Extremes 8, 5–26, 2005


# Regular Score Tests of Independence in Multivariate Extreme Values

ALEXANDRA RAMOS
*Faculdade de Economia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-464, Porto, Portugal*
*CMUP, Rua do Campo Alegre, 687, 4169-007, Porto, Portugal*
*E-mail: aramos@fep.up.pt*

ANTHONY LEDFORD
*AHL Research, Man Investments, Sugar Quay, Lower Thames Street, London, EC3R 6DU, UK*
*E-mail: ALedford@maninvestments.com*

[Received 21 July 2004; Revised 29 March 2005; Accepted 18 May 2005]

**Abstract.** The score tests of independence in multivariate extreme values derived by Tawn (Tawn, J.A., "Bivariate extreme value theory: models and estimation," *Biometrika* 75, 397–415, 1988) and Ledford and Tawn (Ledford, A.W. and Tawn, J.A., "Statistics for near independence in multivariate extreme values," *Biometrika* 83, 169–187, 1996) have non-regular properties that arise due to violations of the usual regularity conditions of maximum likelihood. Two distinct types of regularity violation are encountered in each of their likelihood frameworks: independence within the underlying model corresponding to a boundary point of the parameter space and the score function having an infinite second moment. For applications, the second form of regularity violation has the more important consequences, as it results in score statistics with non-standard normalisation and poor rates of convergence. The corresponding tests are difficult to use in practical situations because their asymptotic properties are unrepresentative of their behaviour for the sample sizes typical of applications, and extensive simulations may be needed in order to evaluate adequately their null distribution. Overcoming this difficulty is the primary focus of this paper.

We propose a modification to the likelihood based approaches used by Tawn (Tawn, J.A., "Bivariate extreme value theory: models and estimation," *Biometrika* 75, 397–415, 1988) and Ledford and Tawn (Ledford, A.W. and Tawn, J.A., "Statistics for near independence in multivariate extreme values," *Biometrika* 83, 169–187, 1996) that provides asymptotically normal score tests of independence with regular normalisation and rapid convergence. The resulting tests are straightforward to implement and are beneficial in practical situations with realistic amounts of data.

**Key words.** independence testing, multivariate extreme value distribution, non-regular likelihood inference

AMS 2000 Subject Classification. Primary—60G70
Secondary—62H15

## 1. Introduction

Univariate and multivariate extreme value (MEV) distributions arise as the limiting distributions of linearly normalised componentwise maxima (Leadbetter et al., 1983; Resnick, 1987). The study and practical application of these distributions is becoming increasingly important in a wide range of application areas such as environmental modelling, flood defence analysis and financial risk assessment. See, for example, Smith (1989), Davison and Smith (1990), Coles and Tawn (1994), Embrechts et al. (1997), Coles (2001), Kotz and Nadarajah (2000), Joe (1997) and Heffernan and Tawn (2004). In the case of MEV distributions there are two separate aspects that have to be considered: the component marginal distributions, which can be handled using univariate extreme value methods, and the dependence structure that relates them. A particularly important dependence structure within the class of MEV distributions is independence, which arises when the limiting joint distribution as $n \to \infty$ of the suitably normalised componentwise maxima of $n$ independent and identically distributed vector observations factorises into its marginal distributions. This behaviour is exhibited for example by all multivariate normal variables with correlations in the open interval $(-1, 1)$, see Sibuya (1960). The independence case is important also from a practical perspective, allowing significant model simplification when it is supported by the underlying data.

The need for statistical tools for assessing when independence within componentwise maxima is appropriate has motivated the development of various techniques for testing independence between marginal extremes. These include the score statistics developed by Tawn (1988) and Ledford and Tawn (1996), the dependence function approaches of Capéraà et al. (1997) and Deheuvels (1980), a Cramer-von Mises-type statistic by Deheuvels and Martynov (1996), and a test based on the number of points below certain thresholds by Dorea and Miasaki (1993). Recently, the behaviour of Kendall's-$\tau$ as a measure of dependence within extremes has been examined, see Capéraà et al. (2000) and Genest and Rivest (2001). An alternative likelihood based approach that uses additional occurrence time information is also provided by Stephenson and Tawn (2005).

In this paper we will focus on extending the score statistic results of Tawn (1988) and Ledford and Tawn (1996). Their score statistics are constructed through examining the independence case exclusively from within the class of MEV distributions. We adopt the same approach here but note that alternative techniques based on families outside the MEV class have been developed, e.g., those based on the coefficient of tail dependence (Ledford and Tawn, 1996, 1997). Although the Tawn (1988) and Ledford and Tawn (1996) results provide score statistics with asymptotically normal null distributions, close inspection reveals their tests to have poor rates of convergence, a problem which leads to difficulties in applications. For example, very extensive simulation may be required in order to obtain good estimates of critical points of the test statistic for each particular sample size of interest. The aim of this paper is to overcome such difficulties by constructing score tests of independence with rapid convergence to standard normality.

An outline of the structure of this paper now follows. In Section 2 a brief introduction focusing on dependence in multivariate extremes is presented. The problematic Tawn (1988) and Ledford and Tawn (1996) score tests of independence together with our modified versions are discussed in Section 3. A comparison of practical aspects of the original and modified tests together with the impact of deviations from our assumed model structure are given in Section 4. A brief conclusion and some summary discussion are provided in Section 5, together with comments on an alternative scheme for modifying the original score tests. Some results for an alternative dependence structure to that used in the main paper together with asymptotic properties of likelihood inference under our modified framework are provided in appendices.

## 2. Dependence in multivariate extreme value distributions

As we are primarily interested in dependence issues for this paper, we will follow the usual approach of working with standardised marginal variables. For simplicity, we choose these to be unit Fréchet distributed, i.e., having distribution function $F_i(x) = \exp(-1/x)$ on $x \geq 0$ where $i$ indexes the dimension of the multivariate variable we are considering. This approach admits no loss of generality, as probability integral transformations may be used in order to extend this framework to arbitrary marginal distributions.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed (iid) $d$-dimensional random variables with joint distribution function $F$ and unit Fréchet marginal distributions, and define the $d$-vector $\mathbf{M}_n$ to be the vector of componentwise maxima, i.e., $\mathbf{M}_n = (M_n^{(1)}, \ldots, M_n^{(d)})$ where $M_n^{(j)} = \max(\mathbf{X}_1^{(j)}, \ldots, \mathbf{X}_n^{(j)})$ for $j = 1, \ldots, d$, where $\mathbf{X}_i^{(j)}$ represents the $j$th component of $\mathbf{X}_i$. We assume that $F$ is a MEV distribution so that

$$\Pr\{\mathbf{M}_n/n \leq \mathbf{x}\} = \mathrm{F}(\mathbf{x}) \text{ for all } n \geq 1$$

with $F(\mathbf{x}) = \exp\{-V(\mathbf{x})\}$, where

$$V(\mathbf{x}) = V(x_1, \ldots, x_d) = \int_{S_d} \max_{1 \leq j \leq d} (w_j/x_j) \mathrm{d}H(\mathbf{w})$$

for a non-negative measure $H$ on the $(d-1)$-dimensional unit simplex

$$S_d = \left\{ (w_1, \ldots, w_d); \sum_{j=1}^{d} w_j = 1, w_j \geq 0,\ j = 1, \ldots, d \right\}.$$

The non-negative measure $H$ determines the dependence structure of $F$ and satisfies

$$\int_{S_d} w_j \,\mathrm{d}H(\mathbf{w}) = 1 \quad \text{for each } \ j = 1, \ldots, d \tag{1}$$

but is otherwise arbitrary (see Coles and Tawn, 1991, 1994). Independence arises when the measure $H$ places atoms of unit mass at each of the $d$-vertices of $S_d$.

From the representation above we see that no finite parametrisation exists for the general MEV distribution as the non-negative measure $H$ is arbitrary apart from having to satisfy the normalisation condition (1). To address this, several flexible parametric families have been developed that satisfy this condition and cover a broad range of possible behaviours (Coles and Tawn, 1991, 1994). In this paper we concentrate on the two most widely used of these, the logistic and the mixed models. In the logistic model, which provides the main dependence structure discussed in this paper, the dependence function is defined by $V(\mathbf{x}) = V(x_1, \ldots, x_d) = (\sum_{i=1}^{d} x_i^{-1/\alpha})^{\alpha}$ for some dependence parameter $\alpha \in (0, 1]$. Clearly, when $\alpha = 1$, which is a boundary of the parameter space for $\alpha$, we obtain $V(\mathbf{x}) = \sum_{i=1}^{d} x_i^{-1}$ and hence that the marginal variables are independent. The mixed model is given by $V(\mathbf{x}) = \sum_{i=1}^{d} x_i^{-1} + (-1)^{d+1}\theta(\sum_{i=1}^{d} x_i)^{-1}$ where $\theta \in (0, 1]$ is a dependence parameter. Independence here arises when $\theta = 0$, which is again a boundary point of the parameter space. Other examples, such as the asymmetric logistic model or the bilogistic model can be found in Coles and Tawn (1991, 1994) and Joe (1989) however these models do not yield tractable score tests due to parameter unidentifiability problems at independence.

## 3. Score tests of independence

As noted above, independence arises for the logistic and mixed dependence structures in the special cases when $\alpha = 1$ and $\theta = 0$ respectively. Score tests based on these specified parameter values were developed by Tawn (1988) and Ledford and Tawn (1996) under differing likelihood frameworks. A unified treatment of these frameworks will be our starting point in what follows. For simplicity of presentation, we restrict attention to the bivariate case, $d = 2$, primarily because doing so admits no loss of generality when testing for independence in MEV distributions (Tiago de Oliveira, 1962, 1963), and additionally because corresponding modifications in the multivariate ($d \geq 3$) case are straightforward. We will focus on the logistic dependence structure for $V(\mathbf{x})$ in our derivation and analysis. Corresponding results for the mixed model will be examined later.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote iid bivariate random variables with joint distribution function $F(x, y)$ and unit Fréchet marginal distributions. It is convenient to partition the outcome space $R = \{(x, y) : (x, y) \in \mathbb{R}^+ \times \mathbb{R}^+\}$ into the four regions given by

$$\{R_{kl} : k = I(x \geq u), l = I(y \geq u)\} \tag{2}$$

where $u \geq 0$ is a threshold and $I$ denotes the indicator function. The different modelling frameworks adopted by Tawn (1988) and Ledford and Tawn (1996) may be described using the following common representation: the joint distribution function $F(x, y)$ is assumed to satisfy

$$F(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\} \text{ for } (x, y) \text{ in some specified region, } R_\alpha \text{ say.} \tag{3}$$

The Tawn (1988) approach takes $R_\alpha = R$, and thus treats the distribution function (3) as a model for the entire outcome space of $(X, Y)$. In contrast, the approach adopted by Ledford and Tawn (1996) takes $R_\alpha = R_{11}$ as in definition (2) where $u$ is chosen to be a high quantile of the unit Fréchet distribution, and so assumes that the distribution function (3) holds only within a joint tail region of $R$. In order to obtain score tests of independence under both approaches we need to consider the corresponding likelihood functions. We will denote the log-likelihoods under the Tawn (1988) and Ledford and Tawn (1996) approaches by $l_n^{(1)}$ and $l_n^{(2)}$ respectively.

Under the Tawn (1988) approach, since the distribution function (3) is assumed to hold over the whole of $R$, the likelihood contribution of each observation is the joint density $f_\alpha(x,y) = \partial^2 F(x,y)/\partial x \partial y$, and thus $l_n^{(1)}(\alpha) = \sum_{i=1}^n \log f_\alpha(X_i, Y_i)$. The corresponding score function for $\alpha = 1$ is given by

$$\begin{aligned} U_n^{(1)} &= \left\{ \frac{\partial}{\partial \alpha} l_n^{(1)}(\alpha) \right\}\bigg|_{\alpha=1} \\ &= \sum_{i=1}^n \left\{ \begin{aligned} &(1 - X_i^{-1}) \log X_i + (1 - Y_i^{-1}) \log Y_i \\ &\quad + (2 - X_i^{-1} - Y_i^{-1}) \log (X_i^{-1} + Y_i^{-1}) - (X_i^{-1} + Y_i^{-1})^{-1} \end{aligned} \right\}. \end{aligned}$$

Under the Ledford and Tawn (1996) approach, the model for $R_{11}$ is extended to regions $R_{00}$, $R_{01}$ and $R_{10}$ by partial and full censoring over these regions. The resulting log-likelihood comprises contributions that are region dependent and is given by

$$l_n^{(2)}(\alpha) = \sum_{i=1}^n \log \left\{ \begin{aligned} &I\{(X_i, Y_i) \in R_{00}\} F(u,u) + I\{(X_i, Y_i) \in R_{10}\} \partial F(X_i, u)/\partial x \\ &\quad + I\{(X_i, Y_i) \in R_{01}\} \partial F(u, Y_i)/\partial y + I\{(X_i, Y_i) \in R_{11}\} f_\alpha(X_i, Y_i) \end{aligned} \right\}$$

where $\partial F(u, Y_i)/\partial y$ means the partial derivative with respect to $y$ of $F(x, y)$ evaluated at $(x, y) = (u, Y_i)$. Proceeding as above, the corresponding score function for $\alpha = 1$ is given by

$$U_n^{(2)} = \sum_{i=1}^n \sum_{k,l \in \{0,1\}} I\{(X_i, Y_i) \in R_{kl}\} S_{kl}(X_i, Y_i)$$

where

$$\begin{aligned} S_{00}(x,y) &= -2u^{-1} \log 2 \\ S_{01}(x,y) &= -u^{-1} \log u + (1 - y^{-1}) \log y + (1 - u^{-1} - y^{-1}) \log (u^{-1} + y^{-1}) \\ S_{10}(x,y) &= -u^{-1} \log u + (1 - x^{-1}) \log x + (1 - x^{-1} - u^{-1}) \log (x^{-1} + u^{-1}) \\ S_{11}(x,y) &= (1 - x^{-1}) \log x + (1 - y^{-1}) \log y + (2 - x^{-1} - y^{-1}) \log (x^{-1} + y^{-1}) \\ &\quad - (x^{-1} + y^{-1})^{-1}. \end{aligned} \tag{4}$$

Note that the likelihood contributions obtained under both the Tawn (1988) and Ledford and Tawn (1996) approaches coincide for observations in region $R_{11}$, and the same is true for the corresponding score contributions.

When $\alpha = 1$, i.e., when the marginal variables are independent, it can be shown that both $U_n^{(1)}$ and $U_n^{(2)}$ have expectation zero, consistent with regular likelihood theory, and infinite variance, which is inconsistent with regular likelihood theory. This infinite variance produces non-regular behaviour in the score statistics based on $U_n^{(1)}$ and $U_n^{(2)}$, as shown by the following:

**Proposition 1 (Tawn, 1988; Ledford and Tawn, 1996):** *Let* $c_n = \{(n \log n)/2\}^{\frac{1}{2}}$. *If the marginal variables are independent, then, as* $n \to \infty$, *both of the following hold:*

$$-U_n^{(1)}/c_n \xrightarrow{w} N(0,1) \quad \textit{and} \quad -U_n^{(2)}/c_n \xrightarrow{w} N(0,1),$$

*where* $\xrightarrow{w}$ *denotes convergence in distribution.*

The minus signs in the above test statistics are present so that positive dependence yields positive values of the test statistics. The normalisation here, $c_n = \{(n \log n)/2\}^{\frac{1}{2}}$, is heavier than that of regular cases (which have normalisation $n^{\frac{1}{2}}$) and provides the extra scaling necessary to counter the infinite variance of $U_n^{(1)}$ and $U_n^{(2)}$ in order to obtain a normal limit. It is important to note that these asymptotic results say nothing about the rate of convergence of $c_n^{-1}U_n^{(1)}$ and $c_n^{-1}U_n^{(2)}$ to standard normal distributions. For the above tests to be most useful in practice, we would hope for fast convergence, but simulation shows that this is not the case and in fact convergence to the standard normal distribution is very slow in both cases. Thus, to apply these tests in practical situations extensive simulations may be required to evaluate the appropriate critical points of their null distributions for each sample size of interest. Similar non-regular and practically problematic behaviour is encountered for tests based on the mixed model dependence structure under both the Tawn (1988) and the Ledford and Tawn (1996) approaches.

### *3.1. Regularised score tests of independence*

From a practical perspective, implementation of the test statistics given in Proposition 1 is problematic because of their slow convergence to standard normal distributions, a property that arises because of the infinite variances of $U_n^{(1)}$ and $U_n^{(2)}$. Careful analysis shows that it is the presence of the $(x^{-1} + y^{-1})^{-1}$ term in the score function for a region which extends to $(\infty, \infty)$ that leads to the infinite variance in both cases. This, in turn, may be attributed to the joint density $f_\alpha(x, y)$ being the likelihood contribution of observations within such a region. Repeating the analysis for the mixed model shows that the infinite variance of the score statistic again is attributable to the joint density being the likelihood contribution for this region. Our proposal for obtaining score tests with regular normalisation is to change this feature of the Tawn (1988) and Ledford and Tawn (1996) frameworks by censoring region $R_{11}$ under both approaches, so that the only information exploited within the resulting likelihoods about the observations in $R_{11}$ is how many of them there are. With this censoring in place we obviously lose information, but this is done as a compromise to obtain tests that have much faster convergence to standard normal distributions and are thus more applicable in practice. A contrasting technique would be to modify the test statistics by discarding the term with the infinite variance, an approach investigated by Crowder (1990), Kimber and Zhu (1999) and Kimber et al. (2005), however this results in a test statistic with non-zero expectation.

Applying this censoring procedure to the Tawn (1988) framework we obtain the following log-likelihood

$$l_n^{(1*)}(\alpha) = \sum_{i=1}^{n} \left[ I\{(X_i, Y_i) \notin R_{11}\} f_\alpha(X_i, Y_i) + I\{(X_i, Y_i) \in R_{11}\} \overline{F}(u, u) \right] \tag{5}$$

where $\overline{F}(u,u) = \Pr(X > u, Y > u) = 1 - 2\exp(-u^{-1}) + \exp(-2^\alpha u^{-1})$. The corresponding score function for $\alpha = 1$ is given by

$$U_n^{(1*)} = \sum_{i:(X_i,Y_i)\notin R_{11}} \left\{ \begin{array}{l} (1 - X_i^{-1})\log X_i + (1 - Y_i^{-1})\log Y_i \\ \quad + (2 - X_i^{-1} - Y_i^{-1}) \log (X_i^{-1} + Y_i^{-1}) - (X_i^{-1} + Y_i^{-1})^{-1} \end{array} \right\}$$
$$+ \frac{2u^{-1} \log 2 \ \exp(-2u^{-1}) N_{11}}{2\exp(-u^{-1}) - \exp(-2u^{-1}) - 1}$$

where $N_{11}$ denotes how many of the $n$ observations fall in region $R_{11}$ Similarly, censoring $R_{11}$ under the Ledford and Tawn (1996) framework, we obtain the log-likelihood

$$l_n^{(2*)}(\alpha) = \sum_{i=1}^{n} \log \left\{ \begin{array}{l} I\{(X_i, Y_i) \in R_{00}\} F(u, u) + I\{(X_i, Y_i) \in R_{10}\} \partial F(X_i, u)/\partial x \\ + I\{(X_i, Y_i) \in R_{01}\} \partial F(u, Y_i)/\partial y + I\{(X_i, Y_i) \in R_{11}\} \overline{F}(u, u) \end{array} \right\} \tag{6}$$

and hence the score statistic

$$U_n^{(2*)} = \sum_{i:(X_i,Y_i)\notin R_{11}} I\{(X_i, Y_i) \in R_{kl}\} S_{kl}(X_i, Y_i) + \frac{2u^{-1} \log 2 \ \exp(-2u^{-1}) N_{11}}{2\exp(-u^{-1}) - \exp(-2u^{-1}) - 1}$$

for $S_{kl}$ as defined in equation (4). The score functions $U_n^{(1*)}$ and $U_n^{(2*)}$ have zero expectation, as before, but more importantly for our purposes have finite variances, which we denote by $n\sigma_1^2$ and $n\sigma_2^2$ respectively. More precisely, we have the following:

**Proposition 2:** *If the variables are independent, then, as $n \to \infty$, both of the following hold:*

$$-U_n^{(1*)}/\sqrt{n\sigma_1^2} \xrightarrow{w} N(0,1) \quad \text{and} \quad -U_n^{(2*)}/\sqrt{n\sigma_2^2} \xrightarrow{w} N(0,1)$$

*where $\sigma_1^2$ and $\sigma_2^2$ denote the variances of the corresponding modified score statistics.*

Thus using the logistic dependence structure we have obtained tests of independence with regular normalisation. In Figure 1 the convergence rates of these test statistics are examined informally and compared to those of the existing Tawn (1988) and Ledford and Tawn (1996) approaches. The test statistics corresponding to the unmodified approaches have empirical distributions that are some considerable way from their limiting $N(0,1)$ law, whereas those corresponding to the modified approaches are much closer to the $N(0,1)$ distribution. The results here are for the single sample size $n = 300$, but additional

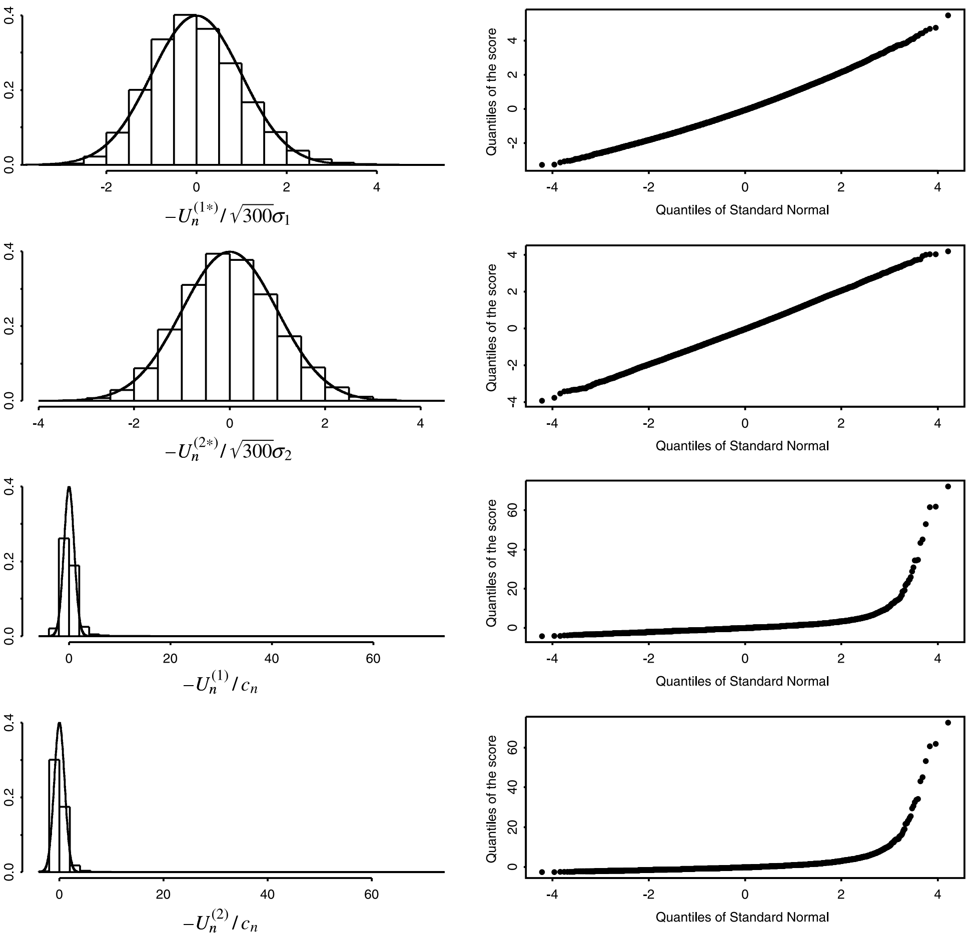


*Figure 1.* Histograms, with standard normal densities superimposed, and standard normal QQ-plots showing the empirical distributions of score statistics based on the logistic dependence structure for $U_n^{(1*)}$ (top), $U_n^{(2*)}$, $U_n^{(1)}$ and $U_n^{(2)}$ (bottom) under independence of the marginal distributions. The sample size is $n = 300$ in each case, and the threshold that defines the boundaries of $R_{11}$ is the 90% quantile of the unit Fréchet distribution, i.e., $u = -1/\log(0.9)$. The results shown were obtained through 20,000 repeated simulations.

simulations (not reported) show that these findings remain true even for much larger sample sizes. Similar results may be obtained for tests based on the mixed dependence structure. These findings suggest that our $R_{11}$ censoring modification provides tests with improved convergence to $N(0,1)$ that should be much more straightforward to implement practically.

For the $d$-dimensional logistic and mixed models, infinite variance of the score statistic arises because the likelihood contribution for a point in the region extending to $(\infty, \ldots, \infty)$ contains a term obtained from the joint density

$$\frac{\partial^d F(x_1, \ldots, x_d)}{\partial x_1 \ldots \partial x_d}.$$

Censoring the region where all margins exceed $u$, as was done in the bivariate case, modifies this contribution and yields a score test with finite variance for both of these dependence models.

### 3.2. *Evaluating $\sigma_1$ and $\sigma_2$ for the logistic dependence structure*

In order to construct the test statistics given in Proposition 2, values for the standard deviations $\sigma_1$ and $\sigma_2$, which depend implicitly on the value of the threshold $u$, are required. Detailed numerical integration is needed in order to calculate these standard deviations accurately. Clearly, the harder it is to evaluate these quantities for a given threshold $u$ then the more problematic it will be to implement the corresponding test statistics. To overcome this potential problem, we calculated $\sigma_1$ and $\sigma_2$ using accurate numerical quadrature for a range of $u$ values, and then examined several easily computed nonlinear approximations of the observed relationships. Corresponding approximations are given for the mixed model in the appendix.

For simplicity, we will work with the threshold probability $p = F(u) = \exp(-1/u)$ rather than with the threshold $u$ directly and will consider values of $p$ satisfying $p \in [0.75, 1 - \epsilon]$ for some small $\epsilon \geq 0$. Taking $\epsilon = 10^{-7}$, we found that the standard deviation $\sigma_1$ is well approximated by the function

$$\begin{aligned}\hat{\sigma}_1(p) = 1.463193 + 0.312435 \log\{-\log(1-p)\} + 0.132315\,[\log\{-\log(1-p)\}]^2 \\ + 0.035713\,[\log\{-\log(1-p)\}]^3,\end{aligned}$$

and taking $\epsilon = 0$, that the relationship between $\sigma_2$ and $p$ is well approximated by

$$\hat{\sigma}_2(p) = 1.107767 + 0.362784 p - 0.084381 p^2.$$

The accurately calculated standard deviations together with these approximations are shown in Figure 2. Clearly the approximations provide high accuracy. Further analysis of the underlying analytic expressions shows that $\sigma_1 \to \infty$ and $\sigma_2 \to 2 \log 2$ as $u \to \infty$.

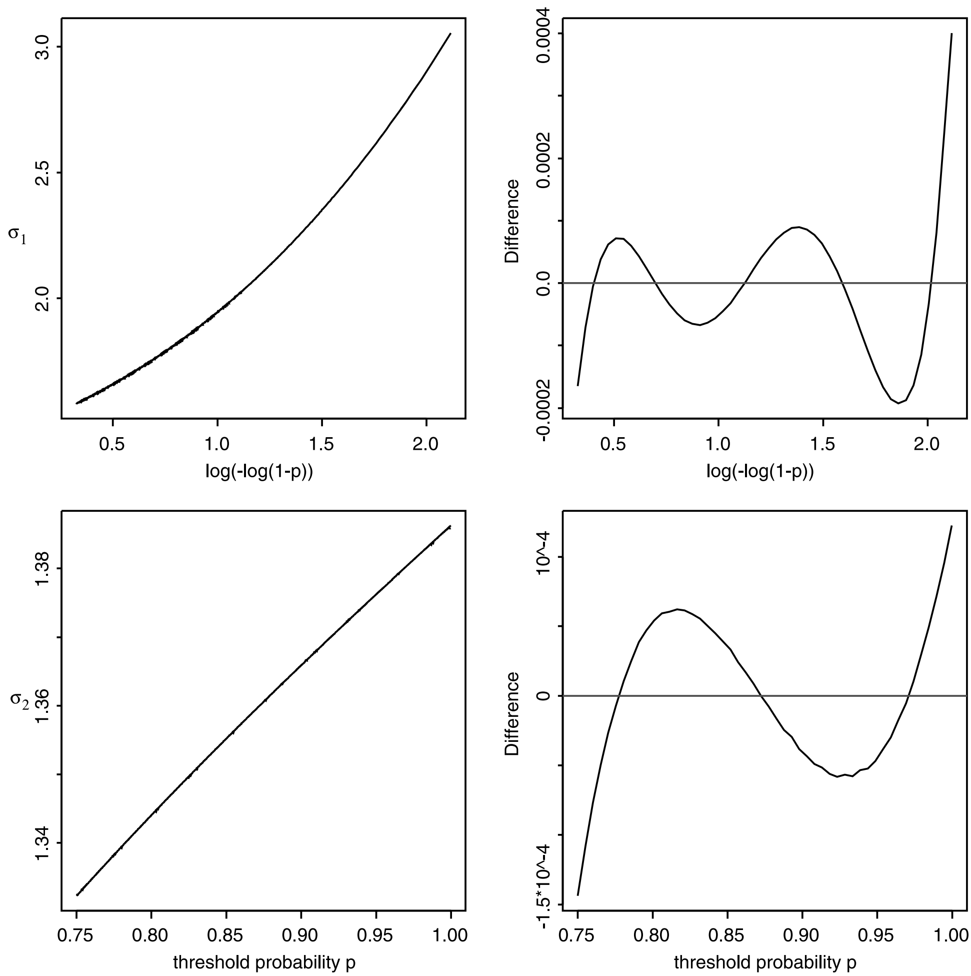


*Figure 2.* The left-hand panels show graphs of $\sigma_1$ and $\sigma_2$ for the logistic dependence structure as the threshold probability $p$ varies from 0.75 to $1–10^{-7}$ and from 0.75 to 1, respectively, with the corresponding approximations superimposed (dotted lines). In both left-hand panels the two lines nearly coincide, indicating that the approximations have a high degree of accuracy. The right-hand panels show the differences between the accurately calculated values and the approximations.

## 4. Properties of the tests

In this section we examine several features of the original and modified tests in order to better quantify the benefits provided by our censoring approach.

### 4.1. Convergence, power and comparison with Kendall's $\tau$

Our first assessment is based on comparing critical values of the null distributions of the logistic dependence structure test statistics under the original and modified frameworks for sample sizes typical of applications. We undertake this by simulating independent Fréchet pairs on $R$. Table 1 shows a range of critical values for the normalised scores based on $U_n^{(1)}$, $U_n^{(2)}$, $U_n^{(1*)}$ and $U_n^{(2*)}$ for threshold probabilities $p = 0.9$ and $p = 0.95$ and sample sizes $n = 1000$ and $n = 2500$. Comparing critical values under the modified and unmodified approaches, we see that the modified approach yields critical values that are always closer to the asymptotic values, and furthermore, that this benefit becomes more pronounced as increasingly extreme critical points are considered. The tests based on $U_n^{(1*)}$, which utilise more of the available information, appear to converge more quickly than those based on $U_n^{(2*)}$, as would be expected.

We now focus on the power functions of the modified and unmodified tests. We again use simulation, but now obtain dependent Fréchet pairs on $R$ using the scheme for generating from a bivariate extreme value distribution with logistic dependence outlined by Shi et al. (1992) and more recently Stephenson (2003). In Figure 3 we compare the power functions of the different score tests for threshold probabilities $p = 0.9$ and $p = 0.95$ and sample sizes $n = 1000$ and $n = 2500$ using the appropriate 95% empirical critical values given in Table 1. Near independence, i.e., $\alpha \approx 1$, the test based on $U_n^{(1)}$ has the highest power, and is followed in order of diminishing power by the $U_n^{(2)}$ and

*Table 1.* Simulated and asymptotic critical values of normalized score statistics based on the logistic dependence structure for **a**) $U_n^{(1)}$, **b**) $U_n^{(2)}$, **c**) $U_n^{(1*)}$, and **d**) $U_n^{(2*)}$. Standard errors are given in parenthesis. The simulation involved 5,000,000 replications of the normalised scores.

| | | | Critical points | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | 10% | 5% | 2.5% | 1% | 0.5% | 0.1% |
| **a**) | 1,000 | | 1.52 (0.01) | 2.17 (0.01) | 2.91 (0.03) | 4.23 (0.07) | 5.72 (0.15) | 12.26 (0.81) |
| | 2,500 | | 1.49 (0.01) | 2.12 (0.02) | 2.83 (0.03) | 4.07 (0.07) | 5.45 (0.15) | 11.47 (0.68) |
| **b**) | 1,000 | 0.9 | 1.25 (0.01) | 1.9 (0.02) | 2.67 (0.04) | 4.06 (0.08) | 5.6 (0.19) | 12.03 (0.81) |
| | | 0.95 | 1.13 (0.01) | 1.77 (0.02) | 2.57 (0.04) | 3.97 (0.09) | 5.52 (0.18) | 12.02 (0.80) |
| | 2,500 | 0.9 | 1.26 (0.01) | 1.87 (0.02) | 2.59 (0.03) | 3.88 (0.06) | 5.31 (0.15) | 11.33 (0.73) |
| | | 0.95 | 1.15 (0.01) | 1.76 (0.02) | 2.49 (0.03) | 3.8 (0.08) | 5.24 (0.16) | 11.59 (0.70) |
| **c**) | 1,000 | 0.9 | 1.29 (0.009) | 1.66 (0.01) | 1.99 (0.01) | 2.38 (0.02) | 2.64 (0.02) | 3.19 (0.05) |
| | | 0.95 | 1.29 (0.009) | 1.69 (0.01) | 2.04 (0.01) | 2.45 (0.02) | 2.74 (0.02) | 3.36 (0.04) |
| | 2,500 | 0.9 | 1.29 (0.009) | 1.66 (0.01) | 1.98 (0.01) | 2.36 (0.02) | 2.62 (0.02) | 3.15 (0.03) |
| | | 0.95 | 1.29 (0.009) | 1.67 (0.01) | 2.01 (0.01) | 2.4 (0.02) | 2.67 (0.02) | 3.24 (0.05) |
| **d**) | 1,000 | 0.9 | 1.3 (0.009) | 1.7 (0.01) | 2.05 (0.01) | 2.47 (0.02) | 2.75 (0.03) | 3.35 (0.05) |
| | | 0.95 | 1.33 (0.009) | 1.78 (0.01) | 2.18 (0.01) | 2.67 (0.02) | 3.02 (0.03) | 3.77 (0.07) |
| | 2,500 | 0.9 | 1.29 (0.009) | 1.68 (0.01) | 2.02 (0.01) | 2.42 (0.02) | 2.69 (0.03) | 3.26 (0.04) |
| | | 0.95 | 1.31 (0.009) | 1.73 (0.01) | 2.1 (0.01) | 2.55 (0.02) | 2.86 (0.03) | 3.51 (0.05) |
| | $\infty$ | | 1.28 | 1.645 | 1.96 | 2.33 | 2.57 | 3.09 |

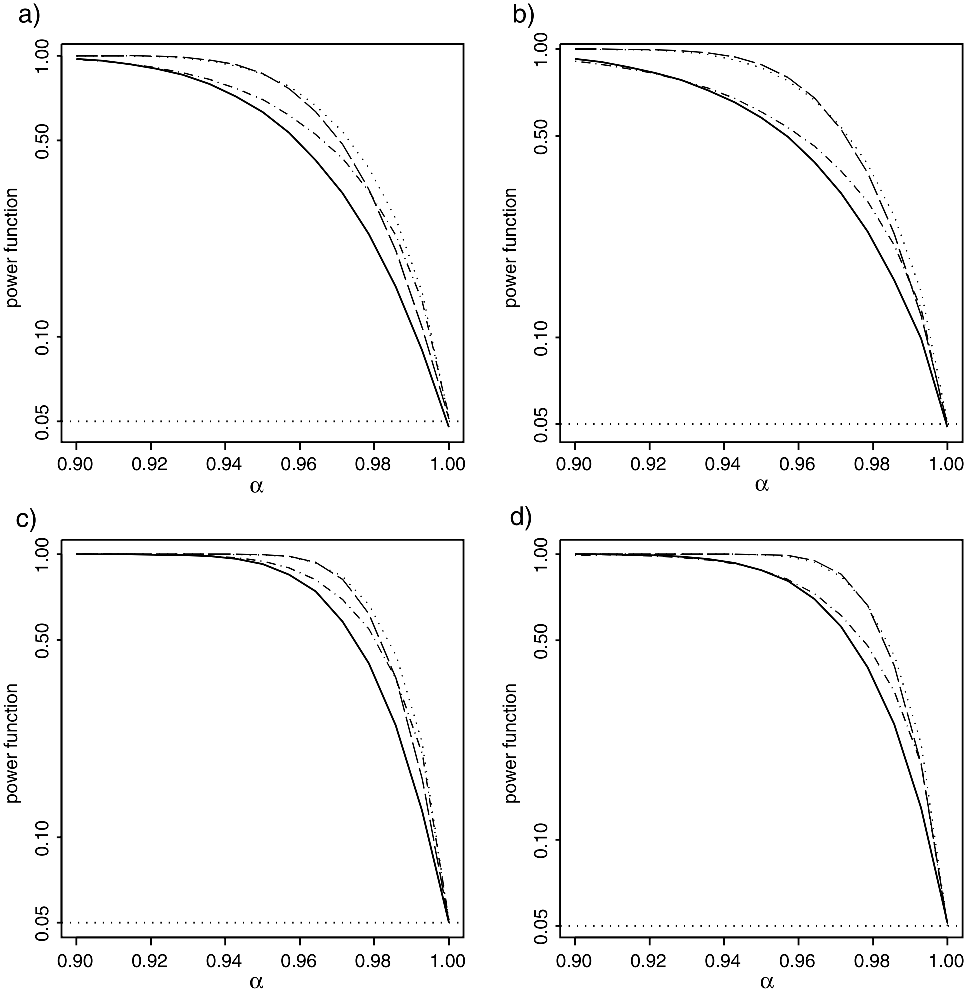


*Figure 3.* Power functions of normalized score statistics based on the logistic dependence structure for $U_n^{(1*)}$ (– – –), $U_n^{(1)}$ (· · ·), $U_n^{(2*)}$ (——) and $U_n^{(2)}$ (– • –), based on the 95% empirical critical values. The underlying data were simulated from a bivariate extreme value distribution on $R$ with unit Fréchet margins and logistic dependence with parameter $\alpha$. The thresholds and sample sizes used were: a) $n = 1{,}000$ and $p = 0.9$, b) $n = 1{,}000$ and $p = 0.95$, c) $n = 2{,}500$ and $p = 0.9$ and d) $n = 2{,}500$ and $p = 0.95$. The number of simulations was 4,000,000. The size of the tests (=0.05) is indicated by a horizontal dotted line in each panel.

$U_n^{(1*)}$ tests, with the $U_n^{(2*)}$ test having the least power. The loss of power resulting from censoring region $R_{11}$ is greatest for some $\alpha < 1$ and is of comparable magnitude to the loss of power between the unmodified Tawn (1988) and Ledford and Tawn (1996) tests. That the power function of $U_n^{(1*)}$ dominates that of $U_n^{(2*)}$ is as expected since the much stronger conditions that underpin $U_n^{(1*)}$ are consistent with the data generation scheme used here. Results under an alternative simulation scheme motivated by the modelling assumptions of $U_n^{(2*)}$ are reported later.

Recent work in Capéraà et al. (2000) and Genest and Rivest (2001) has examined Kendall's-$\tau$ as a measure of dependence within extremes. For comparison, we now examine the power functions of our regular score tests and tests based on Kendall's-$\tau$ using dependent data generated from the logistic model on $R$ for sample size $n = 1000$ and using the 95% point of the standard normal distribution as the critical value throughout, see Figure 4. The modified Tawn score test has greater power than the test based on Kendall's-$\tau$, and this in turn has greater power than the Ledford and Tawn modified score test. The dominance of the modified Tawn test over the Kendall-$\tau$ test is as expected since it is based on assumptions consistent with the underlying data generation method. The dominance of the Kendall-$\tau$ test over the modified Ledford and Tawn test is more interesting, and probably arises because the censoring of the modified Ledford and Tawn framework results in less of the available information being used. However, near independence, the test based on Kendall's-$\tau$ has the least power.

### 4.2. *The mixed model, robustness and marginal estimation*

Tawn (1988) and Ledford and Tawn (1996) report that the mixed model yields score and likelihood ratio tests of independence with essentially the same non-regular behaviour as

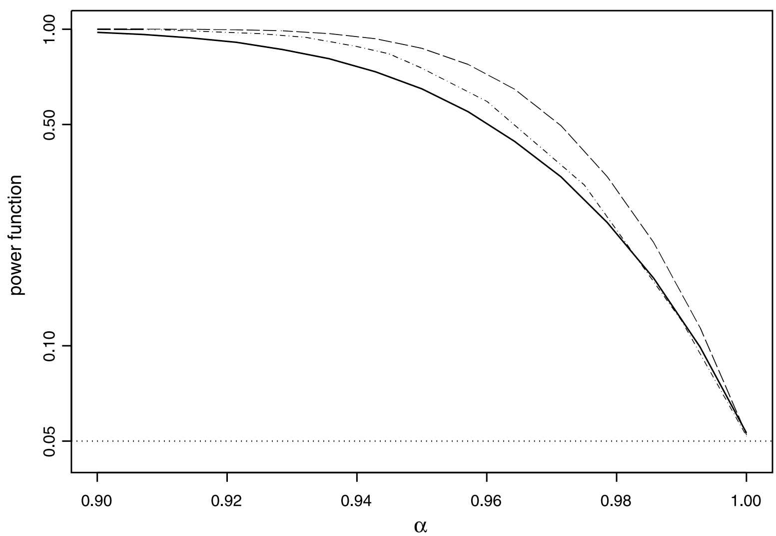


*Figure 4.* Power functions of normalised score statistics for the logistic dependence structure based on $U_n^{(1*)}$ (– – –), $U_n^{(2*)}$ (——) and Kendall's-$\tau$ (– • –). Dependent data were generated over $R$ using the logistic model with dependence parameter $\alpha$. The 95% asymptotic critical value was used and the threshold probability and sample size were $p = 0.9$ and $n = 1{,}000$ respectively. The number of simulations was 4,000,000.

those for the logistic model. Further investigation reveals that these score tests have slow convergence to $N(0,1)$. Employing the same threshold censoring scheme as used previously produces the finite variance score expressions given in Appendix A. From these it is straightforward to obtain score tests of independence based on the mixed model which have faster convergence to $N(0,1)$ than those of Tawn (1988) and Ledford and Tawn (1996).

Here we examine how the modified logistic and mixed score tests perform when evaluated on unit Fréchet marginally distributed dependent data from the following range of bivariate structures:

**A:** the extreme value logistic model,
**B:** the extreme value mixed model,
**C:** the extreme value asymmetric logistic model (see Stephenson, 2003) with dependence function

$$V(x,y) = \left\{ \left(\frac{x}{\theta}\right)^{-1/\alpha} - \left(\frac{y}{\phi}\right)^{-1/\alpha} \right\}^{\alpha} - \frac{1-\theta}{x} - \frac{1-\phi}{y}$$

where $\theta$, $\phi$, $\alpha \in (0,1]$, for $V$ as in Section 2, and
**D:** the joint lower tail of the Clayton (1978) distribution given by

$$F(x,y) = F(x) + F(y) - 1 + \left[ \{\overline{F}(x)\}^{-1/\alpha} + \{\overline{F}(y)\}^{-1/\alpha} - 1 \right]^{-\alpha},$$

where $\alpha > 0$ and $F(x)$ and $\overline{F}(x)$ represent the unit Fréchet distribution and survivor functions respectively.

Case **C**, which provides an asymmetric data generation mechanism, and case **D**, which provides a data generation mechanism in the domain of attraction of a bivariate extreme value distribution, are included in order to assess informally the robustness of the various tests. Power functions obtained under these scenarios are shown in Figure 5 and suggest that the modified logistic and mixed score tests behave similarly to each other and are able to detect departures from independence that are outside the underlying parametric family from which they were derived. This is useful in practice as for many extreme value dependence models, e.g., the asymmetric logistic given above, it is not possible to obtain tractable tests using either the usual score statistic approach or our modification. Our findings also demonstrate that rejecting independence using score tests based on either the logistic or mixed models does not mean that either of these dependence models is appropriate for describing the structure of the underlying data. In this situation, the much harder question of how to model the observed dependence remains open. We also note that since the original and modified score tests are all based on families within the class of MEV dependence structures, they can reject independence when evaluated on

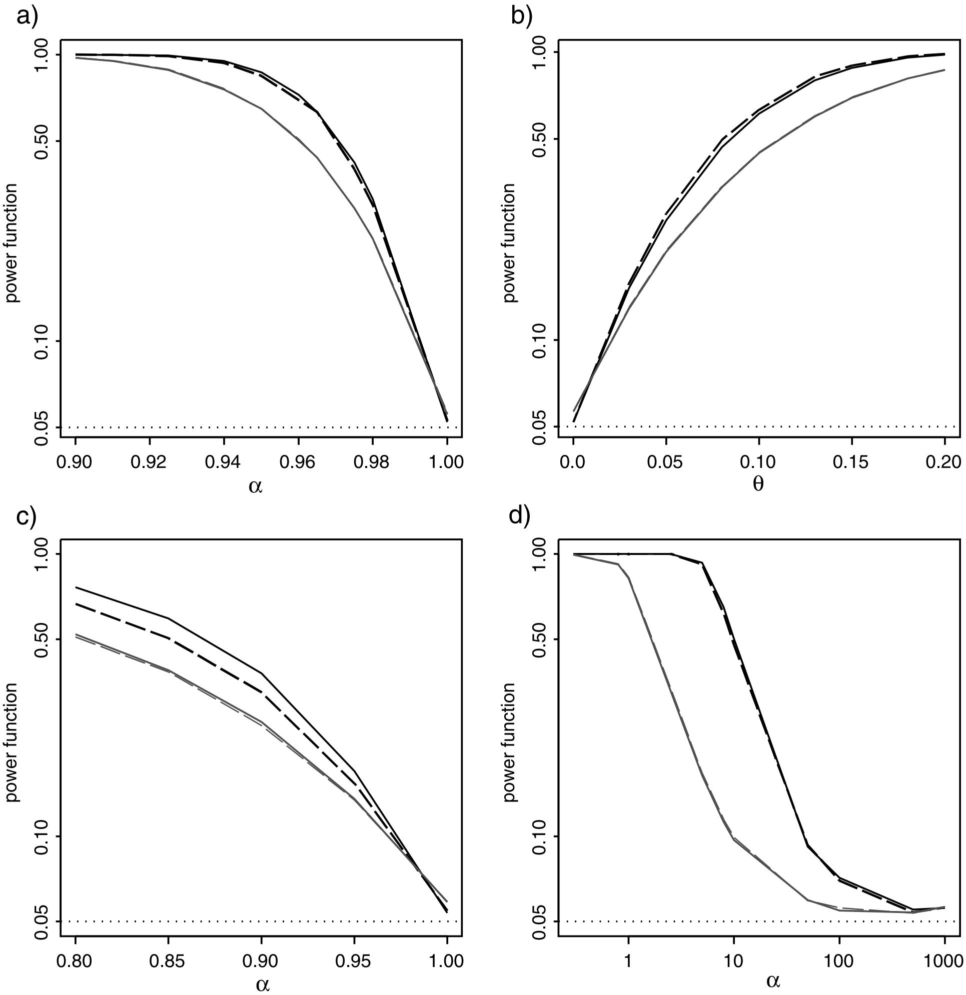


*Figure 5.* Power functions of normalised score statistics based on the logistic $U_n^{(1*)}$ (——), the mixed model version of $U_n^{(1*)}$ (– – –), the logistic $U_n^{(2*)}$ (——, in grey) and the mixed model version of $U_n^{(2*)}$ (– – –, in grey). The underlying data were generated over $R$ from the a) bivariate logistic model with dependence parameter $\alpha$; b) bivariate mixed model with dependence parameter $\theta$; c) asymmetric logistic model with parameters $\alpha$ and $(\theta, \phi) = (0.1, 0.9)$; and d) Clayton distribution with dependence parameter $\alpha$. The threshold probability was $p = 0.9$, the sample size was $n = 1{,}000$ and the asymptotic 95% critical value was used. The number of simulations was 4,000,000.

asymptotically independent data. For a detailed treatment of this point see Proposition 2 of Ledford and Tawn (1996).

So far the simulations that underpin our power function calculations have been based on generating unit Fréchet marginally distributed bivariate data from a dependence

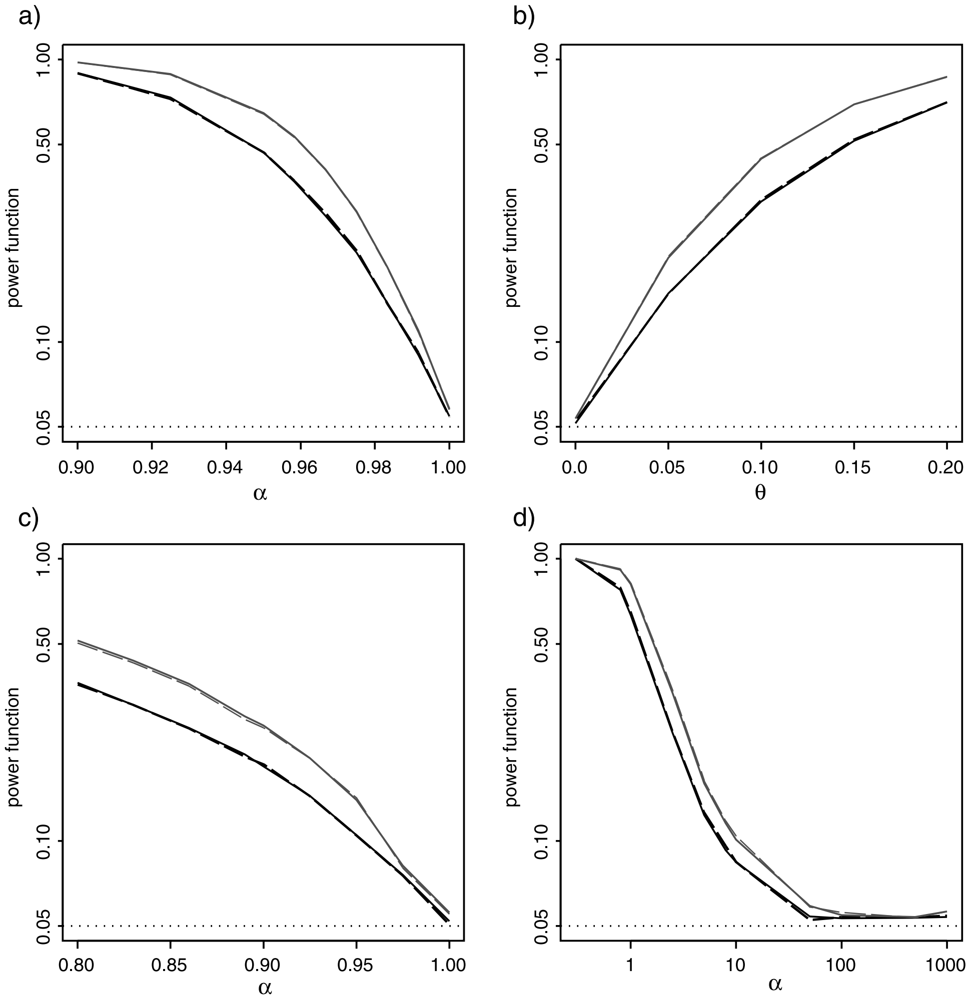


*Figure 6.* Power functions of normalised score statistics based on the logistic $U_n^{(1*)}$ (——), the mixed model version of $U_n^{(1*)}$ (– – –), the logistic $U_n^{(2*)}$ (——, in grey) and the mixed model version of $U_n^{(2*)}$ (– – –, in grey). The underlying data were generated in order to reflect the model assumptions of $U_n^{(2*)}$ but not the stronger model assumptions of $U_n^{(1*)}$, and simulated from the a) bivariate logistic model with dependence parameter $\alpha$; b) bivariate mixed model with dependence parameter $\theta$; c) asymmetric logistic model with parameters $\alpha$ and $(\theta, \phi) = (0.1, 0.9)$; and d) Clayton distribution with dependence parameter $\alpha$. The threshold probability was $p = 0.9$, the sample size was $n = 1{,}000$ and the asymptotic 95% critical value was used. The number of simulations was 4,000,000.

model that extends over the whole of $R$. As might be expected, this approach favours the $U_n^{(1*)}$ tests over the $U_n^{(2*)}$ tests as the modelling framework used for $U_n^{(1*)}$ is closer in structure to the data generation mechanism. An alternative basis for constructing power functions is to simulate data with unit Fréchet margins and a specified dependence structure only in region $R_{11}$, so that the strong model assumptions of $U_n^{(1*)}$ are not satisfied but the weaker assumptions of $U_n^{(2*)}$ still hold. The following describes such a scheme:

1. Generate $n$ points over $R$ from the joint model of interest (that has unit Fréchet margins).
2. Replace the $y$ values of points in $R_{10}$ with independent Fréchet values restricted to $(0,u)$.
3. Replace the $x$ values of points in $R_{01}$ with independent Fréchet values restricted to $(0,u)$.
4. Replace the $x$ and $y$ values of points in $R_{00}$ with independent Fréchet values restricted to $(0,u)$.

Figure 6 shows power functions for the logistic $U_n^{(1*)}$ and $U_n^{(2*)}$ tests and their mixed model versions obtained under this scheme for the dependence structures **A** to **D** used previously. As expected, the $U_n^{(2*)}$ tests now dominate the $U_n^{(1*)}$ tests, illustrating that tests based on the weaker modelling assumptions may be advantageous when the rather restrictive assumptions of the Tawn (1988) approach are not justified. We note that the censoring employed within the Ledford and Tawn (1996) approach ensures that the true power functions of the $U_n^{(2*)}$ tests are invariant to the

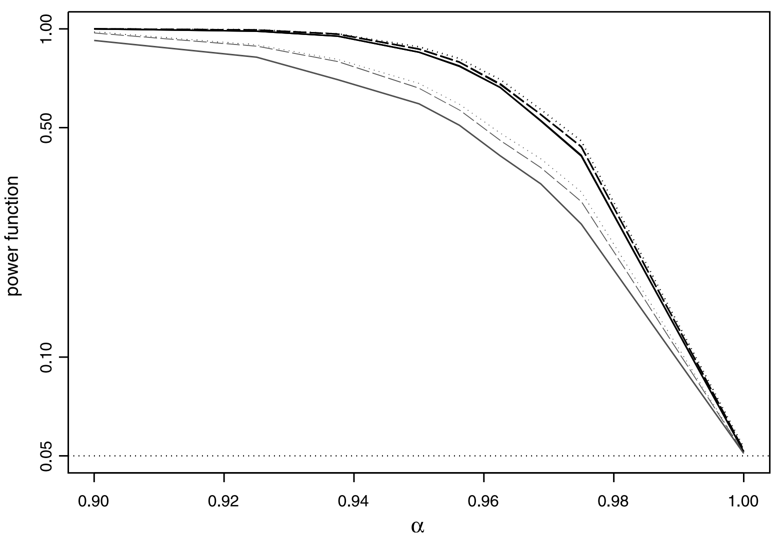


*Figure 7.* Power functions of normalised logistic score tests based on $U_n^{(1*)}$ (black lines) and $U_n^{(2*)}$ (grey lines). The underlying data have bivariate logistic dependence on $R$ with dependence parameter $\alpha$, and either unit Fréchet or Gumbel margins. The lines are as follows: using the GPD for underlying unit Fréchet margins (——), using the GPD for underlying Gumbel margins (– – –), and using the empirical distribution function ( · · · ). Results are identical for the empirical distribution function method for both the unit Fréchet and Gumbel cases due to the underlying data generation mechanism. The threshold probability was $p = 0.9$, the sample size was $n = 1{,}000$ and the asymptotic 95% critical value was used. The number of simulations was 4,000,000.

above region dependent transformations. Thus the strong similarity of the $U_n^{(2*)}$ power functions in Figures 5 and 6 is as expected.

Until now we have assumed that the marginal distributions are known to be unit Fréchet, however this is clearly unrepresentative of applications. In practice, it is necessary to transform marginally the observed data to unit Fréchet via probability integral transformations. These transformations can either be based on marginal empirical distribution functions or a more detailed approach involving fitting a tail model such as the generalised Pareto distribution (GPD). We examine these approaches for two different marginal distributions. Our first scheme samples from a bivariate extreme value distribution with logistic dependence structure and unit Fréchet margins. The data for our second scheme are obtained by taking natural logarithms of the points generated via the first scheme, and thus come from a bivariate extreme value distribution with logistic dependence structure and Gumbel margins. We use maximum likelihood to fit the GPD to tail data in each margin. Power functions constructed under both approaches for the two different marginal distributions are shown in Figure 7. The power functions obtained are seen to be very similar regardless of whether the empirical distribution or a fitted GPD is used as the basis of the marginal transformation. In fact the empirical distribution method appears to dominate the GPD method in our findings. This demonstrates that the more detailed approach of modelling the marginal tails has little to add here and our recommendation would be to use the empirical distribution in applications.

## 5. Conclusions

The proposed regularised score tests have been shown to have good performance in practice and to overcome the convergence difficulties encountered when using the existing non-regular score tests. Our test statistics converge rapidly to standard normal distributions and are consequently more useful in practical situations with realistic amounts of data. This benefit is at the expense of some reduction in power, but this loss of power has been shown to be of an acceptable magnitude. Compared to the Tawn (1988) and the Ledford and Tawn (1996) score tests, our new score tests require less extensive simulations in order to estimate actual critical values to a given precision, and have closer agreement between their actual and nominal sizes when asymptotic critical values are used.

Another approach we investigated is censoring the data in the whole plane according to the four regions used previously. Obviously, the only information such a censoring scheme conveys is the number of points which fall into each region. The likelihood contribution given by an observation in $R_{kl}$ is then just the probability of falling in that region, i.e., $F(u,\ u)$ for $R_{11}$, $F(u)\ -\ F(u,\ u)$ for $R_{01}$ and $R_{10}$, and $\overline{F}(u,u) = \Pr(X > u, Y > u)$ for $R_{11}$. This approach results in a regular score test too, and it can be shown that the variance of the resulting score statistic for the logistic dependence structure converges to $4 \log^2 2$ as $u \to \infty$. This approach can be used to obtain a test of independence within the class of MEV distributions based on counts.

**Acknowledgments**

We wish to thank the referees for their constructive comments. The first author is grateful to FCT/PRAXIS XXI/FEDER for their support of this work.

**Appendix**

**A: The mixed model**

Censoring region $R_{11}$ within the Tawn (1988) framework for the mixed model with dependence function $V(x, y) = x^{-1} + y^{-1} - \theta(x + y)^{-1}$, for $0 \leq \theta \leq 1$, the score function for $\theta = 0$ satisfies

$$U_n^{(m1*)} = \sum_{i:(X_i,Y_i)\notin R_{11}} \left\{ \begin{array}{l} (x+y)^{-1} - (x^2+y^2)(x+y)^{-2} \\ +2(xy)^2(x+y)^{-3} \end{array} \right\} + \frac{(2u)^{-1}\exp(-2u^{-1})N_{11}}{1-2\exp(-u^{-1})+\exp(-2u^{-1})}$$

where $N_{11}$ denotes how many of the $n$ observations fall in region $R_{11}$. Censoring $R_{11}$ under the Ledford and Tawn (1996) framework the score function for $\theta = 0$ is given by

$$U_n^{(m2*)} = \sum_{i=1}^{n} \sum_{k,l\in\{0,1\}} I\{(X_i, Y_i) \in R_{kl}\} S_{kl}^{(m)}(X_i, Y_i)$$

where

$$S_{00}^{(m)}(x,y) = (2u)^{-1} \qquad S_{10}^{(m)}(x,y) = (x+u)^{-1} - x^2(x+u)^{-2}$$

$$S_{01}^{(m)}(x,y) = (u+y)^{-1} - y^2(u+y)^{-2} \quad S_{11}^{(m)}(x,y) = \frac{(2u)^{-1}\exp(-2u^{-1})}{1-2\exp(-u^{-1})+\exp(-2u^{-1})}.$$

As for the logistic model, the score functions $U_n^{(m1*)}$ and $U_n^{(m2*)}$ have zero expectation and finite variances. Hence Proposition 2 also holds for these score statistics. Corresponding approximations for the standard deviations are as follows:

$$\hat{\sigma}_{m1} = 0.4888141 + 0.150803\log\{-\log(1-p)\} + 0.03601739[\log\{-\log(1-p)\}]^2$$
$$+ 0.01425557[\log\{-\log(1-p)\}]^3, \text{ and}$$
$$\hat{\sigma}_{m2} = 0.3178286 + 0.198727p - 0.01658025p^2,$$

for the threshold probability $p \in [0.75, 1 - 10^{-7}]$. The maximum discrepancies that arise between the approximations and the accurately calculated standard deviations are $6 \times 10^{-4}$ for $\sigma_{\mathrm{m1}}$ and $3 \times 10^{-5}$ for $\sigma_{\mathrm{m2}}$, for $p$ within the respective range. Further analysis of the analytic expressions for the standard deviations shows that $\sigma_{\mathrm{m1}} \to \infty$ and $\sigma_{\mathrm{m2}} \to 1/2$ as $u \to \infty$.

## B: Further asymptotic properties of likelihood inference

The asymptotic distribution of the maximum likelihood estimator of the dependence parameter $\alpha$ for the logistic model is noted here. We also examine results for the case when the marginal distributions are not known to be unit Fréchet, see Smith (1985), and consider the related issue of likelihood ratio tests under our modified framework. Corresponding results under the original framework are given in Tawn (1988) and Ledford and Tawn (1996). In practice, tests based on the maximum likelihood estimator or the likelihood ratio are harder to implement than score tests as numerical optimisation is required. However they do provide an important theoretical complement.

**Theorem 3:** *For the log-likelihoods given in equations* (5) *and* (6)*, under independence* (*i.e.,* $\alpha = 1$)*, we have that* $\hat{\alpha}$ *satisfies*

$$(1-\hat{\alpha})\sigma_1\sqrt{n} \stackrel{w}{\rightarrow} Z \quad \text{and} \quad (1-\hat{\alpha})\sigma_2\sqrt{n} \stackrel{w}{\rightarrow} Z$$

*as* $n \rightarrow \infty$*, where the non-negative random variable* $Z$ *has law*

$$\Pr(Z \leq z) = h(z)\Phi(z) \tag{7}$$

*for* $h(\cdot)$ *the Heaviside step function and* $\Phi(\cdot)$ *the standard normal distribution function.*

**Proof:** By Theorem 2 of Self and Liang (1987). □

When the marginal distributions are no longer known to be unit Fréchet the issue of joint estimation of the marginal and dependence parameters arises. Let the marginal distribution have parameters $\phi = (\phi^1, \ldots, \phi^q)$ and let $\phi_0$ be the vector of true marginal parameters. Denote the joint maximum likelihood estimator by $(\hat{\alpha}, \hat{\phi})$ and let $I(1, \phi_0)$ denote the expectation of $n^{-1}I_n(1, \phi_0)$ with respect to the true joint density, where $-I_n(1, \phi_0)$ is the matrix of second derivatives with respect to $\phi$ of the corresponding joint log-likelihood function evaluated at the point $(\alpha, \phi) = (1, \phi_0)$, i.e., the matrix with entries $\partial^2 l_n^{(k*)}(1, \phi_0)/\partial\phi^i\partial\phi^j$ for $i, j = 1, \ldots, q$ and $k = 1$ or 2 corresponding to the modified Tawn or modified Ledford and Tawn frameworks. The following regularity conditions (see Self and Liang, 1987) are required:

i) almost sure existence of the first three derivatives of $l_n^{(k*)}(\theta)$ with respect to $\theta = (\alpha, \phi)$ for $k = 1$ or 2 on the intersection of neighbourhoods of the true parameter value $\theta_0 = (1, \phi_0)$ and the parameter space $\Omega$ where $\theta$ takes values,
ii) on the intersection of neighbourhoods of $\theta_0$ and $\Omega$, $n^{-1}$ times the absolute value of the third derivative of $l_n^{(k*)}(\theta)$ is bounded by a function of $(X_1, Y_1), \ldots, (X_n, Y_n)$ whose expectation exists, and
iii) $I(1, \phi)$ is positive definite on neighbourhoods of $\phi_0$ and the $\phi$ parameter space can be approximated by a cone with vertex at $\phi_0$.

**Theorem 4:** *Under the regularity conditions above and under independence we have*

$$\left\{(1-\hat{\alpha})\sigma_k\sqrt{n}, (\hat{\phi}-\phi_0)\sqrt{n}\right\} \xrightarrow{w} (Z, Z_1, \ldots, Z_q) \text{ for } k = 1, 2$$

*as $n \to \infty$, where $Z$ is as defined in equation* (7)*, $Z_1, \ldots, Z_q$ are zero mean normal random variables and $Z$ is independent of $Z_1, \ldots, Z_q$. The covariance matrix of $(Z, Z_1, \ldots, Z_q)$ is block-diagonal with non-null entries $\{1, I^{-1}(1, \phi_0)\}$. Additionally, for the modified Tawn or Ledford and Tawn frameworks, letting $L_n^{(k*)}(\hat{\alpha})$ denote the maximum of the likelihood taken over the dependence parameter $\alpha \in (0, 1]$ and the marginal parameters, and $L_n^{(k*)}(1)$ the maximum taken over the marginal parameters under the constraint $\alpha = 1$, we have*

$$2\log\left\{L_n^{(k*)}(\hat{\alpha})/L_n^{(k*)}(1)\right\} \xrightarrow{w} Z^2 \text{ as } n \to \infty$$

*where $Z$ is as defined in equation* (7)*. When the marginal variables are known to be unit Fréchet then the same result holds except that $L_n^{(k*)}(\hat{\alpha})$ and $L_n^{(k*)}(1)$ denote the maximum of the likelihood taken over $\alpha$ and the value of the likelihood when $\alpha = 1$.*

## References


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

Capéraà, P., Fougères, A.-L. and Genest, C., "Bivariate distributions with given extreme value attractors," *J. Multivar. Anal.* 72, 30–49, (2000).

Clayton, D.G., "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika* 65, 141–151, (1978).

Coles, S.G., *An Introduction to Statistical Modelling of Extreme Values*, Springer, London, UK, 2001.

Coles, S.G. and Tawn, J.A., "Modelling extreme multivariate events," *J. R. Stat. Soc., B* 53, 377–392, (1991).

Coles, S.G. and Tawn, J.A., "Statistical methods for multivariate extremes: an application to structural design (with discussion)," *Appl. Stat.* 43, 1–48, (1994).

Crowder, M., "On some nonregular tests for a modified Weibull model," *Biometrika* 77, 449–506, (1990).

Davison, A.C. and Smith, R.L., "Models for exceedances over high thresholds (with discussion)," *J. R. Stat. Soc., B* 52, 393–442, (1990).

Deheuvels, P., "Some applications of the dependence functions to statistical inference: nonparametric estimates of extreme value distributions, and a Kiefer type universal bound for the uniform test of independence," *Colloq. Math. Societ. János Bolyai. 32. Nonparam. Stat. Infer., Budapest (Hungary)* 32, 183–201, (1980).

Deheuvels, P. and Martynov, G., "Cramér-Von Mises-Type tests with applications to tests of independence for multivariate extreme value distributions," *Commun. Stat., Theory Methods* 25(4), 871–908, (1996).

Dorea, C. and Miasaki, E., "Asymptotic test for independence of extreme values," *Acta Math. Hung.* 62(3–4), 343–347, (1993).

Embrechts, P., Klüppelberg, C. and Mikosch, T., *Modelling Extremal Events*, Springer, New York, 1997.

Genest, C. and Rivest, L.-P., "On the multivariate probability integral transform," *Stat. Probab. Lett.* 53, 391–399, (2001).

Heffernan, J.E. and Tawn, J.A., "A conditional approach for multivariate extreme values (with discussion)," *J. R. Stat. Soc., B* 66(Part 3), 497–546, (2004).

Joe, H., "Families of min-stable multivariate exponential and multivariate extreme value distributions," *Stat. Probab. Lett.* 9, 75–82, (1989).

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

Kimber, A.C. and Zhu, C.Q., "Diagnostic for a Weibull frailty model," in *Statistical Inference and Design of Experiments* (U.J. Dixit and M.R. Satam, eds.), Mumbai, Narosa, 36–46, (1999).

Kimber, A.C., Sarker, M.J. and Zhu, C.Q., "Some tests for frailty in bivariate Weibull lifetime data," Submitted, (2005).

Kotz, S. and Nadarajah, S., *Extreme Value Distributions*, Imperial College Press, London, UK, 2000.

Leadbetter, M.R., Lindgren, G. and Rootzén, H., *Extremes and Related Properties of Random Sequences and Series*, Springer, New York, 1983.

Ledford, A.W. and Tawn, J.A., "Statistics for near independence in multivariate extreme values," *Biometrika* 83, 169–187, (1996).

Ledford, A.W. and Tawn, J.A., "Modelling dependence within joint tail regions," *J. R. Stat. Soc., B* 59, 475–499, (1997).

Resnick, S.I., *Extreme Values, Regular Variation and Point Processes*, Springer, New York, 1987.

Self, S.G. and Liang, K.-Y., "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions," *J. Am. Stat. Assoc.* 82, 605–610, (1987).

Shi, D., Smith, R.L. and Coles, S.G., *Joint versus marginal estimation for bivariate extremes, Tech. Rep. 2074*, Department of Statistics, University of North Carolina, Chapel Hill, 1992.

Sibuya, M., "Bivariate extreme statistics," *Ann. Inst. Stat. Math.* 11, 195–210, (1960).

Smith, R.L., "Maximum likelihood estimation in a class of nonregular cases," *Biometrika* 72, 67–90, (1985).

Smith, R.L., "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion)," *Stat. Sci.* 4, 367–393, (1989).

Stephenson, A., "Simulating multivariate extreme value distributions of logistic type," *Extremes* 6, 49–59, (2003).

Stephenson, A. and Tawn, J.A., "Exploiting occurrence times in likelihood inference for componentwise maxima," *Biometrika* 92(1), 213–227, (2005).

Tawn, J.A., "Bivariate extreme value theory: models and estimation," *Biometrika* 75, 397–415, (1988).

Tiago de Oliveira, J. "Structure theory of bivariate extremes, extensions," *Est. Mat., Estat. e Econ* 7, 165–195, (1962/1963).