the t-copula, we obtain multivariate distribution function F which have been termed meta t-distribution functions, for instance, the case where $F_1, ...F_d$ are univariate t-distributions with different degree of freedom parameters $\nu_1...\nu_d$.

Let $\boldsymbol{\Theta} := \{\nu_1, \nu_2, \nu, \rho | \nu_1, \nu_2, \nu \in [1, \infty), \rho \in (-1, 1)\}$ denote the parameter space in the bivariate meta t-distribution. Let ν_1, ν_2 be the parameters of the marginal distributions and ν, ρ be the parameters of the copula. The subsequent definition follows Demarta and McNeil (2004).

Definition 2.19 (Bivariate Meta t-Distribution) *Let $\boldsymbol{X} = (X_1, X_2)^t$ be independent random vectors. If the observation $\boldsymbol{x} = (x_1, x_2)^t$ has the joint density function*

$$h(x_1, x_2; \nu_1, \nu_2, \nu, \rho) = c(F(x_1; \nu_1), F(x_2; \nu_2); \nu, R) f(x_1; \nu_1) f(x_2; \nu_2), \tag{2.25}$$

$R := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$, *where $c(\cdot, \cdot; \nu, R)$ is the bivariate t-Copula defined in (2.12) with parameters ν and R and $f(\cdot; \nu_j)$ is the univariate t-distribution with degrees of freedom ν_j for $j = 1, 2$. We then call $\boldsymbol{X}$ to have a bivariate meta t-distribution or bivariate asymmetric t-distribution with (ν, ν_1, ν_2) degrees of freedoms and use the notation $\boldsymbol{X} \sim MT_2(\nu_1, \nu_2, \nu, R)$.*

With the univariate t-density function in (2.6) and the bivariate t-copula density function in (2.12) we have the joint density function of bivariate meta t-distributed observation $\boldsymbol{x} = (x_1, x_2)^t$ defined as

$$\begin{aligned} h(\boldsymbol{x}; \nu_1, \nu_2, \nu, \rho) &= c(F(x_1; \nu_1), F(x_2; \nu_2); \nu, \rho) f(x_1; \nu_1) f(x_2; \nu_2) \\ &= \frac{\frac{\nu}{2}\Gamma(\frac{\nu}{2})^2 (1 + \frac{z_1^2 + z_2^2 - 2z_1 z_2 \rho}{\nu(1-\rho^2)})^{-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu+1}{2})^2 \sqrt{1-\rho^2} (1 + \frac{z_1^2}{\nu})^{-\frac{\nu+1}{2}} (1 + \frac{z_2^2}{\nu})^{-\frac{\nu+1}{2}}} \cdot \\ &\quad \cdot \frac{\Gamma(\frac{\nu_1+1}{2})}{\sqrt{\nu_1 \pi}\Gamma(\frac{\nu_1}{2})} (1 + \frac{x_1^2}{\nu_1})^{-\frac{\nu_1+1}{2}} \cdot \frac{\Gamma(\frac{\nu_2+1}{2})}{\sqrt{\nu_2 \pi}\Gamma(\frac{\nu_2}{2})} (1 + \frac{x_2^2}{\nu_2})^{-\frac{\nu_2+1}{2}} \end{aligned} \tag{2.26}$$

with the notation $(z_1, z_2)^t := (t_\nu^{-1}(F(x_1, \nu_1)), t_\nu^{-1}(F(x_2, \nu_2)))^t$.

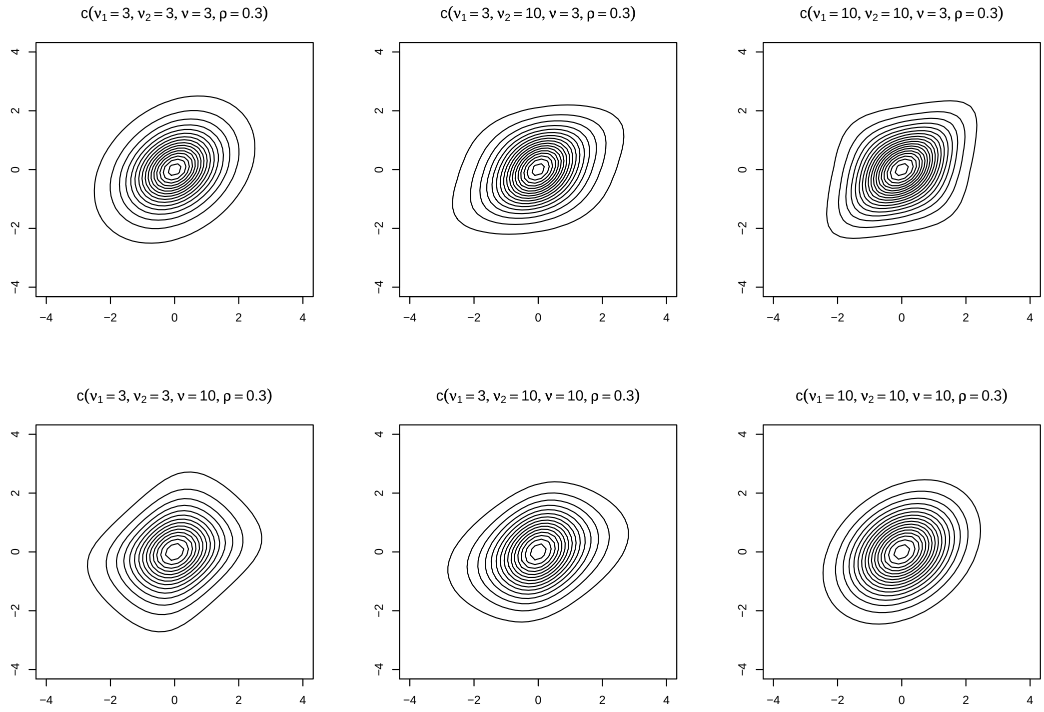


Figure 2.1: contour plots of bivariate meta t-density functions with $\rho = 0.3$ and varying degrees of freedom parameters ν

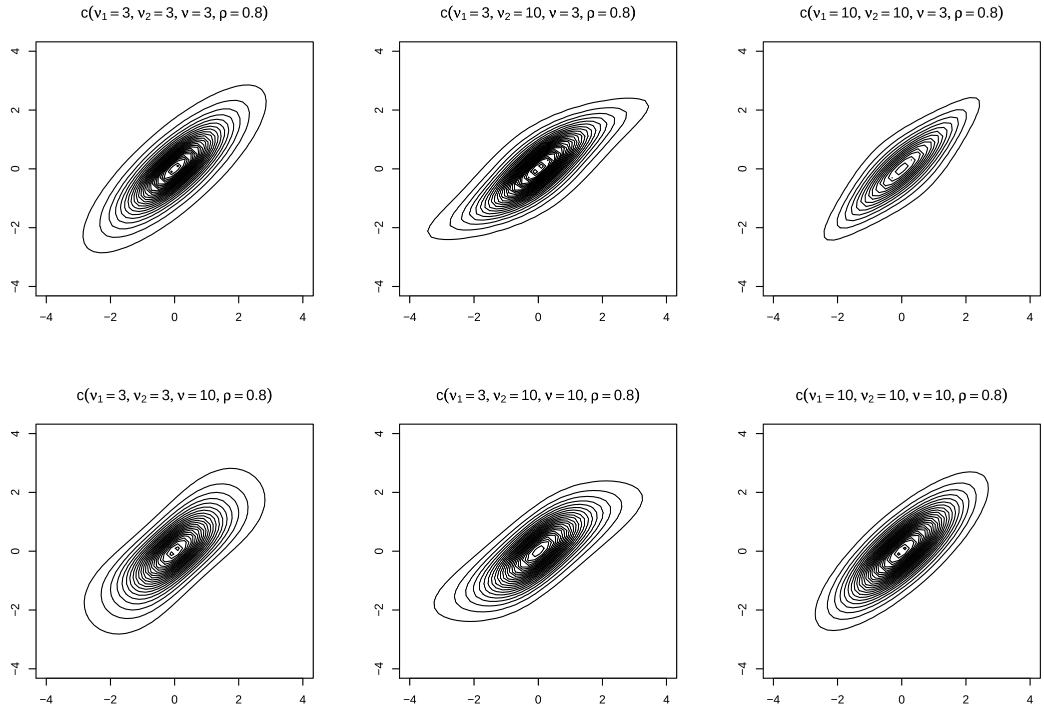


Figure 2.2: contour plots of bivariate meta t-density functions with $\rho = 0.8$ and varying degrees of freedom parameters ν

2.3 Maximum by Parts in Likelihood Inference

The first part of this section is based on the books from van der Vaart (1998) and Bickel and Doksum (1977). The second part on the paper of Song et al. (2005).

2.3.1 Maximum Likelihood Estimation

Let $\boldsymbol{Y} = (\boldsymbol{Y}_1, ...\boldsymbol{Y}_n)^t$ be independent random vectors and suppose $\boldsymbol{Y}_i$ has density in the parametric family $\{f_i(\boldsymbol{y};\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subset \mathcal{R}^p\}$. The likelihood function is given by

$$L(\boldsymbol{\theta};\boldsymbol{y}) := \prod_{i=1}^{n} f_i(\boldsymbol{y}_i;\boldsymbol{\theta})$$

The corresponding log-likelihood function is

$$\ell(\boldsymbol{\theta};\boldsymbol{y}) := \sum_{i=1}^{n} \ln f_i(\boldsymbol{y}_i;\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta};\boldsymbol{y}_i)$$

The Maximum Likelihood Estimator (abbr. ML-estimator) is defined as

$$\hat{\boldsymbol{\theta}} := \arg\max\{\ell(\boldsymbol{\theta};\boldsymbol{y}) : \boldsymbol{\theta} \in \Theta\}.$$

The ML-estimator may not be always existed and it may not be unique in case of existence. To discuss asymptotic properties of ML-estimator, we need the following *regularity conditions*

(i) The random vectors $\boldsymbol{Y}_i$ for $i = 1, 2, ...$ are independent and identically distributed with probability density function $f(\boldsymbol{y};\boldsymbol{\theta})$.

(ii) The parameter space Θ is compact.

(iii) The true but unknown parameter value $\boldsymbol{\theta}_0$ is identified as $\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}\in\Theta} E_{\boldsymbol{\theta}_0}[\ln f(\boldsymbol{Y}_i;\boldsymbol{\theta})]$

(iv) The log-likelihood function $\ell(\boldsymbol{\theta};\boldsymbol{y}) := \sum_{i=1}^{n} \ln f(\boldsymbol{y}_i;\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$.

(v) $E_{\boldsymbol{\theta}_0}[\ln f(\boldsymbol{Y}_i;\boldsymbol{\theta})]$ exists.

(vi) The log likelihood function is such that $\frac{1}{n}\ell(\boldsymbol{\theta};\boldsymbol{y})$ converges almost surely (in probability) to $E_{\boldsymbol{\theta}_0} \ln f(\boldsymbol{Y}_i, \boldsymbol{\theta})$ *uniformly* in $\boldsymbol{\theta} \in \Theta$, i.e.,

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\frac{1}{n}\ell(\boldsymbol{\theta};\boldsymbol{y}) - E_{\boldsymbol{\theta}_0} \ln f(\boldsymbol{Y}_i;\boldsymbol{\theta})\right| < \delta \text{ almost surely (in probability) for some } \delta > 0.$$

In order to find the ML-estimators $\hat{\boldsymbol{\theta}}$, it is necessary to solve the null of the partial derivative of the log likelihood function $\ell(\boldsymbol{\theta};\boldsymbol{y})$ with respect to $\boldsymbol{\theta}$, more precisely, the *score equations*

$$\mathbf{0} \stackrel{!}{=} \dot{\ell}(\boldsymbol{\theta};\boldsymbol{y}) := \frac{\partial \ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial \boldsymbol{\theta}}$$

Unfortunately, the sore equations do not often possess analytical solutions and in that case iteration methods such as the *Newton-Raphson Method* or the *Fisher-Scoring Method* are coming into consideration. First, we need to define some essential terms.

Definition 2.20 (Fisher Information) *The observed Fisher information matrix is given by*

$$\mathcal{I}_{obs}(\boldsymbol{\theta}; \boldsymbol{Y}) := -\left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\ell(\boldsymbol{\theta}; \boldsymbol{Y})\right\}$$

where $\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t$ denotes vectorial differentiation. The expected Fisher information matrix is defined by

$$\mathcal{I}(\boldsymbol{\theta}) := Cov\left[\dot{\ell}(\boldsymbol{\theta}; \boldsymbol{Y})\right] = E\left\{\dot{\ell}(\boldsymbol{\theta}; \boldsymbol{Y})\dot{\ell}(\boldsymbol{\theta}; \boldsymbol{Y})^T\right\}$$

Theorem 2.21 (Fisher Information) *If the regularity condition, $\int \frac{\frac{\partial^2}{\partial\boldsymbol{\theta}^t\boldsymbol{\theta}} f(Y;\boldsymbol{\theta})}{f(Y;\boldsymbol{\theta})^2} d\boldsymbol{\theta} = 0$, is met, then the expected Fisher information matrix may also be described directly by the observed Fisher information matrix as*

$$\mathcal{I}(\boldsymbol{\theta}) = E\left\{\mathcal{I}_{obs}(\boldsymbol{\theta}; \boldsymbol{Y})\right\} = -E\left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\ell(\boldsymbol{\theta}; \boldsymbol{Y})\right\}$$

Proof. Bickel and Doksum (1977).

For simplicity, let $\ddot{\ell}(\boldsymbol{\theta}; \boldsymbol{Y}) := \left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\ell(\boldsymbol{\theta}; \boldsymbol{Y})\right\} \in \mathbb{R}^{p\times p}$, the negative observed Fisher information matrix. It is also known as the second order derivative of the log likelihood with respect to $\boldsymbol{\theta}$ or the *Hessian matrix*. Let $\hat{\boldsymbol{\theta}}$ be the solution of the score equations $\dot{\ell}(\boldsymbol{\theta}; \boldsymbol{Y}) = \mathbf{0} \in \mathbb{R}^p$ and $\boldsymbol{\theta}_0$ be a point near $\hat{\boldsymbol{\theta}}$. According to the first order Taylor expansion around $\boldsymbol{\theta}_0$, the score function can be written by:

$$\mathbf{0} = \dot{\ell}(\hat{\boldsymbol{\theta}}; \boldsymbol{Y}) \approx \dot{\ell}(\boldsymbol{\theta}_0; \boldsymbol{Y}) + \ddot{\ell}(\boldsymbol{\theta}_0; \boldsymbol{Y})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

and it follows

$$\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^0 - \left\{\ddot{\ell}(\boldsymbol{\theta}_0; \boldsymbol{Y})\right\}^{-1}\left[\dot{\ell}(\boldsymbol{\theta}_0; \boldsymbol{Y})\right].$$

Therefore, we need to calculate the inverse of the observed Fisher information matrix and the score function in point $\boldsymbol{\theta}_0$.

Let $\boldsymbol{\theta}^0$ be a given initial value. In general, if $\ddot{\ell}(\boldsymbol{\theta}; \boldsymbol{Y})$ is not singular, then the kth iteration step of the Newton-Raphson algorithm is:

$$\boldsymbol{\theta}^{k+1} := \boldsymbol{\theta}^k - \left\{\ddot{\ell}(\boldsymbol{\theta}^k; \boldsymbol{Y})\right\}^{-1}\left[\dot{\ell}(\boldsymbol{\theta}^k; \boldsymbol{Y})\right], \quad k = 0, 1... \tag{2.27}$$

A variation on this is the *Fisher scoring algorithm*, in which the Hessian matrix $\ddot{\ell}(\boldsymbol{\theta}^k; \boldsymbol{Y})$ is replaced by its expectation or Fisher Information $E_{\boldsymbol{\theta}}\left\{\ddot{\ell}(\boldsymbol{\theta}^k; \boldsymbol{Y})\right\}$, because of its randomness.

Song et al. (2005) pointed out that for a complicated likelihood function, the calculation of the Hessian matrix is often troublesome due to problems with the invertibility or positive definiteness of the Hessian matrix at updated values. One approach to solve this problem is to replace the Fisher information by an estimate, such as

$$\frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_i(\boldsymbol{\theta}^{k-1};\boldsymbol{Y}_i)\dot{\ell}_i(\boldsymbol{\theta}^{k-1};\boldsymbol{Y}_i)^t, \tag{2.28}$$

instead to calculate the Fisher Information or the second order derivatives. This is feasible for a large n, but it can be very unstable, when n is small. Another approach is to approximate first and second order derivatives in the Newton-Raphson algorithm empirically. But when the likelihood is changing only slowly, the empirical approximation is very sensitive to the choice of grid points for differencing, and an algorithm built on this approximation may be fragile, especially in higher dimension.

One of the statistical tests based on the likelihood is the *likelihood ratio test* for the simple null hypothesis, $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. The statistic of the test is given by

$$\chi^2_{LR} := -2 \cdot (\ell(\boldsymbol{\theta}_0;\boldsymbol{Y}) - \ell(\hat{\boldsymbol{\theta}};\boldsymbol{Y})) \tag{2.29}$$

This statistic has an asymptotic chi-squared distribution with p degrees of freedom under the null hypothesis. The quantity p is equal to the difference in dimensionality of Θ_0 and Θ whereby Θ_0 denotes the parameter space of null hypothesis, a subset of parameter space Θ.

2.3.2 Maximization by Parts (MbP)

This section follows the presentation in Song et al. (2005). Song et al. (2005) proposed a new fix-point algorithm for maximizing complicated likelihood functions needed for maximum likelihood estimates by avoiding the calculation of the Hessian matrix.
Let $\boldsymbol{\theta} \in \mathbb{R}^p$ denote all the parameters of the log likelihood $\ell(\boldsymbol{\theta};\boldsymbol{Y})$. The main idea is to decompose the likelihood function $\ell(\boldsymbol{\theta};\boldsymbol{Y})$ into two parts:

$$\ell(\boldsymbol{\theta};\boldsymbol{Y}) = \ell_w(\boldsymbol{\theta};\boldsymbol{Y}) + \ell_e(\boldsymbol{\theta};\boldsymbol{Y}), \tag{2.30}$$

where $\ell_w(\boldsymbol{\theta};\boldsymbol{Y})$ is so-called *the working model* and $\ell_e(\boldsymbol{\theta};\boldsymbol{Y})$ *the error model*. Both models depend on the full parameters $\boldsymbol{\theta}$. The working model is a log likelihood function, relatively simple to be calculated and the error model is needed to update the estimates from the working model. In order to find the maximum likelihood estimates for $\boldsymbol{\theta}$, we need to solve the score equation of the likelihood function $\dot{\ell}(\boldsymbol{\theta};\boldsymbol{Y}) = \mathbf{0}$. Hence,

$$\dot{\ell}(\boldsymbol{\theta};\boldsymbol{Y}) = \dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{Y}) + \dot{\ell}_e(\boldsymbol{\theta};\boldsymbol{Y}) = 0 \tag{2.31}$$

The algorithm requires that $\dot{\ell}_w$ and $\dot{\ell}_e$ are unbiased inference functions, i.e. $E_{\boldsymbol{\theta}}[\dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{Y})] = 0$ and $E_{\boldsymbol{\theta}}[\dot{\ell}_e(\boldsymbol{\theta};\boldsymbol{Y})] = 0$, and solving $\dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{Y}) = \mathbf{0}$ is simple.

The proposed algorithm is given as follows:

Step 1. *Find* $\boldsymbol{\theta}^1_n$*, the solution to* $\dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{Y}) = \mathbf{0}$

Step k. *Find* $\boldsymbol{\theta}_n^k$*, the solution to* $\dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{Y}) = -\dot{\ell}_e(\boldsymbol{\theta}_n^{k-1};\boldsymbol{Y})$ *for* $k = 2, 3, ...$

Let $\boldsymbol{\theta}_0$ be the true value of the parameter. In step 1, let $\boldsymbol{\theta}_n^1$ denote the solution to $\dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{Y}) = \mathbf{0}$. In order to meet consistency and asymptotic normality of the estimator $\boldsymbol{\theta}_n^1$, $\dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{Y})$ is assumed to satisfy sufficient regularity conditions. The the consistency and asymptotic normality of $\boldsymbol{\theta}_n^1$ is stated in particular as: $\boldsymbol{\theta}_n^1$ is consistent, that is

$$\boldsymbol{\theta}_n^1 \xrightarrow{P} \boldsymbol{\theta}_0,$$

and asymptotically normal,

$$\sqrt{n}(\boldsymbol{\theta}_n^1 - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, J(\boldsymbol{\theta}_0)),$$

where $J(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}}\left\{\ddot{\ell}_w^t(\boldsymbol{\theta}_0;\boldsymbol{Y})\right\}^{-1} E_{\boldsymbol{\theta}}\left\{\dot{\ell}_w(\boldsymbol{\theta}_0;\boldsymbol{Y})\dot{\ell}_w^t(\boldsymbol{\theta}_0;\boldsymbol{Y})\right\} E_{\boldsymbol{\theta}}\left\{\ddot{\ell}_w(\boldsymbol{\theta}_0;\boldsymbol{Y})\right\}^{-1}$.

Since the full log likelihood function $\ell(\boldsymbol{\theta};\boldsymbol{Y})$ is not involved in the estimation of $\boldsymbol{\theta}_n^1$, it can have low efficiency. Song et al. (2005) suggests to use the information in the error model $\dot{\ell}_e$ in order to increase the efficiency in step k for $k = 2, 3, ...$ Therefore, it follows that the equation

$$\dot{\ell}_w(\boldsymbol{\theta};\boldsymbol{y}) = -\dot{\ell}_e(\boldsymbol{\theta}_n^{k-1};\boldsymbol{y})) \tag{2.32}$$

is to be solved for $\boldsymbol{\theta}_n^k$, in order to meet the score equation $\dot{\ell}_w(\boldsymbol{\theta}_n^k;\boldsymbol{y}) + \dot{\ell}_e(\boldsymbol{\theta}_n^{k-1};\boldsymbol{y}) \stackrel{!}{=} \mathbf{0}$. Equation (2.32) is also equivalent to the iteration

$$\boldsymbol{\theta}_n^k = \dot{\ell}_w^{-1}(-\dot{\ell}_e(\boldsymbol{\theta}_n^{k-1};\boldsymbol{y})). \tag{2.33}$$

Song et al. (2005) have demonstrated the consistency of the estimators $\boldsymbol{\theta}_n^k$ and its asymptotic normality given the proposed algorithm Their theory is based on the following regularity conditions:
In particular for a neighborhood of the true parameter $\boldsymbol{\theta}_0$, defined as
$\mathcal{U}_0 := \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 < \delta\}$, assume that

(I) $\ell(\boldsymbol{\theta};\boldsymbol{y}), \ell_w(\boldsymbol{\theta};\boldsymbol{y})$,and $\ell_e(\boldsymbol{\theta};\boldsymbol{y})$ are twice continuously differentiable for $\boldsymbol{\theta} \in \mathcal{U}_0$

(II) $\|\mathcal{I}_w^{-1}\mathcal{I}_e\|_2 < 1$, where $\mathcal{I}_w := -n^{-1}E_{\boldsymbol{\theta}}\{\ddot{\ell}_w(\boldsymbol{\theta}_0;\boldsymbol{Y})\}$, $\mathcal{I}_e := -n^{-1}E_{\boldsymbol{\theta}}\{\ddot{\ell}_e(\boldsymbol{\theta}_0;\boldsymbol{Y})\}$ and $\ddot{\ell}_j(\boldsymbol{\theta}_0;\boldsymbol{Y}) := \left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\ell_j(\boldsymbol{\theta};\boldsymbol{Y}) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\}$ for $j = w, e$ and $\|\cdot\|_2$ denotes the Euclidian norm.

The latter condition is referred to as *information dominance*. It implies that $(\mathcal{I}_w^{-1}\mathcal{I}_e)^k \to \mathbf{0}$ as $k \to \infty$, in other words the working model ℓ_w contains more information on $\boldsymbol{\theta}_0$ than the error model ℓ_e does.

Theorem 2.22 *Under condition (**I**), it follows that if* $\boldsymbol{\theta}_n^1$ *is consistent, then* $\boldsymbol{\theta}_n^2$ *is consistent.*

Proof. See Song et al. (2005), Theorem 1.

This theorem implies that if $\boldsymbol{\theta}_n^1$ is consistent, then $\boldsymbol{\theta}_n^k$ is consistent for each $k = 2, 3, ...$ Furthermore, for any integers k and m, the difference between two updates, $\boldsymbol{\theta}_n^{k+m}$ and $\boldsymbol{\theta}_n^k$ with m-step apart, will disappear in probability for $n \to \infty$. The fact that $\boldsymbol{\theta}_n^1$ is a consistent estimator of the true parameter $\boldsymbol{\theta}_0$ is essential for the theorem 2.22. However, as pointed out in Song et al. (2005), if $\boldsymbol{\theta}_n^1$ is not a consistent estimator of $\boldsymbol{\theta}_0$, based on the data and the given sample size n, $\boldsymbol{\theta}_n^k$ will still converge to the ML-estimator, as long as there exists an asymptotic contraction mapping[1]for $\boldsymbol{\theta}_n^k$.

Let $\boldsymbol{M}_n(\cdot)$ be the function which satisfies $\boldsymbol{\theta}_n^k = \boldsymbol{M}_n(\boldsymbol{\theta}_n^{k-1})$. If $\boldsymbol{M}_n(\cdot)$ is an asymptotic contraction mapping, then according to the Banach fixed point theorem there exists a fixed point, that is, $\hat{\boldsymbol{\theta}} = \lim_{k\to\infty} \boldsymbol{\theta}_n^k$. The fixed point $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{M}_n(\cdot)$ is the ML-estimator of $\boldsymbol{\theta}$. In our case, if $\dot{\ell}_w^{-1}$ exist,

$$\boldsymbol{M}_n(\boldsymbol{\theta}_n^{k-1}) := \dot{\ell}_w^{-1}(-\dot{\ell}_e(\boldsymbol{\theta}_n^{k-1}; \boldsymbol{y}))$$

as stated in (2.33). $\boldsymbol{M}_n(\cdot)$ is an asymptotic contraction mapping if the derivative of $\dot{\ell}_w^{-1}(-\dot{\ell}_e(\cdot))$, i. e.

$$-\{\ddot{\ell}_w(\dot{\ell}_w^{-1}(-\dot{\ell}_e(\cdot)))\}^{-1}\{\ddot{\ell}_e(\cdot)\}, \tag{2.34}$$

is bounded asymptotically in absolute value by a constant C_0 with probability 1, where $0 < C_0 < 1$. This is because given a consistent estimator $\hat{\boldsymbol{\theta}}$ such that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = o_p(1),$$ [2]

This derivative (2.34) can be expressed as $\mathcal{I}_w^{-1}\mathcal{I}_e + o_p(1)$.

Let

$$\tau_n(\boldsymbol{\theta}_0, \boldsymbol{Y}) = \left\{-\frac{1}{n}\ddot{\ell}_w(\boldsymbol{\theta}_0, \boldsymbol{Y})\right\}^{-1}\left\{\frac{1}{n}\ddot{\ell}_e(\boldsymbol{\theta}_0, \boldsymbol{Y})\right\},$$

so that

$$I_p - \tau_n(\boldsymbol{\theta}_0, \boldsymbol{Y}) = \left\{\frac{1}{n}\ddot{\ell}_w(\boldsymbol{\theta}_0, \boldsymbol{Y})\right\}^{-1}\left\{\frac{1}{n}\ddot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{Y})\right\},$$

where I_p is the identity matrix of dimension $p \times p$. Let $\tau = -\mathcal{I}_w^{-1}\mathcal{I}_e$. Then

$$\lim_{n\to\infty} \tau_n(\boldsymbol{\theta}_0, \boldsymbol{Y}) = \tau.$$

To establish the asymptotic distribution of $\boldsymbol{\theta}_n^k$, we need to consider the following regularity conditions,

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\dot{\ell}_w \\ \dot{\ell}_e\end{pmatrix} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Omega), \text{ as} \to \infty$$

[1]Let $(\Omega, \mathcal{A}, P)$ denote a probability space. For each $n \geqslant 1$ and $\omega \in \Omega$, let $K_n^\omega(\cdot)$ be a function defined on a set $\mathcal{X}$ where $(\mathcal{X}, d)$ is a metric space. The collection $\{K_n^\omega(\cdot) : n \geqslant 1, \omega \in \Omega\}$ is an *asymptotic contraction mapping* on $(\mathcal{X}, d)$ if there exist a constant c in $[0, 1)$ that does not depend on n or ω, and sets $\{A_n\}$ with each $A_n \in \Omega$ and $P(A_n) \to 1$ as $n \to \infty$, such that for each $\omega \in A_n$, $K_n^\omega(\cdot)$ maps $\mathcal{X}$ to itself and for all $x, y \in \mathcal{X}$. (Reference: Dominitz and Sherman (2005))

[2]Let Y_n and Y be random variables. $Y_n \xrightarrow{P} Y$, *if and only if* for all $\varepsilon > 0$: $\lim_{n\to\infty} P(|Y_n - Y| < \varepsilon) = 1$. $o_p(1)$ is defined as: $u_n = o_p(1) \iff u_n \xrightarrow{P} 0$, for $n \to \infty$.

where

$$\Omega := \lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} E\{\dot\ell_w \dot\ell_w^t\} & E\{\dot\ell_w \dot\ell_e^t\} \\ E\{\dot\ell_e \dot\ell_w^t\} & E\{\dot\ell_e \dot\ell_e^t\} \end{pmatrix} \in \mathbb{R}^{(p_1+p_2)\times(p_1+p_2)}$$

with $\dot\ell_w := \dot\ell_w(\boldsymbol{\theta}_0, \boldsymbol{Y})$ and $\dot\ell_e := \dot\ell_e(\boldsymbol{\theta}_0, \boldsymbol{Y})$.

Theorem 2.23 *Under regularity conditions (i) to (iv) as required in the ML-estimations, for any fixed k, $\boldsymbol{\theta}_n^k$ is asymptotically normally distributed with mean $\boldsymbol{\theta}_0$ and variance $\frac{1}{n}\Sigma_k$ with*

$$\Sigma_k = A_k^t \Omega A_k,$$

$$\mathcal{I} := E\{\dot\ell(\boldsymbol{\theta};\boldsymbol{Y})\dot\ell(\boldsymbol{\theta};\boldsymbol{Y})^t\} \textit{the expected fisher information matrix}$$

where

$$A_k = \begin{pmatrix} [I_p - \tau^k]\mathcal{I}^{-1} \\ \left[I_p - \tau^{k-1}\right]\mathcal{I}^{-1} \end{pmatrix}.$$

*Moreover, under the condition of information dominance in (**II**), as $k \to \infty$, $\Sigma_k \to \mathcal{I}^{-1}$, the inverse of the Fisher information.*

Proof. Song et al. (2005), Appendix A.

Under condition (**II**), $\tau^k \to \mathbf{0}$ as $k \to \infty$. Hence

$$\lim_{k\to\infty} \Sigma_k = \mathcal{I}^{-1}\left[\lim_{n\to\infty} \frac{1}{n} E\{\{\dot\ell_w + \dot\ell_e\}\{\dot\ell_w + \dot\ell_e\}^t\}\right]\mathcal{I}^{-1} = \mathcal{I}.$$

According to Theorem 2.23, as $k \to \infty$, the asymptotic variance matrix of $\boldsymbol{\theta}_n^k$ converges at an exponential rate to the asymptotic variance matrix of the ML-estimator $\hat{\boldsymbol{\theta}}$.

At each iteration, let

$$\hat{\mathcal{I}}^k := \frac{1}{n}\sum_{i=1}^n \dot\ell_i(\boldsymbol{\theta}_n^k; \boldsymbol{y}_i)\dot\ell_i(\boldsymbol{\theta}_n^k; \boldsymbol{y}_i)^t,$$

where $\ell_i(\cdot; \boldsymbol{y}_i)$ denotes the log likelihood with respect to observation $\boldsymbol{y}_i$. At convergence, $\boldsymbol{\theta}_n^k$ can be replaced by the ML-estimator $\hat{\boldsymbol{\theta}}$ and $\hat{\mathcal{I}} = \hat{\mathcal{I}}^k$ for $k \to \infty$. An estimate of $\mathcal{I}_w$ is $\hat{\mathcal{I}}_w = -\frac{1}{n}\sum_{i=1}^n \ddot\ell_{w,i}(\hat{\boldsymbol{\theta}}; \boldsymbol{y}_i)$ where $\ddot\ell_{w,i}(\hat{\boldsymbol{\theta}}; \boldsymbol{y}_i) := \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\ell_w(\boldsymbol{\theta}; \boldsymbol{y}_i)\mid_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. An estimate of $\mathcal{I}_e$ is then given by $\hat{\mathcal{I}}_e = \hat{\mathcal{I}} - \hat{\mathcal{I}}_w$. As pointed out in Song et al. (2005), Theorem 2.22 and 2.23 hold for any fixed iteration k, due to the fact that we use a consistent estimator $\boldsymbol{\theta}_n^1$ of $\boldsymbol{\theta}_0$ as the initial value.

An advantage of MbP framework is that the estimate of asymptotic variance matrix is available at the completion of the first iteration, which can be easily used to form a numerical check on the condition of information dominance. This is a crucial step, because it determines if the algorithm will converge to the ML-estimator based on the chosen working model. In effect, we compute

$$\hat\tau(\boldsymbol{\theta}_n^1) := \{\hat{\mathcal{I}}_w(\boldsymbol{\theta}_n^1)\}^{-1}\hat{\mathcal{I}}_e(\boldsymbol{\theta}_n^1),$$

where

$$\hat{\mathcal{I}}_w(\boldsymbol{\theta}) := -\frac{1}{n}\sum_{i=1}^{n} \ddot{\ell}_{w,i}(\boldsymbol{\theta}; \boldsymbol{y}_i),$$
$$\hat{\mathcal{I}}_e(\boldsymbol{\theta}) := \hat{\mathcal{I}}(\boldsymbol{\theta}) - \hat{\mathcal{I}}_w(\boldsymbol{\theta}),$$

and

$$\hat{\mathcal{I}}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_i(\boldsymbol{\theta}; \boldsymbol{y}_i)\dot{\ell}_i(\boldsymbol{\theta}; \boldsymbol{y}_i)^t.$$

Because $\boldsymbol{\theta}_n^1$ is consistent,

$$\hat{\tau}(\boldsymbol{\theta}_n^1) \xrightarrow{P} \tau(\boldsymbol{\theta}_0) \text{ as } n \to \infty.$$

If $\| \hat{\tau}(\boldsymbol{\theta}_n^1) \|_2 < 1$, then the information dominance condition is confirmed.

2.3.3 Modification of the MbP algorithm

In case that the error model ℓ_e contains all parameters of the log likelihood while the working model ℓ_w only includes parts of them, we need a modification of the previous algorithm.

Let $\boldsymbol{\theta}^t = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^t$, $\boldsymbol{\theta}_1 \in \mathbb{R}^{p_1}$, $\boldsymbol{\theta}_0$ be the true parameter and $\boldsymbol{\theta}_2 \in \mathbb{R}^{p_2}$, $p_1 + p_2 = p$ and assume the partition of the log likelihood $\ell(\boldsymbol{\theta}; \boldsymbol{Y}) = \ell_w(\boldsymbol{\theta}_1; \boldsymbol{Y}) + \ell_e(\boldsymbol{\theta}; \boldsymbol{Y})$. In this case the score function is given by:

$$\begin{pmatrix} \dot{\ell}_{w(1)}(\boldsymbol{\theta}_1; \boldsymbol{Y}) + \dot{\ell}_{e(1)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \boldsymbol{Y}) \\ \dot{\ell}_{e(2)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \boldsymbol{Y}) \end{pmatrix} = \mathbf{0}$$

where

$$\dot{\ell}_{i(j)}(\boldsymbol{\theta}_0; \boldsymbol{Y}) := \frac{\partial \ell_i(\boldsymbol{\theta}_0; \boldsymbol{Y})}{\partial \boldsymbol{\theta}_j} \quad \text{for } i = w, e \text{ and } j = 1, 2. \tag{2.35}$$

The maximization by parts now proceeds as follows:

STEP 1. *Find* $\boldsymbol{\theta}_{1,n}^1$, *the solution to* $\dot{\ell}_{w(1)}(\boldsymbol{\theta}_1; \boldsymbol{Y}) = \mathbf{0}$
Find $\boldsymbol{\theta}_{2,n}^1$, *the solution to* $\dot{\ell}_{e(2)}(\boldsymbol{\theta}_{1,n}^1, \boldsymbol{\theta}_2; \boldsymbol{Y}) = \mathbf{0}$

STEP 2. *Find* $\boldsymbol{\theta}_{1,n}^2$, *the solution to* $\dot{\ell}_{w(1)}(\boldsymbol{\theta}_1; \boldsymbol{Y}) = -\dot{\ell}_{e(1)}(\boldsymbol{\theta}_{1,n}^1, \boldsymbol{\theta}_{2,n}^1; \boldsymbol{Y})$
Find $\boldsymbol{\theta}_{2,n}^2$, *the solution to* $\dot{\ell}_{e(2)}(\boldsymbol{\theta}_{1,n}^1, \boldsymbol{\theta}_2; \boldsymbol{Y}) = \mathbf{0}$

In general:

STEP k. *Find* $\boldsymbol{\theta}_{1,n}^k$, *the solution to* $\dot{\ell}_{w(1)}(\boldsymbol{\theta}_1; \boldsymbol{Y}) = -\dot{\ell}_{e(1)}(\boldsymbol{\theta}_{1,n}^{k-1}, \boldsymbol{\theta}_{2,n}^{k-1}; \boldsymbol{Y})$
Find $\boldsymbol{\theta}_{2,n}^k$, *the solution to* $\dot{\ell}_{e(2)}(\boldsymbol{\theta}_{1,n}^{k-1}, \boldsymbol{\theta}_2; \boldsymbol{Y}) = \mathbf{0}$ *with* $k = 2, 3, ...;$

In Step 1, we search for the solution of the score equations of the working model. The solution, denoted by $\boldsymbol{\theta}_{1,n}^1$, is the ML-estimator of parameter in the working model. Then we use the ML-estimator from the working model to find the ML-estimators of the remaining parameters in the error model. With this way, we reduce the number of parameters that are to be maximized in the error model, thus in some cases it reduces the required computational cost. With this approach we may gain estimators with low efficiency. To solve the problem, Song et al. (2005) proposes to insert some information from the error model to the parameter calculation of the working model in Step 2. The updated estimate, denoted by $\boldsymbol{\theta}_{1,n}^2$, is then calculated by $\boldsymbol{\theta}_{1,n}^2 = \dot{\ell}_{w(1)}^{-1}(-\dot{\ell}_{e(1)}(\boldsymbol{\theta}_{1,n}^1, \boldsymbol{\theta}_{2,n}^1; \boldsymbol{Y}))$, where $\boldsymbol{\theta}_{1,n}^1$ and $\boldsymbol{\theta}_{2,n}^1$ are the results from Step 1. $\boldsymbol{\theta}_{1,n}^2$ is supposed to contain more information about the parameters than $\boldsymbol{\theta}_{1,n}^1$ from Step 1. The remainder of the parameters in the error model are updated using the estimates from Step 1 as well. In summary, for each Step k with $k = 2, 3, ...$, we use the estimates from the previous step to update the estimates of the working and error models in the current step.

Similar to the arguments in Theorem 2.22, if $\boldsymbol{\theta}_n^1 = (\boldsymbol{\theta}_{1,n}^1, \boldsymbol{\theta}_{2,n}^1)$ is consistent, then $\boldsymbol{\theta}_n^k = (\boldsymbol{\theta}_{1,n}^k, \boldsymbol{\theta}_{2,n}^k)$ for each k, $k = 2, 3, ...$ is consistent as well. Under regularity conditions, $\boldsymbol{\theta}_n^k$ will converge to its ML-estimator $\hat{\boldsymbol{\theta}}$ as $k \to \infty$.

In the following theorem, Song et al. (2005) established the asymptotic variance of $\boldsymbol{\theta}_n^k$ for each fixed k. Let

$$\begin{aligned}
\mathcal{I}_{w(11)} &:= -n^{-1} E\left\{\ddot{\ell}_{w(11)}(\boldsymbol{\theta}_0; \boldsymbol{Y})\right\} \in \mathbb{R}^{p_1+p_1}, \\
\mathcal{I}_{e(11)} &:= -n^{-1} E\left\{\ddot{\ell}_{e(11)}(\boldsymbol{\theta}_0; \boldsymbol{Y})\right\} \in \mathbb{R}^{p_2+p_2}, \\
\mathcal{I}_{e(12)} &:= -n^{-1} E\left\{\ddot{\ell}_{e(12)}(\boldsymbol{\theta}_0; \boldsymbol{Y})\right\} \in \mathbb{R}^{p_1+p_2}, \\
\mathcal{I}_{e(21)} &:= \mathcal{I}_{e(12)}^t, \\
\mathcal{I}_{e(22)} &:= -n^{-1} E\{\ddot{\ell}_{e(22)}(\boldsymbol{\theta}_0; \boldsymbol{Y})\} \in \mathbb{R}^{p_2+p_2},
\end{aligned}$$

where $\ddot{\ell}_{i(jk)}(\boldsymbol{\theta}; \boldsymbol{Y}) = \partial^2 \dot{\ell}_i(\boldsymbol{\theta}_0; \boldsymbol{Y}) / \partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k^t$ for $i = w, e$ and $j, k = 1, 2$. Using the abreviation

$\ddot{\ell}_{i(jk)} := \ddot{\ell}_{i(jk)}(\boldsymbol{\theta}_0; \boldsymbol{Y})$ define

$$\begin{aligned}
D_n &:= \begin{pmatrix} -n^{-1}\ddot{\ell}_{w(11)} & \mathbf{0} \\ \mathbf{0} & -n^{-1}\ddot{\ell}_{e(22)} \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{I}_{w(11)} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{e(22)} \end{pmatrix} + o_p(1) \\
&=: D + o_p(1), \\
T_n &:= \begin{pmatrix} n^{-1}\ddot{\ell}_{e(11)} & n^{-1}\ddot{\ell}_{e(12)} \\ n^{-1}\ddot{\ell}_{e(21)} & \mathbf{0} \end{pmatrix} \\
&= -\begin{pmatrix} \mathcal{I}_{e(11)} & \mathcal{I}_{e(12)} \\ \mathcal{I}_{e(21)} & \mathbf{0} \end{pmatrix} + o_p(1) \\
&=: T + o_p(1), \\
L_n &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ n(\ddot{\ell}_{e(22)})^{-1}(\ddot{\ell}_{e(21)})(\ddot{\ell}_{w(11)})^{-1} & \mathbf{0} \end{pmatrix} + o_p(1) \\
&=: L + o_p(1),
\end{aligned}$$

and $D_n^{-1}T_n$ converges in probability to $D^{-1}T$. Therefore, define

$$\Gamma := \operatorname*{plim}_{n\to\infty} D_n^{-1}T_n = D^{-1}T = -\begin{pmatrix} \mathcal{I}_{w(11)}^{-1}\mathcal{I}_{e(11)} & \mathcal{I}_{w(11)}^{-1}\mathcal{I}_{e(12)} \\ \mathcal{I}_{e(22)}^{-1}\mathcal{I}_{e(21)} & \mathbf{0} \end{pmatrix}$$

Using $\dot{\ell}_{i(j)} := \dot{\ell}_{i(j)}(\boldsymbol{\theta}_0; \boldsymbol{Y})$ and letting Ω_V be the asymptotic variance matrix of the estimating functions $n^{-1/2}[\dot{\ell}_{w(1)}, \dot{\ell}_{e(2)}, \dot{\ell}_{e(1)}, \mathbf{0}]^t$. Then Ω_V can be expressed as

$$\Omega_V = \lim_{n\to\infty}\frac{1}{n}\begin{pmatrix} E\{\dot{\ell}_{w(1)}\dot{\ell}_{w(1)}^t\} & E\{\dot{\ell}_{w(1)}\dot{\ell}_{e(2)}^t\} & E\{\dot{\ell}_{w(1)}\dot{\ell}_{e(1)}^t\} & \mathbf{0} \\ E\{\dot{\ell}_{e(2)}\dot{\ell}_{w(1)}^t\} & E\{\dot{\ell}_{e(2)}\dot{\ell}_{e(2)}^t\} & E\{\dot{\ell}_{e(2)}\dot{\ell}_{e(1)}^t\} & \mathbf{0} \\ E\{\dot{\ell}_{e(1)}\dot{\ell}_{w(1)}^t\} & E\{\dot{\ell}_{e(1)}\dot{\ell}_{e(2)}^t\} & E\{\dot{\ell}_{e(1)}\dot{\ell}_{e(1)}^t\} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{2p\times 2p} \tag{2.36}$$

Theorem 2.24 *Under the regularity conditions (i) to (vi),* $\boldsymbol{\theta}_n^k$ *is asymptotically normal with mean* $\boldsymbol{\theta}_0$ *and variance* $n^{-1}\Sigma_k$*, where*

$$\Sigma_k := B_k^t \Omega_V B_k \text{ with } B_k := [B_{k1}, B_{k2}]^t,$$

where

$$B_{k1} := \{I_p - \Gamma^k\}\mathcal{I}^{-1} + \Gamma^{k-1}L$$

and

$$B_{k2} := \{I_p - \Gamma^{k-1}\}\mathcal{I}^{-1}$$

Moreover, if $\Gamma^k \to \mathbf{0}$ *as* $k \to \infty$*, then* $\Sigma \to \mathcal{I}^{-1}$*.*

Proof. Song et al. (2005) Appendix A.

The condition

$$\lim_{k\to\infty} \Gamma^k \to \mathbf{0} \tag{2.37}$$

from Theorem 2.24 can be referred as *information dominance* of the modified MbP algorithm. This condition is fulfilled, if and only if the spectral radius is less than 1, e.g.

$$\lim_{k\to\infty} \Gamma^k \to \mathbf{0} \quad \Longleftrightarrow \quad \rho(\Gamma) < 1$$

where $\rho(\Gamma)$ is the spectral radius of Γ. However, this condition can be hardly verified in practice since the number of data is limited. An application of the MbP algorithm to a bivariate t-copula will be presented in the next chapter.

2.4 Quality of the Estimates

In order to know how good a estimation method works we need to measure the goodness of the estimators results from the respective algorithm. A good estimator of a parameter is expected to be close to the true parameter value. An algorithm is considered to have a good performance, if it consistently delivers good estimators after a number of runs. In the following we look at the statistics that determine the goodness of the estimators. We denote r as the number of estimates generated by an algorithm, $\hat{\theta}$ the estimator of a parameter θ and θ_{tr} its true value. To begin with we consider the under- and overestimation of an estimator that is evaluated by the bias. The applied notations in this section follow the chapter 2.6 of Kastenmeier (2008), Ripley (2002) and Brown (1982).

Definition 2.25 (Bias and Relative Bias) *Let $\hat{\theta}$ be the estimator of the parameter that possesses a true value of θ_{tr}. The bias of the estimator $\hat{\theta}$ is then defined as*

$$b(\hat{\theta}) := E[\hat{\theta}] - \theta_{tr} \tag{2.38}$$

and the relative bias is then given by

$$rb(\hat{\theta}) := \frac{b(\hat{\theta})}{\theta_{tr}} \tag{2.39}$$

If the expected value of the estimator is equal to the true parameter, i.e

$$b(\hat{\theta}) = 0 \Longleftrightarrow E[\hat{\theta}] = \theta_{tr}$$

then the estimator $\hat{\theta}$ is called unbiased.

An accurate estimator should not only have a small bias but also a small variance. Hence we expect of an good estimator that its expectation is greater than twice its standard deviation, i.e.

$$\frac{|E[\hat{\theta}]|}{\sqrt{Var(\hat{\theta})}} > 2.$$

Furthermore the expectation $E[\hat{\theta}]$ can be estimated using the sample mean which is defined as

$$\bar{\theta} := \frac{1}{r}\sum_{i=1}^{r}\hat{\theta}_i, \tag{2.40}$$

where r is the number of estimates. Then we are interested in the variance of the sample mean $\bar{\theta}$. This is given by

$$Var(\bar{\theta}) = Var\left(\frac{1}{r}\sum_{i=1}^{r}\hat{\theta}_i\right) = \frac{1}{r^2}\sum_{i=1}^{r}Var(\hat{\theta}_i) = \frac{1}{r}Var(\hat{\theta}) \tag{2.41}$$

where the last step results from the i.i.d. property of the estimates $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_r$. Moreover, an unbiased and consistent estimator for the variance of an estimate , i.e. $Var(\hat{\theta})$, is the sample variance $s^2(\hat{\theta})$ (Georgii (2002), p.194 and p.202) which is given by

$$s^2(\hat{\theta}) := \frac{1}{r-1}\sum_{i=1}^{r}(\hat{\theta}_i - \bar{\theta})^2. \tag{2.42}$$

Thus the unbiased and consistent estimator for $Var(\bar{\theta})$ are calculated as follows:

$$s^2(\bar{\theta}) := \frac{1}{r}\left(s^2(\hat{\theta})\right) = \frac{1}{r}\left(\frac{1}{r-1}\sum_{i=1}^{r}(\hat{\theta}_i - \bar{\theta})^2\right) \tag{2.43}$$

In addition, the condition $\bar{\theta}/s(\bar{\theta}) > 2$ must be satisfied.

By employing the sample mean, we can specify an estimator for the bias of $\hat{\theta}$ by

$$\hat{b}(\hat{\theta}) := \left(\frac{1}{r}\sum_{i=1}^{r}\hat{\theta}_i\right) - \theta_{tr} = \bar{\theta} - \theta_{tr}. \tag{2.44}$$

Then the estimator of relative bias of $\bar{\theta}$ can be expressed as

$$\widehat{rb}(\bar{\theta}) := \frac{\bar{\theta} - \theta_{tr}}{\theta_{tr}} = \frac{\hat{b}(\hat{\theta})}{\theta_{tr}} \tag{2.45}$$

The variance of estimated relative bias of $\bar{\theta}$ can be written as

$$Var(\widehat{rb}(\bar{\theta})) := Var\left(\frac{\hat{b}(\hat{\theta})}{\theta_{tr}}\right) = Var\left(\frac{\bar{\theta} - \theta_{tr}}{\theta_{tr}}\right) = \frac{1}{\theta_{tr}^2}Var(\bar{\theta} - \theta_{tr}) = \frac{1}{\theta_{tr}^2}Var(\bar{\theta}) \tag{2.46}$$

Then the sample variance of estimated relative bias of $\bar{\theta}$ is given by

$$s_{rb}^2(\bar{\theta}) := \frac{1}{\theta_{tr}^2}s^2(\bar{\theta}) = \frac{1}{\theta_{tr}^2 \cdot r}\left(s^2(\hat{\theta})\right) = \frac{1}{\theta_{tr}^2 \cdot r}\left(\frac{1}{r-1}\sum_{i=1}^{r}(\hat{\theta}_i - \bar{\theta})^2\right). \tag{2.47}$$

Another way to quantify the amount by which an estimator differs from the true value of the quantity being estimated is the *mean squared error* (abbr. *mse*) of an estimator. The quantity *mse* measures the average of the square of the error which is also the bias as defined in the previous definition.

Definition 2.26 *(Mean squared error) Let $\hat{\theta}$ be the estimator of the parameter with its true value θ. The mean squared error (mse) of the estimator is then defined as*

$$mse(\hat{\theta}) := E[(\hat{\theta} - \theta_{tr})^2] \tag{2.48}$$

The mean square error can be also rewritten as the sum of the variance of the estimator $\hat{\theta}$ and its squared bias:

$$\begin{aligned}
mse(\hat{\theta}) &= E[(\hat{\theta} - \theta_{tr})^2] \\
&= E[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta_{tr})^2] \\
&= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2E[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta_{tr})] + E[(E[\hat{\theta} - \theta_{tr}])^2] \\
&= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2E[\hat{\theta}E[\hat{\theta}] - E[\hat{\theta}]^2 - \hat{\theta}\theta_{tr} + E[\hat{\theta}]\theta_{tr}] + E[(E[\hat{\theta}] - \theta_{tr})^2] \\
&= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2 \cdot (E[\hat{\theta}]E[\hat{\theta}] - E[\hat{\theta}]^2 - E[\hat{\theta}]\theta_{tr} + E[\hat{\theta}]\theta_{tr}) + E[(E[\hat{\theta}] - \theta_{tr})^2] \\
&= E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta_{tr})^2 \\
&= Var(\hat{\theta}) + b(\hat{\theta})^2
\end{aligned}$$

Due to this feature, we can estimate the mean squared error of $mse(\hat{\theta})$ by

$$\widehat{mse}(\hat{\theta}) := s^2(\hat{\theta}) + \hat{b}(\hat{\theta})^2 = s^2(\hat{\theta}) + (\bar{\theta} - \theta_{tr})^2 \tag{2.49}$$

To obtain the variance of the mean squared error, we first need an estimator for the random variable mse. This is usually done by the sample mean, i.e.

$$\widehat{mse}^*(\hat{\theta}) = \frac{1}{r}\sum_{i=1}^{r}(\hat{\theta}_i - \theta_{tr})^2$$

The variance of the estimator is calculated as follows:

$$\begin{aligned}
Var[\widehat{mse}^*(\hat{\theta})] &= Var\left[\frac{1}{r}\sum_{i=1}^{r}(\hat{\theta}_i - \theta_{tr})^2\right] \\
&= \frac{1}{r^2}Var\left[\sum_{i=1}^{r}(\hat{\theta}_i - \theta_{tr})^2\right] \\
&= \frac{1}{r}Var\left[(\hat{\theta} - \theta_{tr})^2\right] \\
&= \frac{1}{r}Var\left[\hat{\theta}^2 - 2\hat{\theta}\theta_{tr} + \theta_{tr}^2\right] \\
&= \frac{1}{r}Var\left[\hat{\theta}^2 - 2\hat{\theta}\theta_{tr}\right] \\
&= \frac{1}{r}\left(E[(\hat{\theta}^2 - 2\hat{\theta}\theta_{tr})^2] - (E[\hat{\theta}^2] - 2E[\hat{\theta}]\theta_{tr})^2\right) \\
&= \frac{1}{r}\left(E[\hat{\theta}^4 - 4\hat{\theta}^3\theta_{tr} + 4\hat{\theta}^2\theta_{tr}^2] - (E[\hat{\theta}^2]^2 - 4E[\hat{\theta}^2]E[\hat{\theta}]\theta_{tr} + 4E[\hat{\theta}]^2\theta_{tr}^2)\right) \\
&= \frac{1}{r}\left(E[\hat{\theta}^4] - 4E[\hat{\theta}^3]\theta_{tr} + 4E[\hat{\theta}^2]\theta_{tr}^2 - E[\hat{\theta}^2]^2 + 4E[\hat{\theta}^2]E[\hat{\theta}]\theta_{tr} - 4E[\hat{\theta}]^2\theta_{tr}\right)
\end{aligned}$$

Moreover, to gain the variance of the estimated mse we substitute the moments $E[\hat{\theta}^k]$ for $k = 1, ..., 4$ by its the empirical moments given by

$$m_k := \frac{1}{r}\sum_{i=1}^{r}\hat{\theta}_i^k \tag{2.50}$$

Thus, the standard error of the estimated mse can be estimated as

$$s_{mse}(\hat{\theta}) := \frac{1}{\sqrt{r}}\sqrt{(m_4 - 4\theta_{tr}m_3 + 4\theta_{tr}^2 m_2 - m_2^2 + 4\theta_{tr}m_2 m_1 - 4\theta_{tr}^2 m_1^2)} \tag{2.51}$$

where the true parameter θ_{tr} is estimated by the sample mean $\bar{\theta}$.

One drawback of the estimators such as sample mean, sample variance or the mean squared error is that these estimators might be distractive if there are few outliers. Hence we need a resistant measure of the variability of the samples. The next definitions and descriptions follow Ripley (2002) and Brown (1982).

Definition 2.27 (Median absolute deviation) *Let $\hat{\theta}_1, ..., \hat{\theta}_r$ be r independent estimates of θ_{tr}. The median absolute deviation (or mad) is defined as*

$$mad(\hat{\theta}_1, ..., \hat{\theta}_r) := median(|\hat{\theta}_i - median(\hat{\theta}_1, ..., \hat{\theta}_r)|), \quad i = 1...r \tag{2.52}$$

The mad has advantages over the standard error, since it is more resilient to outliers in a data set. In the standard error the distance from the mean are squared, in the average large deviations are weighted more heavily. In the mad, due to the use of median, the magnitude of the distances of a small number of outliers is irrelevant and negligible.
In addition, by exploring a relationship to the standard deviation we can receive an estimate for the standard deviation σ from mad, i.e.

$$\hat{\sigma} = c \cdot mad(\hat{\theta}_1, ..., \hat{\theta}_r)$$

where c is a constant. The constant ensures consistency, if it takes the value $c = 1.4826$ (approximately $c \approx \frac{1}{\Phi^{-1}(\frac{3}{4})}$, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution), i.e.

$$E[c \cdot mad(\hat{\theta}_1, ..., \hat{\theta}_r)] = \sigma$$

for $\hat{\theta}_i$, i.i.d. $\forall i = 1, ..., r$ distributed as $\mathcal{N}(\mu, \sigma^2)$ for large r.

Definition 2.28 (Estimated robust mean squared error) *The estimated robust mean squared error of $\hat{\theta}$ is defined as*

$$\widehat{rmse}(\hat{\theta}) := (median(\hat{\theta}_1, ..., \hat{\theta}_r) - \theta_{tr})^2 + \left(\frac{mad(\hat{\theta}_1, ..., \hat{\theta}_r)}{\Phi^{-1}(\frac{3}{4})}\right)^2 \tag{2.53}$$

Another way to investigate the statistical dispersion of the independent estimates $\hat{\theta}_1, ..., \hat{\theta}_r$ we can use the estimated *interquartile range*, denoted by $\widehat{IQR}(\hat{\theta})$, which is equal to the difference between the empirical 75th and the 25th percentile of the estimator $\hat{\theta}$, i.e.

$$\widehat{IQR}(\hat{\theta}) := \hat{\theta}_{0.75r} - \hat{\theta}_{0.25r} \tag{2.54}$$

Here $\hat{\theta}_{0.75r}$ ($\hat{\theta}_{0.25r}$) denotes the empirical 75th (25th) quantile of the random variable $\hat{\theta}$. The $\widehat{IQR}(\hat{\theta})$ is essentially the range of the middle 50% of the data and is not affected by outliers or extreme values. The smaller the range of $\widehat{IQR}(\hat{\theta})$ is, the smaller is variability of the estimates of the corresponding method. According to Ripley (2002) the relationship between IQR and the standard deviation of a normal distributed $\hat{\theta}_i$, $\hat{\theta}_i \sim \mathcal{N}(\mu, \sigma^2)$, i.i.d. $\forall i = 1, ..., r$, is described by

$$IQR(\hat{\theta}) \rightarrow \sigma[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)] \approx 1.35\sigma \tag{2.55}$$

for large r. Then we define the *estimated robust standard error* $\hat{\sigma}_r$ of $\hat{\theta}$ by

$$\hat{\sigma}_r(\hat{\theta}) := \widehat{IQR}(\hat{\theta})/1.35 \tag{2.56}$$

In the later chapter, we are going to compare the stability and efficiency of the estimation methods. To achieve this we can calculate the ratio of the mse and $rmse$ values from estimates generated by the alternative method to those generated by the reference method. In particular, assume we compare the methods $M_1, ..., M_d$. Without restriction we define M_1 as the reference estimation method, then the efficiencies of the estimates from M_j, $j = 1, ..., d$, based on mse and $rmse$, are defined respectively as follows:

$$\text{eff}_{mse}(M_j) := \frac{mse_{M_j}(\hat{\theta})}{mse_{M_1}(\hat{\theta})}$$

and

$$\text{eff}_{rmse}(M_j) := \frac{rmse_{M_j}(\hat{\theta})}{rmse_{M_1}(\hat{\theta})}$$

Then their estimated efficiencies are

$$\widehat{\text{eff}}_{mse}(M_j) := \frac{\widehat{mse}_{M_j}(\hat{\theta})}{\widehat{mse}_{M_1}(\hat{\theta})} \tag{2.57}$$

and

$$\widehat{\text{eff}}_{rmse}(M_j) := \frac{\widehat{rmse}_{M_j}(\hat{\theta})}{\widehat{rmse}_{M_1}(\hat{\theta})} \tag{2.58}$$

A smaller ratio of $\text{eff}_{mse}(M_j)$ or $\text{eff}_{rmse}(M_j)$ indicates that the alternative method M_j has a smaller mse or $rmse$, thus the method M_j operates more steadily and efficiently toward the reference method M_1. More precisely, if $\text{eff}_{mse}(M_j) < 1$ then the method M_j performs better than M_1 and similarly if $\text{eff}_{mse}(M_j) > 1$ the reference method M_1 produces more robust estimates.

The statistical quantities defined in this section are summed up in an overview in the Table 2.1 on the following page.

Overview		
Name	Theoretical	Empirical
mean of $\hat{\theta}$	$E[\hat{\theta}]$	$\bar{\theta} := \frac{1}{r}\sum_{i=1}^{r}\hat{\theta}_i$, $\hat{\theta}_i$ i.i.d. $\forall i$, def. in (2.40)
bias of $\hat{\theta}$	$b(\hat{\theta}) := E[\hat{\theta}] - \theta_r$, def. in (2.38)	$\hat{b}(\hat{\theta}) := \frac{1}{r}\sum_{i=1}^{r}\hat{\theta}_i - \theta_{tr} = \bar{\theta} - \theta_{tr}$, def. in (2.44)
relative bias of $\hat{\theta}$	$rb(\bar{\theta}) := \dfrac{b(\hat{\theta})}{\theta_{tr}}$, def. in (2.39)	$\widehat{rb}(\bar{\theta}) := \dfrac{\hat{b}(\hat{\theta})}{\theta_{tr}} = \dfrac{\bar{\theta} - \theta_{tr}}{\theta_{tr}}$, def. in (2.45)
Variance of $\hat{\theta}$	$Var(\hat{\theta}) := E[(\hat{\theta} - E[\hat{\theta}])^2]$	$s^2(\hat{\theta}) := \dfrac{1}{r-1}\sum_{i=1}^{r}(\hat{\theta}_i - \bar{\theta})^2$, def. in (2.42)
Variance of $\bar{\theta}$	$Var(\bar{\theta}) := \dfrac{1}{r}Var(\hat{\theta})$, def. in (2.41)	$s^2(\bar{\theta}) := \dfrac{1}{r}s^2(\hat{\theta}) = \dfrac{1}{r(r-1)}\sum_{i=1}^{r}(\hat{\theta}_i - \bar{\theta})^2$, def. in (2.43)
Standard error of $\bar{\theta}$	$\sqrt{Var(\bar{\theta})}$	$s(\bar{\theta}) := \sqrt{s^2(\bar{\theta})} = \dfrac{1}{\sqrt{r(r-1)}}\sqrt{\sum_{i=1}^{r}(\hat{\theta}_i - \bar{\theta})^2}$
Variance of $\widehat{rb}(\bar{\theta})$	$Var(\widehat{rb}(\bar{\theta})) := \frac{1}{\theta_{tr}^2}Var(\bar{\theta})$, def. in (2.46)	$s_{rb}^2(\bar{\theta}) := \frac{1}{\theta_{tr}^2}s^2(\bar{\theta})$, def. in (2.47)
Standard error of $\widehat{rb}(\bar{\theta})$	$\sqrt{Var(\widehat{rb}(\hat{\theta}))}$	$s_{rb}(\bar{\theta}) := \frac{1}{\theta_{tr}}s(\bar{\theta})$
Mean squared error of $\hat{\theta}$	$mse(\hat{\theta}) := Var(\hat{\theta}) + b(\hat{\theta})^2$, def. in (2.48)	$\widehat{mse}(\hat{\theta}) := s^2(\hat{\theta}) + \hat{b}(\hat{\theta})^2$, def. in (2.49)
kth moment of $\hat{\theta}$	$E[\hat{\theta}^k]$, $k = 1, 2, 3, ...$	$m_k := \frac{1}{r}\sum_{i=1}^{r}\hat{\theta}_i^k$, def. in (2.50)
Variance of $mse(\hat{\theta})$	$Var(mse(\hat{\theta}))$	$s_{mse}^2(\hat{\theta}) := \frac{1}{r}\left(m_4 - 4\theta_{tr}m_3 + 4\theta_{tr}^2 m_2 - m_2^2 + 4\theta_{tr}m_2m_1 - 4\theta_{tr}^2 m_1^2\right)$
Standard error of $mse(\hat{\theta})$	$\sqrt{Var(mse(\hat{\theta}))}$	$s_{mse}(\hat{\theta}) := \frac{1}{\sqrt{r}}\sqrt{\left(m_4 - 4\theta_{tr}m_3 + 4\theta_{tr}^2 m_2 - m_2^2 + 4\theta_{tr}m_2m_1 - 4\theta_{tr}^2 m_1^2\right)}$, def. in (2.51)
Est. robust mean squared error of $\hat{\theta}$		$\widehat{rmse}(\hat{\theta}) := (median(\hat{\theta}_1, ..., \hat{\theta}_r) - \theta_{tr})^2 + \left(\dfrac{mad(\hat{\theta}_1, ..., \hat{\theta}_r)}{\Phi^{-1}(\frac{3}{4})}\right)^2$. def in (2.53)
Interquartile range of θ	$IQR(\hat{\theta}) := \sigma[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)]$ for normal i.i.d. $\hat{\theta}$, def. in (2.55)	$\widehat{IQR}(\hat{\theta}) := \hat{\theta}_{0.75r} - \hat{\theta}_{0.25r}$, def. in (2.54)
Est. robust standard error of $\hat{\theta}$		$\hat{\sigma}_r(\hat{\theta}) := \widehat{IQR}(\hat{\theta})/1.35$, def. in (2.56)

Table 2.1: Statistics needed to assess the quality of estimators

2.5 One-Dimensional Bisection Method

The one-dimensional bisection method that we utilize to find the root of the score functions of the modified working and error models process as follows.

Suppose that given two real-valued points a and b with $a < b$, $f(a)$ and $f(b)$ have opposite signs. According to intermediate value theorem, function f must have at least one root in the interval $[a, b]$ as long as f is continuous.

The bisection method divides the interval $[a, b]$ in two by computing $c = \frac{a+b}{2}$. Now either $f(a)$ and $f(c)$ have opposite signs or $f(b)$ and $f(c)$ have opposite signs. The bisection algorithm is then applied recursively to the sub interval where the sign change occurs.

Algorithm 2.1 Bisection Algorithm

Input: f, a, b, tol
Output: the root of f
Set $c := \frac{a+b}{2}$
if $f(a) \cdot f(b) > 0$ or $a > b$ **then**
    return (improper start values)
else
    while $\| f(c) \| > tol$ **do**
        if $f(a) \cdot f(c) > 0$ **then**
            $a := c$ and $c := \frac{a+b}{2}$
        else
            $b := c$ and $c := \frac{a+b}{2}$
        end if
    end while
end if

When the tolerance $tol := 10^{-8}$ is reached, the bisection algorithm is said to be converged. The number $tol := 10^{-8}$ also implies that the distance between the updated points a and b in the algorithm is less than 10^{-8} at convergence.

2.6 Gamma Function and its Derivatives

The Gamma function is a component in various probability distribution functions, such as the t- and meta t-distributions. In the later analysis we require definitions of the gamma function and its first and second derivatives.

Definition 2.29 *For $z \in \mathbb{R}$ the digamma function is $\psi(z)$ defined as*

$$\psi(z) := \frac{\partial}{\partial z} \ln \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}.$$

where $\Gamma'(z)$ is the first order derivative of $\Gamma(\cdot)$ with respect to z.

For $\ln \Gamma(f(z))$ we have

$$\begin{aligned}\frac{\partial}{\partial z} \ln \Gamma(f(z)) &= \frac{1}{\Gamma(f(z))} \cdot \frac{\partial}{\partial z} \Gamma(f(z)) \\ &= \frac{1}{\Gamma(f(z))} \cdot \psi(f(z)) \cdot \Gamma(f(z)) \cdot f'(z) \\ &= f'(z) \cdot \psi(f(z))\end{aligned}$$

The first order derivative of the digamma function is denoted as the trigamma function.

Definition 2.30 *For $z \in \mathbb{R}$ the trigamma function $\psi^{(1)}$ is defined as*

$$\psi^{(1)}(z) := \frac{\partial^2}{\partial z^2} \ln \Gamma(z).$$

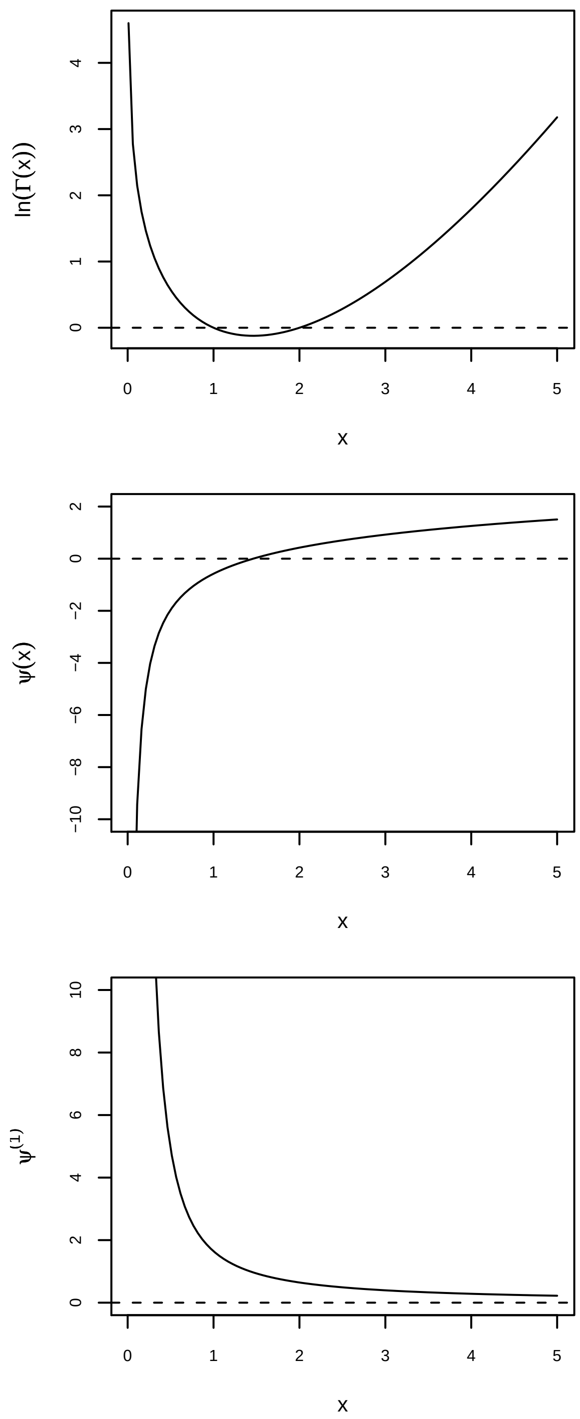


Figure 2.3: Log gamma function $\ln \Gamma(z)$, digamma function $\psi(z)$ and trigamma function $\psi^{(1)}(z)$

2.7 Introduction to Time Series Analysis

2.7.1 Basic Definitions

A brief introduction to the theory of time series analysis is based on the description in Brockwell and Davis (1996) and Zivot (2003).

Definition 2.31 (Stochastic process) *A stochastic process is a family of random variables* $\{X_t, t \in \mathcal{T}\}$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$.

The index T in previous definition is a set of time points, which can be $\{0, \pm 1, \pm 2, ...\}$, $\{0, 1, 2, ...\}$, $[0, \infty)$ or $(-\infty, \infty)$. Furthermore, the definition of a random variable implies that if $t \in \mathcal{T}$ is fixed, X_t is a function of $X_t(\cdot)$ on the set Ω. Reversely, for each fixed $\omega \in \Omega$, $X_t(\omega)$ is a function on T.

Definition 2.32 (Realizations of a stochastic process) *The functions* $\{X_t(\omega), \omega \in \Omega\}$ *on* T *are known as the realizations or sample-paths of the process* $\{X_t, \omega \in \Omega\}$.

The term *time series* refers to both the data and the process of which it is a realization.

Definition 2.33 (Autocovariance function) *Let* $(X_t)_{t\in T}$ *be a time series with* $E[X_t^2] < \infty$. *The covariance function of* $(X_t)_{t\in T}$ *is*

$$\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

for all integers r *and* s.

Definition 2.34 *A time series* $(X_t)_{t\in\mathbb{Z}}$ *with* $E[X_t^2] < \infty$ *is (weakly) stationary if*

(i) $E[X(t)] = c$, c *a constant, independent of* t, $\forall t \in \mathbb{Z}$.

(ii) $\gamma_X(r, s) = \gamma_X(r + h, s + h)$, $\forall r, s, h \in \mathbb{Z}$.

Remark: In view of condition (ii), if $(X_t)_{t\in\mathbb{Z}}$ is stationary then $\gamma_X(r, s) = \gamma_X(r - s, 0)$. Thus, whenever we refer to a *stationary* time series $(X_t)_{t\in\mathbb{Z}}$ we mean the function γ_X of *one* variable, defined by

$$\gamma_X(h) := \gamma_X(h, 0) = \gamma_X(t + h, t)$$

We say that the function $\gamma_X(\cdot)$ is the autocovariance function and $\gamma_X(h)$ is its value at *lag* h.

Definition 2.35 (Autocovariance- and autocorrelation function) *Let* $(X_t)_{t\in\mathbb{Z}}$ *be a stationary time series. The autocovariance function (ACVF) of* $(X_t)_{t\in\mathbb{Z}}$ *is*

$$\gamma_X(h) = Cov(X_{t+h}, X_t).$$

The autocorrelation function (ACF) of $(X_t)_{t\in\mathbb{Z}}$ *is*

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t). \tag{2.59}$$

Definition 2.36 (Strict Stationarity) *The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, ..., X_{t_k})^t$ and $(X_{t_1+h}, ..., X_{t_k+h})^t$ are the same for all positive integers k and for all $t_1, ..., t_k$, $h \in \mathbb{Z}$.*

A strict stationary process with finite second moments is stationary, however statement is not valid backwards. For instance, consider a sequence of independent random variables such that X_t is exponentially distributed with mean one when t is odd and normally distributed with mean one and variance one when t is even. Then the time series $\{X_t\}$ is weakly stationary with $\gamma_X(0) = 1$ and $\gamma_X(h) = 0$ for $h \neq 0$. However, $\{X_t\}$ cannot be strictly stationary because the distributions of X_1 and X_2 are completely different.

Definition 2.37 (IID Noise) *If $(X_t)_{t\in\mathbb{Z}}$ is IID noise and $E[X_t^2] = \sigma^2 < \infty$, then $\mu_X(t)$ is independent of t. By the assumed independence,*

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2 & if \quad h = 0 \\ 0 & if \quad h \neq 0 \end{cases}$$

which does not depend on t. Hence IID noise with finite second moment is stationary. We use the notation

$$(X_t)_{t\in\mathbb{Z}} \sim IID(0, \sigma^2)$$

to indicate that the random variables X_t are independent and identically distributed random variables, each with mean 0 and variance σ^2.

Definition 2.38 (White Noise) *If $(X_t)_{t\in\mathbb{Z}}$ is a sequence of uncorrelated random variables, each with zero mean and variance σ^2, then $(X_t)_{t\in\mathbb{Z}}$ is stationary with the same covariance function as the IID noise. Such a sequence is referred to as white noise with mean 0 and variance σ^2. This is indicated by the notation*

$$(X_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2).$$

Every $IID(0, \sigma^2)$ sequence is $WN(0, \sigma^2)$ but not conversely.

2.7.2 ARMA Processes

Definition 2.39 (ARMA(p,q)) *$(X_t)_{t\in\mathbb{Z}}$ is an ARMA(p,q) process if $(X_t)_{t\in\mathbb{Z}}$ is stationary and if for every t,*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{2.60}$$

where $(Z_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$ and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factors.

The process $(X_t)_{t\in\mathbb{Z}}$ is said to be an *ARMA(p,q) process with mean μ* if $(X_t - \mu)_{t\in\mathbb{Z}}$ is an ARMA(p,q) process. The equation (2.60) can be written as

$$\phi(B)X_t = \theta(B)Z_t,$$

where $\phi(\cdot)$ and $\theta(\cdot)$ are the pth and qth degree polynomials,

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \tag{2.61}$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

and B is the *backward shift operator* defined by

$$B^j X_t = X_{t-j}, B^j Z_t = Z_{t-j}, j = 0, \pm 1, \pm 2, ...$$

The polynomials ϕ and θ will be referred to as the *autoregressive* and *moving average polynomials*, respectively, of the difference equations (2.61).

Existence and Uniqueness

A stationary solution $(X_t)_{t\in\mathbb{Z}}$ of the equation (2.60) exists (and is also the unique stationary solution) if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0, \quad \forall |z| = 1$$

Proposition 2.40 *An ARMA(p,q) process $(X_t)_{t\in\mathbb{Z}}$ is causal, or a causal function of $(Z_t)_{t\in\mathbb{Z}}$, if there exist constants (ψ_j) such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \forall t.$$

Causality is equivalent to the condition

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad |z| \leqslant 1,$$

Proof: Brockwell and Davis (1996).

Example 2.41 (ARMA(1,1) Process) *The time series $(X_t)_{t\in\mathbb{Z}}$ is an ARMA(1,1) process if it is stationary and satisfies for every t,*

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \tag{2.62}$$

where $(Z_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$ and $\phi + \theta \neq \theta$.

Using the backward shift operator B, the equation (2.62) can be expressed more concisely as

$$\phi(B) X_t = \theta(B) Z_t \tag{2.63}$$

where $\phi(B)$ and $\theta(B)$ are the linear filters

$$\phi(B) = 1 - \phi B$$

and

$$\theta(B) = 1 + \theta B,$$

respectively.

2.7.3 GARCH Processes

According to Zivot (2003) the GARCH processes are defined as follows:

Definition 2.42 (GARCH(p,q)) *The generalized ARCH model, GARCH(p,q), which was developed independently by Bollerslev (1986) and Taylor (1986), has the form*

$$\begin{aligned} y_t &= c + \epsilon_t \\ \sigma_t^2 &= a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \end{aligned} \tag{2.64}$$

for $t = 1, ..., T$*, where* $\sigma_t^2 = Var_{t-1}(\epsilon_t)$*.*

For a GARCH(1,1) model we have the case as

$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2, \tag{2.65}$$

where σ_t^2 is known as the conditional variance since it is a one-period ahead estimate for the variance calculated based on any past relevant information. A GARCH(1,1) model with three parameters in the conditional variance equation is often adequate to obtain a good model fit for financial time series.

Chapter 3

Maximum Likelihood Estimation for Bivariate t-Data using Maximization by Parts (MbP)

In this chapter we will apply the maximization by parts algorithm by Song et al. (2005) to the bivariate t-distribution using the copula construction. In bivariate t-model we assume that the univariate t-marginals have the same degrees of freedom as the t-copula's. Thus we have two parameters, one for the degrees of freedom and one for the correlation. The definition of a bivariate copula, as well as the univariate and bivariate t-density function can be inferred from Chapter 2. Note that the bivariate t-distribution is a special case of bivariate meta t-distributions.

Let $\boldsymbol{\Theta} := \{\nu, \rho : \nu \in [1, \infty), \rho \in (-1, 1)\}$ denote the parameter space in the bivariate t-distribution. Let $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ be a bivariate t-data set, where $\boldsymbol{x}_1 = (x_{11}, ..., x_{n1})^t$ and $\boldsymbol{x}_2 = (x_{12}, ..., x_{n2})^t$ are independent observations and let $\boldsymbol{x}_i = (x_{i1}, x_{i2})^t$ be the ith observation of data set $\boldsymbol{x}$, for $i = 1, ..., n$.

The likelihood function of the bivariate t-distribution can be written as

$$L(\nu, \rho; \boldsymbol{x}) = \prod_{i=1}^{n} c(F(x_{i1}; \nu), F(x_{i2}; \nu); \nu, \rho) f(x_{i1}, \nu) f(x_{i2}, \nu) \tag{3.1}$$

and the corresponding log likelihood is

$$\ell(\nu, \rho; \boldsymbol{x}) := \ln L(\nu, \rho; \boldsymbol{x}) = \sum_{i=1}^{n} \ln c(F(x_{i1}; \nu), F(x_{i2}; \nu); \nu, \rho) + \sum_{i=1}^{n} \sum_{j=1}^{2} \ln f(x_{ij}; \nu). \tag{3.2}$$

As pointed out in the previous chapter, the crucial idea of the MbP algorithm introduced by Song et al. (2005) is the decomposition of the log likelihood function into an appropriate working model and an error model. Usually the working model needs to meet the following criteria:

- It contains less parameters to be estimated while the error model includes all parameters of the joint distribution.

- One can calculate its first order derivative easily.
- It must achieve the information dominance condition defined in Theorem 2.24.

3.1 Choices for the Working Model

3.1.1 Independence Working Model

One decomposition possibility of the log likelihood of bivariate t-distribution is employing the following working model consisting of two independent marginal t-distributions:

$$\ell_w(\nu; \boldsymbol{x}) := \ell_m(\nu; \boldsymbol{x}_1) + \ell_m(\nu; \boldsymbol{x}_2) \tag{3.3}$$

where $\ell_m(\nu; \boldsymbol{x}_j)$ is the log likelihood of the univariate t-distribution with degrees of freedom ν, i.e.

$$\ell_m(\nu; \boldsymbol{x}_j) := \sum_{i=1}^{n} \ln f(x_{ij}; \nu), \text{ for } j = 1, 2. \tag{3.4}$$

The error model is defined as the log likelihood of the bivariate t-copula with degrees of freedom ν and correlation coefficient ρ, in particular,

$$\ell_e(\nu, \rho; \boldsymbol{x}) := \sum_{i=1}^{n} \ln c(F(x_{i1}; \nu), F(x_{i2}; \nu); \nu, \rho). \tag{3.5}$$

Thus the sum of these two particular models yields the log likelihood of the bivariate t distribution, e.g.

$$\ell(\nu, \rho; \boldsymbol{x}) = \ell_w(\nu; \boldsymbol{x}) + \ell_e(\nu, \rho; \boldsymbol{x}).$$

The working model $\ell_w(\cdot; \cdot)$ defined in (3.3) is denoted as the *independence working model.* The error model in (3.5) is referred as the *copula error model.*

Since the independence working model is composed of two univariate t-marginals which possess only one common parameter ν, its first order derivative with respect to ν is effortless to calculate. However, the drawback of the independence working is that the model does not disclose any information about the correlation of the data from the error model. Thus, we anticipate that the algorithm will not converge if there is large correlation in the data. We can modify the independence working model in a way that the correlation of the error model is taken into consideration.

3.1.2 Alternative Working Model

Inspired by the idea given in Song et al. (2005), we add a new term denoted as $\phi_i(\boldsymbol{x}_i, \nu, \rho_0)$ into the independence working model ℓ_w in order to achieve the information dominance. The function $\phi_i(\boldsymbol{x}_i, \nu, \rho_0)$ is determined by a term that is proportional to the log likelihood of the bivariate t-distribution. Thus, $\phi_i(\boldsymbol{x}_i, \nu, \rho_0)$ is given by

$$\phi_i(\boldsymbol{x}_i; \nu, \rho_0) := |R_0|^{-\frac{1}{2}} \big(1 + \frac{\boldsymbol{x}_i^t R_0^{-1} \boldsymbol{x}_i}{\nu}\big)^{-\frac{\nu+2}{2}} \tag{3.6}$$

where $R_0 := \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ for a fixed ρ_0. The corresponding log likelihood is then given by

$$\ell_{new}(\nu, \rho_0; \boldsymbol{x}) := \ln(\prod_{i=1}^{n} \phi_i(\boldsymbol{x}_i; \nu, \rho_0)) = -\frac{1}{2}\ln(|R_0|) - \frac{\nu+2}{2}\sum_{i=1}^{n} \ln\big(1 + \frac{\boldsymbol{x}_i^t R_0^{-1} \boldsymbol{x}_i}{\nu}\big).$$

We calculate the value of ρ_0 using the empirical estimate of Kendall's τ correlation $\hat{\tau}$ defined in (2.17) in Chapter 2, i.e.

$$\hat{\rho}_\tau := \sin(\frac{\pi}{2}\hat{\tau}).$$

In the later section we will investigate the influence of ρ_0 on the models. With the additional term $\phi_i(\boldsymbol{x}_i; \nu, \rho_0)$ we can rewrite the bivariate t-density for the ith observation $\boldsymbol{x}_i$ as

$$f_i(\boldsymbol{x}_i; \nu, \rho; \rho_0) = f(x_{i1}; \nu) f(x_{i2}; \nu) \phi_i(\boldsymbol{x}_i; \nu, \rho_0) \frac{c(F(x_{i1}; \nu), F(x_{i2}; \nu); \nu, \rho)}{\phi_i(\boldsymbol{x}_i; \nu, \rho_0)}$$

and the log likelihood analogously as follows:

$$\ell(\nu, \rho; \rho_0; \boldsymbol{x}) = \underbrace{\ell_w(\nu; \boldsymbol{x}) + \ell_{new}(\nu, \rho_0; \boldsymbol{x})}_{\text{modified working model}} + \underbrace{\ell_e(\nu, \rho; \boldsymbol{x}) - \ell_{new}(\nu, \rho_0; \boldsymbol{x})}_{\text{modified error model}}.$$

with the new working model given by

$$\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) := \ell_w(\nu; \boldsymbol{x}) + \ell_{new}(\nu, \rho_0; \boldsymbol{x}) \tag{3.7}$$

and the error model

$$\ell_{e,mod}(\nu, \rho; \rho_0; \boldsymbol{x}) := \ell_e(\nu, \rho; \boldsymbol{x}) - \ell_{new}(\nu, \rho_0; \boldsymbol{x}). \tag{3.8}$$

The modified working model leads to an estimation procedure similar to that obtained with the independence working model and ν remains the unique parameter to be estimated in the working model. In the following MbP algorithm we only utilize the modified working and error models.

In order to process the MbP algorithm we calculate the score functions. For further analysis we also derive the second order derivative of both modified working and error models.

To examine the model more closely we rewrite the models. Since the additional model ℓ_{new} is the log likelihood of the bivariate t-distribution with fixed correlation coefficient ρ_0, we can also express the term as:

$$\ell_{new}(\nu, \rho_0; \boldsymbol{x}) = \ell_w(\nu; \boldsymbol{x}) + \ell_e(\nu, \rho_0; \boldsymbol{x}).$$

It follows that the score functions of the modified working and error models can be written as:

$$\begin{aligned}
\frac{\partial}{\partial\nu}\ell_{w,mod}(\nu, \rho_0; \boldsymbol{x}) &= \frac{\partial}{\partial\nu}\ell_w(\nu; \boldsymbol{x}) + \frac{\partial}{\partial\nu}\ell_{new}(\nu, \rho_0; \boldsymbol{x}) \\
&= \frac{\partial}{\partial\nu}\ell_w(\nu; \boldsymbol{x}) + \frac{\partial}{\partial\nu}\ell_w(\nu; \boldsymbol{x}) + \frac{\partial}{\partial\nu}\ell_e(\nu, \rho_0; \boldsymbol{x}) \\
&= 2 \cdot \frac{\partial}{\partial\nu}\ell_w(\nu; \boldsymbol{x}) + \frac{\partial}{\partial\nu}\ell_e(\nu, \rho_0; \boldsymbol{x}) \qquad (3.9)
\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial \nu}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) &= \frac{\partial}{\partial \nu}\ell_e(\nu,\rho;\boldsymbol{x}) - \frac{\partial}{\partial \nu}\ell_{new}(\nu,\rho_0;\boldsymbol{x}) \\ &= \frac{\partial}{\partial \nu}\ell_e(\nu,\rho;\boldsymbol{x}) - \frac{\partial}{\partial \nu}\ell_e(\nu,\rho_0;\boldsymbol{x}) - \frac{\partial}{\partial \nu}\ell_w(\nu;\boldsymbol{x}). \end{aligned} \tag{3.10}$$

Moreover,

$$\frac{\partial}{\partial \rho}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) = \frac{\partial}{\partial \rho}\ell_e(\nu,\rho;\boldsymbol{x}). \tag{3.11}$$

In the following section we calculate the derivative explicitly.

3.2 Log Likelihoods of the Working and Error Models and its Derivatives

3.2.1 Log Likelihoods of the Working and Error Models

To avoid unclarity in the following formulas we recall the some notations that can be found in chapter 2:

- $\Gamma(\cdot)$: the gamma function
- $\psi(\cdot)$: the digamma function
- $\psi^{(1)}(\cdot)$: the trigamma function
- $|\cdot|$: the absolute value of the determinant of a matrix.

During the calculation of the analytical expressions of the derivatives of the models we often encounter the following terms:
For $R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ we obtain $|R| = 1-\rho^2$ and $\boldsymbol{x}_i^t R^{-1} \boldsymbol{x}_i = (x_{i1}^2 + x_{i2}^2 - 2x_{i1}x_{i2}\rho)/(1-\rho^2)$.
These calculations are identical for matrix $R_0 = \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix}$.

1) Log likelihood of the Independence Working Model:

With

$$\begin{aligned}\ell_m(\nu;\boldsymbol{x}_j) &:= \sum_{i=1}^n \ln f(x_{ij};\nu) \quad \text{for } j=1,2 \\ &= \sum_{i=1}^n \ln\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x_{ij}^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right) \\ &= n\ln\Gamma(\frac{\nu+1}{2}) - \frac{n}{2}\ln(\nu\pi) - n\ln\Gamma(\frac{\nu}{2}) - \frac{\nu+1}{2}\sum_{i=1}^n \ln\left(1+\frac{x_{ij}^2}{\nu}\right)\end{aligned}$$

it follows that the log likelihood of the independence working model is given by

$$\begin{aligned}\ell_w(\nu;\boldsymbol{x}) &:= \ell_m(\nu;\boldsymbol{x}_1)+\ell_m(\nu;\boldsymbol{x}_2)\\ &= 2n\ln\Gamma(\frac{\nu+1}{2}) - 2n\ln\Gamma\big(\frac{\nu}{2}\big) - n\ln(\nu\pi)\\ &\quad -\frac{\nu+1}{2}\Bigg(\sum_{i=1}^{n}\ln\big(1+\frac{x_{i1}^2}{\nu}\big)+\sum_{i=1}^{n}\ln\big(1+\frac{x_{i2}^2}{\nu}\big)\Bigg).\end{aligned}$$

2) Log Likelihood of the (modified) Working Model

Since

$$\begin{aligned}\ell_{new}(\nu,\rho_0;\boldsymbol{x}) &:= \prod_{i=1}^{n}\ln\phi_i(\boldsymbol{x}_i;\nu)\\ &= \prod_{i=1}^{n}\ln\Bigg(-\frac{1}{2}\ln(|R_0|) - \frac{\nu+2}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\big)\Bigg)\\ &= -\frac{n}{2}\ln(1-\rho_0^2) - \frac{\nu+2}{2}\sum_{i=1}^{n}\ln\big(1+\frac{x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho_0}{\nu(1-\rho_0^2)}\big)\end{aligned}$$

we have that the log likelihood of the modified working model is given by

$$\begin{aligned}\ell_{w,mod}(\nu;\rho_0;\boldsymbol{x}) &= \ell_w(\nu;\boldsymbol{x})+\ell_{new}(\nu,\rho_0;\boldsymbol{x})\\ &= 2n\ln\Gamma(\frac{\nu+1}{2}) - 2n\ln\Gamma\big(\frac{\nu}{2}\big) - n\ln(\nu\pi)\\ &\quad -\frac{\nu+1}{2}\Bigg(\sum_{i=1}^{n}\ln\big(1+\frac{x_{i1}^2}{\nu}\big)+\sum_{i=1}^{n}\ln\big(1+\frac{x_{i2}^2}{\nu}\big)\Bigg)\\ &\quad -\frac{n}{2}\ln(1-\rho_0^2) - \frac{\nu+2}{2}\sum_{i=1}^{n}\ln\big(1+\frac{x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho_0}{\nu(1-\rho_0^2)}\big).\end{aligned}$$

3) Log Likelihood of the Copula Error Model

$$\begin{aligned}
\ell_e(\nu,\rho;\boldsymbol{x}) &:= \sum_{i=1}^{n} \ln c(F(x_{i1};\nu),F(x_{i2};\nu);\nu,\rho) \\
&= \sum_{i=1}^{n} \ln\left(\frac{\frac{\nu}{2}\cdot\Gamma(\frac{\nu}{2})^2\big(1+\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu}\big)^{-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu+1}{2})^2\sqrt{|R|}(1+\frac{x_{i1}^2}{\nu})^{-\frac{\nu+1}{2}}(1+\frac{x_{i2}^2}{\nu})^{-\frac{\nu+1}{2}}} \right) \\
&= 2n\ln\Gamma(\frac{\nu}{2}) + n\ln(\frac{\nu}{2}) - 2n\ln\Gamma\big(\frac{\nu+1}{2}\big) - \frac{n}{2}\ln(|R|) \\
&\quad - \frac{\nu+2}{2}\sum_{i=1}^{n}\ln(1+\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu}) + \frac{\nu+1}{2}\left(\sum_{i=1}^{n}\ln\big(1+\frac{x_{i1}^2}{\nu}\big) + \sum_{i=1}^{n}\ln(1+\frac{x_{i2}^2}{\nu})\right)
\end{aligned}$$

4) Log Likelihood of the (modified) Error Model

$$\begin{aligned}
\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) &= \ell_e(\nu,\rho;\boldsymbol{x}) - \ell_{new}(\nu,\rho_0;\boldsymbol{x}) \\
&= 2n\ln\Gamma(\frac{\nu}{2}) + n\ln(\frac{\nu}{2}) - 2n\ln\Gamma\big(\frac{\nu+1}{2}\big) - \frac{n}{2}\ln(|R|) \\
&\quad - \frac{\nu+2}{2}\sum_{i=1}^{n}\ln(1+\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu}) + \frac{\nu+1}{2}\left(\sum_{i=1}^{n}\ln\big(1+\frac{x_{i1}^2}{\nu}\big) + \sum_{i=1}^{n}\ln(1+\frac{x_{i2}^2}{\nu})\right) \\
&\quad + \frac{n}{2}\ln(|R_0|) + \frac{\nu+2}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\big)
\end{aligned}$$

3.2.2 Score Functions of the Working and Error Models

1) Score Function for the Independence Working Model

$$\begin{aligned}
\frac{\partial}{\partial\nu}\ell_m(\nu;\boldsymbol{x}_j) &:= \frac{\partial}{\partial\nu}\sum_{i=1}^{n}\ln f(x_{ij};\nu) \\
&= \frac{n}{2}\psi(\frac{\nu+1}{2}) - \frac{n}{2}\frac{1}{\nu\pi}\pi - \frac{n}{2}\psi(\frac{\nu}{2}) \\
&\quad - \left[\frac{1}{2}\left(\sum_{i=1}^{n}\ln(1+\frac{x_{ij}^2}{\nu})\right) + \frac{\nu+1}{2}\sum_{i=1}^{n}\frac{1}{1+\frac{x_{ij}^2}{\nu}}x_{ij}^2(-\frac{1}{\nu^2})\right] \\
&= \frac{n}{2}\psi(\frac{\nu+1}{2}) - \frac{n}{2\nu} - \frac{n}{2}\psi(\frac{\nu}{2}) - \frac{1}{2}\sum_{i=1}^{n}\ln(1+\frac{x_{ij}^2}{\nu}) + \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\frac{x_{ij}^2}{\nu+x_{ij}^2}
\end{aligned}$$

Thus, the score function of the independence working model is calculated as:

$$\begin{aligned}\frac{\partial}{\partial\nu}\ell_w(\nu;\boldsymbol{x}) &:= \frac{\partial}{\partial\nu}\ell_m(\nu;\boldsymbol{x}_1)+\frac{\partial}{\partial\nu}\ell_m(\nu;\boldsymbol{x}_2)\\ &= n\psi(\frac{\nu+1}{2})-\frac{n}{\nu}-n\psi(\frac{\nu}{2})-\frac{1}{2}\sum_{i=1}^{n}\left(\ln(1+\frac{x_{i1}^2}{\nu})+\ln(1+\frac{x_{i2}^2}{\nu})\right)\\ &\quad+\frac{\nu+1}{2\nu}\sum_{i=1}^{n}\left(\frac{x_{i1}^2}{\nu+x_{i1}^2}+\frac{x_{i2}^2}{\nu+x_{i2}^2}\right)\end{aligned}$$

2) Score Function for the Copula Error Model

The score function of the copula error model is calculated as:

$$\begin{aligned}\frac{\partial}{\partial\rho}\ell_e(\nu,\rho;\boldsymbol{x}) &:= -\frac{n}{2}\frac{\partial}{\partial\rho}\ln(1-\rho^2)-\frac{\nu+2}{2}\sum_{i=1}^{n}\frac{\partial}{\partial\rho}\ln\left(1+\frac{x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}{\nu(1-\rho^2)}\right)\\ &= -\frac{n}{2}\frac{1}{1-\rho^2}(-2\rho)-\frac{\nu+2}{2}\sum_{i=1}^{n}\frac{(-2x_{i1}x_{i2})\nu(1-\rho^2)-\nu(-2\rho)(x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)}{\left(1+\frac{x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}{\nu(1-\rho^2)}\right)\cdot\nu^2(1-\rho^2)^2}\\ &= \frac{n\rho}{1-\rho^2}-\frac{\nu+2}{1-\rho^2}\sum_{i=1}^{n}\frac{-x_{i1}x_{i2}(1-\rho^2)+x_{i1}^2\rho+x_{i2}^2\rho-2x_{i1}x_{i2}\rho^2}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}\\ &= \frac{n\rho}{1-\rho^2}-\frac{\nu+2}{1-\rho^2}\sum_{i=1}^{n}\frac{x_{i1}^2\rho+x_{i2}^2\rho-(1+\rho^2)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho} \qquad (3.12)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\nu}\ell_e(\nu,\rho;\boldsymbol{x}) &:= 2n\cdot\frac{1}{2}\psi(\frac{\nu}{2})-2n\frac{1}{2}\psi(\frac{\nu+1}{2})+n\frac{1}{\nu/2}\cdot\frac{1}{2}\\ &\quad-\left[\frac{1}{2}\left(\sum_{i=1}^{n}\ln(1+\frac{\boldsymbol{x}_i^tR^{-1}\boldsymbol{x}_i}{\nu})\right)+\frac{\nu+2}{2}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR^{-1}\boldsymbol{x}_i}{1+\frac{\boldsymbol{x}_i^tR^{-1}\boldsymbol{x}_i}{\nu}}(-\frac{1}{\nu^2})\right]\\ &\quad+\left[\frac{1}{2}\sum_{i=1}^{n}\left(\ln(1+\frac{x_{i1}^2}{\nu})+\ln(1+\frac{x_{i2}^2}{\nu})\right)+\frac{\nu+1}{2}\sum_{i=1}^{n}\left(\frac{(-\frac{1}{\nu^2})x_{i1}^2}{1+\frac{x_{i1}^2}{\nu}}+\frac{(-\frac{1}{\nu^2})x_{i2}^2}{1+\frac{x_{i2}^2}{\nu}}\right)\right]\\ &= \frac{n}{\nu}+n\psi(\frac{\nu}{2})-n\psi(\frac{\nu+1}{2})-\frac{1}{2}\sum_{i=1}^{n}\ln(1+\frac{\boldsymbol{x}_i^tR^{-1}\boldsymbol{x}_i}{\nu})\\ &\quad+\frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^tR^{-1}\boldsymbol{x}_i}+\frac{1}{2}\left(\sum_{i=1}^{n}\ln(1+\frac{x_{i1}}{\nu})+\ln(1+\frac{x_{i2}}{\nu})\right)\\ &\quad-\frac{\nu+1}{2\nu}\left(\sum_{i=1}^{n}\frac{x_{i1}^2}{\nu+x_{i1}^2}+\frac{x_{i2}^2}{\nu+x_{i2}^2}\right)\end{aligned}$$

3) Score Function of the (modified) Working and Error Models

With

$$\begin{aligned}
\frac{\partial}{\partial\nu}\ell_{new}(\nu,\rho_0;\boldsymbol{x}) &:= \frac{\partial}{\partial\nu}\left(-\frac{n}{2}\ln(1-\rho_0^2) - \frac{\nu+2}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\big)\right)\\
&= -\frac{1}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\big) - \frac{\nu+2}{2}\sum_{i=1}^{n}\frac{1}{1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}}\big(-\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu^2}\big)\\
&= -\frac{1}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\big) + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}
\end{aligned}$$

the score function of the modified working model is given by

$$\begin{aligned}
\frac{\partial}{\partial\nu}\ell_{w,mod}(\nu;\rho_0;\boldsymbol{x}) &= \frac{\partial}{\partial\nu}\ell_w(\nu;\boldsymbol{x}) + \frac{\partial}{\partial\nu}\ell_{new}(\nu,\rho_0;\boldsymbol{x})\\
&= n\psi(\frac{\nu+1}{2}) - \frac{n}{\nu} - n\psi(\frac{\nu}{2}) - \frac{1}{2}\sum_{i=1}^{n}\left(\ln(1+\frac{x_{i1}^2}{\nu}) + \ln(1+\frac{x_{i2}^2}{\nu})\right)\\
&\quad + \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\left(\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}\right)\\
&\quad - \frac{1}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\big) + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i},
\end{aligned} \tag{3.13}$$

while the score functions of the modified error model are given by

$$\begin{aligned}
\frac{\partial}{\partial\nu}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) &= \frac{\partial}{\partial\nu}\ell_{e(1)}(\nu,\rho;\boldsymbol{x}) - \frac{\partial}{\partial\nu}\ell_{new}(\nu,\rho_0;\boldsymbol{x})\\
&= \frac{n}{\nu} + n\psi(\frac{\nu}{2}) - n\psi(\frac{\nu+1}{2}) - \frac{1}{2}\sum_{i=1}^{n}\ln(1+\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu})\\
&\quad + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i} + \frac{1}{2}\left(\sum_{i=1}^{n}\ln(1+\frac{x_{i1}}{\nu}) + \ln(1+\frac{x_{i2}}{\nu})\right)\\
&\quad - \frac{\nu+1}{2\nu}\left(\sum_{i=1}^{n}\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}\right)\\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\ln\big(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\big) - \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}.
\end{aligned} \tag{3.14}$$

$$\frac{\partial}{\partial\rho}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) = \frac{\partial}{\partial\rho}\ell_e(\nu,\rho;\boldsymbol{x})$$

3.2.3 Second Order Derivative of the (modified) Working and Error Models

1) Second Order Derivative of the Independence Working Model

The second order derivative of the independence working model with respect to ν is given by:

$$\begin{aligned}
\frac{\partial^2}{\partial\nu^2}\ell_m(\nu;\boldsymbol{x}_j) &= \frac{\partial}{\partial\nu}(\frac{\partial}{\partial\nu}\ell_m(\nu;\boldsymbol{x}_j)) \\
&= \frac{n}{2}\frac{1}{2}\psi^{(1)}(\frac{\nu+1}{2}) - \frac{n}{2}(-\frac{1}{\nu^2}) - \frac{n}{2}\frac{1}{2}\psi^{(1)}(\frac{\nu}{2}) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{1}{1+\frac{x_{ij}^2}{\nu}}\cdot\frac{-x_{ij}^2}{\nu^2}\right) \\
&\quad + \left[\left(\frac{\nu+1}{2\nu}\right)'\sum_{i=1}^{n}\frac{x_{ij}^2}{\nu+x_{ij}^2} + \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\left(\frac{x_{ij}^2}{\nu+x_{ij}^2}\right)'\right] \\
&= \frac{n}{4}\psi^{(1)}(\frac{\nu+1}{2}) + \frac{n}{2\nu^2} - \frac{n}{4}\psi^{(1)}(\frac{\nu}{2}) + \frac{1}{2\nu}\sum_{i=1}^{n}\frac{x_{ij}^2}{\nu+x_{ij}^2} \\
&\quad + \left[-\frac{1}{2\nu^2}\sum_{i=1}^{n}\frac{x_{ij}^2}{\nu+x_{ij}^2} + \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\frac{-x_{ij}^2}{(\nu+x_{ij}^2)^2}\right] \\
&= \frac{n}{4}\psi^{(1)}(\frac{\nu+1}{2}) + \frac{n}{2\nu^2} - \frac{n}{4}\psi^{(1)}(\frac{\nu}{2}) + \frac{1}{2\nu}(1-\frac{1}{\nu})\sum_{i=1}^{n}\frac{x_{ij}^2}{\nu+x_{ij}^2} - \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\frac{x_{ij}^2}{(\nu+x_{ij}^2)^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\nu^2}\ell_w(\nu,\boldsymbol{x}) &= \frac{\partial}{\partial\nu}(\frac{\partial}{\partial\nu}\ell_m(\nu;\boldsymbol{x}_1)) + \frac{\partial}{\partial\nu}(\frac{\partial}{\partial\nu}\ell_m(\nu;\boldsymbol{x}_2)) \\
&= \frac{n}{2}\psi^{(1)}(\frac{\nu+1}{2}) + \frac{n}{\nu^2} - \frac{n}{2}\psi^{(1)}(\frac{\nu}{2}) + \frac{1}{2\nu}(1-\frac{1}{\nu})\sum_{i=1}^{n}\left(\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}\right) \\
&\quad - \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\left(\frac{x_{i1}^2}{(\nu+x_{i1}^2)^2} + \frac{x_{i2}^2}{(\nu+x_{i2}^2)^2}\right)
\end{aligned}$$

2) Second Order Derivative of the Copula Error Models

The second order derivative of the copula error model with respect to ν is given by:

$$
\begin{aligned}
\frac{\partial^2}{\partial\nu^2}\ell_e(\nu,\rho,\boldsymbol{x}) &= \frac{n}{2}\psi^{(1)}(\frac{\nu}{2}) - \frac{n}{2}\psi^{(1)}(\frac{\nu+1}{2}) - \frac{n}{\nu^2} \\
&- \left[\frac{1}{2}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i(-\frac{1}{\nu^2})}{1+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i/\nu} - \left(\frac{-1}{\nu^2}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i} + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{-\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i)^2}\right)\right] \\
&+ \frac{1}{2}\sum_{i=1}^{n}(\frac{-\frac{x_{i1}^2}{\nu^2}}{1+\frac{x_{i1}^2}{\nu}} + \frac{-\frac{x_{i2}^2}{\nu^2}}{1+\frac{x_{i2}^2}{\nu}}) - (-\frac{1}{2\nu^2})\sum_{i=1}^{n}(\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}) \\
&- \frac{\nu+1}{2\nu}\sum_{i=1}^{n}(\frac{-x_{i1}^2}{(\nu+x_{i2}^2)^2} + \frac{-x_{i2}^2}{(\nu+x_{i2}^2)^2}) \\
&= \frac{n}{2}\psi^{(1)}(\frac{\nu}{2}) - \frac{n}{2}\psi^{(1)}(\frac{\nu+1}{2}) - \frac{n}{\nu^2} + \frac{1}{\nu}(\frac{1}{2}-\frac{1}{\nu})\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i} \\
&- \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i)^2} - \frac{1}{2\nu}(1-\frac{1}{\nu})\sum_{i=1}^{n}(\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}) \\
&+ \frac{\nu+1}{2\nu}\sum_{i=1}^{n}(\frac{x_{i1}^2}{(\nu+x_{i1}^2)^2} + \frac{x_{i2}^2}{(\nu+x_{i2}^2)^2})
\end{aligned}
$$

The second order derivative of the copula error model with respect to ρ is calculated as

$$
\begin{aligned}
\frac{\partial^2}{\partial\rho^2}\ell_e(\nu,\rho,\boldsymbol{x}) &= \frac{\partial}{\partial\rho}\left(\frac{n\rho}{1-\rho^2} - \frac{\nu+2}{1-\rho^2}\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (1+\rho^2)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}\right) \\
&= \frac{n(1-\rho^2)-n\rho(-2\rho)}{(1-\rho^2)^2} - \left(\frac{-(\nu+2)(-2\rho)}{(1-\rho^2)^2}\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (1+\rho^2)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}\right) \\
&- \frac{\nu+2}{1-\rho^2}\sum_{i=1}^{n}\frac{(x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)(\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)}{(\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)^2} \\
&\qquad - \frac{(x_{i1}^2\rho + x_{i2}^2\rho - (\rho^2+1)x_{i1}x_{i2})(-2\nu\rho-2x_{i1}x_{i2})}{(\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)^2} \\
&= \frac{n}{\rho^2-1} - \frac{2\rho(\nu+2)}{(1-\rho^2)^2}\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (1+\rho^2)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho} \\
&- \frac{\nu+2}{1-\rho^2}\sum_{i=1}^{n}\frac{(x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)\left(\nu(1-\rho^2)+1\right)}{\left(\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho\right)^2} \\
&\qquad + \frac{2\left(x_{i1}^2\rho + x_{i2}^2\rho - (\rho^2+1)x_{i1}x_{i2}\right)\left(\nu\rho + x_{i1}x_{i2}\right)}{\left(\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho\right)^2}.
\end{aligned}
$$

The second order derivatives of the copula error model with respect to ρ and ν are expressed as

$$\begin{aligned}
\frac{\partial^2}{\partial\nu\partial\rho}\ell_e(\nu,\rho;\boldsymbol{x}) &= \frac{\partial}{\partial\nu}\left(\frac{n\rho}{1-\rho^2} - \frac{\nu+2}{1-\rho^2}\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (1+\rho^2)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}\right)\\
&= -\frac{1}{1-\rho^2}\left(\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (1+\rho^2)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}\right)\\
&\quad + (-\frac{\nu+2}{1-\rho^2})\sum_{i=1}^{n}\frac{-(x_{i1}^2\rho + x_{i2}^2\rho - (\rho^2+1)x_{i1}x_{i2})(1-\rho^2)}{(\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)^2}\\
&= -\frac{1}{1-\rho^2}\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (\rho^2+1)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}\\
&\quad + (\nu+2)\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (\rho^2+1)x_{i1}x_{i2}}{(\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho)^2}
\end{aligned}$$

and $\frac{\partial^2}{\partial\nu\partial\rho}\ell_e(\nu,\rho;\boldsymbol{x}) = \frac{\partial^2}{\partial\rho\partial\nu}\ell_e(\nu,\rho;\boldsymbol{x})$.

3) Second order derivative of (modified) working and error models:

Since

$$\begin{aligned}
\frac{\partial^2}{\partial\nu^2}\ell_{new}(\nu,\rho_0;\boldsymbol{x}) &= \frac{\partial}{\partial\nu}\left(-\frac{1}{2}\sum_{i=1}^{n}\ln\left(1+\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu}\right) + \frac{\nu+2}{2}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}\right)\\
&= -\frac{1}{2}\sum_{i=1}^{n}\frac{-\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu^2}}{1+\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu}} + (-\frac{2\nu-2(\nu+2)}{(2\nu)^2})\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}\\
&\quad + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{-\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i)^2}\\
&= \frac{1}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i} + (-\frac{1}{\nu^2})\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}\\
&\quad - \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i)^2}\\
&= \frac{1}{\nu}(\frac{1}{2}-\frac{1}{\nu})\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i} - \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^tR_0^{-1}\boldsymbol{x}_i)^2},
\end{aligned}$$

it follows that

$$
\begin{aligned}
\frac{\partial^2}{\partial\nu^2}\ell_{w,mod}(\nu;\rho_0;\boldsymbol{x}) &= \frac{\partial^2}{\partial\nu^2}\ell_w(\nu;\boldsymbol{x}) + \frac{\partial^2}{\partial\nu^2}\ell_{new}(\nu,\rho_0;\boldsymbol{x}) \\
&= \frac{n}{2}\psi^{(1)}(\frac{\nu+1}{2}) + \frac{n}{\nu^2} - \frac{n}{2}\psi^{(1)}(\frac{\nu}{2}) + \frac{1}{2\nu}(1-\frac{1}{\nu})\sum_{i=1}^{n}\left(\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}\right) \\
&\quad - \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\left(\frac{x_{i1}^2}{(\nu+x_{i1}^2)^2} + \frac{x_{i2}^2}{(\nu+x_{i2}^2)^2}\right) \\
&\quad + \frac{1}{\nu}(\frac{1}{2}-\frac{1}{\nu})\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i} - \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i)^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\nu^2}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) &= \frac{\partial^2}{\partial\nu^2}\ell_e(\nu,\rho;\boldsymbol{x}) - \frac{\partial^2}{\partial\nu^2}\ell_{new}(\nu,\rho_0;\boldsymbol{x}) \\
&= \frac{n}{2}\psi^{(1)}(\frac{\nu}{2}) - \frac{n}{2}\psi^{(1)}(\frac{\nu+1}{2}) - \frac{n}{\nu^2} + \frac{1}{\nu}(\frac{1}{2}-\frac{1}{\nu})\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i} \\
&\quad - \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i)^2} - \frac{1}{2\nu}(1-\frac{1}{\nu})\sum_{i=1}^{n}(\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}) \\
&\quad + \frac{\nu+1}{2\nu}\sum_{i=1}^{n}(\frac{x_{i1}^2}{(\nu+x_{i1}^2)^2} + \frac{x_{i2}^2}{(\nu+x_{i2}^2)^2}) \\
&\quad - \frac{1}{\nu}(\frac{1}{2}-\frac{1}{\nu})\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i} + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{(\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i)^2}
\end{aligned}
$$

$$
\frac{\partial^2}{\partial\rho\partial\nu}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) = \frac{\partial^2}{\partial\rho\partial\nu}\ell_e(\nu,\rho;\boldsymbol{x})
$$

and

$$
\frac{\partial^2}{\partial\rho^2}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) = \frac{\partial^2}{\partial\rho^2}\ell_e(\nu,\rho;\boldsymbol{x})
$$

Figures 3.1, 3.2 and 3.3 are based on a randomly generated data set from a bivariate t-distribution with parameters $\nu = 10$, $\rho = 0.6$ and sample size $n = 100$. The parameter $\rho_0 = \hat{\rho}_\tau \approx 0.70643$.

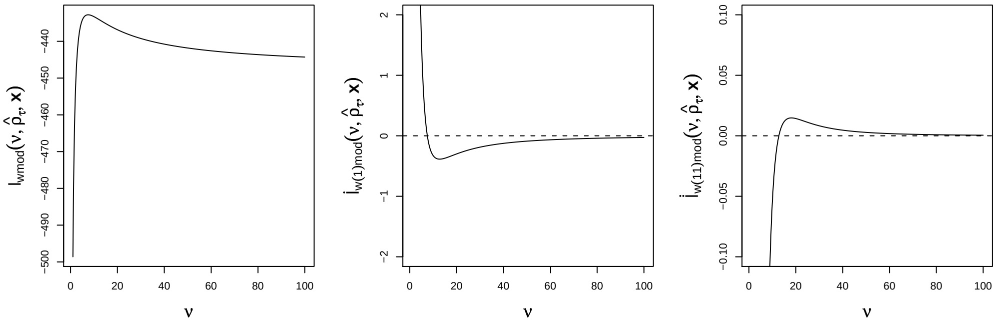


Figure 3.1: Modified working model and its derivatives with respect to ν when ρ is fixed to $\hat{\rho}_\tau \approx 0.70643$. ($\ell_{w,mod}$ (left), $\dot{\ell}_{w(1),mod} := \frac{\partial}{\partial \nu}\ell_{w,mod}$ (middle) and $\ddot{\ell}_{w(11),mod} := \frac{\partial^2}{\partial \nu^2}\ell_{w,mod}$ (right)).

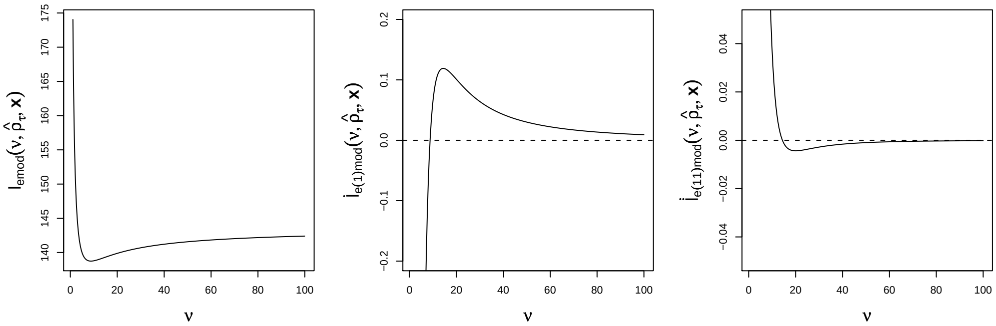


Figure 3.2: Modified error model and its derivatives with respect to ν when ρ is fixed to $\hat{\rho}_\tau \approx 0.70643$. ($\ell_{e,mod}$ (left), $\dot{\ell}_{e(1),mod} := \frac{\partial}{\partial \nu}\ell_{e,mod}$ (middle) and $\ddot{\ell}_{e(11),mod} := \frac{\partial^2}{\partial \nu^2}\ell_{e,mod}$ (right)).

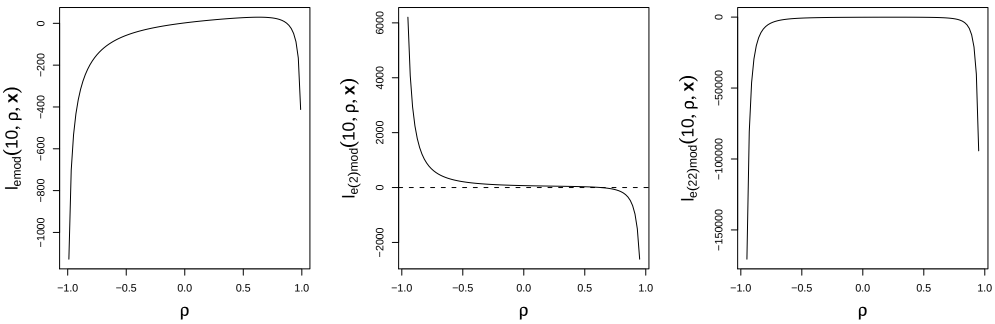


Figure 3.3: Modified error model and its derivatives with respect to ρ when ν is fixed to 10. ($\ell_{e,mod}$ (left), $\dot{\ell}_{e(2),mod} := \frac{\partial}{\partial \rho}\ell_e$ (middle) and $\ddot{\ell}_{e(22),mod} := \frac{\partial^2}{\partial \rho^2}\ell_e$ (right)).

3.3 The Choice of Interchangeable Parameter ρ_0

In this section we want to investigate the influence of the interchangeable parameter ρ_0 in the working and error models.

As specified before, the MbP algorithm applied to the bivariate t model requires to find the root of one-dimensional score equations. However, it is not guaranteed that a solution has to exist. If there does not exist a solution, the algorithm is said to be failed to converge. It implies as well that the information dominance condition stated in Theorem 2.24 is failed to meet.

Originally we had the idea to set the interchangeable parameter ρ_0 equal to the estimated correlation coefficient using estimated Kendall's τ, i.e., $\hat{\rho}_\tau = \sin(\frac{\pi}{2}\hat{\tau})$, since Kendall's τ correlation is demonstrated as a solid estimate for the rank correlation coefficient and by exploring the equality with the linear correlation we obtain a robust estimator for the linear correlation as well.

In order to achieve the information dominance we change the values of ρ_0. In the following illustration we can observe that the score function of the modified working model $\partial/\partial\nu(\ell_{w,mod}(\nu;\rho_0;\boldsymbol{x}))$ does not have a null point when $\rho_0 = \hat{\rho}_\tau$. However as the parameter ρ_0 increases by a certain degree, the curve twists in a way that a null point emerges.

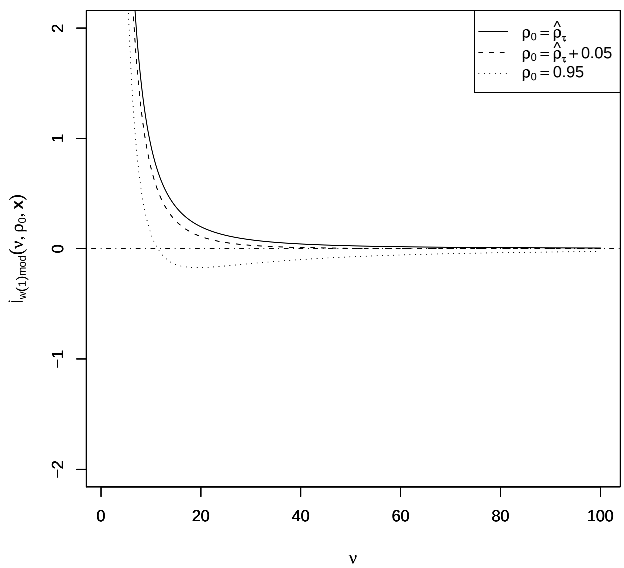


Figure 3.4: Graphic description of the influence of ρ_0. Data $\boldsymbol{X}$ is a random sample from bivariate t distribution with true parameters $\nu = 10$ and $\rho = 0.9$. Sample size is $n = 100$. $\hat{\rho}_\tau \approx 0.8615$

Thus the choice of the parameter ρ_0 can be crucial for the convergence of the MbP algorithm, since it helps to achieve a solution in the first Step of the algorithm. According to in Theorem 2.22, under the condition that the modified working and error models are twice differentiable, if the solution $\boldsymbol{\theta}_n^1$ of step 1 of the MbP algorithm is consistent, then the solution $\boldsymbol{\theta}_n^2$ of step 2 is consistent. However, a solution in 1st step does not guarantee the convergence of the algorithm. Information dominance may disappear after several MbP-steps.

3.4 The Framework of the MbP Algorithm for Bivariate t-Distribution

We use the 1-dimensional bisection algorithm described in Chapter 2 to find the root of score functions in each MbP step. With the expressions of the score functions stated in (3.9), (3.10) and (3.11) the framework of the MbP algorithm process as follows. The implemented MbP algorithm will be presented in the next chapter.

STEP 0. Generating initial values
Via the 1-dimensional bisection algorithm we find the root ν_n^1 of the equation

$$p_0(\nu) := \frac{\partial}{\partial \nu}\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) = 2 \cdot \frac{\partial}{\partial \nu}\ell_w(\nu; \boldsymbol{x}) + \frac{\partial}{\partial \nu}\ell_e(\nu, \rho_0; \boldsymbol{x}) = 0$$

which is required to update the score function of the error model, specifically, we need to solve

$$q_0(\rho) := \frac{\partial}{\partial \rho}\ell_e(\nu_n^1, \rho; \boldsymbol{x}) = 0$$

for ρ. Let ρ_n^1 denote the solution. Overall the solution of the first step is ν_n^1 and ρ_n^1. Actually this step generates the initial values for further MbP steps.

STEP k. k=1,2,... (MbP steps)
Let ν_n^k the solution to

$$p_k(\nu) := \frac{\partial}{\partial \nu}\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) + \frac{\partial}{\partial \nu}\ell_{e,mod}(\nu_n^{k-1}, \rho_n^{k-1}; \rho_0; \boldsymbol{x}) = 0 \tag{3.15}$$

and ρ_n^k the solution to

$$q_k(\rho) := \frac{\partial}{\partial \rho}\ell_e(\nu_n^{k-1}, \rho; \boldsymbol{x}) = 0 \tag{3.16}$$

for $k = 1, 2, ...$ In the kth step, we apply again the bisection algorithm to solve both equations (3.15) and (3.16) in ν and ρ respectively.

The function $\frac{\partial}{\partial \nu}\ell_{e,mod}(\nu_n^k, \rho_n^k; \rho_0; \boldsymbol{x})$ in (3.15) is a constant that depends on the updated values of ν_n^k, ρ_n^k and the fixed parameter ρ_0. It implies that in the parameter range [1,100] for ν the updated equation (3.15) may have two roots since the function $\frac{\partial}{\partial \nu}\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x})$ is not strictly monotone. Thus the bisection algorithm would have problems to find the correct root, if it searches in the range [1,100]. Figure (3.5) demonstrates the two null points of the updated equation in (3.15).

One method to solve the problem is to restrict the parameter range of ν $[1, 100]$ to a smaller range $[1, \tilde{\nu}]$ from Step 2 on. The quantity $\tilde{\nu}$ is the approximate minimum of the score function of the modified working model $\frac{\partial}{\partial \nu}\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x})$ in the parameter space $[1, 100]$. We can use the R-function `optim` to find the minimum. The value $\tilde{\nu}$ is set to the upper limit of the search range of the bisection algorithm for the parameter ν in MbP step k for $k = 2, 3, ...$

Illustration of STEP k, k=0,1.

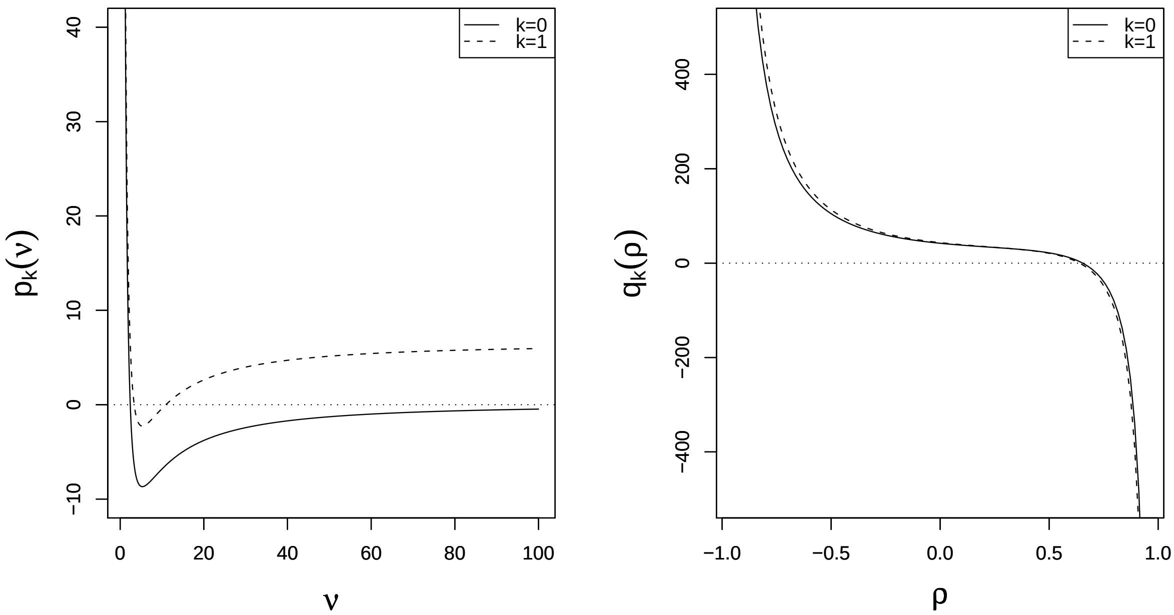


Figure 3.5: Illustration of step k, k=0,1, of the MbP algorithm (left panel: $p_k(\nu)$, right panel: $q_k(\rho)$). Data $\boldsymbol{x}$ is a bivariate t-distributed sample set with true parameters $\nu = 10$ and $\rho = 0.6$. Sample size is $n = 100$. $\rho_0 := 0.95$. The solution for $\nu_n^{k=1}$ is 2.17 and for $\rho_n^{k=1}$ is 0.64

Chapter 4

Small Sample Properties of the MbP Algorithm and its Competitors for Bivariate t-Data

4.1 Data and Likelihood

Before we start with the simulation study we need to produce random t-distributed data for bivariate t-distribution and uniformly distributed data for the bivariate t-copulas.

4.1.1 Generation of bivariate t-Data

To generate independent identically distributed (i.i.d.) bivariate t-data of size n we utilize the R-function `rmvt` which randomly generates multivariate t-distributed data of size n, i.e.

$$\boldsymbol{X} := (\boldsymbol{X}_1, \boldsymbol{X}_2) \sim \texttt{rmvt}(n, R, \texttt{df} = \nu), \tag{4.1}$$

where $R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ with $\boldsymbol{x}_j := (x_{1j}, ..., x_{nj})^t$, $j = 1, 2$, is then bivariate t-distributed with degrees of freedom ν and correlation ρ.

4.1.2 Generation of uniform data from bivariate t-Data

If we have a set bivariate t-data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ and know its true degrees, we can simply use the probability-integral transformation defined in (2.3) to generate the uniformly distributed data, i.e. $(\boldsymbol{u}_1, \boldsymbol{u}_2) := (F(\boldsymbol{x}_1; \nu), F(\boldsymbol{x}_2; \nu)) \in [0, 1]^{n\times 2}$ with the univariate t-distribution function $F(\cdot; \nu)$. However, this parameter is needed to be estimated and for the transformation we can use the concept of the *marginal empirical probability integral transformation.* This transformation approximates the *unknown* parametric marginals $F(\cdot)$ with the empirical cumulative distribution functions $\hat{F}(\cdot)$ defined as follows:

$$\hat{F}(t) := \frac{1}{n+1} \sum_{i=1}^{n} 1_{(-\infty \leqslant x_i]}(t), \quad \text{for } \boldsymbol{x} := (x_1, ..., x_n) \tag{4.2}$$

where $1_{(-\infty \leqslant x_i]}(t)$ represents the indicator function with value 1 if $t \leqslant x_i$ and with value 0 otherwise. The approximate i, jth element, $j = 1, 2$, of the bivariate uniform data $\boldsymbol{u}$ is defined by:

$$u_{ij} := \hat{F}(x_{ij}) \quad \text{with } \boldsymbol{x}_j := (x_{1j}, ..., x_{nj})^t, \text{ for each } i = 1, ..., n$$

Then we have the approximate bivariate uniform data $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2)$ with $\boldsymbol{u}_j := (u_{1j}, ..., u_{nj})^t$, $j = 1, 2$.

4.1.3 The Log Likelihoods and Score Functions

Consider the bivariate i.i.d. t-data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ and the parameter space $\Theta := \{\nu, \rho : \nu \in [1, 100], \rho \in (-1, 1)\}$. Given $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2) \in [0, 1]^{n\times 2}$ the log likelihood of the t-copula is given by

$$\begin{aligned}
\ell_c(\nu, \rho; \boldsymbol{u}) &:= \sum_{i=1}^{n} \ln c(\boldsymbol{u}_1, \boldsymbol{u}_2; \nu, \rho) \\
&= \sum_{i=1}^{n} \ln \left(\frac{\frac{\nu}{2} \cdot \Gamma(\frac{\nu}{2})^2 \big(1 + \frac{\boldsymbol{x}_i^t R^{-1} \boldsymbol{x}_i}{\nu}\big)^{-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu+1}{2})^2 \sqrt{|R|} (1 + \frac{x_{i1}^2}{\nu})^{-\frac{\nu+1}{2}} (1 + \frac{x_{i2}^2}{\nu})^{-\frac{\nu+1}{2}}} \right) \\
&= 2n \ln \Gamma(\frac{\nu}{2}) + n \ln(\frac{\nu}{2}) - 2n \ln \Gamma\big(\frac{\nu+1}{2}\big) - \frac{n}{2} \ln(|R|) \\
&\quad - \frac{\nu+2}{2} \sum_{i=1}^{n} \ln(1 + \frac{\boldsymbol{x}_i^t R^{-1} \boldsymbol{x}_i}{\nu}) + \frac{\nu+1}{2} \left(\sum_{i=1}^{n} \ln\big(1 + \frac{x_{i1}^2}{\nu}\big) + \sum_{i=1}^{n} \ln(1 + \frac{x_{i2}^2}{\nu}) \right)
\end{aligned} \tag{4.3}$$

whereby $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) := (t_\nu^{-1}(\boldsymbol{u}_1), t_\nu^{-1}(\boldsymbol{u}_2)) \in \mathbb{R}^{n\times 2}$ and $t_\nu^{-1}(\cdot)$ denotes the quantile function of Student's t-distribution. Furthermore, given the bivariate t-data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ the log likelihood of the bivariate t-distribution is defined by

$$\ell(\nu, \rho; \boldsymbol{x}) := n \ln(\frac{\nu}{2}) - \frac{n}{2} \ln(1 - \rho^2) - \frac{\nu+2}{2} \sum_{i=1}^{n} \ln(1 + \frac{\boldsymbol{x}_i^t R^{-1} \boldsymbol{x}_i}{\nu}). \tag{4.4}$$

In the later section we will estimate the parameters ν and ρ by maximizing the log likelihoods $\ell_c(\nu, \rho; \boldsymbol{\mu})$ and $\ell(\nu, \rho; \boldsymbol{x})$ with respect to both parameters directly. However, for the MbP algorithm we require the following score functions of the modified working and error models which are calculated in (3.13), (3.12) and (3.14) in Section 3.2 of Chapter 3. Let $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ be the i.i.d. bivariate t-data,

$$\begin{aligned}
\frac{\partial}{\partial \nu} \ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) &:= n\psi(\frac{\nu+1}{2}) - \frac{n}{\nu} - n\psi(\frac{\nu}{2}) - \frac{1}{2} \sum_{i=1}^{n} \left(\ln(1 + \frac{x_{i1}^2}{\nu}) + \ln(1 + \frac{x_{i2}^2}{\nu}) \right) \\
&\quad + \frac{\nu+1}{2\nu} \sum_{i=1}^{n} \left(\frac{x_{i1}^2}{\nu + x_{i1}^2} + \frac{x_{i2}^2}{\nu + x_{i2}^2} \right) \\
&\quad - \frac{1}{2} \sum_{i=1}^{n} \ln\big(1 + \frac{\boldsymbol{x}_i^t R^{-1} \boldsymbol{x}_i}{\nu}\big) + \frac{\nu+2}{2\nu} \sum_{i=1}^{n} \frac{\boldsymbol{x}_i^t R_0^{-1} \boldsymbol{x}_i}{\nu + \boldsymbol{x}_i^t R_0^{-1} \boldsymbol{x}_i},
\end{aligned}$$

$$\frac{\partial}{\partial \rho}\ell_e(\nu,\rho;\boldsymbol{x}) := \frac{n\rho}{1-\rho^2} - \frac{\nu+2}{1-\rho^2}\sum_{i=1}^{n}\frac{x_{i1}^2\rho + x_{i2}^2\rho - (1+\rho^2)x_{i1}x_{i2}}{\nu(1-\rho^2)+x_{i1}^2+x_{i2}^2-2x_{i1}x_{i2}\rho}$$

and

$$\begin{aligned}\frac{\partial}{\partial \nu}\ell_{e,mod}(\nu,\rho;\rho_0;\boldsymbol{x}) := {} & \frac{n}{\nu} + n\psi(\frac{\nu}{2}) - n\psi(\frac{\nu+1}{2}) - \frac{1}{2}\sum_{i=1}^{n}\ln(1+\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu}) \\ & + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R^{-1}\boldsymbol{x}_i} + \frac{1}{2}\left(\sum_{i=1}^{n}\ln(1+\frac{x_{i1}}{\nu}) + \ln(1+\frac{x_{i2}}{\nu})\right) \\ & - \frac{\nu+1}{2\nu}\left(\sum_{i=1}^{n}\frac{x_{i1}^2}{\nu+x_{i1}^2} + \frac{x_{i2}^2}{\nu+x_{i2}^2}\right) \\ & + \frac{1}{2}\sum_{i=1}^{n}\ln\left(1+\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu}\right) - \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i}{\nu+\boldsymbol{x}_i^t R_0^{-1}\boldsymbol{x}_i},\end{aligned}$$

whereby $R_0 := \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix}$. In each MbP step k, $k = 2, 3, ...$, the bisection algorithm searches for a solution of the function

$$\frac{\partial}{\partial \nu}\ell_{w,mod}(\nu;\rho_0;\boldsymbol{x}) + \frac{\partial}{\partial \nu}\ell_{e,mod}(\nu_n^{k-1},\rho_n^{k-1};\rho_0;\boldsymbol{x}) = 0$$

and

$$\frac{\partial}{\partial \rho}\ell_e(\nu_n^{k-1},\rho;\boldsymbol{x}) = 0.$$

as stated in the equations (3.15) and (3.16) in Chapter 3.

4.2 General Settings for the MbP Algorithm in the Simulation Study

The settings for the MbP algorithm utilized in the simulation study in this chapter are stated as:

- Given i.i.d. bivariate t-data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$, which is randomly generated using the R-function `rmvt` defined in (4.1).
- Parameter space Θ: $\nu \in [1, 100]$ and $\rho \in (-1, 1)$.
- Given ρ_0. The fixed parameter ρ_0 can be chosen as $\hat{\rho}_\tau := \sin(\frac{\pi}{2}\hat{\tau})$ or other values in the range $(-1, 1)$.
- Given the convergence criterion denoted as *tols*. If the convergence criterion is met, i.e. $\|\cdot| < tols$, the MbP algorithm is said to be *converged*. Otherwise the algorithm is said to be *failed*. Note that $|\cdot|$ is the absolute value and indicates the distance between the updated values. In the simulation study the *tols* for MbP algorithm is 0.01 for ν and 10^{-4} for ρ. In particular, the MbP algorithm converges if $|\nu^k - \nu^{k-1}| < 0.01$ and $|\rho^k - \rho^{k-1}| < 10^{-4}$.

- The maximal number of iterations k is limited to $k := 100$. If the maximal number of iterations is reached but the convergence criterion of MbP algorithm is not met, the MbP algorithm is considered to be failed to converge.

Algorithm 4.1 MbP Algorithm for Bivariate t-Distribution

Input: $\boldsymbol{x}$, ρ_0, $tols = (10^{-2}, 10^{-4})$
Output: $\hat{\nu}$, $\hat{\rho}$

$k := 1$
Find solution ν_n^1 of $\frac{\partial}{\partial \nu}\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) = 0$ in a range of $[1, 100]$ using bisection algorithm.
Find solution ρ_n^1 of $\frac{\partial}{\partial \rho}\ell_e(\nu_n^1, \rho; \boldsymbol{x}) = 0$ in a range $(-1, 1)$ of using bisection algorithm.
Find a minimum $\tilde{\nu}$ of $\frac{\partial}{\partial \nu}\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x})$ using R-function `optim` in the parameter range $[1, 100]$. The new value $\tilde{\nu}$ is set to the upper limit of the search range of bisection algorithm for parameter ν in the step $k = 2, 3, ...$
while $k < 100$ and $|\nu_n^k - \nu_n^{k-2}| > 10^{-2}$ and $|\rho_n^k - \rho_n^{k-1}| > 10^{-4}$ **do**
  $k := k + 1;$
  Find solution ν_n^k of $\frac{\partial}{\partial \nu}\ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) = -\frac{\partial}{\partial \nu}\ell_{e,mod}(\nu_n^{k-1}, \rho_n^{k-1}; \rho_0; \boldsymbol{x})$ in a range of $[1, \tilde{\nu}]$ via bisection algorithm.
  Find solution ρ_n^k of $\frac{\partial}{\partial \rho}\ell_e(\nu_n^{k-1}, \rho; \boldsymbol{x}) = 0$ in a range $(-1, 1)$ using bisection algorithm.
end while
return $\hat{\nu}$, $\hat{\rho}$

It is to note that the implemented MbP algorithm used in the simulation study always returns an estimate for the parameters ν and ρ. However the algorithm is not necessarily converged, i.e. the convergence criterion is not fulfilled. If the MbP algorithm converges, the returned values are the maximum likelihood estimates for the parameters. Otherwise the implemented algorithm returns an approximation of the maximum likelihood estimates or a suggestion of the estimates. In the next table we classify the outputs $\hat{\nu}$ and $\hat{\rho}$ from the implemented MbP algorithm. We assign each outcome to an indicator FF with a specified number 0-3. The indicator FF:=0 implies a successful convergence of the MbP algorithm, whereas other numbers indicate a failure of convergence. A flow chart of the MbP algorithm is illustrated in Figure 4.1.

FF	Outputs $\hat{\nu}$ and $\hat{\rho}$ generated by MbP algorithm and its description
0:	The predefined convergence criterion is met, MbP algorithm converges successfully and returns $\hat{\nu} := \hat{\nu}_n^k$ and $\hat{\rho} := \hat{\rho}_n^k$
1:	The bisection algorithm for ν, which is embedded in the MbP algorithm, does not find a solution in the first step. MbP algorithm fails to converge, but returns $\hat{\nu} := 100$ and $\hat{\rho} := \hat{\rho}_\tau$ as an output
2:	The bisection algorithm does not find a solution in step $k.$,$k \in \{2, ..., 100\}$. The MbP algorithm fails to converge, but returns estimates from the previous step $\hat{\nu} := \hat{\nu}_n^{k-1}$ and $\hat{\rho} := \hat{\rho}_n^{k-1}$
3:	The maximal iteration number $k = 100$ is reached but the convergence criterion has not met. The MbP algorithm fails to converge and returns $\hat{\nu} := \hat{\nu}_n^{100}$ and $\hat{\rho} := \hat{\rho}_n^{100}$.

Table 4.1: Definition of convergence indicator FF and the corresponding output for $\hat{\nu}$ and $\hat{\rho}$

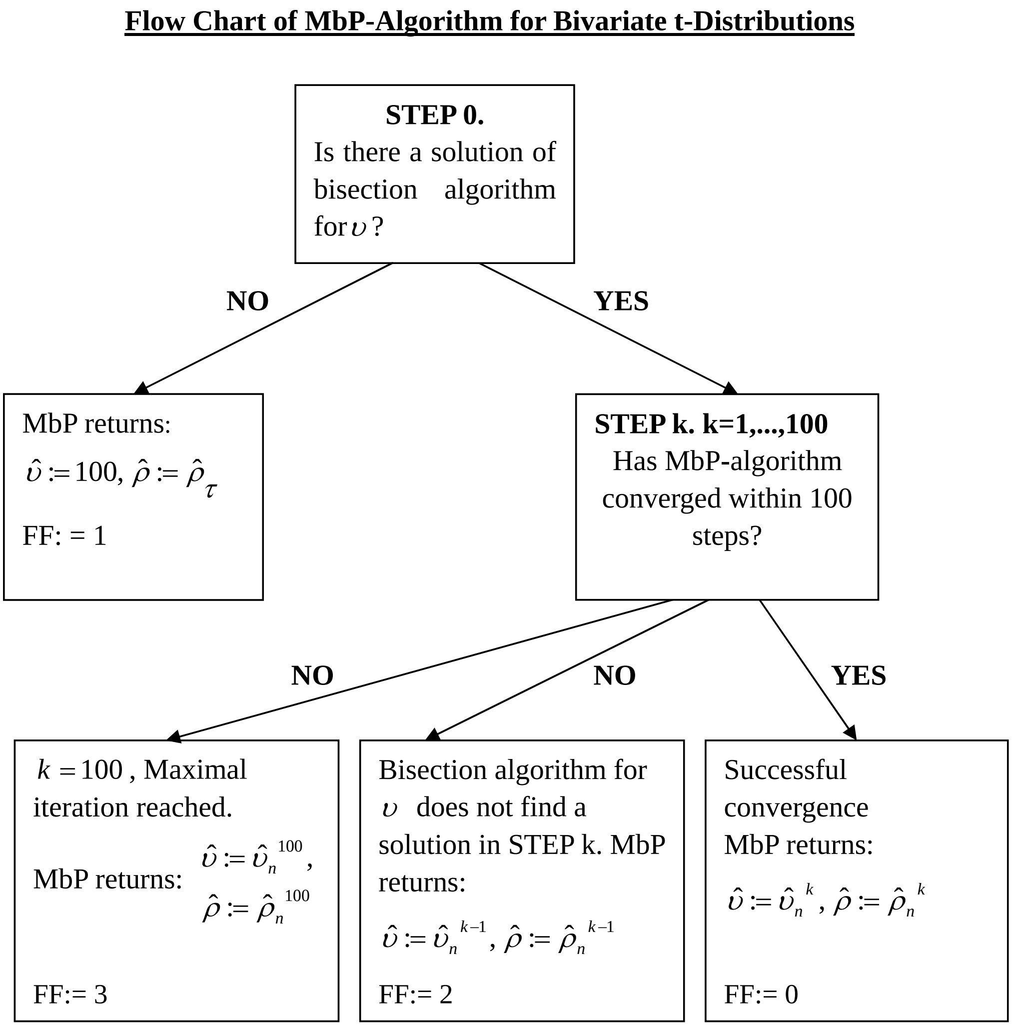


Figure 4.1: Flow chart for determining the convergence indicator FF for MbP algorithm applied to bivariate t-distribution

4.3 Illustration of the Performance of the MbP Algorithm for Different Choices of ρ_0

As described in Section 3.3, the idea of ρ_0 in the modified models is to put some information of the data correlation into the original working model in order to achieve an information dominance of the working model towards the error model. Under the condition of information dominance the MbP algorithm converges to a fixed point that is the maximum likelihood estimator.

By default we set ρ_0 equal to the estimated correlation coefficient $\hat{\rho}_\tau$ via estimated Kendall's τ of the corresponding data defined in (2.17), since the Kendall's τ is a robust estimator of the correlation of a bivariate data set and independent from the marginal distributions of the data. However this choice does not necessarily assure the convergence of the MbP algorithm. It is to note that in the simulation study the convergence of the MbP algorithm means that the convergence criterion, i.e. $|\nu^k - \nu^{k-1}| < 0.01$ and $|\rho^k - \rho^{k-1}| < 10^{-4}$, is reached. Otherwise the algorithm fails to converge.

In the following sections we look at some representative data sets for the analysis of the MbP algorithm under the influence of ρ_0.

Example 1: Convergence with $\rho_0 = \hat{\rho}_\tau$

Consider the following specific data set $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ (R-file: `x_data_10_06.RData`) of size $n = 100$ which is generated using `rmvt` function defined in (4.1) with true parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.6$. The estimated correlation using Kendall's τ of the data amounts approximately to 0.6845, i.e. $\hat{\rho}_\tau \approx 0.6845$. With $\rho_0 = \hat{\rho}_\tau$, the convergence criterion of the MbP algorithm, i.e. $|\nu^k - \nu^{k-1}| < 0.01$ and $|\rho^k - \rho^{k-1}| < 10^{-4}$, is reached after 7 steps ($k = 7$). In Figure 4.2 we illustrate the estimated ν, ρ's and the corresponding log likelihood of each step.

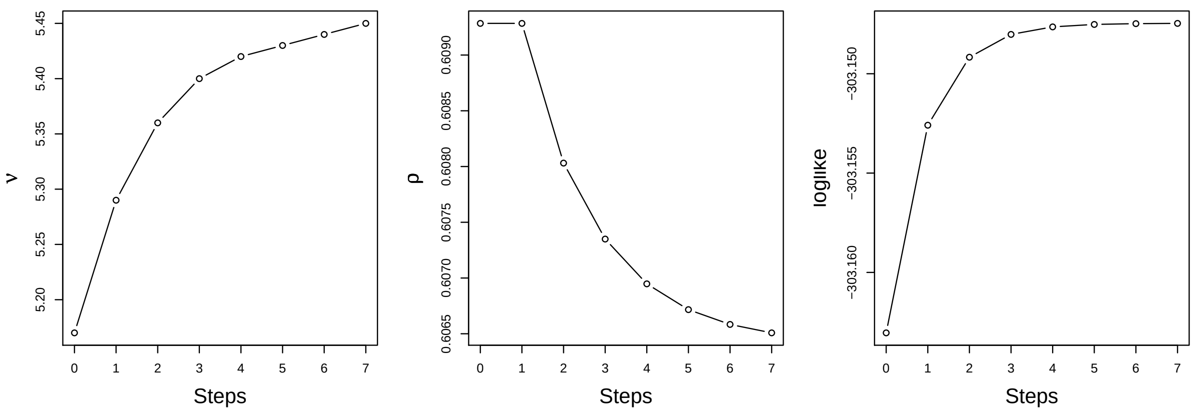


Figure 4.2: Estimates via MbP algorithm with $\rho_0 = \hat{\rho}_\tau$ on a randomly generated bivariate t-data $\boldsymbol{x}$ of size $n = 100$ with true parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.6$

With this data set $\boldsymbol{x}$ we can make any small shifts to ρ_0, i.e. $\rho_0 = \hat{\rho}_\tau + 0.025$ or

$\rho_0 = \hat{\rho}_\tau + 0.05$ with $\hat{\rho}_\tau \approx 0.6845$, the algorithm still converges to the same fixed point. However if we change the value of ρ_0 to 0.95 ($\rho_0 = 0.95$) which is far from the value of estimated correlation $\hat{\rho}_\tau \approx 0.6845$, the algorithm fails to converge, since the bisection algorithm cannot find a solution of the function

$$p_k(\nu) := \frac{\partial}{\partial \nu} \ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) + \frac{\partial}{\partial \nu} \ell_{e,mod}(\nu_n^{k-1}, \rho_n^{k-1}; \rho_0; \boldsymbol{x}) = 0$$

for ν in the 3th MbP step (k=3). The MbP step k with k=0,1,2,3 with respect to the parameter ν is illustrated in Figure 4.3.

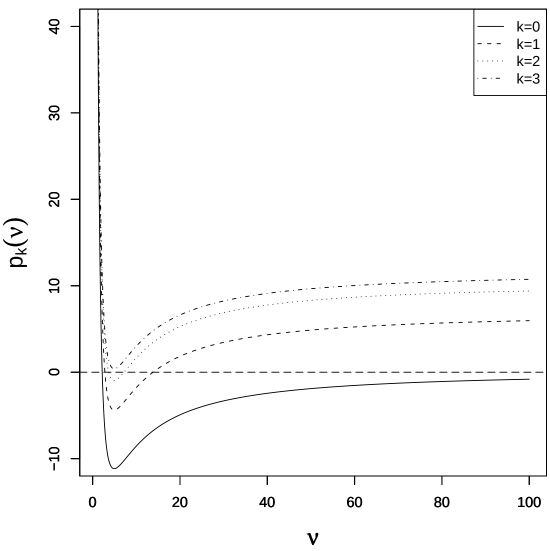


Figure 4.3: Estimates via MbP algorithm with $\rho_0 = \hat{\rho}_\tau$ on a randomly generated bivariate t-data $\boldsymbol{x}$ of size $n = 100$ with true parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.6$

Example 2: Convergence with a shifted ρ_0 from the value of $\hat{\rho}_\tau$

In Figure 3.4 in section 3.3 we presented an example of the influence of ρ_0 on the score function of the modified working model. We considered the following specific data set which was randomly generated using R-function `rmvt` in (4.1) (R-file: `x_data_10_09.RData`). The sample size is $n = 100$, the true parameters are $\nu_{tr} = 10$, $\rho_{tr} = 0.9$ and the estimated correlation using Kendall's τ is $\hat{\rho}_\tau \approx 0.8615$. With the same data set we continue the estimation in this example. In our first trials, the MbP algorithm fails to find a ML-estimate with $\rho_0 = \hat{\rho}_\tau$ and $\rho_0 = \hat{\rho}_\tau + 0.025$ in the step 0. The parameter ν was estimated as 100. This result indicates that the bisection algorithm fails to find the root of the score function of the modified working model $p_0(\nu) := \frac{\partial}{\partial \nu} \ell_{w,mod}(\nu; \rho_0; \boldsymbol{x}) = 0$ defined in (3.13).

In the next trial we shift the default of ρ_0 by 0.05 higher, i.e. $\rho_0 = \hat{\rho}_\tau + 0.05$, then the MbP algorithm converges after 17 steps. In this case the bisection algorithm found a root of the modified working model in the first MbP step as well as in the subsequent MbP iterations. We display the outcomes of MbP algorithm with $\rho_0 = \hat{\rho}_\tau + 0.05$ in Figure 4.4.

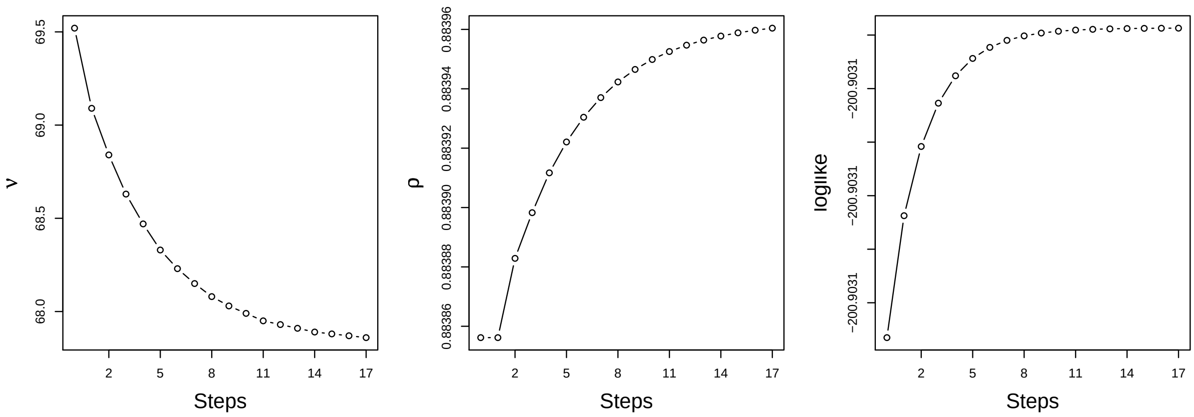


Figure 4.4: Estimates via MbP algorithm with $\rho_0 = \hat{\rho}_\tau + 0.05$ on a randomly generated bivariate t-data $\boldsymbol{x}$ of size $n = 100$ with true parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.9$

We can observe that the value of the log likelihood rises upwards to its maximum and the parameters converge to their fixed values.

Example 3: Failure after k MbP steps, $k < 100$

Nevertheless it is to point out that a solution in the first step does not guarantee the convergence of the MbP algorithm. To illustrate this case we consider a randomly generated bivariate t-data $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ using `rmvt` function in (4.1) (R-file: `x_10_06_2.RData`). The sample size is $n = 100$ and the data is generated by the true parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.6$. With $\rho_0 = \hat{\rho}_\tau \approx 0.6859$ no solution can be found by bisection algorithm for ν in the 12th step. The log likelihood becomes flat as displayed in Figure 4.5.

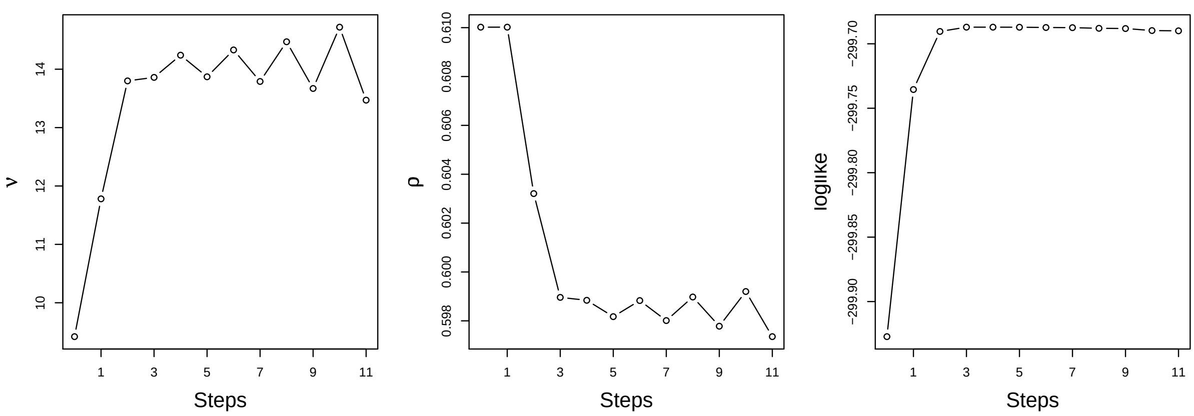


Figure 4.5: Estimation via MbP algorithm with $\rho_0 = \hat{\rho}_\tau$ on a randomly generated bivariate t-data $\boldsymbol{x}$ of size $n = 100$ with true parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.6$

However, if we choose the value of ρ_0 as $\rho_0 = \hat{\rho}_\tau - 0.025$ the MbP algorithm converges within 6 steps. Figure 4.6 shows the estimates of the parameters and the values of its log likelihood values.

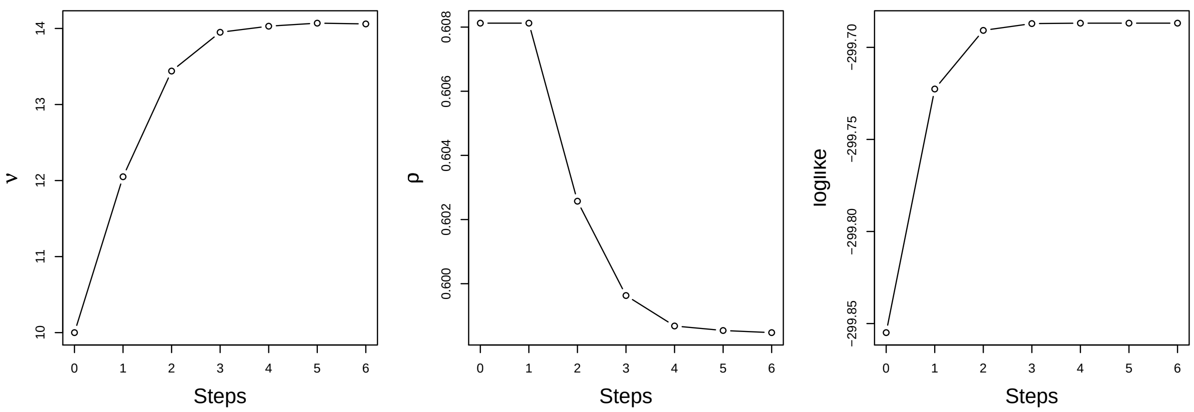


Figure 4.6: Estimation via MbP algorithm with $\rho_0 = \hat{\rho}_\tau - 0.025$ on a randomly generated bivariate t-data $\boldsymbol{x}$ of size $n = 100$ with true parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.6$

Conclusions

The choice of ρ_0 is crucial for the convergence of the MbP algorithm. If ρ_0 is chosen relatively far from the value of the ML-estimator of ρ, the MbP algorithm might hardly converge. If the ρ_0 is close to the ML-estimator of ρ, it is more likely that the estimates from MbP algorithm converges to the ML-estimators of the parameters. When we receive a data set with unknown parameters, it is in some cases hard to find an adequate ρ_0, since there is no general way to determine appropriate values for ρ_0. In conclusion we need to look at every data carefully.

In the simulation study in the next section we are going to generate 500 replicates using MbP algorithm and alternative computational methods for estimating the maximum likelihood estimates.

4.4 Alternative Computational Methods for Determining the MLE

4.4.1 Direct: Direct Maximization of Likelihood of Bivariate t-Distribution

In the direct maximization, the optimization routine `optim` in R will be directly applied to the log likelihood of bivariate t-distribution $\ell(\nu, \rho; \boldsymbol{x})$ defined in function (4.4). In this method we estimate the parameters ν and ρ simultaneously, while in MbP algorithm they are treated separately.

First we set the initial value (ν^0, ρ^0) for the `optim` function as $(\nu^0, \rho^0) := (3, \hat{\rho}_\tau)$. Furthermore, to process a 2-dimensional direct maximization we use the `optim`-function based on `L-BFGS-B`: the quasi-Newton method given a lower and/or upper bound for the parameters (Ref. R-help). Again the searching region for the `optim`-function is in the parameter space $\Theta := \{(\nu, \rho) : \nu \in [1, 100], \rho \in (-1, 1)\}$. Moreover, it is to mention that `optim` minimizes the objective function. Thus, to find the maximum of a function we simply minimize the negative of it. Finally, for the direct maximization algorithm we need to generate a set of bivariate t-distributed data using the R-function `rmvt` stated in (4.1).

Algorithm 4.2 Direct Maximization for bivariate t-Data

Input: data $\boldsymbol{x}$, initial vector (ν^0, ρ^0)

Output: a 2-dim. vector $(\hat{\nu}, \hat{\rho})$

1: Find 2-dimensional maximum $(\hat{\nu}, \hat{\rho})$ of $\ell(\nu, \rho; \boldsymbol{x})$ via `optim`-function with `option=L-BFGS-B` and search range for (ν, ρ) `lower=c(1,-0.99999)`, `upper=c(100,0.99999)`

2: **return** $(\hat{\nu}, \hat{\rho})$

4.4.2 CSP: Semi-parametric Method using Kendall's τ applied to Likelihood of Bivariate t-Copula

Unlike the MbP algorithm and the direct maximization method the idea of the semi-parametric method relies on the concept of the marginal empirical probability integral transformation. Given the bivariate t-data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ with $\boldsymbol{x}_j := (x_{1j}, ..., x_{nj})^t$, $j = 1, 2$ we use the empirical cumulative distribution function defined in (4.2) to transform the bivariate t-data to the approximate uniform data $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2)^t$ with $\boldsymbol{u}_j := (u_{1j}, ..., u_{nj})^t$, $j = 1, 2$. Furthermore, we estimate the parameters ν and ρ of the log likelihood of the bivariate t-copula separately. Let $\ell_c(\nu, \hat{\rho}_u; \boldsymbol{u})$ be the log likelihood of the bivariate t-copula, i.e. $\ell_c(\nu, \hat{\rho}_u; \boldsymbol{u})$ as defined in (4.3). First we estimate ρ by Kendall's τ denoted as $\hat{\rho}_\tau$ then we insert the estimator $\hat{\rho}_\tau$ into the log likelihood of the bivariate t-copula function to estimate the parameter ν. The CSP algorithm is stated as follows:

Algorithm 4.3 CSP Algorithm, Semi-parametric Method for bivariate t-Data

Input: data $\boldsymbol{x}$, initial value ν^0

Output: estimates of ν and $\hat{\rho}_u$

1: Transform $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2)$ to approximate uniform data $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2)$ using empirical cumulative function defined in (4.2)

2: Estimate correlation coefficient $\hat{\rho}$ via empirical data estimate of Kendall's τ of the uniform data $\boldsymbol{u}$, i.e. $\hat{\rho}_u := \sin(\frac{\pi}{2}\hat{\tau}(\boldsymbol{u}_1, \boldsymbol{u}_2))$

3: Find 1-dimensional maximum ν of $\ell_c(\nu, \hat{\rho}_u; \boldsymbol{u})$ via `optim` function with initial value $\nu^0 = 3$, method `L-BFGS-B` and search range $[1, 100]$ for ν.

4: **return** $\hat{\nu}$ and $\hat{\rho}_u$

4.4.3 CML: The Canonical Maximum Likelihood Method

Similar to the CSP method we do not make any assumptions about the distributions of the marginals and adapt the notion of the *marginal empirical probability integral transformation* to transform the bivariate t-data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ into approximate uniformly distributed data $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2)$ using the empirical cumulative distribution function defined in (4.2). Then we estimate the parameters ν and ρ of the log likelihood of the bivariate t-copula $\ell_c(\nu, \rho; \boldsymbol{u})$, defined in (4.3), jointly. The CML method is implemented via a two-stage procedure:

Algorithm 4.4 CML Algorithm, Canonical Maximum Likelihood for bivariate t-Data

Input: $\boldsymbol{x}$, initial values (ν^0, ρ^0)
Output: a 2-dim. vector $(\hat{\nu}, \hat{\rho})$

1: Transform $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2)$ to approximate uniform data $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2)$ using empirical cumulative function defined in (4.2)
2: Maximize 2-dimensional parameter vector (ν, ρ) of the log likelihood of the bivariate t-copula $\ell_c(\nu, \rho; \hat{\boldsymbol{u}})$ in the parameter space
$\Theta := \{(\nu, \rho) : \nu \in [1, 100], \rho \in (-1, 1)\}$ via the 2-dimensional `optim` function with initial values (ν^0, ρ^0), `option=L-BFGS-B`.
3: **return** $(\hat{\nu}, \hat{\rho})$

4.4.4 Outputs of Direct, CSP and CML Methods

Since the direct maximization and CML methods adopt the `optim` function on a 2-dimensional optimization problem using `L-BFGS-B` method and the CSP uses `optim` for a one-dimensional optimization, the successful convergence of the `optim` function is crucial for the estimation of all three methods.

Quoted from the help file on `optim` provided by R, the convergence of `optim` using `L-BFGS-B` method occurs when the reduction in the objective is within this factor of the machine tolerance, at default 10^7, and that is a tolerance of about 10^{-8}.

Due to some circumstances the `optim` function returns a solution with the an error message `ABNORMAL_TERMINATION_IN_LNSRCH`, which means an unsuccessful convergence of the `optim` function. This message is supposedly indicates that while the optimizer tries to do the line search in the steepest descent direction, it finds that the derivative along this line is not negative, which is impossible because the direction would not go downwards in this case. This problem could happen with numerical gradients used in `optim` when the surface of the function is almost flat. Despite of the unsuccessful convergence `optim` gives back an estimate that locates near the optimum point.

When there are many complicated terms in the objective function, such as the log-gamma functions and the digamma functions, `optim` might generate error messages which states that `L-BFGS-B` needs finite 'fn'. This error is presumably caused by some numerical problems, such as overflow or underflow, a condition that occurs when a calculation produces a result that is greater than what a given register can represent. In this case `optim` fails completely to process and does not give back any estimates.

In addition, we classify the outcome possibilities for the estimation $\hat{\nu}$ and $\hat{\rho}$ in the implemented Direct maximization, CSP and the CML methods. Similar to the implemented MbP algorithm we assign each type of the results to the indicator FF with numbers $0-2$. Flow charts of the direct maximization and CML methods, as well as the CML method, are illustrated in Figure 4.7.

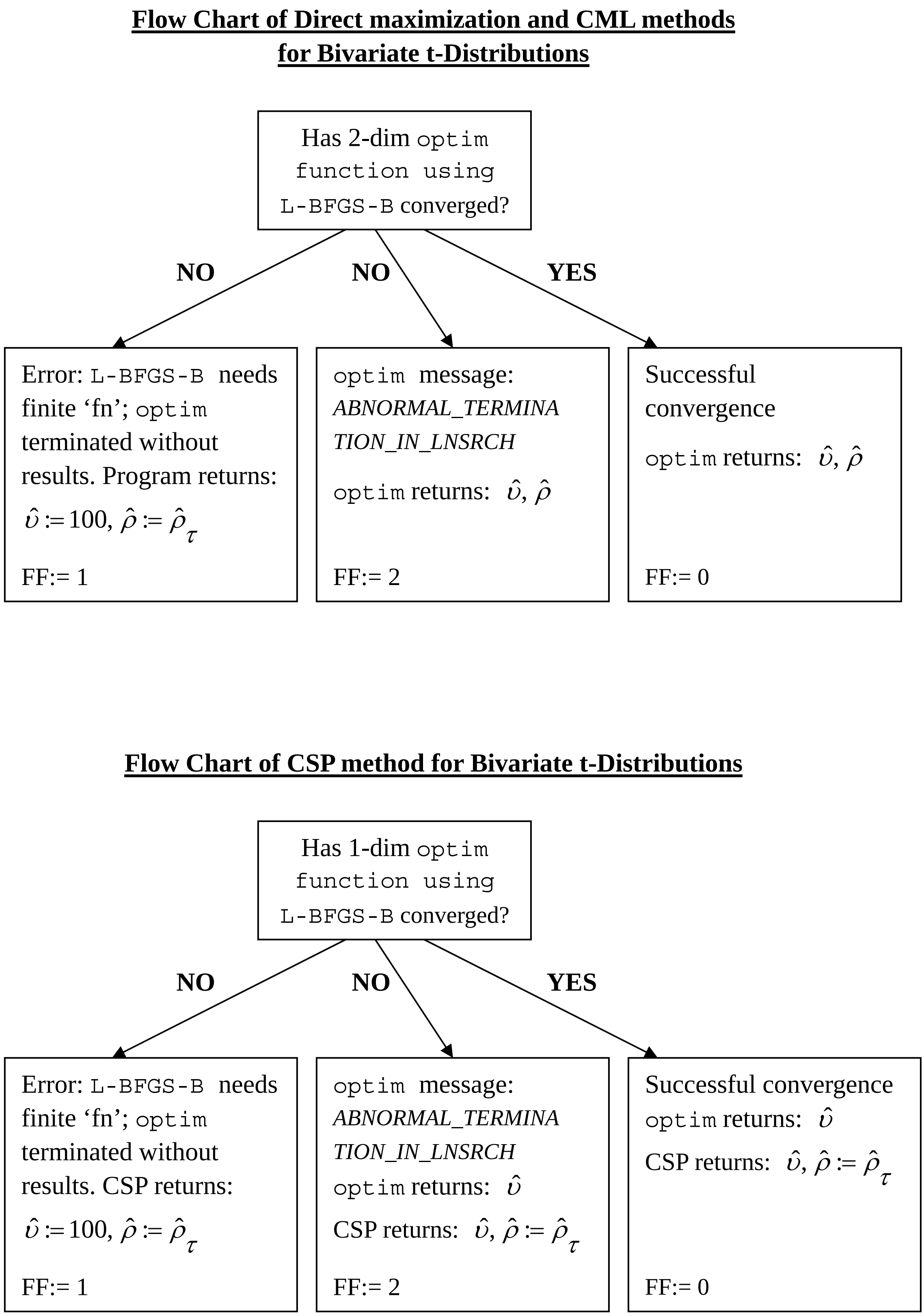


Figure 4.7: Flow charts for determining the convergence indicator FF for Direct maximization, CML and CSP methods applied to bivariate t-distribution

FF	The outputs $\hat{\nu}$ and $\hat{\rho}$ of Direct maximization, CSP and CML methods
0:	`optim` function converges successfully, i.e. `optim$convergence`:=0, The implemented programs return the estimates from `optim`.
1:	`optim` function generates an error message: `L-BFGS-B` needs finite 'fn', and does not give back any values. The programs return $\hat{\nu} := 100$ and $\hat{\rho} := \hat{\rho}_\tau$.
2:	`optim` function generates the message: `ABNORMAL_TERMINATION_IN_LNSRCH`, The implemented programs return the estimates from `optim`.

Table 4.2: Definition of convergence indicator FF and the corresponding output of Direct maximization, CSP and CML methods for $\hat{\nu}$ and $\hat{\rho}$

4.5 Simulation Setups and Performance Measures

In the subsequent simulation studies we are examining scenarios with different parameter combinations in order to compare the MbP algorithm with three other estimation methods - the direct maximization of the likelihood (Direct), the semi-parametric method (CSP) and the canonical maximum likelihood method (CML) which are frequently adopted in practice. For MbP algorithm and direct method we assume that the marginals are t-distributed while we do not make this assumption for the CSP method and CML method. The simulation studies are composed of the following parameter combinations:

- sample size $n \in \{100, 500\}$
- degrees of freedom $\nu_{tr} \in \{5, 10, 20\}$
- correlation $\rho_{tr} \in \{0.5, 0.9\}$
- number of replicates $r = 500$

Given these parameter values we have a total of 12 scenarios. For each of them we compute 500 replicates ($r = 500$) using the four estimation methods. With these replicates we can look at the summary statistics in order to find out the advantages and drawback of MbP algorithm in comparison with other estimation methods.

For the MbP algorithm we investigate three different versions by choosing different ρ_0. We denote these three scenarios as MbP0, MbP1 and MbP2 we take the value of ρ_0 as:

- MbP0: $\rho_0 := \hat{\rho}_\tau$
- MbP1: $\rho_0 := \hat{\rho}_\tau + 0.025$
- MbP2: $\rho_0 := \hat{\rho}_\tau + 0.05$

where $\hat{\rho}_\tau := \sin(\frac{\pi}{2}\hat{\tau})$ and $\hat{\tau}$ is the empirically estimated Kendall's τ from the data. Regarding the three versions of the MbP algorithm with different ρ_0 as three different methods we then have in total 6 estimation methods to be examined, i.e. MbP0, MbP1, MbP2, DIR, CSP and CML methods.

In order to inspect the estimates and evaluate the estimation methods we dismiss the ith estimates, $i \in \{1, ..., r\}$ with $r = 500$, if all six estimation methods return $\hat{\nu} = 100$

using the ith sample set. Let $|\{i\}|$ be the total number of dismissed samples, whereby $|\{\cdot\}|$ indicates the cardinality of a set. Then $r_c := r - |\{i\}|$ amounts the total number of samples we use for comparison of the 6 estimation methods.

Using the output indicator FF defined in Table 4.1 and 4.2, we can investigate the frequency of failure, type of the failure and the convergence of the corresponding algorithm. For each estimation method we denote r_g as the number of the replicates whose output indicator FF $\neq 1$. This means the implemented algorithms have iterated some steps till they converge or fail. The maximum number of r_g is the number of the replicates $r = 500$. The higher r_g is, the more reliable and stable is the estimation method. Thus, the number r_g is the an indicator for the reliability and robustness of the corresponding algorithm.

To measure of the goodness of algorithm we investigate the estimates generated by respective method. The summary statistics are defined in Section 2.4 of Chapter 2. The sample mean of the estimates should be close to the true parameter value. Moreover, the the sample mean should be greater than twice its estimated standard error. We also calculate the estimated relative bias and its estimated standard error, the estimated mean squared error and its estimated standard error. These quantities measure the variability of the estimates, though they are extremely vulnerable to outliers. To examine the goodness of the estimates without the influence of the outliers we can calculate the estimated robust standard error approximately by 1/1.35 times the interquartile range IQR of the estimates. Furthermore, the robust mean squared error measures the variability of the estimates by eliminating the outliers, since its calculation is adapted with the median and the estimated median absolute deviation instead of the sample variance and the estimated bias of the samples.

If we want to compare the 6 estimation methods with one another, an simple way is to look at the estimated efficiency based the estimated mean squared error and the estimated robust mean squared error of an algorithm to a predefined reference algorithm as defined in (2.57) and (2.58) in Section 2.4 of Chapter 2. Without any restriction we identify the MbP algorithm with fixed $\rho_0 := \hat{\rho}_\tau$, i.e. MbP0, as the reference method. In the simulation study if the estimated efficiency of an alternative algorithm compared to MbP0 is smaller than 1, i.e. $\text{eff}_{mse}(altern.method) < 1$ then the alternative method performs better than MbP0 method and similarly if $\text{eff}_{mse}(altern.method) > 1$ the reference method MbP0 produces more robust estimates and operates more stable.

In addition we calculate the log likelihood value, denoted as *loglike*, to examine if the algorithms return an good approximation for the ML-estimators of the parameters. We also record the computing time which is regarded as a crucial factor for the performance of an algorithm. An efficient algorithm requires the minimal computing cost and delivers accurate results.

Finally, we record the averaged iterations of MbP algorithm, denoted as *avg.it* to make a detailed look of how fast the MbP algorithm reaches the convergence criterion for each scenario. This number *avg.it* does not apply to Direct, CSP and CML methods, since they are not iterative algorithms.

n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*
$\nu_{tr}=5,\quad \rho_{tr}=0.5$																	
100	ν	MbP0	6.036	0.229	1.609	0.207	0.046	27.384	18.236	1.000	4.921	1.000	−309.308	500	499	0.503	7.393
		MbP1	5.945	0.164	1.608	0.189	0.033	14.322	5.746	0.523	4.951	1.006	−309.307	500	500	0.510	7.494
		MbP2	5.944	0.164	1.609	0.189	0.033	14.318	5.744	0.523	5.000	1.016	−309.307	500	500	0.513	7.548
		Direct	5.946	0.164	1.618	0.189	0.033	14.315	5.742	0.523	4.941	1.004	−309.307	500	500	0.048	
		CSP	16.211	1.345	3.657	2.242	0.269	1029.645	125.523	37.600	13.643	2.772	−313.414	500	500	0.361	
		CML	17.636	1.411	4.295	2.527	0.282	1155.703	132.030	42.204	15.052	3.059	−313.767	500	499	1.218	
	ρ	MbP0	0.485	0.004	0.087	−0.030	0.008	0.007	0.000	1.000	0.017	1.000	−309.308	500	499	0.503	7.393
		MbP1	0.485	0.004	0.087	−0.030	0.008	0.007	0.001	1.007	0.017	1.000	−309.307	500	500	0.510	7.494
		MbP2	0.485	0.004	0.087	−0.030	0.008	0.007	0.001	1.006	0.017	1.000	−309.307	500	500	0.513	7.548
		Direct	0.485	0.004	0.087	−0.030	0.008	0.007	0.001	1.006	0.017	1.000	−309.307	500	500	0.048	
		CSP	0.489	0.004	0.091	−0.023	0.008	0.009	0.001	1.168	0.018	1.100	−313.414	500	500	0.361	
		CML	0.497	0.004	0.093	−0.006	0.008	0.009	0.001	1.218	0.019	1.147	−313.767	500	499	1.218	
500	ν	MbP0	5.124	0.032	0.686	0.025	0.006	0.526	0.040	1.000	0.918	1.000	−1546.512	500	500	1.280	5.800
		MbP1	5.122	0.032	0.688	0.024	0.006	0.525	0.040	0.999	0.900	0.979	−1546.512	500	500	1.305	5.950
		MbP2	5.120	0.032	0.691	0.024	0.006	0.524	0.039	0.997	0.887	0.966	−1546.512	500	500	1.441	6.552
		Direct	5.124	0.032	0.692	0.025	0.006	0.527	0.040	1.001	0.914	0.995	−1546.512	500	500	0.047	
		CSP	5.087	0.088	1.333	0.017	0.018	3.913	0.724	7.441	3.820	4.159	−1549.151	500	500	1.116	
		CML	5.131	0.091	1.404	0.026	0.018	4.126	0.752	7.846	3.803	4.141	−1549.186	500	500	3.479	
	ρ	MbP0	0.499	0.002	0.036	−0.001	0.003	0.001	0.000	1.000	0.003	1.000	−1546.512	500	500	1.280	5.800
		MbP1	0.500	0.002	0.036	−0.001	0.003	0.001	0.000	0.999	0.003	1.000	−1546.512	500	500	1.305	5.950
		MbP2	0.500	0.002	0.036	−0.001	0.003	0.001	0.000	0.999	0.003	1.000	−1546.512	500	500	1.441	6.552
		Direct	0.499	0.002	0.036	−0.001	0.003	0.001	0.000	0.998	0.003	0.960	−1546.512	500	500	0.047	
		CSP	0.500	0.002	0.042	−0.001	0.003	0.001	0.000	1.137	0.004	1.301	−1549.151	500	500	1.116	
		CML	0.502	0.002	0.042	0.004	0.003	0.001	0.000	1.116	0.004	1.259	−1549.186	500	500	3.479	
$\nu_{tr}=5,\quad \rho_{tr}=0.9$																	
100	ν	MbP0	6.237	0.243	1.752	0.247	0.049	30.992	18.453	1.000	5.710	1.000	−240.636	500	499	0.585	8.467
		MbP1	6.134	0.226	1.742	0.227	0.045	26.797	18.088	0.865	5.596	0.980	−240.639	500	499	0.597	8.625
		MbP2	6.039	0.288	1.602	0.208	0.058	42.538	25.529	1.373	4.544	0.796	−240.839	500	498	0.624	9.000
		Direct	6.235	0.243	1.750	0.247	0.049	30.982	18.453	1.000	5.563	0.974	−240.636	500	500	0.068	
		CSP	16.906	1.302	5.410	2.381	0.260	988.711	119.645	31.902	24.984	4.375	−245.849	500	500	0.364	
		CML	17.687	1.339	5.555	2.537	0.268	1058.018	122.761	34.138	28.451	4.982	−245.992	500	500	1.725	
	ρ	MbP0	0.896	0.001	0.021	−0.005	0.001	0.000	0.000	1.000	0.001	1.000	−240.636	500	499	0.585	8.467
		MbP1	0.896	0.001	0.021	−0.005	0.001	0.000	0.000	1.000	0.001	1.000	−240.639	500	499	0.597	8.625
		MbP2	0.897	0.001	0.021	−0.004	0.001	0.000	0.000	1.023	0.001	0.928	−240.839	500	498	0.624	9.000
		Direct	0.896	0.001	0.021	−0.005	0.001	0.000	0.000	0.999	0.001	1.000	−240.636	500	500	0.068	
		CSP	0.896	0.001	0.025	−0.004	0.001	0.001	0.000	1.736	0.001	1.462	−245.849	500	500	0.364	
		CML	0.895	0.001	0.025	−0.006	0.001	0.001	0.000	1.721	0.001	1.298	−245.992	500	500	1.725	
500	ν	MbP0	5.141	0.034	0.711	0.028	0.007	0.592	0.044	1.000	1.065	1.000	−1203.251	500	500	1.348	6.022
		MbP1	5.129	0.034	0.711	0.026	0.007	0.590	0.043	0.996	1.104	1.037	−1203.251	500	500	1.844	8.194
		MbP2	5.128	0.034	0.707	0.026	0.007	0.585	0.043	0.988	1.107	1.039	−1203.253	500	500	2.277	10.104
		Direct	5.141	0.034	0.713	0.028	0.007	0.592	0.044	1.001	1.109	1.041	−1203.250	500	500	0.077	
		CSP	5.317	0.090	1.548	0.063	0.018	4.155	0.546	7.020	5.014	4.708	−1206.857	500	500	1.072	
		CML	5.321	0.091	1.529	0.064	0.018	4.240	0.658	7.163	4.713	4.425	−1206.695	500	500	5.106	
	ρ	MbP0	0.900	0.000	0.007	−0.000	0.000	0.000	0.000	1.000	0.000	1.000	−1203.251	500	500	1.348	6.022
		MbP1	0.900	0.000	0.007	−0.000	0.000	0.000	0.000	1.001	0.000	1.000	−1203.251	500	500	1.844	8.194
		MbP2	0.900	0.000	0.008	−0.000	0.000	0.000	0.000	1.008	0.000	1.000	−1203.253	500	500	2.277	10.104
		Direct	0.900	0.000	0.008	−0.000	0.000	0.000	0.000	1.000	0.000	1.000	−1203.250	500	500	0.077	
		CSP	0.900	0.000	0.011	−0.000	0.001	0.000	0.000	1.599	0.000	2.560	−1206.857	500	500	1.072	
		CML	0.899	0.000	0.011	−0.001	0.001	0.000	0.000	1.524	0.000	1.960	−1206.695	500	500	5.106	

Table 4.3: Estimated ν's and ρ's comparing MbP algorithm adapted with different ρ_0 (MbP0, MbP1, MbP2), direct maximization (Direct), semi-parametric (CSP) and canonical maximum likelihood (CML) methods over 500 replicates. Samples are generated with parameters $\nu_{tr} = 5$ and $\rho_{tr} = 0.5, 0.9$.

n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*
$\nu_{tr}=10,\quad \rho_{tr}=0.5$																	
100	ν	MbP0	16.863	0.870	6.100	0.686	0.087	416.789	72.523	1.000	54.524	1.000	−289.106	488	471	0.514	8.635
		MbP1	16.362	0.805	6.102	0.636	0.081	356.794	64.820	0.856	54.539	1.000	−289.106	488	475	0.509	8.516
		MbP2	15.975	0.760	6.102	0.598	0.076	317.442	59.888	0.762	54.524	1.000	−289.107	488	478	0.494	8.201
		Direct	16.817	0.853	6.094	0.682	0.085	401.377	69.410	0.963	54.485	0.999	−289.102	488	500	0.062	
		CSP	30.874	1.779	32.737	2.087	0.178	1979.461	153.331	4.749	132.489	2.430	−291.817	488	500	0.450	
		CML	33.252	1.822	70.656	2.325	0.182	2161.233	157.796	5.185	176.128	3.230	−291.920	488	495	1.430	
	ρ	MbP0	0.488	0.004	0.080	−0.024	0.007	0.007	0.000	1.000	0.014	1.000	−289.106	488	471	0.514	8.635
		MbP1	0.488	0.004	0.080	−0.024	0.007	0.007	0.000	0.999	0.014	1.000	−289.106	488	475	0.509	8.516
		MbP2	0.488	0.004	0.080	−0.024	0.007	0.007	0.000	0.999	0.014	1.000	−289.107	488	478	0.494	8.201
		Direct	0.488	0.004	0.080	−0.023	0.007	0.007	0.000	0.993	0.015	1.049	−289.102	488	500	0.062	
		CSP	0.492	0.004	0.088	−0.017	0.008	0.008	0.001	1.189	0.016	1.165	−291.817	488	500	0.450	
		CML	0.504	0.004	0.089	0.009	0.008	0.008	0.001	1.187	0.017	1.223	−291.920	488	495	1.430	
500	ν	MbP0	10.845	0.140	2.231	0.084	0.014	10.552	1.565	1.000	9.336	1.000	−1445.990	500	500	1.412	6.822
		MbP1	10.841	0.140	2.234	0.084	0.014	10.546	1.565	0.999	9.290	0.995	−1445.990	500	500	1.451	7.028
		MbP2	10.838	0.140	2.235	0.084	0.014	10.542	1.564	0.999	9.310	0.997	−1445.990	500	500	1.562	7.586
		Direct	10.844	0.140	2.234	0.084	0.014	10.550	1.565	1.000	9.330	0.999	−1445.990	500	500	0.065	
		CSP	14.403	0.818	4.791	0.440	0.082	354.349	67.344	33.582	38.858	4.162	−1449.098	500	500	1.270	
		CML	14.747	0.839	5.006	0.475	0.084	374.454	69.716	35.488	40.088	4.294	−1449.036	500	499	4.222	
	ρ	MbP0	0.499	0.002	0.036	−0.003	0.003	0.001	0.000	1.000	0.003	1.000	−1445.990	500	500	1.412	6.822
		MbP1	0.499	0.002	0.036	−0.003	0.003	0.001	0.000	1.000	0.003	1.000	−1445.990	500	500	1.451	7.028
		MbP2	0.499	0.002	0.037	−0.002	0.003	0.001	0.000	1.000	0.003	1.000	−1445.990	500	500	1.562	7.586
		Direct	0.499	0.002	0.037	−0.003	0.003	0.001	0.000	1.001	0.003	1.000	−1445.990	500	500	0.065	
		CSP	0.499	0.002	0.041	−0.002	0.003	0.001	0.000	1.142	0.004	1.166	−1449.098	500	500	1.270	
		CML	0.503	0.002	0.040	0.005	0.003	0.001	0.000	1.089	0.004	1.169	−1449.036	500	499	4.222	
$\nu_{tr}=10,\quad \rho_{tr}=0.9$																	
100	ν	MbP0	20.319	1.113	7.478	1.032	0.111	725.148	96.778	1.000	71.870	1.000	−220.096	499	462	0.655	11.002
		MbP1	16.066	0.818	5.708	0.607	0.082	371.087	67.256	0.512	46.295	0.644	−220.133	499	482	0.554	8.942
		MbP2	11.133	0.530	3.564	0.113	0.053	141.639	38.437	0.195	24.105	0.335	−220.573	499	494	0.387	5.998
		Direct	20.070	1.073	7.696	1.007	0.107	675.635	92.075	0.932	75.379	1.049	−220.061	499	500	0.081	
		CSP	28.863	1.660	20.894	1.886	0.166	1731.516	142.743	2.388	133.148	1.853	−223.240	499	500	0.432	
		CML	30.938	1.715	29.165	2.094	0.172	1906.239	147.886	2.629	146.744	2.042	−223.165	499	499	1.892	
	ρ	MbP0	0.896	0.001	0.019	−0.005	0.001	0.000	0.000	1.000	0.001	1.000	−220.096	499	462	0.655	11.002
		MbP1	0.897	0.001	0.019	−0.003	0.001	0.000	0.000	0.915	0.001	0.986	−220.133	499	482	0.554	8.942
		MbP2	0.901	0.001	0.020	0.002	0.001	0.000	0.000	0.952	0.001	1.148	−220.573	499	494	0.387	5.998
		Direct	0.896	0.001	0.018	−0.004	0.001	0.000	0.000	0.869	0.001	0.847	−220.061	499	500	0.081	
		CSP	0.896	0.001	0.024	−0.004	0.001	0.001	0.000	1.548	0.001	1.487	−223.240	499	500	0.432	
		CML	0.896	0.001	0.022	−0.004	0.001	0.001	0.000	1.425	0.001	1.317	−223.165	499	499	1.892	
500	ν	MbP0	10.995	0.159	2.454	0.100	0.016	13.593	2.148	1.000	11.142	1.000	−1102.730	500	500	1.880	8.880
		MbP1	10.880	0.145	2.417	0.088	0.015	11.306	1.579	0.832	10.886	0.977	−1102.738	500	500	2.162	10.332
		MbP2	9.419	0.082	1.651	−0.058	0.008	3.687	0.226	0.271	6.500	0.583	−1103.192	500	500	1.712	8.250
		Direct	10.995	0.159	2.449	0.100	0.016	13.591	2.148	1.000	11.090	0.995	−1102.730	500	500	0.086	
		CSP	14.151	0.770	4.591	0.415	0.077	313.882	62.887	23.092	42.015	3.771	−1106.244	500	500	1.187	
		CML	14.323	0.781	4.639	0.432	0.078	323.818	63.996	23.823	42.847	3.846	−1105.907	500	500	5.845	
	ρ	MbP0	0.900	0.000	0.008	−0.001	0.000	0.000	0.000	1.000	0.000	1.000	−1102.730	500	500	1.880	8.880
		MbP1	0.900	0.000	0.008	−0.000	0.000	0.000	0.000	0.998	0.000	1.000	−1102.738	500	500	2.162	10.332
		MbP2	0.902	0.000	0.009	0.003	0.000	0.000	0.000	1.279	0.000	1.473	−1103.192	500	500	1.712	8.250
		Direct	0.900	0.000	0.008	−0.001	0.000	0.000	0.000	1.000	0.000	1.000	−1102.730	500	500	0.086	
		CSP	0.900	0.000	0.010	−0.000	0.000	0.000	0.000	1.554	0.000	1.960	−1106.244	500	500	1.187	
		CML	0.900	0.000	0.010	−0.001	0.000	0.000	0.000	1.375	0.000	1.692	−1105.907	500	500	5.845	

Table 4.4: Estimated ν's and ρ's comparing MbP algorithm adapted with different ρ_0 (MbP0, MbP1, MbP2), direct maximization (Direct), semi-parametric (CSP) and canonical maximum likelihood (CML) methods over 500 replicates. Samples are generated with parameters $\nu_{tr} = 10$ and $\rho_{tr} = 0.5, 0.9$.

n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\bar{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*
$\nu_{tr} = 20, \quad \rho_{tr} = 0.5$																	
100	ν	MbP0	35.698	1.463	22.419	0.785	0.073	1229.310	107.210	1.000	439.302	1.000	−280.012	459	391	0.436	9.476
		MbP1	34.691	1.419	20.197	0.735	0.071	1139.632	103.049	0.927	419.022	0.954	−280.013	459	399	0.429	9.228
		MbP2	32.854	1.339	17.786	0.643	0.067	987.593	95.239	0.803	372.929	0.849	−280.017	459	412	0.404	8.364
		Direct	36.609	1.469	27.129	0.830	0.073	1266.716	106.115	1.030	539.108	1.227	−279.988	459	500	0.075	–
		CSP	40.903	1.985	69.871	1.045	0.099	2246.308	135.436	1.827	541.965	1.234	−282.390	459	500	0.488	–
		CML	43.668	2.007	69.497	1.183	0.100	2408.259	137.908	1.959	840.402	1.913	−282.338	459	494	1.579	–
	ρ	MbP0	0.491	0.004	0.077	−0.017	0.007	0.006	0.000	1.000	0.013	1.000	−280.012	459	391	0.436	9.476
		MbP1	0.491	0.004	0.079	−0.017	0.007	0.006	0.000	1.001	0.014	1.037	−280.013	459	399	0.429	9.228
		MbP2	0.492	0.004	0.078	−0.017	0.007	0.006	0.000	0.994	0.013	0.996	−280.017	459	412	0.404	8.364
		Direct	0.493	0.004	0.080	−0.014	0.007	0.006	0.000	0.965	0.014	1.034	−279.988	459	500	0.075	–
		CSP	0.494	0.004	0.083	−0.012	0.008	0.007	0.001	1.137	0.015	1.111	−282.390	459	500	0.488	–
		CML	0.509	0.004	0.083	0.017	0.008	0.007	0.001	1.136	0.016	1.198	−282.338	459	494	1.579	–
500	ν	MbP0	25.188	0.749	8.036	0.259	0.037	305.373	48.179	1.000	120.316	1.000	−1396.745	496	486	1.701	9.121
		MbP1	24.908	0.709	8.040	0.245	0.035	273.204	43.361	0.895	120.343	1.000	−1396.745	496	488	1.628	8.770
		MbP2	24.365	0.648	8.042	0.218	0.032	227.349	37.789	0.744	120.725	1.003	−1396.747	496	491	1.648	8.853
		Direct	25.120	0.740	8.036	0.256	0.037	297.581	47.014	0.974	120.182	0.999	−1396.742	496	500	0.081	–
		CSP	34.212	1.534	25.414	0.711	0.077	1369.245	109.377	4.484	350.365	2.912	−1399.270	496	500	1.410	–
		CML	35.055	1.550	27.608	0.753	0.078	1418.309	110.842	4.645	362.128	3.010	−1399.150	496	500	4.953	–
	ρ	MbP0	0.499	0.001	0.036	−0.001	0.003	0.001	0.000	1.000	0.003	1.000	−1396.745	496	486	1.701	9.121
		MbP1	0.499	0.001	0.036	−0.001	0.003	0.001	0.000	0.997	0.003	1.000	−1396.745	496	488	1.628	8.770
		MbP2	0.500	0.001	0.035	−0.001	0.003	0.001	0.000	0.992	0.003	0.919	−1396.747	496	491	1.648	8.853
		Direct	0.500	0.001	0.035	−0.001	0.003	0.001	0.000	0.986	0.003	0.919	−1396.742	496	500	0.081	–
		CSP	0.500	0.002	0.038	−0.000	0.003	0.001	0.000	1.175	0.003	1.085	−1399.270	496	500	1.410	–
		CML	0.504	0.002	0.037	0.009	0.003	0.001	0.000	1.119	0.003	1.101	−1399.150	496	500	4.953	–
$\nu_{tr} = 20, \quad \rho_{tr} = 0.9$																	
100	ν	MbP0	41.880	1.632	54.334	1.094	0.082	1789.666	122.133	1.000	602.665	1.000	−210.642	492	371	0.484	10.612
		MbP1	31.538	1.428	16.897	0.577	0.071	1136.982	101.933	0.635	215.651	0.358	−210.752	492	421	0.374	7.223
		MbP2	15.946	0.802	6.220	−0.203	0.040	332.843	48.932	0.186	159.309	0.264	−211.521	492	478	0.236	3.902
		Direct	43.828	1.634	59.791	1.191	0.082	1881.313	122.466	1.051	1058.161	1.756	−210.511	492	500	0.094	–
		CSP	39.696	1.890	69.414	0.985	0.094	2144.808	129.733	1.198	479.382	0.795	−213.286	492	500	0.467	–
		CML	41.344	1.902	69.191	1.067	0.095	2234.931	131.057	1.249	637.943	1.059	−213.155	492	498	1.980	–
	ρ	MbP0	0.896	0.001	0.019	−0.005	0.001	0.000	0.000	1.000	0.001	1.000	−210.642	492	371	0.484	10.612
		MbP1	0.898	0.001	0.019	−0.002	0.001	0.000	0.000	0.935	0.001	0.915	−210.752	492	421	0.374	7.223
		MbP2	0.905	0.001	0.019	0.005	0.001	0.000	0.000	0.980	0.001	1.019	−211.521	492	478	0.236	3.902
		Direct	0.898	0.001	0.019	−0.002	0.001	0.000	0.000	0.725	0.001	0.844	−210.511	492	500	0.094	–
		CSP	0.897	0.001	0.024	−0.003	0.001	0.001	0.000	1.317	0.001	1.503	−213.286	492	500	0.467	–
		CML	0.897	0.001	0.021	−0.003	0.001	0.001	0.000	1.222	0.001	1.158	−213.155	492	498	1.980	–
500	ν	MbP0	26.531	0.866	8.801	0.327	0.043	418.044	59.629	1.000	138.057	1.000	−1053.164	500	481	2.233	11.790
		MbP1	20.228	0.404	5.544	0.011	0.020	81.801	19.743	0.196	63.376	0.459	−1053.334	500	498	1.690	8.725
		MbP2	11.117	0.131	2.674	−0.444	0.007	87.469	2.018	0.209	100.452	0.728	−1055.902	500	500	0.485	2.458
		Direct	26.719	0.881	8.797	0.336	0.044	433.568	61.511	1.037	138.133	1.001	−1053.144	500	500	0.098	–
		CSP	31.540	1.388	19.381	0.577	0.069	1096.265	98.135	2.622	367.529	2.662	−1055.931	500	500	1.371	–
		CML	32.248	1.402	20.066	0.612	0.070	1132.696	99.182	2.710	361.737	2.620	−1055.614	500	498	6.340	–
	ρ	MbP0	0.900	0.000	0.008	−0.000	0.000	0.000	0.000	1.000	0.000	1.000	−1053.164	500	481	2.233	11.790
		MbP1	0.901	0.000	0.008	0.001	0.000	0.000	0.000	0.985	0.000	1.014	−1053.334	500	498	1.690	8.725
		MbP2	0.908	0.000	0.009	0.009	0.000	0.000	0.000	2.208	0.000	1.492	−1055.902	500	500	0.485	2.458
		Direct	0.900	0.000	0.007	−0.000	0.000	0.000	0.000	0.946	0.000	0.825	−1053.144	500	500	0.098	–
		CSP	0.900	0.000	0.010	−0.000	0.000	0.000	0.000	1.527	0.000	1.617	−1055.931	500	500	1.371	–
		CML	0.900	0.000	0.009	−0.000	0.000	0.000	0.000	1.335	0.000	1.195	−1055.614	500	498	6.340	–

Table 4.5: Estimated ν's and ρ's comparing MbP algorithm adapted with different ρ_0 (MbP0, MbP1, MbP2), direct maximization (Direct), semi-parametric (CSP) and canonical maximum likelihood (CML) methods over 500 replicates. Samples are generated with parameters $\nu_{tr} = 20$ and $\rho_{tr} = 0.5, 0.9$.

4.6 Simulation Results

In Table 4.3, 4.4 and 4.10 we demonstrate the summary statistics for each scenario with different parameter settings given in Section 4.5. The estimates of the compared replicates of size r_c are illustrated via boxplots in Figure 4.8, 4.9 and 4.10. We refer the definitions of the relevant summery statistics to Section 2.4 of Chapter 2.

(i) *Scenarios with* $\nu_{tr} = 5$, $\rho_{tr} = 0.5, 0.9$ *and* $n = 100, 500$:
Table 4.3 and Figure 4.8 give an overview of the estimates. In these scenarios the true value of degrees of freedom ν is relatively low. For the scenario $n = 100$, $\nu_{tr} = 5$ and $\rho_{tr} = 0.5$, the estimated mean $\bar{\theta}$, estimated relative bias $\widehat{rb}(\bar{\theta})$ and the estimated mean squared error (mse) $\widehat{mse}(\hat{\theta})$ for parameter ν from MbP0 method are higher than MbP1, MbP2 and Direct methods, as well as the numbers of the estimated standard error of mean $s(\bar{\theta})$, the estimated standard error of relative bias $s_{rb}(\bar{\theta})$ and the estimated standard error of $\widehat{mse}(\hat{\theta})$. These high statistic numbers of MbP0 are caused by one outlier where the MbP0 method fails to converge. We can observe it in the quantity r_g as well as in the first panel of Figure 4.8. The influence of the outlier is also reflected in the number of estimated efficiency $\widehat{\text{eff}}_{mse}$. If we look at the estimated robust statistics such as the estimated robust standard error $\hat{\sigma}_r(\hat{\theta})$ and the estimated robust mean squared error $\widehat{rmse}(\hat{\theta})$ the results of MbP0 are similar to those of the MbP1, MbP2 and Direct methods. In contrary the estimates produced by CSP and CML methods are more variable and the sample means are far from the true parameter values.
For the scenario $n = 100$, $\nu_{tr} = 5$ and $\rho_{tr} = 0.9$ we can observe that the MbP0 performs similarly well as the MbP1, MbP2 and Direct methods whereas the variability of the estimates from CSP and CML methods remains high.
As the number of sample size of the data rises to $n = 500$ (in scenario $n = 500$, $\nu_{tr} = 5$, $\rho_{tr} = 0.5$ and scenario $n = 500$, $\nu_{tr} = 5$, $\rho_{tr} = 0.9$) none of the estimation methods fails in the entire 500 replications. Thus we can see that $r_g = 500$ is stated for all methods. The summary statistics of all estimation methods are similar. They demonstrate the low variability of the estimates generated by all method as well as the accuracy of the estimations. Especially for the parameter ρ, we can observe that the sample mean shows almost the exact number as the true parameter value.

(ii) *Scenarios with* $\nu_{tr} = 10$ *and* $\rho_{tr} = 0.5, 0.9$ *and* $n = 100, 500$:
As the true parameter value for ν becomes higher, $\nu_{tr} = 10$, we can observe in Table 4.4 or in Figure 4.9 more outliers of $\hat{\nu}$ as well as the number of fails $(500 - r_g)$ for MbP0, MbP1 and MbP2 methods. For the scenario with $n = 100$, $\rho_{tr} = 0.5$ the summary statistics of MbP0, MbP1, MbP2 and Direct methods are almost identical. While the estimated robust statistic $\widehat{\text{eff}}_{rmse}$ of those methods are nearly the same, the smaller number of the statistics $\widehat{\text{eff}}_{mse}$ shows that the variability and the bias of the estimates from MbP2 are smaller. We can observe that the sample mean of parameter ν from MbP2 is closer to the true parameter value than the estimates from any other methods.
However the advantage of MbP2 becomes more evident when the data samples have high correlations, i.e. $\rho_{tr} = 0.9$. In the scenario $n = 100$, $\nu_{tr} = 10$, $\rho_{tr} = 0.9$,

the sample mean of estimates $\hat{\nu}$ from MbP2 is very close to its true parameter value whereas the statistic from other methods significantly show the bias of their estimates. Furthermore, the statistics $\widehat{\text{eff}}_{mse}$ and $\widehat{\text{eff}}_{rmse}$ of MbP2 are much smaller than any other methods, which demonstrate the accuracy and robustness of the MbP2 estimates. In addition, the value of r_g of MbP2 is also close to the total of replicates which means the lower number of failures in comparison to the scenario $n = 100$, $\nu_{tr} = 10$, $\rho_{tr} = 0.5$.

As the sample size increases, i.e. $n = 500$, the estimations become more accurate as demonstrated in the given statistics in the result table. The number of failures also disappears completely. While the statistics of the MbP and Direct methods in scenario $n = 500$, $\nu_{tr} = 10$, $\rho_{tr} = 0.5$ are very similar, the sample mean $\bar{\theta}$ of parameter ν from MbP2 slightly underestimates its true parameter value in the scenario $n = 500$, $\nu_{tr} = 10$, $\rho_{tr} = 0.9$. However the estimated efficiency $\widehat{\text{eff}}_{mse}$ and $\widehat{\text{eff}}_{rmse}$ of MbP2 method in this scenario are superior to the other methods.

(iii) *Scenarios with* $\nu_{tr} = 20$ *and* $\rho_{tr} = 0.5, 0.9$ *and* $n = 100, 500$:

As the true parameter value of ν becomes higher, it is evidently more difficult to estimate the parameters since the generated data samples are close to normal and its likelihood become more even. In these scenarios we can observe high number of failures $(500 - r_g)$ of MbP methods in Table 4.5 as well as the large variability of the estimates in Figure 4.10.

In the scenario $n = 100$, $\nu_{tr} = 20$, $\rho_{tr} = 0.5$, under the 459 compared replicates, the MbP2 method has smaller numbers in the estimated statistics $\bar{\theta}$, $s(\bar{\theta})$, $\hat{\sigma}_r(\hat{\theta})$, $rb(\bar{\theta})$, $s_{rb}(\bar{\theta})$, eff_{mse} and eff_{rmse} for parameter ν in comparison to other methods. Even though these statistic numbers for MbP2 method in scenario $n = 100$, $\nu_{tr} = 20$, $\rho_{tr} = 0.9$ are even smaller and the number of failures becomes lower, we can see in Figure 4.10 that in both scenarios with sample size $n = 100$ and $n = 500$ the range between the first quartile and third quartile in the boxplot of MbP2 for estimated parameter $\hat{\nu}$ does not cover its true parameter value $\nu_{tr} = 20$. Moreover, the box-plots for parameter ρ show that MbP2 tends to overestimate the parameter and the 75% quartile range does not cover the true parameter value $\rho_{tr} = 0.9$ with sample size $n = 500$. Therefore MbP2 method is not recommendable in the case of $\nu_{tr} = 20$ and $\rho_{tr} = 0.9$

However, the MbP1 method proofs as the most stable and reliable estimation method in these scenarios with $\nu_{tr} = 20$, $\rho_{tr} = 0.9$ and $n = 100, 500$. The statistics numbers $\bar{\theta}$, $s(\bar{\theta})$, $\hat{\sigma}_r(\hat{\theta})$, $rb(\bar{\theta})$, $s_{rb}(\bar{\theta})$, eff_{mse} and eff_{rmse} for parameter ν illustrate that the estimates of MbP1 are less biased and less variable than these from MbP0, Direct, CSP and CMl methods. In the scenario with $n = 500$, both sample mean of the estimates $\hat{\nu}$ and $\hat{\rho}$ almost exactly represent their true parameter values while the statistics from other methods have a significant deflection.

4.7 Conclusions and Summary

In the parameter estimation of ν and ρ for the bivariate t-distribution we first note that the semi-parametric (CSP) and direct estimation method (CML) using the marginal empirical probability integral transformation are unstable and inaccurate estimation methods, the log likelihood values in all the scenarios are not at its maximum as displayed in the result tables. Compared to the MbP algorithm, the CSP method separately estimates the parameters in one step. Moreover, both CSP and CML methods include only few information about the marginal distributions due to the approximative nature of the empirical cumulative transformation, especially if the sample size is relatively small such as $n = 100$, the empirical transformation narrowly approaches true distribution of their margins. Therefore, both method are not recommendable.

Based on the estimated summary statistics in Table 4.3, 4.4 and 4.10 and illustrated boxplots in Figure 4.8, 4.9 and 4.10 we can observe that the correlation parameter ρ in the bivariate t-model is estimated accurately by all estimation methods in the most scenarios. The log likelihood values of MbP0, MbP1, MbP2 and Direct maximization methods do not have significant differences which means the MbP algorithm iteratively converges to the maximum likelihood estimators. Based on the estimated summary statistics we can conclude that the MbP algorithm performs stably and accurately and can be considered as an good alternative to the direct maximization method, even though the MbP algorithm has significant disadvantages in *computing time* and *the number of failures* especially if the true parameter values of the sampled data are high. In order to reduce the the number of failure we need to change the fixed parameter ρ_0 in the MbP algorithm adequately. In general the MbP algorithm with $\rho_0 := \hat{\rho}_\tau$ (MbP0) is recommendable for scenarios with low parameter value for ν. That is the case when the joint distributions have fat tails. If the joint distributions are close to normal, i.e the true parameter values are higher, the MbP algorithm with a different value for ρ_0 instead of $\rho_0 := \hat{\rho}_\tau$ is more likely to succeed in the estimations. Particularly, the MbP algorithm with $\rho_0 := \hat{\rho}_\tau + 0.05$ (MbP2) is suitable for scenarios with $\nu_{tr} = 10$ and the combination $\nu_{tr} = 20$, $\rho_{tr} = 0.5$. If both degrees of freedom and correlations are high, i.e. $\nu_{tr} = 20$ and $\rho_{tr} = 0.9$ the MbP algorithm with $\rho_0 := \hat{\rho}_\tau + 0.025$ (MbP1) is the most recommendable method on average.

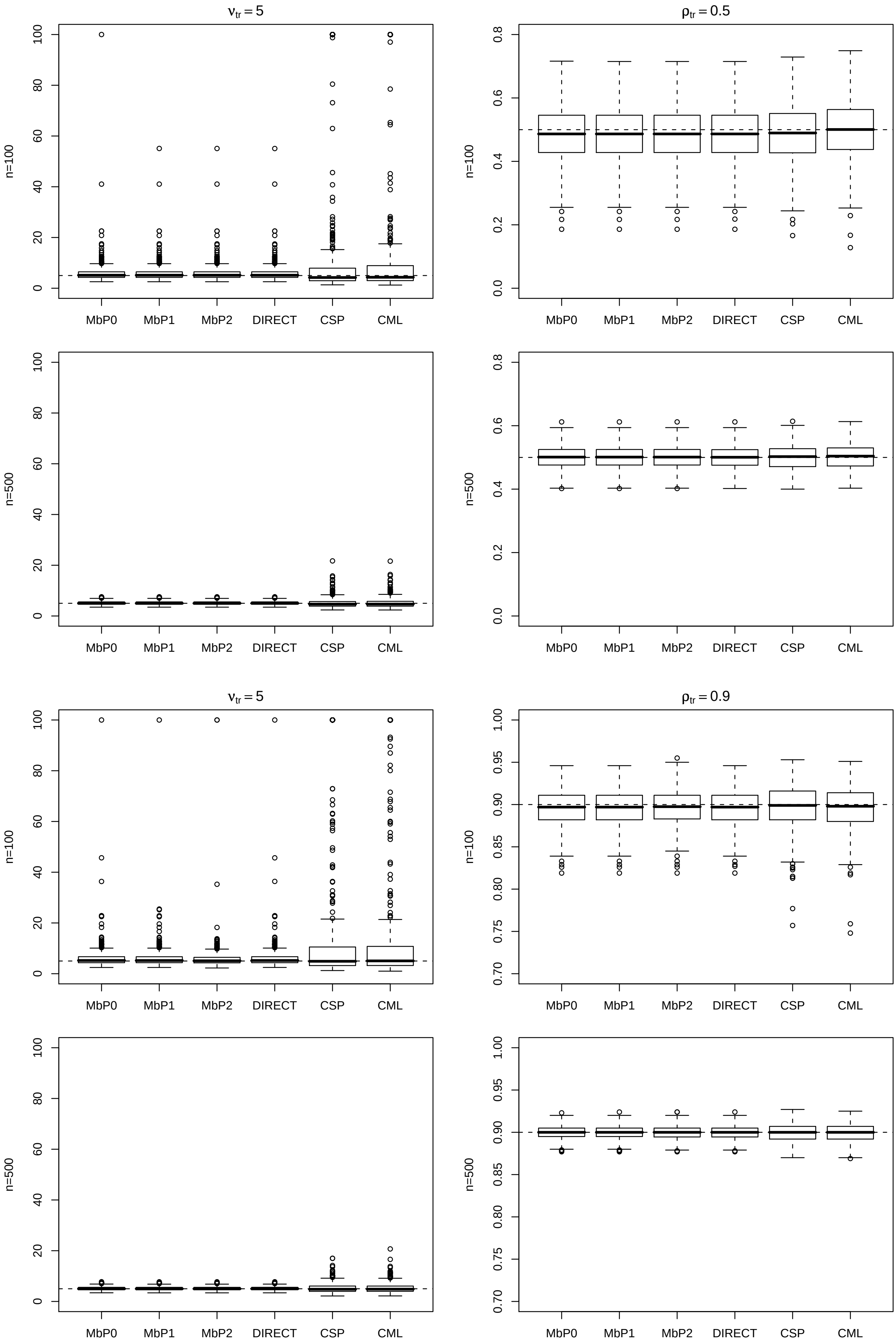


Figure 4.8: Boxplot of parameter estimates $(\hat{\nu}, \hat{\rho})$ for r_c samples when $\nu_{tr} = 5$ and $\rho_{tr} = 0.5, 0.9$.

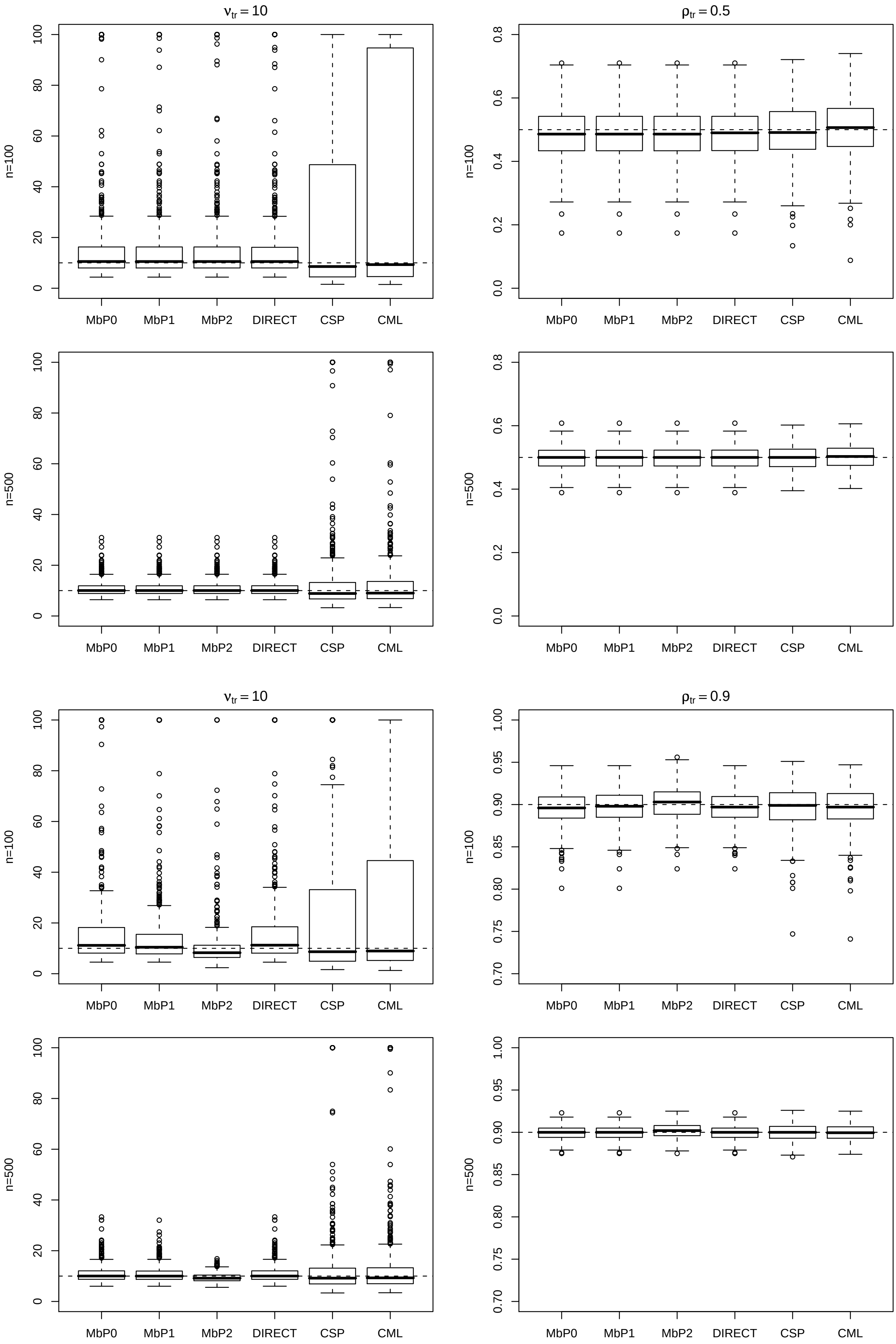


Figure 4.9: Boxplot of parameter estimates $(\hat{\nu}, \hat{\rho})$ for r_c samples when $\nu_{tr} = 10$ and $\rho_{tr} = 0.5, 0.9$.

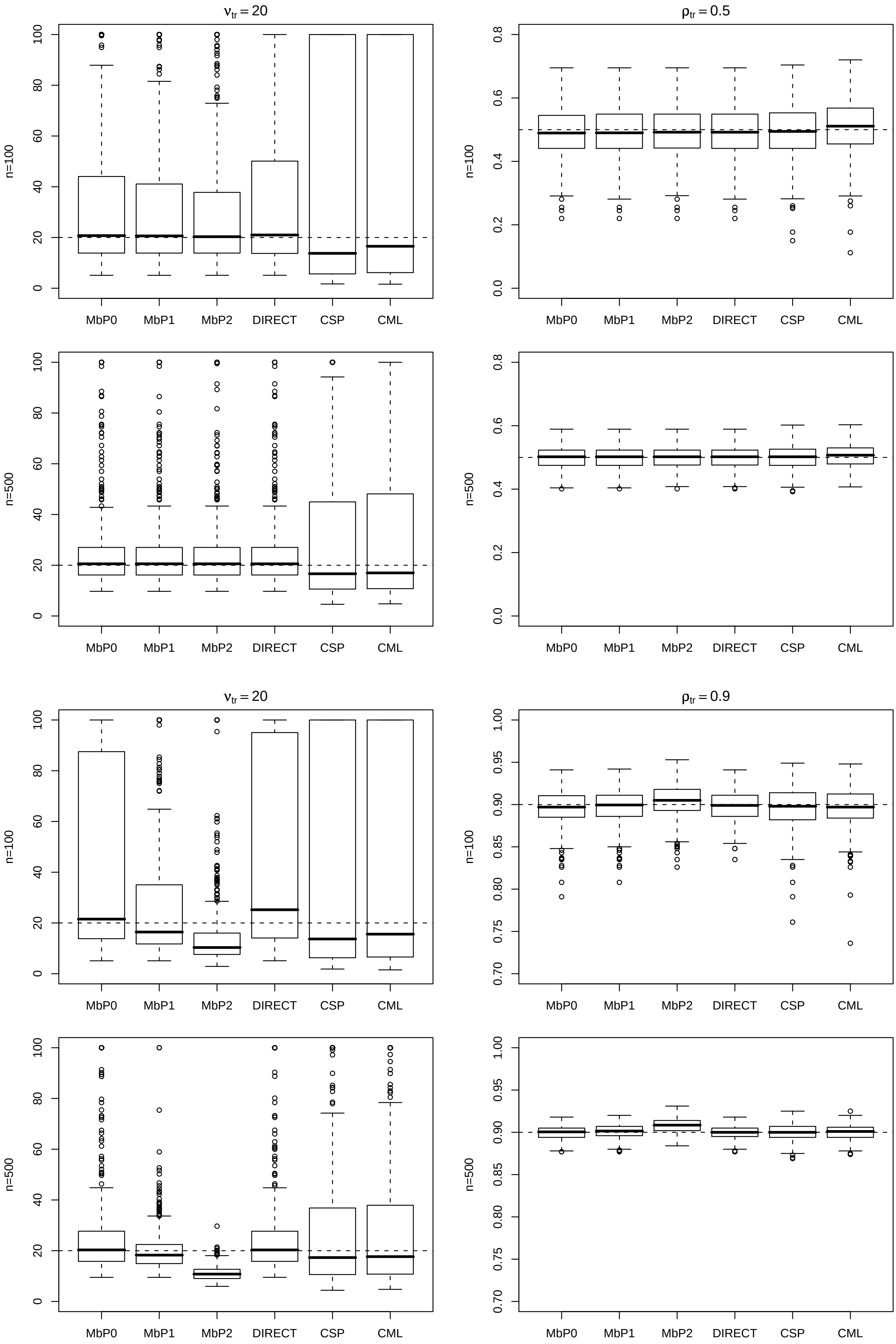


Figure 4.10: Boxplot of parameter estimates $(\hat{\nu}, \hat{\rho})$ for r_c samples when $\nu_{tr} = 20$ and $\rho_{tr} = 0.5, 0.9$.

Chapter 5

Maximum Likelihood Estimation for Bivariate Meta t-Data using Maximization by Parts (MbP)

5.1 MbP Partitions for Meta t-Distribution

Consider a bivariate meta t-distributed data set $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ of size n, $\boldsymbol{x}_j := (x_{1j}, ..., x_{nj})^t$ for $j = 1, 2$. As defined in (2.25), $h(\boldsymbol{x}; \nu_1, \nu_2, \nu, \rho)$ denotes the density function of bivariate meta t-distribution with marginal parameters ν_1 and ν_2 and copula parameters ν and ρ. Then the likelihood function of bivariate meta t-distribution is defined by:

$$
\begin{aligned}
L(\nu_1, \nu_2, \nu, \rho, \boldsymbol{x}) &:= \prod_{i=1}^{n} h_i(\boldsymbol{x}; \nu_1, \nu_2, \nu, \rho) \\
&= \prod_{i=1}^{n} \left(c(F_1(x_{i1}, \nu_1), F_2(x_{i2}, \nu_2); \nu, \rho) \prod_{j=1}^{2} f_j(x_{ij}; \nu_j) \right) \qquad (5.1)
\end{aligned}
$$

Then the log likelihood of bivariate meta t-distribution, denoted as $\ell(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x})$, is defined by

$$
\begin{aligned}
\ell(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}) &:= \sum_{i=1}^{n} \ln h_i(x_{i1}, x_{i2}; \nu_1, \nu_2, \nu, \rho) \\
&= \underbrace{\sum_{i=1}^{n} \ln c(F_1(x_{i1}, \nu_1), F_2(x_{i2}, \nu_2); \nu, \rho)}_{=:\ell_e(\nu_1,\nu_2,\nu,\rho;\boldsymbol{x}) \text{ copula error model}} + \underbrace{\sum_{i=1}^{n} \ln f(x_{i1}; \nu_1) + \ln f(x_{i2}; \nu_2)}_{=:\ell_w(\nu_1,\nu_2,\nu,\rho;\boldsymbol{x}) \text{ independence working model}} \quad , \qquad (5.2)
\end{aligned}
$$

which consists of the independence working model with two marginal distributions and the copula error model. In particular we define the following log likelihood functions as follows:

$$\ell_e(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}) := \sum_{i=1}^{n} \ln c(F_1(x_{i1}, \nu_1), F_2(x_{i2}, \nu_2); \nu, \rho) \tag{5.3}$$

$$\ell_m(\nu_j; \boldsymbol{x}_j) := \sum_{i=1}^{n} \ln f_j(x_{ij}; \nu_j), \quad j = 1, 2$$

$$\ell_w(\nu_1, \nu_2; \boldsymbol{x}) := \ell_m(\nu_1; \boldsymbol{x}_1) + \ell_m(\nu_2; \boldsymbol{x}_2) \tag{5.4}$$

As pointed out in Chapter 2, the convergence of the MbP algorithm hinges on a condition concerning the decomposition of the log-likelihood function. Since the independence working model $\ell_w(\cdot)$ from the previous decomposition represents only the univariate marginal distributions and does not contain any information about their dependence, one can hardly expect that the information dominance condition from Theorem 2.24 will be satisfied and MbP fails to work. Thus in order to achieve the information dominance we need to choose an additional term which amends the drawback of the independence working model and accounts for some degree of correlation between the marginals. Following the idea in Song et al. (2005) the proposed term, denoted as $\phi(\boldsymbol{x}; \nu_1, \nu_2, \nu_0, \rho_0)$, can be the density function of bivariate meta t-distribution with marginal parameters ν_1 and ν_2, interchangeable degrees of freedom ν_0 and correlation coefficient ρ_0. The function $\phi(\boldsymbol{x}; \nu_1, \nu_2, \nu_0, \rho_0)$ is given by

$$\phi(\boldsymbol{x}; \nu_1, \nu_2, \nu_0, \rho_0) := |R_0|^{-\frac{1}{2}} (1 + \frac{\boldsymbol{z}_i^t R_0^{-1} \boldsymbol{z}}{\nu_0})^{\frac{\nu_0+2}{2}} \frac{f(x_1; \nu_1) f(x_2; \nu_2)}{f(t_{\nu_0}^{-1}(F(x_1; \nu_1)); \nu_0) f(t_{\nu_0}^{-1}(F(x_2; \nu_2)); \nu_0)} \tag{5.5}$$

where $\boldsymbol{z} = (t_{\nu_0}^{-1}(F(x_1; \nu_1)), t_{\nu_0}^{-1}(F(x_2; \nu_2)))^t \in \mathbb{R}^2$ and $R_0 := \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$. The function $F(\cdot; \nu)$ denotes the univariate cumulative t-distribution function defined in (2.7).

With the additional term we can rewrite the log likelihood of the bivariate meta t-distribution as

$$\begin{aligned} \ell(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}) = & \underbrace{\sum_{i=1}^{n} \ln c(F_1(x_{i1}, \nu_1), F_2(x_{i2}, \nu_2); \nu, \rho) - \ln \phi_i(\boldsymbol{x}_i; \nu_1, \nu_2, \nu_0, \rho_0)}_{\text{modified error model}} \\ & \underbrace{+ \sum_{i=1}^{n} \ln f(x_{i1}; \nu_1) + \ln f(x_{i2}; \nu_2) + \ln \phi_i(\boldsymbol{x}_i; \nu_1, \nu_2, \nu_0, \rho_0)}_{\text{modified working model}} \end{aligned}$$

With the definition of the log likelihood of the addition model, i.e.

$$\begin{aligned} \ell_{new}(\nu_1, \nu_2; \nu_0, \rho_0; \boldsymbol{x}) :&= \sum_{i=1}^{n} \ln \phi_i(\boldsymbol{x}_i; \nu_1, \nu_2, \nu_0, \rho_0) \\ &= \ell_e(\nu_1, \nu_2, \nu_0, \rho_0; \boldsymbol{x}) + \ell_m(\nu_1; \boldsymbol{x}) + \ell_m(\nu_2; \boldsymbol{x}), \end{aligned}$$

we then have our modified working and error models composed of these log likelihood functions as follows:

$$\begin{aligned}\ell_{e,mod}(\nu_1,\nu_2,\nu,\rho;\nu_0,\rho_0;\boldsymbol{x}) &:= \ell_e(\nu_1,\nu_2;\nu,\rho;\boldsymbol{x}) - \ell_{new}(\nu_1,\nu_2;\nu_0,\rho_0;\boldsymbol{x})\\ &= \ell_e(\nu_1,\nu_2;\nu,\rho;\boldsymbol{x}) - \ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x}) - \ell_m(\nu_1;\boldsymbol{x}_1) - \ell_m(\nu_2;\boldsymbol{x}_2)\end{aligned} \tag{5.6}$$

$$\begin{aligned}\ell_{w,mod}(\nu_1,\nu_2;\nu_0,\rho_0;\boldsymbol{x}) &:= \ell_w(\nu_1,\nu_2;\boldsymbol{x}) + \ell_{new}(\nu_1,\nu_2;\nu_0,\rho_0;\boldsymbol{x})\\ &= \ell_m(\nu_1;\boldsymbol{x}_1) + \ell_m(\nu_2;\boldsymbol{x}_2) + \ell_{new}(\nu_1,\nu_2;\nu_0,\rho_0;\boldsymbol{x})\\ &= \ell_m(\nu_1;\boldsymbol{x}_1) + \ell_m(\nu_2;\boldsymbol{x}_2) + \ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x}) + \ell_m(\nu_1;\boldsymbol{x}_1) + \ell_m(\nu_2;\boldsymbol{x}_2)\\ &= 2\left(\ell_m(\nu_1;\boldsymbol{x}_1) + \ell_m(\nu_2;\boldsymbol{x}_2)\right) + \ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x})\end{aligned} \tag{5.7}$$

Let $\boldsymbol{\eta}_1 := (\nu_1,\nu_2)$ with $\boldsymbol{\eta}_2 := (\nu,\rho)$ and $\boldsymbol{\eta} := (\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$. The resulting score equations will be iteratively solved in the MbP algorithm for bivariate meta t-distributions:

$$\begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\eta}_1}\ell_{w,mod}(\nu_1,\nu_2;\nu_0,\rho_0;\boldsymbol{x}) + \frac{\partial}{\partial \boldsymbol{\eta}_1}\ell_{e,mod}(\nu_1,\nu_2,\nu,\rho;\nu_0,\rho_0;\boldsymbol{x}) \\ \frac{\partial}{\partial \boldsymbol{\eta}_2}\ell_{e,mod}(\nu_1,\nu_2,\nu,\rho;\nu_0,\rho_0;\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix} \tag{5.8}$$

Furthermore, we can rewrite the score equations in order to diminish the computational effort. The first order derivative of the modified working model and error model with respect to $\boldsymbol{\eta}_1$ can be expressed by:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\eta}_1}\ell_{w,mod}(\nu_1,\nu_2;\nu_0,\rho_0;\boldsymbol{x}) &\overset{(5.7)}{=} 2\frac{\partial}{\partial \boldsymbol{\eta}_1}\left(\ell_m(\nu_1;\boldsymbol{x}_1) + \ell_m(\nu_2;\boldsymbol{x}_2)\right) + \frac{\partial}{\partial \boldsymbol{\eta}_1}\ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x})\\ \frac{\partial}{\partial \boldsymbol{\eta}_1}\ell_{e,mod}(\nu_1,\nu_2,\nu,\rho;\nu_0,\rho_0;\boldsymbol{x}) &\overset{(5.6)}{=} \frac{\partial}{\partial \boldsymbol{\eta}_1}\ell_e(\nu_1,\nu_2;\nu,\rho;\boldsymbol{x}) - \frac{\partial}{\partial \boldsymbol{\eta}_1}\ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x})\\ &\quad - \frac{\partial}{\partial \boldsymbol{\eta}_1}\left(\ell_m(\nu_1;\boldsymbol{x}_1) + \ell_m(\nu_2;\boldsymbol{x}_2)\right)\end{aligned} \tag{5.9}$$

And the first order derivative of the modified error model with respect to $\boldsymbol{\eta}_2$ is the same as the first order derivative of the original error model, i.e.

$$\frac{\partial}{\partial \boldsymbol{\eta}_2}\ell_{e,mod}(\nu_1,\nu_2,\nu,\rho;\nu_0,\rho_0;\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{\eta}_2}\ell_e(\nu_1,\nu_2,\nu,\rho;\boldsymbol{x}). \tag{5.10}$$

We note that the modified working and error models and their derivatives only associate with the copula error model and the marginal model.

5.2 Derivatives of Working and Error Models

In this section, we assume that $\boldsymbol{x} := (\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ is bivariate meta t-distributed, $\boldsymbol{x}_j = (x_{1j},...,x_{nj})^t$ for $j = 1,2$.

5.2.1 Derivatives of the Quantile Function of t-Distribution

For solving the score equation of the log likelihood functions required in the MbP algorithm, we cannot avoid to calculate the following two derivatives:

$$\frac{\partial}{\partial \nu_j} t_\nu^{-1}(F(x_{ij}, \nu_j)), \quad j = 1, 2 \tag{5.11}$$

and

$$\frac{\partial}{\partial \nu} t_\nu^{-1}(F(x_{ij}, \nu_j)). \tag{5.12}$$

We can also rewrite both terms as:

$$\begin{aligned}
\frac{\partial}{\partial \nu_j} t_\nu^{-1}(F(x_{ij}, \nu_j)) &= \frac{\partial t_\nu^{-1}(F(x_{ij}, \nu_j))}{\partial F(x_{ij}, \nu_j)} \cdot \frac{\partial F(x_{ij}, \nu_j)}{\partial \nu_j} \\
&=: \frac{\partial t_\nu^{-1}(u)}{\partial u} \cdot \frac{\partial F(x_{ij}, \nu_j)}{\partial \nu_j} \qquad u := F(x_{ij}, \nu_j) \\
&= \frac{1}{f(c; \nu)} \cdot \frac{\partial F(x_{ij}, \nu_j)}{\partial \nu_j}, \qquad c := t_\nu^{-1}(u) \\
&= \frac{\frac{\partial}{\partial \nu_j} F(x_{ij}; \nu_j)}{f(t_\nu^{-1}(F(x_{ij}, \nu_j)); \nu)}
\end{aligned}$$

and

$$\frac{\partial}{\partial \nu} t_\nu^{-1}(F(x_{ij}, \nu_j)) = -\frac{\frac{\partial}{\partial \nu} F(t_\nu^{-1}(F(x_{ij}; \nu_j)); \nu)}{f(t_\nu^{-1}(F(x_{ij}; \nu_j)); \nu)}, \quad j = 1, 2$$

where t_ν^{-1} is the quantile function, $f(\cdot; \nu)$ and $F(\cdot; \nu)$ are the p.d.f. and c.d.f. of the univariate t-distribution with degrees of freedom ν defined in (2.8), (2.6) and (2.7), respectively.

In both cases we need to solve the first derivative of the c.d.f. of univariate t-distribution function with respect to ν, i.e. $\frac{\partial}{\partial \nu} F(\cdot; \nu)$, which does not have analytical expressions. Two numerical approaches of calculating the derivative are presented here:

1st Approach: by the means of Taylor series

The Taylor series of a real number or complex number function $f(x)$ that is infinitely differentiable function in a neighborhood of a real number or complex number number a, is the power series

$$f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f^{(3)}(a)}{3!}(x - a)^3 + \cdots$$

which in a more compact form can be written as $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n$ where $n!$ is the factorial of n and $f^{(n)}(a)$ denotes the nth derivative of f at the point a.

Similarly, we can write an arbitrary function $q : \mathbb{R} \to \mathbb{R}$ via the Taylor series as follows:

$$q(\nu + h) = q(\nu) + \frac{q'(\nu)}{1!} \cdot h + \frac{q''(\nu)}{2!} \cdot h^2 + O(h^3)$$
$$q(\nu - h) = q(\nu) - \frac{q'(\nu)}{1!} \cdot h + \frac{q''(\nu)}{2!} \cdot h^2 + O(h^3).$$

It follows that $q(\nu + h) - q(\nu - h) = 2h \cdot q'(\nu) + O(h^3)$. After dividing both sides of the equation by $2h$, we obtain

$$q'(\nu) + O(h^2) = \frac{q(\nu + h) - q(\nu - h)}{2h}.$$

For $h \to 0$, $O(h^2)$ converges to zero at a rate of order 2. The convergence order shows that the given differential quotient,

$$\frac{q(\nu + h) - q(\nu - h)}{2h},$$

is a good approximation for the derivative. Moreover, for solving the function $q(\cdot)$, we employ the equivalent point arrangement in an interval [a,b] by setting $h := \dfrac{b-a}{N}$ with N denotes the number of the interval between a and b. For a large N, h converges to zero.

Remarks:

(i) With this approach, we can skip the calculation of $\frac{\partial}{\partial \nu} F(\cdot; \nu)$ and go direct to the calculation of the derivative of the t-quantile function with respect to ν, i.e. $\frac{\partial}{\partial \nu} t_\nu^{-1}(x)$.

(ii) Practical simulations show that we can even adopt this method to calculate the derivatives of the working and error models directly. However, we have to take a good care of handling the differential quotient because of the closed interval for the parameter of interest.

2nd Approach: Numerical Integration

An alternative way to calculate $\frac{\partial}{\partial \nu} F(\cdot; \nu)$ is to integrate part of the c.d.f. after some appropriate transformations. The calculation steps are derived in the following way:

Since $\dfrac{\partial}{\partial \nu} \ln(f(t; \nu)) = \dfrac{1}{f(t; \nu)} \dfrac{\partial f(t; \nu)}{\partial \nu}$, we have

$$\frac{\partial}{\partial \nu} F(x; \nu) = \frac{\partial}{\partial \nu} \int_{-\infty}^{x} f(t; \nu) dt = \int_{-\infty}^{x} \frac{\partial}{\partial \nu} \ln(f(t; \nu)) \cdot f(t; \nu) dt$$

In addition the function $\dfrac{\partial}{\partial \nu} \ln f(t; \nu)$ can be expressed as:

$$\frac{\partial}{\partial \nu} \ln f(t; \nu) = \frac{1}{2} \psi(\frac{\nu + 1}{2}) - \frac{1}{2\nu} - \frac{1}{2} \psi(\frac{\nu}{2}) - \frac{1}{2} \ln(1 + \frac{t^2}{\nu}) + \frac{\nu + 1}{2\nu} \frac{t^2}{\nu + t^2}$$

With the further definitions

$$g_1(\nu) := \frac{1}{2} \cdot \left(\psi(\frac{\nu+1}{2}) - \psi(\frac{\nu}{2}) - \frac{1}{\nu} \right)$$

and

$$g_2(t,\nu) := -\frac{1}{2}\ln(1+\frac{t^2}{\nu}) + \frac{\nu+1}{2\nu} \cdot \frac{t^2}{\nu+t^2},$$

we obtain

$$\frac{\partial}{\partial \nu} F(x,\nu) = \int_{-\infty}^{x} g_1(\nu) f(t,\nu) dt + \int_{-\infty}^{x} g_2(t,\nu) f(t,\nu) dt$$

Moreover,

$$\int_{-\infty}^{x} g_1(\nu) f(t,\nu) dt = g(\nu) \cdot \int_{-\infty}^{x} f(t,\nu) dt = g_1(\nu) \cdot F(x,\nu)$$

$$\begin{aligned}
\int_{-\infty}^{x} g_2(t,\nu) f(t,\nu) dt &= \int_{-\infty}^{x} g_2(\nu) dF(t,\nu) \quad \text{Riemann-Integral} \\
&= \Big[g_2(t,\nu) \cdot F(t,\nu) \Big]_{-\infty}^{x} - \int_{-\infty}^{x} F(t,\nu) dg_2(t,\nu) \\
&= g_2(x,\nu) \cdot F(x,\nu) - \int_{-\infty}^{x} F(t,\nu) \cdot \big[\frac{\partial}{\partial t} g_2(t,\nu) \big] dt \\
&= g_2(x,\nu) \cdot F(x,\nu) - \int_{-\infty}^{x} F(t,\nu) \cdot \big[-\frac{t}{\nu+t^2} + \frac{(\nu+1)t}{(\nu+t^2)^2} \big] dt
\end{aligned}$$

All together, we receive

$$\begin{aligned}
\frac{\partial}{\partial \nu} F(x,\nu) &= g_1(\nu) \cdot F(x,\nu) + g_2(x,\nu) \cdot F(x,\nu) - \int_{-\infty}^{x} F(t,\nu) \cdot \big[-\frac{t}{\nu+t^2} + \frac{(\nu+1)t}{(\nu+t^2)^2} \big] dt \\
&= \frac{\partial}{\partial \nu} \ln f(x,\nu) \cdot F(x,\nu) - \int_{-\infty}^{x} F(t,\nu) \cdot \big[-\frac{t}{\nu+t^2} + \frac{(\nu+1)t}{(\nu+t^2)^2} \big] dt
\end{aligned}$$

The last integral has no analytical expression and can be calculated using Gauss-Quadrature, specifically the Gauss-Laguerre method. In R, the base function `integrate(...)` can be utilized for the one dimensional integration.

Remark: In comparison to the first approach, the second approach is more time consuming and unstable.

5.2.2 Derivatives of the Modified Working and Error Model

To simplify the expressions of the derivatives of the modified working and error models, we begin with some definitions:

$$A(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}_i) := \boldsymbol{z}_i^t R^{-1} \boldsymbol{z}_i \tag{5.13}$$

$$= \frac{1}{1-\rho^2}\left(z_{i1}^2(\nu_1, \nu; x_{i1}) + z_{i2}^2(\nu_2, \nu; x_{i2}) - 2z_{i1}(\nu_1, \nu; x_{i1}) z_{i2}(\nu_2, \nu; x_{i2})\rho\right)$$

$$B(\nu_j, \nu; x_{ij}) := z_{ij}^2(\nu_j, \nu; x_{ij}), \quad j = 1, 2 \tag{5.14}$$

where $\boldsymbol{z}_i := (z_{i1}(\nu_1, \nu; x_{i1}), z_{i2}(\nu_2, \nu; x_{i2}))^t := (t_\nu^{-1}(F(x_{i1}, \nu_1)), t_\nu^{-1}(F(x_{i2}, \nu_2)))^t$.

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

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

$$\frac{\partial B(\nu_j, \nu; x_{ij})}{\partial \nu_j} = 2z_{ij}(\nu_j, \nu; x_{ij})\frac{\partial}{\partial \nu_j} z_{ij}(\nu_j, \nu; x_{ij})$$

where $k \neq j$, $k, j = 1, 2$.

Let $z_{ij} := z_{ij}(\nu_j, \nu; x_{ij})$ for $j = 1, 2$

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

$$= \frac{1}{1-\rho^2}\left(2\sum_{j=1}^{2} z_{ij}\frac{\partial z_{ij}}{\partial \nu} - 2\rho \sum_{j\neq k; j,k=1}^{2} z_{ik}\frac{\partial z_{ij}}{\partial \nu}\right)$$

$$\frac{\partial B(\nu_j, \nu; x_{ij})}{\partial \nu} = 2z_{ij}\frac{\partial z_{ij}}{\partial \nu}$$

where

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

By the same token,

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

$$= \frac{-2z_{i1}z_{i2} + 2z_{i1}z_{i2}\rho^2 + 2\rho(z_{i1}^2 + z_{i2}^2) - 4z_{i1}z_{i2}\rho^2}{(1-\rho^2)^2}$$

$$= \frac{-2z_{i1}z_{i2}(1+\rho^2) + 2\rho(z_{i1}^2 + z_{i2}^2)}{(1-\rho^2)^2}$$

Henceforth, the analytic expressions for the score equations of the working and error models can be expressed in the following way:

The Marginal Model and its Derivatives

With the density function of univariate t-Distribution defined in (2.6) we have the marginal model and its derivatives with respect to ν_j, $j = 1, 2$, as follows:

$$\begin{aligned}\ell_m(\nu_j; \boldsymbol{x}_j) &:= \sum_{i=1}^{n} \ln f(x_{ij}; \nu) \quad \text{for } j = 1, 2\\ &:= \sum_{i=1}^{n} \ln \left(\frac{\Gamma(\frac{\nu_j+1}{2})}{\sqrt{\nu_j \pi}\Gamma(\frac{\nu_j}{2})} \left(1 + \frac{x^2}{\nu_j}\right)^{-\frac{\nu_j+1}{2}} \right)\\ &= n \ln \Gamma(\frac{\nu_j+1}{2}) - \frac{n}{2} \ln(\nu_j \pi) - n \ln \Gamma(\frac{\nu_j}{2}) - \frac{\nu_j+1}{2} \sum_{i=1}^{n} \ln \left(1 + \frac{x_{ij}^2}{\nu_j}\right) \qquad (5.15)\end{aligned}$$

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

The Error Model and its Derivatives

With the density function of the bivariate t-Copula defined in (2.12) the error model is given by:

$$\begin{aligned}\ell_e(\nu_1, \nu_2, \nu, \rho) &:= \sum_{i=1}^{n} \ln c(F(x_{i1}; \nu), F(x_{i2}; \nu); \nu, \rho)\\ &= \sum_{i=1}^{n} \ln \left(\frac{\frac{\nu}{2} \cdot \Gamma(\frac{\nu}{2})^2 \left(1 + \frac{z_1^2+z_2^2-2z_1z_2\rho}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu+1}{2})^2 \sqrt{1-\rho^2} (1 + \frac{z_1^2}{\nu})^{-\frac{\nu+1}{2}} (1 + \frac{z_2^2}{\nu})^{-\frac{\nu+1}{2}}} \right)\\ &= 2n \ln \Gamma(\frac{\nu}{2}) + n \ln(\frac{\nu}{2}) - 2n \ln \Gamma(\frac{\nu+1}{2}) - \frac{n}{2} \ln(1-\rho^2)\\ &\quad - \frac{\nu+2}{2} \sum_{i=1}^{n} \ln \left(1 + \frac{A(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}_i)}{\nu}\right) + \frac{\nu+1}{2} \sum_{i=1}^{n} \sum_{j=1}^{2} \ln \left(1 + \frac{B(\nu_j, \nu, x_{ij})}{\nu}\right)\end{aligned} \qquad (5.16)$$

with $A(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}_i)$ and $B(\nu_j, \nu, x_{ij})$, $j = 1, 2$, defined in (5.13). The first order derivatives of the error model with respect to the parameters ν_1, ν_2, ν and ρ are calculated as follows:

$$\begin{aligned}\frac{\partial}{\partial \nu_j} \ell_e(\nu_1, \nu_2, \nu, \rho) = &-\frac{\nu+2}{2} \sum_{i=1}^{n} \frac{1}{1 + \frac{A(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}_i)}{\nu}} \cdot \frac{\partial}{\partial \nu_j} A(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}_i) \frac{1}{\nu}\\ &+ \frac{\nu+1}{2} \sum_{i=1}^{n} \frac{1}{1 + \frac{B(\nu_1, \nu; x_{ij})}{\nu}} \cdot \frac{\partial}{\partial \nu_j} B(\nu_1, \nu; x_{ij}) \frac{1}{\nu}\end{aligned}$$

Let $A(\cdot) := A(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}_i)$, then we receive

$$\begin{aligned}
\frac{\partial}{\partial \nu}\ell_e(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}) &= \frac{n}{\nu} + n\psi(\frac{\nu}{2}) - n\psi(\frac{\nu+1}{2}) \\
&- \left[\frac{1}{2}(\sum_{i=1}^{n} \ln(1 + \frac{A(\cdot)}{\nu}) + \frac{\nu+2}{2}\sum_{i=1}^{n}\frac{1}{1 + A(\cdot)/\nu}\cdot\frac{\partial}{\partial \nu}(\frac{A(\cdot)}{\nu}))\right] \\
&+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}(\ln(1 + \frac{B(\nu_j, \nu; x_{ij})}{\nu})) \\
&+ \frac{\nu+1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}\frac{1}{1 + B(\nu_j, \nu; x_{ij})/\nu}\cdot\frac{\partial}{\partial \nu}(\frac{B(\nu_j, \nu; x_{ij})}{\nu}) \\
&= \frac{n}{\nu} + n\psi(\frac{\nu}{2}) - n\psi(\frac{\nu+1}{2}) \\
&- \left[\frac{1}{2}(\sum_{i=1}^{n} \ln(1 + \frac{A(\cdot)}{\nu}) + \frac{\nu+2}{2\nu}\sum_{i=1}^{n}\frac{\frac{\partial}{\partial \nu}A(\cdot)\cdot\nu - A(\cdot)}{\nu + A(\cdot)})\right] \\
&+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}(\ln(1 + \frac{B(\nu_j, \nu; x_{ij})}{\nu})) \\
&+ \frac{\nu+1}{2\nu}\sum_{i=1}^{n}\sum_{j=1}^{2}\frac{\frac{\partial}{\partial \nu}B(\nu_j, \nu; x_{ij})\cdot\nu - B(\nu_j, \nu; x_{ij})}{\nu + B(\nu_j, \nu; x_{ij})}
\end{aligned}$$

whereby

$$\frac{\partial}{\partial \nu}\left(\frac{A(\cdot)}{\nu}\right) = \frac{\frac{\partial}{\partial \nu}A(\cdot)\cdot\nu - A(\cdot)}{\nu^2}.$$

and it follows similarly that

$$\frac{\partial}{\partial \nu}\left(\frac{B(\nu_j, \nu; x_{ij})}{\nu}\right) = \frac{\frac{\partial}{\partial \nu}B(\nu_j, \nu; x_{ij})\cdot\nu - B(\nu_j, \nu; x_{ij})}{\nu^2}, \quad j = 1, 2$$

and

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

5.3 The MbP Algorithm for the Bivariate Meta t-Distribution

5.3.1 Framework of the MbP Algorithm for Bivariate Meta t-Distribution

Let $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ be a bivariate meta t-distributed data set of size n with $\boldsymbol{x}_j := (x_{1j}, ..., x_{nj})^t$ for $j = 1, 2$. In the MbP algorithm for bivariate meta t-distribution we are going to solve the following score functions as given by (5.8).

Given the initial values $(\nu_{1,n}^0, \nu_{2,n}^0, \nu_n^0, \rho_n^0)$ and using the extended forms of the score equations stated in (5.9) and (5.10), the framework of the MbP algorithm states as follows:

STEP k.1, $k = 1, 2, ...$
Find the root of the equation

$$\begin{pmatrix} 2\frac{\partial}{\partial\nu_1}\ell_m(\nu_1;\boldsymbol{x}_1) + \frac{\partial}{\partial\nu_1}\ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x}) \\ 2\frac{\partial}{\partial\nu_2}\ell_m(\nu_2;\boldsymbol{x}_2) + \frac{\partial}{\partial\nu_2}\ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x}) \end{pmatrix} + \boldsymbol{C}(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu_n^{k-1}, \rho_n^{k-1}, \nu_0, \rho_0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where the $\boldsymbol{C}(\cdot)$ is a function depending on $\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu_n^{k-1}, \rho_n^{k-1}, \nu_0, \rho_0$. In particular,

$$\boldsymbol{C}(\cdot) := \begin{pmatrix} \frac{\partial}{\partial\nu_1}\ell_e(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu_n^{k-1}, \rho_n^{k-1};\boldsymbol{x}) - \frac{\partial}{\partial\nu_1}\ell_e(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu_0, \rho_0;\boldsymbol{x}) - \frac{\partial}{\partial\nu_1}\ell_m(\nu_{1,n}^{k-1};\boldsymbol{x}_1) \\ \frac{\partial}{\partial\nu_2}\ell_e(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu_n^{k-1}, \rho_n^{k-1};\boldsymbol{x}) - \frac{\partial}{\partial\nu_2}\ell_e(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu_0, \rho_0;\boldsymbol{x}) - \frac{\partial}{\partial\nu_2}\ell_m(\nu_{2,n}^{k-1};\boldsymbol{x}_2) \end{pmatrix} \tag{5.17}$$

The result of STEP k.1 is denoted as $(\nu_{1,n}^k, \nu_{2,n}^k)$.

STEP k.2
Find the root of the equation

$$\begin{pmatrix} \frac{\partial}{\partial\nu}\ell_e(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu, \rho;\boldsymbol{x}) \\ \frac{\partial}{\partial\rho}\ell_e(\nu_{1,n}^{k-1}, \nu_{1,n}^{k-1}, \nu, \rho;\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The result of STEP k.2 is indicated as (ν_n^k, ρ_n^k).

5.3.2 Choices of Interchangeable Parameters ν_0 and ρ_0

We can set the interchangeable Parameters ν_0 and ρ_0 in two ways - the non-adaptive and adaptive way. Thus it results two versions of the MbP algorithm:

MBP1: Non-adaptive

If we hold the interchangeable fixed during the entire MbP iterations, i.e. $\nu_0 := \nu^{\text{fix}}$, $\rho_0 := \rho^{\text{fix}}$, then the STEP k.1 of the MbP algorithm operates as follows:

STEP k.1, $k = 1, 2, ...$
Find the root $(\nu_{1,n}^k, \nu_{2,n}^k)$ of the equation

$$\begin{pmatrix} 2\frac{\partial}{\partial \nu_1}\ell_m(\nu_1;\boldsymbol{x}_1) + \frac{\partial}{\partial \nu_1}\ell_e(\nu_1,\nu_2,\nu^{\text{fix}},\rho^{\text{fix}};\boldsymbol{x}) \\ 2\frac{\partial}{\partial \nu_2}\ell_m(\nu_2;\boldsymbol{x}_2) + \frac{\partial}{\partial \nu_2}\ell_e(\nu_1,\nu_2,\nu^{\text{fix}},\rho^{\text{fix}};\boldsymbol{x}) \end{pmatrix} + \boldsymbol{C}(\nu_{1,n}^{k-1},\nu_{2,n}^{k-1},\nu_n^{k-1},\rho_n^{k-1},\nu^{\text{fix}},\rho^{\text{fix}}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where

$$\boldsymbol{C}(\cdot) := \begin{pmatrix} \frac{\partial}{\partial \nu_1}\ell_e(\nu_{1,n}^{k-1},\nu_{2,n}^{k-1},\nu_n^{k-1},\rho_n^{k-1};\boldsymbol{x}) - \frac{\partial}{\partial \nu_1}\ell_e(\nu_{1,n}^{k-1},\nu_{2,n}^{k-1},\nu^{\text{fix}},\rho^{\text{fix}};\boldsymbol{x}) - \frac{\partial}{\partial \nu_1}\ell_m(\nu_{1,n}^{k-1};\boldsymbol{x}_1) \\ \frac{\partial}{\partial \nu_2}\ell_e(\nu_{1,n}^{k-1},\nu_{2,n}^{k-1},\nu_n^{k-1},\rho_n^{k-1};\boldsymbol{x}) - \frac{\partial}{\partial \nu_2}\ell_e(\nu_{1,n}^{k-1},\nu_{2,n}^{k-1},\nu^{\text{fix}},\rho^{\text{fix}};\boldsymbol{x}) - \frac{\partial}{\partial \nu_2}\ell_m(\nu_{2,n}^{k-1};\boldsymbol{x}_2) \end{pmatrix}$$

MBP2: Adaptive

If we let the values of ν_0 and ρ_0 be updated in each MbP iterations by the estimated parameters from previous step, i.e. $\nu_0 := \nu_n^{k-1}$, $\rho_0 := \rho_n^{k-1}$, then the STEP k.1 of the MbP algorithm appears as:

STEP k.1, $k = 1, 2, ...$
Find the root $(\nu_{1,n}^k, \nu_{2,n}^k)$ of the equation

$$\begin{pmatrix} 2\frac{\partial}{\partial \nu_1}\ell_m(\nu_1;\boldsymbol{x}_1) + \frac{\partial}{\partial \nu_1}\ell_e(\nu_1,\nu_2,\nu_n^{k-1},\rho_n^{k-1};\boldsymbol{x}) \\ 2\frac{\partial}{\partial \nu_2}\ell_m(\nu_2;\boldsymbol{x}_2) + \frac{\partial}{\partial \nu_2}\ell_e(\nu_1,\nu_2,\nu_n^{k-1},\rho_n^{k-1};\boldsymbol{x}) \end{pmatrix} + \boldsymbol{C}(\nu_{1,n}^{k-1},\nu_{2,n}^{k-1}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where

$$\boldsymbol{C}(\cdot) := -\begin{pmatrix} \frac{\partial}{\partial \nu_1}\ell_m(\nu_{1,n}^{k-1};\boldsymbol{x}_1) \\ \frac{\partial}{\partial \nu_2}\ell_m(\nu_{2,n}^{k-1};\boldsymbol{x}_2) \end{pmatrix}$$

Remark: We can also utilize a mixed structure for the interchangeable variables, i.e. $\nu_0 := \nu_n^{k-1}$ and $\rho_0 := \rho^{\text{fix}}$ or vice versa. However, the result after several simulation is not satisfied, since the likelihood values does not converge to its maximum. Thus in the following work, this mixed case will not be analyzed.

Chapter 6

Small Sample Properties of the MbP Algorithm and its Competitors for Bivariate Meta t-Data

6.1 Data and the Log Likelihood

6.1.1 Generation of bivariate Meta t-Data

To generate bivariate meta t-distributed data we can perform a two-step procedure as follows:

- Generate bivariate n-dimensional uniform distributed data $\boldsymbol{U} := (\boldsymbol{U}_1, \boldsymbol{U}_2) \in [0,1]^{n\times 2}$, $\boldsymbol{U}_j := (U_{1j}, ..., U_{nj})^t$ for $j = 1, 2$, jointly from the bivariate t-Copula distribution with known degree of freedom and correlation coefficient ν_{tr} and ρ_{tr}. In R, the procedure can be achieved using the function `tCopula` from the R package `copula`. Particularly,

$$\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2) \sim \texttt{tCopula}(\rho_{tr}, \texttt{dim} = 2, \texttt{df} = \nu_{tr}). \tag{6.1}$$

- Inversely transform the generated data $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2)$ into bivariate meta t-data using the t-quantile function with known degrees of freedom $\nu_{1,tr}$ and $\nu_{2,tr}$ of the marginal distributions, i.e. $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) = (t^{-1}_{\nu_{1,tr}}(\boldsymbol{u}_1), t^{-1}_{\nu_{2,tr}}(\boldsymbol{u}_2))$ and thus

$$\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \sim MT_2(\nu_{1,tr}, \nu_{2,tr}, \nu_{tr}, \rho_{tr})$$

  with $\boldsymbol{x}_j := (x_{1j}, ..., x_{nj})^t$ for $j = 1, 2$.

6.1.2 The Log Likelihood

Let $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ be the bivariate meta t-data of size n. With the decomposition defined in (2.26) the log likelihood of the bivariate meta t-distribution is given by:

$$\ell(\boldsymbol{x}; \nu_1, \nu_2, \nu, \rho) := \ell_e(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x}) + \ell_m(\nu_1; \boldsymbol{x}_1) + \ell_m(\nu_2; \boldsymbol{x}_2) \tag{6.2}$$

with error model $\ell_e(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x})$ and marginal model $\ell_m(\nu_j; \boldsymbol{x}_j)$ defined in (5.16) and (5.15).

6.2 General Setting for the MbP Algorithm in the Simulation Study

Denote the parameter set of interest $\boldsymbol{\eta} := (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ with $\boldsymbol{\eta}_1 := (\nu_1, \nu_2)$ and $\boldsymbol{\eta}_2 := (\nu, \rho)$.

1) Initial Settings for the MbP algorithm

- Input: bivariate meta t-data $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$
- Input: Initial values: $(\nu_1^0, \nu_2^0, \nu^0, \rho^0)$
- Parameter space $\Theta := \{\boldsymbol{\eta} := (\nu_1, \nu_2, \nu, \rho) | \nu_1, \nu_2, \nu \in [1, 100] \text{ and } \rho \in (-1, 1)\}$
- Maximum number of iterations: $k = 50$
- Convergence criteria: $\| \boldsymbol{\eta}_n^k - \boldsymbol{\eta}_n^{k-1} \|_2 < 10^{-3}$, where $\| \cdot \|_2$ is the Euclidean norm.

Furthermore we investigate two versions of MbP algorithm in the simulation study:

(i) **MBP1**: non-adaptive with fixed ν_0 and ρ_0

(ii) **MBP2**: adaptive method with updated ν_0 and ρ_0 in each MbP step

2) MbP Steps:

The Quasi-Newton method with box constraints (L-BFGS-B) is used in the R function `optim` for the non-adaptive MbP algorithm (MBP1), the adaptive MbP algorithm (MBP2) and direct maximization method (DIR). The interval for parameters for degrees of freedom ν_1, ν_2 and ν is in this case constraint to $[1, 100]$. If the `optim` function with `L-BFGS-B` method either produces an error message or fails to converge (that is the case if the indicator `convergence` in the `optim` is not equal 0.) within the parameter space $\Theta := \{(\nu_1, \nu_2, \nu, \rho) : \nu_1, \nu_2, \nu \in [1, 100] \text{ and } \rho \in (-1, 1)\}$, the implemented MbP algorithm will reset and run the iterations using `optim` with the Quasi-Newton method without box constraints (BFGS) method. Since upper limit of the searching area is unconstraint in this method, the algorithm may find estimates that are far beyond 100 for parameter ν, for instance 100000. Eventually, if the algorithm with BFGS method still generates error messages, the implemented algorithm will return 100 for parameter ν and $0.999 \cdot \text{sign}(\rho_0)$ for parameter ρ, where $\text{sign}(\rho_0)$ indicates the sign of the interchangeable parameter ρ_0. Adapting both `L-BFGS-B` and `BFGS` methods the framework of the MbP steps states as follows:

- `optim` using Quasi-Newton method with box constraints

  STEP k.1 Following the described MbP algorithm in the last section, we find the root of the following score equation for (ν_1, ν_2):

$$\begin{pmatrix} 2\frac{\partial}{\partial \nu_1}\ell_m(\nu_1; \boldsymbol{x}_1) + \frac{\partial}{\partial \nu_1}\ell_e(\nu_1, \nu_2, \nu_0, \rho_0; \boldsymbol{x}) \\ 2\frac{\partial}{\partial \nu_2}\ell_m(\nu_2; \boldsymbol{x}_2) + \frac{\partial}{\partial \nu_2}\ell_e(\nu_1, \nu_2, \nu_0, \rho_0; \boldsymbol{x}) \end{pmatrix} + \boldsymbol{C}(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu_n^{k-1}, \rho_n^{k-1}, \nu_0, \rho_0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

with $\boldsymbol{C}(\cdot)$ as defined in (5.17). Moreover, let denote the first element of $\boldsymbol{C}(\cdot)$ as $\boldsymbol{C}_a(\cdot)$ and second element as $\boldsymbol{C}_b(\cdot)$, i.e. $\boldsymbol{C}(\cdot) = (\boldsymbol{C}_a(\cdot), \boldsymbol{C}_b(\cdot))^t$. Then we calculate the root of the score equation above by minimizing sum of the squared elements of the score equation. The result is demonstrated as follows:

$$(\nu_{1,n}^k, \nu_{2,n}^k) = \text{argmin}\{(s_{1,a}(\nu_1,\nu_2))^2 + (s_{1,b}(\nu_1,\nu_2))^2 | \nu_1, \nu_2 \in [1,100]\} \tag{6.3}$$

where

$$\begin{aligned} s_{1,a}(\nu_1,\nu_2) &:= 2\frac{\partial}{\partial \nu_1}\ell_m(\nu_1;\boldsymbol{x}_1) + \frac{\partial}{\partial \nu_1}\ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x}) + \boldsymbol{C}_a(\cdot), \\ s_{1,b}(\nu_1,\nu_2) &:= 2\frac{\partial}{\partial \nu_2}\ell_m(\nu_2;\boldsymbol{x}_2) + \frac{\partial}{\partial \nu_2}\ell_e(\nu_1,\nu_2,\nu_0,\rho_0;\boldsymbol{x}) + \boldsymbol{C}_b(\cdot). \end{aligned}$$

To process (6.3) we first utilize the R-function `optim` using primarily the *Quasi-Newton Method with box constraints*, abbr. `L-BFGS-B`, that operates with limited interval for the parameters of interest $(\nu_1,\nu_2) \in [1,100] \times [1,100]$.

STEP k.2 In this step we find the root of the equation

$$\begin{pmatrix} \frac{\partial}{\partial \nu}\ell_e(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu, \rho; \boldsymbol{x}) \\ \frac{\partial}{\partial \rho}\ell_e(\nu_{1,n}^{k-1}, \nu_{2,n}^{k-1}, \nu, \rho; \boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

by maximizing the log likelihood of the copula error model $\ell_e(\boldsymbol{\eta}_{1,n}^{k-1}, \boldsymbol{\eta}_2; \boldsymbol{x})$ or minimizing the negative log likelihood. We have again two parameters $\nu \in [1,100]$ and $\rho \in (-1,1)$ for this step.

- `optim` using Quasi-Newton method without box constraints

  If the algorithm fails to converge, meaning that the `optim`-procedure was terminated and delivers an error message, we will delete all the previous results calculated by `L-BFGS-B` method and adopt the *Quasi-Newton method without box constraints*, abbr. `BFGS` method, for **STEP k.1** and **STEP k.2**, $k = 1, 2, ...$ Since there are no defined limits for the parameters of interest, we need to undertake a parameter transformation such as the fisher's z-transformation for the BFGS method. The Fisher's z-transformation operates as follows:

$$\begin{aligned} \nu_\gamma &= \ln(\nu - 1) \in \mathbb{R} \\ \rho_\gamma &= \ln\left(\frac{1+\rho}{1-\rho}\right) \in \mathbb{R} \end{aligned} \tag{6.4}$$

  with the reverse transformation given by

$$\begin{aligned} \nu &= 1 + e^{\nu_\gamma} \in [1,\infty] \quad \text{and} \\ \rho &= \frac{e^{\rho_\gamma - 1}}{e^{\rho_\gamma + 1}} \in (-1,1) \end{aligned}$$

The flow chart of the MbP algorithm and the description of the convergence indictor FF are given in Figure 6.1 and Table 6.1.

Flow Chart of MbP-Algorithm for Bivariate Meta t-Distributions

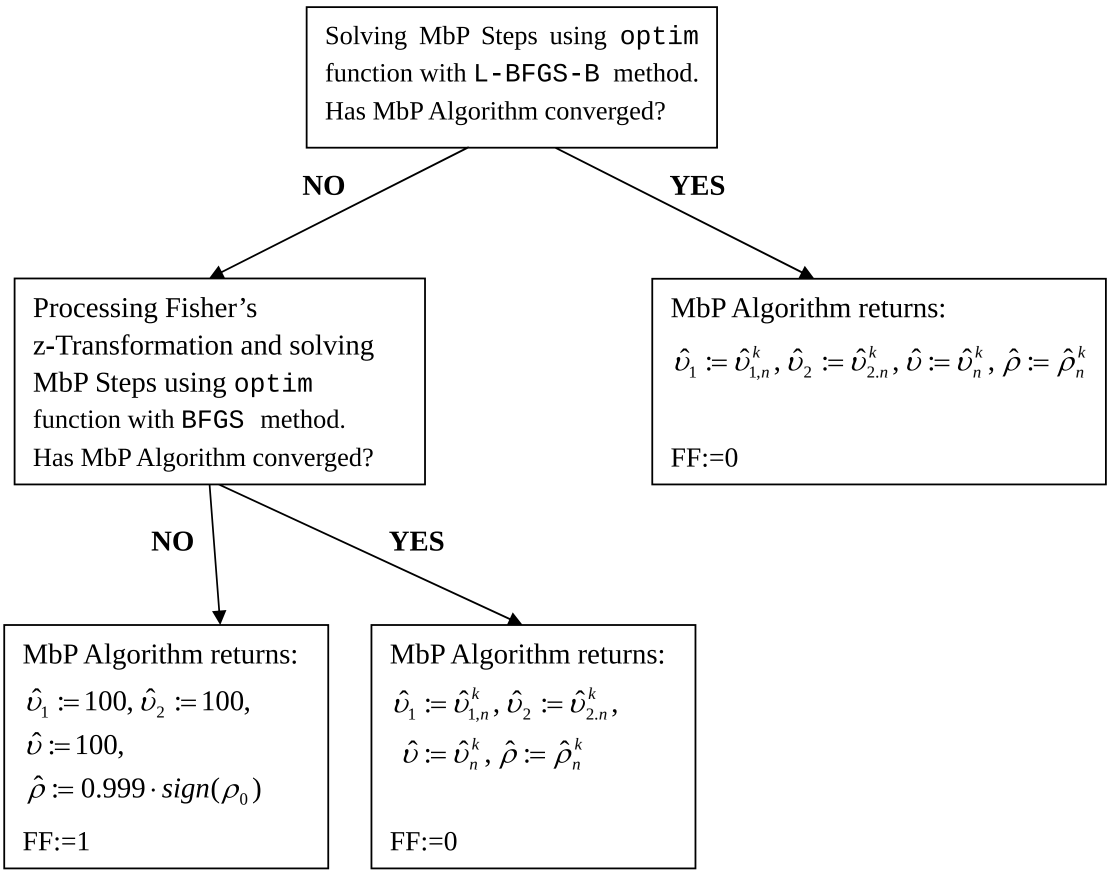


Figure 6.1: Flow chart of determining convergence indicator FF and the corresponding output for $\hat{\nu}$ and $\hat{\rho}$ for MbP algorithm applied to bivariate meta t-distribution

Algorithm 6.1 MbP Algorithm for Bivariate Meta t-Distribution

Input: bivariate data set $\boldsymbol{x}$ of size n, initial values $\boldsymbol{\eta}^0 := (\nu_1^0, \nu_2^0, \nu^0, \rho^0)$
Output: $\hat{\boldsymbol{\eta}}_n^k := (\hat{\nu}_{1,n}^k, \hat{\nu}_{2,n}^k, \hat{\nu}_n^k, \hat{\rho}_n^k)$

Initialize $k := 0$, fail $:= 0$
while $k < 50$ and $\|\boldsymbol{\eta}^k - \boldsymbol{\eta}^{k-1}\|_2 > 10^{-3}$, fail $:= 0$ **do**
  $k := k + 1$;
  if non-adaptive **then**
    $(\nu_0, \rho_0) := (\nu^0, \rho^0)$
  else
    $(\nu_0, \rho_0) := (\hat{\nu}_n^{k-1}, \hat{\rho}_n^{k-1})$
  end if
  Find minimum $\boldsymbol{\eta}_{1,n}^k := (\nu_{1,n}^k, \nu_{2,n}^k)$ of function $s_{1,a}^2(\boldsymbol{\eta}_1)+s_{1,b}^2(\boldsymbol{\eta}_1)$ using `optim`-function adapted with `L-BFGS-B` method in the parameter space $\Theta_1 := \{(\nu_1, \nu_2) : \nu_1, \nu_2 \in [1, 100]\}$
  Find the maximum $\boldsymbol{\eta}_{2,n}^k := (\nu_n^k, \rho_n^k)$ of $\ell_e(\boldsymbol{\eta}_{1,n}^{k-1}, \boldsymbol{\eta}_2; \boldsymbol{x})$ via `optim`-function adapted with `L-BFGS-B` method in the parameter space $\Theta_2 := \{\boldsymbol{\eta}_2 := (\nu, \rho) : \nu \in [1, 100], \rho \in (-1, 1)\}$
  if optim fails to converge **then**
    fail $:= 1$
  end if
end while
(MbP algorithm adapted with Fisher's z-transformation)
if fail $:= 1$ **then**
  Reset $k := 0$, fail $:= 0$
  while $k < 50$ and $\|\boldsymbol{\eta}_n^k - \boldsymbol{\eta}_n^{k-1}\|_2 > 10^{-3}$, fail $:= 0$ **do**
    $k := k + 1$;
    if non-adaptive **then**
      $(\nu_0, \rho_0) := (\nu^0, \rho^0)$
    else
      $(\nu_0, \rho_0) := (\hat{\nu}_n^{k-1}, \hat{\rho}_n^{k-1})$
    end if
    Fisher's z-transformation $\boldsymbol{\eta}_{\gamma,1} := (\nu_{\gamma,1}, \nu_{\gamma,2})$ with $\nu_{\gamma,j} := \ln(\nu_j - 1)$ for $j = 1, 2$
    Find minimum $\boldsymbol{\eta}_{\gamma,1,n}^k := (\nu_{\gamma,1,n}^k, \nu_{\gamma,2,n}^k)$ of function $s_{1,a}^2(\boldsymbol{\eta}_1) + s_{1,b}^2(\boldsymbol{\eta}_1)$ using `optim`-function adapted with `BFGS` method in the parameter space $\Theta_{\gamma,1} := \{(\nu_{\gamma,1}, \nu_{\gamma,2}) : \nu_{\gamma,1}, \nu_{\gamma,2} \in [1, \infty]\}$
    $\boldsymbol{\eta}_{\gamma,2} := (\nu_\gamma, \rho_\gamma)$ with $\nu_\gamma := \ln(\nu - 1)$ and $\rho_\gamma := \ln(\frac{1+\rho}{1-\rho})$
    Find the maximum $\boldsymbol{\eta}_{\gamma,2,n}^k := (\nu_n^k, \rho_n^k)$ of $\ell_e(\boldsymbol{\eta}_{\gamma,1,n}^{k-1}, \boldsymbol{\eta}_{\gamma,2}; \boldsymbol{x})$ via `optim`-function adapted with `BFGS` method in the parameter space $\Theta_{\gamma,2} := \{\boldsymbol{\eta}_{\gamma,2} := (\nu_\gamma, \rho_\gamma) : \nu \in [1, \infty], \rho \in (-1, 1)\}$
    $\boldsymbol{\eta}_n^k := \boldsymbol{\eta}_{\gamma,n}^k$
  end while
end if
return $\hat{\boldsymbol{\eta}}_n^k := (\nu_{1,n}^k, \nu_{2,n}^k, \nu_n^k, \rho_n^k)$

6.3 Alternative Computational Methods for Determining the MLE

6.3.1 Direct Optimization Method (DIR)

In the direct maximization method (DIR) we process a 4-dim.maximization of log-likelihood $\ell(\nu_1, \nu_2, \nu, \rho; \boldsymbol{x})$ using `optim`-function with the `L-BFGS-B` method for parameters ν_1, ν_2, $\nu \in [1, 100]$ and $\rho \in (-1, 1)$. If `optim` fails to converge, we adopt the `BFGS` method in which the parameters are transformed using Fisher's z-transformation defined in (6.4) in Section 6.2. The flow chart of the algorithm, as well as the description of the output indictor FF, are given in Figure 6.2 and Table 6.1.

Algorithm 6.2 Direct Maximization for bivariate t-Data

Input: data $\boldsymbol{x}$, initial vector $(\nu_1^0, \nu_2^0, \nu^0, \rho_n^0)$

Output: a 4-dim. vector $\hat{\boldsymbol{\eta}} := (\hat{\nu}_1, \hat{\nu}_2, \hat{\nu}, \hat{\rho})$

1: Find 4-dimensional maximum of $\ell(\boldsymbol{\eta}; \boldsymbol{x})$ via `optim`-function adapted with `L-BFGS-B` method in a parameter space $\Theta := \{\boldsymbol{\eta} : \nu_1, \nu_2, \nu \in [1, 100], \rho \in (-1, 1)\}$

2: **if** optim with L-BFGS-B method fails to converge **then**

3: Fisher's z-transformation $\boldsymbol{\eta}_\gamma := (\nu_{\gamma,1}, \nu_{\gamma,2}, \nu_\gamma, \rho_\gamma)$ with $\nu_{\gamma,j} := \ln(\nu_j - 1)$ for $j = 1, 2$, $\nu_\gamma := \ln(\nu - 1)$ and $\rho_\gamma := \ln(\frac{1+\rho}{1-\rho})$

4: Find 4-dimensional maximum of $\ell(\boldsymbol{\eta}_\gamma; \boldsymbol{x})$ via `optim`-function adapted with `BFGS` method in a parameter space $\Theta := \{\boldsymbol{\eta}_\gamma : \nu_1, \nu_2, \nu \in [1, \infty], \rho \in (-1, 1)\}$

5: **end if**

6.3.2 Inference for Margins Method (IFM)

Let $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ be the bivariate data set. In the IFM method, we first estimate the parameter of the marginal distributions separately.

STEP 0.1

Solve the equation $\dfrac{\partial}{\partial \boldsymbol{\eta}_1} \ell_w(\nu_1, \nu_2; \boldsymbol{x}) = \mathbf{0}$. In particular we find the root $\nu_{1,n}^{\text{ifm}}$ of the equation

$$\frac{\partial}{\partial \nu_1} \ell_m(\nu_1; \boldsymbol{x}_1) = 0 \tag{6.5}$$

and the root $\nu_{2,n}^{\text{ifm}}$ of

$$\frac{\partial}{\partial \nu_2} \ell_m(\nu_2; \boldsymbol{x}_2) = 0 \tag{6.6}$$

using 1-dimensional bisection method.

STEP 0.2

Estimate the coefficient ρ of t-copula using empirical estimated Kendall's τ $\hat{\rho}_\tau$. To recall that the coefficient ρ of the t-copula is given by

$$\rho := Cor(t_\nu^{-1}(F(\boldsymbol{x}_1; \nu_1)), t_\nu^{-1}(F(\boldsymbol{x}_2; \nu_2)))$$

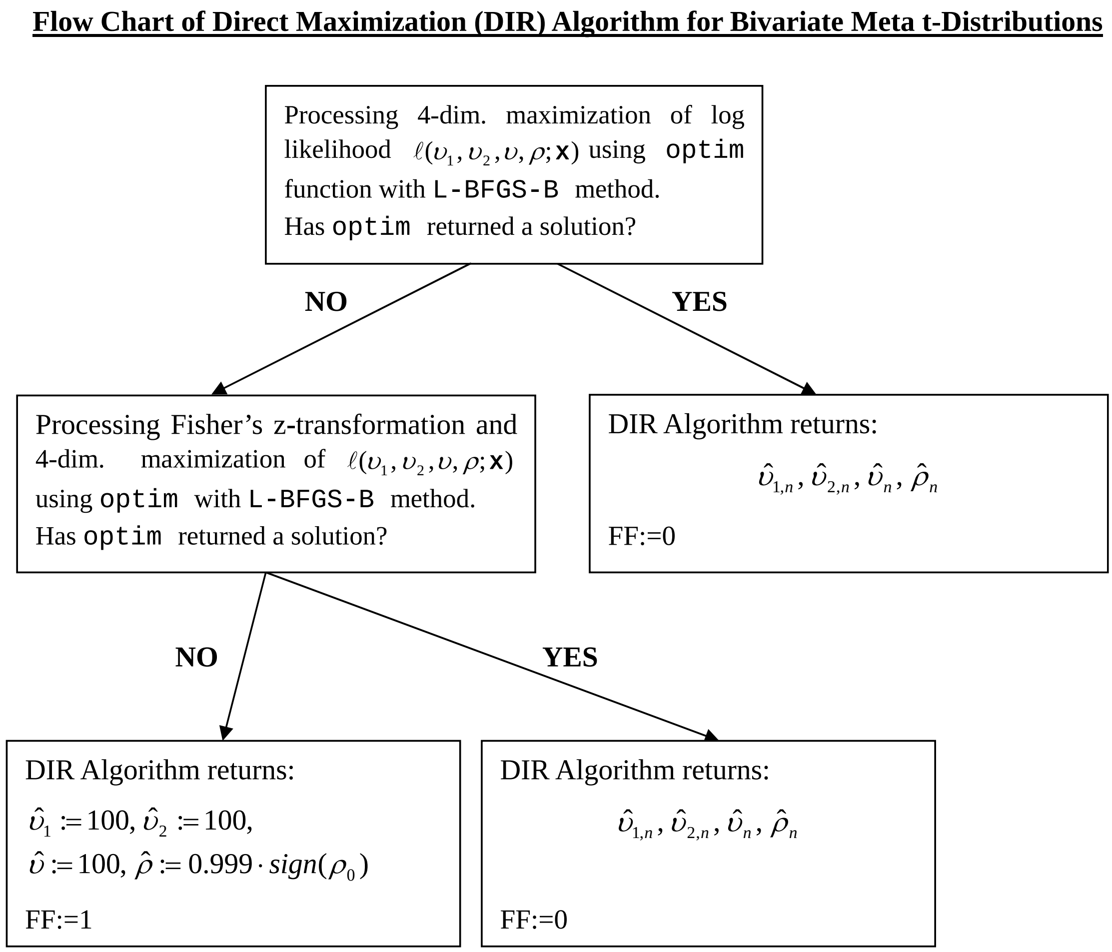


Figure 6.2: Flow chart of determining convergence indicator FF and the corresponding output for $\hat{\nu}$ and $\hat{\rho}$ for Direct Maximization algorithm applied to bivariate meta t-distribution

It then follows that

$$\begin{aligned}
\hat{\rho} &:= Cor(t_\nu^{-1}(F(\boldsymbol{x}_1; \nu_{1,n}^{\text{ifm}})), t_\nu^{-1}(F(\boldsymbol{x}_2; \nu_{2,n}^{\text{ifm}}))) \\
&= \sin\left(\frac{\pi}{2}\hat{\rho}_\tau\left(t_\nu^{-1}(F(\boldsymbol{x}_1; \nu_{1,n}^{\text{ifm}})), t_\nu^{-1}(F(\boldsymbol{x}_2; \nu_{2,n}^{\text{ifm}}))\right)\right) \\
&= \sin(\frac{\pi}{2}\hat{\rho}_\tau(\boldsymbol{x}_1, \boldsymbol{x}_2)) \\
&=: \hat{\rho}_n^{\text{ifm}}
\end{aligned}$$

STEP 0.3

Estimate the remaining parameter $\nu \in [1, 100]$ by maximum likelihood with the estimated coefficient $\hat{\rho}_n^{\text{ifm}}$. Find the root of the equation

$$\frac{\partial}{\partial \nu}\ell_e(\hat{\nu}_{1,n}^{\text{ifm}}, \hat{\nu}_{2,n}^{\text{ifm}}, \nu, \hat{\rho}_n^{\text{ifm}}; \boldsymbol{x}) = \mathbf{0}.$$

Flow Chart of Inference for Margins (IFM) Algorithm for Bivariate Meta t-Distributions

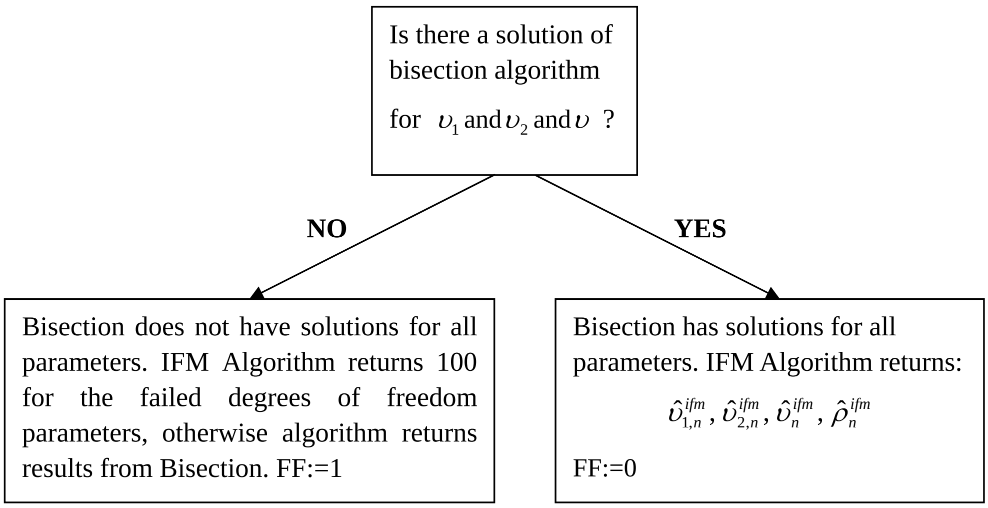


Figure 6.3: Flow chart of determining convergence indicator FF and the corresponding output for $\hat{\nu}$ and $\hat{\rho}$ for Inference for Margins algorithm applied to bivariate meta t-distribution

We denote the estimates from IFM method as: $(\hat{\nu}_{1,n}^{\text{ifm}}, \hat{\nu}_{2,n}^{\text{ifm}}, \hat{\nu}_{n}^{\text{ifm}}, \hat{\rho}_{n}^{\text{ifm}})$. The flow chart of the IFM algorithm and the description of the output indictor FF are illustrated in Figure 6.3 and in Table 6.1.

Algorithm 6.3 IFM: Inference for Margins Method for Bivariate Meta t-Data

Input: data $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2)$ of size n
Output: $\hat{\nu}_{1,n}^{\text{ifm}}$, $\hat{\nu}_{2,n}^{\text{ifm}}$, $\hat{\nu}_{n}^{\text{ifm}}$, $\hat{\rho}_{n}^{\text{ifm}}$

Find the root $\hat{\nu}_{1,n}^{\text{ifm}}$ of $\frac{\partial}{\partial \nu_1}\ell_m(\nu_1; \boldsymbol{x}_1) = 0$ in the range $\nu_1 \in [1, 100]$ using bisection algorithm
Find the root $\hat{\nu}_{2,n}^{\text{ifm}}$ of $\frac{\partial}{\partial \nu_2}\ell_m(\nu_2; \boldsymbol{x}_2) = 0$ in the range $\nu_2 \in [1, 100]$ using bisection algorithm
Calculate $\hat{\rho}_{n}^{\text{ifm}}$ by $\hat{\rho} := \sin(\frac{\pi}{2}\hat{\rho}_\tau(\boldsymbol{x}_1, \boldsymbol{x}_2))$
Find the root $\hat{\nu}_{n}^{\text{ifm}}$ of $\frac{\partial}{\partial \nu}\ell_e(\hat{\nu}_{1,n}^{\text{ifm}}, \hat{\nu}_{2,n}^{\text{ifm}}, \nu, \hat{\rho}_{n}^{\text{ifm}}; \boldsymbol{x}_2) = 0$ in the range $\nu \in [1, 100]$ using bisection algorithm
return $\hat{\nu}_{1,n}^{\text{ifm}}$, $\hat{\nu}_{2,n}^{\text{ifm}}$, $\hat{\nu}_{n}^{\text{ifm}}$, $\hat{\rho}_{n}^{\text{ifm}}$

FF:=0	
MbP	Convergence, MbP returns $\hat{\nu}^k_{1,n}$,$\hat{\nu}^k_{2,n}$,$\hat{\nu}^k_n$,$\hat{\rho}^k_n$
DIR	Algorithm returns results $\hat{\nu}^k_{1,n}$,$\hat{\nu}^k_{2,n}$,$\hat{\nu}_n$,$\hat{\rho}_n$ from `optim`
IFM	Bisection finds solutions for all parameters. Algorithm returns: $\hat{\nu}^{\text{ifm}}_{1,n}$,$\hat{\nu}^{\text{ifm}}_{2,n}$,$\hat{\nu}^{\text{ifm}}_n$,$\hat{\rho}^{\text{ifm}}_n$
FF:=1	
MbP	In step k, $k = 1, ..., 50$ 2-dim. `optim` embedded MbP algorithm fails to returns a solution. Algorithm returns $\hat{\nu}_1 := 100$, $\hat{\nu}_2 := 100$, $\hat{\nu} := 100$, $\hat{\nu}_1 := 0.999 \cdot \text{sign}(\rho_0)$.
DIR	`optim` fails to return a solution. Algorithm returns $\hat{\nu}_1 := 100$, $\hat{\nu}_2 := 100$, $\hat{\nu} := 100$ and $\hat{\rho} := 0.999 \cdot \text{sign}(\rho_0)$
IFM	Bisection algorithm does not have solutions for all parameters. Algorithm returns 100 for the failed degrees of freedom parameters, otherwise algorithm returns results from Bisection algorithm.

Table 6.1: Overview of convergence indicator FF for MbP, DIR and IFM algorithms

6.4 Illustration of the Performance of MbP Algorithm for Non-adaptive and Adaptive ν_0 and ρ_0

6.4.1 Example for One Randomly Generated Data Set

With the initial settings as stated in Section 6.2 we first examine the performance of MbP algorithm for non-adaptive and adaptive ν_0 and ρ_0 with one randomly generated meta-t data set using (6.1). Let $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n\times 2}$ be the bivariate meta t-data of size n with $\boldsymbol{x}_j := (x_{1j}, ..., x_{nj})^t$ for $j = 1, 2$.

The calculation of initial values for MbP algorithm occurs via IFM method described in the previous section. As result for the initial values, we have:

$$(\nu_1^0, \nu_2^0, \nu^0, \rho^0) := (\hat{\nu}^{\text{ifm}}_{1,n}, \hat{\nu}^{\text{ifm}}_{2,n}, \hat{\nu}^{\text{ifm}}_n, \hat{\rho}^{\text{ifm}}_n)$$

If the IFM method fails to converge, we start the MbP iteration with a fixed value, i.e.

$$(\nu_1^0, \nu_2^0, \nu^0, \rho^0) = (5, 5, 20, 0.9)$$

Then we use previously generated data $\boldsymbol{x}$ and initial value $(\nu_1^0, \nu_2^0, \nu^0, \rho^0)$ to execute the direct maximization algorithm and MbP algorithm.

Example:

In our example, we consider a specific sample of size $n = 100$ with parameters $\nu_{1,tr} = 3$, $\nu_{2,tr} = 10$, $\nu_{tr} = 3$ and $\rho_{tr} = 0.8$ randomly generated by (6.1).

1) Result from **IFM method**, as initial values for MbP and direct maximization method:

- estimated values:
  $\hat{\nu}^{\text{ifm}}_{1,n} = 2.103401$ $\hat{\nu}^{\text{ifm}}_{2,n} = 4.878466$ $\hat{\nu}^{\text{ifm}}_n = 2.761441$ and $\hat{\rho}^{\text{ifm}}_n = 0.874158$
- log-likelihood value: -284.899879

- calculating time: 6.54 sec

2) Result from **direct maximization method**:

- estimated values: $\hat{\nu}_{1,n} = 2.524670$ $\hat{\nu}_{2,n} = 6.937399$ $\hat{\nu}_n = 2.775408$ $\hat{\rho}_n = 0.852053$
- log-likelihood value: -284.319566
- calculating time: 3.54 sec
- `optim` method: Quasi-Newton method with box constraints

3) Result from **MBP1: the non-adaptive MbP algorithm with fixed $\nu_0 := \nu_n^{\mathbf{ifm}}$ and $\rho_0 := \rho_n^{\mathbf{ifm}}$**

- estimates: $\hat{\nu}_{1,n}^k = 2.524274$, $\hat{\nu}_{2,n}^k = 6.934209$, $\hat{\nu}_n^k = 2.774007$ and $\hat{\rho}_n^k = 0.852066$
- log-likelihood value: -284.31956510
- number of iterations: 16
- calculating time: 354.63 sec
- `optim` method: `L-BFGS-B` (Quasi-Newton method with box constraints)
- estimated values:

k	$\hat{\nu}_{1,n}^k$	$\hat{\nu}_{2,n}^k$	$\hat{\nu}_n^k$	$\hat{\rho}_n^k$
0	2.103401	4.878465	2.761440	0.874157
1	2.217496	5.326820	2.501675	0.861310
2	2.304016	5.802664	2.591402	0.858938
3	2.374666	6.163409	2.639314	0.856600
4	2.425509	6.426413	2.682687	0.855026
⋮	⋮	⋮	⋮	⋮
13	2.523123	6.928168	2.773583	0.852110
14	2.523727	6.931333	2.773728	0.852087
15	2.524083	6.933250	2.773872	0.852073
16	2.524274	6.934209	2.774007	0.852065

- The convergence in graphics, see Figure 6.4.

4) Result from **MBP2: the adaptive MbP algorithm with updated ν_0 and ρ_0**

- estimates: $\hat{\nu}_{1,n}^k = 2.524104$, $\hat{\nu}_{2,n}^k = 6.932947$, $\hat{\nu}_n^k = 2.774225$ and $\hat{\rho}_n^k = 0.852072$
- log-likelihood value: -284.31956493
- number of iterations: 17
- calculating time: 323.00 sec
- `optim` method: `L-BFGS-B` (Quasi-Newton method with box constraints)

- estimated values:

k	$\nu_{1,n}^k$	$\nu_{2,n}^k$	ν_n^k	ρ_n^k
0	2.103401	4.878465	2.761440	0.874157
1	2.217496	5.326820	2.501675	0.861310
2	2.300988	5.784421	2.591402	0.858938
3	2.369773	6.132445	2.637863	0.856686
4	2.419597	6.389630	2.680518	0.855159
⋮	⋮	⋮	⋮	⋮
13	2.522516	6.924836	2.773341	0.852147
14	2.523232	6.928466	2.773886	0.852109
15	2.523699	6.930872	2.774005	0.852089
16	2.523994	6.932490	2.774117	0.852078
17	2.524104	6.932947	2.774225	0.852072

- The convergence in graphics, see Figure 6.5.

Conclusion

In this specific example, we observe that the direct maximization method and both MbP methods deliver similar estimates of the parameters $(\nu_1, \nu_2, \nu, \rho)$, while the estimates from IFM method largely differ from the true values. Furthermore, we can also observe that MbP algorithms and direct maximization have similar log-likelihood values. On the other hand the direct maximization method terminates after a moderate time within few seconds, while both MbP methods require 50× to 100× more computing time than the IFM and direct maximization method.

One explanation for extreme time costliness is that the MbP algorithm is an iterative algorithm which requires $2 \times k$, $k > 1$, optimization procedures over complicated score functions while the direct maximization executes once over a simpler log-likelihood function. In the IFM method, only 1-dimensional bisection method is adopted which operates in a moderate time frame as well.

In order to scrutinize the MbP algorithm and find out its advantages, we need to simulate more scenarios in the next section.

6.5 Simulation Setups and Performance Measures

6.5.1 Simulation Setups

In the following simulations we scrutinize all the combinations with the parameters

- $\nu_1, \nu_2, \nu \in \{3, 10\}$
- $\rho \in \{0.3, 0.8\}$
- sample size $n \in \{100, 500\}$
- number of replications $r = 100$.

Plus the combination for $(\nu_1, \nu_2, \nu) = (20, 20, 20)$ it results in total 28 scenarios. For each scenario we estimate the parameters $(\nu_1, \nu_2, \nu, \rho)$ using the non-adaptive MbP method (MBP1), the adaptive MbP method (MBP2), direct maximization method (DIR) and

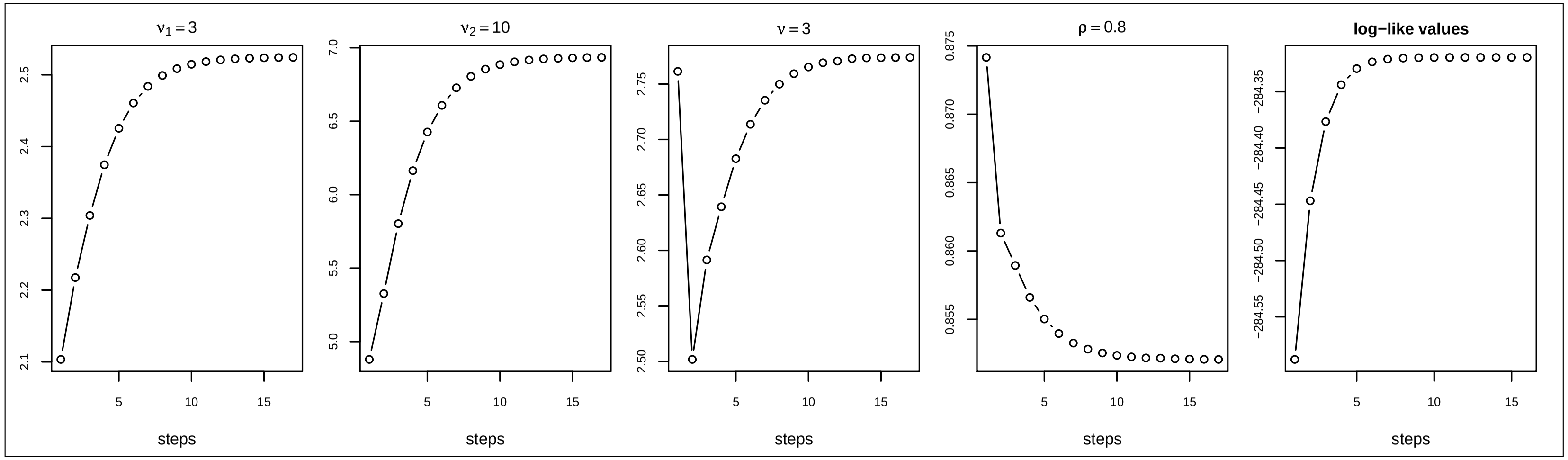


Figure 6.4: MBP1: non-adaptive method. Estimated values of the parameters ν_1, ν_2, ν and $\rho = 0.3$ and the resulting log-likelihood values

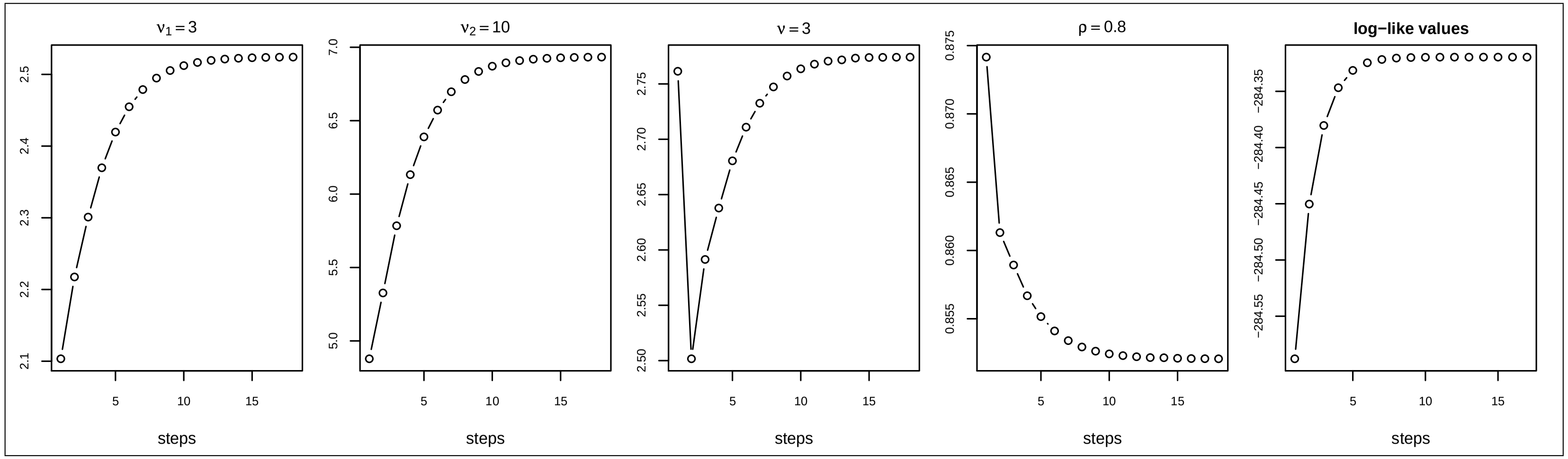


Figure 6.5: MBP2: adaptive method. Estimated values of the parameters ν_1, ν_2, ν and $\rho = 0.3$ and the resulting log-likelihood values

inference for margins method (IFM) over 100 replications ($r = 100$). Thus, in total we have four estimation methods in the simulation study.

6.5.2 Performance Measures

Like the evaluation of the estimation method in Chapter 4 we utilize the output indicator FF to investigate the frequency of failure, type of the failure, the convergence of the corresponding algorithm and to determine the total number of compared replicates in the simulation study. The indicator FF for the meta case is defined in Table 6.1. The quantity FF := 1 indicates a failure in the replication. Thus, for each estimation method we denote r_g as the number of the replicates whose output indicator FF $\neq 1$. The maximum number of r_g is the number of the replicates $r = 500$. The higher r_g is, the more reliable and stable is the estimation method. Therefore, the number r_g implies the reliability and robustness of the corresponding algorithm.

To compare the four estimation methods, MBP1, MBP2, DIR and IFM methods, with each other we dismiss the ith estimates, $i \in \{1, ..., r\}$ with $r = 500$, if all four estimation methods return $\hat{\nu} = 100$ using the ith sample set. Let $|\{i\}|$ be the total number of dismissed samples, whereby $|\{\cdot\}|$ indicates the cardinality of a set. Then $r_c := r - |\{i\}|$ amounts the total number of samples we use for comparison of the four estimation methods.

Based r_c we compare the four estimations methods with the summary statistics sample mean $\bar{\theta}$, estimated standard error of mean $s(\bar{\theta})$, estimated robust standard error $\hat{\sigma}_r(\hat{\theta})$, estimated relative bias $\widehat{rb}(\bar{\theta})$, estimated standard error of relative bias $s_{rb}(\bar{\theta})$, estimated mean squared error (mse) $\widehat{mse}(\hat{\theta})$, estimated standard error of mse $s_{mse}(\hat{\theta})$, estimated efficiency based on mse $\widehat{eff}_{mse}$, estimated robust mse (rmse) $\widehat{rmse}(\hat{\theta})$ and estimated efficiency based on rmse $\widehat{eff}_{rmse}$. These summary statistics are defined in Section 2.4. We also calculate their log likelihood value *loglike* to examine if the estimates on average are at their maximum.

For stopping time measurement we adopt the R-function `proc.time()` to measure the procedure time of `optim` evaluations in each MbP-step. Since we have 2× 2-dim. `optim` operations in each step, $2 \times k$ measured time values are summed up and averaged over k steps. In the summary table we denote *time* for the consuming time.

Moreover, we examine averaged iterations *avg.it* to reach the convergence criterion for the MbP methods. However, this number does not apply to Direct, CSP and CML methods, since they are not iterative algorithms.

In the summary Table 6.2 in the simulation study we also display the averaged number of function evaluations *f.eval* in the `optim` function of each estimation methods. For MBP algorithm, *f.eval* is the averaged function evaluations of both working and error models during one MBP iteration.

Ranking Method

Since there are in total 28 scenarios in the simulation study and the advantages of MbP algorithm cannot be easily identified, we want to construct a relative fair way to evaluate the four maximization methods and discover benefits and disadvantages of each method based on all scenarios. The idea is that we rank all four methods based on the statistics

$\widehat{\text{eff}}_{rmse}$ of the estimates.

How have the methods been ranked?
The ranking is completed by `rank`-function in R using the *averaged rank* method. Here is an example of the ranking method based on $\widehat{\text{eff}}_{rmse}$, a statistic which describes the ratio of the robust MSE of each method to MBP1. In the following points, we illustrate the ranking method for the scenario $(\nu_1, \nu_2, \nu, \rho) = (3, 3, 3, 0.3)$ with sample size $n = 100$.

(i) Foremost we rank the methods for each parameter based on $\widehat{\text{eff}}_{rmse}$. Since this figure is the ratio of the robust MSE of other method to MBP1, i.e. $\frac{\widehat{rmse}(method)}{\widehat{rmse}(\text{MBP1})}$, the smaller number means that the method has a smaller RMSE estimate than the method MBP1 and thus it is more stable method than MBP1. Therefore, the smaller the number is, the better is the ranking. For example:

n	θ	Meth.		$\widehat{\text{eff}}_{rmse}$	*the rank:*
		MBP1		1.000	*3*
	ν_1	MBP2		1.000	*3*
		DIR	···	1.000	*3*
		IFM		0.714	*1*
		MBP1		1.000	*1.5*
	ν_2	MBP2	···	1.000	*1.5*
		DIR		1.001	*3*
100		IFM		1.042	*4*
		MBP1	···	1.000	*2*
	ν	MBP2		1.000	*2*
		DIR		1.000	*2*
		IFM	···	1.150	*4*
		MBP1		1.000	*2.5*
	ρ	MBP2		1.000	*2.5*
		DIR	···	1.001	*4*
		IFM		0.830	*1*

(ii) From (i) we obtain a rank matrix. Then we can either sum the ranked number of all parameters or rank the numbers by regarding the parameters of marginal distributions and these of the copula model separately.

- all parameters

	MBP1	MBP2	DIR	IFM
ν_1	3	3	3	1
ν_2	1.5	1.5	3	4
ν	2	2	2	4
ρ	2.5	2.5	4	1
$\sum$	*9*	*9*	*12*	*10*

- separated

	MBP1	MBP2	DIR	IFM
ν_1	3	3	3	1
ν_2	1.5	1.5	3	4
$\sum$	*4.5*	*4.5*	*6*	*5*
ν	2	2	2	4
ρ	2.5	2.5	4	1
$\sum$	*4.5*	*4.5*	*6*	*5*

(iii) then we rank the result from (ii):
- all parameters

	MBP1	MBP2	DIR	IFM
	9	9	12	10
rank:	*1.5*	*1.5*	*4*	*3*

- separated

	MBP1	MBP2	DIR	IFM
ν_1, ν_2	4.5	4.5	6	5
rank:	*1.5*	*1.5*	*4*	*3*
ν, ρ	4.5	4.5	6	5
rank:	*1.5*	*1.5*	*4*	*3*

The ranked results are then illustrated in the table as follows:

Scenario	ν_1	ν_2	ν	ρ	n		MBP1	MBP2	DIR	IFM
S01	3.0	3.0	3.0	0.3	100	Marginal	1.5	1.5	4.0	3.0
						Copula	1.5	1.5	4.0	3.0
						Total	1.5	1.5	4.0	3.0

(iv) In the following tables we can investigate the algorithms by summing up the ranking of different scenarios. In Table 6.7 we have summed up the ranked number of all scenarios and then ranked the sum. For example, we sum up the ranked numbers from all scenarios. The labels Marginal and Copula indicate the sum of ranked number from marginal and copula parameters of all scenarios respectively:

				MBP1	MBP2	DIR	IFM
			$\sum$ Marginal	66.5	69.5	74.5	70.0
			$\sum$ Copula	66.5	74.5	67.0	72.0
			$\sum$ Total	64.5	75.5	69.0	72.0

Again, we rank the results and receive the final ranking as stated in Table 6.7.

				MBP1	MBP2	DIR	IFM
			$\sum$ Marginal	1.0	2.0	4.0	3.0
			$\sum$ Copula	1.0	4.0	2.0	3.0
			$\sum$ Total	1.0	4.0	2.0	3.0

To examine the behaviors of a specified parameter combination, we add the ranking number of the certain scenarios and rank the sum. In Table 6.5 and 6.6, some specified combinations are ranked and we can then easily observe at which combination the MBP1 or the MBP2 method exceed direct maximization and IFM method.

6.6 Simulation Results

6.6.1 Result from Example Scenarios

As an example, we regard two scenarios:

(i) Scenario $(\nu_1, \nu_2, \nu, \rho) = (10, 10, 3, 0.3)$ and $n = 100$:
In the summary table it shows that the sample mean $\bar{\theta}$ of the estimated parameters $(\nu_1, \nu_2, \nu, \rho)$ from MBP1 and MBP2 methods are closer to the true parameters than the results from the DIR and IFM methods. While the estimated standard error of mean $s(\bar{\theta})$ and estimated relative bias $\widehat{rb}(\bar{\theta})$ of the copula parameters ν and ρ from MBP1 and MBP2 are similar to those from the DIR and IFM methods, those estimates for the marginal parameters are considerably smaller, which means that the estimated marginal parameters from MBP1 and MBP2 methods are less variable than those from the DIR and IFM methods. The low variability of the estimates from MBP1 and MBP2 methods also reflects in the numbers of the estimated *mean squared error* ($\widehat{mse}(\hat{\theta})$) or the estimated *robust mean squared error* ($\widehat{rmse}(\hat{\theta})$). As described in Chapter 2, the latter is calculated based on the median while the other uses the average of the estimates which is much sensitive towards outliers. To compare the $\widehat{mse}(\hat{\theta})$ and $\widehat{rmse}(\hat{\theta})$ of the four estimation methods we look at the ratio between each of the four method to the reference method MBP1. Since the smaller the $\widehat{mse}(\hat{\theta})$ or $\widehat{rmse}(\hat{\theta})$ are, the closer and the better are the estimates. Thus a preferable estimation method is the one with smaller $\widehat{\text{eff}}_{mse}$ or smaller $\widehat{\text{eff}}_{rmse}$ values. In the column of $\widehat{\text{eff}}_{mse}$ we observe that for the marginal parameters ν_1 and ν_2 the ratio for MBP2 is significantly smaller than MBP1 while the ratio for DIR and IFM methods are 2-3 times larger, which indicates that the estimates from these methods are very variable and the value are far from true parameter values. However the ratio for the copula parameters of the DIR and IFM are smaller, especially for the degrees of freedom. In contrary the column of $\widehat{\text{eff}}_{rmse}$ demonstrates much stable ratio numbers which clearly disclose that there are more large estimates for the marginal parameters ν_1 and ν_2. Recall that the number of failed replicates is described by $100 - r_g$. In this scenario none of replicates was failed with MBP1, MBP2 and DIR algorithms while IFM methods fails to deliver estimates calculated by bisection algorithm 22 times.

Although both MbP methods (MBP1 and MBP2) deliver nearly the same log likelihood values as the direct maximization method (DIR), the drawback of the MbP method in comparison to the DIR method is clearly in its huge computing intensity and time consumption. The averaged calculating time for one replicate is in this case 37 sec. and that is 18 times higher than the time consumed by the DIR and IFM methods. The reason may be obviously described by the quantity *avg.it* in the summary table which explains that due to the iterative nature of the MbP algorithm it needs on average 13 steps per replicate to convergence while DIR and IFM calculate the estimates in one step.

(ii) Scenario $(\nu_1, \nu_2, \nu, \rho) = (10, 10, 3, 0.3)$ and $n = 500$:
In this scenario as the sample size rises to 500, we can see in the summary table there are no failed replicates being observed and estimated efficiency stated in the column $\widehat{\text{eff}}_{mse}$ and $\widehat{\text{eff}}_{rmse}$ are very similar, as well as the estimated sample $\bar{\theta}$, estimated standard error of mean $s(\bar{\theta})$, the estimated robust standard error $\hat{\sigma}_r(\hat{\theta})$, estimated relative bias $\widehat{rb}(\bar{\theta})$ and the log likelihood values. This result is within the expectation, since the population of a larger sample size contains more information

which facilitates the maximum likelihood estimation in general. The MbP algorithm is in this case more disadvantageous in the matter of time consumption. We can see in column *time* that MBP1 and MBP2 require on average 170 seconds per replicate while the DIR and IFM need ca. 7 and 18 seconds for the ML-estimation which is about 25 times and 9 times faster. The lower boxplot panel in Figure 6.6 demonstrates that with the increased sample size the variability of the estimates of MBP1 and MBP2 diminishes.

It is to note that we can obtain a different result if the scenario setting varies. For the results from other scenarios we refer to the summary tables and boxplots for all 28 scenarios in the Appendix.

			$\nu_{1,tr}=10,\ \nu_{2,tr}=10,\ \nu_{tr}=3,\ \rho_{tr}=0.3$															
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	17.717	1.904	9.790	0.772	0.190	400.456	127.844	1.000	83.937	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	17.454	1.796	9.793	0.745	0.180	358.834	109.498	0.896	84.000	1.001	−294.565	94	100	37.563	12.777	17.337
		DIR	23.640	3.069	10.012	4.547	0.307	1071.586	263.009	2.676	88.558	1.055	−294.565	94	100	1.698	1.000	33.340
		IFM	23.954	3.211	8.996	1.395	0.321	1164.066	284.065	2.907	75.543	0.900	−294.619	94	78	2.863	1.000	46.027
	ν_2	MBP1	16.779	1.838	6.388	0.678	0.184	363.406	117.756	1.000	51.684	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	16.815	1.847	6.385	0.682	0.185	367.119	120.500	1.010	51.634	0.999	−294.565	94	100	37.563	12.777	17.337
		DIR	20.919	2.810	6.451	1.092	0.281	861.665	237.432	2.371	53.358	1.032	−294.565	94	100	1.698	1.000	33.340
		IFM	19.508	2.679	5.849	0.951	0.268	764.898	228.374	2.105	55.333	1.071	−294.619	94	78	2.863	1.000	46.027
	ν	MBP1	3.812	0.278	1.511	0.271	0.093	7.951	4.183	1.000	4.044	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	3.812	0.278	1.511	0.271	0.093	7.945	4.183	0.999	4.050	1.001	−294.565	94	100	37.563	12.777	17.337
		DIR	3.814	0.255	1.513	0.271	0.085	6.753	2.912	0.849	4.114	1.017	−294.565	94	100	1.698	1.000	33.340
		IFM	3.773	0.242	1.336	0.258	0.081	6.123	2.754	0.770	4.387	1.085	−294.619	94	78	2.863	1.000	46.027
	ρ	MBP1	0.313	0.011	0.121	0.042	0.037	0.012	0.002	1.000	0.032	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	0.313	0.011	0.121	0.042	0.037	0.012	0.002	1.000	0.032	1.000	−294.565	94	100	37.563	12.777	17.337
		DIR	0.312	0.011	0.121	0.041	0.037	0.011	0.002	0.998	0.032	1.003	−294.565	94	100	1.698	1.000	33.340
		IFM	0.310	0.011	0.111	0.034	0.037	0.012	0.002	1.001	0.027	0.846	−294.619	94	78	2.863	1.000	46.027
500	ν_1	MBP1	11.458	0.402	3.153	0.146	0.040	18.299	5.975	1.000	19.273	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	11.458	0.402	3.153	0.146	0.040	18.290	5.972	1.000	19.280	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	11.487	0.412	3.177	0.496	0.041	19.224	6.611	1.051	19.048	0.988	−1476.409	100	100	6.879	1.000	28.700
		IFM	11.511	0.419	3.142	0.151	0.042	19.832	6.679	1.084	16.427	0.852	−1476.487	100	100	17.737	1.000	59.715
	ν_2	MBP1	11.821	0.513	2.816	0.182	0.051	29.583	10.013	1.000	14.477	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	11.821	0.512	2.816	0.182	0.051	29.572	10.004	1.000	14.484	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	11.882	0.538	2.942	0.188	0.054	32.477	11.676	1.098	14.625	1.010	−1476.409	100	100	6.879	1.000	28.700
		IFM	11.898	0.550	3.040	0.190	0.055	33.891	11.835	1.146	13.812	0.954	−1476.487	100	100	17.737	1.000	59.715
	ν	MBP1	3.093	0.064	0.603	0.031	0.021	0.412	0.144	1.000	0.703	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	3.093	0.064	0.603	0.031	0.021	0.412	0.144	1.000	0.703	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	3.093	0.063	0.605	0.031	0.021	0.403	0.134	0.977	0.710	1.011	−1476.409	100	100	6.879	1.000	28.700
		IFM	3.106	0.066	0.607	0.035	0.022	0.444	0.152	1.076	0.856	1.218	−1476.487	100	100	17.737	1.000	59.715
	ρ	MBP1	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.020	−1476.409	100	100	6.879	1.000	28.700
		IFM	0.292	0.005	0.042	−0.027	0.016	0.002	0.000	1.019	0.004	0.986	−1476.487	100	100	17.737	1.000	59.715

Table 6.2: Summary statistics sample mean $\bar{\theta}$, estimated standard error of mean $s(\bar{\theta})$, estimated robust standard error $\hat{\sigma}_r(\hat{\theta})$, estimated relative bias $\widehat{rb}(\bar{\theta})$, estimated standard error of relative bias $s_{rb}(\bar{\theta})$, estimated mean squared error (mse) $\widehat{mse}(\hat{\theta})$, estimated standard error of mse $s_{mse}(\hat{\theta})$, estimated efficiency based on mse $\widehat{eff}_{mse}$, estimated robust mse (rmse) $\widehat{rmse}(\hat{\theta})$, estimated efficiency based on rmse $\widehat{eff}_{rmse}$, log likelihood value *loglike*, number of non-failed replicates r_g, computing time *time*, averaged MbP steps *avg.it*, function evaluations *func.eval.* of 100 replicates (r=100) and samples of size $n = 100, 500$ using MbP algorithm with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters ν_0 and ρ_0, direct maximization (DIR) and inference for margins method (IFM).

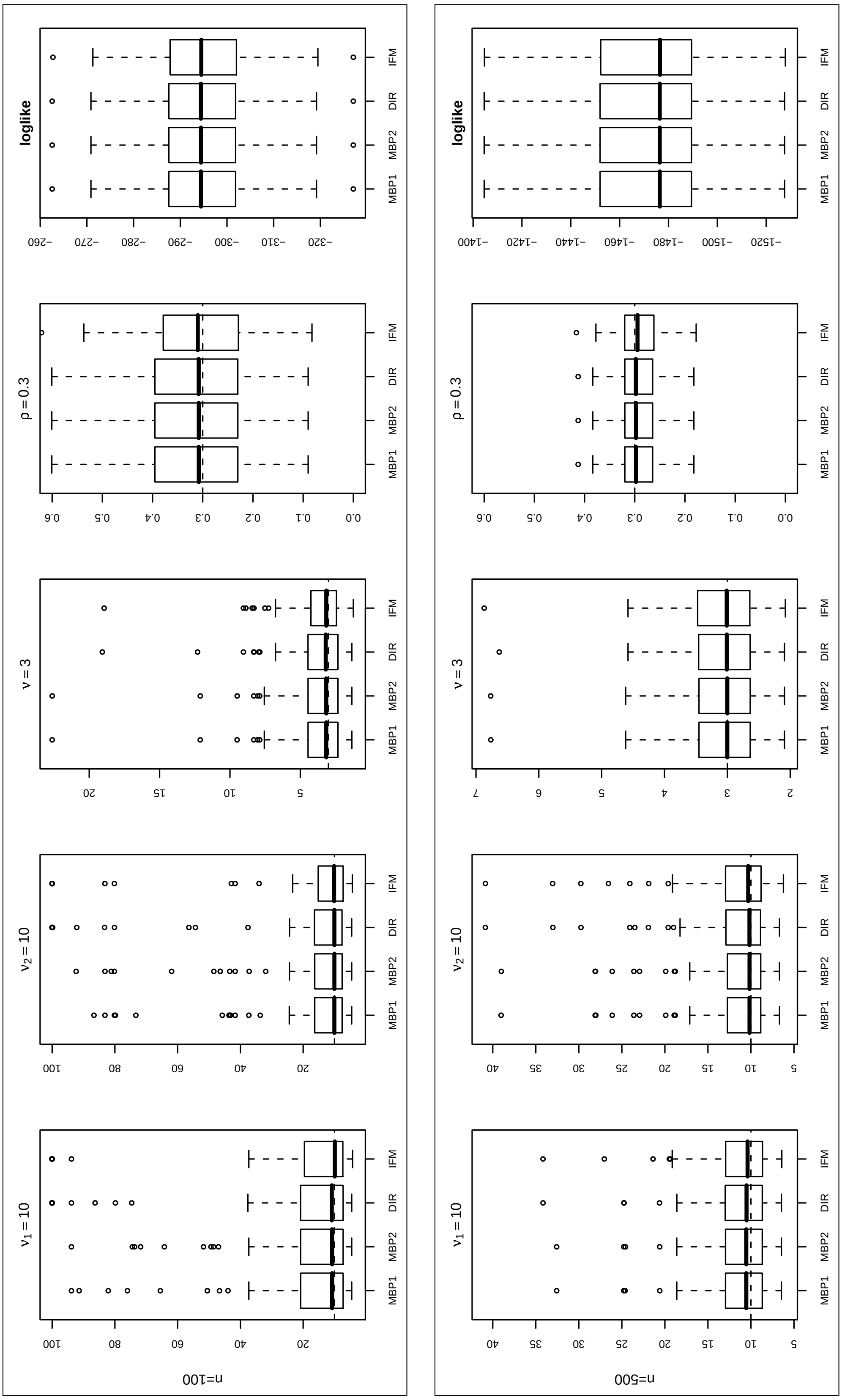


Figure 6.6: Boxplots of estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.

6.6.2 Results and Conclusions from the Simulation Ranking over all Scenarios

Table 6.7 illustrates the ranking of all 28 scenario using ranking method described in Section 6.5.2. The statistics which the ranking is based on is the efficiency $\widehat{\text{eff}}_{rmse}$ calculated by estimated robust mean squared error.

Table 6.3 gives an overview of the parameter combination being setup in the scenario. In Table 6.4 we assign each scenario to the method with the highest ranking (lowest number). Then we receive 10 scenarios for the non-adaptive MbP method (MbP1), 5 for the adaptive MbP method (MbP2), 12 for direct maximization method (DIR) and 11 for inference for margins method (IFM). It is to note that two method can perform equally well and receive the highest ranking at the same time.

$\nu_{1,tr}$	$\nu_{2,tr}$	ν_{tr}	Scenarios
3	3	3	S01 − S04
10	10	10	S05 − S08
20	20	20	S09 − S12
3	3	10	S13 − S16
3	10	3	S17 − S20
3	10	10	S21 − S24
10	10	3	S25 − S28

Table 6.3: Scenario allocation: for instance, in scenarios S01 to S04 we have parameter combination $(\nu_1, \nu_2, \nu) = (3, 3, 3)$, $\rho \in \{0.3, 0.8\}$ and $n \in \{100, 500\}$

	MBP1	MBP2	DIR	IFM
	S01	S01	S02	S05
	S02	S02	S03	S06
	S08	S13	S04	S07
	S09	S17	S10	S13
	S11	S20	S16	S14
	S12		S17	S15
	S13		S18	S17
	S17		S19	S21
	S20		S20	S22
	S25		S24	S23
			S27	S26
			S28	
$\sum$	10	5	12	11

Table 6.4: First column gives the scenarios which have MBP1 as best method. Similarly the other columns can be interpreted.

Furthermore, Table 6.5 and Table 6.6 show the ranking of scenarios with certain parameter combinations based on $\widehat{\text{eff}}_{rmse}$. The quantity *Num* indicates the number of scenarios in the specific combination group.

Depend on ranking in Table 6.5 and Table 6.6 we note that the non-adaptive MbP method (MBP1) outperforms the direct maximization method (DIR) and the IFM method, if the sample size n and the copula parameter ν are relatively small ($n = 100$ and $\nu = 3$). This result indicates that the MbP method estimate the likelihood better if the distributions have heavier tails. The iterative algorithm is preferable in both simulations with low and high correlations ($\rho \in \{0.3, 0.8\}$). In case of a larger sample size ($n = 500$), MBP1 estimate the copula parameter ν better, if the correlation ρ is smaller ($\rho = 0.3$). Nevertheless, by all means, if the sample size is large, the direct maximization method and IFM method are superior to the MbP algorithm.

As an overall result of all 28 scenarios stated in Table 6.7 we conclude that the MBP1 method delivers the most accurate and robust estimates and therefore is the best method for estimating the maximum likelihood estimators.

		Num		MBP1	MBP2	DIR	IFM
$n = 100$	$\rho = 0.3$	7	Marginal	2.0	1.0	4.0	3.0
			Copula	2.0	3.0	4.0	1.0
			Total	1.0	2.0	4.0	3.0
	$\rho = 0.8$	7	Marginal	1.0	4.0	3.0	2.0
			Copula	4.0	3.0	1.0	2.0
			Total	2.5	4.0	2.5	1.0
	$\nu = 3$	6	Marginal	2.0	1.0	3.5	3.5
			Copula	1.0	3.0	2.0	4.0
			Total	1.0	3.0	2.0	4.0
	$\nu = 10$	6	Marginal	2.0	3.0	4.0	1.0
			Copula	2.0	3.0	4.0	1.0
			Total	2.0	3.0	4.0	1.0
	overall	14	Marginal	1.0	2.0	4.0	3.0
			Copula	2.0	4.0	3.0	1.0
			Total	1.0	3.0	4.0	2.0
$n = 500$	$\rho = 0.3$	7	Marginal	3.0	4.0	1.5	1.5
			Copula	1.0	2.0	4.0	3.0
			Total	3.5	3.5	2.0	1.0
	$\rho = 0.8$	7	Marginal	2.0	4.0	1.0	3.0
			Copula	2.0	3.0	1.0	4.0
			Total	2.0	3.0	1.0	4.0
	$\nu = 3$	6	Marginal	2.0	4.0	1.0	3.0
			Copula	1.0	2.0	3.0	4.0
			Total	2.0	3.0	1.0	4.0
	$\nu = 10$	6	Marginal	4.0	3.0	2.0	1.0
			Copula	3.0	1.0	2.0	4.0
			Total	4.0	3.0	2.0	1.0
	overall	14	Marginal	3.0	4.0	1.0	2.0
			Copula	2.0	3.0	1.0	4.0
			Total	2.0	3.0	1.0	4.0

Table 6.5: Ranking of scenarios with certain parameter combinations, based on $\widehat{\text{eff}}_{rmse}$.

		Num		MBP1	MBP2	DIR	IFM
$n = 100$	$\nu = 3$, $\rho = 0.3$	3	Marginal	2.0	1.0	4.0	3.0
			Copula	1.0	4.0	3.0	2.0
			Total	1.0	2.0	4.0	3.0
	$\nu = 3$, $\rho = 0.8$	3	Marginal	1.0	2.0	4.0	3.0
			Copula	2.5	4.0	2.5	1.0
			Total	1.0	4.0	3.0	2.0
	$\nu = 10$, $\rho = 0.3$	3	Marginal	2.0	1.0	4.0	3.0
			Copula	2.0	3.0	4.0	1.0
			Total	2.0	3.0	4.0	1.0
	$\nu = 10$, $\rho = 0.8$	3	Marginal	1.0	3.0	4.0	2.0
			Copula	3.5	3.5	2.0	1.0
			Total	2.0	3.5	3.5	1.0
$n = 500$	$\nu = 3$, $\rho = 0.3$	3	Marginal	3.0	4.0	1.0	2.0
			Copula	1.0	2.0	3.0	4.0
			Total	2.0	3.0	1.0	4.0
	$\nu = 3$, $\rho = 0.8$	3	Marginal	2.0	4.0	1.0	3.0
			Copula	2.0	3.0	1.0	4.0
			Total	2.0	3.0	1.0	4.0
	$\nu = 10$, $\rho = 0.3$	3	Marginal	4.0	3.0	1.0	2.0
			Copula	1.5	1.5	3.0	4.0
			Total	4.0	3.0	1.0	2.0
	$\nu = 10$, $\rho = 0.8$	3	Marginal	3.0	4.0	2.0	1.0
			Copula	2.0	3.0	1.0	4.0
			Total	2.0	4.0	1.0	3.0

Table 6.6: Ranking of scenarios with certain parameter combinations, based on $\widehat{\text{eff}}_{rmse}$

Scenario	ν_1	ν_2	ν	ρ	n		MBP1	MBP2	DIR	IFM
						Marginal	1.5	1.5	4.0	3.0
S01	3.0	3.0	3.0	0.3	100	Copula	1.5	1.5	4.0	3.0
						Total	1.5	1.5	4.0	3.0
						Marginal	2.5	2.5	1.0	4.0
S02	3.0	3.0	3.0	0.3	500	Copula	1.5	1.5	4.0	3.0
						Total	2.0	2.0	2.0	4.0
						Marginal	1.0	4.0	2.5	2.5
S03	3.0	3.0	3.0	0.8	100	Copula	2.5	2.5	1.0	4.0
						Total	2.0	3.0	1.0	4.0
						Marginal	3.5	3.5	1.0	2.0
S04	3.0	3.0	3.0	0.8	500	Copula	1.0	3.0	2.0	4.0
						Total	2.0	3.0	1.0	4.0
						Marginal	1.0	2.5	4.0	2.5
S05	10.0	10.0	10.0	0.3	100	Copula	3.0	3.0	3.0	1.0
						Total	2.0	3.0	4.0	1.0
						Marginal	3.5	3.5	1.5	1.5
S06	10.0	10.0	10.0	0.3	500	Copula	2.0	1.0	4.0	3.0
						Total	4.0	3.0	2.0	1.0
						Marginal	1.0	4.0	2.5	2.5
S07	10.0	10.0	10.0	0.8	100	Copula	2.5	2.5	4.0	1.0
						Total	2.0	3.0	4.0	1.0
						Marginal	1.5	1.5	3.0	4.0
S08	10.0	10.0	10.0	0.8	500	Copula	2.0	3.0	1.0	4.0
						Total	1.0	3.0	2.0	4.0
						Marginal	1.5	1.5	3.5	3.5
S09	20.0	20.0	20.0	0.3	100	Copula	3.0	4.0	1.0	2.0
						Total	1.0	2.0	3.0	4.0
						Marginal	1.0	2.0	3.5	3.5
S10	20.0	20.0	20.0	0.3	500	Copula	4.0	3.0	1.0	2.0
						Total	2.5	2.5	1.0	4.0
						Marginal	1.0	2.0	3.5	3.5
S11	20.0	20.0	20.0	0.8	100	Copula	3.5	3.5	2.0	1.0
						Total	1.5	3.5	3.5	1.5
						Marginal	1.0	2.0	4.0	3.0
S12	20.0	20.0	20.0	0.8	500	Copula	2.5	4.0	1.0	2.5
						Total	1.0	3.5	2.0	3.5
						Marginal	1.5	1.5	3.0	4.0
S13	3.0	3.0	10.0	0.3	100	Copula	2.5	2.5	4.0	1.0
						Total	2.0	2.0	4.0	2.0
						Marginal	2.5	2.5	4.0	1.0
S14	3.0	3.0	10.0	0.3	500	Copula	2.5	2.5	2.5	2.5
						Total	2.5	2.5	4.0	1.0
						Marginal	2.5	2.5	4.0	1.0
S15	3.0	3.0	10.0	0.8	100	Copula	3.0	3.0	3.0	1.0
						Total	2.5	2.5	4.0	1.0
						Marginal	4.0	3.0	1.0	2.0
S16	3.0	3.0	10.0	0.8	500	Copula	4.0	2.0	1.0	3.0
						Total	4.0	2.5	1.0	2.5
						Marginal	4.0	1.0	2.5	2.5
S17	3.0	10.0	3.0	0.3	100	Copula	1.0	4.0	2.5	2.5
						Total	2.5	2.5	2.5	2.5
						Marginal	2.5	2.5	1.0	4.0
S18	3.0	10.0	3.0	0.3	500	Copula	1.0	4.0	2.5	2.5
						Total	2.0	3.0	1.0	4.0
						Marginal	2.0	4.0	1.0	3.0
S19	3.0	10.0	3.0	0.8	100	Copula	4.0	2.0	1.0	3.0
						Total	2.0	4.0	1.0	3.0
						Marginal	2.5	2.5	2.5	2.5
S20	3.0	10.0	3.0	0.8	500	Copula	2.0	2.0	2.0	4.0
						Total	2.0	2.0	2.0	4.0
						Marginal	3.5	1.5	3.5	1.5
S21	3.0	10.0	10.0	0.3	100	Copula	2.0	3.0	4.0	1.0
						Total	3.0	2.0	4.0	1.0
						Marginal	3.5	3.5	2.0	1.0
S22	3.0	10.0	10.0	0.3	500	Copula	1.5	1.5	4.0	3.0
						Total	3.0	3.0	3.0	1.0
						Marginal	4.0	2.0	3.0	1.0
S23	3.0	10.0	10.0	0.8	100	Copula	3.0	4.0	2.0	1.0
						Total	3.5	3.5	2.0	1.0
						Marginal	4.0	3.0	1.0	2.0
S24	3.0	10.0	10.0	0.8	500	Copula	3.0	3.0	1.0	3.0
						Total	4.0	3.0	1.0	2.0
						Marginal	1.5	1.5	4.0	3.0
S25	10.0	10.0	3.0	0.3	100	Copula	1.0	2.0	4.0	3.0
						Total	1.0	2.0	4.0	3.0
						Marginal	3.0	3.0	3.0	1.0
S26	10.0	10.0	3.0	0.3	500	Copula	1.5	1.5	4.0	3.0
						Total	2.5	2.5	4.0	1.0
						Marginal	4.0	1.0	2.5	2.5
S27	10.0	10.0	3.0	0.8	100	Copula	2.5	2.5	1.0	4.0
						Total	3.0	2.0	1.0	4.0
						Marginal	1.0	4.0	2.5	2.5
S28	10.0	10.0	3.0	0.8	500	Copula	3.0	2.0	1.0	4.0
						Total	2.0	3.0	1.0	4.0
						$\sum$ Marginal	1.0	2.0	4.0	3.0
						$\sum$ Copula	1.0	4.0	2.0	3.0
						$\sum$ Total	1.0	4.0	2.0	3.0

Table 6.7: Ranking of all 28 Scenarios, based on $\widehat{\text{eff}}_{rmse}$

Chapter 7

Application: Swap Rates

In this chapter we consider the swap data from Kuhn (2001). This data was originally provided by the risk management department of HypoVereinsbank. It is about an annually compounded zero coupon swap rate of the currency Euro from 12/07/1988 to 05/21/2001 with maturities of 2 years, 3 years, 5 years, 7 years and 10 years. A *Zero-Coupon Swap* is an interest rate swap agreement with the fixed-rate side based on a zero-coupon bond. With agreement of the counterparty, the swap agreement may call for a single fixed payment at maturity by the holder of the zero. The payments on the other side may follow typical swap interim payment schedules. Because of the payment mismatch, a zero-coupon swap exposes one of the counterparties to significant credit risk and is the functional equivalent of a loan.

In risk management the analysis of dependence between the swap rate of different maturities of a currency is one of the main interests. In Figure 7.1 the historical prices of the swap data are illustrated.

7.1 Data Transformation

In order to apply the bivariate meta t-distribution for the swap data, we need i.i.d. data. The data transformation described in this section follows the master thesis Baumann (2007). The correlations of a paired data set with conditioned expectations and the conditioned variance are expressed by a ARMA(1,1) defined in (2.62) and a GARCH(1,1) model defined in (2.65). We receive the time series i, $i = 1, ..., 5$, with the following equations for the value of $X_{i,t}$:

$$\begin{aligned} X_{i,t} &= \mu_i + \psi_i X_{i,t-1} + \sigma_{i,t} Z_{i,t} + \theta_i \sigma_{i,t-1} Z_{i,t-1} \\ \sigma^2_{i,t} &= a_{i,0} + a_{i,1}\epsilon^2_{i,t-1} + b_{i,1}\sigma^2_{i,t-1}, \end{aligned}$$

whereby $E[Z_{i,t}] = 0$, $Var[Z_{i,t}] = 1$ and $\epsilon_{i,t-1} = \sigma_{i,t} Z_{i,t}$.
For further analysis we need the standardized residuals $\widehat{\boldsymbol{Z}}_i$ in this model, where $\widehat{\boldsymbol{Z}}_i := (\widehat{Z}_{1,1}, ..., \widehat{Z}_{1,n})$ with $t = 1, ..., n$. After the marginal parameter estimation we receive:

$$\hat{Z}_{i,t} = \frac{1}{\hat{\sigma}_{i,t}}(X_{i,t} - \hat{\mu_i} - \hat{\psi}_i X_{i,t-1} - \hat{\theta}_i \hat{\sigma}_{i,t-1} \hat{Z}_{i,t-1}) \tag{7.1}$$

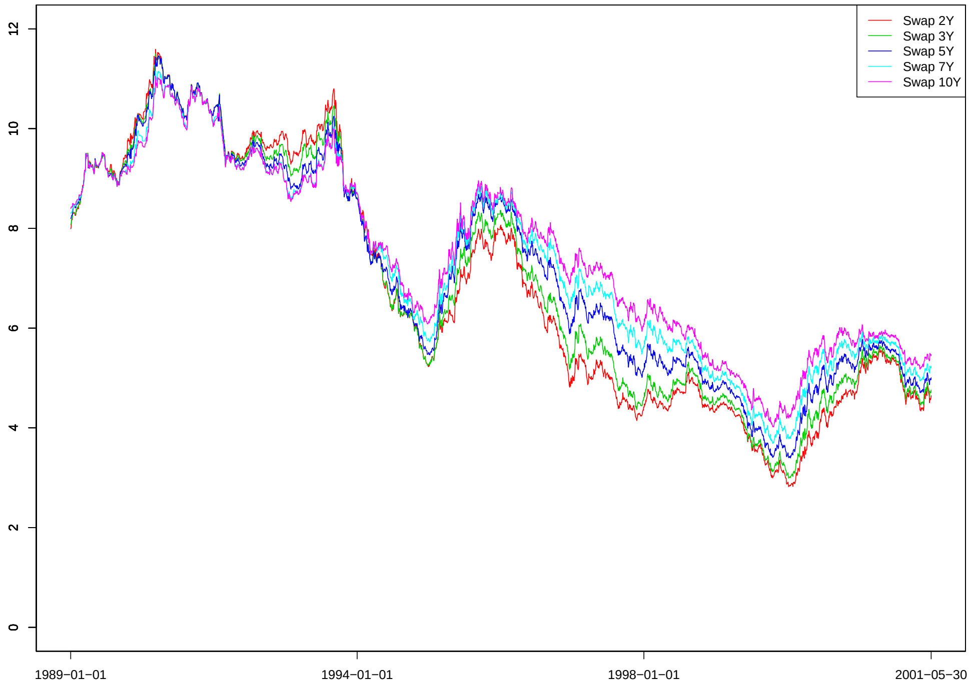


Figure 7.1: Time series of swap data

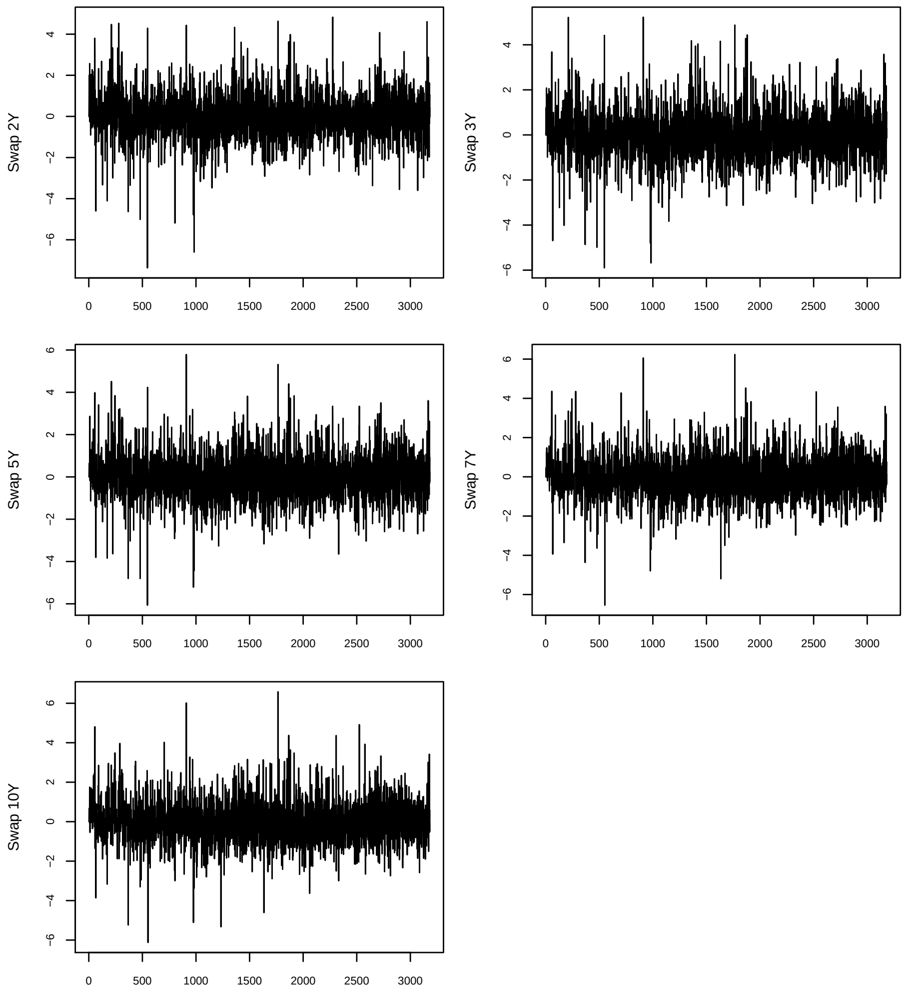


Figure 7.2: GARCH filtered residuals of swap data

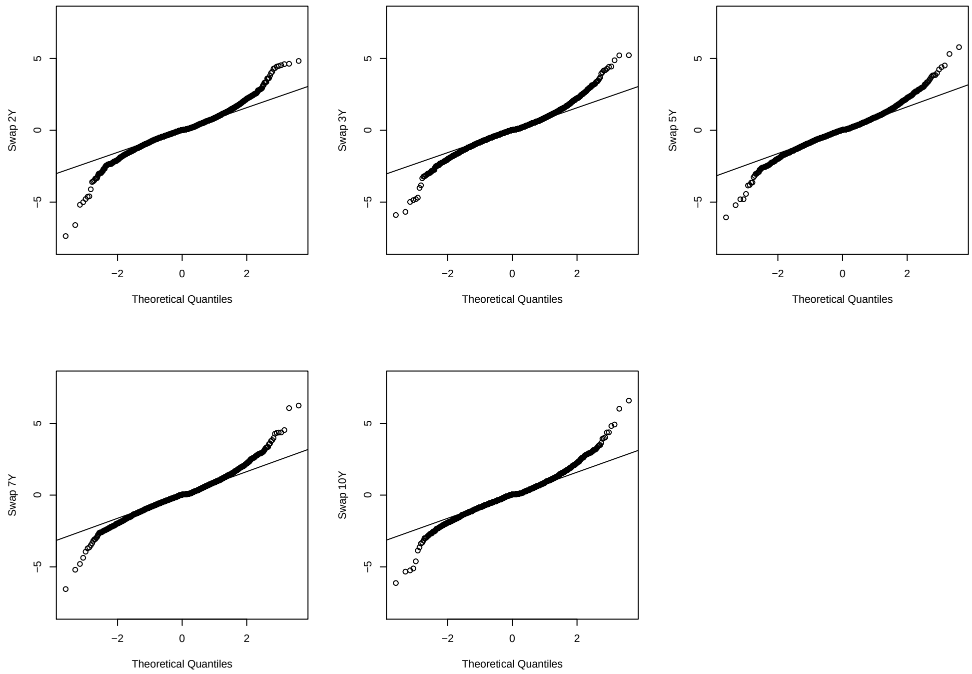


Figure 7.3: Normal QQ-plots of GARCH filtered residuals of swap data

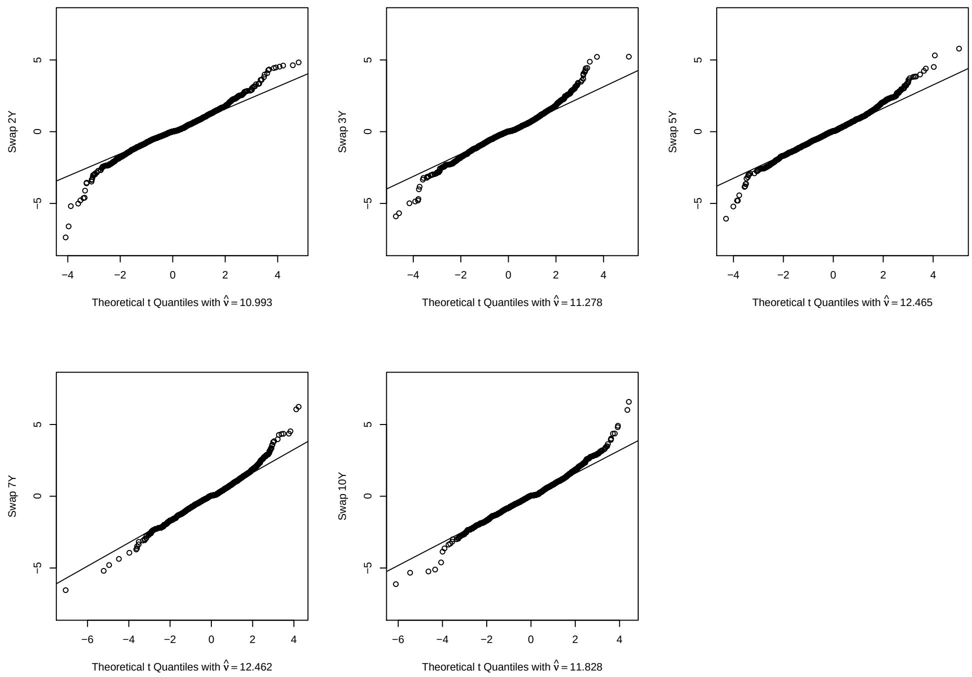


Figure 7.4: t QQ-plots of GARCH filtered residuals of swap data using estimated degrees of freedom by IFM method

In addition, Baumann (2007) has examined the independence of the residuals can be examined using the Ljung-Box Test. For more information on the test, see Baumann (2007). Figure 7.2 displays the GARCH filtered residuals of the swap data.

7.2 Results and Interpretations

For the analysis we have five margins where each margin is set as one data set with maturities of 2 years, 3 years, 5 years, 7 years and 10 years. Thus there are 10 paired data sets in total to be analyzed. We apply the bivariate meta t-model to the data sets and estimate the parameters ν_1, ν_2, ν and ρ using MBP1 (non-adaptive), MBP2 (adaptive), DIR (direct maximization) and IFM (inference for margins) method. In Table 7.1 we can see the results of the estimated parameters $\hat{\nu}_1$, $\hat{\nu}_2$, $\hat{\nu}$ and $\hat{\rho}$ of the 10 pairs of data sets. The notation S2Y:S3Y indicates that a bivariate data set consisting of a filtered swap data set with maturity of 2 years and a filtered swap data set with maturity of 3 years. Similarly the notation is applied to other bivariate data sets. Furthermore, the results demonstrate that MBP1 and MBP2 yield similar results as the DIR method. The most data pairs have similar marginal distributions with parameter ν_1 and ν_2 which lie in a range of [6.2, 7.5], while their joint parameter ν is relatively small, lying between [1.9, 5.1]. In addition the joint swap data sets with different maturity are highly correlated. They are in the range of [0.84, 0.97]. The similar values of the log likelihood of the MBP1, MBP2, DIR affirm that the MBP1 and MBP2 methods calculate 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 make an investigation using the likelihood ratio test. In the likelihood ratio test we also look at the bivariate t-model compared to the meta t-model. In the bivariate meta t-model we assume that the marginal distributions of each data pair have different degrees of freedom as its joint distribution while in the bivariate t-model we assume that the degrees of freedom of the marginal distribution and the joint distribution are the same, thus the bivariate t-model has only two parameters ν and ρ. For the bivariate t-model we estimate the parameters using the MbP algorithm (MbP0) with $\rho_0 := \hat{\rho}_\tau$, the estimated correlation using Kendall's τ, direct maximization (DIR), semi-parametric method (CSP) and the canonical maximization method (CML). Table 7.2 displays the estimated degrees of freedom $\hat{\nu}$ is in a range of [4.5, 6.5] and the estimated correlations lying in a range [0.78, 0.96] with the bivariate t-model. Furthermore, both the estimation results using the bivariate t-model and the meta t-model demonstrate that the log likelihood values by the MbP algorithm and direct maximization method are evidently at their maximum while the IFM method for the meta t-model and both CSP and CML for the bivariate t-model have significant deviations in their estimated log likelihood values.

With both bivariate t and bivariate meta t-model we construct a test with the null hypothesis $H_0 : \nu_1 = \nu_2 = \nu$ versus the alternative $H_1 : \nu_1 \neq \nu_2 \neq \nu$. We use the MBP1 (non-adaptive) method for parameter estimation in the meta t-model and The test statistics is defined by

$$\chi_2^2 \sim -2(\ell(\hat{\nu}, \hat{\rho}; \boldsymbol{x}) - \ell(\hat{\nu}_1, \hat{\nu}_2, \hat{\nu}, \hat{\rho}; \boldsymbol{x}))$$

where $\ell(\hat{\nu}, \hat{\rho}; \boldsymbol{x})$ is the log likelihood of the bivariate t-model defined in (3.2) and $\ell(\hat{\nu}_1, \hat{\nu}_2, \hat{\nu}, \hat{\rho}; \boldsymbol{x})$ is the log likelihood of the bivariate meta t-model defined in (5.2). In the Table 7.3 we can observe that based on different α-levels $\alpha = 0.01$ $\alpha = 0.025$ and $\alpha = 0.05$ the null hypothesis is accepted for the pairs S2Y:S10Y and S3Y:S10Y and rejected for the rest of the data pairs.

	Method	S2Y : S3Y	S2Y : S5Y	S2Y : S7Y	S2Y : S10Y	S3Y : S5Y	S3Y : S7Y	S3Y : S10Y	S5Y : S7Y	S5Y : S10Y	S7Y : S10Y
$\hat{\nu}_1$	MBP1	6.366	6.884	6.731	6.680	6.984	6.688	6.845	7.132	7.511	6.339
	MBP2	6.366	6.881	6.728	6.680	6.983	6.689	6.844	7.130	7.510	6.339
	DIR	6.362	6.877	6.725	6.677	6.978	6.684	6.838	7.127	7.506	6.336
	IFM	10.993	10.993	10.993	10.993	11.278	11.278	11.278	12.465	12.465	12.462
$\hat{\nu}_2$	MBP1	6.241	6.846	6.488	6.228	6.849	6.684	6.785	7.170	7.212	6.399
	MBP2	6.241	6.843	6.485	6.229	6.848	6.684	6.784	7.168	7.210	6.398
	DIR	6.234	6.839	6.482	6.225	6.843	6.679	6.780	7.166	7.207	6.396
	IFM	11.278	12.465	12.462	11.828	12.465	12.462	11.828	12.462	11.828	11.828
$\hat{\nu}$	MBP1	3.167	4.063	4.465	5.093	3.111	4.159	4.983	2.991	3.611	1.954
	MBP2	3.166	4.054	4.455	5.102	3.109	4.162	4.977	2.988	3.606	1.954
	DIR	3.165	4.048	4.450	5.090	3.106	4.151	4.961	2.986	3.604	1.952
	IFM	2.436	3.466	4.293	5.176	2.701	3.797	4.411	2.623	3.059	1.760
$\hat{\rho}$	MBP1	0.958	0.915	0.865	0.841	0.954	0.909	0.884	0.942	0.921	0.968
	MBP2	0.958	0.915	0.865	0.841	0.954	0.909	0.884	0.942	0.921	0.968
	DIR	0.958	0.915	0.865	0.841	0.954	0.909	0.884	0.942	0.921	0.968
	IFM	0.938	0.880	0.816	0.785	0.935	0.875	0.841	0.920	0.892	0.955
loglike	MBP1	−5374.532	−6438.849	−7121.341	−7352.185	−5501.220	−6557.607	−6897.655	−5841.373	−6295.258	−4904.250
	MBP2	−5374.532	−6438.848	−7121.341	−7352.185	−5501.220	−6557.607	−6897.655	−5841.372	−6295.258	−4904.250
	DIR	−5374.532	−6438.848	−7121.341	−7352.185	−5501.220	−6557.607	−6897.655	−5841.372	−6295.258	−4904.249
	IFM	−5426.637	−6483.500	−7163.900	−7396.200	−5543.411	−6599.868	−6939.750	−5877.965	−6332.553	−4946.939
time in sec.	MBP1	4186.870	4183.275	2986.636	3849.041	3919.012	4416.001	3711.954	5260.335	2944.865	4511.678
	MBP2	3614.028	4049.451	2937.919	3783.164	3732.953	4620.764	3698.060	3967.770	3582.316	5268.276
	DIR	84.728	79.457	80.149	72.228	86.027	70.827	75.919	72.150	63.655	83.185
	IFM	91.155	74.179	80.760	78.893	75.567	79.406	65.865	83.119	73.080	86.879
iter	MBP1	29	29	28	30	30	29	28	29	28	30
	MBP2	26	27	27	26	27	26	25	27	26	27
	DIR	1	1	1	1	1	1	1	1	1	1
	IFM	1	1	1	1	1	1	1	1	1	1

Table 7.1: Estimated parameters using 4 different estimation methods applied to *bivariate meta t-model* for all bivariate combinations of the filtered Euro swap data from year 1989-01-01 to 2001-05-30

	Method	S2Y : S3Y	S2Y : S5Y	S2Y : S7Y	S2Y : S10Y	S3Y : S5Y	S3Y : S7Y	S3Y : S10Y	S5Y : S7Y	S5Y : S10Y	S7Y : S10Y
$\hat{\nu}$	MbP0	5.399	6.171	6.111	6.210	5.657	5.997	6.376	5.552	6.055	4.531
	DIR	5.369	6.144	6.085	6.188	5.633	5.969	6.355	5.526	6.030	4.501
	CSP	2.827	3.560	4.458	5.162	2.815	4.115	4.867	2.967	3.526	2.053
	CML	2.827	3.560	4.458	5.162	2.815	4.115	4.867	2.967	3.526	2.053
$\hat{\rho}$	MbP0	0.962	0.919	0.869	0.844	0.958	0.913	0.887	0.948	0.927	0.974
	DIR	0.962	0.919	0.870	0.844	0.958	0.913	0.887	0.948	0.927	0.974
	CSP	0.938	0.880	0.816	0.785	0.935	0.875	0.841	0.920	0.892	0.955
	CML	0.938	0.880	0.816	0.785	0.935	0.875	0.841	0.920	0.892	0.955
loglike	MbP0	−5392.700	−6448.136	−7126.642	−7354.132	−5526.481	−6565.511	−6900.629	−5873.417	−6314.195	−4975.232
	DIR	−5392.696	−6448.134	−7126.640	−7354.131	−5526.479	−6565.509	−6900.628	−5873.414	−6314.193	−4975.224
	CSP	−5721.463	−6659.272	−7254.745	−7449.286	−5829.937	−6726.084	−7026.014	−6161.453	−6539.489	−5478.124
	CML	−5735.056	−6663.230	−7260.247	−7453.515	−5845.894	−6734.152	−7030.895	−6174.264	−6547.772	−5522.019
time in sec.	MbP0	54.270	65.920	54.230	53.240	60.980	53.840	51.330	57.770	57.200	62.200
	DIR	0.740	0.560	0.510	1.020	0.810	0.530	0.470	0.640	0.660	0.800
	CSP	13.620	13.610	17.890	21.910	12.830	16.260	19.670	9.590	12.850	15.270
	CML	101.120	42.840	39.780	62.970	95.990	40.030	45.750	70.890	40.040	154.140
iter	MbP0	16	16	15	15	17	15	15	16	15	17
func.eval	DIR	21	20	18	16	30	18	18	23	19	27
	CSP	7	7	10	5	7	9	11	5	7	8
	CML	31	13	14	18	34	14	15	25	14	51

Table 7.2: Estimated parameters using 4 different estimation methods applied to *bivariate t-model* for all bivariate combinations of the filtered Euro swap data from year 1989-01-01 to 2001-05-30

		S2Y : S3Y	S2Y : S5Y	S2Y : S7Y	S2Y : S10Y	S3Y : S5Y	S3Y : S7Y	S3Y : S10Y	S5Y : S7Y	S5Y : S10Y	S7Y : S10Y
$\hat{\nu}$	Statistic	36.337	18.575	10.601	3.895	50.523	15.809	5.949	64.089	37.875	141.965
	p-value	0.000	0.000	0.005	0.143	0.000	0.000	0.051	0.000	0.000	0.000
	H_0, $\alpha = 0.01$	rejected	rejected	rejected	accepted	rejected	rejected	accepted	rejected	rejected	rejected
	H_0, $\alpha = 0.025$	rejected	rejected	rejected	accepted	rejected	rejected	accepted	rejected	rejected	rejected
	H_0, $\alpha = 0.05$	rejected	rejected	rejected	accepted	rejected	rejected	accepted	rejected	rejected	rejected

Table 7.3: Result from likelihood ratio test with $H_0 : \nu_1 = \nu_2 = \nu$ and different α-levels. Likelihood values are based on the results of MBP1 (non-adaptive) method for parameter estimation in the meta t-model and MbP0 method for parameter estimations in the bivariate t-model.

7.3 Rolling Windows

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 for the entire time horizon. However the economic environment often changes considerably and it may not reasonable to assume the constancy of a model's parameters. For this reason, it would be interesting to see how the parameters of the model vary with respect to the time. We will catch the changes of the parameters by the rolling window analysis of a time series model.

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 size of a window amounts up to 250 data points which implies the number of trading days in a year. The setting for the analysis states as follows:

Initial Settings:

- time span: December 07, 1988 to Jan 21, 2001
- data size: 3150
- number of the moving windows: 59
- data size in each moving window: 250
- overlapped data points of sequential windows: 50

Results and Interpretations:

The sequence of estimated values will depict the evolution of the parameters of interest in our models over time. The 59 rolling windows for each parameters ν_1, ν_2, ν and ρ are illustrated in Figures 7.5, 7.6, 7.7 and 7.8, respectively.

Apparently, the 59 rolling parameter estimates of each ν_1, ν_2, ν and ρ vary considerably over the data windows. In Figure 7.5 and 7.6 we can observe that for the most data pairs the estimated rolling marginal degrees of freedom parameters $\hat{\nu}_{1,h}$ and $\hat{\nu}_{2,h}$, $h = 1, ..., 59$ start at a very low value close to 3-5 in the beginning time period in 1989, rise to about 15-20 in 1991/1992, fall again close to 10 at the time around 1995 and increase in the period between 1997 and 1999 up to 40. For parameter ν_1 the rolling estimates eventually plummet to 10-20 at 2001 while for ν_2 the rolling estimates jumps up to 60-100. Furthermore, we can also observe that the rolling estimates fluctuate up to 100 occasionally during the entire time horizon. The fluctuations may be caused by the difficulty of estimations by each methods due to the small data size (n=250) or it could also indicate that data at those specific time intervals is almost normal distributed. Compared to the marginal degrees of freedom the rolling estimated joint degrees of freedom $\hat{\nu}_h$, $h = 1, ..., 59$, show much less extreme fluctuations over time. Figure 7.7 reveals that the values start out low around 3-5, move upward to 20 at the time around 1993-1995, stay volatile during the period from 1993 to 1997, until a decline of values in 1999. The estimated rolling degree of freedom $\hat{\nu}_h$'s of each data differ significantly at the end of time period. In Figure 7.8 we can take notice of coincidental movement of the rolling correlations estimated by MBP1, MBP2 and DIR methods while the estimated rolling correlations by IFM method differ from those estimated by other methods significantly.

Furthermore, in Figure 7.9 we compare the estimated parameters $\hat{\nu}_1$, $\hat{\nu}_2$ and $\hat{\nu}$ in bivariate meta t-model with the estimated degrees of freedom $\hat{\nu}_t$ in bivariate t-model. Figure 7.9 and its detailed plots in Figure 7.10 show 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 disaccord between each other by a considerable amount over the entire time interval. This also means that any significant deviation from the solid line (rolling estimations of ν_t) affirms that a meta t-model is needed. For rolling correlations ρ and ρ_t, respectively estimated by bivariate meta t-model and bivariate t-model, the differences are marginal and can be caught particularly in the periods between 1993-1995 and between 1997-1999. This result is displayed in Figure 7.11.

Finally, we calculate the estimated tail dependence defined in (2.13) of the data using the estimated joint parameters ν and ρ . Figure 7.12 discloses the significant difference of the estimated tail dependence by the bivariate meta t and bivariate t-models. Especially during the period between 1997 and 1999, the estimated tail dependence of the meta t-model is high, while the result of the bivariate t-model shows very low values. Besides we also notice that the rolling estimates of tail dependence of meta t-model has a higher volatility than the rolling estimates of the bivariate t-model which may indicates the accuracy of the estimation using bivariate meta t-model for the corresponding data pairs.

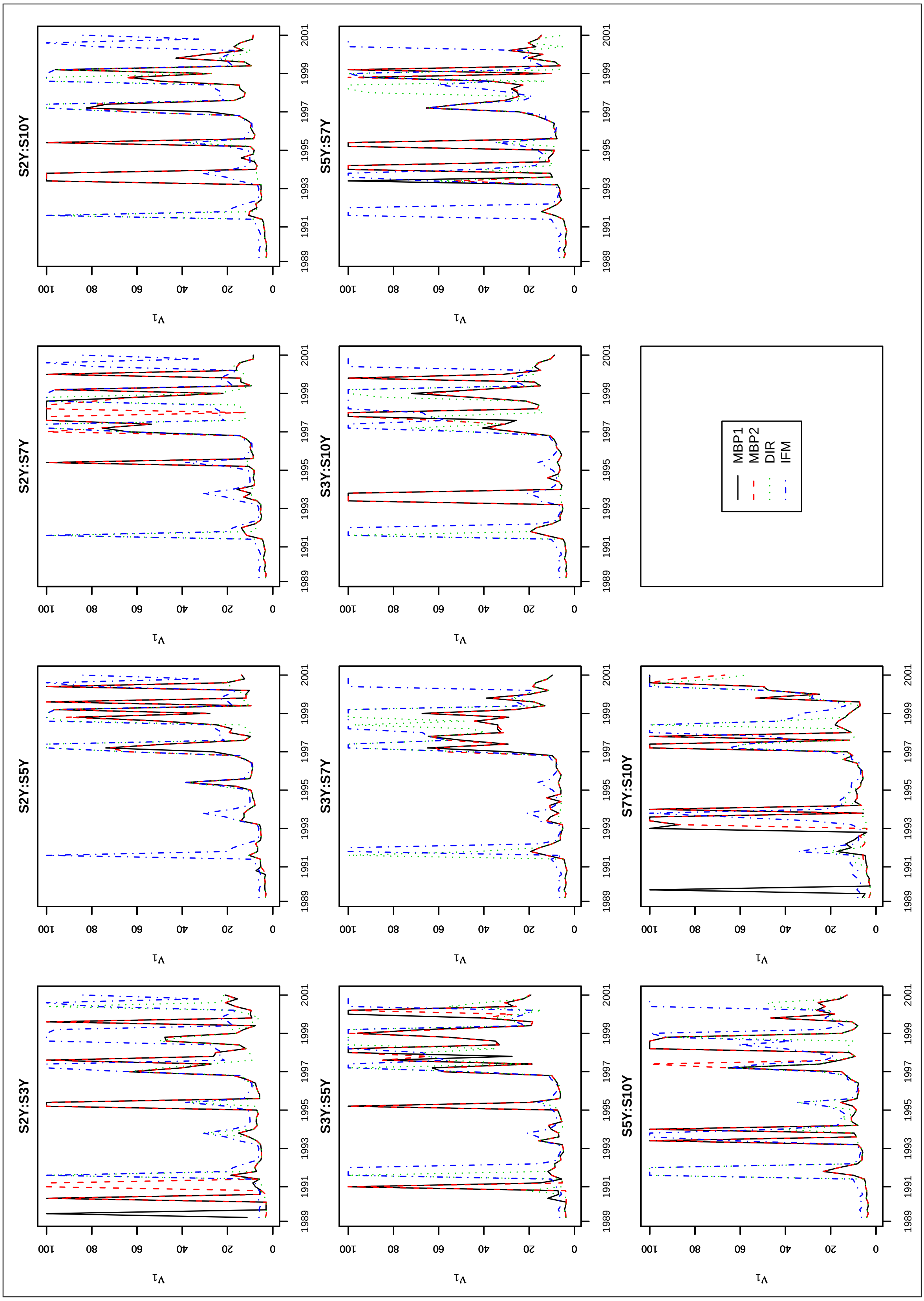


Figure 7.5: 59 rolling marginal degrees of freedom ν_1 estimated by MBP1, MBP2, Direct and IFM methods

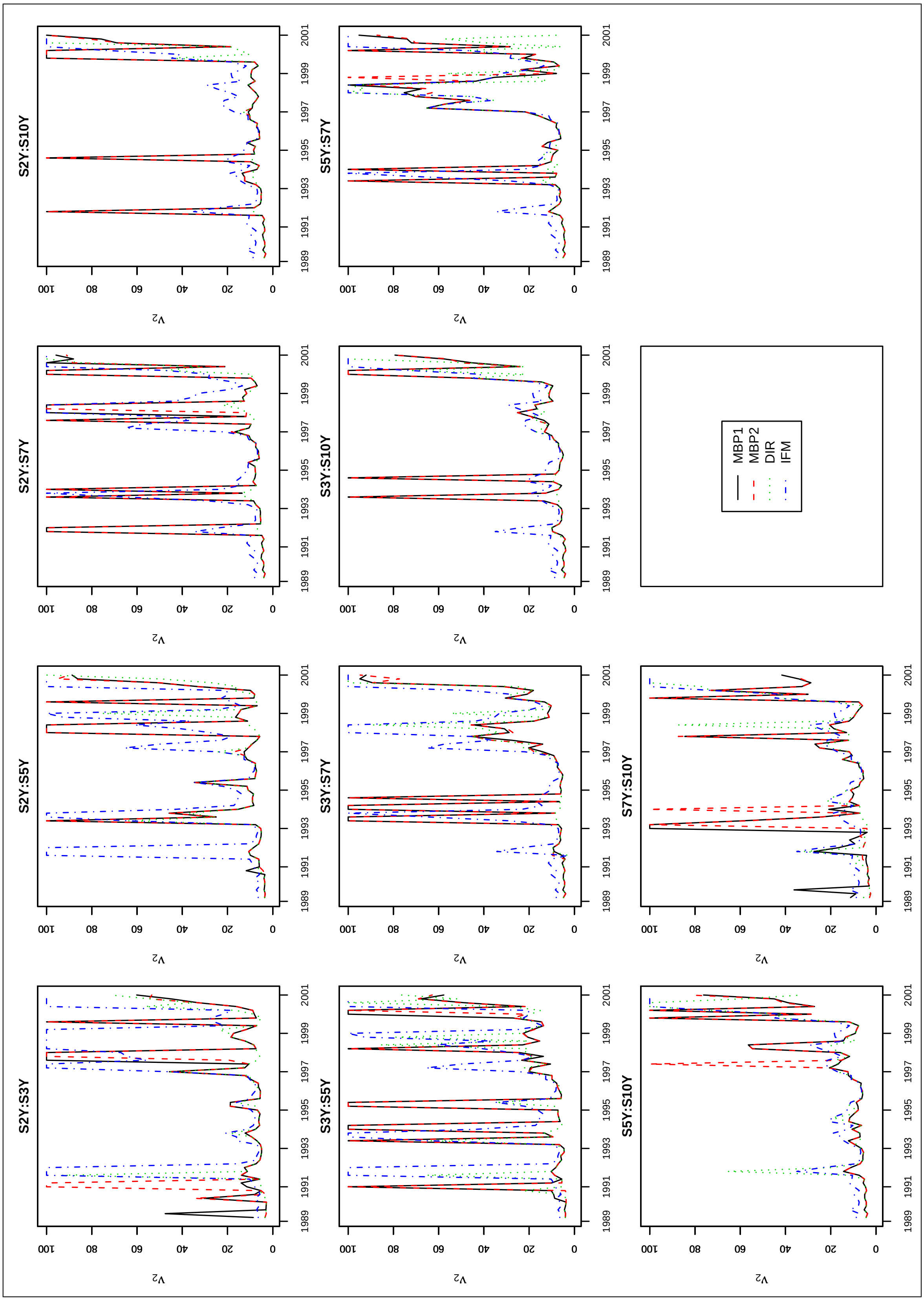


Figure 7.6: 59 rolling marginal degrees of freedom ν_2 estimated by MBP1, MBP2, Direct and IFM methods

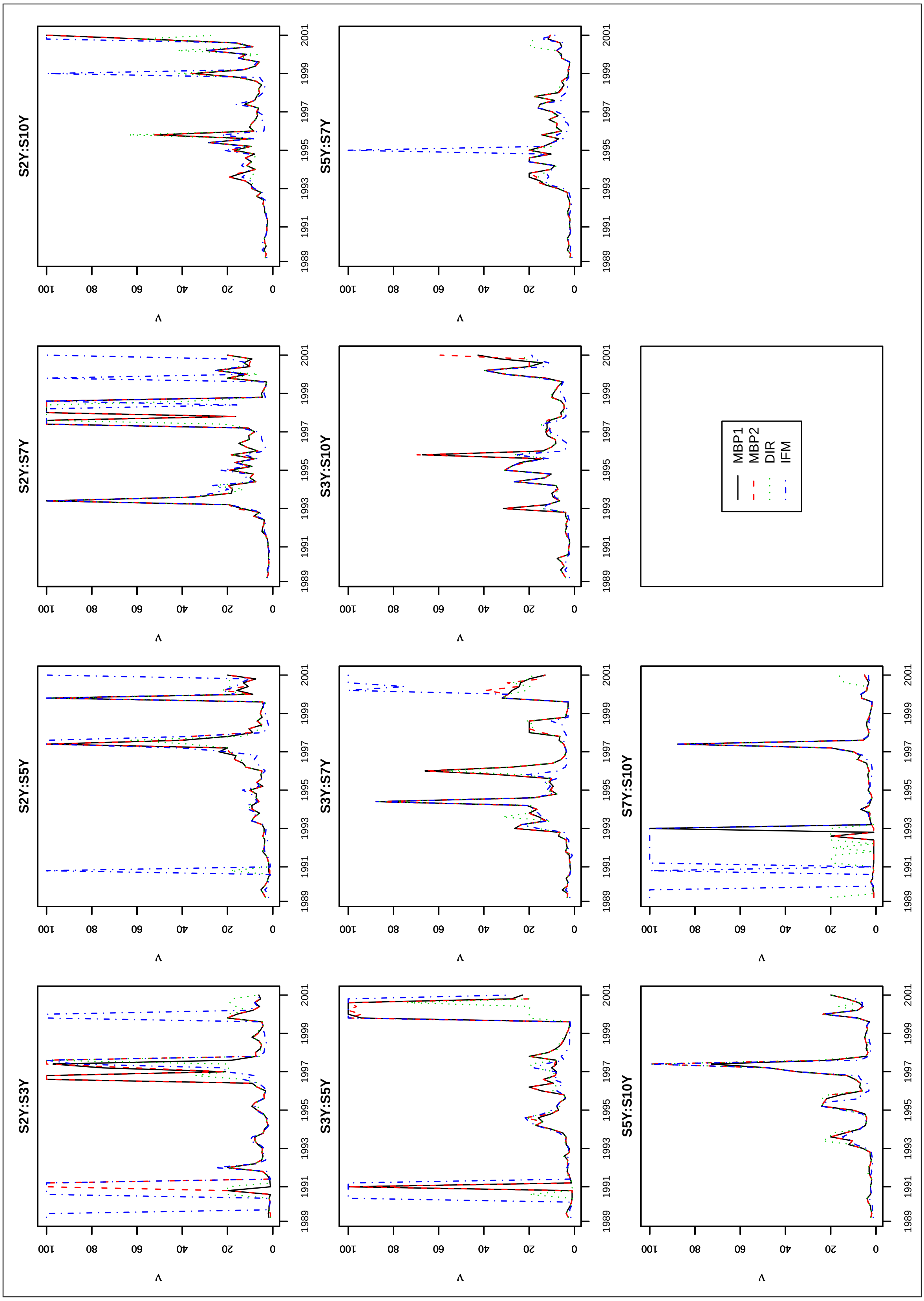


Figure 7.7: 59 rolling joint degrees of freedom ν estimated by MBP1, MBP2, Direct and IFM methods

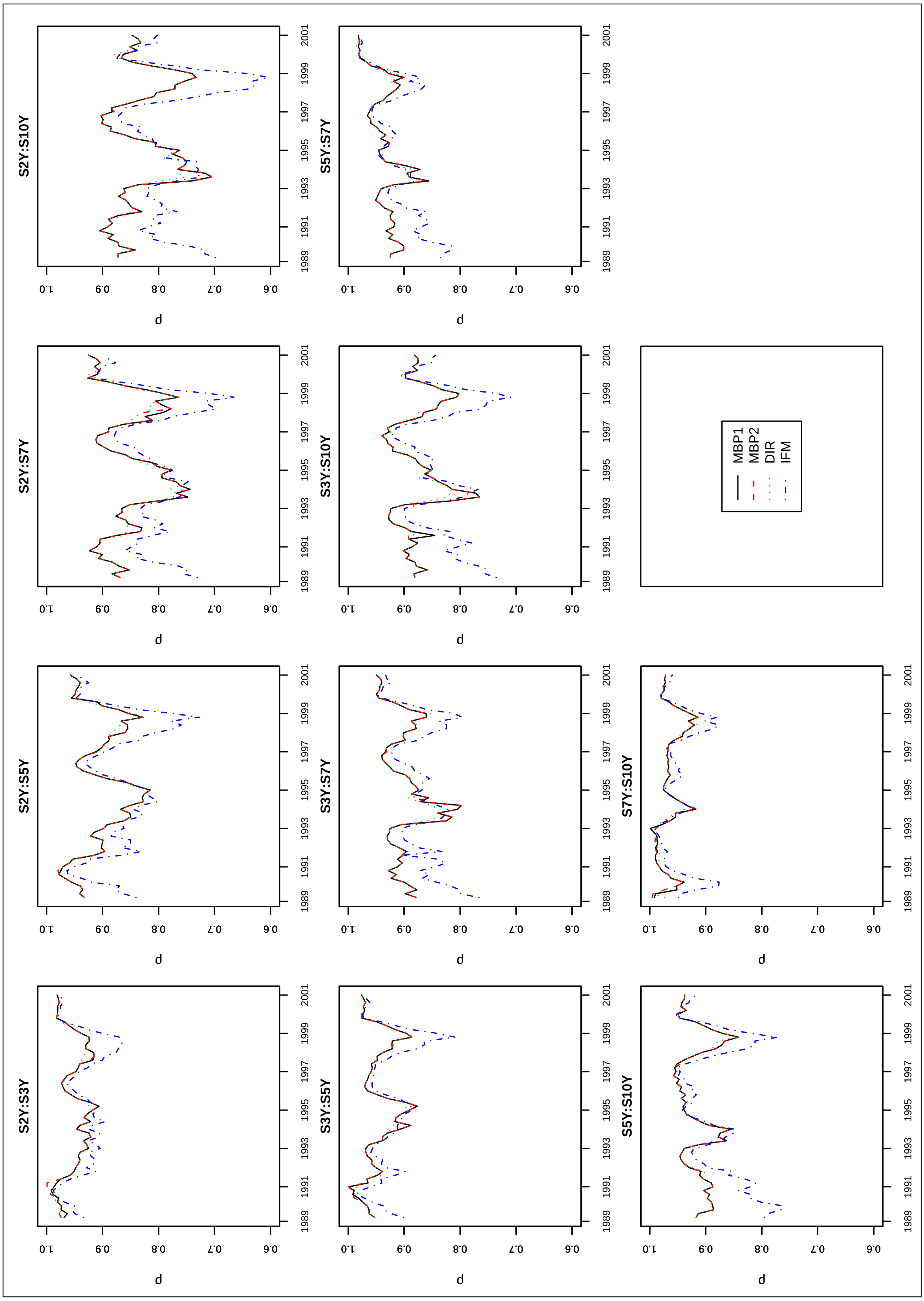


Figure 7.8: 59 rolling correlation ρ estimated by MBP1, MBP2, Direct and IFM methods

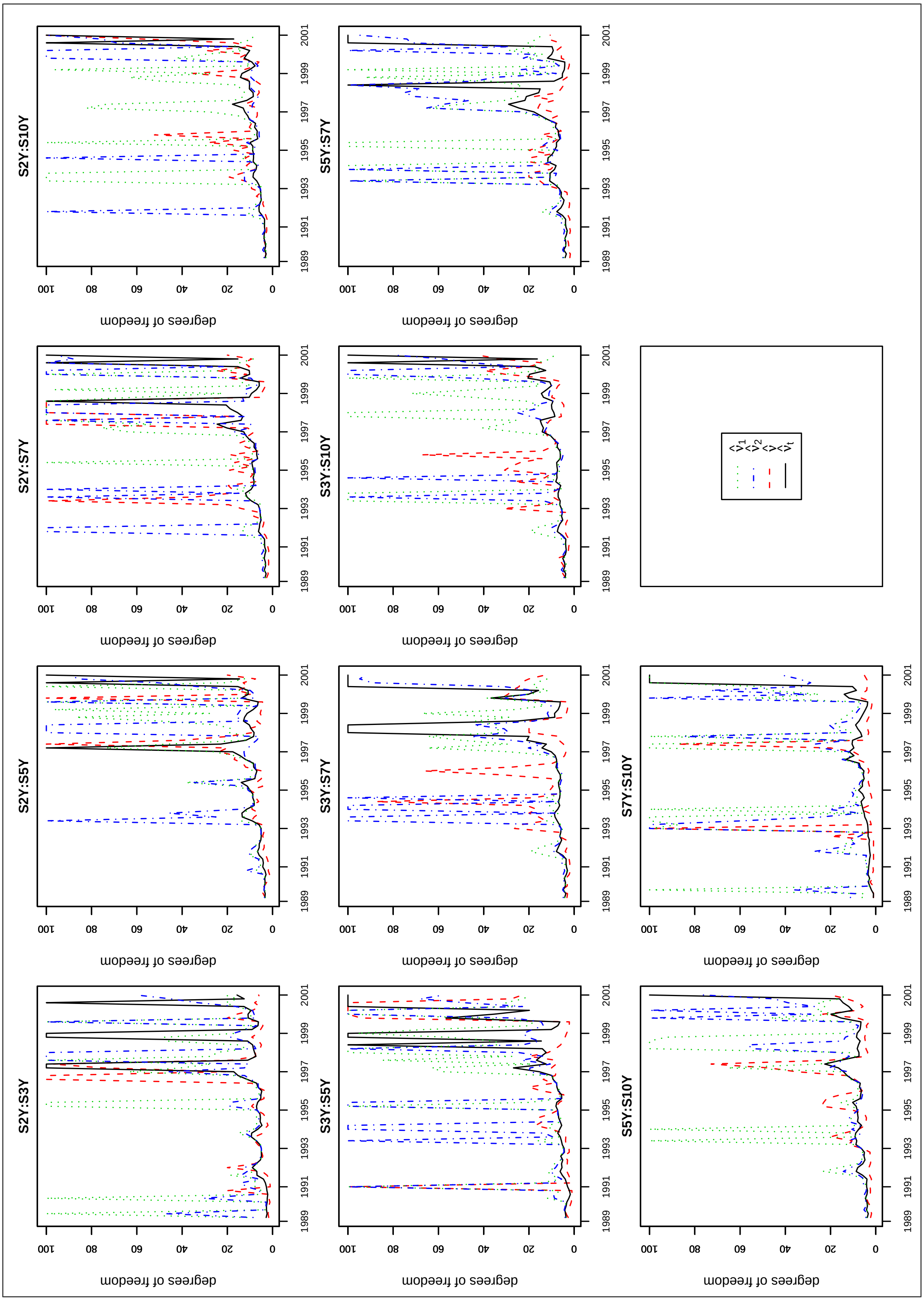


Figure 7.9: Comparison of estimated rolling degrees of freedom ν_1, ν_2, ν's of meta t-model based on MBP1 method and estimated rolling parameter $\hat{\nu}_t$ of bivariate t-model based on MbP0 method

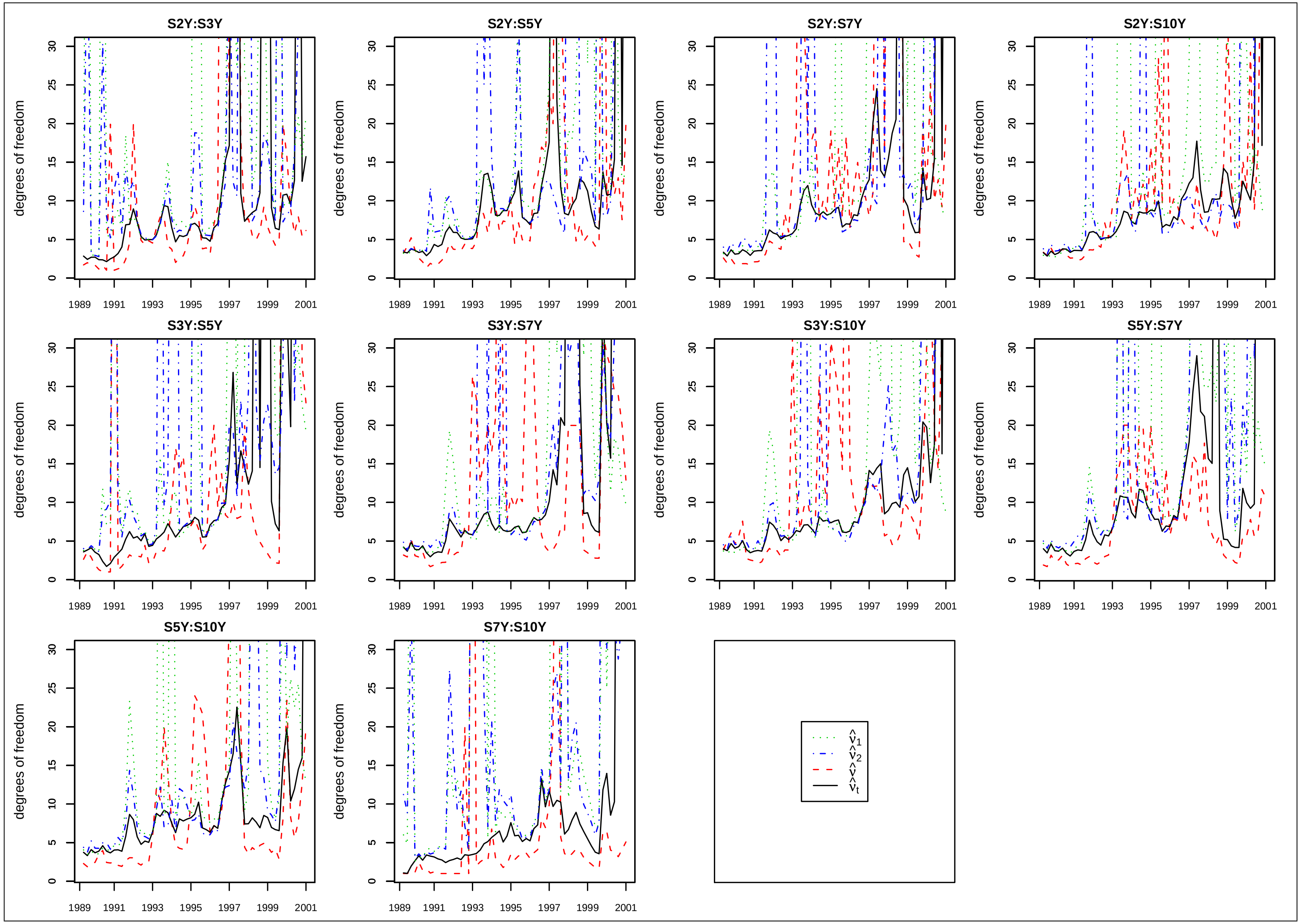


Figure 7.10: Detail plots; Comparison of estimated rolling degrees of freedom ν_1, ν_2, ν's of meta t-model based on MBP1 method and estimated rolling parameter $\hat{\nu}_t$ of bivariate t-model based on MbP0 method

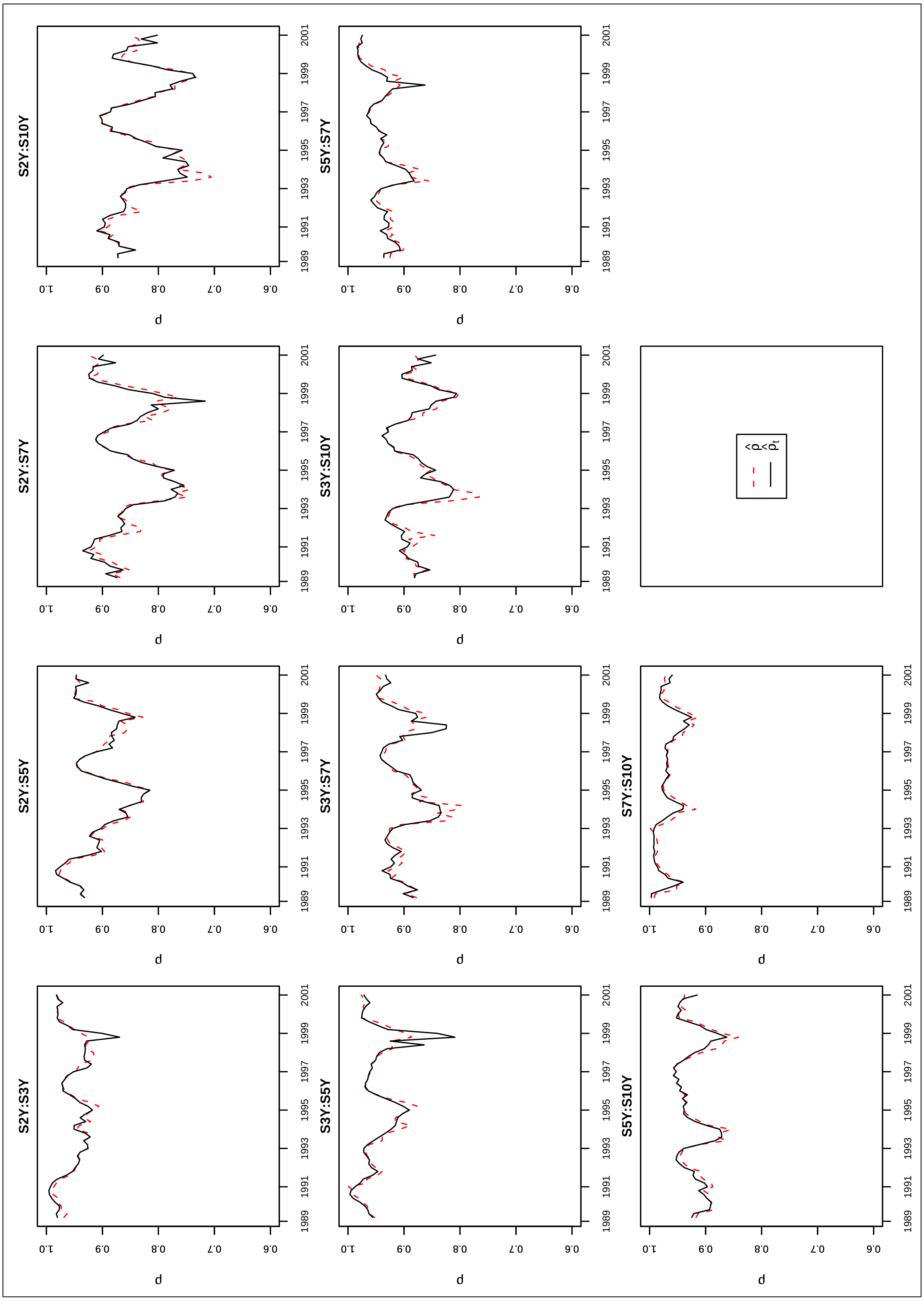


Figure 7.11: Comparison of rolling correlations ρ of meta t-model based on MBP1 method and rolling correlations ρ_t of bivariate t-model based on MbP0 method

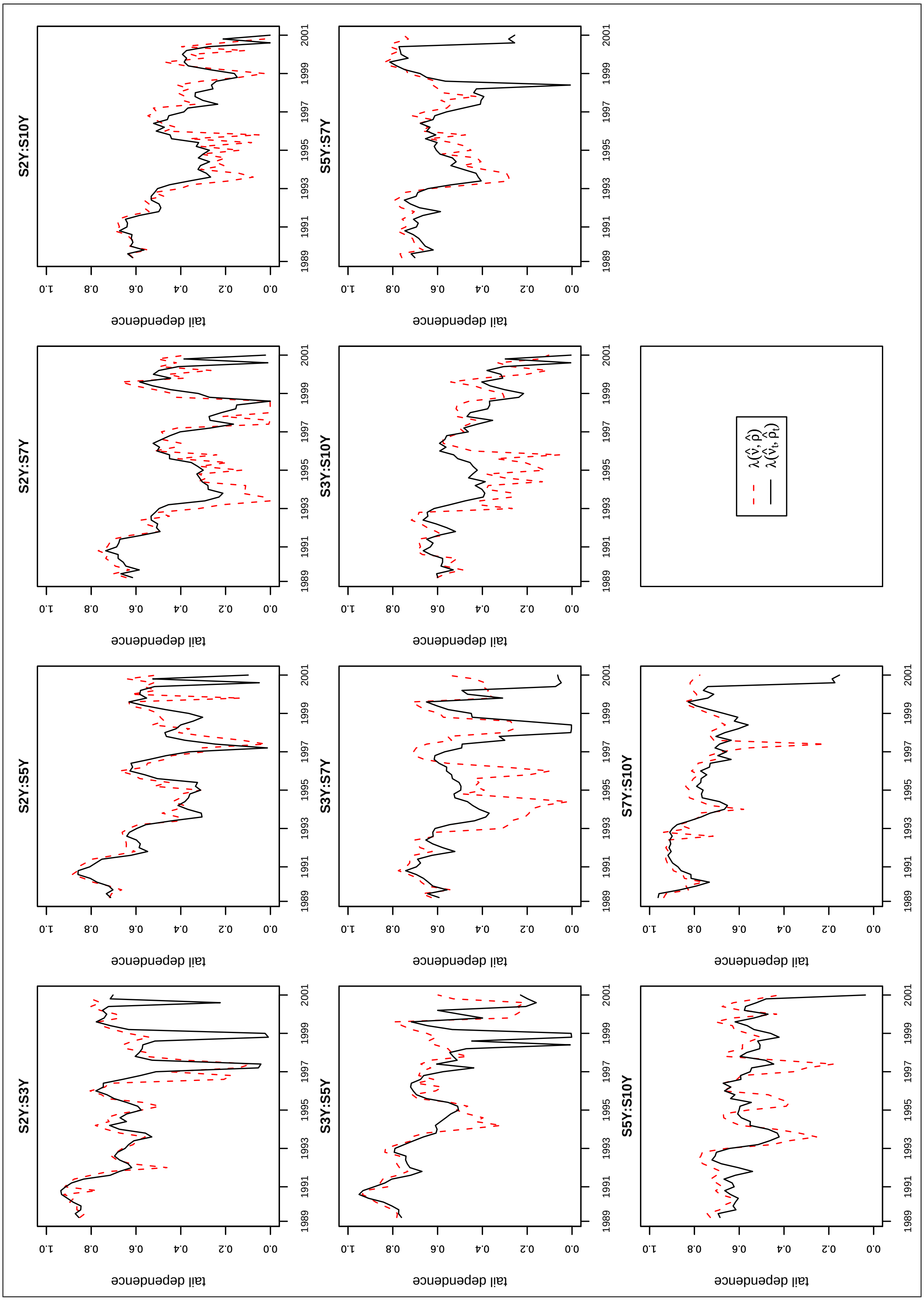


Figure 7.12: Comparison of rolling tail dependence between estimates of meta t-model based on MBP1 method and estimates of bivariate t-model based on MbP0 method

Chapter 8

Summary

In this thesis we have adopted the concept of Maximization by Parts from Song et al. (2005) to calculate the maximum likelihood estimators for parameters of the bivariate t and bivariate meta t-distributions. We were confronted with several difficulties in the calculation.

First, for the bivariate t case the challenge was to find the root of the score function of the modified working model to generate an initial value for the MbP iterations in Step k, $k \in \mathbb{N}$. Since the null of the function does not necessarily exist and we want the algorithm to iteratively find the MLE, we changed the shape of the modified working model by shifting the interchangeable parameter ρ_0 to guarantee a solution of the modified working model.

In the bivariate meta t-model the difficulty was more about choosing a simple and dominant modified working models to ensure the *information dominance*. For this reason one needs to find an appropriate working model. Adopted the idea of Song et al. (2005) we added the log likelihood function of the bivariate meta t-distribution with fixed copula parameters. Thus, the additional model was very complicated and caused lots of computing efforts. In order to reduce the computing cost we might choose a different working model. However, it is not an easy task, since the additional model has to satisfy the information condition, otherwise the algorithm will not iteratively converge to a fixed point which is the MLE. Therefore, for the study of the MbP algorithm it may be interesting to develop a concept to construct a simpler function for the working model which also guarantees the information dominance.

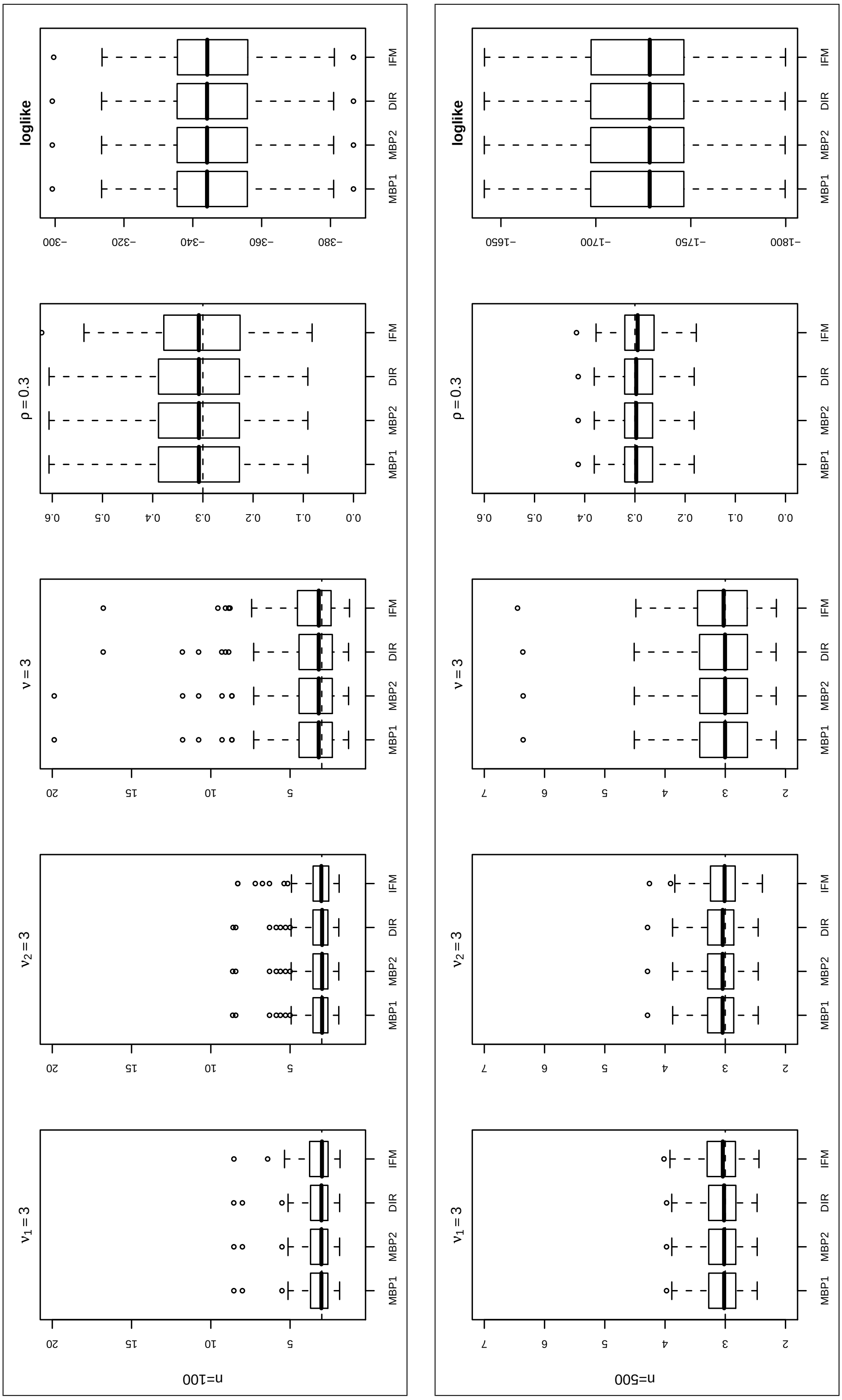
In general for the bivariate t and meta t-distributions the computing effort of the MbP algorithm is very intensive. However, based on the ranking presented in Chapter 6, the MbP algorithm is the most robust estimation method for small sample size in comparison to the direct maximization and IFM methods. With this advantage we can adopt the MbP algorithm in many applications such as the rolling window analysis where small samples are needed.

Appendix A

The Summary Tables of Simulation Studies in Chapter 6

$\nu_{1,tr}=3,\ \nu_{2,tr}=3,\ \nu_{tr}=3,\ \rho_{tr}=0.3$																	
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	3.285	0.109	0.799	0.095	0.036	1.226	0.413	1.000	1.656	1.000	−344.140	96	100	44.544	11.146	22.347
		MBP2	3.285	0.109	0.799	0.095	0.036	1.226	0.413	1.000	1.656	1.000	−344.140	96	100	44.149	11.156	22.452
		DIR	3.285	0.109	0.797	0.095	0.036	1.226	0.414	1.001	1.657	1.000	−344.140	96	100	1.024	1.000	20.052
		IFM	3.276	0.104	0.856	0.092	0.035	1.113	0.347	0.908	1.181	0.714	−344.205	96	96	3.374	1.000	61.109
	ν_2	MBP1	3.297	0.115	0.685	0.099	0.038	1.360	0.467	1.000	1.077	1.000	−344.140	96	100	44.544	11.146	22.347
		MBP2	3.297	0.115	0.685	0.099	0.038	1.360	0.467	1.000	1.078	1.000	−344.140	96	100	44.149	11.156	22.452
		DIR	3.296	0.115	0.684	0.099	0.038	1.357	0.465	0.998	1.078	1.001	−344.140	96	100	1.024	1.000	20.052
		IFM	3.294	0.122	0.722	0.098	0.041	1.505	0.476	1.107	1.123	1.042	−344.205	96	96	3.374	1.000	61.109
	ν	MBP1	3.862	0.264	1.533	0.287	0.088	7.427	3.135	1.000	4.572	1.000	−344.140	96	100	44.544	11.146	22.347
		MBP2	3.862	0.264	1.533	0.287	0.088	7.427	3.135	1.000	4.572	1.000	−344.140	96	100	44.149	11.156	22.452
		DIR	3.837	0.246	1.533	0.279	0.082	6.515	2.267	0.877	4.574	1.000	−344.140	96	100	1.024	1.000	20.052
		IFM	3.810	0.232	1.524	0.270	0.077	5.836	2.111	0.786	5.258	1.150	−344.205	96	96	3.374	1.000	61.109
	ρ	MBP1	0.308	0.011	0.117	0.027	0.036	0.011	0.001	1.000	0.031	1.000	−344.140	96	100	44.544	11.146	22.347
		MBP2	0.308	0.011	0.117	0.027	0.036	0.011	0.001	1.000	0.031	1.000	−344.140	96	100	44.149	11.156	22.452
		DIR	0.308	0.011	0.117	0.027	0.036	0.011	0.001	1.000	0.031	1.001	−344.140	96	100	1.024	1.000	20.052
		IFM	0.306	0.011	0.112	0.021	0.036	0.011	0.002	1.030	0.026	0.830	−344.205	96	96	3.374	1.000	61.109
500	ν_1	MBP1	3.083	0.032	0.330	0.028	0.011	0.109	0.017	1.000	0.214	1.000	−1728.356	100	100	225.411	9.460	35.247
		MBP2	3.083	0.032	0.330	0.028	0.011	0.109	0.017	1.000	0.214	1.000	−1728.356	100	100	224.516	9.460	34.605
		DIR	3.083	0.032	0.332	0.028	0.011	0.109	0.017	1.000	0.211	0.985	−1728.356	100	100	4.015	1.000	16.780
		IFM	3.083	0.033	0.342	0.028	0.011	0.115	0.018	1.052	0.260	1.213	−1728.394	100	100	17.745	1.000	66.605
	ν_2	MBP1	3.081	0.033	0.320	0.027	0.011	0.119	0.022	1.000	0.180	1.000	−1728.356	100	100	225.411	9.460	35.247
		MBP2	3.081	0.033	0.320	0.027	0.011	0.119	0.022	1.000	0.180	1.000	−1728.356	100	100	224.516	9.460	34.605
		DIR	3.081	0.033	0.321	0.027	0.011	0.119	0.022	1.000	0.180	0.997	−1728.356	100	100	4.015	1.000	16.780
		IFM	3.079	0.034	0.297	0.026	0.011	0.122	0.021	1.025	0.184	1.021	−1728.394	100	100	17.745	1.000	66.605
	ν	MBP1	3.081	0.061	0.583	0.027	0.020	0.375	0.115	1.000	0.724	1.000	−1728.356	100	100	225.411	9.460	35.247
		MBP2	3.081	0.061	0.583	0.027	0.020	0.375	0.115	1.000	0.724	1.000	−1728.356	100	100	224.516	9.460	34.605
		DIR	3.081	0.061	0.586	0.027	0.020	0.375	0.116	1.001	0.726	1.002	−1728.356	100	100	4.015	1.000	16.780
		IFM	3.090	0.062	0.609	0.030	0.021	0.395	0.122	1.054	0.837	1.156	−1728.394	100	100	17.745	1.000	66.605
	ρ	MBP1	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.000	−1728.356	100	100	225.411	9.460	35.247
		MBP2	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.000	−1728.356	100	100	224.516	9.460	34.605
		DIR	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.001	−1728.356	100	100	4.015	1.000	16.780
		IFM	0.292	0.005	0.042	−0.027	0.016	0.002	0.000	1.022	0.004	0.976	−1728.394	100	100	17.745	1.000	66.605

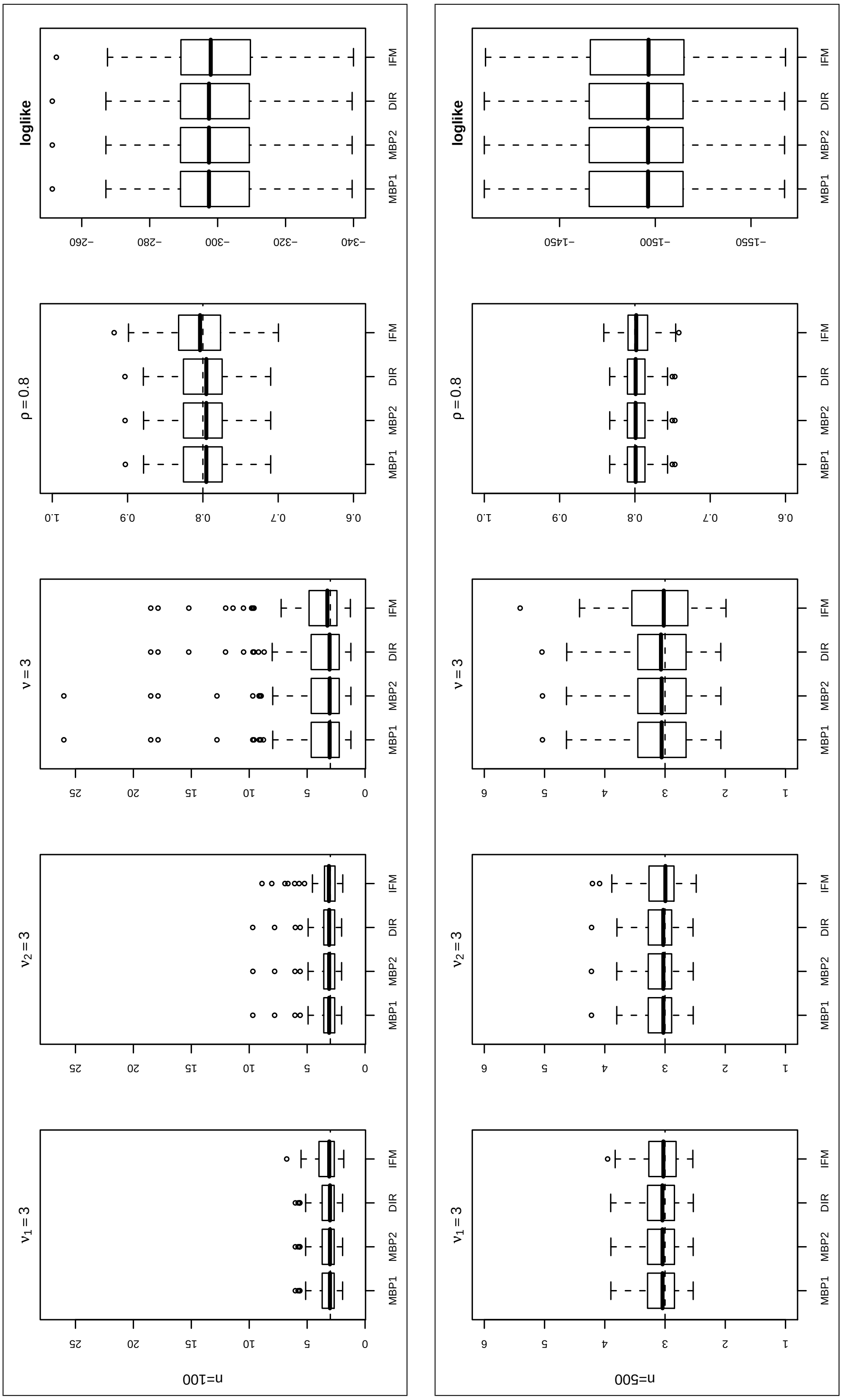
Table A.1: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
ν1 = 3
ν2 = 3
ν = 3
ρ = 0.3
loglike
MBP1
MBP2
DIR
IFM

$\nu_{1,tr} = 3, \quad \nu_{2,tr} = 3, \quad \nu_{tr} = 3, \quad \rho_{tr} = 0.8$

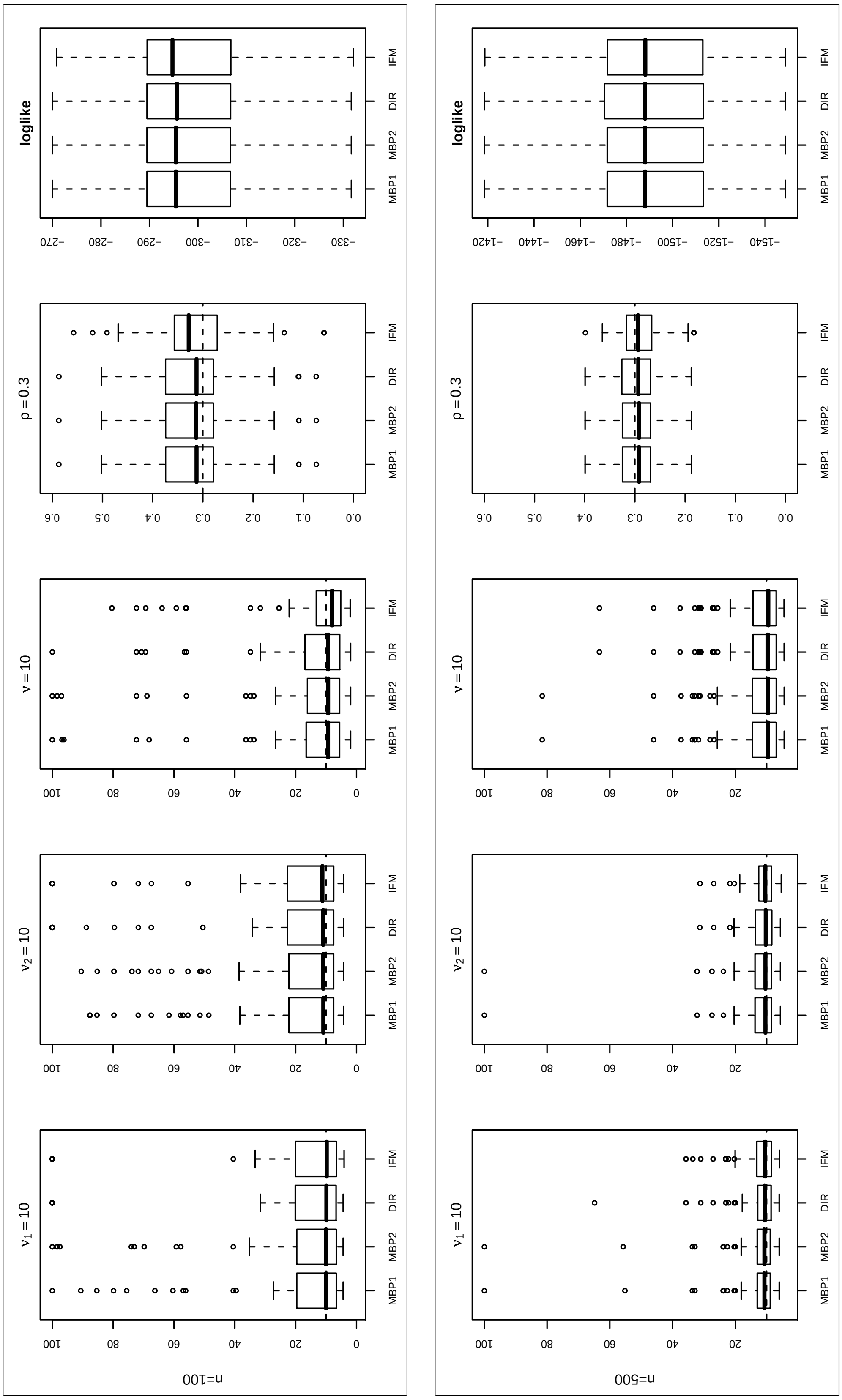
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	loglike	r_c	r_g	time	avg.it	func.eval
100	ν_1	MBP1	3.283	0.093	0.764	0.094	0.031	0.925	0.186	1.000	1.426	1.000	−297.443	98	100	72.844	14.469	28.653
		MBP2	3.283	0.093	0.764	0.094	0.031	0.925	0.186	1.000	1.428	1.001	−297.443	98	100	71.546	14.469	28.529
		DIR	3.282	0.093	0.765	0.094	0.031	0.923	0.186	0.997	1.434	1.006	−297.443	98	100	1.426	1.000	27.837
		IFM	3.303	0.097	0.974	0.101	0.032	1.007	0.191	1.088	1.414	0.991	−297.964	98	98	3.337	1.000	60.520
	ν_2	MBP1	3.291	0.110	0.684	0.097	0.037	1.265	0.518	1.000	1.068	1.000	−297.443	98	100	72.844	14.469	28.653
		MBP2	3.291	0.110	0.684	0.097	0.037	1.265	0.517	0.999	1.069	1.001	−297.443	98	100	71.546	14.469	28.529
		DIR	3.291	0.110	0.680	0.097	0.037	1.265	0.519	0.999	1.065	0.997	−297.443	98	100	1.426	1.000	27.837
		IFM	3.298	0.118	0.665	0.099	0.039	1.464	0.486	1.157	1.094	1.024	−297.964	98	98	3.337	1.000	60.520
	ν	MBP1	4.290	0.378	1.758	0.430	0.126	15.633	6.314	1.000	6.011	1.000	−297.443	98	100	72.844	14.469	28.653
		MBP2	4.290	0.377	1.758	0.430	0.126	15.604	6.312	0.998	6.009	1.000	−297.443	98	100	71.546	14.469	28.529
		DIR	4.195	0.326	1.771	0.398	0.109	11.826	3.747	0.757	5.955	0.991	−297.443	98	100	1.426	1.000	27.837
		IFM	4.309	0.325	1.757	0.436	0.108	12.047	3.788	0.771	6.792	1.130	−297.964	98	98	3.337	1.000	60.520
	ρ	MBP1	0.797	0.004	0.038	−0.003	0.005	0.002	0.000	1.000	0.003	1.000	−297.443	98	100	72.844	14.469	28.653
		MBP2	0.797	0.004	0.038	−0.003	0.005	0.002	0.000	1.000	0.003	1.000	−297.443	98	100	71.546	14.469	28.529
		DIR	0.797	0.004	0.038	−0.003	0.005	0.002	0.000	1.000	0.003	1.000	−297.443	98	100	1.426	1.000	27.837
		IFM	0.802	0.004	0.041	0.002	0.005	0.002	0.000	1.042	0.004	1.208	−297.964	98	98	3.337	1.000	60.520
500	ν_1	MBP1	3.080	0.028	0.330	0.027	0.009	0.087	0.013	1.000	0.246	1.000	−1496.223	100	100	356.709	12.070	52.303
		MBP2	3.080	0.028	0.330	0.027	0.009	0.087	0.013	1.000	0.246	1.000	−1496.223	100	100	360.506	12.040	53.351
		DIR	3.080	0.028	0.328	0.027	0.009	0.087	0.013	1.001	0.245	0.998	−1496.223	100	100	5.617	1.000	23.400
		IFM	3.070	0.031	0.328	0.023	0.010	0.102	0.015	1.165	0.236	0.960	−1496.485	100	100	17.611	1.000	66.060
	ν_2	MBP1	3.088	0.032	0.285	0.029	0.011	0.108	0.020	1.000	0.129	1.000	−1496.223	100	100	356.709	12.070	52.303
		MBP2	3.088	0.032	0.285	0.029	0.011	0.108	0.020	1.000	0.129	1.000	−1496.223	100	100	360.506	12.040	53.351
		DIR	3.089	0.032	0.285	0.030	0.011	0.108	0.020	1.001	0.128	0.995	−1496.223	100	100	5.617	1.000	23.400
		IFM	3.077	0.034	0.299	0.026	0.011	0.123	0.022	1.143	0.215	1.663	−1496.485	100	100	17.611	1.000	66.060
	ν	MBP1	3.137	0.061	0.584	0.046	0.020	0.391	0.063	1.000	0.760	1.000	−1496.223	100	100	356.709	12.070	52.303
		MBP2	3.137	0.061	0.584	0.046	0.020	0.391	0.063	1.000	0.760	1.000	−1496.223	100	100	360.506	12.040	53.351
		DIR	3.138	0.061	0.588	0.046	0.020	0.392	0.063	1.002	0.754	0.991	−1496.223	100	100	5.617	1.000	23.400
		IFM	3.109	0.062	0.684	0.036	0.021	0.393	0.069	1.005	0.903	1.189	−1496.485	100	100	17.611	1.000	66.060
	ρ	MBP1	0.797	0.002	0.017	−0.003	0.002	0.000	0.000	1.000	0.001	1.000	−1496.223	100	100	356.709	12.070	52.303
		MBP2	0.797	0.002	0.017	−0.003	0.002	0.000	0.000	1.000	0.001	1.001	−1496.223	100	100	360.506	12.040	53.351
		DIR	0.797	0.002	0.017	−0.003	0.002	0.000	0.000	1.000	0.001	1.005	−1496.223	100	100	5.617	1.000	23.400
		IFM	0.796	0.002	0.019	−0.005	0.002	0.000	0.000	1.241	0.001	1.307	−1496.485	100	100	17.611	1.000	66.060

Table A.2: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.8
ν = 3
ν2 = 3
ν1 = 3
MBP1
MBP2
DIR
IFM

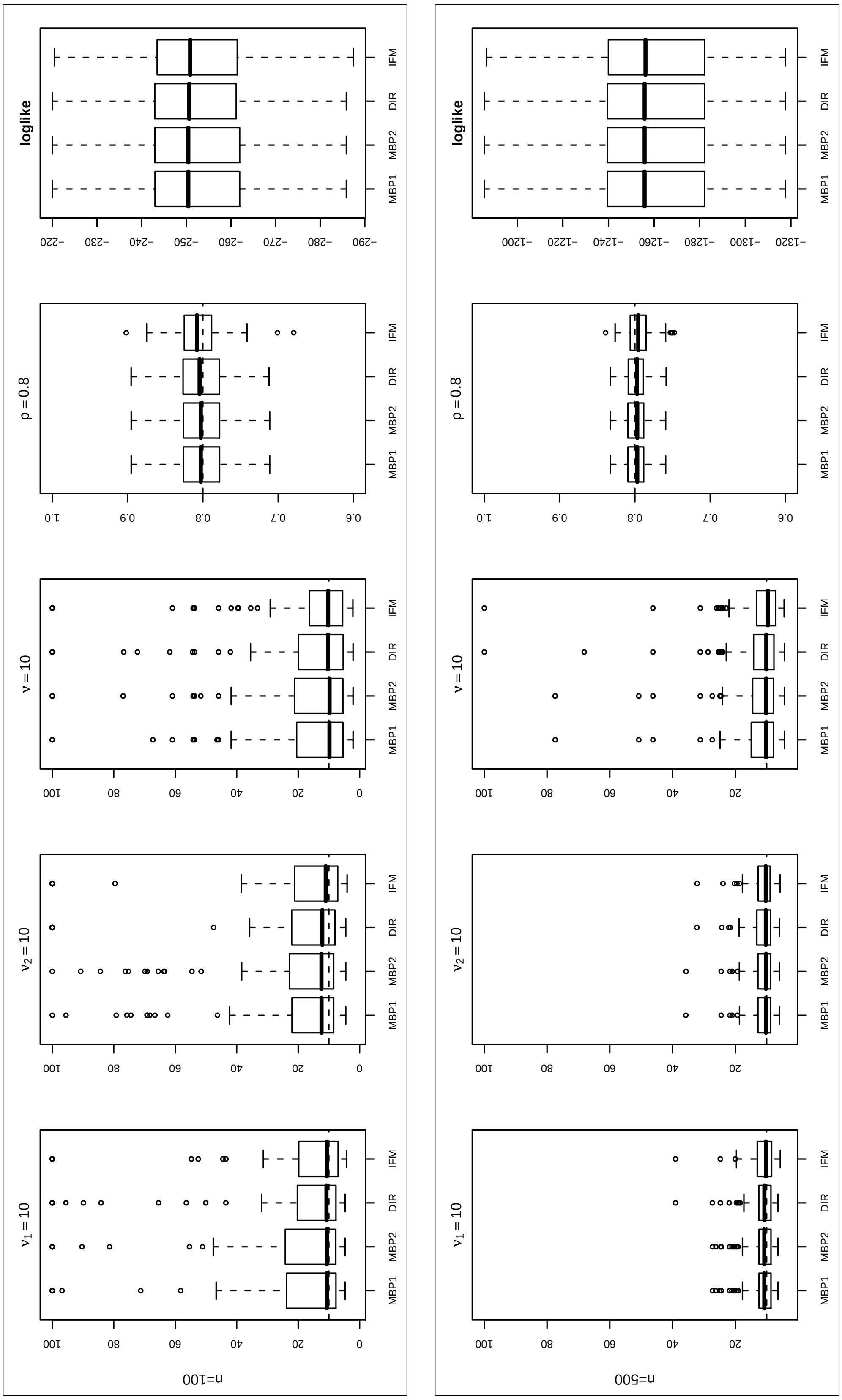
$\nu_{1,tr}=10,\ \nu_{2,tr}=10,\ \nu_{tr}=10,\ \rho_{tr}=0.3$																		
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	19.161	2.524	9.612	0.916	0.252	574.459	176.845	1.000	76.984	1.000	−296.095	77	100	35.130	11.403	18.846
		MBP2	19.869	2.759	9.612	0.987	0.276	683.521	214.937	1.190	78.185	1.016	−296.096	77	100	36.350	11.481	18.818
		DIR	22.891	3.490	9.946	1.289	0.349	1104.041	308.682	1.922	79.904	1.038	−296.491	77	100	1.683	1.000	32.130
		IFM	22.978	3.503	9.963	1.298	0.350	1113.234	308.495	1.938	69.349	0.901	−294.918	77	60	2.254	1.000	38.565
	ν_2	MBP1	21.098	2.505	10.913	1.110	0.250	606.160	161.445	1.000	92.362	1.000	−296.095	77	100	35.130	11.403	18.846
		MBP2	20.955	2.465	10.912	1.095	0.246	587.869	155.518	0.970	92.975	1.007	−296.096	77	100	36.350	11.481	18.818
		DIR	25.326	3.466	11.231	1.533	0.347	1159.939	294.673	1.914	94.514	1.023	−296.491	77	100	1.683	1.000	32.130
		IFM	23.689	3.235	11.272	1.369	0.323	993.056	272.479	1.638	113.274	1.226	−294.918	77	60	2.254	1.000	38.565
	ν	MBP1	17.370	2.636	8.153	0.737	0.264	589.255	205.164	1.000	100.791	1.000	−296.095	77	100	35.130	11.403	18.846
		MBP2	17.404	2.651	7.845	0.740	0.265	595.798	207.768	1.011	99.883	0.991	−296.096	77	100	36.350	11.481	18.818
		DIR	15.230	2.065	8.471	0.523	0.206	355.649	134.445	0.604	105.040	1.042	−296.491	77	100	1.683	1.000	32.130
		IFM	14.781	1.991	5.964	0.478	0.199	327.996	107.112	0.557	67.870	0.673	−294.918	77	60	2.254	1.000	38.565
	ρ	MBP1	0.315	0.011	0.071	0.050	0.036	0.009	0.002	1.000	0.014	1.000	−296.095	77	100	35.130	11.403	18.846
		MBP2	0.315	0.011	0.070	0.050	0.036	0.009	0.002	0.999	0.015	1.025	−296.096	77	100	36.350	11.481	18.818
		DIR	0.314	0.011	0.071	0.048	0.035	0.009	0.002	0.994	0.014	0.968	−296.491	77	100	1.683	1.000	32.130
		IFM	0.314	0.011	0.063	0.048	0.035	0.009	0.002	0.967	0.010	0.680	−294.918	77	60	2.254	1.000	38.565
500	ν_1	MBP1	13.118	1.107	3.083	0.312	0.111	130.936	83.829	1.000	23.975	1.000	−1488.801	99	100	132.746	10.222	15.213
		MBP2	13.124	1.109	3.083	0.312	0.111	131.475	83.953	1.004	23.976	1.000	−1488.801	99	100	133.643	10.212	15.186
		DIR	12.272	0.735	3.147	0.227	0.074	58.697	31.213	0.448	21.315	0.889	−1488.801	99	100	4.578	1.000	17.960
		IFM	11.957	0.557	3.429	0.196	0.056	34.518	10.281	0.264	21.600	0.901	−1488.891	99	99	15.002	1.000	53.348
	ν_2	MBP1	12.362	0.991	3.842	0.236	0.099	102.888	81.433	1.000	25.690	1.000	−1488.801	99	100	132.746	10.222	15.213
		MBP2	12.362	0.991	3.842	0.236	0.099	102.887	81.433	1.000	25.695	1.000	−1488.801	99	100	133.643	10.212	15.186
		DIR	11.395	0.418	3.895	0.139	0.042	19.277	5.646	0.187	24.338	0.947	−1488.801	99	100	4.578	1.000	17.960
		IFM	11.321	0.421	3.025	0.132	0.042	19.255	5.624	0.187	20.483	0.797	−1488.891	99	99	15.002	1.000	53.348
	ν	MBP1	12.947	1.056	5.651	0.295	0.106	119.062	53.942	1.000	40.190	1.000	−1488.801	99	100	132.746	10.222	15.213
		MBP2	12.930	1.053	5.651	0.293	0.105	118.339	53.892	0.994	40.160	0.999	−1488.801	99	100	133.643	10.212	15.186
		DIR	12.625	0.929	5.489	0.263	0.093	92.407	33.087	0.776	40.586	1.010	−1488.801	99	100	4.578	1.000	17.960
		IFM	12.566	0.932	5.480	0.257	0.093	92.532	33.078	0.777	41.062	1.022	−1488.891	99	99	15.002	1.000	53.348
	ρ	MBP1	0.291	0.005	0.041	−0.028	0.015	0.002	0.000	1.000	0.003	1.000	−1488.801	99	100	132.746	10.222	15.213
		MBP2	0.291	0.005	0.041	−0.028	0.015	0.002	0.000	1.000	0.003	1.000	−1488.801	99	100	133.643	10.212	15.186
		DIR	0.292	0.005	0.042	−0.027	0.015	0.002	0.000	1.011	0.004	1.433	−1488.801	99	100	4.578	1.000	17.960
		IFM	0.290	0.005	0.038	−0.035	0.015	0.002	0.000	1.026	0.003	0.996	−1488.891	99	99	15.002	1.000	53.348

Table A.3: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.3
ν = 10
ν2 = 10
ν1 = 10
MBP1
MBP2
DIR
IFM

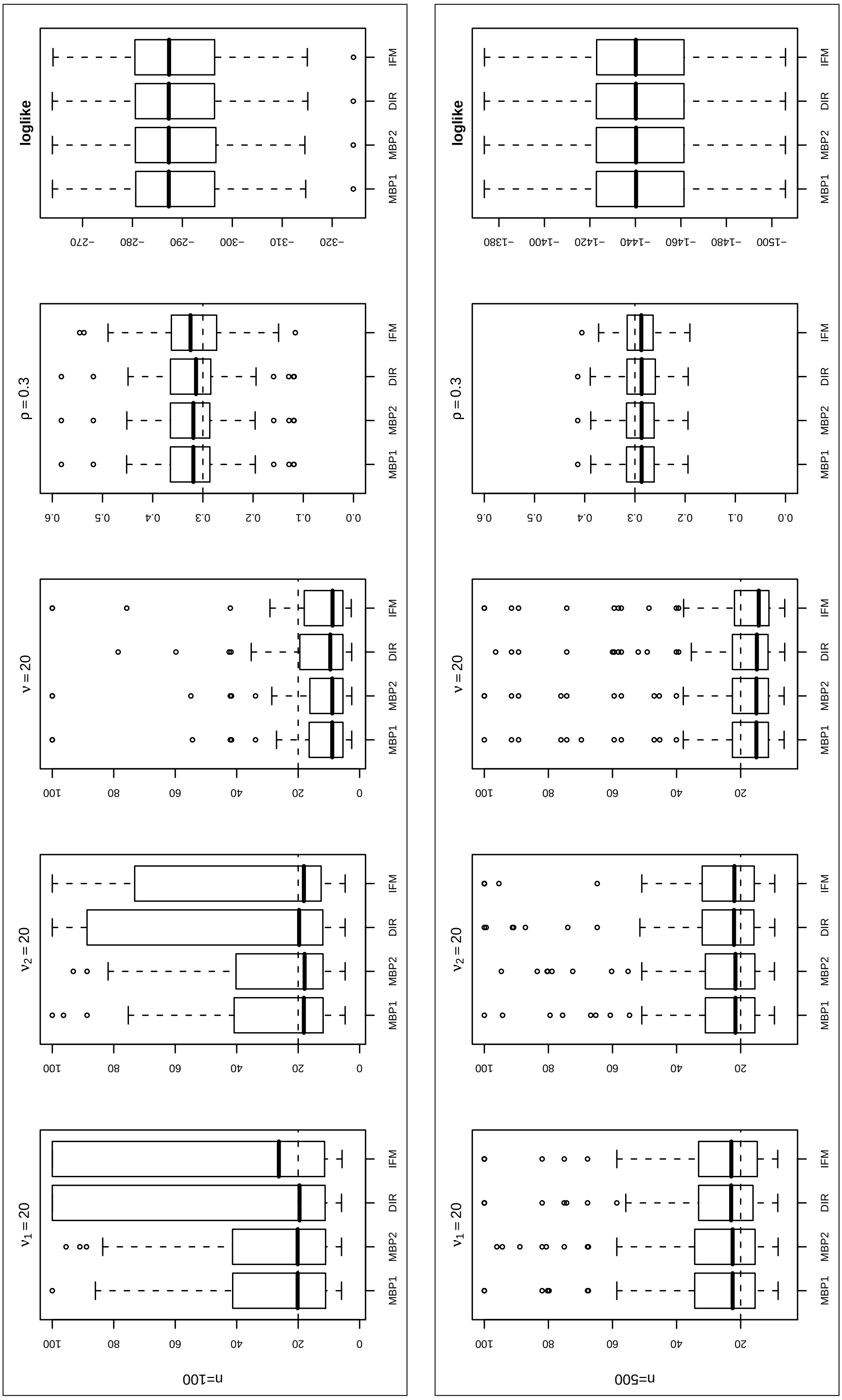
$\nu_{1,tr}=10$, $\nu_{2,tr}=10$, $\nu_{tr}=10$, $\rho_{tr}=0.8$																	
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	22.232	2.886	11.265	1.223	0.289	816.005	237.640	1.000	104.746	1.000	−250.357	80	100	44.453	13.650	20.322
		MBP2	22.172	2.880	12.011	1.217	0.288	811.754	236.279	0.995	104.732	1.000	−250.357	80	100	46.580	13.750	19.611
		DIR	23.253	3.195	9.220	1.325	0.319	992.058	267.968	1.216	89.336	0.853	−250.456	80	100	2.180	1.000	41.763
		IFM	22.840	3.288	9.401	1.284	0.329	1029.690	284.453	1.262	105.939	1.011	−249.295	80	58	2.032	1.000	35.725
	ν_2	MBP1	22.043	2.543	9.764	1.204	0.254	662.389	177.938	1.000	118.531	1.000	−250.357	80	100	44.453	13.650	20.322
		MBP2	22.302	2.566	10.663	1.230	0.257	678.246	177.050	1.024	126.678	1.069	−250.357	80	100	46.580	13.750	19.611
		DIR	26.239	3.584	10.201	1.624	0.358	1291.317	321.051	1.949	137.242	1.158	−250.456	80	100	2.180	1.000	41.763
		IFM	25.000	3.557	10.190	1.500	0.356	1237.473	312.942	1.868	107.582	0.908	−249.295	80	58	2.032	1.000	35.725
	ν	MBP1	17.451	2.190	11.074	0.745	0.219	439.306	151.668	1.000	151.356	1.000	−250.357	80	100	44.453	13.650	20.322
		MBP2	17.694	2.241	11.643	0.769	0.224	460.782	156.290	1.049	150.120	0.992	−250.357	80	100	46.580	13.750	19.611
		DIR	19.694	2.711	10.712	0.969	0.271	682.061	209.012	1.553	165.255	1.092	−250.456	80	100	2.180	1.000	41.763
		IFM	19.113	2.753	7.916	0.911	0.275	689.494	220.187	1.570	125.048	0.826	−249.295	80	58	2.032	1.000	35.725
	ρ	MBP1	0.801	0.004	0.035	0.001	0.005	0.001	0.000	1.000	0.003	1.000	−250.357	80	100	44.453	13.650	20.322
		MBP2	0.801	0.004	0.035	0.001	0.005	0.001	0.000	0.999	0.003	1.003	−250.357	80	100	46.580	13.750	19.611
		DIR	0.801	0.004	0.035	0.001	0.005	0.001	0.000	1.026	0.003	1.037	−250.456	80	100	2.180	1.000	41.763
		IFM	0.806	0.004	0.027	0.007	0.005	0.001	0.000	1.285	0.002	0.554	−249.295	80	58	2.032	1.000	35.725
500	ν_1	MBP1	11.617	0.443	2.790	0.162	0.044	21.884	5.485	1.000	17.865	1.000	−1255.847	98	100	209.674	15.327	14.814
		MBP2	11.612	0.442	2.791	0.161	0.044	21.763	5.436	0.994	17.510	0.980	−1255.847	98	100	206.920	15.204	14.640
		DIR	11.693	0.492	2.812	0.169	0.049	26.573	9.626	1.214	19.213	1.075	−1255.847	98	100	7.434	1.000	28.929
		IFM	11.548	0.475	3.394	0.155	0.047	24.502	8.992	1.120	23.367	1.308	−1256.230	98	97	14.936	1.000	52.806
	ν_2	MBP1	11.456	0.428	2.953	0.146	0.043	20.095	7.290	1.000	14.315	1.000	−1255.847	98	100	209.674	15.327	14.814
		MBP2	11.454	0.428	2.953	0.145	0.043	20.056	7.258	0.998	14.446	1.009	−1255.847	98	100	206.920	15.204	14.640
		DIR	11.424	0.405	3.129	0.142	0.040	18.086	5.765	0.900	14.901	1.041	−1255.847	98	100	7.434	1.000	28.929
		IFM	11.228	0.398	2.818	0.123	0.040	17.013	5.534	0.847	15.502	1.083	−1256.230	98	97	14.936	1.000	52.806
	ν	MBP1	13.069	1.021	4.961	0.307	0.102	111.584	50.554	1.000	36.243	1.000	−1255.847	98	100	209.674	15.327	14.814
		MBP2	13.052	1.021	4.708	0.305	0.102	111.471	50.584	0.999	36.548	1.008	−1255.847	98	100	206.920	15.204	14.640
		DIR	13.484	1.256	4.629	0.348	0.126	166.838	89.402	1.495	32.468	0.896	−1255.847	98	100	7.434	1.000	28.929
		IFM	12.597	1.129	4.495	0.260	0.113	131.575	83.074	1.179	37.728	1.041	−1256.230	98	97	14.936	1.000	52.806
	ρ	MBP1	0.798	0.002	0.015	−0.003	0.002	0.000	0.000	1.000	0.001	1.000	−1255.847	98	100	209.674	15.327	14.814
		MBP2	0.798	0.002	0.015	−0.003	0.002	0.000	0.000	0.999	0.001	1.000	−1255.847	98	100	206.920	15.204	14.640
		DIR	0.798	0.002	0.015	−0.003	0.002	0.000	0.000	1.002	0.001	0.980	−1255.847	98	100	7.434	1.000	28.929
		IFM	0.795	0.002	0.015	−0.007	0.002	0.000	0.000	1.354	0.001	1.066	−1256.230	98	97	14.936	1.000	52.806

Table A.4: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.8
ν = 10
ν2 = 10
ν1 = 10
MBP1
MBP2
DIR
IFM

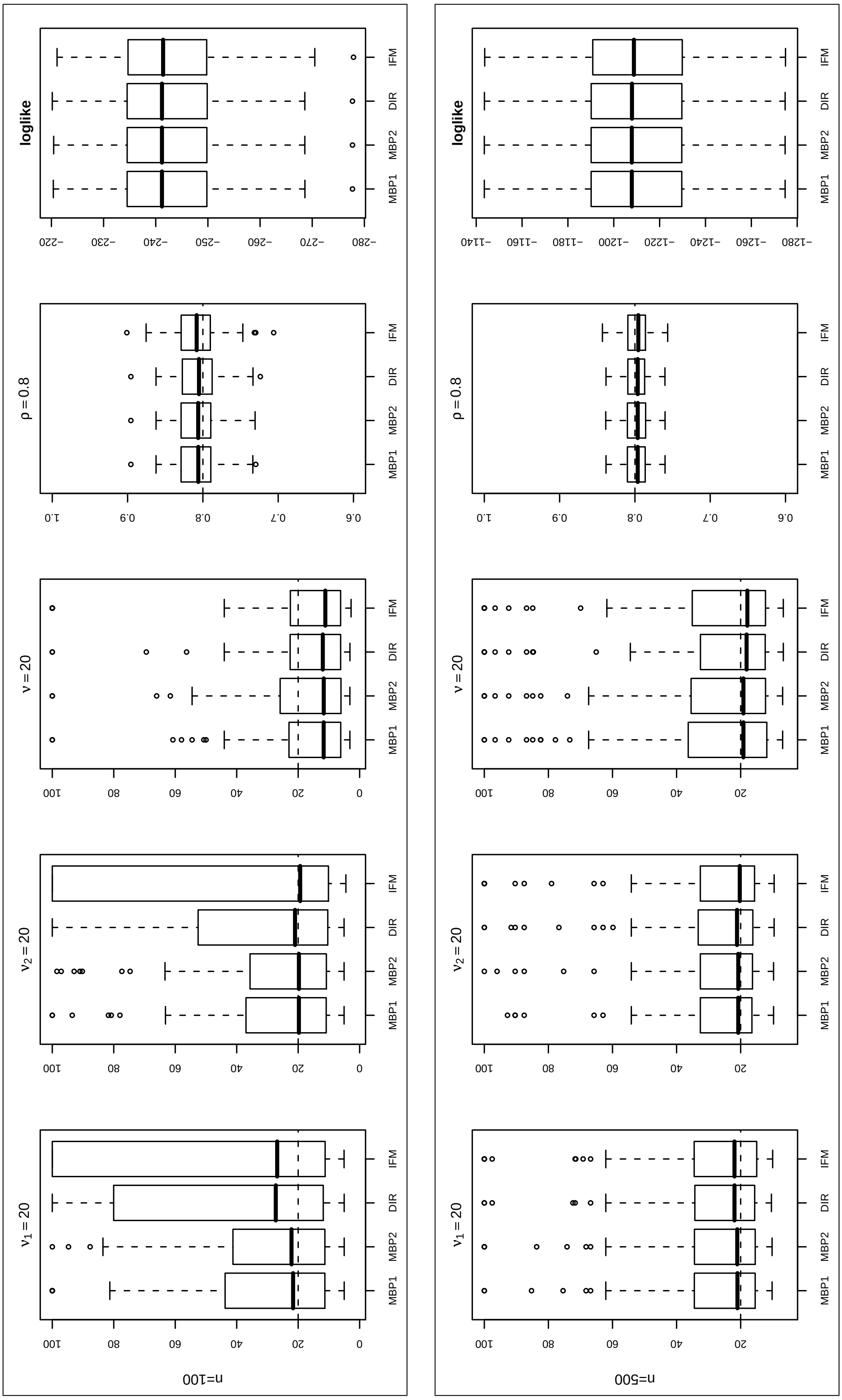
$\nu_{1,tr} = 20,\ \nu_{2,tr} = 20,\ \nu_{tr} = 20,\ \rho_{tr} = 0.3$																		
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	31.487	3.430	22.386	0.574	0.172	755.516	184.469	1.000	705.795	1.000	−288.297	53	100	37.680	11.792	21.152
		MBP2	31.980	3.503	22.434	0.599	0.175	793.864	192.517	1.051	726.130	1.029	−288.297	53	100	39.048	11.245	21.270
		DIR	46.864	5.360	65.756	1.343	0.268	2244.362	379.306	2.971	900.323	1.276	−288.297	53	100	1.907	1.000	37.189
		IFM	49.573	5.680	65.630	1.479	0.284	2584.618	413.954	3.421	1796.546	2.545	−288.310	53	26	1.412	1.000	23.509
	ν_2	MBP1	29.099	3.443	21.439	0.455	0.172	711.233	197.393	1.000	405.863	1.000	−288.297	53	100	37.680	11.792	21.152
		MBP2	28.621	3.243	20.981	0.431	0.162	631.573	165.606	0.888	395.357	0.974	−288.297	53	100	39.048	11.245	21.270
		DIR	43.798	5.265	56.832	1.190	0.263	2035.327	360.768	2.862	724.821	1.786	−288.297	53	100	1.907	1.000	37.189
		IFM	39.997	5.005	44.942	1.000	0.250	1727.543	352.291	2.429	501.751	1.236	−288.310	53	26	1.412	1.000	23.509
	ν	MBP1	15.852	2.747	8.139	−0.207	0.137	417.246	164.746	1.000	212.278	1.000	−288.297	53	100	37.680	11.792	21.152
		MBP2	15.907	2.755	7.972	−0.205	0.138	418.903	164.767	1.004	212.724	1.002	−288.297	53	100	39.048	11.245	21.270
		DIR	14.504	2.007	10.393	−0.275	0.100	243.629	68.144	0.584	209.781	0.988	−288.297	53	100	1.907	1.000	37.189
		IFM	15.723	2.824	9.318	−0.214	0.141	440.816	172.105	1.056	197.706	0.931	−288.310	53	26	1.412	1.000	23.509
	ρ	MBP1	0.321	0.012	0.058	0.069	0.041	0.008	0.002	1.000	0.009	1.000	−288.297	53	100	37.680	11.792	21.152
		MBP2	0.321	0.012	0.058	0.069	0.041	0.008	0.002	0.998	0.009	1.000	−288.297	53	100	39.048	11.245	21.270
		DIR	0.318	0.012	0.060	0.061	0.040	0.008	0.002	0.981	0.008	0.861	−288.297	53	100	1.907	1.000	37.189
		IFM	0.323	0.012	0.067	0.076	0.040	0.008	0.002	0.989	0.010	1.119	−288.310	53	26	1.412	1.000	23.509
500	ν_1	MBP1	29.432	2.395	13.960	0.472	0.120	553.599	154.090	1.000	299.378	1.000	−1440.899	81	100	112.202	7.864	16.218
		MBP2	29.141	2.291	13.961	0.457	0.115	508.807	135.326	0.919	299.574	1.001	−1440.899	81	99	135.438	7.877	16.952
		DIR	30.168	2.491	12.578	0.508	0.125	605.887	165.929	1.094	362.698	1.212	−1440.699	81	100	3.820	1.000	14.914
		IFM	29.584	2.442	13.564	0.479	0.122	574.805	163.308	1.038	373.519	1.248	−1440.650	81	70	11.762	1.000	40.302
	ν_2	MBP1	27.977	2.087	11.458	0.399	0.104	416.260	121.422	1.000	219.019	1.000	−1440.899	81	100	112.202	7.864	16.218
		MBP2	28.085	2.093	11.458	0.404	0.105	420.350	115.409	1.010	219.002	1.000	−1440.899	81	99	135.438	7.877	16.952
		DIR	28.982	2.305	11.961	0.449	0.115	510.976	150.402	1.228	263.923	1.205	−1440.699	81	100	3.820	1.000	14.914
		IFM	29.305	2.509	12.033	0.465	0.125	596.562	181.707	1.433	262.212	1.197	−1440.650	81	70	11.762	1.000	40.302
	ν	MBP1	22.941	2.272	8.320	0.147	0.114	426.653	127.438	1.000	144.304	1.000	−1440.899	81	100	112.202	7.864	16.218
		MBP2	23.316	2.396	8.320	0.166	0.120	475.884	144.979	1.115	144.288	1.000	−1440.899	81	99	135.438	7.877	16.952
		DIR	22.554	2.186	8.234	0.128	0.109	393.676	118.054	0.923	134.330	0.931	−1440.699	81	100	3.820	1.000	14.914
		IFM	22.719	2.348	7.973	0.136	0.117	453.958	141.628	1.064	143.889	0.997	−1440.650	81	70	11.762	1.000	40.302
	ρ	MBP1	0.288	0.005	0.041	−0.040	0.016	0.002	0.000	1.000	0.004	1.000	−1440.899	81	100	112.202	7.864	16.218
		MBP2	0.288	0.005	0.041	−0.040	0.016	0.002	0.000	1.000	0.004	0.999	−1440.899	81	99	135.438	7.877	16.952
		DIR	0.288	0.005	0.042	−0.041	0.016	0.002	0.000	1.005	0.004	0.952	−1440.699	81	100	3.820	1.000	14.914
		IFM	0.285	0.005	0.038	−0.051	0.016	0.002	0.000	1.075	0.004	0.969	−1440.650	81	70	11.762	1.000	40.302

Table A.5: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.3
ν = 20
ν2 = 20
ν1 = 20
MBP1
MBP2
DIR
IFM

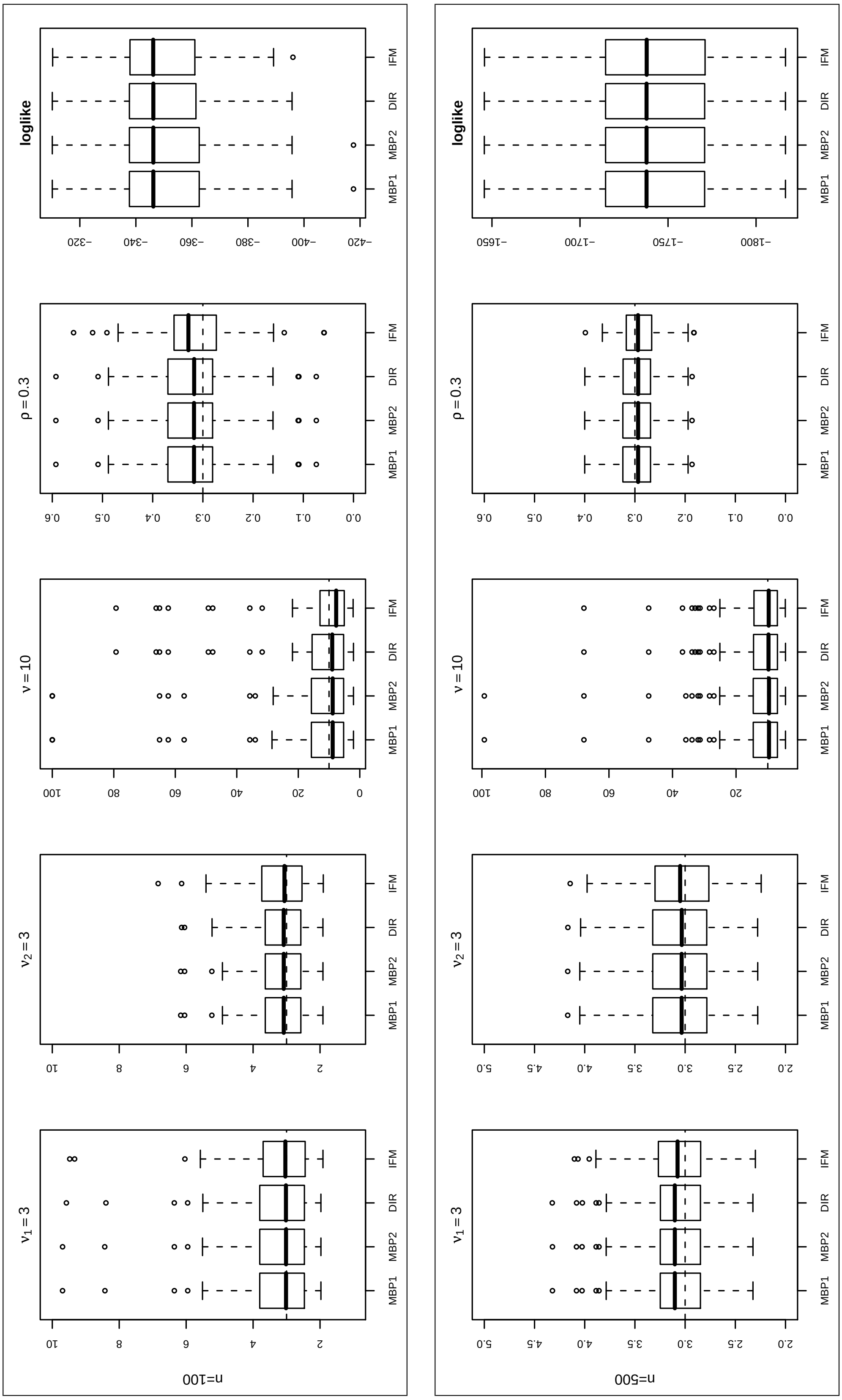
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
			$\nu_{1,tr}=20,\ \nu_{2,tr}=20,\ \nu_{tr}=20,\ \rho_{tr}=0.8$															
100	ν_1	MBP1	32.767	3.624	23.737	0.638	0.181	951.186	236.481	1.000	641.166	1.000	−241.807	60	100	44.886	13.800	21.107
		MBP2	31.592	3.326	21.375	0.580	0.166	798.009	187.906	0.839	665.188	1.037	−241.809	60	100	45.307	13.117	21.006
		DIR	42.672	4.648	48.572	1.134	0.232	1810.432	326.093	1.903	1580.720	2.465	−241.856	60	100	2.165	1.000	42.167
		IFM	46.838	5.095	65.655	1.342	0.255	2277.582	367.119	2.394	1897.063	2.959	−242.235	60	33	1.514	1.000	26.175
	ν_2	MBP1	28.816	3.262	18.800	0.441	0.163	716.275	193.287	1.000	499.858	1.000	−241.807	60	100	44.886	13.800	21.107
		MBP2	29.052	3.356	17.815	0.453	0.168	757.789	202.722	1.058	507.123	1.015	−241.809	60	100	45.307	13.117	21.006
		DIR	37.310	4.496	30.289	0.866	0.225	1512.249	318.126	2.111	704.938	1.410	−241.856	60	100	2.165	1.000	42.167
		IFM	40.242	4.941	66.546	1.012	0.247	1874.681	358.095	2.617	639.260	1.279	−242.235	60	33	1.514	1.000	26.175
	ν	MBP1	19.669	2.739	12.122	−0.017	0.137	450.276	148.972	1.000	303.828	1.000	−241.807	60	100	44.886	13.800	21.107
		MBP2	19.876	2.791	13.917	−0.006	0.140	467.557	150.759	1.038	282.729	0.931	−241.809	60	100	45.307	13.117	21.006
		DIR	19.228	2.642	11.690	−0.039	0.132	419.253	150.343	0.931	281.043	0.925	−241.856	60	100	2.165	1.000	42.167
		IFM	22.142	3.609	11.613	0.107	0.180	786.229	242.543	1.746	228.173	0.751	−242.235	60	33	1.514	1.000	26.175
	ρ	MBP1	0.808	0.004	0.029	0.010	0.005	0.001	0.000	1.000	0.002	1.000	−241.807	60	100	44.886	13.800	21.107
		MBP2	0.808	0.004	0.029	0.010	0.005	0.001	0.000	1.002	0.002	1.050	−241.809	60	100	45.307	13.117	21.006
		DIR	0.806	0.004	0.029	0.008	0.005	0.001	0.000	1.003	0.002	0.983	−241.856	60	100	2.165	1.000	42.167
		IFM	0.808	0.005	0.028	0.010	0.006	0.001	0.000	1.286	0.002	0.868	−242.235	60	33	1.514	1.000	26.175
500	ν_1	MBP1	28.214	1.997	14.068	0.411	0.100	430.417	118.613	1.000	208.099	1.000	−1210.401	91	100	161.226	10.549	16.824
		MBP2	28.369	2.087	14.050	0.418	0.104	466.438	133.366	1.084	209.735	1.008	−1210.401	91	100	250.588	11.429	18.574
		DIR	28.750	2.051	13.835	0.438	0.103	459.395	126.158	1.067	246.849	1.186	−1210.371	91	100	4.945	1.000	19.549
		IFM	29.381	2.241	14.458	0.469	0.112	545.148	142.830	1.267	352.285	1.693	−1210.671	91	79	11.593	1.000	40.247
	ν_2	MBP1	27.522	1.881	11.930	0.376	0.094	378.669	109.102	1.000	172.141	1.000	−1210.401	91	100	161.226	10.549	16.824
		MBP2	27.799	1.972	12.061	0.390	0.099	414.767	122.291	1.095	172.157	1.000	−1210.401	91	100	250.588	11.429	18.574
		DIR	28.830	2.143	12.632	0.442	0.107	495.983	137.479	1.310	197.250	1.146	−1210.371	91	100	4.945	1.000	19.549
		IFM	29.286	2.434	12.539	0.464	0.122	625.514	169.747	1.652	172.155	1.000	−1210.671	91	79	11.593	1.000	40.247
	ν	MBP1	28.718	2.558	18.168	0.436	0.128	671.498	156.866	1.000	317.867	1.000	−1210.401	91	100	161.226	10.549	16.824
		MBP2	28.470	2.558	17.206	0.424	0.128	667.020	163.007	0.993	359.678	1.132	−1210.401	91	100	250.588	11.429	18.574
		DIR	28.438	2.626	14.974	0.422	0.131	698.781	173.090	1.041	295.179	0.929	−1210.371	91	100	4.945	1.000	19.549
		IFM	29.707	2.919	16.932	0.485	0.146	869.394	200.467	1.295	262.407	0.826	−1210.671	91	79	11.593	1.000	40.247
	ρ	MBP1	0.798	0.002	0.018	−0.003	0.002	0.000	0.000	1.000	0.001	1.000	−1210.401	91	100	161.226	10.549	16.824
		MBP2	0.798	0.002	0.018	−0.003	0.002	0.000	0.000	0.994	0.001	1.046	−1210.401	91	100	250.588	11.429	18.574
		DIR	0.797	0.002	0.017	−0.003	0.002	0.000	0.000	1.000	0.001	0.973	−1210.371	91	100	4.945	1.000	19.549
		IFM	0.796	0.002	0.017	−0.005	0.002	0.000	0.000	1.251	0.001	1.106	−1210.671	91	79	11.593	1.000	40.247

Table A.6: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.8
ν = 20
ν_2 = 20
ν_1 = 20
MBP1
MBP2
DIR
IFM

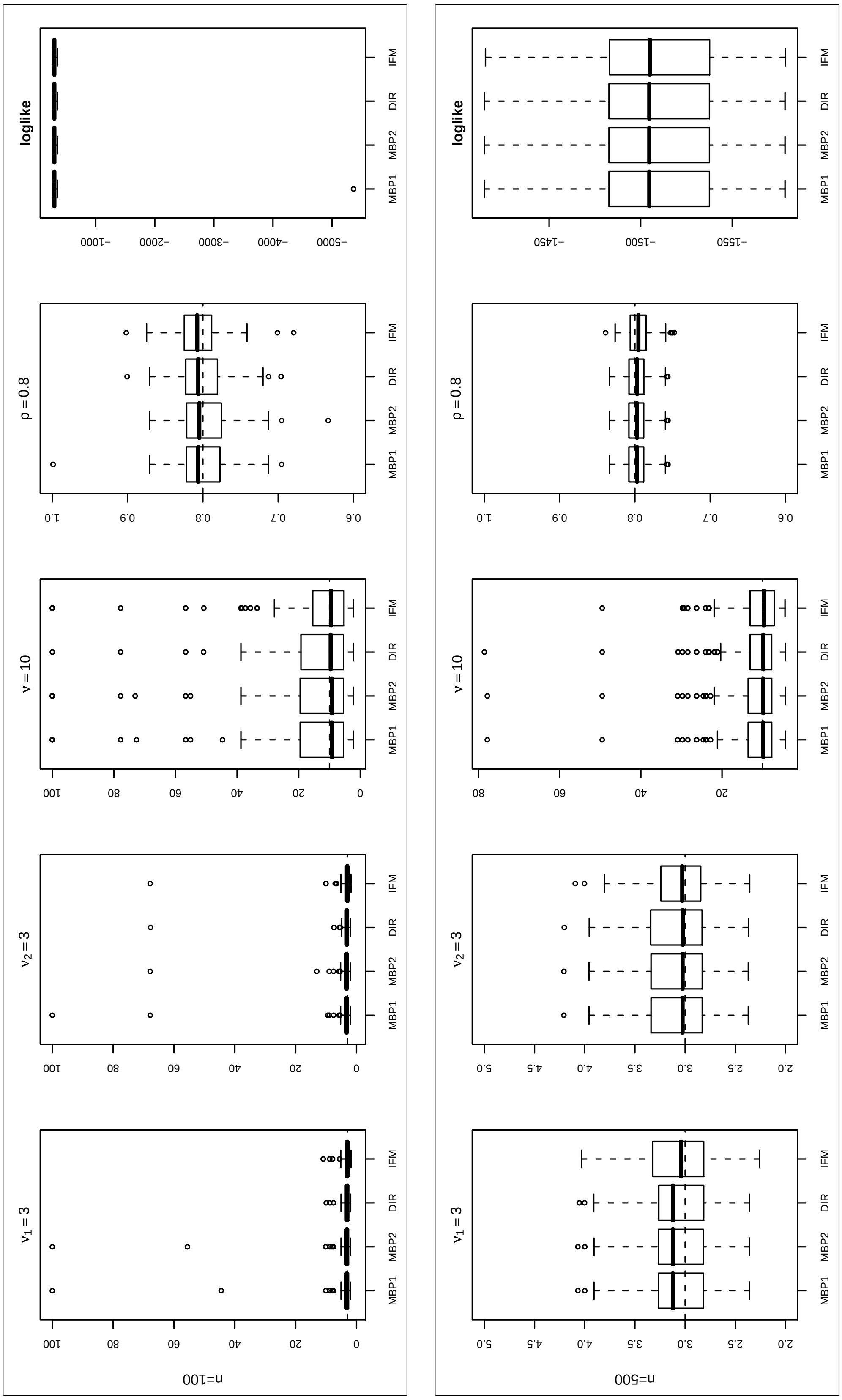
			$\nu_{1,tr}=3,\ \nu_{2,tr}=3,\ \nu_{tr}=10,\ \rho_{tr}=0.3$															
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	3.323	0.153	0.982	0.108	0.051	1.876	0.711	1.000	1.691	1.000	−347.138	76	99	42.481	9.868	22.325
		MBP2	3.324	0.153	0.982	0.108	0.051	1.876	0.711	1.000	1.691	1.000	−347.138	76	99	39.127	9.303	21.511
		DIR	3.321	0.152	0.984	0.032	0.051	1.851	0.693	0.987	1.703	1.007	−347.138	76	100	1.159	1.000	22.066
		IFM	3.316	0.155	0.894	0.105	0.052	1.924	0.759	1.026	1.696	1.003	−347.200	76	76	2.841	1.000	55.599
	ν_2	MBP1	3.247	0.100	0.772	0.082	0.033	0.827	0.192	1.000	1.442	1.000	−347.138	76	99	42.481	9.868	22.325
		MBP2	3.248	0.100	0.772	0.083	0.033	0.827	0.192	1.000	1.442	1.000	−347.138	76	99	39.127	9.303	21.511
		DIR	3.247	0.100	0.772	0.082	0.033	0.824	0.191	0.997	1.434	0.994	−347.138	76	100	1.159	1.000	22.066
		IFM	3.234	0.106	0.879	0.078	0.035	0.908	0.241	1.098	1.614	1.119	−347.200	76	76	2.841	1.000	55.599
	ν	MBP1	15.720	2.419	7.519	0.572	0.242	477.422	187.293	1.000	88.262	1.000	−347.138	76	99	42.481	9.868	22.325
		MBP2	15.716	2.419	7.519	0.572	0.242	477.242	187.295	1.000	88.267	1.000	−347.138	76	99	39.127	9.303	21.511
		DIR	13.963	1.803	7.460	0.396	0.180	262.717	92.856	0.550	89.490	1.014	−347.138	76	100	1.159	1.000	22.066
		IFM	13.454	1.806	5.741	0.345	0.181	259.834	92.944	0.544	58.533	0.663	−347.200	76	76	2.841	1.000	55.599
	ρ	MBP1	0.316	0.010	0.064	0.052	0.035	0.009	0.002	1.000	0.012	1.000	−347.138	76	99	42.481	9.868	22.325
		MBP2	0.316	0.010	0.064	0.052	0.035	0.009	0.002	1.000	0.012	1.000	−347.138	76	99	39.127	9.303	21.511
		DIR	0.315	0.010	0.065	0.052	0.035	0.009	0.002	0.999	0.012	1.011	−347.138	76	100	1.159	1.000	22.066
		IFM	0.316	0.011	0.062	0.052	0.035	0.009	0.002	1.018	0.009	0.777	−347.200	76	76	2.841	1.000	55.599
500	ν_1	MBP1	3.097	0.040	0.296	0.032	0.013	0.164	0.027	1.000	0.275	1.000	−1737.932	99	100	167.069	7.051	25.343
		MBP2	3.097	0.040	0.296	0.032	0.013	0.164	0.027	1.000	0.275	1.000	−1737.932	99	100	167.006	7.071	25.842
		DIR	3.097	0.039	0.295	0.010	0.013	0.164	0.027	0.999	0.275	1.000	−1737.933	99	100	4.172	1.000	16.354
		IFM	3.092	0.039	0.312	0.031	0.013	0.161	0.024	0.982	0.230	0.838	−1738.010	99	99	14.901	1.000	59.404
	ν_2	MBP1	3.072	0.037	0.399	0.024	0.012	0.141	0.021	1.000	0.359	1.000	−1737.932	99	100	167.069	7.051	25.343
		MBP2	3.072	0.037	0.399	0.024	0.012	0.141	0.021	1.000	0.359	1.000	−1737.932	99	100	167.006	7.071	25.842
		DIR	3.072	0.037	0.398	0.024	0.012	0.141	0.021	0.999	0.359	1.001	−1737.933	99	100	4.172	1.000	16.354
		IFM	3.063	0.037	0.397	0.021	0.012	0.141	0.021	0.995	0.351	0.978	−1738.010	99	99	14.901	1.000	59.404
	ν	MBP1	13.443	1.289	5.643	0.344	0.129	176.361	87.558	1.000	41.675	1.000	−1737.932	99	100	167.069	7.051	25.343
		MBP2	13.443	1.289	5.643	0.344	0.129	176.361	87.558	1.000	41.675	1.000	−1737.932	99	100	167.006	7.071	25.842
		DIR	12.729	0.967	5.575	0.273	0.097	100.038	37.940	0.567	44.271	1.062	−1737.933	99	100	4.172	1.000	16.354
		IFM	12.647	0.969	5.453	0.265	0.097	99.967	37.941	0.567	45.639	1.095	−1738.010	99	99	14.901	1.000	59.404
	ρ	MBP1	0.292	0.005	0.041	−0.027	0.015	0.002	0.000	1.000	0.004	1.000	−1737.932	99	100	167.069	7.051	25.343
		MBP2	0.292	0.005	0.041	−0.027	0.015	0.002	0.000	1.000	0.004	1.000	−1737.932	99	100	167.006	7.071	25.842
		DIR	0.292	0.005	0.040	−0.027	0.015	0.002	0.000	0.998	0.004	0.995	−1737.933	99	100	4.172	1.000	16.354
		IFM	0.290	0.005	0.038	−0.035	0.015	0.002	0.000	1.018	0.003	0.722	−1738.010	99	99	14.901	1.000	59.404

Table A.7: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


loglike
ρ = 0.3
ν = 10
$\nu_2 = 3$
$\nu_1 = 3$
MBP1
MBP2
DIR
IFM
n=100
n=500

n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
			$\nu_{1,tr}=3,\ \nu_{2,tr}=3,\ \nu_{tr}=10,\ \rho_{tr}=0.8$															
100	ν_1	MBP1	5.158	1.332	0.858	0.719	0.444	144.824	120.005	1.000	1.680	1.000	−300.667	79	99	57.010	12.772	21.597
		MBP2	5.298	1.391	0.858	0.766	0.464	158.225	122.895	1.093	1.675	0.997	−300.630	79	100	59.305	13.025	22.579
		DIR	3.320	0.149	0.838	0.032	0.050	1.861	0.773	0.013	1.747	1.040	−300.629	79	100	1.569	1.000	30.076
		IFM	3.304	0.163	0.817	0.101	0.054	2.181	0.936	0.015	1.249	0.744	−300.754	79	75	2.649	1.000	52.468
	ν_2	MBP1	5.499	1.467	0.775	0.833	0.489	176.334	128.957	1.000	1.344	1.000	−300.667	79	99	57.010	12.772	21.597
		MBP2	4.332	0.833	0.775	0.444	0.278	56.545	52.825	0.321	1.347	1.002	−300.630	79	100	59.305	13.025	22.579
		DIR	4.083	0.822	0.735	0.361	0.274	54.592	52.662	0.310	1.423	1.059	−300.629	79	100	1.569	1.000	30.076
		IFM	4.088	0.828	0.833	0.363	0.276	55.410	52.809	0.314	1.323	0.985	−300.754	79	75	2.649	1.000	52.468
	ν	MBP1	21.767	3.215	10.499	1.177	0.322	955.225	264.626	1.000	122.682	1.000	−300.667	79	99	57.010	12.772	21.597
		MBP2	20.837	3.054	10.500	1.084	0.305	854.268	248.545	0.894	122.785	1.001	−300.630	79	100	59.305	13.025	22.579
		DIR	15.364	1.850	10.302	0.536	0.185	299.108	120.400	0.313	122.211	0.996	−300.629	79	100	1.569	1.000	30.076
		IFM	17.944	2.612	7.504	0.794	0.261	602.266	206.493	0.630	97.673	0.796	−300.754	79	75	2.649	1.000	52.468
	ρ	MBP1	0.801	0.005	0.033	0.001	0.006	0.002	0.001	1.000	0.003	1.000	−300.667	79	99	57.010	12.772	21.597
		MBP2	0.796	0.004	0.034	−0.005	0.006	0.002	0.000	0.911	0.002	0.903	−300.630	79	100	59.305	13.025	22.579
		DIR	0.801	0.004	0.031	0.001	0.005	0.001	0.000	0.756	0.003	1.017	−300.629	79	100	1.569	1.000	30.076
		IFM	0.806	0.004	0.027	0.007	0.005	0.001	0.000	0.820	0.002	0.582	−300.754	79	75	2.649	1.000	52.468
500	ν_1	MBP1	3.094	0.037	0.333	0.031	0.012	0.140	0.022	1.000	0.267	1.000	−1504.775	97	100	298.211	10.629	42.156
		MBP2	3.094	0.037	0.333	0.031	0.012	0.140	0.022	1.001	0.267	1.000	−1504.775	97	100	297.659	10.557	42.290
		DIR	3.094	0.037	0.331	0.009	0.012	0.140	0.022	1.001	0.248	0.928	−1504.775	97	100	5.700	1.000	22.165
		IFM	3.085	0.037	0.375	0.028	0.012	0.143	0.020	1.022	0.324	1.215	−1505.105	97	97	15.044	1.000	59.778
	ν_2	MBP1	3.087	0.036	0.378	0.029	0.012	0.135	0.022	1.000	0.264	1.000	−1504.775	97	100	298.211	10.629	42.156
		MBP2	3.087	0.036	0.378	0.029	0.012	0.135	0.022	1.001	0.264	0.999	−1504.775	97	100	297.659	10.557	42.290
		DIR	3.087	0.036	0.379	0.029	0.012	0.135	0.022	1.000	0.256	0.970	−1504.775	97	100	5.700	1.000	22.165
		IFM	3.072	0.036	0.295	0.024	0.012	0.131	0.021	0.974	0.178	0.673	−1505.105	97	97	15.044	1.000	59.778
	ν	MBP1	12.821	0.982	4.311	0.282	0.098	101.448	49.914	1.000	36.576	1.000	−1504.775	97	100	298.211	10.629	42.156
		MBP2	12.839	0.983	4.312	0.284	0.098	101.860	49.908	1.004	36.565	1.000	−1504.775	97	100	297.659	10.557	42.290
		DIR	12.771	0.982	3.947	0.277	0.098	101.263	50.861	0.998	32.304	0.883	−1504.775	97	100	5.700	1.000	22.165
		IFM	11.815	0.711	4.409	0.181	0.071	52.389	17.459	0.516	33.186	0.907	−1505.105	97	97	15.044	1.000	59.778
	ρ	MBP1	0.797	0.002	0.015	−0.003	0.002	0.000	0.000	1.000	0.001	1.000	−1504.775	97	100	298.211	10.629	42.156
		MBP2	0.797	0.002	0.015	−0.003	0.002	0.000	0.000	1.000	0.001	0.999	−1504.775	97	100	297.659	10.557	42.290
		DIR	0.797	0.002	0.015	−0.003	0.002	0.000	0.000	0.998	0.001	0.995	−1504.775	97	100	5.700	1.000	22.165
		IFM	0.795	0.002	0.016	−0.007	0.002	0.000	0.000	1.290	0.001	1.037	−1505.105	97	97	15.044	1.000	59.778

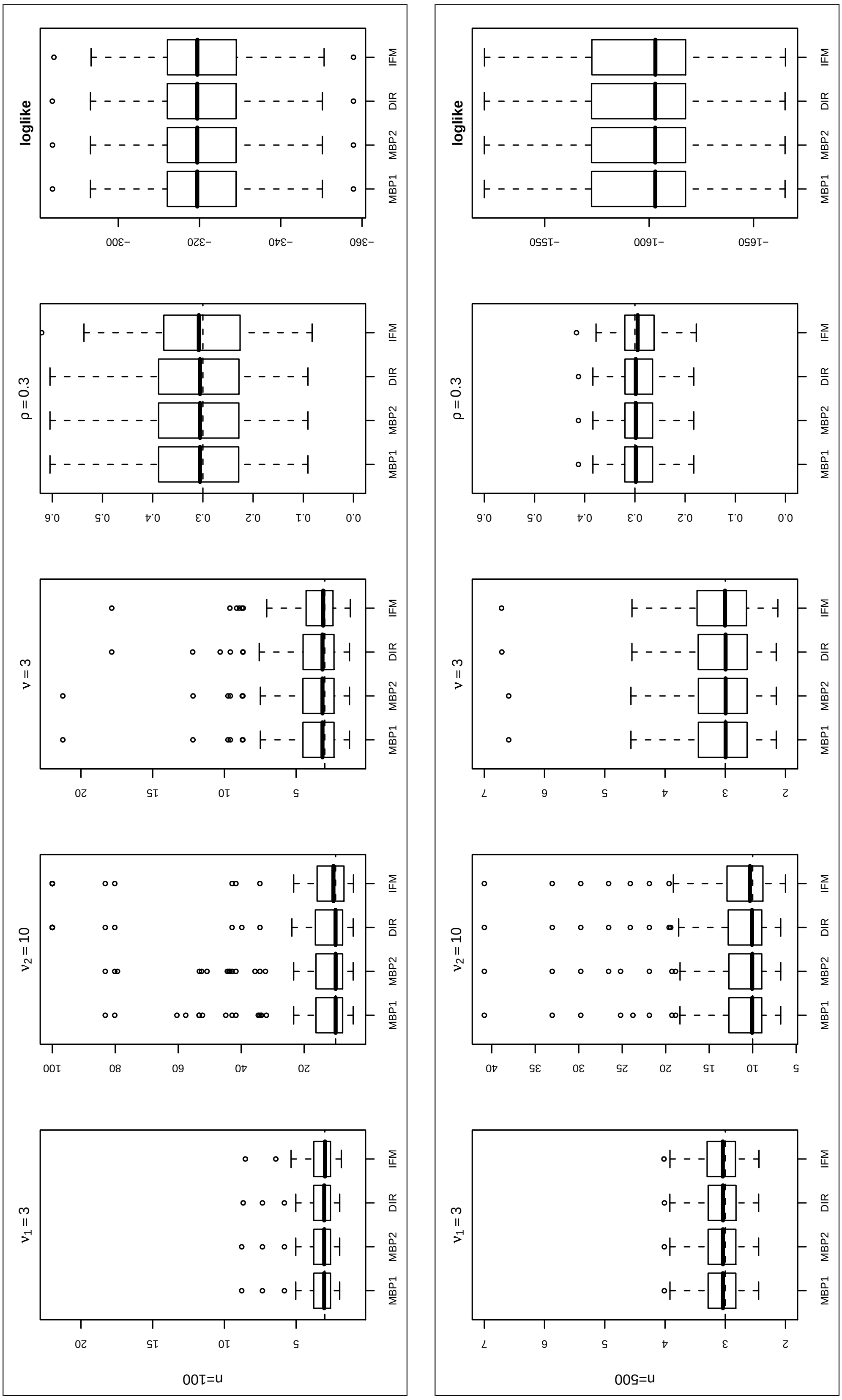
Table A.8: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
ν1 = 3
ν2 = 3
ν = 10
ρ = 0.8
loglike
MBP1
MBP2
DIR
IFM

$\nu_{1,tr} = 3, \quad \nu_{2,tr} = 10, \quad \nu_{tr} = 3, \quad \rho_{tr} = 0.3$

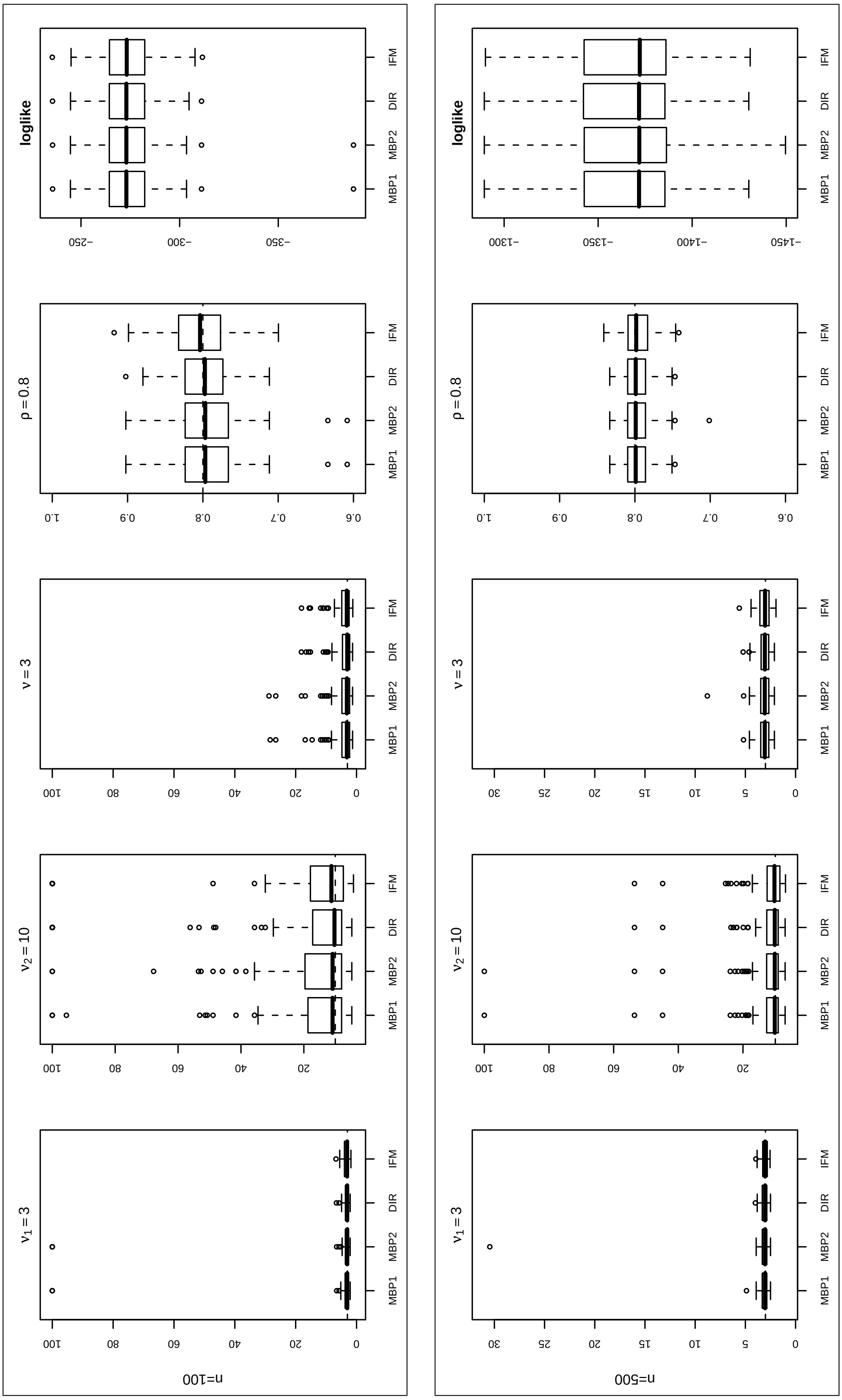
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	3.265	0.106	0.843	0.088	0.035	1.139	0.404	1.000	1.524	1.000	−319.438	96	100	51.863	11.833	22.341
		MBP2	3.266	0.106	0.843	0.089	0.035	1.140	0.404	1.001	1.523	0.999	−319.438	96	100	52.457	11.760	22.193
		DIR	3.273	0.105	0.845	0.091	0.035	1.137	0.394	0.998	1.508	0.990	−319.438	96	100	1.389	1.000	27.177
		IFM	3.276	0.104	0.856	0.092	0.035	1.113	0.347	0.977	1.181	0.775	−319.463	96	87	3.056	1.000	52.562
	ν_2	MBP1	16.458	1.616	6.219	0.646	0.162	292.309	88.238	1.000	52.459	1.000	−319.438	96	100	51.863	11.833	22.341
		MBP2	16.601	1.661	6.217	0.660	0.166	308.545	95.389	1.056	52.180	0.995	−319.438	96	100	52.457	11.760	22.193
		DIR	22.246	2.980	6.398	1.225	0.298	1002.477	258.670	3.430	52.996	1.010	−319.438	96	100	1.389	1.000	27.177
		IFM	21.247	2.873	6.069	1.125	0.287	918.992	247.908	3.144	68.136	1.299	−319.463	96	87	3.056	1.000	52.562
	ν	MBP1	3.830	0.271	1.590	0.277	0.090	7.720	3.610	1.000	3.846	1.000	−319.438	96	100	51.863	11.833	22.341
		MBP2	3.831	0.271	1.590	0.277	0.090	7.722	3.610	1.000	3.870	1.006	−319.438	96	100	52.457	11.760	22.193
		DIR	3.815	0.250	1.583	0.272	0.083	6.652	2.547	0.862	3.741	0.973	−319.438	96	100	1.389	1.000	27.177
		IFM	3.793	0.236	1.363	0.264	0.079	5.998	2.403	0.777	4.170	1.084	−319.463	96	87	3.056	1.000	52.562
	ρ	MBP1	0.308	0.011	0.117	0.028	0.036	0.011	0.001	1.000	0.031	1.000	−319.438	96	100	51.863	11.833	22.341
		MBP2	0.308	0.011	0.117	0.028	0.036	0.011	0.001	1.000	0.031	1.007	−319.438	96	100	52.457	11.760	22.193
		DIR	0.308	0.011	0.117	0.027	0.036	0.011	0.001	0.993	0.031	1.013	−319.438	96	100	1.389	1.000	27.177
		IFM	0.306	0.011	0.112	0.021	0.036	0.011	0.002	1.022	0.026	0.844	−319.463	96	87	3.056	1.000	52.562
500	ν_1	MBP1	3.087	0.032	0.334	0.029	0.011	0.113	0.018	1.000	0.248	1.000	−1602.927	100	100	344.851	11.530	30.351
		MBP2	3.087	0.033	0.334	0.029	0.011	0.113	0.018	1.003	0.248	1.000	−1602.927	100	100	346.896	11.480	30.823
		DIR	3.088	0.033	0.337	0.029	0.011	0.114	0.018	1.010	0.246	0.992	−1602.927	100	100	5.655	1.000	23.600
		IFM	3.083	0.033	0.342	0.028	0.011	0.115	0.018	1.018	0.260	1.047	−1602.984	100	100	17.765	1.000	63.210
	ν_2	MBP1	11.853	0.540	2.767	0.185	0.054	32.541	11.736	1.000	13.644	1.000	−1602.927	100	100	344.851	11.530	30.351
		MBP2	11.880	0.546	2.767	0.188	0.055	33.390	11.879	1.026	13.644	1.000	−1602.927	100	100	346.896	11.480	30.823
		DIR	11.892	0.546	2.796	0.189	0.055	33.380	11.835	1.026	13.458	0.986	−1602.927	100	100	5.655	1.000	23.600
		IFM	11.898	0.550	3.040	0.190	0.055	33.891	11.835	1.041	13.812	1.012	−1602.984	100	100	17.765	1.000	63.210
	ν	MBP1	3.090	0.063	0.588	0.030	0.021	0.400	0.132	1.000	0.712	1.000	−1602.927	100	100	344.851	11.530	30.351
		MBP2	3.090	0.063	0.589	0.030	0.021	0.400	0.132	1.000	0.712	1.000	−1602.927	100	100	346.896	11.480	30.823
		DIR	3.092	0.063	0.591	0.031	0.021	0.409	0.140	1.024	0.718	1.008	−1602.927	100	100	5.655	1.000	23.600
		IFM	3.101	0.064	0.606	0.034	0.021	0.426	0.141	1.066	0.875	1.229	−1602.984	100	100	17.765	1.000	63.210
	ρ	MBP1	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.000	−1602.927	100	100	344.851	11.530	30.351
		MBP2	0.292	0.005	0.040	−0.027	0.016	0.002	0.000	1.000	0.004	1.001	−1602.927	100	100	346.896	11.480	30.823
		DIR	0.292	0.005	0.040	−0.027	0.016	0.002	0.000	1.000	0.004	0.994	−1602.927	100	100	5.655	1.000	23.600
		IFM	0.292	0.005	0.042	−0.027	0.016	0.002	0.000	1.019	0.004	0.962	−1602.984	100	100	17.765	1.000	63.210

Table A.9: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.3
ν = 3
ν2 = 10
ν1 = 3
MBP1
MBP2
DIR
IFM

n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
								$\nu_{1,tr}=3,\quad \nu_{2,tr}=10,\quad \nu_{tr}=3,\quad \rho_{tr}=0.8$										
100	ν_1	MBP1	5.220	1.391	0.765	0.740	0.464	194.658	134.376	1.000	1.224	1.000	−272.996	98	100	72.071	12.561	29.252
		MBP2	5.221	1.391	0.720	0.740	0.464	194.656	134.376	1.000	1.279	1.045	−272.996	98	100	74.035	12.388	29.124
		DIR	3.243	0.082	0.723	0.081	0.027	0.714	0.165	0.004	1.247	1.019	−272.996	98	100	1.824	1.000	35.786
		IFM	3.303	0.097	0.974	0.101	0.032	1.007	0.191	0.005	1.414	1.155	−273.181	98	90	3.105	1.000	53.301
	ν_2	MBP1	17.814	1.837	7.820	0.781	0.184	391.924	138.736	1.000	75.192	1.000	−272.996	98	100	72.071	12.561	29.252
		MBP2	17.550	1.732	8.391	0.755	0.173	350.866	122.969	0.895	75.273	1.001	−272.996	98	100	74.035	12.388	29.124
		DIR	20.657	2.609	6.823	1.066	0.261	780.659	223.570	1.992	46.720	0.621	−272.996	98	100	1.824	1.000	35.786
		IFM	19.701	2.551	7.454	0.970	0.255	731.671	222.891	1.867	72.347	0.962	−273.181	98	90	3.105	1.000	53.301
	ν	MBP1	4.572	0.441	1.821	0.524	0.147	21.512	8.870	1.000	6.051	1.000	−272.996	98	100	72.071	12.561	29.252
		MBP2	4.618	0.453	1.820	0.539	0.151	22.741	9.188	1.057	5.989	0.990	−272.996	98	100	74.035	12.388	29.124
		DIR	4.189	0.329	1.727	0.396	0.110	12.019	3.737	0.559	5.036	0.832	−272.996	98	100	1.824	1.000	35.786
		IFM	4.244	0.311	1.739	0.415	0.104	11.003	3.358	0.511	5.718	0.945	−273.181	98	90	3.105	1.000	53.301
	ρ	MBP1	0.794	0.005	0.041	−0.007	0.006	0.002	0.000	1.000	0.004	1.000	−272.996	98	100	72.071	12.561	29.252
		MBP2	0.794	0.005	0.041	−0.007	0.006	0.002	0.000	0.999	0.004	1.000	−272.996	98	100	74.035	12.388	29.124
		DIR	0.798	0.004	0.037	−0.003	0.005	0.002	0.000	0.710	0.003	0.779	−272.996	98	100	1.824	1.000	35.786
		IFM	0.802	0.004	0.041	0.002	0.005	0.002	0.000	0.778	0.004	1.015	−273.181	98	90	3.105	1.000	53.301
500	ν_1	MBP1	3.102	0.035	0.334	0.034	0.012	0.130	0.038	1.000	0.277	1.000	−1371.742	100	100	568.682	14.180	49.856
		MBP2	3.358	0.275	0.334	0.119	0.092	7.707	7.499	59.265	0.274	0.988	−1371.879	100	100	548.594	14.110	46.703
		DIR	3.091	0.030	0.335	0.030	0.010	0.100	0.016	0.766	0.279	1.007	−1371.742	100	100	7.109	1.000	29.630
		IFM	3.070	0.031	0.328	0.023	0.010	0.102	0.015	0.782	0.236	0.853	−1372.135	100	100	17.817	1.000	63.150
	ν_2	MBP1	12.851	1.088	2.596	0.285	0.109	126.458	83.213	1.000	8.054	1.000	−1371.742	100	100	568.682	14.180	49.856
		MBP2	12.859	1.088	2.596	0.286	0.109	126.608	83.212	1.001	8.069	1.002	−1371.879	100	100	548.594	14.110	46.703
		DIR	12.034	0.643	2.562	0.203	0.064	45.510	22.381	0.360	8.004	0.994	−1371.742	100	100	7.109	1.000	29.630
		IFM	12.003	0.663	2.918	0.200	0.066	48.006	22.473	0.380	17.145	2.129	−1372.135	100	100	17.817	1.000	63.150
	ν	MBP1	3.154	0.062	0.580	0.051	0.021	0.411	0.066	1.000	0.773	1.000	−1371.742	100	100	568.682	14.180	49.856
		MBP2	3.204	0.084	0.580	0.068	0.028	0.743	0.336	1.808	0.773	1.000	−1371.879	100	100	548.594	14.110	46.703
		DIR	3.147	0.062	0.559	0.049	0.021	0.409	0.068	0.995	0.697	0.902	−1371.742	100	100	7.109	1.000	29.630
		IFM	3.119	0.063	0.670	0.040	0.021	0.416	0.078	1.014	1.052	1.361	−1372.135	100	100	17.817	1.000	63.150
	ρ	MBP1	0.797	0.002	0.017	−0.003	0.002	0.000	0.000	1.000	0.001	1.000	−1371.742	100	100	568.682	14.180	49.856
		MBP2	0.797	0.002	0.017	−0.004	0.003	0.000	0.000	1.250	0.001	1.000	−1371.879	100	100	548.594	14.110	46.703
		DIR	0.797	0.002	0.017	−0.003	0.002	0.000	0.000	0.975	0.001	1.005	−1371.742	100	100	7.109	1.000	29.630
		IFM	0.796	0.002	0.019	−0.005	0.002	0.000	0.000	1.200	0.001	1.211	−1372.135	100	100	17.817	1.000	63.150

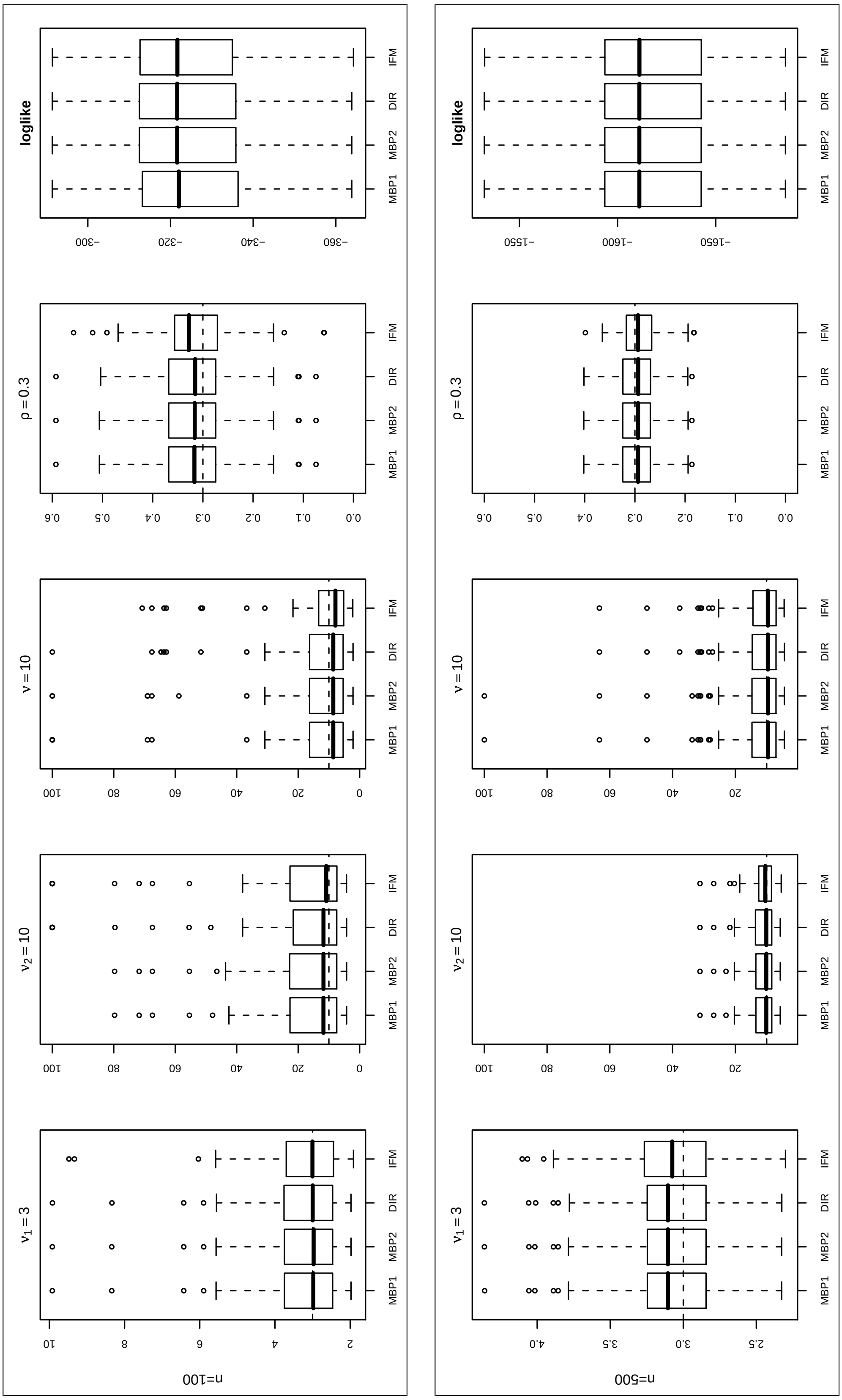
Table A.10: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
ν1 = 3
ν2 = 10
ν = 3
ρ = 0.8
loglike
MBP1
MBP2
DIR
IFM

$\nu_{1,tr} = 3, \quad \nu_{2,tr} = 10, \quad \nu_{tr} = 10, \quad \rho_{tr} = 0.3$

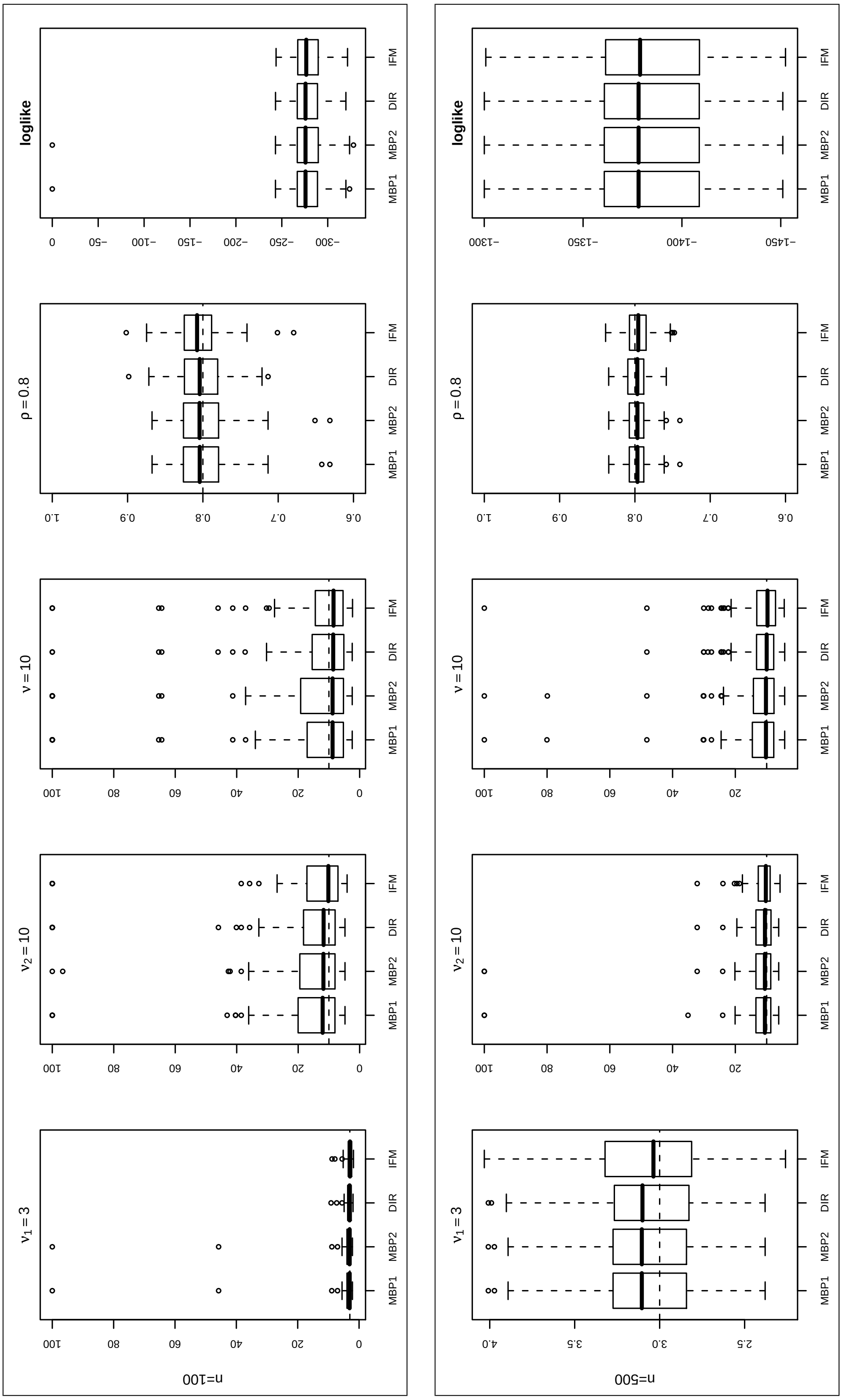
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	loglike	r_c	r_g	time	avg.it	func.eval
100	ν_1	MBP1	3.330	0.155	0.945	0.110	0.052	1.927	0.739	1.000	1.639	1.000	−322.492	76	99	44.753	9.776	21.766
		MBP2	3.330	0.155	0.946	0.110	0.052	1.927	0.739	1.000	1.630	0.994	−322.492	76	100	50.134	9.711	21.753
		DIR	3.333	0.155	0.952	0.033	0.052	1.930	0.738	1.002	1.726	1.053	−322.492	76	100	1.516	1.000	29.000
		IFM	3.307	0.155	0.894	0.102	0.052	1.924	0.759	0.998	1.648	1.006	−322.542	76	70	2.646	1.000	48.296
	ν_2	MBP1	18.243	1.849	11.230	0.824	0.185	327.797	95.518	1.000	125.897	1.000	−322.492	76	99	44.753	9.776	21.766
		MBP2	18.428	1.874	11.297	0.843	0.187	337.969	96.136	1.031	125.639	0.998	−322.492	76	100	50.134	9.711	21.753
		DIR	23.591	3.224	10.089	1.359	0.322	974.713	273.392	2.974	125.235	0.995	−322.492	76	100	1.516	1.000	29.000
		IFM	22.672	3.116	11.258	1.267	0.312	898.725	259.468	2.742	100.138	0.795	−322.542	76	70	2.646	1.000	48.296
	ν	MBP1	16.495	2.637	8.027	0.649	0.264	570.543	213.169	1.000	78.255	1.000	−322.492	76	99	44.753	9.776	21.766
		MBP2	15.573	2.301	8.031	0.557	0.230	433.486	166.054	0.760	78.242	1.000	−322.492	76	100	50.134	9.711	21.753
		DIR	14.710	2.027	8.024	0.471	0.203	334.541	129.769	0.586	80.759	1.032	−322.492	76	100	1.516	1.000	29.000
		IFM	13.571	1.779	5.937	0.357	0.178	253.394	85.970	0.444	60.416	0.772	−322.542	76	70	2.646	1.000	48.296
	ρ	MBP1	0.312	0.011	0.068	0.040	0.035	0.009	0.002	1.000	0.013	1.000	−322.492	76	99	44.753	9.776	21.766
		MBP2	0.312	0.011	0.068	0.040	0.035	0.009	0.002	1.000	0.013	1.025	−322.492	76	100	50.134	9.711	21.753
		DIR	0.312	0.011	0.068	0.039	0.035	0.009	0.002	0.993	0.013	1.054	−322.492	76	100	1.516	1.000	29.000
		IFM	0.313	0.011	0.063	0.043	0.035	0.009	0.002	1.001	0.009	0.697	−322.542	76	70	2.646	1.000	48.296
500	ν_1	MBP1	3.098	0.040	0.298	0.033	0.013	0.165	0.028	1.000	0.296	1.000	−1611.805	99	100	226.181	8.404	25.312
		MBP2	3.099	0.040	0.298	0.033	0.013	0.165	0.028	1.000	0.296	1.000	−1611.805	99	100	226.671	8.424	25.252
		DIR	3.099	0.040	0.298	0.010	0.013	0.165	0.028	1.000	0.292	0.984	−1611.806	99	100	5.309	1.000	20.788
		IFM	3.092	0.039	0.312	0.031	0.013	0.161	0.024	0.976	0.230	0.777	−1611.811	99	99	14.877	1.000	56.318
	ν_2	MBP1	11.413	0.422	3.752	0.141	0.042	19.601	5.684	1.000	22.112	1.000	−1611.805	99	100	226.181	8.404	25.312
		MBP2	11.414	0.422	3.759	0.141	0.042	19.607	5.684	1.000	22.101	1.000	−1611.805	99	100	226.671	8.424	25.252
		DIR	11.421	0.421	3.841	0.142	0.042	19.562	5.626	0.998	21.860	0.989	−1611.806	99	100	5.309	1.000	20.788
		IFM	11.321	0.421	3.025	0.132	0.042	19.255	5.624	0.982	20.483	0.926	−1611.811	99	99	14.877	1.000	56.318
	ν	MBP1	13.403	1.271	5.671	0.340	0.127	171.454	87.057	1.000	39.655	1.000	−1611.805	99	100	226.181	8.404	25.312
		MBP2	13.403	1.271	5.671	0.340	0.127	171.453	87.057	1.000	39.654	1.000	−1611.805	99	100	226.671	8.424	25.252
		DIR	12.669	0.936	5.626	0.267	0.094	93.947	33.644	0.548	40.048	1.010	−1611.806	99	100	5.309	1.000	20.788
		IFM	12.598	0.938	5.465	0.260	0.094	93.915	33.643	0.548	43.397	1.094	−1611.811	99	99	14.877	1.000	56.318
	ρ	MBP1	0.292	0.005	0.041	−0.027	0.015	0.002	0.000	1.000	0.004	1.000	−1611.805	99	100	226.181	8.404	25.312
		MBP2	0.292	0.005	0.041	−0.027	0.015	0.002	0.000	1.000	0.004	1.000	−1611.805	99	100	226.671	8.424	25.252
		DIR	0.292	0.005	0.041	−0.027	0.015	0.002	0.000	0.998	0.004	1.009	−1611.806	99	100	5.309	1.000	20.788
		IFM	0.290	0.005	0.038	−0.035	0.015	0.002	0.000	1.023	0.003	0.770	−1611.811	99	99	14.877	1.000	56.318

Table A.11: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.3
ν = 10
ν2 = 10
ν1 = 3
MBP1
MBP2
DIR
IFM

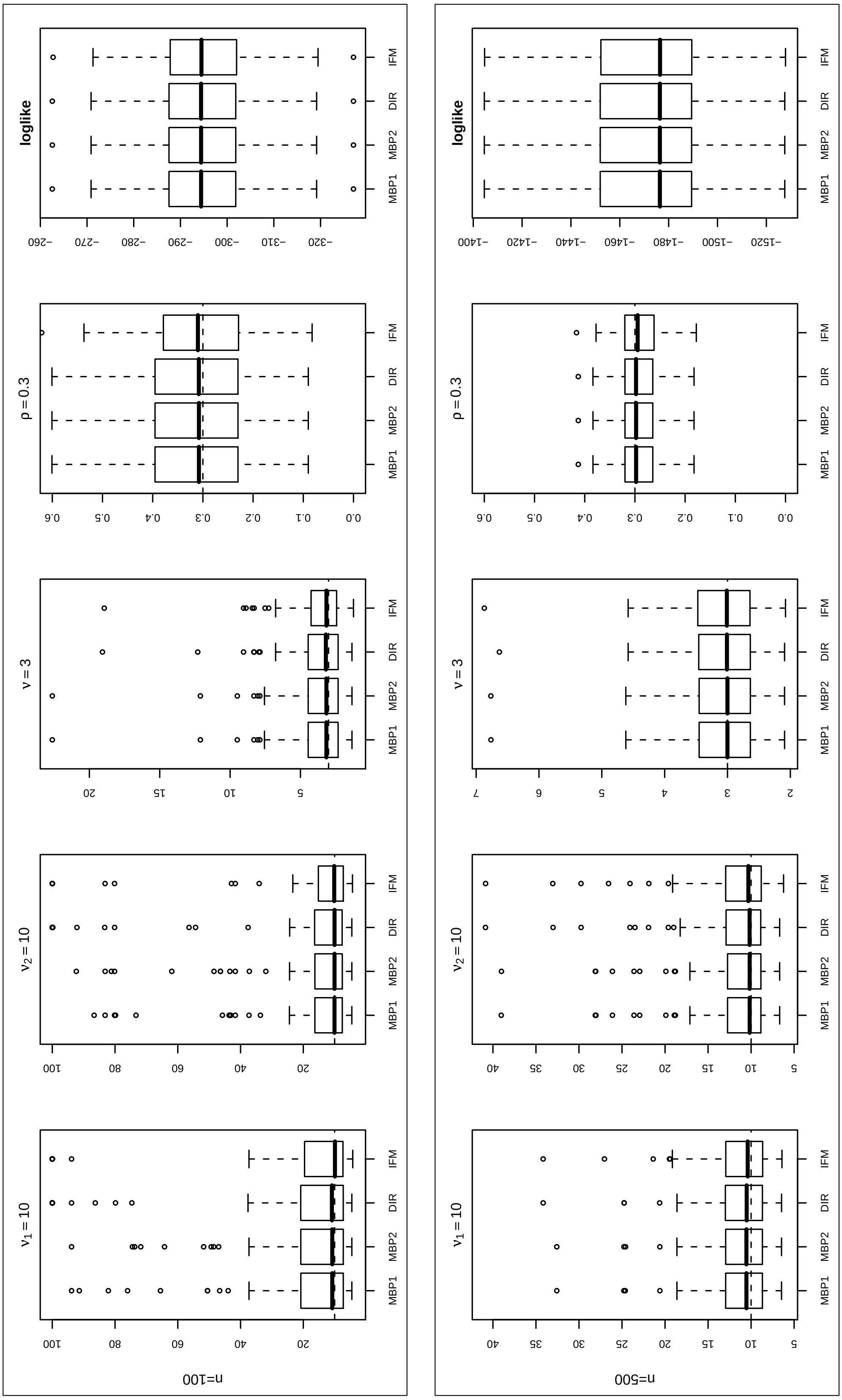
			$\nu_{1,tr}=3,\quad \nu_{2,tr}=10,\quad \nu_{tr}=10,\quad \rho_{tr}=0.8$															
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	5.271	1.465	0.941	0.757	0.488	159.646	131.831	1.000	1.923	1.000	−277.114	72	99	76.994	13.944	27.129
		MBP2	5.253	1.465	0.938	0.751	0.488	159.573	131.833	1.000	1.789	0.930	−277.114	72	99	71.654	13.042	25.950
		DIR	3.296	0.137	0.848	0.030	0.046	1.432	0.575	0.009	1.623	0.844	−276.757	72	100	1.738	1.000	33.639
		IFM	3.194	0.141	0.819	0.065	0.047	1.464	0.568	0.009	1.504	0.782	−276.910	72	64	2.541	1.000	47.125
	ν_2	MBP1	19.072	2.342	8.430	0.907	0.234	477.117	189.982	1.000	106.115	1.000	−277.114	72	99	76.994	13.944	27.129
		MBP2	17.731	2.012	8.100	0.773	0.201	351.338	151.572	0.736	105.692	0.996	−277.114	72	99	71.654	13.042	25.950
		DIR	19.732	2.792	7.539	0.973	0.279	656.005	241.322	1.375	109.896	1.036	−276.757	72	100	1.738	1.000	33.639
		IFM	18.141	2.787	7.368	0.814	0.279	625.447	241.453	1.311	68.416	0.645	−276.910	72	64	2.541	1.000	47.125
	ν	MBP1	20.939	3.388	8.681	1.094	0.339	946.193	283.518	1.000	109.422	1.000	−277.114	72	99	76.994	13.944	27.129
		MBP2	21.949	3.556	9.909	1.195	0.356	1053.119	300.125	1.113	113.109	1.034	−277.114	72	99	71.654	13.042	25.950
		DIR	15.310	2.255	7.514	0.531	0.226	394.337	165.578	0.417	93.406	0.854	−276.757	72	100	1.738	1.000	33.639
		IFM	16.147	2.552	6.529	0.615	0.255	506.853	196.764	0.536	84.020	0.768	−276.910	72	64	2.541	1.000	47.125
	ρ	MBP1	0.796	0.005	0.034	−0.005	0.006	0.002	0.001	1.000	0.003	1.000	−277.114	72	99	76.994	13.944	27.129
		MBP2	0.796	0.005	0.034	−0.005	0.006	0.002	0.001	0.978	0.003	1.010	−277.114	72	99	71.654	13.042	25.950
		DIR	0.801	0.004	0.032	0.001	0.005	0.001	0.000	0.633	0.002	0.896	−276.757	72	100	1.738	1.000	33.639
		IFM	0.805	0.004	0.027	0.006	0.005	0.001	0.000	0.807	0.002	0.593	−276.910	72	64	2.541	1.000	47.125
500	ν_1	MBP1	3.108	0.036	0.317	0.036	0.012	0.141	0.022	1.000	0.229	1.000	−1379.024	98	100	495.129	12.694	43.501
		MBP2	3.106	0.036	0.317	0.035	0.012	0.139	0.021	0.983	0.229	0.999	−1379.024	98	100	488.501	12.612	43.827
		DIR	3.085	0.036	0.323	0.008	0.012	0.132	0.021	0.936	0.225	0.981	−1379.026	98	100	6.979	1.000	27.235
		IFM	3.079	0.038	0.373	0.026	0.013	0.144	0.020	1.023	0.326	1.423	−1379.675	98	97	15.137	1.000	56.383
	ν_2	MBP1	13.277	1.338	3.501	0.328	0.134	186.101	115.614	1.000	22.474	1.000	−1379.024	98	100	495.129	12.694	43.501
		MBP2	13.244	1.333	3.484	0.324	0.133	184.675	115.568	0.992	22.484	1.000	−1379.024	98	100	488.501	12.612	43.827
		DIR	11.331	0.393	3.495	0.133	0.039	16.888	5.471	0.091	21.274	0.947	−1379.026	98	100	6.979	1.000	27.235
		IFM	11.166	0.400	2.789	0.117	0.040	17.056	5.533	0.092	15.492	0.689	−1379.675	98	97	15.137	1.000	56.383
	ν	MBP1	13.817	1.326	4.877	0.382	0.133	186.884	96.374	1.000	41.376	1.000	−1379.024	98	100	495.129	12.694	43.501
		MBP2	13.766	1.324	4.728	0.377	0.132	185.853	96.286	0.994	41.376	1.000	−1379.024	98	100	488.501	12.612	43.827
		DIR	12.236	0.714	4.025	0.224	0.071	54.955	16.672	0.294	30.062	0.727	−1379.026	98	100	6.979	1.000	27.235
		IFM	12.695	1.136	4.391	0.269	0.114	133.820	83.314	0.716	35.736	0.864	−1379.675	98	97	15.137	1.000	56.383
	ρ	MBP1	0.797	0.002	0.014	−0.004	0.002	0.000	0.000	1.000	0.000	1.000	−1379.024	98	100	495.129	12.694	43.501
		MBP2	0.797	0.002	0.014	−0.004	0.002	0.000	0.000	0.993	0.000	1.000	−1379.024	98	100	488.501	12.612	43.827
		DIR	0.798	0.002	0.015	−0.003	0.002	0.000	0.000	0.915	0.000	0.917	−1379.026	98	100	6.979	1.000	27.235
		IFM	0.795	0.002	0.016	−0.007	0.002	0.000	0.000	1.255	0.001	1.150	−1379.675	98	97	15.137	1.000	56.383

Table A.12: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
$\nu_1 = 3$
$\nu_2 = 10$
$\nu = 10$
$\rho = 0.8$
loglike
MBP1
MBP2
DIR
IFM

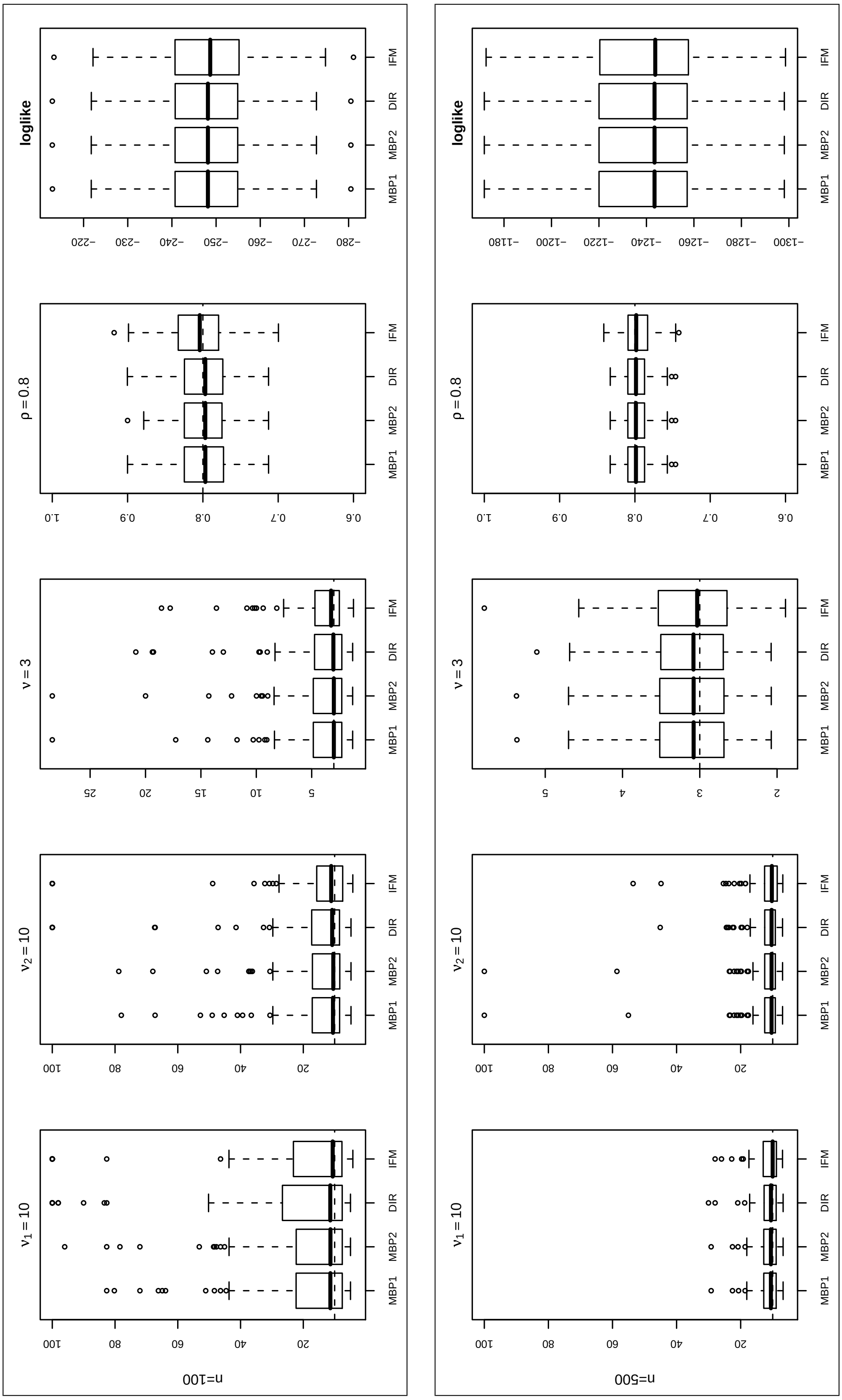
$\nu_{1,tr}=10,\ \nu_{2,tr}=10,\ \nu_{tr}=3,\ \rho_{tr}=0.3$																	
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	17.717	1.904	9.790	0.772	0.190	400.456	127.844	1.000	83.937	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	17.454	1.796	9.793	0.745	0.180	358.834	109.498	0.896	84.000	1.001	−294.565	94	100	37.563	12.777	17.337
		DIR	23.640	3.069	10.012	4.547	0.307	1071.586	263.009	2.676	88.558	1.055	−294.565	94	100	1.698	1.000	33.340
		IFM	23.954	3.211	8.996	1.395	0.321	1164.066	284.065	2.907	75.543	0.900	−294.619	94	78	2.863	1.000	46.027
	ν_2	MBP1	16.779	1.838	6.388	0.678	0.184	363.406	117.756	1.000	51.684	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	16.815	1.847	6.385	0.682	0.185	367.119	120.500	1.010	51.634	0.999	−294.565	94	100	37.563	12.777	17.337
		DIR	20.919	2.810	6.451	1.092	0.281	861.665	237.432	2.371	53.358	1.032	−294.565	94	100	1.698	1.000	33.340
		IFM	19.508	2.679	5.849	0.951	0.268	764.898	228.374	2.105	55.333	1.071	−294.619	94	78	2.863	1.000	46.027
	ν	MBP1	3.812	0.278	1.511	0.271	0.093	7.951	4.183	1.000	4.044	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	3.812	0.278	1.511	0.271	0.093	7.945	4.183	0.999	4.050	1.001	−294.565	94	100	37.563	12.777	17.337
		DIR	3.814	0.255	1.513	0.271	0.085	6.753	2.912	0.849	4.114	1.017	−294.565	94	100	1.698	1.000	33.340
		IFM	3.773	0.242	1.336	0.258	0.081	6.123	2.754	0.770	4.387	1.085	−294.619	94	78	2.863	1.000	46.027
	ρ	MBP1	0.313	0.011	0.121	0.042	0.037	0.012	0.002	1.000	0.032	1.000	−294.565	94	100	37.292	12.862	17.522
		MBP2	0.313	0.011	0.121	0.042	0.037	0.012	0.002	1.000	0.032	1.000	−294.565	94	100	37.563	12.777	17.337
		DIR	0.312	0.011	0.121	0.041	0.037	0.011	0.002	0.998	0.032	1.003	−294.565	94	100	1.698	1.000	33.340
		IFM	0.310	0.011	0.111	0.034	0.037	0.012	0.002	1.001	0.027	0.846	−294.619	94	78	2.863	1.000	46.027
500	ν_1	MBP1	11.458	0.402	3.153	0.146	0.040	18.299	5.975	1.000	19.273	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	11.458	0.402	3.153	0.146	0.040	18.290	5.972	1.000	19.280	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	11.487	0.412	3.177	0.496	0.041	19.224	6.611	1.051	19.048	0.988	−1476.409	100	100	6.879	1.000	28.700
		IFM	11.511	0.419	3.142	0.151	0.042	19.832	6.679	1.084	16.427	0.852	−1476.487	100	100	17.737	1.000	59.715
	ν_2	MBP1	11.821	0.513	2.816	0.182	0.051	29.583	10.013	1.000	14.477	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	11.821	0.512	2.816	0.182	0.051	29.572	10.004	1.000	14.484	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	11.882	0.538	2.942	0.188	0.054	32.477	11.676	1.098	14.625	1.010	−1476.409	100	100	6.879	1.000	28.700
		IFM	11.898	0.550	3.040	0.190	0.055	33.891	11.835	1.146	13.812	0.954	−1476.487	100	100	17.737	1.000	59.715
	ν	MBP1	3.093	0.064	0.603	0.031	0.021	0.412	0.144	1.000	0.703	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	3.093	0.064	0.603	0.031	0.021	0.412	0.144	1.000	0.703	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	3.093	0.063	0.605	0.031	0.021	0.403	0.134	0.977	0.710	1.011	−1476.409	100	100	6.879	1.000	28.700
		IFM	3.106	0.066	0.607	0.035	0.022	0.444	0.152	1.076	0.856	1.218	−1476.487	100	100	17.737	1.000	59.715
	ρ	MBP1	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.000	−1476.409	100	100	169.577	13.850	15.426
		MBP2	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.000	−1476.409	100	100	167.756	13.930	15.110
		DIR	0.292	0.005	0.041	−0.027	0.016	0.002	0.000	1.000	0.004	1.020	−1476.409	100	100	6.879	1.000	28.700
		IFM	0.292	0.005	0.042	−0.027	0.016	0.002	0.000	1.019	0.004	0.986	−1476.487	100	100	17.737	1.000	59.715

Table A.13: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
loglike
ρ = 0.3
ν = 3
ν2 = 10
ν1 = 10
MBP1
MBP2
DIR
IFM

$\nu_{1,tr}=10,\ \nu_{2,tr}=10,\ \nu_{tr}=3,\ \rho_{tr}=0.8$																		
n	θ	Method	$\bar{\theta}$	$s(\bar{\theta})$	$\hat{\sigma}_r(\hat{\theta})$	$\widehat{rb}(\bar{\theta})$	$s_{rb}(\bar{\theta})$	$\widehat{mse}(\hat{\theta})$	$s_{mse}(\hat{\theta})$	$\widehat{eff}_{mse}$	$\widehat{rmse}(\hat{\theta})$	$\widehat{eff}_{rmse}$	*loglike*	r_c	r_g	*time*	*avg.it*	*func.eval*
100	ν_1	MBP1	18.701	1.791	10.867	0.870	0.179	383.597	100.278	1.000	113.923	1.000	−248.268	96	100	47.738	14.479	19.588
		MBP2	18.550	1.821	10.756	0.855	0.182	391.570	114.854	1.021	112.860	0.991	−248.268	96	100	48.876	14.469	19.453
		DIR	24.266	2.959	14.030	4.755	0.296	1043.839	251.554	2.721	110.269	0.968	−248.268	96	100	2.156	1.000	42.479
		IFM	24.749	3.133	11.102	1.475	0.313	1159.646	274.309	3.023	91.587	0.804	−248.672	96	83	2.902	1.000	47.302
	ν_2	MBP1	15.375	1.315	6.436	0.537	0.132	194.946	64.751	1.000	56.917	1.000	−248.268	96	100	47.738	14.479	19.588
		MBP2	15.175	1.281	6.431	0.517	0.128	184.254	64.236	0.945	56.314	0.989	−248.268	96	100	48.876	14.469	19.453
		DIR	19.190	2.401	6.498	0.919	0.240	637.662	203.193	3.271	65.631	1.153	−248.268	96	100	2.156	1.000	42.479
		IFM	18.028	2.310	5.804	0.803	0.231	576.860	199.306	2.959	67.329	1.183	−248.672	96	83	2.902	1.000	47.302
	ν	MBP1	4.222	0.379	1.897	0.407	0.126	15.295	7.137	1.000	5.778	1.000	−248.268	96	100	47.738	14.479	19.588
		MBP2	4.249	0.391	1.899	0.416	0.130	16.208	7.433	1.060	5.806	1.005	−248.268	96	100	48.876	14.469	19.453
		DIR	4.297	0.374	1.795	0.432	0.125	15.093	5.265	0.987	5.763	0.997	−248.268	96	100	2.156	1.000	42.479
		IFM	4.192	0.315	1.611	0.397	0.105	10.944	3.656	0.716	6.146	1.064	−248.672	96	83	2.902	1.000	47.302
	ρ	MBP1	0.798	0.004	0.037	−0.003	0.005	0.002	0.000	1.000	0.004	1.000	−248.268	96	100	47.738	14.479	19.588
		MBP2	0.798	0.004	0.037	−0.003	0.005	0.002	0.000	0.999	0.003	0.982	−248.268	96	100	48.876	14.469	19.453
		DIR	0.797	0.004	0.037	−0.003	0.005	0.002	0.000	1.000	0.003	0.951	−248.268	96	100	2.156	1.000	42.479
		IFM	0.803	0.004	0.039	0.003	0.005	0.002	0.000	1.114	0.004	1.033	−248.672	96	83	2.902	1.000	47.302
500	ν_1	MBP1	11.370	0.346	2.808	0.137	0.035	13.862	4.211	1.000	15.365	1.000	−1243.409	100	100	227.135	16.260	16.042
		MBP2	11.373	0.347	2.809	0.137	0.035	13.912	4.221	1.004	15.396	1.002	−1243.409	100	100	227.625	16.230	16.008
		DIR	11.384	0.369	2.765	0.461	0.037	15.517	5.254	1.119	14.896	0.969	−1243.409	100	100	9.031	1.000	37.660
		IFM	11.288	0.378	3.037	0.129	0.038	15.966	4.509	1.152	13.800	0.898	−1243.718	100	100	17.652	1.000	59.650
	ν_2	MBP1	12.663	1.048	2.409	0.266	0.105	116.944	82.779	1.000	9.838	1.000	−1243.409	100	100	227.135	16.260	16.042
		MBP2	12.701	1.063	2.410	0.270	0.106	120.368	83.615	1.029	9.844	1.001	−1243.409	100	100	227.625	16.230	16.008
		DIR	11.811	0.508	2.412	0.181	0.051	29.115	12.783	0.249	10.191	1.036	−1243.409	100	100	9.031	1.000	37.660
		IFM	12.003	0.663	2.918	0.200	0.066	48.006	22.473	0.410	17.145	1.743	−1243.718	100	100	17.652	1.000	59.650
	ν	MBP1	3.154	0.063	0.606	0.051	0.021	0.415	0.071	1.000	0.788	1.000	−1243.409	100	100	227.135	16.260	16.042
		MBP2	3.154	0.063	0.606	0.051	0.021	0.416	0.072	1.001	0.791	1.004	−1243.409	100	100	227.625	16.230	16.008
		DIR	3.150	0.062	0.584	0.050	0.021	0.401	0.064	0.967	0.755	0.959	−1243.409	100	100	9.031	1.000	37.660
		IFM	3.120	0.065	0.654	0.040	0.022	0.440	0.087	1.059	1.005	1.276	−1243.718	100	100	17.652	1.000	59.650
	ρ	MBP1	0.797	0.002	0.016	−0.003	0.002	0.000	0.000	1.000	0.001	1.000	−1243.409	100	100	227.135	16.260	16.042
		MBP2	0.797	0.002	0.016	−0.003	0.002	0.000	0.000	1.000	0.001	0.996	−1243.409	100	100	227.625	16.230	16.008
		DIR	0.797	0.002	0.016	−0.003	0.002	0.000	0.000	0.989	0.001	0.999	−1243.409	100	100	9.031	1.000	37.660
		IFM	0.796	0.002	0.019	−0.005	0.002	0.000	0.000	1.254	0.001	1.350	−1243.718	100	100	17.652	1.000	59.650

Table A.14: Estimated ν_1's, ν_2's, ν's and ρ's comparing MbP with both nonadaptive (MBP1) and adaptive (MBP2) interchangeable parameters, direct maximization (DIR) and inference for margins method (IFM) over 100 replicates ($r = 100$) and samples of size $n = 100, 500$.


n=100
n=500
$\nu_1 = 10$
$\nu_2 = 10$
$\nu = 3$
$\rho = 0.8$
loglike
MBP1
MBP2
DIR
IFM

Bibliography

Baumann, T. (Dezember 2007). Modellierung von stochastischen abhängigkeitsstrukturen auf graphen. Master's thesis, Technische Universität München.

Bickel, P. J. and K. A. Doksum (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day, Inc.

Brockwell, P. J. and R. A. Davis (1996). *Introduction to time series and forecasting*. Springer Texts in Statistics. New York: Springer-Verlag. With 1 IBM-PC floppy disk (3.5 inch; HD).

Brown, M. (1982, March). *Robust Line Estimation with Errors in Both Variables*, Volume 77. American Statistical Association.

Demarta, S. and A. McNeil (2004). The t copula and related copulas. Technical report, ETH Zuerich.

Dominitz, J. and R. P. Sherman (2005). Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory 21*(4), 838–863.

Embrechts, Paul, F. L. A. M. (2001). Modelling dependence with copulas and applications to risk management.

Fang, H.-B., K.-T. Fang, and S. Kotz (2005). Corrigendum to: "The meta-elliptical distributions with given marginals" [J. Multivariate Anal. **82** (2002), no. 1, 1–16]. *J. Multivariate Anal. 94*(1), 222–223.

Fang, K., S. K. and K. Ng (1987). *Symmetric Multivariate and Related Distributions*. Chapman & Hall.

Georgii, H.-O. (2002). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Volume 139. de Gruyter.

Joe, X. (1996). The estimation method of inference functions for margins for multivariate models. pp. 166.

Kastenmeier (April 2008). Joint regression analysis of insurance claims and claim size. Master's thesis, Technische Universität München.

Kibria, B., A. J. (2006). The multivairate t-distribution and some of its application. Technical report, Department of Mathematical Science, King Fahd University of Petroleum & Minerals.

Kuhn, G. (November 2001). Multivariate value at risk schätzung von zins swap s"atzen. Master's thesis, Technische Universität München.

Lindskog, Filip, A. M. U. S. (2001). Kendall's tau for elliptical distribution.

Nelsen, R. B. (1999). *An introduction to copulas* (Second ed.). Springer Series in Statistics. New York: Springer.

Ripley, Brian D., W. N. V. (2002). *Modern Applied Statistics with S.* Birkhäuser.

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

Song, P. X.-K., Y. Fan, and J. D. Kalbfleisch (2005). Maximization by parts in likelihood inference. *J. Amer. Statist. Assoc. 100*(472), 1145–1167. With comments and a rejoinder by the authors.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.

Zivot, E., W. J. (2003). *Modeling Financial Time Series With S-Plus*. Springer.