Discussion Paper, Series A, No. 2010-230

A Class of Nonparametric Estimators for Bivariate Extreme Value Copulas

Akio Suzukawa

November 22, 2010

Graduate School of Economics & Business Administration
Hokkaido University
Kita 9 Nishi 7, Kita-Ku, Sapporo 060-0809, JAPAN

A class of nonparametric estimators for bivariate extreme value copulas

Akio Suzukawa

Graduate School of Economics and Business Administration
Hokkaido University

Abstract

Extreme value copulas are the limiting copulas of component-wise maxima. A bivariate extreme value copulas can be represented by a convex function called Pickands dependence function. In this paper we consider nonparametric estimation of the Pickands dependence function. Several estimators have been proposed. They can be classified into two types: Pickands-type estimators and Capéraà-Fougères-Genest-type estimators. We propose a new class of estimators, which contains these two types of estimators. Asymptotic properties of the estimators are investigated, and asymptotic efficiencies of them are discussed under Marshall-Olkin copulas.

Key words and phrases: bivariate exponential distribution; extreme value distribution; Pickands dependence function.

1 Introduction

Copulas are functions that join multivariate distribution functions to their one-dimensional margins. A class of copulas derived from the limiting behavior of component-wise maxima of independent, identically distributed samples is that of extreme value copulas. Unlike the univariate case, there is no finite-dimensional parametrization in the multivariate extreme value distributions. In other words, the class of extreme value copulas cannot be represented by a finite-dimensional parameters. The multivariate extreme value distributions have been discussed in many textbooks of extreme value theory or copulas theory, for example, Galambos(1978, 1987), Resnick (1987), Joe (1997), Kotz and Nadarajah (2000), Beirlant, *et al.* (2004), Castillo, *et al.* (2005), de Haan and Ferreira (2006) and Nelsen (2006).

In this paper we consider bivariate extreme value distributions. Without loss of generality, we can assume that marginal distributions are exponentials with unit means. Let X and Y be random variables with survival functions $\bar{F}(x) = P(X > x)$ and $\bar{G}(y) = P(Y > y)$, respectively. When (X, Y) follows a bivariate extreme value distribution, its joint survival function can be represented as

$$S(x, y) = \exp\left\{-(x+y)A\left(\frac{y}{x+y}\right)\right\} \tag{1}$$

for $0 \le x, y < \infty$ with $x + y > 0$, where $A : [0, 1] \to [1/2, 1]$ is a convex function satisfying $A(0) = A(1) = 1$ and

$$\max(1-t, t) \le A(t) \le 1, \quad t \in [0, 1]. \tag{2}$$

The representation (1) was obtained by Pickands (1981), and the function A is called *Pickands dependence function.* The survival copula corresponding to the survival function S is given by

$$C(u, v) = S(-\log u, -\log v) = \exp\left[\log(uv)A\left\{\frac{\log v}{\log(uv)}\right\}\right], \quad 0 \le u, v \le 1.$$

The copula is determined by the Pickands dependence function A. Important examples of A are the lower and upper bounds of (2). If $A(t) \equiv 1$ (the upper bound), then X and Y are independent.

If $A(t) = \max(1-t, t)$ (the lower bound), then X and Y are completely dependent, that is, $X = Y$ holds with probability one.

Several parametric models for A are presented by Tawn (1988), Joe (1997), Kotz and Nadarajah (2000) and Beirlant, *et al.* (2004). One of the classical parametric models is the so-called logistic model, proposed by Gumbel (1960), and defined by

$$A(t) = \{(1-t)^q + t^q\}^{1/q}, \quad t \in [0,1], \tag{3}$$

where $q \geq 1$ is a dependence parameter. Independence and complete dependence correspond to $r = 1$ and $r = \infty$, respectively. Another classical parametric model is the so-called Marshall and Olkin (1967) model, defined by

$$A(t) = \max(1 - \theta t, 1 - \theta(1-t)), \quad t \in [0,1], \tag{4}$$

where $\theta \in [0,1]$ is a dependence parameter. Independence and complete dependence correspond to $\theta = 0$ and $\theta = 1$, respectively.

We are concerned with nonparametric estimation of the Pickands dependence function A. A nonparametric estimator of A was proposed by Pickands (1981). Modifications of the Pickands estimator were suggested by Tiago de Oliveira (1989), Deheuvels and Tiago de Oliveira (1989), Deheuvels (1991) and Hall and Tajvidi (2000). Another type of nonparametric estimator was proposed by Capéraà, Fougères and Genest (1997) . Modifications to satisfy the constraints of convexity were suggested by Hall and Tajvidi (2000), Jiménz, Villa-Diharce and Flores (2001) and Fils-Villetard, *et al.* (2008).

With the exception of modifications for the convexity, nonparametric estimators can be classified into two families: Pickands-type estimators and Capéraà-Fougères-Genest (CFG)-type estimators. Based on a simulation study, Capéraà, *et al.* (1997) discussed comparison between the Pickands-type estimators and the CFG-type estimators. Their results indicate that the CFG-type estimators are preferable to the Pickands-type estimators under a wide range of dependence structures. Segers (2008) gave a unified treatment of the Pickands-type and CFG-type, and showed that the CFG-type is asymptotically more efficient than Pickands-type under independence of X and Y. The unified treatment and the moment formulas obtained by Segers (2008) are very useful to discuss several properties of the nonparametric estimators. In this paper, we develop the unified treatment a little further and propose a new class of nonparametric estimators for the Pickands dependence function. The class includes both the Pickands-type and the CFG-type estimators.

The outline of the paper is as follows. In Section 2, based on the unified treatment obtained by Segers (2008), we define a new class of nonparametric estimators for the Pickands dependence function. The asymptotic properties of the nonparametric estimators are investigated in Section 3. In Section 4, we discuss comparison of the estimators under Marshall-Olkin model of (4). All proofs are given in Section 5.

2 A class of estimators for Pickands dependence function

Let (X_i, Y_i), $i = 1, 2, \ldots, n$ be random samples from the bivariate survival function (1). For $t \in [0,1]$, define

$$\xi_i(t) = \min\left(\frac{X_i}{1-t}, \frac{Y_i}{t}\right), \; i = 1, 2, \ldots, n.$$

Then, $\xi_i(t)$ is exponentially distributed with mean

$$\frac{1}{A(t)} = E[\xi_i(t)]. \tag{5}$$

Pickands (1981) estimator is an empirical version of this moment equation, and which is defined by

$$\frac{1}{\hat{A}^{\mathrm{P}}(t)} = n^{-1}\sum_{i=1}^{n} \xi_i(t) = n^{-1}\sum_{i=1}^{n} \min\left(\frac{X_i}{1-t}, \frac{Y_i}{t}\right).$$

The estimator does not satisfy the constraints $A(0) = A(1) = 1$.

Deheuvels (1991) proposed an estimator $\hat{A}^{\mathrm{D}}(t)$ defined by

$$\frac{1}{\hat{A}^{\mathrm{D}}(t)} = n^{-1}\sum_{i=1}^{n}\{\xi_i(t) - (1-t)(X_i - 1) - t(Y_i - 1)\},$$

and which satisfies $\hat{A}^{\mathrm{D}}(0) = \hat{A}^{\mathrm{D}}(1) = 1$. This estimator can be considered as an empirical version of a moment equation

$$\frac{1}{A(t)} = E[\xi_i(t) - (1-t)(X_i - 1) - t(Y_i - 1)]. \tag{6}$$

Another nonparametric estimator was proposed by Capéraà, Fougères and Genest (1997). They focused on equations

$$\log A(t) = \int_0^t \frac{\Pr\{Y_i/(X_i+Y_i) \le z\} - z}{z(1-z)}dz = -\int_t^1 \frac{\Pr\{Y_i/(X_i+Y_i) \le z\} - z}{z(1-z)}dz.$$

Replacing the distribution function $\Pr\{Y_i/(X_i+Y_i) \le z\}$ by corresponding empirical distribution function $n^{-1}\sum_{i=1}^{n} I\{Y_i/(X_i+Y_i) \le z\}$, where I is the indicator function, Capéraà, Fougères and Genest (CFG) (1997) estimator $\hat{A}^{CFG}$ is defined by

$$\begin{aligned}\log \hat{A}^{CFG}(t) &= p(t)\int_0^t \frac{n^{-1}\sum_{i=1}^{n} I\{Y_i/(X_i+Y_i) \le z\} - z}{z(1-z)}dz \\ &\quad -\{1-p(t)\}\int_t^1 \frac{n^{-1}\sum_{i=1}^{n} I\{Y_i/(X_i+Y_i) \le z\} - z}{z(1-z)}dz,\end{aligned}$$

where $p(t)$ is an appropriate weight function on$[0,1]$. Beirlant, *et al.* (2004) and Segers (2008) showed that $\hat{A}^{CFG}(t)$ is an empirical version of an equation

$$-\log A(t) = E\left[\log \xi_i(t) - p(t)\log X_i - \{1-p(t)\}\log Y_i\right], \tag{7}$$

and $\hat{A}^{CFG}(t)$ can be expressed as

$$-\log \hat{A}^{CFG}(t) = n^{-1}\sum_{i=1}^{n}\left[\log \xi_i(t) - p(t)\log X_i - \{1-p(t)\}\log Y_i\right].$$

In this paper, we consider Box and Cox's (1964) power-transformation on $[0,\infty)$ defined by

$$\varphi_\lambda(x) = \begin{cases} \lambda^{-1}(x^\lambda - 1), & \lambda > 0, \\ \log x, & \lambda = 0. \end{cases}$$

It can be easily verified that

$$E[\varphi_\lambda\{\xi_i(t)\}] = \begin{cases} \Gamma(1+\lambda)\varphi_\lambda\{1/A(t)\} + \lambda^{-1}\{\Gamma(1+\lambda) - 1\}, & \lambda > 0, \\ -\log A(t) - \gamma, & \lambda = 0, \end{cases}$$

where Γ is the gamma function and γ is the Euler's constant. Let $a(t)$ and $b(t)$ be appropriate weight functions on $[0,1]$. Then, it can be seen that

$$\begin{aligned}&E[\varphi_\lambda\{\xi_i(t)\} - a(t)\varphi_\lambda(X_i) - b(t)\varphi_\lambda(Y_i)] \\ &= \begin{cases} \Gamma(1+\lambda)\varphi_\lambda\{1/A(t)\} + \lambda^{-1}\{\Gamma(1+\lambda) - 1\}\{1 - a(t) - b(t)\}, & \lambda > 0, \\ -\varphi_0\{1/A(t)\} - \gamma\{1 - a(t) - b(t)\}, & \lambda = 0. \end{cases}\end{aligned}$$

From this, we can obtain an equation

$$\begin{aligned}\Gamma(1+\lambda)\varphi_\lambda\{1/A(t)\} &= E[\varphi_\lambda\{\xi_i(t)\} - a(t)\varphi_\lambda(X_i) - b(t)\varphi_\lambda(Y_i)] \\ &\quad -\lambda^{-1}\{\Gamma(1+\lambda) - 1\}\{1 - a(t) - b(t)\}. \end{aligned} \tag{8}$$

This equation is a generalization of equations (5), (6) and (7). When $\lambda = 1$ and $a(t) = b(t) \equiv 0$, the equation (8) reduces to (5). If $\lambda = 1$, $a(t) = 1-t$ and $b(t) = t$, then (8) gives (6). If two weight functions are chosen as $a(t) = 1 - p(t)$, $b(t) = p(t)$ and let $\lambda \to 0$ in (8), then (7) is obtained.

The equation (8) suggests estimating $A(t)$ by replacing the expectation term by sample means. We define an estimator $\hat{A}_\lambda(t; a(t), b(t))$ by

$$\begin{aligned}\varphi_\lambda\{1/\hat{A}_\lambda(t; a(t), b(t))\} &= \frac{1}{n\Gamma(1+\lambda)} \sum_{i=1}^{n} [\varphi_\lambda\{\xi_i(t)\} - a(t)\varphi_\lambda(X_i) - b(t)\varphi_\lambda(Y_i)] \\ &\quad - c_\lambda\{1 - a(t) - b(t)\}, \end{aligned} \tag{9}$$

where

$$c_\lambda = \begin{cases} \lambda^{-1}\{\Gamma(1+\lambda) - 1\}/\Gamma(1+\lambda), & \lambda > 0, \\ -\gamma, & \lambda = 0. \end{cases}$$

Noting that $\lim_{\lambda\to 0} c_\lambda = -\gamma = c_0$, as $\lambda \to 0$ in (9), we have a CFG-type estimator

$$-\log \hat{A}_0(t; a(t), b(t)) = \frac{1}{n} \sum_{i=1}^{n} [\log \xi_i(t) - a(t)\log X_i - b(t)\log Y_i] + \gamma\{1 - a(t) - b(t)\}.$$

From (8) and (9), we have

$$\begin{aligned} E\left[\varphi_\lambda\{1/\hat{A}_\lambda(t; a(t), b(t))\}\right] &= \frac{1}{\Gamma(1+\lambda)} E\left[\varphi_\lambda\{\xi_i(t)\} - a(t)\varphi_\lambda(X_i) - b(t)\varphi_\lambda(Y_i)\right] \\ &\quad - c_\lambda\{1 - a(t) - b(t)\} \\ &= \varphi_\lambda\{1/A(t)\}, \end{aligned}$$

for any fixed $t \in [0,1]$ and $\lambda \geq 0$. Hence, $\varphi_\lambda\{1/\hat{A}_\lambda(t; a(t), b(t))\}$ is an unbiased estimator of $\varphi_\lambda\{1/A(t)\}$. The constant term $-c_\lambda\{1-a(t)-b(t)\}$ in (8) is a bias-correction term. If $a(t)+b(t) = 1$, then this term is not needed.

Estimators $\hat{A}_1(t; 0, 0)$, $\hat{A}_1(t; 1-t, t)$, $\hat{A}_0(t; 1-p(t), p(t))$ are $\hat{A}^{\mathrm{P}}(t)$, $\hat{A}^{\mathrm{D}}(t)$ and $\hat{A}^{CFG}(t)$, respectively. In the next section, we investigate asymptotic properties of $\hat{A}_\lambda(t; a(t), b(t))$ for $\lambda \geq 0$.

3 Asymptotic Properties

Assume that weight functions a and b are bounded on $[0,1]$. Then, for any fixed $t \in [0,1]$ and $\lambda \geq 0$,

$$E[|\varphi_\lambda\{\xi_i(t)\} - a(t)\varphi_\lambda(X_i) - b(t)\varphi_\lambda(Y_i)|] < \infty.$$

Thus, by the strong law of large numbers, for any fixed $t \in [0,1]$ and $\lambda \geq 0$,

$$\lim_{n\to\infty} \varphi_\lambda\{1/\hat{A}_\lambda(t; a(t), b(t))\} = \varphi_\lambda\{1/A(t)\} \quad \text{a.s.}.$$

Since the transformation φ_λ is continuous, $\hat{A}_\lambda(t; a(t), b(t))$ is consistent, that is,

$$\lim_{n\to\infty} \hat{A}_\lambda(t; a(t), b(t)) = A(t) \quad \text{a.s.}.$$

The second moment of the transformed variate $\varphi_\lambda\{\xi_i(\cdot)\}$ is given by the following lemma.

Lemma 1 *For $\lambda > 0$ and $0 \leq s \leq t \leq 1$, covariance between $\varphi_\lambda\{\xi(s)\}$ and $\varphi_\lambda\{\xi(t)\}$ is given by*

$$\begin{aligned} &\mathrm{Cov}\left[\varphi_\lambda\{\xi(s)\}, \varphi_\lambda\{\xi(t)\}\right] \\ &= \frac{\Gamma(1+2\lambda)}{2\lambda^2}\left[\left[\frac{1-t}{(1-s)\{A(t)\}^2}\right]^\lambda + \left[\frac{s}{t\{A(s)\}^2}\right]^\lambda + \frac{\lambda}{(1-s)^\lambda t^\lambda}\int_s^t \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}dw\right] \\ &\quad - \frac{\{\Gamma(1+\lambda)\}^2}{\lambda^2\{A(s)A(t)\}^\lambda}. \end{aligned}$$

Letting $\lambda = 1$ in Lemma 1, we have

$$\begin{aligned}
&\mathrm{Cov}\left[\varphi_1\{\xi(s)\}, \varphi_1\{\xi(t)\}\right] = \mathrm{Cov}\left[\xi(s), \xi(t)\right] \\
&= \frac{1-t}{(1-s)\{A(t)\}^2} + \frac{s}{t\{A(s)\}^2} + \frac{1}{(1-s)t}\int_s^t \frac{1}{\{A(w)\}^2}dw - \frac{1}{A(s)A(t)}.
\end{aligned}$$

This formula has been obtained in Theorem 1 of Segers (2008). Covariance formula between $\varphi_0\{\xi(s)\} = \log \xi(s)$ and $\varphi_0\{\xi(t)\} = \log \xi(t)$ has been given in Theorem 2 of Segers (2008). The formula can be also derived from Lemma 1 as $\lambda \to 0$. The result is given in the next corollary.

Corollary 2 *For $0 \le s \le t \le 1$, covariance between $\varphi_0\{\xi(s)\} = \log \xi(s)$ and $\varphi_0\{\xi(t)\} = \log \xi(t)$ is given by*

$$\begin{aligned}
&\mathrm{Cov}[\varphi_0\{\xi(s)\}, \varphi_0\{\xi(t)\}] \\
&= \lim_{\lambda \to 0} \mathrm{Cov}[\varphi_\lambda\{\xi(s)\}, \varphi_\lambda\{\xi(t)\}] \\
&= \frac{\pi^2}{6} + (\log t)\log\frac{1-t}{1-s} + \int_s^t \frac{\log w}{1-w}dw - \{\log A(t)\}\log\frac{1-t}{1-s} - \{\log A(s)\}\log\frac{s}{t} \\
&\quad + \frac{1}{2}\left\{\log\frac{A(s)}{A(t)}\right\}^2 - \int_s^t \frac{\log A(w)}{w(1-w)}dw.
\end{aligned}$$

The estimator $\hat{A}_\lambda(t; a(t), b(t))$ consists of random samples of 3-dimensional random vector $[\varphi_\lambda\{\xi(t)\}, \varphi_\lambda(X), \varphi_\lambda(Y)]$. Covariance matrix of it is directly obtained from Lemma 1.

Corollary 3 *For any fixed $t \in [0,1]$ and $\lambda > 0$, covariance matrix of 3-dimensional random vector $[\varphi_\lambda\{\xi(t)\}, \varphi_\lambda(X), \varphi_\lambda(Y)]$ is given by*

$$\mathbf{\Sigma}_\lambda(t) = \begin{bmatrix} \sigma_\lambda(t,t) & \sigma_\lambda(0,t) & \sigma_\lambda(t,1) \\ \sigma_\lambda(0,t) & \sigma_\lambda(0,0) & \sigma_\lambda(0,1) \\ \sigma_\lambda(t,1) & \sigma_\lambda(0,1) & \sigma_\lambda(1,1) \end{bmatrix}, \tag{10}$$

where

$$\begin{aligned}
\sigma_\lambda(t,t) &= \mathrm{Var}[\varphi_\lambda\{\xi(t)\}] = \frac{\Gamma(1+2\lambda) - \{\Gamma(1+\lambda)\}^2}{\lambda^2\{A(t)\}^{2\lambda}}, \\
\sigma_\lambda(0,0) &= \sigma_\lambda(1,1) = \mathrm{Var}[\varphi_\lambda(X)] = \mathrm{Var}[\varphi_\lambda(Y)] = \frac{\Gamma(1+2\lambda) - \{\Gamma(1+\lambda)\}^2}{\lambda^2}, \\
\sigma_\lambda(0,t) &= \mathrm{Cov}[\varphi_\lambda(X), \varphi_\lambda\{\xi(t)\}] \\
&= \frac{\Gamma(1+2\lambda)}{2\lambda^2}\left[\left[\frac{1-t}{\{A(t)\}^2}\right]^\lambda + \frac{\lambda}{t^\lambda}\int_0^t \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}dw\right] - \frac{\{\Gamma(1+\lambda)\}^2}{\lambda^2\{A(t)\}^\lambda}, \\
\sigma_\lambda(t,1) &= \mathrm{Cov}[\varphi_\lambda\{\xi(t)\}, \varphi_\lambda(Y)] \\
&= \frac{\Gamma(1+2\lambda)}{2\lambda^2}\left[\left[\frac{t}{\{A(t)\}^2}\right]^\lambda + \frac{\lambda}{(1-t)^\lambda}\int_t^1 \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}dw\right] - \frac{\{\Gamma(1+\lambda)\}^2}{\lambda^2\{A(t)\}^\lambda}, \\
\sigma_\lambda(0,1) &= \mathrm{Cov}[\varphi_\lambda(X), \varphi_\lambda(Y)] \\
&= \frac{\Gamma(1+2\lambda)}{2\lambda}\int_0^1 \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}dw - \frac{\{\Gamma(1+\lambda)\}^2}{\lambda^2}.
\end{aligned}$$

Tawn (1988) has stated that correlation between X and Y, whose margins are exponential, is given by

$$\mathrm{Cor}[X,Y] = \int_0^1 \frac{1}{\{A(w)\}^2}dw - 1.$$

From Corollary 3, we can obtain a generalized formula

$$\begin{aligned}\mathrm{Cor}[X^\lambda, Y^\lambda] &= \mathrm{Cor}[\varphi_\lambda(X), \varphi_\lambda(Y)] = \sigma_\lambda(0,1)/\sigma_\lambda(0,0) \\ &= \frac{2^{-1}\lambda\Gamma(1+2\lambda)\int_0^1 \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}dw - \{\Gamma(1+\lambda)\}^2}{\Gamma(1+2\lambda) - \{\Gamma(1+\lambda)\}^2},\end{aligned}$$

which is correlation under Weibull margins. For any $\lambda > 0$, the two extreme case $\mathrm{Cor}[X^\lambda, Y^\lambda] = 0$ and 1 correspond to independence $A(w) \equiv 1$ and complete dependence $A(w) = \max(1-w, w)$.

Fundamental asymptotic properties of $\hat{A}_\lambda(t; a(t), b(t))$ are given by the following theorem.

Theorem 4 *Assume that weight functions a and b are bounded on* $[0,1]$. *Then, for any fixed* $t \in [0,1]$ *and* $\lambda \geq 0$,

$$E\left[\hat{A}_\lambda(t; a(t), b(t))\right] = A(t) + \frac{(1+\lambda)\{A(t)\}^{2\lambda+1}\, \tau_\lambda^2(t; a(t), b(t))}{2n} + O\left(\frac{1}{n^2}\right), \quad (11)$$

$$\mathrm{Var}\left[\hat{A}_\lambda(t; a(t), b(t))\right] = \frac{\{A(t)\}^{2(\lambda+1)}\, \tau_\lambda^2(t; a(t), b(t))}{n} + O\left(\frac{1}{n^2}\right) \quad (12)$$

and

$$\sqrt{n}\left\{\hat{A}_\lambda(t; a(t), b(t)) - A(t)\right\} \xrightarrow{L} N\left(0, \{A(t)\}^{2(\lambda+1)}\, \tau_\lambda^2(t; a(t), b(t))\right), \quad (13)$$

where

$$\tau_\lambda^2(t; a(t), b(t)) = \frac{\boldsymbol{\eta}'(t)\boldsymbol{\Sigma}_\lambda(t)\boldsymbol{\eta}(t)}{\{\Gamma(1+\lambda)\}^2},$$

$\boldsymbol{\eta}(t)$ *is a 3-dimensional column vector defined by* $\boldsymbol{\eta}'(t) = [1, -a(t), -b(t)]$ *and* $\boldsymbol{\Sigma}_\lambda(t)$ *is the covariance matrix defined in Corollary 3.*

From Theorem 4, expected squared error of $\hat{A}_\lambda(t; a(t), b(t))$ is given by

$$E\left[\left\{\hat{A}_\lambda(t; a(t), b(t)) - A(t)\right\}^2\right] = \mathrm{Var}\left[\hat{A}_\lambda(t; a(t), b(t))\right] + O\left(\frac{1}{n^2}\right),$$

and its main contribution comes from the variance term.

When $a(t) = b(t) \equiv 0$, the asymptotic variance is

$$\{A(t)\}^{2(\lambda+1)}\, \tau_\lambda^2(t; 0, 0) = \frac{\sigma_\lambda(0,0)}{\{\Gamma(1+\lambda)\}^2}\{A(t)\}^2.$$

Thus, asymptotic relative efficiency (ARE) of $\hat{A}_\lambda(t; 0, 0)$ with respect to Pickands estimator $\hat{A}^{\mathrm{P}}(t) = \hat{A}_1(t; 0, 0)$ is given by

$$\mathrm{ARE}\left[\hat{A}_\lambda(t; 0, 0), \hat{A}^{\mathrm{P}}(t)\right] = \frac{\{\Gamma(1+\lambda)\}^2}{\sigma_\lambda(0,0)},$$

which does not depend on t. Figure 1 shows the ARE for $\lambda \geq 0$. We can see that the ARE is less than one for $\lambda \neq 1$. If weight functions are not used, then Pickands estimator is asymptotically preferable. The estimator $\hat{A}_0(t; 0, 0)$ is an CFG-type estimator without weight functions, and its ARE is $6/\pi^2 \approx 0.608$. From these results, it can be seen that weight functions $a(t)$ and $b(t)$ play an important role in $\hat{A}_\lambda(t; a(t), b(t))$.

We now consider some conditions for the weight functions. From Theorem 4, asymptotic variances of $\hat{A}_\lambda(0; a(0), b(0))$ and $\hat{A}_\lambda(1; a(1), b(1))$ are given by

$$\tau_\lambda^2(0; a(0), b(0)) = \frac{\{(1-a(0))^2 + (b(0))^2\}\sigma_\lambda(0,0) - 2b(0)(1-a(0))\sigma_\lambda(0,1)}{\{\Gamma(1+\lambda)\}^2} \quad \text{and}$$

$$\tau_\lambda^2(1; a(1), b(1)) = \frac{\{(a(1))^2 + (1-b(1))^2\}\sigma_\lambda(1,1) - 2a(1)(1-b(1))\sigma_\lambda(0,1)}{\{\Gamma(1+\lambda)\}^2},$$

respectively. Thus, if $a(t)$ and $b(t)$ satisfy conditions $a(0) = b(1) = 1$ and $a(1) = b(0) = 0$, then these variances vanish. This is natural because, under these conditions, $\hat{A}_\lambda(t; a(t), b(t))$ satisfies a preferable property

$$\hat{A}_\lambda(0; a(0), b(0)) = \hat{A}_\lambda(1; a(1), b(1)) = 1.$$

Partition $\mathbf{\Sigma}_\lambda(t)$ of (10) as

$$\mathbf{\Sigma}_\lambda(t) = \left[\begin{array}{c|cc} \sigma_\lambda(t,t) & \sigma_\lambda(0,t) & \sigma_\lambda(t,1) \\ \hline \sigma_\lambda(0,t) & \sigma_\lambda(0,0) & \sigma_\lambda(0,1) \\ \sigma_\lambda(t,1) & \sigma_\lambda(0,1) & \sigma_\lambda(1,1) \end{array}\right] = \left[\begin{array}{cc} \sigma_\lambda(t,t) & \tilde{\boldsymbol{\sigma}}'_\lambda(t) \\ \tilde{\boldsymbol{\sigma}}_\lambda(t) & \tilde{\mathbf{\Sigma}}_\lambda \end{array}\right]$$

and put $\tilde{\boldsymbol{\eta}}'(t) = [a(t), b(t)]$ Then, the asymptotic variance of $\hat{A}_\lambda(t; a(t), b(t))$ can be expressed as

$$\frac{\{A(t)\}^{2(1+\lambda)}}{\{\Gamma(1+\lambda)\}^2}\left\{\sigma_\lambda(t,t) - 2\tilde{\boldsymbol{\sigma}}'_\lambda(t)\tilde{\boldsymbol{\eta}}(t) + \tilde{\boldsymbol{\eta}}'(t)\tilde{\mathbf{\Sigma}}_\lambda\tilde{\boldsymbol{\eta}}(t)\right\}.$$

For any fixed $t \in [0,1]$ and $\lambda \geq 0$, this is minimized at

$$\tilde{\boldsymbol{\eta}}(t) = \left[\begin{array}{c} a^*_\lambda(t) \\ b^*_\lambda(t) \end{array}\right] = \tilde{\mathbf{\Sigma}}_\lambda^{-1}\tilde{\boldsymbol{\sigma}}_\lambda(t) = \left[\begin{array}{cc} \sigma_\lambda(0,0) & \sigma_\lambda(0,1) \\ \sigma_\lambda(0,1) & \sigma_\lambda(1,1) \end{array}\right]^{-1}\left[\begin{array}{c} \sigma_\lambda(0,t) \\ \sigma_\lambda(t,1) \end{array}\right] \tag{14}$$

if $\tilde{\mathbf{\Sigma}}_\lambda$ is nonsingular. Corresponding minimal variance is

$$\frac{\{A(t)\}^{2(1+\lambda)}}{\{\Gamma(1+\lambda)\}^2}\left\{\sigma_\lambda(t,t) - \tilde{\boldsymbol{\sigma}}'_\lambda(t)\tilde{\mathbf{\Sigma}}_\lambda^{-1}\tilde{\boldsymbol{\sigma}}(t)\right\}. \tag{15}$$

We call $a^*_\lambda(t)$ and $b^*_\lambda(t)$ defined by (14) as optimal weight functions. These functions satisfy $a^*_\lambda(0) = b^*_\lambda(1) = 1$ and $a^*_\lambda(1) = b^*_\lambda(0) = 0$, and hence, $\hat{A}_\lambda(t; a^*_\lambda(t), b^*_\lambda(t))$ satisfy

$$\hat{A}_\lambda(0; a^*(0), b^*(0)) = \hat{A}_\lambda(1; a^*(1), b^*(1)) = 1.$$

Singularity of $\tilde{\mathbf{\Sigma}}_\lambda$ occurs only in complete dependence case because $\sigma_\lambda(0,1) = \sigma_\lambda(0,0)$ if and only if $A(w) = \max(1-w, w)$ for all $w \in [0,1]$. In other words, except for complete dependence case, optimal weight functions are uniquely determined by (14). However, we can not know them because they depend on the unknown dependence function A.

4 Asymptotic Comparison under Marshall-Olkin models

The purpose here is to explore a little further into optimalities of λ, $a(t)$ and $b(t)$ in the estimator $\hat{A}_\lambda(t; a(t), b(t))$. It seems reasonable to consider that the optimalities depend on the unknown $A(t)$. It is quite likely that dependence structure between X and Y influences the optimalities of λ, $a(t)$ and $b(t)$. In order to investigate such an influence, we assume the Marshall-Olkin model defined by (4). It is a symmetric case of the nondifferentiable asymmetric logistic model introduced by Tawn (1988), and is just the Marshall and Olkin's (1967) bivariate exponential model transformed to have unit exponential margins. When $\theta = 0$, we have independence. Complete dependence corresponds to $\theta = 1$. The joint distribution of X and Y is singular on the line $x = y$, and

$$\mathrm{Cor}[X, Y] = \Pr[X = Y] = \theta/(2-\theta).$$

Define a function f_λ on $[0,1]$ by

$$f_\lambda(t) = \lambda t^{-\lambda}\int_0^t w^{\lambda-1}(1-w)^{\lambda-1}\,dw.$$

This is one of the Gaussian hypergeometric function, that is,

$$f_\lambda(t) = {}_2F_1[\lambda, 1-\lambda; \lambda+1; t] = \sum_{j=0}^{\infty}\frac{(\lambda)_j(1-\lambda)_j\,t^j}{(1+\lambda)_j\,j!},$$

where $(c)_j$ is Pochhammer's symbol defined by $(c)_j = c(c+1)\cdots(c+j-1)$, cf. Johnson, *et al.* (2005).

Under (4), it holds that

$$\lambda \int_s^t \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}\, dw$$
$$= \begin{cases} \left(\frac{t}{A(t)}\right)^{\lambda} f_\lambda\left(\frac{(1-\theta)t}{A(t)}\right) - \left(\frac{s}{A(s)}\right)^{\lambda} f_\lambda\left(\frac{(1-\theta)s}{A(s)}\right), & t < \frac{1}{2}, \\ \frac{2}{(2-\theta)^\lambda} f_\lambda\left(\frac{1-\theta}{2-\theta}\right) - \left(\frac{s}{A(s)}\right)^{\lambda} f_\lambda\left(\frac{(1-\theta)s}{A(s)}\right) - \left(\frac{1-t}{A(t)}\right)^{\lambda} f_\lambda\left(\frac{(1-\theta)(1-t)}{A(t)}\right), & s < \frac{1}{2} \le t, \\ \left(\frac{1-s}{A(s)}\right)^{\lambda} f_\lambda\left(\frac{(1-\theta)(1-s)}{A(s)}\right) - \left(\frac{1-t}{A(t)}\right)^{\lambda} f_\lambda\left(\frac{(1-\theta)(1-t)}{A(t)}\right), & \frac{1}{2} \le s, \end{cases} \tag{16}$$

for $0 \le s \le t \le 1$. A proof of this is in Section 5.

From (16) and Corollary 3, we have, under (4),

$$\begin{aligned}
\frac{2\lambda^2\sigma_\lambda(0,0)}{\Gamma(1+2\lambda)} &= \frac{2\lambda^2\sigma_\lambda(1,1)}{\Gamma(1+2\lambda)} = 2 - f_\lambda(1), \\
\frac{2\lambda^2\sigma_\lambda(t,t)}{\Gamma(1+2\lambda)} &= \frac{2-f_\lambda(1)}{\{A(t)\}^{2\lambda}}, \\
\frac{2\lambda^2\sigma_\lambda(0,t)}{\Gamma(1+2\lambda)} &= \frac{(1-t)^\lambda}{\{A(t)\}^{2\lambda}} - \frac{f_\lambda(1)}{\{A(t)\}^\lambda} + \frac{I(t<1/2)}{\{A(t)\}^\lambda} f_\lambda\left(\frac{(1-\theta)t}{A(t)}\right) \\
&\quad + I(t \ge 1/2)\left\{\frac{2}{t^\lambda(2-\theta)^\lambda} f_\lambda\left(\frac{1-\theta}{2-\theta}\right) - \left(\frac{1-t}{tA(t)}\right)^\lambda f_\lambda\left(\frac{(1-\theta)(1-t)}{A(t)}\right)\right\}, \qquad (17) \\
\frac{2\lambda^2\sigma_\lambda(t,1)}{\Gamma(1+2\lambda)} &= \frac{t^\lambda}{\{A(t)\}^{2\lambda}} - \frac{f_\lambda(1)}{\{A(t)\}^\lambda} + \frac{I(t>1/2)}{\{A(t)\}^\lambda} f_\lambda\left(\frac{(1-\theta)(1-t)}{A(t)}\right) \\
&\quad + I(t \le 1/2)\left\{\frac{2}{(1-t)^\lambda(2-\theta)^\lambda} f_\lambda\left(\frac{1-\theta}{2-\theta}\right) - \left(\frac{t}{(1-t)A(t)}\right)^\lambda f_\lambda\left(\frac{(1-\theta)t}{A(t)}\right)\right\}, \\
\frac{2\lambda^2\sigma_\lambda(0,1)}{\Gamma(1+2\lambda)} &= \frac{2}{(2-\theta)^\lambda} f_\lambda\left(\frac{1-\theta}{2-\theta}\right) - f_\lambda(1),
\end{aligned}$$

where I is the indicator function.

The correlation coefficient between X^λ and Y^λ is given by

$$\mathrm{Cor}[X^\lambda, Y^\lambda] = \mathrm{Cor}[\varphi_\lambda(X), \varphi_\lambda(Y)] = \frac{\sigma_\lambda(0,1)}{\sigma_\lambda(0,0)} = \frac{2f_\lambda\left(\frac{1-\theta}{2-\theta}\right) - (2-\theta)^\lambda f_\lambda(1)}{(2-\theta)^\lambda\{2-f_\lambda(1)\}},$$

and which is non-increasing with $\lambda \ge 0$.

(Under complete dependence: $\theta = 1$)

Substitute $\theta = 1$ into (17). Then, we have covariance matrix of (10) as

$$\mathbf{\Sigma}_\lambda(t) = \frac{\Gamma(1+2\lambda)\{2-f_\lambda(1)\}}{2\lambda^2}\begin{bmatrix} \{A(t)\}^{-2\lambda} & \{A(t)\}^{-\lambda} & \{A(t)\}^{-\lambda} \\ \{A(t)\}^{-\lambda} & 1 & 1 \\ \{A(t)\}^{-\lambda} & 1 & 1 \end{bmatrix}$$

under complete dependence. Because of singularity, the optimal weight functions are not determined by (14). In this case, the asymptotic variance of $\hat{A}_\lambda(t; a(t), b(t))$ is given by

$$\frac{\{2-f_\lambda(1)\}\{A(t)\}^2}{\lambda^2 f_\lambda(1)}\left[1 - \{A(t)\}^\lambda\{a(t)+b(t)\}\right]^2. \tag{18}$$

Under complete dependence, if two weight functions satisfy a condition

$$a(t) + b(t) = \frac{1}{\{\max(1-t,t)\}^{\lambda}} \quad \text{for all } t \in [0,1], \tag{19}$$

then, (18) vanishes. This is a natural result. If (19) holds, then we have

$$\hat{A}_{\lambda}(t; a(t), b(t)) = A(t) = \max(1-t,t) \quad \text{for all } t \in [0,1] \tag{20}$$

with probability one. This is shown in Section 5. When $\lambda = 0$, (19) reduces to a simple condition $a(t) + b(t) \equiv 1$. This shows that, in CFG-type estimator ($\lambda = 0$), the simple weight functions $a(t) = p(t)$ and $b(t) = 1 - p(t)$ are optimal under complete dependence.

(Under independence: $\theta = 0$)

When $\theta = 0$, the covariance matrix of (10) is given by

$$\boldsymbol{\Sigma}_{\lambda}(t) = \frac{\Gamma(1+2\lambda)}{2\lambda^2} \begin{bmatrix} 2 - f_{\lambda}(1) & (1-t)^{\lambda} - f_{\lambda}(1) + f_{\lambda}(t) & t^{\lambda} - f_{\lambda}(1) + f_{\lambda}(1-t) \\ (1-t)^{\lambda} - f_{\lambda}(1) + f_{\lambda}(t) & 2 - f_{\lambda}(1) & 0 \\ t^{\lambda} - f_{\lambda}(1) + f_{\lambda}(1-t) & 0 & 2 - f_{\lambda}(1) \end{bmatrix}.$$

From (14) and (15), optimal weight functions and corresponding minimal variance are given by

$$a_{\lambda}^*(t) = \frac{(1-t)^{\lambda} - f_{\lambda}(1) + f_{\lambda}(t)}{2 - f_{\lambda}(1)}, \quad b_{\lambda}^*(t) = \frac{t^{\lambda} - f_{\lambda}(1) + f_{\lambda}(1-t)}{2 - f_{\lambda}(1)}$$

and

$$\frac{1}{\lambda^2 f_{\lambda}(1)} \left[2 - f_{\lambda}(1) - \frac{\{(1-t)^{\lambda} - f_{\lambda}(1) + f_{\lambda}(t)\}^2 + \{t^{\lambda} - f_{\lambda}(1) + f_{\lambda}(1-t)\}^2}{2 - f_{\lambda}(1)} \right],$$

respectively. When $\lambda = 1$, these reduce to $a_1^*(t) = 1 - t$, $b_1^*(t) = t$ and $2t(1-t)$, respectively, and which has been already derived by Segers (2008). When $\lambda = 0$, optimal weight functions and corresponding minimal variance are given by

$$a_0^*(t) = 1 - \frac{6}{\pi^2} L_2(t), \quad b_0^*(t) = 1 - \frac{6}{\pi^2} L_2(1-t)$$

and

$$\frac{\pi^2}{6} - \frac{\{\pi^2/6 - L_2(t)\}^2 + \{\pi^2/6 - L_2(1-t)\}^2}{\pi^2/6},$$

where $L_2(t)$ is the dilogarithm function defined by

$$L_2(t) = -\int_0^t \frac{\log(1-w)}{w}\, dw = \sum_{j=1}^{\infty} \frac{t^j}{j^2}.$$

These results for $\lambda = 0$ has been also obtained by Segers (2008).

Figure 2 shows the optimal weight functions a_{λ}^* and b_{λ}^* for $\lambda = 0$ (dotted curve), $\lambda = 1/3$ (broken curve), $\lambda = 1$ (solid line) and $\lambda = 3/2$ (broken-dotted curve). Under independence, the smaller λ is, the larger optimal weights for both marginals are.

The optimal weight functions under independence do not satisfy the optimality condition (19) under complete dependence. (19) is equivalent to $\{a(t) + b(t)\}\{\max(1-t,t)\}^{\lambda} = 1$. We are interested in how different $\{a_{\lambda}^*(t) + b_{\lambda}^*(t)\}\{\max(1-t,t)\}^{\lambda}$ is from one. Figure 3 shows $\{a_{\lambda}^*(t) + b_{\lambda}^*(t)\}\{\max(1-t,t)\}^{\lambda}$ for $\lambda = 0$ (dotted curve), $\lambda = 1/3$ (broken curve), $\lambda = 1$ (solid curve) and $\lambda = 3/2$ (broken-dotted curve). From this figure, we can see that, for $\lambda = 0, 1$ and $3/2$, a_{λ}^* and b_{λ}^*, which are optimal under independence, do not work well under complete dependence. On the other hand, $\{a_{1/3}^*(t) + b_{1/3}^*(t)\}\{\max(1-t,t)\}^{1/3}$ is near one for all $t \in [0,1]$. In other words, $a_{1/3}^*(t)$ and $b_{1/3}^*(t)$ are not so far from the optimality under complete dependence.

Figure 4 shows ARE of $\hat{A}_\lambda(t; a^*_\lambda(t), b^*_\lambda(t))$ with respect to $\hat{A}_0(t; a^*_0(t), b^*_0(t))$ which is CFG-type estimator with optimal weights, for $\lambda = 1/3$ (broken curve), $\lambda = 1/2$ (dotted curve), $\lambda = 1$ (solid curve) and $\lambda = 3/2$ (broken-dotted curve). Under independence, estimators for $\lambda \geq 1$ are inferior to the CFG-type estimator if weight functions are optimally chosen. When $0 < \lambda < 1$, $\hat{A}_\lambda(t; a^*_\lambda(t), b^*_\lambda(t))$ has smaller variance than the CFG-type estimator in neighborhoods of both edges. However, in general, the CFG-type estimator may be better under independence.

(Under the Marshall-Olkin model: $0 \leq \theta \leq 1$)

Covariances for general $0 \leq \theta \leq 1$ are given by (17). Unless $\theta = 1$ (complete dependence), optimal weight functions $a^*_\lambda(t)$ and $b^*_\lambda(t)$ are uniquely determined by (14), and they are symmetric about $t = 1/2$.

Figure 5 shows $a^*_\lambda(t)$ for $\lambda = 0, 1/3, 1$ and $3/2$ under $\theta = 0$ (solid curve), $\theta = 0.5$ (broken curve) and $\theta = 0.8$ (dotted curve). The parameter θ can be considered as one of the global dependence measures between X and Y. From Figure 5, we can see that $a^*_{1/3}(t)$ does not receive its influence so much and $a^*_{1/3}(t) \approx 1 - t$ for any $0 \leq \theta \leq 1$. This is important in the sense that the simple choice $a(t) = 1 - t$ and $b(t) = t$ can be used independently of θ when $\lambda = 1/3$.

For each $\lambda \geq 0$, minimal variance is given by (15). From that, we can obtain asymptotic relative efficiency (ARE) of $\hat{A}_\lambda(t; a^*_\lambda(t), b^*_\lambda(t))$ with respect to the CFG-type estimator $\hat{A}_0(t; a^*_0(t), b^*_0(t))$ with optimal weights. Figure 6 shows the ARE's of $\lambda = 1/3, 1/2, 1$ and $3/2$ under $\theta = 0$ (solid), $\theta = 0.3$ (broken), $\theta = 0.5$(dotted) and $\theta = 0.8$ (broken-dotted). When $\lambda = 1$ or $3/2$, corresponding Estimators for $\lambda = 1$ and $3/2$ are inefficient for all θ. On the other hand, estimators of $\lambda = 1/3$ and $1/2$ are more efficient than the CFG-type if θ is not so small. When $\theta \geq 0.5$ (in this case, correlation coefficient between X and Y is not less than 1/3), the estimator of $\lambda = 1/3$ is more efficient than the CFG-type for all $t \in [0, 1]$.

From Figure 6, we can see that, except for $\theta = 0$, the ARE has a minimum at symmetric points near $t = 1/4$ and $3/4$ and it has a substantial maximum at $t = 1/2$. We shall focus on $t = 1/4$ and $1/2$. Figure 7 shows ARE of $\hat{A}_\lambda(t; a^*_\lambda(t), b^*_\lambda(t))$ with respect to $\hat{A}_0(t; a^*_0(t), b^*_0(t))$ at $t = 1/4$ and $1/2$, for $\theta = 0, 0.3, 0.5$ and 0.8. The left ($t = 1/4$) and right ($t = 1/2$) figures approximately show minimum and maximum efficiencies, respectively. When $\theta = 0.8$ (0.5), the minimum efficiency is greater than one for $0 < \lambda < 0.6$ ($0 < \lambda < 0.4$), and the maximum efficiency is greater than one for $0 < \lambda < 0.9$ ($0 < \lambda < 0.8$). Except for independent case ($\theta = 0$), there exists a optimal $\lambda \in (0, 0.5)$. It seems reasonable to suppose that $0.2 < \lambda < 0.4$ is better than $\lambda = 0$ if there exists dependency to some degree.

The above comparison is based on optimal choice of the weight functions. However, the optimals depend on the unknown dependence function $A(t)$. Segers (2008) proposed an adaptive estimator, in which an initial estimator of $A(t)$ was used for estimating the optimal weight functions. We consider weight functions

$$\tilde{a}_\lambda(t) = \frac{1-t}{\{\max(1-t, t)\}^\lambda} \quad \text{and} \quad \tilde{b}_\lambda(t) = \frac{t}{\{\max(1-t, t)\}^\lambda}. \tag{21}$$

When $\lambda = 0$ (CFG-type), these are the simple weight functions $1 - t$ and t, respectively. For any $\lambda \geq 0$, $\tilde{a}_\lambda(t)$ and $\tilde{b}_\lambda(t)$ of (21) satisfy (19) which is a condition for optimality under complete dependence. This is an important condition in the sense that (20) holds if the bivariate data are completely dependent, that is, $X_i = Y_i$ for $i = 1, 2, \ldots, n$.

Figure 8 shows ARE of $\hat{A}_\lambda(t; \tilde{a}_\lambda(t), \tilde{b}_\lambda(t))$ with respect to $\hat{A}_\lambda(t; a^*_\lambda(t), b^*_\lambda(t))$. It means a loss of information by using $\tilde{a}_\lambda$ and $\tilde{b}_\lambda$ instead of the optimals a^*_λ and b^*_λ. The nearer zero θ is, the larger the loss is. In general, (21) may be suitable under strong dependency. Under independence ($\theta = 0$), about 20% and 50% information is lost in maximum when $\lambda = 0$ and 1, respectively. Efficiencies of the Pickands-type ($\lambda = 1$) and the CFG-type ($\lambda = 0$) are sensitive to choice of the weight functions. On the other hand, when $\lambda = 1/3$, the loss is not so large even if $\theta = 0$. The estimator $\hat{A}_{1/3}(t; \tilde{a}_{1/3}(t), \tilde{b}_{1/3}(t))$ has a stable efficiency.

As we have seen in Figure 6 and 7, if the weight functions can be chosen optimally for each $\lambda \geq 0$, the CFG-type ($\lambda = 0$) is more efficient under independence or weak dependency, and $0.2 < \lambda < 0.4$

is more efficient under strong dependency. As has been pointed out, the CFG-type is sensitive to choice of the weight functions. The simple weights $a(t) = 1 - t$ and $b(t) = t$ are approximately optimal for the CFG-type under strong dependency. However, under weak dependency, the simple choice causes a loss of information.

Practically, the simple weights has been used in the CFG-type and the Pickands-type estimators. Deheuvels estimator is Pickands-type with the simple weights. Segers (2008) has mentioned that the difference between the simple choice and the optimal choice are almost negligible within the class of the CFG-type. We are interested in comparison between

Deheuvels estimator : $\hat{A}_1(t; 1-t, t)$,

the CFG-type estimator with the simple weights : $\hat{A}_0(t; 1-t, t)$,

an estimator of $\lambda = 1/3$ with the simple weights : $\hat{A}_{1/3}(t; 1-t, t)$ and

an estimator of $\lambda = 1/3$ with the weights defined by (21) : $\hat{A}_{1/3}(t; \tilde{a}_{1/3}(t), \tilde{b}_{1/3}(t))$.

Figure 9 shows AREs of $\hat{A}_1(t; 1-t, t)$, $\hat{A}_{1/3}(t; 1-t, t)$ and $\hat{A}_{1/3}(t; \tilde{a}_{1/3}(t), \tilde{b}_{1/3}(t))$ with respect to $\hat{A}_0(t; 1-t, t)$ under $\theta = 0$, 0.5 and 0.8. The solid curve is the ARE of Deheuvels estimator, and which is not greater than one for all θ. Deheuvels estimator is asymptotically inferior to the simply weighted CFG-type estimator under the Marshall-Olkin model. Comparing the dotted and broken curves, it can be seen that the weight functions defined by (21) brings improvement of the simple weight functions when $\lambda = 1/3$. The reason for this is that, when $\lambda = 1/3$, the weight functions defined by (21) is more similar to the optimal weights than the simple weights.

Since the dotted curves in Figure 9 are beyond one, $\hat{A}_{1/3}(t; \tilde{a}_{1/3}(t), \tilde{b}_{1/3}(t))$ is asymptotically more efficient than the simply weighted CFG-type estimator under each θ. As we have seen in Figure 4, under independence, the CFG-type ($\lambda = 0$) is preferable to $\lambda = 1/3$ if the weight functions are optimally selected. However, the dominance relation under independence is reversed in $\hat{A}_0(t; 1-t, t)$ and $\hat{A}_{1/3}(t; \tilde{a}_{1/3}(t), \tilde{b}_{1/3}(t))$. The reason for this is that, under independence, the simple choice is different from the optimal choice of the CFG-type ($\lambda = 0$) as we have seen in Figure 5. Under strong dependency, the simple choice and the weights defined by (21) are approximately optimal for $\lambda = 0$ and $1/3$, respectively. As we have seen in Figure 6, in the case of the optimal choice, $\lambda = 1/3$ is more efficient than $\lambda = 0$ under strong dependency. These are the reason why $\hat{A}_{1/3}(t; \tilde{a}_{1/3}(t), \tilde{b}_{1/3}(t))$ is more efficient than $\hat{A}_0(t; 1-t, t)$.

5 Proofs

Proof of Lemma 1

Essential techniques for derivation of the covariance formula can be found in Segers (2008). For any $\lambda > 0$, some modifications are needed.

For $0 < s \le t < 1$,

$$\xi(s)\xi(t) = \min\left(\frac{X^2}{(1-s)(1-t)}, \frac{XY}{(1-s)t}, \frac{Y^2}{st}\right).$$

Thus, for $\lambda > 0$, we can express as

$$\begin{aligned}
&E\left[\{\xi(s)\xi(t)\}^\lambda\right] \\
&= \int_0^\infty \Pr\left\{\xi(s)\xi(t) > z^{1/\lambda}\right\} dz \\
&= \int_0^\infty \Pr\left\{X^2 > (1-s)(1-t)z^{1/\lambda}, XY > (1-s)tz^{1/\lambda}, Y^2 > stz^{1/\lambda}\right\} dz \\
&= \int_0^\infty E\left[I\left(X^2 > (1-s)(1-t)z^{1/\lambda}\right)\Pr\left\{Y > \max\left(\frac{(1-s)tz^{1/\lambda}}{X}, \sqrt{stz^{1/\lambda}}\right)|X\right\}\right] dz.
\end{aligned}$$

From Lemma 1 of Segers (2008), conditional survival function of Y given $X = x$ can be written as

$$\Pr(Y > y|X = x) = e^x g(x, y),$$

where $g(x, y) = S(x, y)Q(w)$, $Q(w) = A(w) - wA'(w)$, $w = y/(x + y)$ and A' is a right-hand derivative of A.

Hence, we have

$$\begin{aligned} & E\left[\{\xi(s)\xi(t)\}^\lambda\right] \\ = & \int_0^\infty E\left[I\left(X^2 > (1-s)(1-t)z^{1/\lambda}\right)e^X g\left(X, \max\left(\frac{(1-s)tz^{1/\lambda}}{X}, \sqrt{stz^{1/\lambda}}\right)\right)\right] dz \\ = & \int_0^\infty \int_0^\infty I\left(x^2 > (1-s)(1-t)z^{1/\lambda}\right) g\left(x, \max\left(\frac{(1-s)tz^{1/\lambda}}{x}, \sqrt{stz^{1/\lambda}}\right)\right) dxdz. \end{aligned}$$

Let $(x, z) = (x, u^\lambda x^{2\lambda})$ and $v(u) = \max((1-s)tu, \sqrt{stu})$. Then,

$$\begin{aligned} E\left[\{\xi(s)\xi(t)\}^\lambda\right] &= \int_0^\infty \int_0^\infty I((1-s)(1-t)u < 1)g(x, v(u)x)\lambda u^{\lambda-1}x^{2\lambda}\, dxdz. \\ &= \int_0^{\frac{1}{(1-s)(1-t)}} \lambda u^{\lambda-1}\left\{\int_0^\infty g(x, v(u)x)x^{2\lambda}\, dx\right\} du. \end{aligned}$$

Denoting $w(u) = v(u)/\{1 + v(u)\}$, it can be written as

$$\begin{aligned} & E\left[\{\xi(s)\xi(t)\}^\lambda\right] \\ = & \int_0^{\frac{1}{(1-s)(1-t)}} \lambda u^{\lambda-1}Q(w(u))\left[\int_0^\infty x^{2\lambda}\exp\{-(1+v(u))A(w(u))x\}\, dx\right] du. \\ = & \int_0^{\frac{1}{(1-s)(1-t)}} \lambda u^{\lambda-1}Q(w(u))\frac{\Gamma(1+2\lambda)}{\{(1+v(u))A(w(u))\}^{1+2\lambda}}\, du. \\ = & \Gamma(1+2\lambda)\int_0^{\frac{s}{(1-s)^2 t}} \frac{\lambda u^{\lambda-1}Q\left(\frac{\sqrt{stu}}{1+\sqrt{stu}}\right)}{\left\{(1+\sqrt{stu})A\left(\frac{\sqrt{stu}}{1+\sqrt{stu}}\right)\right\}^{1+2\lambda}}\, du \\ & + \Gamma(1+2\lambda)\int_{\frac{s}{(1-s)^2 t}}^{\frac{1}{(1-s)(1-t)}} \frac{\lambda u^{\lambda-1}Q\left(\frac{(1-s)tu}{1+(1-s)tu}\right)}{\left[\{1+(1-s)tu\}A\left(\frac{(1-s)tu}{1+(1-s)tu}\right)\right]^{1+2\lambda}}\, du. \end{aligned} \tag{22}$$

Put $w = \sqrt{stu}/(1 + \sqrt{stu})$ in the first term of (22), then

$$\begin{aligned} \text{the first term of (22)} &= \frac{2\lambda\Gamma(1+2\lambda)}{(st)^\lambda}\int_0^s \frac{w^{2\lambda-1}Q(w)}{\{A(w)\}^{1+2\lambda}}\, dw \\ &= \frac{\Gamma(1+2\lambda)}{(st)^\lambda}\int_0^s \frac{2\lambda w^{2\lambda-1}\{A(w) - wA'(w)\}}{\{A(w)\}^{1+2\lambda}}\, dw \\ &= \frac{\Gamma(1+2\lambda)}{(st)^\lambda}\left[\left\{\frac{w}{A(w)}\right\}^{2\lambda}\right]_0^s = \frac{\Gamma(1+2\lambda)}{\{A(s)\}^{2\lambda}}\left(\frac{s}{t}\right)^\lambda. \end{aligned}$$

Put $w = (1-s)tu/\{1 + (1-s)tu\}$ in the second term of (22), then

$$\begin{aligned} & \text{the second term of (22)} \\ = & \frac{\Gamma(1+2\lambda)}{(1-s)^\lambda t^\lambda}\int_s^t \frac{\lambda w^{\lambda-1}(1-w)^\lambda Q(w)}{\{A(w)\}^{1+2\lambda}}\, dw \\ = & \frac{\Gamma(1+2\lambda)}{2(1-s)^\lambda t^\lambda}\int_s^t \left(\frac{1-w}{w}\right)^\lambda \left\{\left(\frac{w}{A(w)}\right)^{2\lambda}\right\}'\, dw \end{aligned}$$

$$= \frac{\Gamma(1+2\lambda)}{2(1-s)^{\lambda}t^{\lambda}}\left\{\left[\left(\frac{1-w}{w}\right)^{\lambda}\left(\frac{w}{A(w)}\right)^{2\lambda}\right]_s^t + \lambda\int_s^t \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}\,dw\right\}$$
$$= \frac{\Gamma(1+2\lambda)}{2}\left[\left\{\frac{1-t}{(1-s)\{A(t)\}^2}\right\}^{\lambda} - \left\{\frac{s}{t\{A(s)\}^2}\right\}^{\lambda} + \frac{\lambda}{(1-s)^{\lambda}t^{\lambda}}\int_s^t \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}\,dw\right].$$

The result follows from

$$\begin{aligned}\mathrm{Cov}[\varphi_\lambda\{\xi(s)\},\varphi_\lambda\{\xi(t)\}] &= \mathrm{Cov}\left[\frac{\{\xi(s)\}^{\lambda}-1}{\lambda},\frac{\{\xi(t)\}^{\lambda}-1}{\lambda}\right] \\ &= \frac{E\left[\{\xi(s)\xi(t)\}^{\lambda}\right] - E\left[\{\xi(s)\}^{\lambda}\right]E\left[\{\xi(t)\}^{\lambda}\right]}{\lambda^2}. \square\end{aligned}$$

Proof of Corollary 2

Let, for $\lambda > 0$,

$$g(\lambda) = \lambda\int_0^1 u^{\lambda-1}(1-u)^{\lambda-1}\,du = \frac{2\{\Gamma(1+\lambda)\}^2}{\Gamma(1+2\lambda)}.$$

Then, by using L'Hôpital's rule, we can easily verify that, for any fixed $x > 0$,

$$\lim_{\lambda\downarrow 0} g(\lambda) = 2, \quad \lim_{\lambda\downarrow 0}\frac{2-g(\lambda)}{\lambda^2} = \frac{\pi^2}{3}, \tag{23}$$

$$\lim_{\lambda\downarrow 0}\frac{\varphi_\lambda(x)-\log x}{\lambda} = \frac{(\log x)^2}{2} \quad \text{and} \quad \lim_{\lambda\downarrow 0}\frac{\varphi_\lambda(x^2)-g(\lambda)\varphi_\lambda(x)}{\lambda} = (\log x)^2. \tag{24}$$

From Lemma 1, we can express as

$$\begin{aligned}&\frac{g(\lambda)}{\{\Gamma(1+\lambda)\}^2}\mathrm{Cov}[\varphi_\lambda\{\xi(s)\},\varphi_\lambda\{\xi(t)\}] \\ &= \frac{2}{\lambda^2} + \lambda^{-1}\varphi_\lambda\left(\frac{1-t}{(1-s)\{A(t)\}^2}\right) + \lambda^{-1}\varphi_\lambda\left(\frac{s}{t\{A(s)\}^2}\right) \\ &\quad +\lambda^{-1}\varphi_\lambda\left(\frac{1}{t(1-s)}\right)\int_s^t\left\{1+\lambda\varphi_\lambda\left(\frac{w(1-w)}{\{A(w)\}^2}\right)\right\}\frac{dw}{w(1-w)} \\ &\quad -\frac{g(\lambda)}{\lambda^2}\left\{1+\lambda\varphi_\lambda\left(\frac{1}{A(s)A(t)}\right)\right\} \\ &= \frac{2-g(\lambda)}{\lambda^2} \\ &\quad +\frac{1}{\lambda}\left[\varphi_\lambda\left(\frac{1-t}{(1-s)\{A(t)\}^2}\right) + \varphi_\lambda\left(\frac{s}{t\{A(s)\}^2}\right) + \int_s^t\frac{dw}{w(1-w)} - g(\lambda)\varphi_\lambda\left(\frac{1}{A(s)A(t)}\right)\right] \\ &\quad +\left[\varphi_\lambda\left(\frac{1}{t(1-s)}\right)\int_s^t\frac{dw}{w(1-w)} + \int_s^t\varphi_\lambda\left(\frac{w(1-w)}{\{A(w)\}^2}\right)\frac{dw}{w(1-w)}\right] \\ &\quad +\lambda\varphi_\lambda\left(\frac{1}{t(1-s)}\right)\int_s^t\varphi_\lambda\left(\frac{w(1-w)}{\{A(w)\}^2}\right)\frac{dw}{w(1-w)}.\end{aligned} \tag{25}$$

On the left-hand side of (25), $g(\lambda)/\{\Gamma(1+\lambda)\}^2 \to 2$ as $\lambda\downarrow 0$.

On the other hand, from (23), the first term on the right-hand side of (25) converges to $\pi^2/3$ as $\lambda\downarrow 0$. The last term of (25) converges to zero. The third term converges to

$$\begin{aligned}&\log\frac{1}{t(1-s)}\log\frac{t(1-s)}{s(1-t)} + \int_s^t\log\left\{\frac{w(1-w)}{(A(w))^2}\right\}\frac{dw}{w(1-w)} \\ &= -\frac{1}{2}(\log t)^2 - \frac{1}{2}(\log s)^2 - \frac{1}{2}\{\log(1-s)\}^2 - \frac{1}{2}\{\log(1-t)\}^2 \\ &\quad +2(\log t)\{\log(1-t)\} + (\log s)(\log t) + \{\log(1-s)\}\{\log(1-t)\} - 2\{\log(1-s)\}(\log t) \\ &\quad +2\int_s^t\frac{\log w}{1-w}\,dw - 2\int_s^t\frac{\log A(w)}{w(1-w)}\,dw.\end{aligned}$$

The second term can be written as

$$
\begin{aligned}
&\text{2nd term of right-hand side of (25)}\\
&= \frac{1}{\lambda}\left[\varphi_\lambda\left(\frac{1-t}{(1-s)\{A(t)\}^2}\right)+\varphi_\lambda\left(\frac{s}{t\{A(s)\}^2}\right)+\log\frac{t(1-s)}{s(1-t)}-g(\lambda)\varphi_\lambda\left(\frac{1}{A(s)A(t)}\right)\right]\\
&= \frac{1}{\lambda}\left[\varphi_\lambda\left(\frac{1-t}{(1-s)\{A(t)\}^2}\right)-\log\left(\frac{1-t}{(1-s)\{A(t)\}^2}\right)\right]\\
&\quad+\frac{1}{\lambda}\left[\varphi_\lambda\left(\frac{s}{t\{A(s)\}^2}\right)-\log\left(\frac{s}{t\{A(s)\}^2}\right)\right]-\frac{1}{\lambda}\left[\varphi_\lambda\left(\frac{1}{\{A(s)A(t)\}^2}\right)-\log\left(\frac{1}{\{A(s)A(t)\}^2}\right)\right]\\
&\quad+\frac{1}{\lambda}\left[\varphi_\lambda\left(\frac{1}{\{A(s)A(t)\}^2}\right)-g(\lambda)\varphi_\lambda\left(\frac{1}{A(s)A(t)}\right)\right].
\end{aligned}
$$

From (24), it is seen that the second term converges to

$$
\begin{aligned}
&\frac{1}{2}\left[\log\frac{1-t}{(1-s)\{A(t)\}^2}\right]^2+\frac{1}{2}\left[\log\frac{s}{t\{A(s)\}^2}\right]^2-\frac{1}{2}\left[\log\frac{1}{\{A(s)A(t)\}^2}\right]^2+\left[\log\frac{1}{A(s)A(t)}\right]^2\\
&\quad=\frac{1}{2}\{\log(1-t)\}^2+\frac{1}{2}\{\log(1-s)\}^2-\{\log(1-s)\}\{\log(1-t)\}+\frac{1}{2}(\log s)^2+\frac{1}{2}(\log t)^2\\
&\quad\quad-(\log s)(\log t)-2\left(\log\frac{1-t}{1-s}\right)\{\log A(t)\}-2\left(\log\frac{s}{t}\right)\{\log A(s)\}+\left\{\log\frac{A(s)}{A(t)}\right\}^2. \square
\end{aligned}
$$

Proof of Theorem 4

Let, for $i=1,2,\ldots,n$,

$$W_{\lambda,i}(t)=\frac{1}{\Gamma(1+\lambda)}\left[\varphi_\lambda\{\xi_i(t)\}-a(t)\varphi_\lambda(X_i)-b(t)\varphi_\lambda(Y_i)\right]-c_\lambda\{1-a(t)-b(t)\},$$

then, from(9),

$$\varphi_\lambda\{1/\hat{A}_\lambda(t;a(t),b(t))\}=\frac{1}{n}\sum_{i=1}^{n}W_{\lambda,i}(t).$$

It is nothing but a arithmetic average of i.i.d. random variables $W_{\lambda,i}(t)$, $i=1,2,\ldots,n$, with mean

$$E\left[W_{\lambda,i}(t)\right]=\varphi_\lambda\{1/A(t)\}$$

and variance

$$\text{Var}\left[W_{\lambda,i}(t)\right]=\tau_\lambda^2(t;a(t),b(t)).$$

For $\lambda\geq 0$, define a function $h_\lambda(x)$ on $[0,\infty)$ by

$$h_\lambda(x)=\begin{cases}(1+\lambda x)^{-1/\lambda}, & \lambda>0\\ e^{-x}, & \lambda=0.\end{cases}$$

Then, the k-th derivative of $h_\lambda(x)$ is given by

$$h_\lambda^{(k)}(x)=(-1)^k(1+\lambda x)^{-k}h_\lambda(x)\prod_{j=1}^{k-1}\{1+j\lambda\},$$

and which is bounded on $[0,\infty)$ for $k=1,2,\ldots$,. Noting

$$\hat{A}_\lambda(t;a(t),b(t))=h_\lambda\left(\frac{1}{n}\sum_{i=1}^{n}W_{\lambda,i}(t)\right)\quad\text{and}\quad A(t)=h_\lambda\left(\varphi_\lambda\{1/A(t)\}\right),$$

from Theorem 4.2.1 of Lehmann (1999), we obtain

$$
\begin{aligned}
E\left[\hat{A}_\lambda(t;a(t),b(t))\right] &= A(t)+\frac{\tau_\lambda^2(t;a(t),b(t))}{2n}h_\lambda^{(2)}(\varphi_\lambda\{1/A(t)\})+O\left(\frac{1}{n^2}\right),\\
\text{Var}\left[\hat{A}_\lambda(t;a(t),b(t))\right] &= \frac{\tau_\lambda^2(t;a(t),b(t))}{n}\left\{h_\lambda^{(1)}(\varphi_\lambda\{1/A(t)\})\right\}^2+O\left(\frac{1}{n^2}\right)
\end{aligned}
$$

and

$$\sqrt{n}\left\{\hat{A}_\lambda(t;a(t),b(t))-A(t)\right\}\xrightarrow{L} N\left(0,\tau_\lambda^2(t;a(t),b(t))\left\{h_\lambda^{(1)}(\varphi_\lambda\{1/A(t)\})\right\}^2\right).\ \square$$

Proof of equation (16)

We present a proof for $s<1/2\le t$. Equations for other cases can be shown similarly.

$$\lambda\int_s^t \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}\,dw=\lambda\int_s^{\frac{1}{2}}\frac{w^{\lambda-1}(1-w)^{\lambda-1}}{(1-\theta w)^{2\lambda}}\,dw+\lambda\int_{1-t}^{\frac{1}{2}}\frac{w^{\lambda-1}(1-w)^{\lambda-1}}{(1-\theta w)^{2\lambda}}\,dw.$$

Changing variable by $y=(1-\theta)w/(1-\theta w)$, we have

$$\begin{aligned}
&\lambda\int_s^t \frac{w^{\lambda-1}(1-w)^{\lambda-1}}{\{A(w)\}^{2\lambda}}\,dw\\
&= (1-\theta)^{-\lambda}\left\{\lambda\int_{\frac{1-\theta}{2-\theta}}^{\frac{(1-\theta)s}{1-\theta s}} y^{\lambda-1}(1-y)^{\lambda-1}\,dw+\lambda\int_{\frac{1-\theta}{2-\theta}}^{\frac{(1-\theta)(1-t)}{1-\theta(1-t)}} y^{\lambda-1}(1-y)^{\lambda-1}\,dw\right\}\\
&= (1-\theta)^{-\lambda}\left\{2\left(\frac{1-\theta}{2-\theta}\right)^\lambda f_\lambda\left(\frac{1-\theta}{2-\theta}\right)-\left(\frac{(1-\theta)s}{1-\theta s}\right)^\lambda f_\lambda\left(\frac{(1-\theta)s}{1-\theta s}\right)\right.\\
&\quad\left.-\left(\frac{(1-\theta)(1-t)}{1-\theta(1-t)}\right)^\lambda f_\lambda\left(\frac{(1-\theta)(1-t)}{1-\theta(1-t)}\right)\right\}.\ \square
\end{aligned}$$

Proof of equation (20)

Under complete dependence, $X_i=Y_i$, $i=1,2,\ldots,n$ with probability one. In this case, we have

$$\xi_i(t)=\min\left(\frac{X_i}{1-t},\frac{Y_i}{t}\right)=\frac{X_i}{\max(1-t,t)}=\frac{X_i}{A(t)},$$

and

$$\begin{aligned}
\varphi_\lambda(\xi_i(t))-a(t)\varphi_\lambda(X_i)-b(t)\varphi_\lambda(Y_i) &= \lambda^{-1}\left[X^\lambda\left\{(A(t))^{-\lambda}-a(t)-b(t)\right\}-1+a(t)+b(t)\right]\\
&= \lambda^{-1}\{a(t)+b(t)-1\}.
\end{aligned}$$

Thus, from (9), it holds that

$$\begin{aligned}
\varphi_\lambda(1/\hat{A}_\lambda(t;a(t),b(t))) &= \frac{1}{n\Gamma(1+\lambda)}\sum_{i=1}^n \lambda^{-1}\{a(t)+b(t)-1\}-c_\lambda\{1-a(t)-b(t)\}\\
&= \frac{a(t)+b(t)-1}{\lambda}=\frac{(A(t))^{-\lambda}-1}{\lambda}=\varphi_\lambda(1/A(t)).\ \square
\end{aligned}$$

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes, Theory and Applications.* Wiley.

[2] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, **26**, 211-252.

[3] Capéraà, P., Fougères, A. L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, **84**, 567-577.

[4] Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley.

[5] Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, **12**, 429-439.

[6] Deheuvels, P. and Tiago de Oliveira, J. (1989). On the non-parametric estimation of the bivariate extreme-value distributions. *Statistics & Probability Letters*, **8**, 315-323.

[7] Fils-Villetard, A., Guillou, A., and Segers, J. (2008). Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics*, **36**, No.3, 369-382.

[8] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*, Wiley.

[9] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed., Krieger.

[10] Gumbel, E. J. (1960). Bivariate exponential distributions. *J. Am. Statist. Assoc.*, **55**, 698-707.

[11] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory, An Introduction.* Springer.

[12] Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. Bernoulli, **6**, 835-844.

[13] Jiménz, J. R., Villa-Diharce, E. and Flores, M. (2001). Nonparametric estimation of dependence function in bivariate extreme value distributions. *Journal of Multivariate Analysis*, bf 76, 159-191.

[14] Joe, H. (1997). *Multivariate Models and Dependence Concept.* Chapman and Hall.

[15] Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate Discrete Distributions, 3rd ed.* Wiley.

[16] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions*, Imperial College Press.

[17] Lehmann, E. L. (1999). *Elements of Large-Sample Theory.* Springer.

[18] Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution, *Journal of the American Statistical Association*, 62, 30-44.

[19] Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer.

[20] Pickands, J. (1981). Multivariate extreme value distributions. *Proceedings of the 43rd Session of the ISI*, Buenos Aires, **49**, 859-878.

[21] Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*, Springer.

[22] Segers, J. (2008). Non-parametric inference for bivariate extreme-value copulas. *Topics in Extreme Values*, Ahsanullah, M. and Kirmani, S., Ed. Nova Science Publishers, Inc. 185-207.

[23] Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation, *Biometrika*, **75**, 3, 397-415.

[24] Tiago de Oliveira, J. (1989). Intrinsic estimation of the dependence structure for bivariate extremes. *Statistics & Probability Letters*, **8**, 213-218.

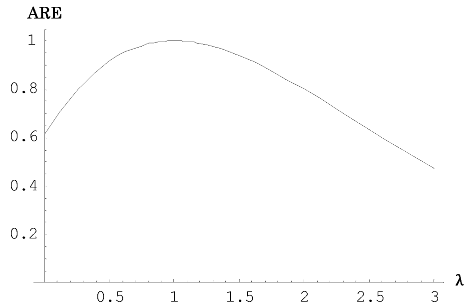

Figure 1. Asymptotic relative efficiency of $\hat{A}_\lambda(t;0,0)$ with respect to Pickands estimator $\hat{A}^{\mathrm{P}}(t) = \hat{A}_1(t;0,0)$

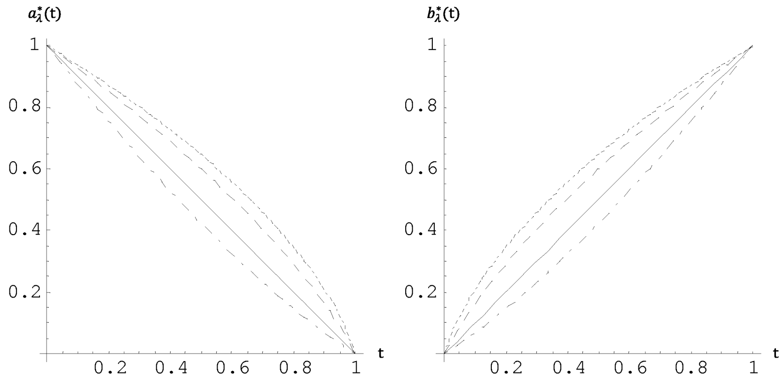

Figure 2. Optimal weight functions $a^*_\lambda(t)$ and $b^*_\lambda(t)$ for $\lambda = 0$ (dotted), $\lambda = 1/3$ (broken), $\lambda = 1$ (solid) and $\lambda = 3/2$ (broken-dotted) under independence.

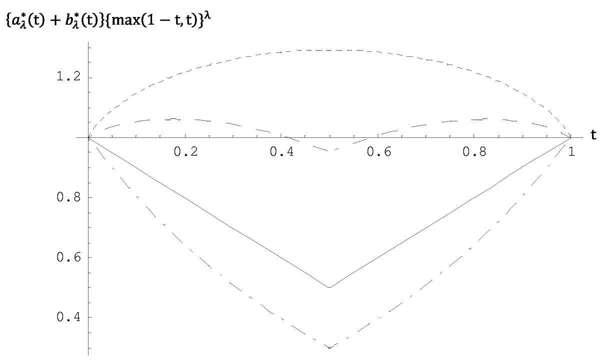

Figure 3. $\{a^*_\lambda(t) + b^*_\lambda(t)\}\{\max(1-t,t)\}^\lambda$ of optimal weight functions $a^*_\lambda(t)$ and $b^*_\lambda(t)$ under independence, for $\lambda = 0$ (dotted), $\lambda = 1/3$ (broken), $\lambda = 1$ (solid) and $\lambda = 3/2$ (broken-dotted).

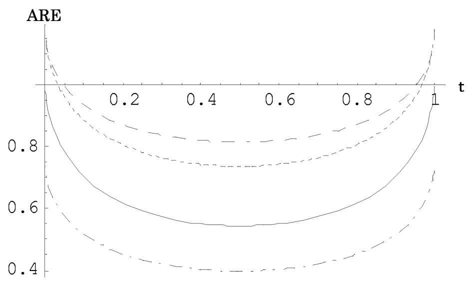

Figure 4. Asymptotic relative efficiency of $\hat{A}_\lambda(t; a^*_\lambda(t), b^*_\lambda(t))$ with respect to $\hat{A}_0(t; a^*_0(t), b^*_0(t))$, for $\lambda = 1/3$ (broken), $\lambda = 1/2$ (dotted), $\lambda = 1$ (solid) and $\lambda = 3/2$ (broken-dotted) under independence.

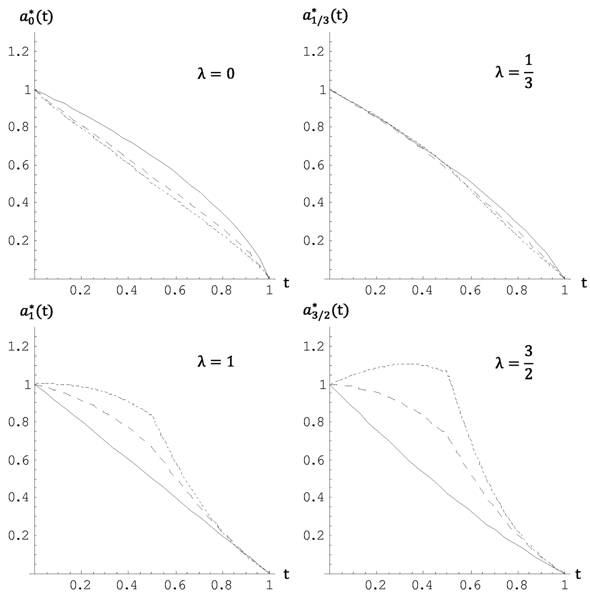

Figure 5. Optimal weight function $a^*_\lambda(t)$ under Marshall and Olkin's Model (4) of $\theta = 0$ (solid), $\theta = 0.5$ (broken) and $\theta = 0.8$ (dotted).

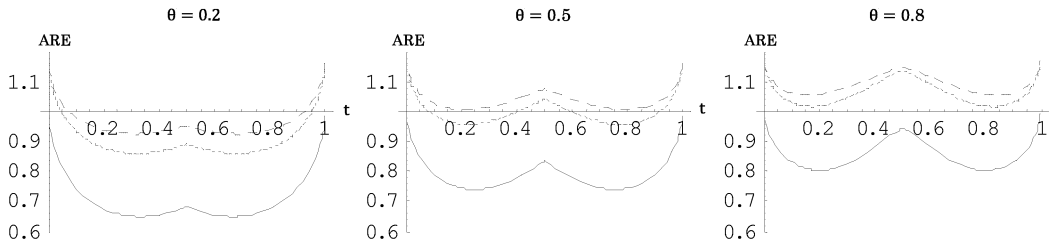

Figure 6. Asymptotic relative efficiency of $\hat{A}_\lambda(t; a_\lambda^*(t), b_\lambda^*(t))$ with respect to $\hat{A}_0(t; a_0^*(t), b_0^*(t))$, for $\lambda = 1/3$ (broken), 1/2 (dotted) and 1 (solid) under Marshall and Olkin's Model (4).

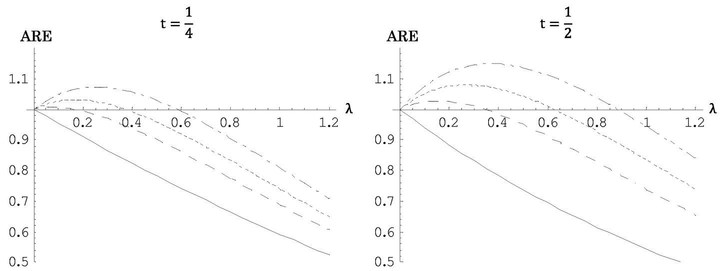

Figure 7. Asymptotic relative efficiency of $\hat{A}_\lambda(t; a_\lambda^*(t), b_\lambda^*(t))$ with respect to $\hat{A}_0(t; a_0^*(t), b_0^*(t))$, at $t = 1/4$ and 1/2, under Marshall and Olkin's Model (4) of $\theta = 0$ (solid), 0.3 (broken), 0.5 (dotted) and 0.8 (broken-dotted).

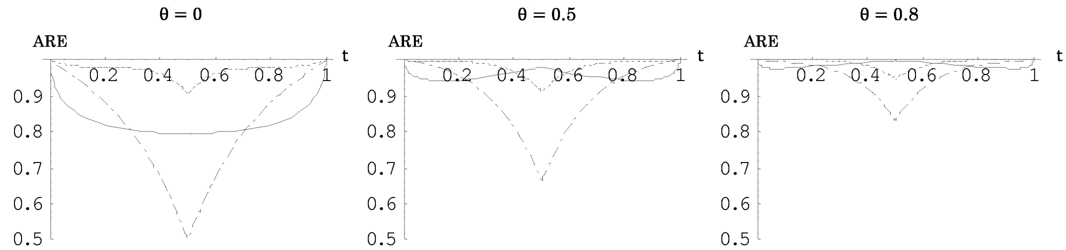

Figure 8. Asymptotic relative efficiency of $\hat{A}_\lambda(t; \tilde{a}_\lambda(t), \tilde{b}_\lambda(t))$ with respect to $\hat{A}_\lambda(t; a_\lambda^*(t), b_\lambda^*(t))$ for $\lambda = 0$ (solid), $\lambda = 1/3$ (dotted) and $\lambda = 1$ (broken-dotted) under Marshall and Olkin's Model (4).

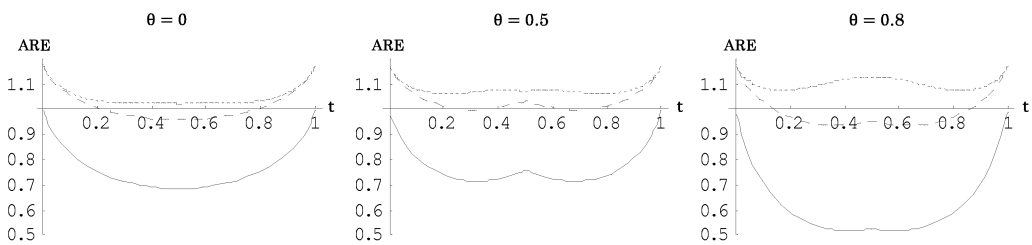

Figure 9. Asymptotic relative efficiency of $\hat{A}_1(t; 1-t, t)$ (solid), $\hat{A}_{1/3}(t; 1-t, t)$ (broken) $\hat{A}_{1/3}(t; \tilde{a}_{1/3}(t), \tilde{b}_{1/3}(t))$ (dotted) with respect to $\hat{A}_0(t; 1-t, t)$ under Marshall and Olkin's Model (4).