ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Multivariate Analysis 94 (2005) 401–419
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Asymptotic efficiency of the two-stage estimation method for copula-based models

Harry Joe*

Department of Statistics, University of British Columbia, Room 333 6356, Agriculture Road, Vancouver, Canada V6T 1Z2

Received 27 March 2003
Available online 13 August 2004

Abstract

For multivariate copula-based models for which maximum likelihood is computationally difficult, a two-stage estimation procedure has been proposed previously; the first stage involves maximum likelihood from univariate margins, and the second stage involves maximum likelihood of the dependence parameters with the univariate parameters held fixed from the first stage. Using the theory of inference functions, a partitioned matrix in a form amenable to analysis is obtained for the asymptotic covariance matrix of the two-stage estimator. The asymptotic relative efficiency of the two-stage estimation procedure compared with maximum likelihood estimation is studied. Analysis of the limiting cases of the independence copula and Fréchet upper bound help to determine common patterns in the efficiency as the dependence in the model increases. For the Fréchet upper bound, the two-stage estimation procedure can sometimes be equivalent to maximum likelihood estimation for the univariate parameters. Numerical results are shown for some models, including multivariate ordinal probit and bivariate extreme value distributions, to indicate the typical level of asymptotic efficiency for discrete and continuous data.

AMS 1991 subject classification: 62H12; 62F12

Keywords: Copula; Multivariate non-normal; Estimating or inference functions; Fréchet bounds; Latent value models; Generalized extreme value

* Fax: +1-604-822-6960.
E-mail address: harry@stat.ubc.ca.

0047-259X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2004.06.003

1. Introduction

We study efficiency properties of a two-stage estimation procedure for copula-based models. This estimation method is intended for situations where maximum likelihood is computationally too difficult or infeasible. Further theoretical results are obtained beyond those in [23,7, Chapter 10].

The use of copulas is a general approach to model multivariate non-normal data [7], with the dependence structure separated from the univariate margins. The general form of a copula-based m–variate cumulative distribution function (cdf) is

$$F(\mathbf{y}; \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\delta}) = C(F_1(y_1, \boldsymbol{\alpha}_1), \ldots, F_m(y_m, \boldsymbol{\alpha}_m); \boldsymbol{\delta}), \quad \mathbf{y} = (y_1, \ldots, y_m)^T, \quad (1.1)$$

where $F_j(\cdot; \boldsymbol{\alpha}_j)$ is the jth univariate margin with parameter $\boldsymbol{\alpha}_j$ and $C(\cdot; \boldsymbol{\delta})$ is a family of copulas or multivariate uniform distributions with dependence parameter $\boldsymbol{\delta}$. In general $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\delta}$ are vectors. We refer to $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ as univariate parameters and $\boldsymbol{\delta}$ as the multivariate parameter.

The two-stage estimation procedure also applies more generally to models $F(\mathbf{y}; \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\delta})$ with multivariate dependence parameter $\boldsymbol{\delta}$ and univariate parameter $\boldsymbol{\alpha}_j$ in the jth univariate margin.

We assume (1.1) is a good model for applications, in that it includes the independence copula and the Fréchet upper bound, and the Fréchet lower bound for $m = 2$ when strong negative dependence is needed.

The Fréchet upper bound with copula $C(u_1, \ldots, u_m) = \min\{u_1, \ldots, u_m\}$ represents the most positive dependence; in the case of continuous variables, it implies perfect dependence with each variable being an increasing monotone transformation of any other variable. For $m = 2$, the Fréchet lower bound with copula $C(u_1, u_2) = \max\{0, u_1 + u_2 - 1\}$ represents the most negative dependence; in the case of continuous variables, it implies perfect negative dependence with one variable being a decreasing monotone transformation of the other variable. See Chapter 3 of [7] for a detailed discussion of Fréchet bounds.

When the joint multivariate likelihood is computationally difficult to work with, researchers have used a two-stage procedure of firstly estimating the univariate parameters from separate univariate likelihoods and then secondly estimating the multivariate parameters from the multivariate likelihood with the univariate parameters given the values from the first stage. Note that this two-stage procedure is the same as maximum likelihood estimation for the multivariate normal distribution (C is multivariate normal copula with correlation matrix R, F_j corresponds to $N(\mu_j, \sigma_j^2)$).

Examples of multivariate models (1.1) where numerical computations become harder as the dimension m increases, and where the two-stage estimation procedure has been or could be used, include the following.

(a) Multivariate normal latent models for categorical data (with a multivariate normal copula and discrete univariate margins) such as multivariate probit; this has mainly been in the psychometrics literature such as [12–14,9,11].

(b) Multivariate extreme value models with generalized extreme value margins; the two-stage estimation method is partly used as a way to get a starting point for the MLE (if the latter is possible) and to compare different models. Some references are [21,22,6].

(c) Multivariate log-normal Poisson model for count data [1,23].
(d) Multivariate logit models for ordinal data [7].
(e) Models for multivariate survival data (extensions of models in [19]).

The relative efficiency of the two-stage estimation procedure has not previously been studied theoretically in its generality. This is not a tractible problem for theoretical analysis, as the asymptotic covariance matrices of the estimators do not have closed form. However, based on analysis of the estimators for the Fréchet bounds and independence copula, we are able to obtain some general results on the efficiency of the two-stage method. Xu [23] has simulations and calculations for multivariate discrete/categorical models that show the two-stage method is highly efficient compared with maximum likelihood (ML). It is actually in the case of continuous variables where the two-stage method may have low efficiency relative to ML for the univariate parameters. The efficiency for the dependence parameters is generally high with the two-stage method.

The organization of the remainder of the paper is as follows. In Section 2, the asymptotic covariance matrix of the two-stage estimator is derived in a partitioned form, using theory of inference functions. For some models it may be possible to get empirical estimates of the asymptotic covariance matrix. Section 3 has analyses of the two-stage estimator for the cases of the Fréchet bounds and the independence copula in (1.1). Section 4 combines the results of Sections 2 and 3 to quantify the patterns in efficiency as a function of dependence, and illustrates this with a few examples. Section 5 consists of conclusions and some discussion.

2. Inference function for margins

In this section, we study the estimation of parameters of a copula-based multivariate model based on the two-stage estimation method. Following Xu [23], we called this the method of inference function for margins (IFM), because the inference or estimating functions correspond to likelihood score functions (univariate or multivariate). This method was proposed in a general framework in Xu [23] and Joe [7, Chapter 10], but had been used much earlier in the psychometrics literature for latent models based on the multivariate normal distributions. The method has been used mainly for multivariate models in which a multi-parameter numerical optimization for maximum likelihood estimation is too time-consuming or infeasible. The theoretical analysis given here is for independent and identically distributed (iid) observations, but the estimation method can also be used when there are covariates.

The usual regularity conditions for maximum likelihood are assumed for our analysis. For notation, we use boldface for vectors but not for matrices; vectors are assumed to be column vectors.

Consider a copula-based parametric model for the random m-vector $\mathbf{Y}$, with cdf given by (1.1). We assume that C has a density c (mixed derivative of order m). The vector $\mathbf{Y}$ could be discrete or continuous, but for notational simplicity, we will not consider the mixed case of some continuous and some discrete variables. In the former discrete case, the joint probability mass function (pmf) $f(\cdot; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m, \boldsymbol{\delta})$ for $\mathbf{Y}$ can be derived from the cdf in (1.1) as rectangle probabilities, and we let the univariate marginal pmfs be denoted by $f_1, \dots, f_m$; in the latter continuous case, we assume that F_j has density f_j for $j = 1, \dots, m$, and that $\mathbf{Y}$ has density

$$f(\mathbf{y};\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m,\boldsymbol{\delta}) = c(F_1(y_1;\boldsymbol{\alpha}_1),\ldots,F_m(y_m;\boldsymbol{\alpha}_m);\boldsymbol{\delta})\prod_{j=1}^{m} f_j(y_j;\boldsymbol{\alpha}_j). \tag{2.1}$$

For a sample of size n, with observed random vectors $\mathbf{y}_1,\ldots,\mathbf{y}_n$, we can consider the m log-likelihood functions for the univariate margins,

$$L_j(\boldsymbol{\alpha}_j) = \sum_{i=1}^{n} \log\, f_j(y_{ij};\boldsymbol{\alpha}_j), \quad j=1,\ldots,m \tag{2.2}$$

and the log-likelihood function for the joint distribution,

$$L(\boldsymbol{\delta},\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m) = \sum_{i=1}^{n} \log\, f(\mathbf{y}_i;\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m,\boldsymbol{\delta}). \tag{2.3}$$

With multivariate models in general, one does not have closed form estimators (maximum likelihood or other methods) and numerical techniques are needed. For ML estimation, the number of parameters increases with the dimension m and numerical optimization becomes more difficult as the total number of parameters increases (e.g., with a quasi-Newton method). Also for some models, m-dimensional numerical integration is needed, and this becomes increasing difficult as m increases. Hence this is the motivation for the two-stage estimation method.

Consider the two-stage process where in the first stage m separate optimizations of the univariate likelihoods, and in the second stage an optimization of the multivariate likelihood as a function of the dependence parameter vector. More specifically,

(a) the log-likelihoods L_j of the m univariate margins are separately maximized to get estimates $\tilde{\boldsymbol{\alpha}}_1,\ldots,\tilde{\boldsymbol{\alpha}}_m$;

(b) the function $L(\boldsymbol{\delta},\tilde{\boldsymbol{\alpha}}_1,\ldots,\tilde{\boldsymbol{\alpha}}_m)$ is maximized over $\boldsymbol{\delta}$ to get $\tilde{\boldsymbol{\delta}}$.

Under regularity conditions, $(\tilde{\boldsymbol{\alpha}}_1,\ldots,\tilde{\boldsymbol{\alpha}}_m,\tilde{\boldsymbol{\delta}})$ is the solution of

$$(\partial L_1/\partial\boldsymbol{\alpha}_1^T,\ldots,\partial L_m/\partial\boldsymbol{\alpha}_m^T,\partial L/\partial\boldsymbol{\delta}^T) = \mathbf{0}^T. \tag{2.4}$$

This procedure is computationally simpler than estimating all parameters $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m,\boldsymbol{\delta}$ simultaneously from L in (2.3). For comparison, we let $(\hat{\boldsymbol{\alpha}}_1,\ldots,\hat{\boldsymbol{\alpha}}_m,\hat{\boldsymbol{\delta}})$ denote the maximum likelihood estimate (MLE). Under regularity conditions, this comes from solving

$$(\partial L/\partial\boldsymbol{\alpha}_1^T,\ldots,\partial L/\partial\boldsymbol{\alpha}_m^T,\partial L/\partial\boldsymbol{\delta}^T) = \mathbf{0}^T. \tag{2.5}$$

For notation, we let $\boldsymbol{\eta}^T = (\boldsymbol{\alpha}_1^T,\ldots,\boldsymbol{\alpha}_m^T,\boldsymbol{\delta}^T)$ be the vector of all parameters (univariate and multivariate), $\tilde{\boldsymbol{\eta}}^T = (\tilde{\boldsymbol{\alpha}}_1^T,\ldots,\tilde{\boldsymbol{\alpha}}_m^T,\tilde{\boldsymbol{\delta}}^T)$ be the IFM estimate, and $\hat{\boldsymbol{\eta}}^T = (\hat{\boldsymbol{\alpha}}_1^T,\ldots,\hat{\boldsymbol{\alpha}}_m^T,\hat{\boldsymbol{\delta}}^T)$ be the MLE.

A natural question to consider is the asymptotic relative efficiency (ARE) of $\tilde{\boldsymbol{\eta}}$ compared with $\hat{\boldsymbol{\eta}}$. This is studied in Sections 3 and 4. We first need to obtain a form for the asymptotic covariance matrix of $\tilde{\boldsymbol{\eta}}$, which can be derived using theory of inference functions [4], where the inference functions are in the left-hand side of (2.5). They can be written as

$$\sum_{i=1}^{n} \mathbf{g}(\mathbf{Y}_i; \boldsymbol{\eta}),$$

where $\mathbf{g}^T = (\mathbf{g}_1^T, \ldots, \mathbf{g}_m^T, \mathbf{g}_d^T)$, $\mathbf{g}_j = \partial\ell_j/\partial\boldsymbol{\alpha}_j$, $\ell_j = \log f_j(\cdot; \boldsymbol{\alpha}_j)$ for $j = 1, \ldots, m$; and $d = m + 1$ indexes the score equation $\mathbf{g}_d = \partial\ell/\partial\boldsymbol{\delta}$ for the multivariate log-likelihood and $\ell = \log f(\cdot; \boldsymbol{\eta})$. Note that the two-stage estimation is only for the computational implementation, not for the theoretical analysis.

Let $\mathcal{I} = \mathcal{I}(\boldsymbol{\eta})$ be the Fisher information matrix so that

$$n^{1/2}[\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}] \to_d N(\mathbf{0}, \mathcal{I}^{-1}), \quad n \to \infty.$$

From the theory of inference functions [4,7, Section 10.1.1],

$$n^{1/2}[\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}] \to_d N(\mathbf{0}, V), \quad n \to \infty,$$

where the asymptotic covariance matrix is given in (2.6) below.

The information matrix can be decomposed as

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}_{11} & \cdots & \mathcal{I}_{1m} & \mathcal{I}_{1d} \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{I}_{m1} & \cdots & \mathcal{I}_{mm} & \mathcal{I}_{md} \\ \mathcal{I}_{d1} & \cdots & \mathcal{I}_{dm} & \mathcal{I}_{dd} \end{pmatrix},$$

where $\mathcal{I}_{jk} = -\mathrm{E}\,[\partial^2\ell/\partial\boldsymbol{\alpha}_j\partial\boldsymbol{\alpha}_k^T]$ for $1 \leqslant j, k \leqslant m$, and $\mathcal{I}_{jd} = -\mathrm{E}\,[\partial^2\ell/\partial\boldsymbol{\alpha}_j\partial\boldsymbol{\delta}^T]$, $\mathcal{I}_{dj} = \mathcal{I}_{jd}^T$ for $j = 1, \ldots, m$. Let its inverse be denoted as $\mathcal{I}^{-1} = (\mathcal{I}^{(jk)})$.

The asymptotic covariance matrix for $\tilde{\boldsymbol{\eta}}$ is

$$V = (-D_{\mathbf{g}}^{-1})M_{\mathbf{g}}(-D_{\mathbf{g}}^{-1})', \tag{2.6}$$

where $M_{\mathbf{g}} = \mathrm{Cov}\,(\mathbf{g}(\mathbf{Y}; \boldsymbol{\eta})) = \mathrm{E}\,[\mathbf{g}\mathbf{g}^T]$ and $D_{\mathbf{g}} = \mathrm{E}\,[\partial\mathbf{g}(\mathbf{Y}, \boldsymbol{\eta})/\partial\boldsymbol{\eta}^T]$. Let $\mathcal{J}_{jk} = \mathrm{Cov}\,(\mathbf{g}_j, \mathbf{g}_k) = \mathrm{E}\,[\mathbf{g}_j\mathbf{g}_k^T]$ for $1 \leqslant j, k \leqslant m$, so that $\mathcal{J}_{jj}$ is the information matrix from the jth univariate log-likelihood. With this notation, the matrices, in partitioned form, become :

$$-D_{\mathbf{g}} = \begin{pmatrix} \mathcal{J}_{11} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathcal{J}_{mm} & 0 \\ \mathcal{I}_{d1} & \cdots & \mathcal{I}_{dm} & \mathcal{I}_{dd} \end{pmatrix}, \quad -D_{\mathbf{g}}^{-1} = \begin{pmatrix} \mathcal{J}_{11}^{-1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathcal{J}_{mm}^{-1} & 0 \\ a_1 & \cdots & a_m & \mathcal{I}_{dd}^{-1} \end{pmatrix},$$

where $a_j = -\mathcal{I}_{dd}^{-1}\mathcal{I}_{dj}\mathcal{J}_{jj}^{-1}$ for $j = 1, \dots, m$,

$$M_{\mathbf{g}}=\begin{pmatrix} \mathcal{J}_{11} & \cdots & \mathcal{J}_{1m} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{J}_{m1} & \cdots & \mathcal{J}_{mm} & 0 \\ 0 & \cdots & 0 & \mathcal{I}_{dd} \end{pmatrix}, \quad (-D_{\mathbf{g}}^{-1})M_{\mathbf{g}}=\begin{pmatrix} \mathcal{J}_{11}^{-1}\mathcal{J}_{11} & \cdots & \mathcal{J}_{11}^{-1}\mathcal{J}_{1m} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{J}_{mm}^{-1}\mathcal{J}_{m1} & \cdots & \mathcal{J}_{mm}^{-1}\mathcal{J}_{mm} & 0 \\ \sum_{j=1}^m a_j\mathcal{J}_{j1} & \cdots & \sum_{j=1}^m a_j\mathcal{J}_{jm} & I_d \end{pmatrix},$$

and I_d has the dimension of $\boldsymbol{\delta}$. Finally, V has (j, k) element $\mathcal{J}_{jj}^{-1}\mathcal{J}_{jk}\mathcal{J}_{kk}^{-1}$ for $1 \leqslant j, k \leqslant m$; (j, d) element $\mathcal{J}_{jj}^{-1}\sum_{k=1}^m \mathcal{J}_{jk}a_k^T$; (d, j) element $\sum_{k=1}^m a_k\mathcal{J}_{kj}\mathcal{J}_{jj}^{-1}$ for $j = 1, \dots, m$; (d, d) element $\mathcal{I}_{dd}^{-1} + \sum_{j=1}^m \sum_{k=1}^m a_j\mathcal{J}_{jk}a_k^T$. Note that the diagonal elements of V are $\mathcal{J}_{jj}^{-1}$, for $j = 1, \dots, m$, and these can also be obtained directly from ML theory for the univariate margins.

The derivation of V consists mostly of matrix manipulations, and the only non-trivial calculations are that $\mathrm{Cov}(\mathbf{g}_j, \mathbf{g}_d) = 0$ for $j = 1, \dots, d$. The proof of this calculation is given in the appendix.

Otherwise the entries in the partitioned matrices come from:

1. $\frac{\partial \mathbf{g}_j}{\partial \boldsymbol{\alpha}_j^T} = \frac{\partial^2 \ell}{\partial \boldsymbol{\alpha}_j \partial \boldsymbol{\alpha}_j^T}$, $-\mathrm{E}\,[\frac{\partial \mathbf{g}_j}{\partial \boldsymbol{\alpha}_j^T}] \stackrel{\text{def}}{=} \mathcal{J}_{jj}$, the Fisher information associated with the jth margin.
2. $\frac{\partial \mathbf{g}_j}{\partial \boldsymbol{\alpha}_k^T} = 0$, $j \neq k$, $\frac{f\mathbf{g}_j}{\partial \boldsymbol{\delta}^T} = 0$, $j, k = 1, \dots, m$.
3. $\frac{\partial \mathbf{g}_d}{\partial \boldsymbol{\alpha}_j^T} = \frac{\partial^2 \ell}{\partial \boldsymbol{\delta} \partial \boldsymbol{\alpha}_j^T}$, $-\mathrm{E}\,[\frac{\partial^2 \ell}{\partial \boldsymbol{\delta} \partial \boldsymbol{\alpha}_j^T}] = \mathcal{I}_{dj}$, $j = 1, \dots, m$.
4. $\frac{\partial \mathbf{g}_d}{\partial \boldsymbol{\delta}^T} = \frac{\partial^2 \ell}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T}$, $-\mathrm{E}\,[\frac{\partial^2 \ell}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T}] = \mathcal{I}_{dd}$.

From the above derived form for V, an estimated covariance matrix for $\tilde{\boldsymbol{\eta}}$ is $n^{-1}\tilde{V}$, where $\tilde{V}$ is a consistent estimate of V. For some models, with the help of symbolic manipulation software, it may be possible to obtain analytic forms for the derivatives in V and compute empirical versions of these derivatives. In Xu [23] and Joe [7], the form of V was not given, as $n^{-1}\tilde{V}$ was obtained using the (subset-delete) jackknife. With the jackknife and the use of a quasi-Newton method (e.g., [15]) for the two-stage estimation procedure, only the likelihoods (univariate and multivariate) need to be coded; even the inference functions in (2.4) need not be obtained.

3. Fréchet bounds and independence

The cases of extreme dependence and independence can provide us with an understanding of the efficiency of the IFM method. In the case of the Fréchet upper bound with continuous margins, sometimes the IFM estimator is exactly the same as the MLE, and sometimes it is different (in which case it is less efficient than the MLE in estimating the univariate parameters). Under some conditions, the efficiency of the IFM method is 1 in the limiting case of the independence copula.

For a Fréchet bound analysis, the $m = 2$ case provides insight that generalizes to higher dimensions. For the bivariate continuous case, suppose Y_{i1} are iid $F_1(\cdot; \boldsymbol{\alpha}_1)$ and Y_{i2} are iid $F_2(\cdot; \boldsymbol{\alpha}_2)$. Suppose (Y_{i1}, Y_{i2}) have a joint distribution of the Fréchet upper or lower bound.

The IFM estimates of $\boldsymbol{\alpha}_j$ come from separate likelihoods. The joint bivariate likelihood has the constraint of the functional relationship $Y_{i2} = h(Y_{i1}) = h(Y_{i1}; \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$, where

$$h(x) = F_2^{-1}(F_1(x; \boldsymbol{\alpha}_1); \boldsymbol{\alpha}_2) \tag{3.1}$$

or

$$h(x) = F_2^{-1}(1 - F_1(x; \boldsymbol{\alpha}_1); \boldsymbol{\alpha}_2), \tag{3.2}$$

respectively, for the Fréchet upper and lower bounds.

If α_1, α_2 are scalars, then $h(x_1; \alpha_1, \alpha_2)$ is either a two-parameter family or a one-parameter family of a function of a parameter θ which depends on α_1, α_2. If $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ are vectors, then h depends on a few functions of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$.

Consider the Fréchet upper bound. If $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ are identified from (3.1), then maximum likelihood of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ has no error and the IFM method is inefficient compared with ML. If $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ are not identified, then the IFM method may be the same as ML (see examples below). For the bivariate Fréchet lower bound, more generally $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ are identified from (3.2). Hence we can find examples where the asymptotic relative efficiency of estimating $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ with the IFM method is 0 or 1.

For the bivariate discrete Fréchet upper bound, the bivariate pmf consists of many zero probabilities away from the 'main diagonal'. However this is generally not perfect dependence (if $F_1 \neq F_2$), and the IFM relative efficiency is usually between 0 and 1.

We illustrate what can happen at the Fréchet bounds using a few examples. We start with a location or scale family for the univariate margins (α_j is a scalar location or scale parameter) and then extend to cases where $\boldsymbol{\alpha}_j$ is a vector such as with a generalized extreme value (GEV) margin, which is used for multivariate extreme value analysis.

Example 1. Scale family such as two exponential margins. Let $F_j(y) = F_0(y/\alpha_j)$ for a given F_0, for example, $F_0(z) = 1 - e^{-z}$, $z > 0$. For the Fréchet upper bound, $h(x) = \alpha_2 x/\alpha_1 = \theta x$, where $\theta = \alpha_2/\alpha_1$. The data (Y_{i1}, Y_{i2}), $i = 1, \ldots, n$, would lie on a straight line where only θ is known exactly, but not α_1, α_2. The ML and IFM estimators of α_1, α_2 are the MLEs $\hat{\alpha}_1, \hat{\alpha}_2$ from individual univariate likelihoods, since $\hat{\alpha}_2/\hat{\alpha}_1 = \theta$. For example, $\hat{\alpha}_j = \overline{Y}_j = n^{-1}\sum_{i=1}^n Y_{ij}$, $j = 1, 2$, for exponential margins.

For the Fréchet lower bound, generally α_1, α_2 are identifiable from curve (3.2). For example, with exponential margins, $h(x) = -\alpha_2 \log(1 - e^{-x/\alpha_1})$. The data (Y_{i1}, Y_{i2}) would lie on this curve, and the parameters of this families of curves are identifiable from $n \geqslant 2$ points.

Let $\lambda_j = \alpha_j^{-1}$. We outline a proof that $\lambda_2^{-1}\log(1 - e^{-\lambda_1 x})$ and $\beta_2^{-1}\log(1 - e^{-\beta_1 x})$ are distinct for $(\lambda_1, \lambda_2) \neq (\beta_1, \beta_2)$. If the two curves are equal, then for all $x \geqslant 0$,

$$\frac{\log(1 - e^{-\lambda_1 x})}{\lambda_2} = \frac{\log(1 - e^{-\beta_1 x})}{\beta_2}. \tag{3.3}$$

Equal derivatives imply

$$\frac{\lambda_1}{\lambda_2[e^{\lambda_1 x} - 1]} = \frac{\beta_1}{\beta_2[e^{\beta_1 x} - 1]}. \tag{3.4}$$

Hence

$$\lambda_2^{-1} = \lim_{x\to 0} \frac{x\lambda_1}{\lambda_2[e^{\lambda_1 x} - 1]} = \lim_{x\to 0} \frac{x\beta_1}{\beta_2[e^{\beta_1 x} - 1]} = \beta_2^{-1}.$$

Substitution of this equality into (3.3) leads to $\lambda_1 = \beta_1$.

Example 2. *Location family*: Let $F_j(y) = F_0(y - \alpha_j)$ for a given F_0 with median 0. With the Fréchet upper bound, $h(x) = x - \alpha_1 + \alpha_2 = x - \theta$, where $\theta = \alpha_1 - \alpha_2$. As in Example 1, the data (Y_{i1}, Y_{i2}), $i = 1, \dots, n$, would lie on a straight line where only θ is known exactly, but not α_1, α_2. The ML and IFM estimators of α_1, α_2 are the MLEs $\hat{\alpha}_1, \hat{\alpha}_2$ from individual univariate likelihoods, since $\hat{\alpha}_1 - \hat{\alpha}_2 = \theta$.

For the Fréchet lower bound, (3.2) becomes $h(x) = F_0^{-1}(1 - F_0(x - \alpha_1)) + \alpha_2$. If F_0 is symmetric about 0, then $h(x) = F_0^{-1}(F_0(\alpha_1 - x)) + \alpha_2 = \alpha_1 + \alpha_2 - x$, and (Y_{i1}, Y_{i2}), $i = 1, \dots, n$, would lie on a straight line, where $\theta = \alpha_1 + \alpha_2$ is known but not α_1, α_2. If F_0 is not symmetric about 0, then α_1, α_2 will generally be identifiable from curve (3.2). An analysis similar to that for exponential margins in Example 1 would have to be checked for special cases of F_0.

Example 3. *Location-scale family*: Suppose $F_j(y) = F_0((y - \mu_j)/\sigma_j)$ for a given F_0 and $\boldsymbol{\alpha}_j = (\mu_j, \sigma_j)^T$. For the Fréchet upper bound, $h(x) = \mu_2 + \sigma_2[(x - \mu_1)/\sigma_1] = a + bx$, where $a = \mu_2 - \sigma_2\mu_1/\sigma_1$, $b = \sigma_2/\sigma_1$. Hence the observed data (Y_{i1}, Y_{i2}) would satisfy $Y_{i2} = a + bY_{i1}$ for some constants a, b, and then $\hat{\mu}_2 = a + b\hat{\mu}_1$, $\hat{\sigma}_2 = b\hat{\sigma}_1$, so that the ML and IFM estimators are the same.

For the Fréchet lower bound, $h(x) = \mu_2 + \sigma_2 F_0^{-1}(1 - F_0([x - \mu_1]/\sigma_1))$. This can be a parametric family of functions with 2 to 4 parameters. If F_0 is symmetric about zero, then for (3.2), $h(x) = \mu_2 + \sigma_2(\mu_1 - x)/\sigma_1$ and also $\sigma_1(Y_{i2} - \mu_2) = \sigma_2(\mu_1 - Y_{i1})$ or $\sigma_1 Y_{i2} + \sigma_2 Y_{i1} = \sigma_2\mu_1 + \sigma_1\mu_2$. That is, (Y_{i1}, Y_{i2}) lie on a straight line, $Y_{i2} = (\sigma_2\mu_1 + \sigma_1\mu_2)/\sigma_1 - \sigma_2 Y_{i1}/\sigma_1$, and σ_2/σ_1 and $(\sigma_2\mu_1 + \sigma_1\mu_2)/\sigma_1$ are known. The IFM estimates must satisfy $\tilde{\mu}_2 = (\sigma_2\mu_1 + \sigma_1\mu_2)/\sigma_1 - \sigma_2\tilde{\mu}_1/\sigma_1$, $\tilde{\sigma}_2 = \sigma_2\tilde{\sigma}_1/\sigma_1$, so that the ML and IFM estimators are the same. More generally, if F_0 is not symmetric about zero, the ML and IFM estimators are different.

Example 4. *Families with shape parameters*: Some examples with scalar α_j are (a) Pareto distributions $F(y; \alpha_j) = 1 - (1 + y)^{-\alpha_j}$, $x > 0$, for which (3.1) becomes $h(x) = (1 + x)^{\alpha_1/\alpha_2} - 1$; (b) Weibull distributions $F(y; \alpha_j) = 1 - \exp\{-y^{\alpha_j}\}$, $x > 0$, for which (3.1) becomes $h(x) = x^{\alpha_1/\alpha_2}$; (c) Gamma($\alpha_j$,1) distributions for which $h(x)$ in (3.1) does not reduce to a one-parameter family. Hence for the one-parameter Pareto and Weibull margins, IFM and ML estimators of α_1, α_2 are the same with the Fréchet upper bound, but for Gamma margins, the IFM estimators $\tilde{\alpha}_j$ would have zero asymptotic efficiency for the Fréchet upper bound.

For the Fréchet lower bound, α_1, α_2 are identified from $h(x)$ in (3.2) for the Pareto, Weibull and Gamma margins (analysis is similar to that in Example 1).

Example 5. *GEV univariate margins and the Fréchet upper bound*: Since multivariate extreme value analysis (GEV margins) has been one area where two-stage estimation has been used in the past, we provide some details for this case.

With $(z)_+ = \max\{0, z\}$, let $F_j(y) = \exp\{-(1 + \gamma_j[y - \mu_j]/\sigma_j)_+^{-1/\gamma_j}\}$, $F_j^{-1}(u) = \mu_j + \sigma_j[(-\log u)^{-\gamma_j} - 1]/\gamma_j$, $-\infty < \gamma_j < \infty$, $-\infty < \mu_j < \infty$, $\sigma_j > 0$, $j = 1, 2$, with the limiting case of $\gamma_j \to 0$ corresponding to $\exp\{-\exp[-(y - \mu_j)/\sigma_j]\}$. We assume $\gamma_1, \gamma_2 \neq 0$ below; the analysis of this case follows from a limit.

For the Fréchet upper bound, from (3.1),

$$\begin{aligned} h(x) &= \mu_2 + \sigma_2\gamma_2^{-1}\big\{(1 + \gamma_1[x - \mu_1]/\sigma_1)_+^{\gamma_2/\gamma_1} - 1\big\} \\ &= (\mu_2 - \sigma_2\gamma_2^{-1}) + \sigma_2\gamma_2^{-1}\big(1 - \gamma_1\mu_1\sigma_1^{-1} + \gamma_1 x\sigma_1^{-1}\big)_+^{\gamma_2/\gamma_1} \\ &= (\mu_2 - \sigma_2\gamma_2^{-1}) + \Big[(1 - \gamma_1\mu_1\sigma_1^{-1})(\sigma_2/\gamma_2)^{\gamma_1/\gamma_2} + (\sigma_2/\gamma_2)^{\gamma_1/\gamma_2}\gamma_1 x\sigma_1^{-1}\Big]_+^{\gamma_2/\gamma_1} \end{aligned}$$

This is a 4-parameter family of curves if $\gamma_1 \neq \gamma_2$. Note that for $a > 0$, the parameter transforms $\gamma_1 \to a\gamma_1$, $\gamma_2 \to a\gamma_2$, $\sigma_1 \to a\sigma_1$, $\sigma_2 \to a\sigma_2$ leads to the same curve.

If $\gamma_1 = \gamma_2$, the above curve is linear (2-parameter family) over the support of F_1. Because (Y_{i1}, Y_{i2}) would lie on the straight in this case, the ML and IFM estimators are the same (see Example 3).

For $\gamma_1 \neq \gamma_2$, reparametrize with $\lambda_j = \sigma_j/\gamma_j$; λ_j has the same sign as γ_j. Then, for (3.1),

$$h(x) = \begin{cases} \mu_2 - \lambda_2 + \lambda_2\lambda_1^{-\gamma_2/\gamma_1}[x - (\mu_1 - \lambda_1)]_+^{\gamma_2/\gamma_1}, & \gamma_1/\gamma_2 > 0, \\ \mu_2 - \lambda_2 + |\lambda_2||\lambda_1|^{-\gamma_2/\gamma_1}[(\mu_1 - \lambda_1) - x]_+^{\gamma_2/\gamma_1}, & \gamma_1/\gamma_2 < 0 \end{cases} \tag{3.5}$$

and this is a 4-parameter family of curves with parameters $a_1 = \mu_1 - \lambda_1$, $a_2 = \mu_2 - \lambda_2$, $b = |\lambda_2||\lambda_1|^{-\gamma_2/\gamma_1}$, $d = \gamma_2/\gamma_1$. These parameters can be identified with $n \geqslant 4$ points on the curve. Let $W_{ij} = Y_{ij} - a_j$ with cdf

$$G_j(w) = \exp\{-(w/\lambda_j)_+^{-1/\gamma_j}\} = \exp\{-\beta_j|w|^{-1/\gamma_j}\},$$

where $\beta_j = |\lambda_j|^{1/|\gamma_j|}$ and $w > 0$ for $\gamma_j > 0$, $w < 0$ for $\gamma_j < 0$. That is, if γ_j is positive, then the support of G_j is $(0, \infty)$, and if γ_j is negative, then the support of G_j is $(-\infty, 0)$. G_j is a Fréchet and Weibull extreme value distribution, respectively, in the two cases. The sign of γ_j will be known from the sign of Y_{ij} after a_1, a_2 are deduced. If W has a Weibull distribution on $(-\infty, 0)$ with parameter $\gamma < 0$, then $-W^{-1}$ has a Fréchet distribution on $(0, \infty)$ with parameter $-\gamma > 0$, since

$$\Pr(-(W)^{-1} \leqslant z) = \Pr(W \leqslant -z^{-1}) = \exp\{-\beta z^{1/\gamma}\}, \quad z > 0.$$

Hence we now continue our analysis assuming $\gamma_1, \gamma_2 > 0$.

With the parameter transform to $\beta_j, b = (\beta_2/\beta_1)^{\gamma_2}$. If the data are $\{W_{ij}\}$, for $j = 1, 2$, the univariate ML estimates based on W_{ij} maximized the univariate likelihoods with parameters β_j, γ_j. The density of W_{ij} is

$$g_j(w) = \gamma_j^{-1} w^{-1-1/\gamma_j}\beta_j \exp\{-\beta_j w^{-1/\gamma_j}\}$$

and the log-likelihood is

$$L_j = n \log \beta_j - n \log \gamma_j - (1+\gamma_j^{-1}) \sum_i \log w_{ij} - \beta \sum_i w_{ij}^{-1/\gamma_j}.$$

Solving the derivative equations lead to

$$\begin{aligned}\frac{\partial L_j}{\partial \beta_j} &= n\beta_j^{-1} - \sum w_i^{-1/\gamma_j} = 0,\\ \frac{\partial L_j}{\partial \gamma_j} &= -n\gamma_j^{-1} + \gamma_j^{-2} \sum \log w_{ij}\\ &\qquad -\beta_j \gamma_j^{-2} \sum w_{ij}^{-1/\gamma_j} \log w_{ij} = 0,\end{aligned}$$

$$\hat\beta_j = n \bigg/ \sum w_{ij}^{-1/\hat\gamma_j}\ , \tag{3.6}$$

$$-n\hat\gamma_j + \sum \log w_{ij} - \hat\beta \sum w_{ij}^{-1/\hat\gamma_j} \log w_{ij} = 0,$$

$$\hat\gamma_j = n^{-1} \sum \log w_{ij} - \sum w_{ij}^{-1/\hat\gamma_j} \log w_{ij} \bigg/ \sum w_{ij}^{-1/\hat\gamma_j}\ . \tag{3.7}$$

From (3.5), $w_{i2} = bw_{i1}^d$, and substituting in (3.7) leads to

$$\begin{aligned}\hat\gamma_2 &= n^{-1} \sum \log w_{i2} - \sum w_{i2}^{-1/\hat\gamma_2} \log w_{i2} \bigg/ \sum w_{i2}^{-1/\hat\gamma_2}\\ &= dn^{-1} \sum \log w_{i1} + \log b - \sum w_{i1}^{-d/\hat\gamma_2} [d \log w_{i1} + \log b] \bigg/ \sum w_{i1}^{-d/\hat\gamma_2}\\ &= dn^{-1} \sum \log w_{i1} - d \sum w_{i1}^{-d/\hat\gamma_2} \log w_{i1} \bigg/ \sum w_{i1}^{-d/\hat\gamma_2}\ .\end{aligned}$$

From (3.7), $\hat\gamma_2 = d\hat\gamma_1$ and hence $\hat\gamma_2/\hat\gamma_1 = d$. Substituting in (3.6) leads to

$$\begin{aligned}\hat\beta_2 &= n \bigg/ \sum w_{i2}^{-1/\hat\gamma_2} = nb^{1/\hat\gamma_2} \bigg/ \sum w_{i1}^{-d/\hat\gamma_2}\\ &= n(\beta_2/\beta_1)^{\gamma_2/\hat\gamma_2} \bigg/ \sum w_{i1}^{-1/\hat\gamma_1} = (\beta_2/\beta_1)^{\gamma_2/\hat\gamma_2} \hat\beta_1\ .\end{aligned}$$

Hence $(\hat\beta_2/\hat\beta_1)^{\hat\gamma_2} = (\beta_2/\beta_1)^{\gamma_2} = b$ and the estimates $\hat\beta_1, \hat\beta_2, \hat\gamma_1, \hat\gamma_2$ satisfy the constraints from the curve h.

These estimates with the transformed data correspond to MLE of $\{Y_{ij}\}$ with the Fréchet upper bound. For the IFM estimators, a_j and W_{ij}, $j = 1, 2$, would not be known. Hence the IFM estimators are less efficient than MLEs for the Fréchet upper bound.

The analyses in the above examples extend to $m > 2$. For the Fréchet upper bound with $m > 2$, each bivariate margin is a bivariate Fréchet upper bound, so the above results apply. In general, there is no Fréchet lower bound for $m > 2$.

Next we study the information matrix $\mathcal{I}$ in the limiting case of an independence copula. Let $\boldsymbol{\delta}_I$ be the parameter value in (1.1) for the independence copula. As $\boldsymbol{\delta} \to \boldsymbol{\delta}_I$, then $\mathcal{I}_{jj} \to \mathcal{J}_{jj}$ and $\mathcal{I}_{jk} \to 0$, $j \neq k$, $j, k = 1, \ldots, m$, under some regularity conditions. From this, it follows that a necessary condition for $V - \mathcal{I}^{-1} \to 0$ as $\boldsymbol{\delta} \to \boldsymbol{\delta}_I$ is for $\mathcal{I}_{jd} \to 0$ for $j = 1, \ldots, m$, and this condition is proved below. Note that for many useful copula families which do not extend to negative dependence (for example, multivariate extreme value distributions have positive dependence only), $\boldsymbol{\delta}_I$ is on the boundary of the parameter space and hence asymptotic ML theory would not apply for $\boldsymbol{\delta} = \boldsymbol{\delta}_I$.

Theorem. *As $\boldsymbol{\delta} \to \boldsymbol{\delta}_I$, under the usual regularity conditions for maximum likelihood, $V - \mathcal{I}^{-1} \to 0$. That is, the covariance matrix for the IFM estimator becomes the same as the covariance matrix of the MLE when the independence copula is approached.*

Proof. For the continuous case, from (2.1),

$$\mathcal{I}_{jk} = -\mathrm{E}\left[\frac{\partial^2\{\log c(F_1(Y_1;\boldsymbol{\alpha}_1), \ldots, F_m(Y_m;\boldsymbol{\alpha}_m);\boldsymbol{\delta}) + \sum \log f_j(Y_j;\boldsymbol{\alpha}_j)\}}{\partial\boldsymbol{\alpha}_j\partial\boldsymbol{\alpha}_k^T}\right].$$

Assuming that taking the limit as $\boldsymbol{\delta} \to \boldsymbol{\delta}_I$ and differentiation with respect to $\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_k$ can be interchanged, then $c \to 1$ as $\boldsymbol{\delta} \to \boldsymbol{\delta}_I$, and $\mathcal{I}_{jk} = 0$ for $j \neq k$ and $\mathcal{I}_{jj} = \mathcal{J}_{jj}$ for $j = 1, \ldots, m$.

For the discrete case, letting $I(\mathbf{y})$ be an indicator function, then

$$\mathcal{I}_{jk} = -\mathrm{E}\left[\frac{\partial^2 \ell}{\partial\boldsymbol{\alpha}_j\partial\boldsymbol{\alpha}_k^T}\right],$$

where

$$\ell = \sum_y I(\mathbf{y}) \log f(\mathbf{y};\boldsymbol{\eta}) \to \sum_y I(\mathbf{y})[\log f_1(y_1;\boldsymbol{\alpha}_1) + \cdots + \log f_m(y_m;\boldsymbol{\alpha}_m)],$$

as $\boldsymbol{\delta} \to \boldsymbol{\delta}_I$, so that $\mathcal{I}_{jk} \to 0$ for $j \neq k$ and $\mathcal{I}_{jj} \to \mathcal{J}_{jj}$ for $j = 1, \ldots, m$ (assuming the limit and differentiation can be interchanged).

Similarly, $\mathcal{J}_{jk} \to 0$ for $j \neq k$. From the matrices in (2.6), if $\mathcal{I}_{jd} \to 0$ for $j = 1, \ldots, m$, then $-\boldsymbol{D}_{\mathbf{g}} - \mathcal{I} \to 0$ and $\boldsymbol{M}_{\mathbf{g}} - \mathcal{I} \to 0$, and $V - \mathcal{I}^{-1} \to 0$.

Finally, we show that $\mathcal{I}_{jd} \to 0$ as $\boldsymbol{\delta} \to \boldsymbol{\delta}_I$, $j = 1, \ldots, m$. Under regularity conditions,

$$\mathcal{I}_{jd} = -\mathrm{E}\left[\frac{\partial^2 \ell}{\partial\boldsymbol{\alpha}_j\partial\boldsymbol{\delta}}\right] = \mathrm{E}\left[\frac{\partial \ell}{\partial\boldsymbol{\alpha}_j}\,\frac{\partial \ell}{\partial\boldsymbol{\delta}^T}\right] \overset{\boldsymbol{\delta}\to\boldsymbol{\delta}_I}{\to} \mathrm{E}\left[\frac{\partial \ell_j}{\partial\boldsymbol{\alpha}_j}\,\frac{\partial \ell}{\partial\boldsymbol{\delta}^T}\right] = \mathrm{E}\,[\mathbf{g}_j\mathbf{g}_d^T] = 0.$$

The last equality is shown in the appendix, and the limit step is similar to the above. □

4. Efficiency

In this section, we use results from the preceding sections and study asymptotic relative efficiencies (AREs) of $\tilde{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\eta}}$ based on V and $\mathcal{I}^{-1}$. The bivariate continuous case with scalar quantities $\alpha_1, \alpha_2, \delta$ is used to show some patterns in the efficiency of the IFM estimator and link together some results from Sections 2 and 3. In this case, we get a simpler form for the inverse of Fisher information, and the quantities in V and $\mathcal{I}$ can be more readily evaluated numerically.

We proceed with some analyses in the bivariate continuous case with scalar parameters to indicate the behavior of terms in the Fisher information matrix as the Fréchet bounds are approached. For α_1, α_2 being scalars, the asymptotic variance of $n^{1/2}(\hat{\alpha}_j - \alpha_j)$ is $\mathcal{I}^{(jj)}$ and the asymptotic variance of $n^{1/2}(\hat{\alpha}_j - \alpha_j)$ is $\mathcal{J}_{jj}^{-1}$. From (2.1), $\log f(\mathbf{y}; \alpha_1, \alpha_2, \delta) = \log c(F_1(y_1; \alpha_1), F_2(y_2; \alpha_2); \delta) + \log f_1(y_1; \alpha_1) + \log f_2(y_2; \alpha_2)$ so that $\mathcal{I}_{jj} = \mathcal{J}_{jj} + \zeta_{jj}$, $j = 1, 2$, where

$$\zeta_{jj} = -\mathrm{E}\left\{\frac{\partial^2 \log c(F_1, F_2; \delta)}{\partial \alpha_j^2}\right\}.$$

To understand the behavior of the asymptotic covariance matrix (2.6) for the IFM estimator, and the behavior of ζ_{jj}, we consider the efficiency of estimating α_2 when α_1 is known. In this subproblem, the relevant information matrix for is

$$\mathcal{I}^* = \begin{pmatrix} \mathcal{I}_{22} & \mathcal{I}_{2d} \\ \mathcal{I}_{d2} & \mathcal{I}_{dd} \end{pmatrix},$$

so asymptotic covariance matrix for the MLE $(\hat{\alpha}_2, \hat{\delta})$ is

$$(\mathcal{I}^*)^{-1} = \frac{1}{\mathcal{I}_{22}\mathcal{I}_{dd} - \mathcal{I}_{2d}^2} \begin{pmatrix} \mathcal{I}_{dd} & -\mathcal{I}_{2d} \\ -\mathcal{I}_{d2} & \mathcal{I}_{22} \end{pmatrix}$$

For the Fréchet bounds, α_2 in (3.1) or (3.2) can be identified from the curve, so that α_2 can be estimated without error (the IFM efficiency goes to 0), and

$$\begin{aligned} \mathcal{I}_{dd}/(\mathcal{I}_{22}\mathcal{I}_{dd} - \mathcal{I}_{2d}^2) &= \mathcal{I}_{dd}/([\mathcal{J}_{22} + \zeta_{22}]\mathcal{I}_{dd} - \mathcal{I}_{2d}^2) \\ &= [\mathcal{J}_{22} + \zeta_{22} - \mathcal{I}_{2d}^2/\mathcal{I}_{dd}]^{-1} \to 0, \end{aligned} \tag{4.1}$$

as $\delta \to \delta_U$ or δ_L, the parameter values for the Fréchet upper and lower bounds. $\mathcal{J}_{22}$ does not change with δ, so that from (4.1), $\zeta_{22} - \mathcal{I}_{2d}^2/\mathcal{I}_{dd} \to \infty$ as $\delta \to \delta_U$ or δ_L. Hence $\zeta_{22} \to \infty$ for the Fréchet bounds, and by symmetry $\zeta_{11} \to \infty$ also. This means that $\mathcal{I}_{11} \to \infty$ and $\mathcal{I}_{22} \to \infty$ for the Fréchet bounds.

Now back to estimating α_1, α_2 with both unknown. With the notation $\mathcal{I}^{-1} = (\mathcal{I}^{(jk)})$, $\mathcal{I}^{(22)} = [\mathcal{I}_{11}\mathcal{I}_{dd} - \mathcal{I}_{1d}^2]/\det(\mathcal{I})$; after some algebra we get

$$\mathcal{I}^{(22)} = \left[\mathcal{J}_{22} + \zeta_{22} - \frac{\mathcal{I}_{12}^2 + \mathcal{I}_{11}\mathcal{I}_{2d}^2/\mathcal{I}_{dd} - 2\mathcal{I}_{12}\mathcal{I}_{1d}\mathcal{I}_{2d}/\mathcal{I}_{dd}}{\mathcal{I}_{11} - \mathcal{I}_{1d}^2/\mathcal{I}_{dd}}\right]^{-1}. \tag{4.2}$$

We consider two cases for the behavior of $\mathcal{I}_{jk}$ to explain the possible form of the MLE of α_1, α_2 for the Fréchet bounds, when $\mathcal{I}_{1d}, \mathcal{I}_{2d}, \mathcal{I}_{dd}$ remain bounded. First, assume $|\mathcal{I}_{12}| \to \infty$ as $\delta \to \delta_U$ or δ_L at a rate comparable to ζ_{11}, ζ_{22}. Then from (4.2),

$$\mathcal{I}^{(22)} \approx [\mathcal{J}_{22} + \zeta_{22} - \mathcal{I}_{12}^2/\mathcal{I}_{11}]^{-1}$$

for δ close to the Fréchet bound parameters. With $\zeta_{22}, |\mathcal{I}_{12}| \to \infty$, it is possible for $\zeta_{22} - \mathcal{I}_{12}^2/\mathcal{I}_{11}$ to approach ∞, a positive constant or 0; in the latter case, IFM and ML estimators become equivalent for the Fréchet bound and otherwise the IFM estimator is less efficient. Next, assume that $|\mathcal{I}_{12}|$ remains bounded. Then from (4.2),

$$\mathcal{I}^{(22)} \approx [\mathcal{J}_{22} + \zeta_{22} - \mathcal{I}_{2d}^2/\mathcal{I}_{dd}]^{-1} \to 0.$$

Note that $\mathcal{I}^{(22)} \to 0$ as $\delta \to \delta_U$ or δ_L corresponds to the case where the univariate distributions can be determined when observing data satisfying the Fréchet bound (see the discussion in Section 3).

Next we discuss specific examples where the Fisher information $\mathcal{I}$ and the matrix V in (2.6) have been numerically compared. In the numerical computations, we noted that behavior of $\mathcal{I}_{jk}, \zeta_{jj}$ and $\mathcal{I}^{(jj)}$ referred to above.

Consider a bivariate family

$$C(F_1(y_1, \alpha_1), F_2(y_2; \alpha_2); \delta) \tag{4.3}$$

which includes the independence copula and the Fréchet upper bound, and is increasing in dependence as the scalar parameter δ increases. Assuming some regularity conditions, from the results in Section 3, the pattern in the efficiency of the IFM estimator for α_1, α_2 that we might expect to see are: (a) efficiency of 1 for $\delta \to \delta_I$, (b) efficiency decreases as δ increases from δ_I, (c) efficiency continues to decrease as δ increases to δ_U, in the case where the parameters α_1, α_2 are identifiable from (3.1), or efficiency reaches a minimum for some δ^* and then efficiency increases towards 1, in the case the parameters α_1, α_2 are not identifiable from (3.1) and the IFM and ML estimators are the same in the limit. If furthermore, the bivariate family has negative dependence and extends to the Fréchet lower bound, we might expect to see: (d) efficiency decreases as δ decreases from δ_I and reaching an efficiency of 0 as $\delta \to \delta_L$, or efficiency reaches a minimum for some $\delta^\dagger$ and then efficiency increases towards 1, depending on whether α_1, α_2 are identifiable from (3.2). These are the patterns in the efficiency that we see in all numerical examples for which we have done the computations.

In Table 1, some numerical results to illustrate this is given for (4.3) for the Plackett [17] copula $C(u_1, u_2; \delta) = \frac{1}{2}(\delta - 1)^{-1}\{1 + (\delta - 1)(u_1 + u_2) - [(1 + (\delta - 1)(u_1 + u_2))^2 - 4\delta(\delta - 1)u_1u_2]^{1/2}\}$ ($\delta \to 1$ for independence, $\delta \to \infty$ for the Fréchet upper bound, $\delta \to 0$ for the Fréchet lower bound) and exponential margins with mean parameters α_1, α_2 (because of scale invariance, the results are invariant to the values of α_1, α_2). In Table 1, the δ values correspond to those that yield Kendall tau values from -0.9 to 0.9 in steps of 0.1 (see [7, Table 5.1]), and the ARE values are from the ratio of diagonal elements of $\mathcal{I}^{-1}$ to corresponding diagonal elements of V. In the case for positive dependence, the efficiency for the IFM estimator $\tilde{\alpha}_j$ decreases up to a Kendall tau value of around 0.8 ($\delta = 44$) and then increases. Note that the IFM estimator $\tilde{\delta}$ has very high efficiency; this is also typical of other examples that were studied.

For model (4.3) with continuous margins and $\alpha_1, \alpha_2, \delta$ being scalars, the elements of $\mathcal{I}$ and V are straightforward to compute using two-dimensional numerical integration [2]. For the integrations, the following derivatives of the copula density $c(u_1, u_2; \delta) = \partial^2 C(u_1, u_2; \delta)/\partial u_1 \partial u_2$ are needed: $\partial c/\partial u_j$, $\partial c/\partial\delta$, $\partial^2 c/\partial u_j^2$, $\partial^2 c/\partial u_1\partial u_2$, $\partial^2 c/\partial\delta^2$ and $\partial^2 c/\partial u_j\partial\delta$. With C or c in closed form, these derivatives are easily obtained using symbolic manipulation software. We tried computations with a number of one-parameter family of copulas (see

Table 1
ARE of IFM estimators for Plackett copula, exponential margins

δ	ARE($\tilde{\alpha}_j$)	ARE($\tilde{\delta}$)
0.002	0.064	0.996
0.009	0.238	0.995
0.023	0.444	0.995
0.047	0.621	0.996
0.088	0.753	0.997
0.15	0.852	0.998
0.25	0.921	0.999
0.40	0.968	0.999+
0.64	0.993	0.999+
1.00	1.000	1.000
1.57	0.994	0.999+
2.48	0.978	0.999+
4.00	0.955	0.999+
6.60	0.930	0.999
11.4	0.909	0.999
21.1	0.896	0.999
44.1	0.893	0.999
115	0.895	0.998
530	0.898	0.998

[7, Section 5.1]) that include independence and the Fréchet upper bound, and a number of different univariate one-parameter families: $N(\alpha, 1)$, exponential(α), Weibull(α), Gumbel or extreme value distribution with location parameter α. The pattern is similar to the above table, but of course the efficiency depends on the copula family and the univariate margins. The turning point where efficiency of $\tilde{\alpha}_j$ increases again sometimes corresponds to a Kendall tau value which is closer to 1 than 0.8.

The combination with the worse case of efficiency of $\tilde{\alpha}_j$ for positive dependence, among those mentioned above, is for the Frank [3] copula with exponential margins. In this case, ARE($\tilde{\alpha}_j$) reached 0.57 for δ corresponding to a Kendall tau value of 0.9.

The next Table 2 is based on an artificial model. It is used to compare continuous and discrete univariate margins. The copula is the Frank [3] copula $C(u_1, u_2; \delta) = -\delta^{-1}\log([(1-e^{-\delta}) - (1-e^{-\delta u_1})(1-e^{-\delta u_2})]/(1-e^{-\delta}))$ ($\delta \to 0$ for independence, $\delta \to \infty$ for the Fréchet upper bound) and the univariate margin is $F_j(y; \alpha_j) = 1 - \exp\{-\tau_y/\alpha_j\}$, $\tau_y = -\log(1 - y/(r+1))$, $y = 1, \ldots, r$; this corresponds to discretized exponential with $r+1$ ordered categories. The δ values in Table 2 correspond to Kendall tau values of 0.2, 0.5, 0.7 and 0.9. Note that the efficiency loss for the IFM method slowly worsens as the number of categories increases.

The next example is for bivariate ordinal probit or discretized bivariate normal, where the number of univariate parameters is one less than the number of categories r_j. If the cutpoints or threshold parameters are $\boldsymbol{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{1,r_1-1})$, and $\boldsymbol{\alpha}_2 = (\alpha_{21}, \ldots, \alpha_{2,r_2-1})$ and $\alpha_{j0} = -\infty$ and $\alpha_{j,r_j} = \infty$, the bivariate ordinal probit model has

Table 2
ARE of IFM estimators for Frank copula with discretized exponential margins in r categories, $\alpha_1 = \alpha_2 = 1$

r	δ	ARE($\tilde{\alpha}$)	ARE($\tilde{\delta}$)
2	1.86	0.988	0.999+
3	1.86	0.984	0.999+
5	1.86	0.981	0.999+
10	1.86	0.979	0.999+
20	1.86	0.978	0.999+
2	5.74	0.941	0.999+
3	5.74	0.916	0.999
5	5.74	0.898	0.998
10	5.74	0.888	0.997
20	5.74	0.884	0.997
2	11.4	0.926	0.999+
3	11.4	0.877	0.998
5	11.4	0.825	0.996
10	11.4	0.786	0.994
20	11.4	0.773	0.994
2	20.9	0.950	0.999+
3	20.9	0.903	0.999+
5	20.9	0.826	0.998
10	20.9	0.730	0.995
20	20.9	0.683	0.994

probabilities:

$$\Pr(Y_1 = y_1, Y_2 = y_2) = \Pr(\alpha_{1,y_1-1} < Y_1 \leqslant \alpha_{1,y_1},\ \alpha_{2,y_2-1} < Y_2 \leqslant \alpha_{2,y_2}),$$
$$(Y_1, Y_2) \sim N(\mathbf{0}, R), \quad R = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}. \tag{4.4}$$

Note that this is a copula-based model with the bivariate normal copula $\Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \delta)$ and univariate cdfs $F_j(y) = \Phi(\alpha_{j,y})$ for $y = 1, \ldots, r_j$, where Φ is the univariate standard normal cdf and Φ_2 is the bivariate standard normal cdf with correlation δ. The two-stage estimation method for this model was studied in Olsson [16], but without any comparison of relative efficiency.

Table 3 has some results on asymptotic relative efficiency for this model. Given some $\boldsymbol{\alpha}_j$ vectors, the δ value listed is the positive value leading to the smallest ARE. The computations of $\mathcal{I}$ and V are not difficult because of the discreteness and because the derivatives have simple forms. With the range of r and parameters commonly encountered for this model, two-stage estimation is highly efficient. There is a slow worsening of efficiency of the univariate parameters as r_1, r_2 increase. Similar to Table 1, the efficiency of $\tilde{\delta}$ is high (above 0.98 in all numerical cases that were computed).

As the dimension m increases, intuitively one would expect the ARE of the IFM method to worsen. The next example gives an indication of amount of the decrease in efficiency as m increases. We use the multivariate exchangeable binary probit model (multivariate

Table 3
Minimum AREs for the bivariate ordinal probit model

$r_1 = r_2$	$\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$	δ	min ARE$(\tilde{\alpha}_{j\ell})$
3	$(-.5, 0.3)$	.89	.9907
5	$(-1, -.5, .2, .9)$	.92	.9500
7	$(-1, -.5, -.1, 0.2, .7, 1.2)$	.96	.9157

Table 4
Constrained minimum AREs for the multivariate exchangeable binary probit model, $\rho = 0.9$

m	$\boldsymbol{\alpha}$	min ARE$(\tilde{\alpha}_m)$
3	(.225, .058, .01)	.979
4	(.714, .714, .030, .01)	.938
5	(.657, .657, .033, .033, .01)	.893
6	(.693, .693, .693, .034, .034, .01)	.885
7	(.673, .673, .673, .034, .034, .034, .01)	.857
8	(.794, .794, .598, .598, .035, .035, .035, .01)	.851

version of (4.4) with $r_j = 2$ for all $j = 1, \ldots, m$, correlation matrix $R = (\rho_{jk})$ with $\rho_{jk} = \rho \geqslant 0$ for all $j \neq k$). We parameterize the model so that the univariate parameters are Bernoulli parameters α_j (thresholds $\Phi^{-1}(\alpha_j)$) and the multivariate dependence parameter is $\delta = \rho$. For this model, computations are numerically straightforward because the rectangle probabilities and their derivatives with respect to α_j, ρ can be reduced to 1-dimensional numerical integrals. For a fixed $(\alpha_1, \ldots, \alpha_m)$, the minimum ARE of $\tilde{\alpha}_j$ typically occurs for ρ in the interval (0.9, 1). With $(\alpha_1, \ldots, \alpha_m)$ varying and ρ fixed, the minimum ARE of $\tilde{\alpha}_j$ occurs on the boundary as $\alpha_j \to 0$ or 1. Hence for a more reasonable optimization, we set $\rho = 0.9$, $\alpha_m = 0.01$ and constrain α_j in $[0.01, 0.99]$ for α_j, $j = 1, \ldots, m-1$. Table 4 shows the constrained minimum ARE of $\tilde{\alpha}_m$ for $m = 3, \ldots, 8$ as well the vector $\boldsymbol{\alpha}$ leading to the minimum (the complementary vector $\mathbf{1} - \boldsymbol{\alpha}$ leads to the same value of ARE). The constrained minimum ARE values are much closer to 1 if the interval for each α_j is narrower such as [0.1, 0.9].

Finally, we show a table for bivariate extremes to illustrate Example 5. This example is relevant to extreme value inference. Note that the generalized extreme value distribution is a nonregular case in that the range of support depends on the μ_j parameters; however regular asymptotic maximum likelihood theory is valid for $\gamma_j > -\frac{1}{2}$ ([18;20, Section 7]). In practice, the shape parameter γ is usually in the interval $(-\frac{1}{2}, \frac{1}{2})$.

Because the expectations in $\mathcal{I}$ are computationally difficult and the results in the appendix do not apply, we approximate the asymptotic relative efficiencies based on Monte Carlo simulations. Table 5 has some ARE results for δ values for the Gumbel [5] copula $C(u_1, u_2; \delta) = \exp\{-([-\log u_1]^\delta + [-\log u_2]^\delta)^{1/\delta}\}$ ($\delta = 1$ for independence, $\delta \to \infty$ for the Fréchet upper bound) with δ values corresponding to Kendall tau values 0.1, 0.3, 0.5, 0.7, 0.9. Combinations of (γ_1, γ_2) chosen were (.1, .1), (.1, .2), (.2, .2), (.2, .3), $(-.1, -.1)$, $(-.1, .1)$. The estimated ARE values are based on sample sizes of $n = 500$ and 1000 simulations. The range of ARE for $\tilde{\mu}_j, \tilde{\sigma}_j, \tilde{\gamma}_j, \tilde{\delta}$ is given for each of the δ values. Note that the main loss of efficiency is in $\tilde{\gamma}_j$ as δ increases. In the moderate dependence range, most

Table 5
Range of AREs for bivariate extreme value with Gumbel copula

δ	ARE($\tilde{\mu}_j$)	ARE($\tilde{\sigma}_j$)	ARE($\tilde{\gamma}_j$)	ARE($\tilde{\delta}$)
1.11	.99–1.0	.99–1.0	.94–.97	.99–1.0
1.43	.96–.99	.97–1.0	.77–.85	.98–.99
2.00	.94–.97	.94–.99	.61–.72	.97–.99
3.33	.93–.96	.92–.98	.48–.62	.97–.99
10.0	.93–.96	.92–.98	.30–.53	.97–.99

common in applications, the efficiency loss may be acceptable. For this bivariate model, numerical ML is straightforward as there are just 7 parameters. In multivariate extreme value models where numerical ML is infeasible and the IFM method can be used, the efficiency loss would have to be accepted for computational reasons; see the next section on suggestions on how the IFM method might be modified.

For all of the examples, there was high efficiency for $\tilde{\delta}$ (see in particular Tables 1,2,5). From the elements of the matrices V and $\mathcal{I}$ in Section 2, note that if $\mathcal{I}_{jd} = 0$ for $j = 1, \ldots, m$, then $\mathcal{I}^{(dd)} = \mathcal{I}_{dd}^{-1}$, $V_{jd} = 0$ for $j = 1, \ldots, m$ and $V_{dd} = \mathcal{I}_{dd}^{-1}$. That is, $\tilde{\delta}$ is asymptotically fully efficient in this case. Hence if $\mathcal{I}_{jd}$, $j = 1, \ldots, m$, consist of small absolute values relative to other elements in $\mathcal{I}$, one can expect $\tilde{\delta}$ to be highly efficient. An example where $\mathcal{I}_{jd} = 0$ is a model consisting of a bivariate reflection symmetric copula $c(u, v; \delta) = c(1-u, 1-v; \delta)$ and symmetric univariate margins with a location parameter (proof is straightforward and is omitted).

5. Conclusions and further research

Results obtained in this paper were based on numerical comparisons in combination with derivation of theoretical results to explain the patterns in the efficiency of the two-stage estimation or IFM method. There is tradeoff between computatability and asymptotic relative efficiency of estimators. Generally, IFM has good efficiency except possibly for extreme dependence near the Fréchet bounds. For discrete margins, with few categories, the IFM estimator appears to be highly efficient, and the efficiency slowly worsens with more categories. It is in the case of continuous margins, that there can efficiency loss with strong dependence. Computationally two-stage or IFM estimation is much easier, particularly when the total number of parameters exceeds the 15–20 range. Standard errors of (functions of) parameters can be estimated by the jackknife in general or in some cases with the asymptotic covariance matrix. When maximum likelihood estimation is computationally feasible, the two-stage procedure provides a good starting point.

The two-stage procedure is especially convenient for the comparison of different copulas with the same set of univariate margins. This type of comparison is important in practice for a sensitivity analysis of inferences to the multivariate model.

If one feels that IFM estimation might be inefficient because of strong association, a version of estimating/inference equations based on a combination of univariate and bivariate log-likelihoods could be considered as an alternative to the multivariate log-likelihood. This approach is considered in Jöreskog and Moustaki [8] for a multivariate ordinal probit model with factor analysis, and is an example of what has been called a composite likelihood [10]. Its efficiency for multivariate models is a topic of future research. Other versions of estimation based on log-likelihoods of low-dimensional margins could also be considered. These variations are important for models in which computations of high-dimensional multivariate probabilities are difficult. The general good efficiency properties of the two-stage estimation method can be expected to carry over to these other variations.

Appendix

We prove that $\mathrm{Cov}(\mathbf{g}_j, \mathbf{g}_d) = \mathrm{E}[\mathbf{g}_j \mathbf{g}_d^T] = \mathbf{0}$, first for the continuous case, and then for the discrete case. The usual regularity conditions (including interchange of differentiation and integration/summation) for maximum likelihood theory are assumed. The support of F_j is assumed not to depend on the parameter $\boldsymbol{\alpha}_j$.

In the continuous case, let $\mathbf{y}_{-j}$ be the vector $\mathbf{y}$ omitting the jth component y_j. Then

$$\begin{aligned}
\mathrm{E}[\mathbf{g}_j \mathbf{g}_d^T] &= \int \frac{\partial \ell_j}{\partial \boldsymbol{\alpha}_j} \cdot \frac{\partial \log c}{\partial \boldsymbol{\delta}^T} \cdot f_1 \cdots f_m \cdot c(F_1, \ldots, F_m; \boldsymbol{\delta})\, dy_1 \cdots dy_m \\
&= \int \frac{\partial \ell_j}{\partial \boldsymbol{\alpha}_j} \cdot \frac{\partial c}{\partial \boldsymbol{\delta}^T} \cdot f_1 \cdots f_m \, dy_1 \cdots dy_m \\
&= \int_{y_j} \frac{\partial \ell_j}{\partial \boldsymbol{\alpha}_j} \int_{\mathbf{y}_{-j}} f_1 \cdots f_m \cdot \frac{\partial c(F_1, \ldots, F_m; \boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T}\, d\mathbf{y}_{-j} dy_j = \int_{y_j} \frac{\partial \ell_j}{\partial \boldsymbol{\alpha}_j} \frac{\partial}{\partial \boldsymbol{\delta}^T} f_j dy_j = \mathbf{0},
\end{aligned}$$

where an interchange of integration and differentiation is made in the second to last step.

In the discrete case, let $f_j(y_j; \boldsymbol{\alpha}_j) = \Pr(Y_j = y_j)$ where y_j are the support points of the jth margin. Let $f(\mathbf{y}; \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\delta})$ be the joint probability mass function with univariate margins $f_1, \ldots, f_m$. Let $I_1(y_1), \ldots, I_m(y_m), I(\mathbf{y})$ be the indicator functions. Then

$$\begin{aligned}
\ell_j &= \sum_{y_j} I_j(y_j) \log f_j(y_j; \boldsymbol{\alpha}_j) = \sum_{z_j} \sum_{\mathbf{y}: y_j = z_j} I(\mathbf{y}) \log f_j(z_j; \boldsymbol{\alpha}_j), \\
\ell &= \sum I(\mathbf{y}) \log f(\mathbf{y}; \delta, \boldsymbol{\alpha}), \\
\mathbf{g}_j &= \frac{\partial \ell_j}{\partial \boldsymbol{\alpha}_j} = \sum_{z_j} \sum_{\mathbf{y}: y_j = z_j} I(\mathbf{y})\, f_j^{-1} \frac{\partial f_j}{\partial \boldsymbol{\alpha}_j}, \\
\mathbf{g}_d &= \frac{\partial \ell}{\partial \boldsymbol{\delta}} = \sum_{\mathbf{y}} I(\mathbf{y})\, f^{-1} \frac{\partial f}{\partial \boldsymbol{\delta}}, \\
\mathrm{E}[\mathbf{g}_j \mathbf{g}_d^T] &= \sum_{\mathbf{y}} \mathrm{E}[I(\mathbf{y})]\, f_j^{-1} f^{-1} \frac{\partial f_j}{\partial \boldsymbol{\alpha}_j} \frac{\partial f}{\partial \boldsymbol{\delta}^T} \\
&= \sum_{\mathbf{y}} f_j^{-1} \frac{\partial f_j}{\partial \boldsymbol{\alpha}_j} \frac{\partial f}{\partial \boldsymbol{\delta}^T} = \sum_{z_j} f_j^{-1} \frac{\partial f_j}{\partial \boldsymbol{\alpha}_j} \sum_{\mathbf{y}: y_j = z_j} \frac{\partial f}{\partial \boldsymbol{\delta}^T} = \sum_{z_j} f_j^{-1} \frac{\partial f_j}{\partial \boldsymbol{\alpha}_j} \frac{\partial f_j}{\partial \boldsymbol{\delta}^T} = \mathbf{0}.
\end{aligned}$$

Acknowledgements

This research has been supported by an NSERC Research Grant. I am grateful to the referees for comments leading to an improved presentation.

References

[1] J. Aitchison, C.H. Ho, The multivariate Poisson-log normal distribution, Biometrika 76 (1989) 643–653.
[2] P.J. Davis, P. Rabinowitz, Methods of Numerical Integration, second ed., Academic Press, Orlando, 1984.
[3] M.J. Frank, On the simultaneous associativity of $F(x, y)$ and $x + y + F(x, y)$, Aequationes Math., 19 (1979) 194–226.
[4] V.P. Godambe (Ed.), Estimating Functions, Oxford University Press, Oxford, 1991.
[5] E.J. Gumbel, Distributions des valeurs extrêmes en plusieurs dimensions, Publ. Inst. Statist. Univ. Paris 9 (1960) 171–173.
[6] H. Joe, Multivariate extreme value distributions with applications to environmental data, Canad. J. Statist. 22 (1994) 47–64.
[7] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, London, 1997.
[8] K.G. Jöreskog, I. Moustaki, Factor analysis of ordinal variables: a comparison of three approaches, Multivariate Behav. Res. 36 (2001) 347–387.
[9] S.Y. Lee, W.Y. Poon, P.M. Bentler, A two-stage estimation of structural equation models with continuous and polytomous variables, Br. J. Math. Statist. Psychol. 48 (1995) 339–358.
[10] P.J. Heagerty, S.R. Lele, A composite likelihood approach to binary spatial data, J. Amer. Statist. Assoc. 93 (1998) 1099–1111.
[11] A. Maydeu-Olivares, Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates, J. Educ. Behav. Statist. 26 (2001) 49–69.
[12] B. Muthén, Contributions to factor analysis of dichotomous variables, Psychometrika 43 (1978) 551–560.
[13] B. Muthén, Latent variable structural equation modeling with categorical data, J. Econom. 22 (1979) 43–65.
[14] B. Muthén, A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators, Psychometrika 49 (1984) 115–132.
[15] J.C. Nash, Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, second ed., Hilger, New York, 1990.
[16] U. Olsson, Maximum likelihood estimation of the polychoric correlation coefficient, Psychometrika 44 (1979) 443–460.
[17] R.L. Plackett, A class of bivariate distributions, J. Amer. Statist. Assoc. 60 (1965) 516–522.
[18] P. Prescott, A.T. Walden, Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, Biometrika 67 (1980) 723–724.
[19] J.H. Shih, T.A. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics 51 (1995) 1384–1399.
[20] R.L. Smith, Maximum likelihood estimation in a class of nonregular cases, Biometrika 72 (1985) 67–90.
[21] R.L. Smith, J.A. Tawn, H.-K. Yuen, Statistics of multivariate extremes, Int. Statist. Inst. Rev. 58 (1990) 47–58.
[22] J.A. Tawn, Modelling multivariate extreme value distributions, Biometrika 77 (1990) 245–253.
[23] J.J. Xu, Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data. Ph.D. Thesis, Department of Statistics, University of British Columbia, 1996.