JOINT VERSUS MARGINAL ESTIMATION FOR BIVARIATE EXTREMES

Daoji Shi[(1)], Richard L. Smith[(2)] and Stuart G. Coles[(3)]

First version: February 1992

Summary

Bivariate extreme value distributions contain parameters of two types; those that define the marginal distributions, and parameters defining the dependence between suitably standardized variates. As an alternative to full maximum likelihood based on the joint distribution, we consider a "marginal estimation" method in which the margin and dependence parameters are estimated separately. This method is simpler to implement computationally, but may be inefficient. Asymptotic results allow the inefficiency to be quantified. The concepts are relevant to a large class of families of multivariate distributions, but the detailed analysis is restricted to Gumbel's logistic model with Gumbel or Generalized Extreme Value margins.

Keywords

Bivariate extremes, Fisher information matrix, Generalized Extreme Value distribution, Gumbel distribution, hydrological extremes, maximum likelihood.

(1) Department of Mathematics, Tianjin University, Tianjin 300072, China.
(2) Department of Statistics, University of North Carolina, Chapel Hill, N.C. 27599-3260, U.S.A. Please use this address for all correspondence.
(3) Department of Mathematics, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

1. INTRODUCTION

Bivariate extreme value distributions arise when one is interested in the joint distribution of extremes of two variables, for instance, the heights of a river at two neighbouring stations. The univariate margins are taken to be either the two-parameter Gumbel distribution

$$F(x;\mu,\sigma)=\exp\left[-\exp\{-(x-\mu)/\sigma\}\right] \tag{1.1}$$

(Gumbel 1958), or the three-parameter Generalized Extreme Value distribution

$$F(x;\mu,\sigma,\xi)=\exp\left[-\{1-\xi(x-\mu)/\sigma\}_+^{1/\xi}\right] \tag{1.2}$$

where $y_+ = \max(y,0)$. Equation (1.2) reduces to (1.1) when $\xi \to 0$; see Prescott and Walden (1980) and Smith (1985) for asymptotic estimation theory and Hosking (1985) for a Fortran algorithm to calculate maximum likelihood estimates.

There is a wide literature about bivariate and more generally multivariate extremes, but almost all the information we need is contained in Tawn (1988). Tawn reviewed a number of old and new models for bivariate extremes. A simple model which appears to be widely applicable is the *logistic* model due originally to Gumbel. This is given by the joint distribution function

$$\begin{aligned} F(x,y;\mu,\sigma,\nu,\tau,\alpha) &= \Pr\{X\le x, Y\le y\} \\ &= \exp\left[-\left\{e^{-(x-\mu)/\sigma\alpha}+e^{-(y-\nu)/\tau\alpha}\right\}^{\alpha}\right] \end{aligned} \tag{1.3}$$

in the case with Gumbel margins, or

$$F(x,y;\mu,\sigma,\xi,\nu,\tau,\eta,\alpha)=\exp\left[-\left\{(1-\xi(x-\mu)/\sigma)_+^{1/\alpha\xi}+(1-\eta(y-\nu)/\tau)_+^{1/\alpha\eta}\right\}^{\alpha}\right] \tag{1.4}$$

with Generalized Extreme Value margins. Here $\alpha \in [0,1]$ measures the dependence between X and Y, the limits $\alpha \to 1$, $\alpha \to 0$ corresponding respectively to independence and complete dependence (meaning that there is a deterministic relation between the two variables). Note that (1.3) arises from (1.4) by taking the limits $\xi \to 0, \eta \to 0$ in the same way as (1.2) leads to (1.1). Tawn considered the asymptotic theory of maximum likelihood, focussing in particular on the nonregular behaviour at $\alpha = 1$.

When $0 < \alpha < 1$, maximum likelihood estimators have the usual regular asymptotic properties, and asymptotically efficient estimators may be calculated by maximizing the full likelihood function with respect to all the parameters. We call this *joint estimation* because it is based on the joint distribution of X and Y. There is, however, another method that also seems appealing. This is to estimate μ, σ and ξ (if we do not assume $\xi = 0$) purely from the X values in the sample, ν, τ and η purely from the Y values, and then apply the transformation

$$\widehat{S} = \begin{cases} \exp\{-(X-\widehat{\mu})/\widehat{\sigma}\}, & \text{(Gumbel case)} \\ \{1-\widehat{\xi}(X-\widehat{\mu})/\widehat{\sigma}\}^{1/\widehat{\xi}}, & \text{(GEV case)} \end{cases}$$

$$\widehat{T} = \begin{cases} \exp\{-(Y-\widehat{\nu})/\widehat{\tau}\}, & \text{(Gumbel case)} \\ 1-\widehat{\eta}(Y-\widehat{\nu})/\widehat{\tau}\}^{1/\widehat{\eta}}, & \text{(GEV case)} \end{cases}$$

based on estimates $\widehat{\mu}, \widehat{\sigma}$, etc. If the estimates were replaced by their true values μ, σ, etc., then we could replace $\widehat{S}, \widehat{T}$ by S, T with joint distribution

$$P\{S > s, T > t\} = \exp\{-(s^{1/\alpha} + t^{1/\alpha})^{\alpha}\}, \quad s > 0, t > 0 \tag{1.5}$$

and it would be straightforward to estimate α by maximizing the one-parameter likelihood derived from (1.5). This suggests we do the same thing with the joint distribution of $(\widehat{S}, \widehat{T})$ ignoring the effect of estimating the marginal parameters. We call this method *marginal estimation*.

Computationally, it would simplify things a great deal if we could use marginal in place of joint estimation. One could use established algorithms such as Hosking's (1985) to estimate the margin parameters, and there would be no messy maximization in 5 or 7 dimensions. This advantage is even more clear-cut in more complicated models with more than one dependence parameter, or in multivariate extreme value distributions for more than two variables. However, no study has been made so far of the relative efficiency of marginal as compared with joint estimation. This is the principal motivation of the present paper.

There are some related issues connected with regional methods in hydrology (NERC 1975). Hydrologists typically try to improve their estimates by combining data from sites in the same region, assuming some or all the parameters to be common to the region. This raises questions of (a) testing for equality of parameters, (b) estimation of parameters under the assumption that they are equal at different sites. The traditional hydrological approach ignores dependence between sites. The methods of this paper allow for some theoretical analysis of the effects of that, though only in the case of two sites and logistic dependence. For a more practical approach see Smith (1992), and Coles and Walshaw (1992) for a similar problem involving wind directions.

The detailed study is restricted to (1.3) and (1.4), but the methodology is in principle applicable to any problem in which there are observed variables $(X_1, \ldots, X_p)$ and transformed variables $(S_1, \ldots, S_p)$ (where each S_i is a function of X_i, depending on unknown margin parameters), and the joint distribution of $(S_1, \ldots, S_p)$ depends on additional "dependence" parameters. All the multivariate extreme value distributions are of this form.

The study is also restricted to the traditional approach to extreme values based on the extreme value distributions. The alternative "threshold" approach propounded for univariate extremes by Davison and Smith (1990), has been extended to the bivariate case by Coles and Tawn (1991) and Joe, Smith and Weissman (1992). This raises similar issues of joint versus marginal estimation.

A key part of the analysis is the calculation of the Fisher information matrix for the model (1.3) or (1.4). In the case of (1.3), this has been done, with minor changes in notation, in a recent paper of Oakes and Manatunga (1992). The extension to (1.4) is considerably more complicated. However this appears to be worthwhile because, whereas our results show that the marginal method of estimation is satisfactory for (1.3), in the sense that the asymptotic variances of the two methods are generally very close, the same is not true of (1.4), where for certain combinations of parameter values, the marginal method can be very inefficient indeed. The latter result is supported by simulations; in fact it was simulations of the problem which originally prompted us to pursue the theoretical investigation in this paper.

2. COMPUTING THE FISHER INFORMATION

Let (X, Y) denote a bivariate random variable with joint distribution function $F(x, y)$ given by (1.4). Define the successive transformations

$$S = \{1 - \xi(X - \mu)/\sigma\}^{1/\xi}, \; T = \{1 - \eta(Y - \nu)/\tau\}^{1/\eta}, \tag{2.1}$$

$$S = Z\cos^{2\alpha} V, \; T = Z\sin^{2\alpha} V, \tag{2.2}$$

into first (S, T) and then (Z, V). In accordance with standard convention, upper-case roman letters denote random variables, and their lower-case equivalents denote numerical values. The joint density of (X, Y), written as a function of (s, t), is

$$\begin{aligned} f(x, y) &= \partial^2 F(x, y)/\partial x \partial y \\ &= (\sigma\tau)^{-1} s^{1/\alpha - \xi} t^{1/\alpha - \eta} (s^{1/\alpha} + t^{1/\alpha})^{\alpha - 2}. \\ &\quad . \left\{ (s^{1/\alpha} + t^{1/\alpha})^{\alpha} - 1 + 1/\alpha \right\} \exp\left\{ -(s^{1/\alpha} + t^{1/\alpha})^{\alpha} \right\}, \end{aligned} \tag{2.3}$$

defined when $s > 0, t > 0$. Alternatively, if we transform to (Z, V) and allow for the Jacobean of the transformation, we find that their joint density is

$$(\alpha z + 1 - \alpha) e^{-z} \sin 2v, \quad 0 < v < \pi/2, \; 0 < z < \infty \tag{2.4}$$

which shows that Z and V are independent with easily characterized distributions: V may be represented as (arc sin $U^{1/2}$), where U is uniform on (0, 1), while Z is the $1 - \alpha : \alpha$ mixture of a unit exponential random variable and the sum of two independent unit exponential random variables. This representation, apparently due to Lee (1979), was also the starting point for the paper by Oakes and Manatunga (1992). Amongst other things, it suggests an easy way to simulate from the distribution (1.3) or (1.4).

To compute the Fisher information matrix, we need the components of the score statistic, i.e. the derivatives of $\log f$ with respect to the parameters. These may be expressed in terms of Z, V and

$$W = (1 - 1/\alpha)\{2 + 1/(Z - 1 + 1/\alpha)\} - Z \tag{2.5}$$

as follows:

$$\begin{aligned}
\frac{\partial \log f}{\partial \mu} &= \frac{1}{\sigma} Z^{-\xi} (\cos V)^{-2\alpha\xi} \left(\frac{1}{\alpha} - \xi + W \cos^2 V \right), \\
\frac{\partial \log f}{\partial \sigma} &= -\frac{1}{\sigma} + \frac{1}{\sigma\xi} \left\{ Z^{-\xi} (\cos V)^{-2\alpha\xi} - 1 \right\} \left(\frac{1}{\alpha} - \xi + W \cos^2 V \right), \\
\frac{\partial \log f}{\partial \xi} &= -\log Z - 2\alpha \log \cos V + \frac{1}{\xi^2} \left(\frac{1}{\alpha} - \xi + W \cos^2 V \right) . \\
&\quad . \left\{ 1 - Z^{-\xi} (\cos V)^{-2\alpha\xi} - \xi \log Z - 2\alpha\xi \log \cos V \right\}, \\
\frac{\partial \log f}{\partial \alpha} &= -\frac{2}{\alpha} (\log \sin V + \log \cos V) - \frac{1}{\alpha^2 (Z - 1 + 1/\alpha)} \\
&\quad - 2W (\cos^2 V \log \cos V + \sin^2 V \log \sin V).
\end{aligned}$$

The expressions for $\partial \log f/\partial \nu$, $\partial \log f/\partial \tau$ and $\partial \log f/\partial \eta$ are obtained from those for $\partial \log f/\partial \mu$, $\partial \log f/\partial \sigma$ and $\partial \log f/\partial \xi$ by making the obvious substitutions of τ for σ, η for ξ, and replacing $\cos V$ by $\sin V$ everywhere. In the Gumbel case $\xi = \eta = 0$, these expressions may be replaced by their easily calculated limits as $\xi \to 0$, $\eta \to 0$. The Fisher information matrix is evaluated by calculating the variances and covariances of these quantities. The details of that are extremely tedious, but routine, so we shall not give any details of the calculation. For the Gumbel case, the answer has been given by Oakes and Manatunga (1992). They used a different parametrization from us, considering a bivariate Weibull form

$$\Pr\{T_1 > t_1, T_2 > t_2\} = \exp\left[- \left\{ (\eta_1^{\kappa_1} t_1^{\kappa_1})^{1/\alpha} + (\eta_2^{\kappa_2} t_2^{\kappa_2})^{1/\alpha} \right\}^{\alpha} \right], \quad t_1 > 0, t_2 > 0. \tag{2.6}$$

This may be transformed into (1.3) by writing $\kappa_1 = 1/\sigma$, $\eta_1 = e^{\mu}$, $t_1 = e^{-x}$, $\kappa_2 = 1/\tau$, $\eta_2 = e^{\nu}$, $t_2 = e^{-y}$. To transform the information matrix, note that, for example,

$$\frac{\partial \log f}{\partial \mu} = \frac{\partial \log f}{\partial \eta_1} \frac{\partial \eta_1}{\partial \mu} = e^{\mu} \frac{\partial \log f}{\partial \eta_1}$$

so that

$$\mathrm{E}\left\{ \left(\frac{\partial \log f}{\partial \mu} \right)^2 \right\} = -e^{2\mu} \mathrm{E} \left\{ \frac{\partial^2 \log f}{\partial \eta_1^2} \right\}$$

with similar expressions for the other parameters. In this way the results for the Gumbel case may be obtained directly from those in Oakes and Manatunga. Note that we do need the new parametrization, (1.3) instead of (2.6), because (2.6) is not easily extended to the Generalized Extreme Value case.

For the Generalized Extreme Value case, the needed formulae are given in detail in the Appendix.

We conclude this section by calculating some joint moments of S and T. From (2.2) and (2.4) it may quickly be established that

$$E\{S^pT^q\} = \frac{\Gamma(1+\alpha p)\Gamma(1+\alpha q)\Gamma(1+p+q)}{\Gamma(1+\alpha p+\alpha q)}. \tag{2.7}$$

Expectations involving log S and log T may also be calculated by differentiating with respect to p and q in (2.7). We write $\Psi(x) = (d/dx)\log\Gamma(x)$ for the digamma function, and note that $\gamma = -\Psi(1)$ is Euler's constant, and $\Psi'(1) = \pi^2/6$. It then follows that

$$\mathrm{cov}(S,T) = \frac{2\Gamma^2(1+\alpha)}{\Gamma(1+2\alpha)} - 1, \tag{2.8}$$

$$\begin{aligned} \mathrm{cov}\{(1-S)\log S, T\} &= 2 - (1+\alpha)\gamma - \alpha\Psi(1+\alpha) \\ &\quad - \frac{2\Gamma^2(1+\alpha)}{\Gamma(1+2\alpha)}\left[\frac{3}{2} - \gamma + \alpha\Psi(1+\alpha) - \alpha\Psi(1+2\alpha)\right], \end{aligned} \tag{2.9}$$

$$\begin{aligned} \mathrm{cov}\{(1-S)\log S, (1-T)\log T\} &= 1 + \gamma^2 - (1+\alpha^2)\frac{\pi^2}{6} + 2\alpha^2\Psi'(1+\alpha) \\ &\quad - 2(1-\gamma)\{1-\gamma-\alpha\gamma-\alpha\Psi(1+\alpha)\} \\ &\quad + \frac{2\Gamma^2(1+\alpha)}{\Gamma(1+2\alpha)}\left[\frac{\pi^2}{6} - \frac{5}{4} - \alpha^2\Psi'(1+2\alpha)\right. \\ &\quad \left. + \left\{\frac{3}{2} - \gamma + \alpha\Psi(1+\alpha) - \alpha\Psi(1+2\alpha)\right\}^2\right]. \end{aligned} \tag{2.10}$$

These formulae will be used in Section 3.

The results on Fisher information were generated by direct calculation without computer algebra. The algebraic manipulations have been checked several times, and we believe as best as we can that they are completely correct. In addition, extensive numerical checks have been run. All the expected-value formulae were checked by making comparisons with averages of up to a million Monte Carlo observations, and in virtually every case the Monte Carlo mean lay within 2.5 estimated standard errors of the calculated theoretical values. This was true for numerous different combinations of the key parameters ξ, η and α. In addition, the fact that all the computed covariance matrices were positive definite, and all the inequalities between the joint and marginal methods went in the right direction (i.e. the joint method being more efficient), itself served to corroborate that the algebraic results were correct. Finally, as a check against typographical or transcription errors, after the first draft of this manuscript had been completed, all the computer programs were completely rewritten using just the manuscript to generate the code. This produced identical results to the earlier ones that we had used while generating the results.

A separate issue is that of the inherent numerical stability of the formulae. They are unstable as ξ and η tend to 0, where in the limit it would be necessary to derive alternative limiting forms, and also as α tends to 0 or 1, in the latter case because then the Fisher information itself becomes unbounded (Tawn 1988). Numerical experience has suggested that in the middle of the range (α close to 0.5) the results are satisfactory for $|\xi|$ and $|\eta|$ as small as 0.001, but that it is necessary to be much more careful for values of α such as 0.1 or 0.9. All the results were programmed in double precision Fortran, using formulae in Abramowitz and Stegun (1964) to generate the gamma, digamma and trigamma functions to high accuracy, and using our own implementation of Simpson's rule, on a grid of 1000 points, to evaluate the numerical integrals involved in the formulae in the Appendix.

3. COMPARISON OF JOINT AND MARGINAL ESTIMATORS

The general method to be outlined is applicable to any problem of the structure described in Section 1, but the detailed discussion is restricted to the logistic model with Gumbel margins, i.e. (1.3). The method is then directly applicable to the corresponding problem for (1.4).

The model has five parameters $\mu, \sigma, \nu, \tau, \alpha$. Let J denote the 5×5 Fisher information matrix. The "joint" method of estimation consists of maximizing the full five-parameter likelihood function. From standard theory, the covariance matrix of the parameter estimates is given asymptotically by $n^{-1}J^{-1}$ where n is the sample size. This is obtained directly from Oakes and Manatunga (1992) using the transformation described in Section 2.

The "marginal" method consists of estimating μ and σ by the maximum likelihood estimates, $\widehat{\mu}$ and $\widehat{\sigma}$ say, based on the X values alone, and similarly $\widehat{\nu}$ and $\widehat{\tau}$ based on the Y values alone. These estimates are then used to transform X and Y to $\widehat{S}$ and $\widehat{T}$, and the estimator $\widehat{\alpha}$ is obtained by maximizing the one-parameter likelihood based on (1.5). Note that we use $\widehat{\mu}$, $\widehat{\alpha}$, etc., to denote the marginal rather than joint estimators. This will not cause confusion, because we shall nowhere need a separate notation for the joint estimators. Let θ denote the column vector of parameters μ, σ, ν, τ. If h is a function of θ we let $\partial h/\partial\theta, \partial h/\partial\theta^T$ denote respectively the column and row vectors of first-order derivatives, and $\partial^2 h/\partial\theta\partial\theta^T$ the matrix of second-order derivatives with respect to the components of θ. Let $\ell_n(\theta, \alpha)$ denote the log likelihood based on n observations, and let $\ell_n^*(\theta) = \ell_n(\theta, 1)$ denote the log likelihood for θ that arises from assuming X and Y independent. This is obtained by differentiating $\log f^*$, where f^* is the density in this case (i.e. after setting $\alpha = 1$). We also let θ_0, α_0 denote the true values of θ and α. The marginal estimator $\widehat{\theta}$ is the value of θ that maximizes $\ell_n^*(\theta)$, and $\widehat{\alpha}$ maximizes $\ell_n(\widehat{\theta}, \alpha)$ with respect to α.

A standard Taylor expansion shows that, to first order,

$$\widehat{\theta} - \theta_0 = \left(-\frac{\partial^2 \ell_n^*}{\partial\theta\partial\theta^T}\right)^{-1} \frac{\partial \ell_n^*}{\partial\theta} \tag{3.1}$$

where (here and subsequently) derivatives are evaluated at (θ_0, α_0) unless otherwise indicated. Thus

$$\mathrm{Cov}(\widehat{\theta}) \approx \left(-E\left\{\frac{\partial^2 \ell_n^*}{\partial\theta\partial\theta^T}\right\}\right)^{-1} \mathrm{Cov}\left(\frac{\partial \ell_n^*}{\partial\theta}\right) \left(-E\left\{\frac{\partial^2 \ell_n^*}{\partial\theta\partial\theta^T}\right\}\right)^{-1}. \tag{3.2}$$

Recall (from Gumbel 1958 or by setting $\alpha = 1$ in the Fisher information matrix) that the 2×2 Fisher information matrix for (μ, σ) is

$$\frac{1}{\sigma^2}\begin{bmatrix} 1 & \gamma - 1 \\ \gamma - 1 & \pi^2/6 + (1-\gamma)^2 \end{bmatrix} \tag{3.3}$$

and similarly for (ν, τ). This leads at once to $E\{-\partial^2 \ell_n^*/\partial\theta\partial\theta^T\}$, since the expected second-order derivatives with respect to (μ,μ), (σ,σ), (μ,σ), (ν,ν), (τ,τ) and (ν,τ) are obtained from (3.3) (with τ in place of σ where appropriate) and the derivatives with respect to (μ,ν), (μ,τ), (ν,σ) and (σ,τ) are identically 0. Similarly, the diagonal and $(\mu,\sigma), (\nu,\tau)$ off-diagonal elements of $\mathrm{Cov}\,(\partial\ell_n^*/\partial\theta)$ are derived from (3.3). However, the $(\mu,\nu), (\mu,\tau)$, (ν,σ) and (σ,τ) elements are non-zero and these terms require separate computation. In terms of S and T, it is easily checked that

$$\frac{\partial \log f^*}{\partial\mu} = \frac{1-S}{\sigma}, \quad \frac{\partial \log f^*}{\partial\sigma} = -\frac{1}{\sigma}\{1 + (1-S)\log S\},$$

$$\frac{\partial \log f^*}{\partial\nu} = \frac{1-T}{\tau}, \quad \frac{\partial \log f^*}{\partial\tau} = -\frac{1}{\tau}\{1 + (1-T)\log T\}.$$

The covariances of these must, of course, be evaluated under the true model, for which $\alpha \neq 1$. Thus

$$\sigma\tau E\left\{\frac{\partial \log f^*}{\partial\mu}\frac{\partial \log f^*}{\partial\nu}\right\} = \mathrm{Cov}(S,T), \tag{3.4}$$

$$\sigma\tau E\left\{\frac{\partial \log f^*}{\partial\sigma}\frac{\partial \log f^*}{\partial\nu}\right\} = \sigma\tau E\left\{\frac{\partial \log f^*}{\partial\mu}\frac{\partial \log f^*}{\partial\tau}\right\}$$

$$= \mathrm{cov}\{(1-S)\log S, T\}, \tag{3.5}$$

$$\sigma\tau E\left\{\frac{\partial \log f^*}{\partial\sigma}\frac{\partial \log f^*}{\partial\tau}\right\} = \mathrm{cov}\{(1-S)\log S, (1-T)\log T\}. \tag{3.6}$$

Expressions (3.4)-(3.6) are evaluated from (2.8)-(2.10). This allows the evaluation of (3.2) to be completed.

Now we consider the properties of $\widehat{\alpha}$ obtained by setting $\partial\ell_n/\partial\alpha = 0$ at $(\widehat{\theta}, \widehat{\alpha})$. A Taylor expansion of ℓ_n shows that, to first order,

$$\widehat{\alpha} - \alpha_0 = \left(-\frac{\partial^2 \ell_n}{\partial\alpha^2}\right)^{-1}\left\{\frac{\partial \ell_n}{\partial\alpha} + \frac{\partial^2 \ell_n}{\partial\alpha\partial\theta^T}(\widehat{\theta} - \theta_0)\right\}, \tag{3.7}$$

where, again, the partial derivatives are evaluated at the true values (θ_0, α_0). This may be combined with (3.1) to obtain an approximation for $\widehat{\alpha}$ in terms of derivatives of ℓ_n and ℓ_n^*. For first-order asymptotics, it suffices to replace all the second-order derivatives by their expected values. We also have

$$\mathrm{Cov}\left(\frac{\partial \ell_n}{\partial \alpha}, \frac{\partial \ell_n^*}{\partial \theta}\right) = 0. \tag{3.8}$$

To see (3.8), observe that for any measurable function h of x alone, $\int h(x) f(x, y; \theta, \alpha) dx dy$ depends just on the marginal distribution of X and hence is independent of α. Hence, differentiating,

$$\mathrm{Cov}\left\{h(X), \frac{\partial \log f}{\partial \alpha}\right\} = 0.$$

Of course, the same thing also applies to a function of Y alone. Since all the elements of $\partial \ell_n^* / \partial \theta$ are functions of either the X-values alone or the Y-values alone, (3.8) follows.

Combining (3.1), (3.7) and using (3.8), we deduce

$$\mathrm{Cov}(\widehat{\theta}, \widehat{\alpha}) \approx \left\{-E\left(\frac{\partial^2 \ell_n}{\partial \alpha^2}\right)\right\}^{-1} E\left(\frac{\partial^2 \ell_n}{\partial \alpha \partial \theta^T}\right) \mathrm{Cov}(\widehat{\theta}), \tag{3.9}$$

$$\mathrm{var}(\widehat{\alpha}) \approx \left\{-E\left(\frac{\partial^2 \ell_n}{\partial \alpha^2}\right)\right\}^{-1} + \left\{-E\left(\frac{\partial^2 \ell_n}{\partial \alpha^2}\right)\right\}^{-2} E\left(\frac{\partial^2 \ell_n}{\partial \alpha \partial \theta^T}\right) \mathrm{Cov}(\widehat{\theta}) E\left(\frac{\partial^2 \ell_n}{\partial \alpha \partial \theta}\right). \tag{3.10}$$

In (3.9) and (3.10), the means and covariances involving ℓ_n^* are calculated using (3.3)-(3.6) as before, while the remaining quantities are derived from the full Fisher information matrix based on ℓ_n. Combining (3.2), (3.9) and (3.10) gives the full asymptotic covariance matrix of $(\widehat{\theta}, \widehat{\alpha})$.

We now consider the application of these results to a number of specific problems:-

(a) *Efficiency of parameter estimation.* The formulae (3.2), (3.9) and (3.10) lead directly to the asymptotic covariance matrix of the estimators under the marginal method. This can be compared with the corresponding matrix for the joint method, obtained by inverting the Fisher information matrix. In particular, the ratios of diagonal entries of these matrices are a direct measure of the asymptotic relative efficiency of the two methods for individual parameter estimates.

(b) *D-efficiency, Q-efficiency.* To summarize the information in the covariance matrices into a single number, one commonly used measure is the determinant of the matrix. We compute the ratio of the two determinants for the joint and marginal methods of estimation, and raise this to the power $1/m$ where m is the total number of parameters (5 or 7). The last step acts as a scaling transformation, so that the various efficiencies are comparable. This one will be referred to as the D-efficiency of joint vs. marginal estimation.

An alternative approach is to focus on specific functionals. In extreme value theory, the functionals of most interest are usually quantiles of the distribution. For one of the two margins, say X, the p^*-quantile of the distribution can be obtained as a function of μ, σ and if present ξ, by solving $F(x) = p^*$ where F is given by (1.1) or (1.2). The asymptotic variance of the resulting estimate can be calculated via the delta method. The ratio of asymptotic variances for the joint and marginal methods provides another measure of efficiency which we call Q-efficiency. Note that this is something that depends purely on the marginal distribution of one of the two variables. Intuition would suggest that that there is little to be gained from joint estimation in this case. Intuition can be wrong!

(c) *Testing equality of margin parameters.* An important issue in connection with regional methods in hydrology is to test whether all or some of the margin parameters are constant across a region. Here we consider a "region" consisting of just two sites. This can be assessed by testing whether the vector $(\widehat{\mu}-\widehat{\nu}, \widehat{\sigma}-\widehat{\tau})$ is 0. A naive approach would be to do this ignoring the dependence between the two sites. Note that, as a consequence of (3.4)-(3.6), although the marginal distributions of $\widehat{\mu}, \widehat{\sigma}, \widehat{\nu}$ and $\widehat{\tau}$ are unaffected by dependence, the cross-covariances of $(\widehat{\mu}, \widehat{\sigma})$ with $(\widehat{\nu}, \widehat{\tau})$ are non-zero, meaning that covariances derived under the assumption that X and Y are independent will be wrong. Equation (3.2) may be used to evaluate the true covariances and hence to assess how far we would be in error if we ignored the dependence altogether.

To put more substance on this, suppose the estimated covariance matrix of $(\widehat{\mu}-\widehat{\nu}, \widehat{\sigma}-\widehat{\tau})$ is C. That is to say, this is the covariance we would calculate if X and Y were independent (directly from the Fisher information). Also let C_0 denote the true covariance matrix, which we can compute using the formulae for the marginal method. We can consider the test statistic

$$(\widehat{\mu}-\widehat{\nu}, \widehat{\sigma}-\widehat{\tau})^T C^{-1} (\widehat{\mu}-\widehat{\nu}, \widehat{\sigma}-\widehat{\tau})$$

which also approximates the likelihood ratio statistic. The actual (approximate) distribution of this, under the assumption that the true covariance matrix is C_0, is a weighted mixture of chi-squares, but a good summary is obtained from the mean, which we may calculate as $\mathrm{tr}\{C^{-1}C_0\}/2$, the division by 2 being to rescale it relative to the mean of the χ^2_2 statistic which would apply if $C = C_0$. In the Generalized Extreme Value case, with three parameters to test, this 2 is replaced by 3. For want of a better name, the resulting scalar quantity will be called the testing efficiency.

This is not strictly a "joint" versus "marginal" estimation problem, so much as one in which we take account of dependence versus one in which we do not. However, it is so closely related to the main theme of this paper that it seems worth considering as part of the same study. Intuitively one would expect the testing efficiency to be near 1 when α is near 1, since then the two components are approximately independent, but to be near 0 for α near 0, when the components are highly dependent and the true variances of $\widehat{\mu}-\widehat{\nu}$ and $\widehat{\sigma}-\widehat{\tau}$ (under the null hypothesis) near 0. A testing efficiency between 0 and 1 means that the χ^2 statistic will be too small and thus the test will tend to accept the null hypothesis when it should reject.

(d) *Estimation assuming common margin parameters.* Another issue arising in regional methods is the effect of dependence on estimates of margin parameters when these are assumed constant across a region (c.f. Hosking and Wallis 1988). Again considering a "region" of only two sites, we may consider the estimation of μ, σ, ν and τ under the assumption that $\mu=\nu$ and $\sigma=\tau$. Suppose we naively estimate μ and σ from the likelihood constructed by assuming independence. If we partition $\theta=\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}$ where $\theta_1=\begin{pmatrix}\mu\\ \sigma\end{pmatrix}$ and $\theta_2=\begin{pmatrix}\nu\\ \tau\end{pmatrix}$, we have the expansion (about the true values θ_{10}, θ_{20}),

$$
\begin{aligned}
&\ell_n^*(\theta_1,\theta_2)-\ell_n^*(\theta_{10},\theta_{20})= \\
&(\theta_1-\theta_{10})^T\frac{\partial \ell_n^*}{\partial\theta_1}+(\theta_2-\theta_{20})^T\frac{\partial \ell_n^*}{\partial\theta_2}+\frac{1}{2}\left\{(\theta_1-\theta_{10})^T\frac{\partial^2\ell_n^*}{\partial\theta_1\partial\theta_1^T}(\theta_1-\theta_{10})\ldots\right\}.
\end{aligned}
$$

Setting $\theta_1=\theta_2=\theta_3$ say, $\theta_{10}=\theta_{20}=\theta_{30}$ and maximizing with respect to θ_3, we have to first order

$$
\widehat{\theta}_3=\left(-\frac{\partial^2\ell_n^*}{\partial\theta_1\partial\theta_1^T}-\frac{\partial^2\ell_n^*}{\partial\theta_1\partial\theta_2^T}-\frac{\partial^2\ell_n^*}{\partial\theta_2\partial\theta_1^T}-\frac{\partial^2\ell_n^*}{\partial\theta_2\partial\theta_2^T}\right)^{-1}\left(\frac{\partial\ell_n^*}{\partial\theta_1}+\frac{\partial\ell_n^*}{\partial\theta_2}\right)=B^{-1}c \text{ say.} \quad (3.11)
$$

The asymptotic covariance matrix of $\widehat{\theta}_3$ is then given by $B^{-1}E\{cc^T\}B^{-1}$, which reduces to B^{-1} only in the case $\alpha=1$. By comparing these two matrices, we may assess how far we would be in error if we ignored the dependence.

By analogy with our method for problem (c), one way to reduce this to a scalar measure of efficiency is to compute $\mathrm{tr}[E\{cc^T\}B^{-1}]/m$ where m is 2 in the Gumbel case and 3 in the Generalized Extreme Value case. We call this the common estimation efficiency, and a value greater than 1 indicates that the sampling variances are being underestimated by ignoring the dependence.

It should be pointed out that problems (c) and (d) do not correspond exactly to the regional models usually adopted in hydrology. These assume the data over a region are from a common distribution up to a site-dependent scaling constant. This would amount to testing $\mu/\sigma=\nu/\tau$ in problem (c), and estimating under this constraint in problem (d). We do not consider this problem separately, since in the context of our general development it is no more than a minor variant.

4. NUMERICAL AND SIMULATION RESULTS

This section contains the results of numerical computations of the various asymptotic efficiencies described in Section 3. In addition, some limited simulations were performed to try to determine actual efficiencies in finite samples.

Table 1 contains asymptotic efficiencies for the Gumbel model. The only parameter that needs to be varied here is α, and this is given in the first column. The next three columns contain asymptotic efficiencies for the estimation of μ, σ and α. Of course, the asymptotic efficiencies for estimation of ν and τ are the same as those for μ and σ respectively. The next four columns contain the D-efficiency, and the Q-efficiency for p^* =0.9, 0.99 and 0.999. The last two columns contain the testing efficiency and the common estimation efficiency.

Table 1: Efficiencies in the Gumbel case

α	μ-eff	σ-eff	α-eff	D-eff	Q-eff 0.9	Q-eff 0.99	Q-eff 0.999	T-eff	CE-eff
.001	1.000	1.000	1.000	.786	1.000	1.000	1.000	.000	2.000
.010	1.000	1.000	1.000	.789	1.000	1.000	1.000	.000	2.000
.100	.999	.987	.999	.820	.994	.991	.990	.032	1.968
.200	.997	.965	.997	.852	.981	.974	.971	.107	1.893
.300	.994	.945	.994	.882	.968	.957	.953	.207	1.793
.400	.992	.934	.992	.908	.957	.946	.942	.318	1.682
.500	.990	.932	.990	.931	.951	.941	.938	.433	1.567
.600	.988	.938	.988	.951	.950	.943	.941	.549	1.451
.700	.988	.951	.988	.968	.956	.952	.951	.664	1.336
.800	.990	.967	.990	.982	.968	.966	.966	.777	1.223
.900	.995	.986	.995	.993	.985	.985	.985	.889	1.111
.990	1.000	.999	1.000	1.000	.999	.999	.999	.989	1.011
.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	1.001

Ignoring the last two columns for the moment, it can be seen that the efficiencies of the marginal method are very good virtually everywhere. The worst efficiencies for single parameter estimates are those for σ, but even there the lowest value is 0.932. Results for Q-efficiency usually lie between those for the estimation of μ and σ. Given the greater practical ease of the marginal method, we could well conclude that the marginal method is good enough for all practical purposes. Such a conclusion could have been anticipated from Oakes and Manatunga (1992), who found an approximate orthogonality in the Fisher information matrix between α and the margin parameters.

The last two columns of Table 1 tell a different story, but they are concerned with a different question. The T-efficiency represents the extent to which the χ^2 statistic is reduced by ignoring the dependence, and the CE-efficiency represents the extent to which the variances of estimates of common margin parameters are inflated compared with those that would be applicable if the margins were independent. In each case, therefore, the results indicate anti-conservative behaviour, and the effect is quite substantial in the middle of the table where the dependence is moderate.

The conclusion seems to be that it is satisfactory to use the marginal method for estimation, but for the purposes of testing and common estimation, it is not satisfactory to ignore the dependence altogether.

Now we turn to the Generalized Extreme value distribution, and find that, for certain combinations of parameter values, the qualitative conclusion drawn with the Gumbel model, that marginal estimation is satisfactory, fails rather badly! Detailed results are given only for Q-estimation with $p^* = 0.999$, but the other definitions of efficiency have also been computed and will be described briefly. Computations have been carried out for $r = 1/\alpha$ of 1.1, 2 and 10, and for certain combinations of ξ and η listed in Table 2. Here Q_1-efficiency and Q_2-efficiency refer to the first (X) and second (Y) components respectively. Also included are simulation results for estimation in finite samples of sizes n =30, 100 and 500. The rows of Table 2 labelled $n = \infty$ correspond to the asymptotic calculations, while the others are based on simulation. All the simulations were carried out on the Sheffield University Prime computer system and are based on averages of 500 replications in the cases of sample sizes n =30 and 100, and 200 replications for $n = 500$. The relatively small number of replications was necessitated by the large amounts of CPU time needed to run these simulations. Standard errors for the simulation results are not reported in Table 2, but were substantial. Some indication of that may be gained from the cases in which $\xi = \eta$, where the Q_1 and Q_2 efficiencies should be equal, but are not in most cases. Note that the simulation results are reported as they came out with no attempt to correct for the two efficiencies being equal in this case.

The main conclusion from the asymptotic results is that in some cases the efficiency of the marginal estimation method is very low, below 0.1. This seems to occur when ξ and η are of opposite sign, i.e. the long tail of one component corresponds to the short tail of the other. It turns out, counter-intuitively, that it is the short tail (the one with positive shape parameter) that is worse estimated by the marginal method. The effect is most clearly demonstrated for $r = 10$, for which the two components are very highly dependent, but even for $r = 2$ (a more typical value for practical applications) some of the efficiencies are low enough to be of concern.

It seems paradoxical that efficiencies lower than 0.5 can be obtained at all. Imagine the following experiment: you have n observations from the X component, say $X_1, ..., X_n$, and you want to estimate the extreme quantiles of X. Clearly this can be done without any reference to Y, but it is conceivable that knowing the corresponding values of $Y_1, ..., Y_n$, when the two components are highly dependent, would improve the estimation of the distribution of X. So far, there is no paradox. Suppose, however, you are offered a choice: either take $Y_1, ..., Y_n$, or a second sample of X values, say $X_{n+1}, ..., X_{2n}$. Which do you

Table 2: Q-Efficiencies for the GEV model ($r = 1/\alpha$)

ξ	η	n	Q_1-eff $r = 1.1$	Q_2-eff $r = 1.1$	Q_1-eff $r = 2$	Q_2-eff $r = 2$	Q_1-eff $r = 10$	Q_2-eff $r = 10$
-.40	-.40	30	.916	1.18	.541	.170	.230	1.93
-.40	-.40	100	.917	1.05	.671	.571	.443	.476
-.40	-.40	500	.996	.971	.775	.894	.667	.669
-.40	-.40	∞	.968	.968	.783	.783	.631	.631
-.40	-.20	30	1.12	.989	.722	.913	.083	.116
-.40	-.20	100	.891	.931	.685	.677	.482	.468
-.40	-.20	500	.943	.956	.802	.727	.553	.469
-.40	-.20	∞	.969	.959	.799	.703	.591	.429
-.20	-.20	30	.714	.808	.264	.269	.244	.286
-.20	-.20	100	.939	.987	.537	.679	.508	.497
-.20	-.20	500	.894	.951	.690	.710	.444	.514
-.20	-.20	∞	.960	.960	.727	.727	.551	.551
-.40	.20	30	.822	1.18	.026	.805	.028	.356
-.40	.20	100	1.02	.911	.640	.661	.274	.256
-.40	.20	500	1.01	.967	.788	.578	.325	.198
-.40	.20	∞	.970	.918	.782	.486	.324	.061
-.20	.20	30	.837	.926	.420	.757	.170	.404
-.20	.20	100	.949	.955	.589	.731	.398	.387
-.20	.20	500	.979	.957	.760	.635	.369	.245
-.20	.20	∞	.962	.920	.734	.517	.237	.088
.20	.20	30	.979	.950	.467	.565	.409	.381
.20	.20	100	1.01	1.02	.719	.734	.587	.559
.20	.20	500	.945	.916	.652	.806	.592	.577
.20	.20	∞	.924	.924	.614	.614	.460	.460
-.40	.40	30	.776	1.04	.331	.811	.111	.403
-.40	.40	100	.956	.972	.440	.607	.172	.219
-.40	.40	500	.923	.956	.694	.588	.351	.139
-.40	.40	∞	.967	.868	.701	.349	.263	.055
-.20	.40	30	.676	1.03	.262	.818	.103	.417
-.20	.40	100	.889	.920	.552	.634	.254	.261
-.20	.40	500	.966	.924	.724	.597	.350	.170
-.20	.40	∞	.959	.869	.645	.365	.139	.057
.20	.40	30	.928	1.08	.838	.864	.334	.470
.20	.40	100	.931	1.00	.776	.744	.441	.454
.20	.40	500	.889	.896	.723	.665	.410	.388
.20	.40	∞	.923	.875	.568	.447	.072	.079
.40	.40	30	1.00	.950	.715	.794	.538	.479
.40	.40	100	.944	.965	.755	.730	.502	.531
.40	.40	500	.955	.959	.700	.776	.629	.641
.40	.40	∞	.880	.880	.558	.558	.451	.451

choose? Remember that the Y values have a different marginal distribution and their only relevance to the problem at hand is the fact that X and Y are dependent. Nevertheless, in those cases with a marginal vs. joint efficiency less than 0.5, the correct decision is to take the Y sample.

In terms of comparing the simulated and asymptotic efficiencies, it has to be admitted that the direct numerical match-up is not at all good. This is partly because of the very high variability in the simulation results - which perhaps at least acts as a warning against trying to resolve a problem like this entirely by simulation. It also seems likely to us that some of the very small efficiencies in the table - those of .026 and .028 for $\xi = -0.4$, $\eta = 0.2$, $n = 30$ - are spurious. In general, we would expect the simulated efficiencies to be greater than the asymptotic efficiencies, the reason for this being that in small samples the relative simplicity of the marginal method, not requiring nearly so complicated an optimization, ought to be acting in its favour. Indeed, for $n = 30$ (especially) quite a few of the results indicate that the marginal method is slightly the more efficient of the two. However, the simulation results do provide independent evidence that for certain combinations of parameter values the efficiency of the marginal method can be very much less than 1, and they largely mimic the asymptotic results in predicting the combinations of parameters for which this occurs.

Results for estimation of individual parameters show that the efficiencies for μ, σ, ν, τ and α are always high (greater than 0.85 for all cases computed), but those for ξ and η can be much smaller, often in the range 0.1-0.2 for those cases in which the Q-efficiency is less than 0.1. These results have also been examined by simulation, with similar qualitative conclusions as for Q-efficiency, i.e. less dramatic than the asymptotic results but still confirming that the efficiency can be much less than 1.

Table 3: Testing and Common Estimation Efficiencies

ξ	η	T-eff $r = 1.1$	CE-eff $r = 1.1$	T-eff $r = 2$	CE-eff $r = 2$	T-eff $r = 10$	CE-eff $r = 10$
-.40	-.40	.910	1.090	.481	1.519	.043	1.957
-.40	-.20	.920	1.162	.508	1.627	.067	2.047
-.20	-.20	.904	1.096	.460	1.540	.038	1.962
-.40	.20	1.045	1.909	.693	2.531	.271	2.933
-.20	.20	.916	1.580	.548	2.168	.153	2.572
.20	.20	.883	1.117	.403	1.597	.030	1.970
-.40	.40	1.162	4.831	.867	5.615	.469	6.017
-.20	.40	.980	3.551	.683	4.294	.320	4.700
.20	.40	.813	1.710	.450	2.342	.122	2.727
.40	.40	.863	1.137	.364	1.636	.025	1.975

Finally, Table 3 gives asymptotic results for T-efficiency and CE-efficiency. There are some cases in which the T-efficiency is greater than 1 (indicating that the χ^2 test is conservative), but in general the results confirm that the same problems exist as in the Gumbel case but are significantly worse here.

5. SUMMARY AND CONCLUSIONS

The original motivation for this study was to try to demonstrate that the marginal method, which seems intuitively very reasonable, would in fact have high asymptotic efficiency with respect to the theoretically efficient joint method. For the models with Gumbel margins, this is confirmed by the results. For the Generalized Extreme Value cases, however, it turned out that there are certain combinations of parameter values for which the variances of the two methods were dramatically different. Simulation results failed to give good numerical agreement, the result at least in part of the difficulty of obtaining accurate simulation results, but they did confirm the general pattern.

In spite of these conclusions, we are inclined to recommend the marginal method for practical use under most circumstances. The dramatic failures only occur when one tail is short, the other long, and the dependence between the two components is high. This would be an unusual combination of circumstances in practice. Nevertheless, our results serve as a warning of what can go wrong.

The general methodology of the paper is applicable to much broader classes of multivariate models than the one that has been studied in detail. The "joint" versus "marginal" issue arises whenever the multivariate structure is defined in terms of a reduced or standardized form of the marginal distributions. In the case of the multivariate normal the issue is trivial because in this case the joint and marginal estimators coincide, but for most non-normal classes the issue can be expected to be a live one.

The conclusions on T-efficiency and CE-efficiency are of rather a different nature and serve primarily to warn against ignoring the dependence between components under these circumstances.

ACKNOWLEDGEMENTS

The work of the first two authors was carried out primarily at the University of Surrey and supported by the Wolfson Foundation. The Foundation's support was very much appreciated and grateful thanks are given. The work of the third author was carried out as part of a PhD thesis at Sheffield University and was supported by the SERC. Computational facilities were supported by the SERC (Surrey) and the NSF (North Carolina). The authors wish to thank Dr. Jonathan Tawn for his advice and encouragement.

REFERENCES

Abramowitz, M. and Stegun, I.A. (1964), *Handbook of Mathematical Functions.* National Bureau of Standards, Washington D.C., reprinted by Dover, New York.

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

Coles, S.G. and Walshaw (1992), Directional modelling of extreme wind speeds. To appear in *Applied Statistics.*

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

Gumbel, E.J. (1958), *Statistics of Extremes.* Columbia University Press, New York.

Hosking, J.R.M. (1985), Algorithm AS215: Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Appl. Statist.* **34**, 301-310.

Hosking, J.R.M., and Wallis, J.R. (1988), The effect of intersite dependence on regional flood frequency analysis. *Wat. Res. Res.* **24**, 588-600.

Joe, H., Smith, R.L. and Weissman, I. (1992), Bivariate threshold models for extremes. *J.R. Statist. Soc. B* **54**, 171-183.

Lee, L. (1979), Multivariate distributions having Weibull properties. *J. Multivariate Analysis* **9**, 267-277.

N.E.R.C. (1975), *The Flood Studies Report.* The Natural Environment Research Council, London.

Oakes, D. and Manatunga, A.K. (1992), Fisher information for a bivariate extreme value distribution method. To appear, *Biometrika.*

Prescott, P. and Walden, A.T. (1980), Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Biometrika* **67**, 723-724.

Smith, R.L. (1985), Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67-92.

Smith, R.L. (1992), Regional estimation from spatially dependent data. Tentatively accepted for *Applied Statistics.*

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

APPENDIX: GENERALIZED EXTREME VALUE MARGINS

The Fisher information matrix for the univariate Generalized Extreme Value distribution was given by Prescott and Walden (1980); for the regularity conditions we require $\xi < \frac{1}{2}, \eta < \frac{1}{2}$. Here we give (a) the Fisher information matrix for the bivariate case (1.4), (b) the analogue of (3.4)-(3.6). Let f denote the density derived from (1.4), f_0 the corresponding density in the Gumbel case $\xi = \eta = 0$, and f^* the density corresponding to f when $\alpha = 1$. The beta function is denoted $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$; all other notation is as in Section 2.

With Z and W defined by (2.2) and (2.5), let

$$H_k(\xi) = E\left\{Z^{-\xi}(\log Z)^{k-1}W^2\right\}, \quad k = 1, 2,$$

which must be evaluated by numerical integration. Also let

$$\begin{aligned} K_1(\xi) &= E\left\{\frac{Z^{-\xi}W}{Z-1+1/\alpha}\right\} \\ &= -\frac{\alpha}{1-\alpha}H_1(\xi) + \Gamma(1-\xi)\left\{\frac{2(2-\alpha\xi)(1-\alpha\xi)}{\alpha} + \frac{(2+\alpha-\alpha\xi)(1+\alpha-\alpha\xi)(1-\xi)}{1-\alpha}\right\}. \end{aligned}$$

Define

$$\begin{aligned} A_1(\xi) &= \frac{\Gamma(1-\xi)}{\alpha^2}\left[(1-\alpha\xi)\{\gamma+\Psi(3-\alpha\xi)\} + 1 - \frac{1}{2-\alpha\xi}\right] \\ &\quad + \frac{K_1(\xi)}{\alpha^2(2-\alpha\xi)} - \frac{H_1(\xi)\{\gamma+\Psi(3-\alpha\xi)\}}{(2-\alpha\xi)(3-\alpha\xi)}, \\ B_1(\xi) &= \left(\frac{1}{\alpha}-\xi\right)\left(3\xi-\frac{1}{\alpha}\right)\Gamma(1-2\xi) + \frac{H_1(2\xi)}{3-2\alpha\xi}, \\ B_2(\xi,\eta) &= B(2-\alpha\xi, 2-\alpha\eta)H_1(\xi+\eta) \\ &\quad - \frac{(1-\alpha\xi)(1-\alpha\eta)(1-\alpha\xi-\alpha\eta)}{\alpha^2}B(1-\alpha\xi, 1-\alpha\eta)\Gamma(1-\xi-\eta), \\ C_1(\xi) &= \frac{H_2(\xi)}{3-\alpha\xi} - \frac{\alpha H_1(\xi)}{(3-\alpha\xi)^2} - \frac{2-\alpha\xi}{\alpha}\Gamma(1-\xi) - \frac{(1-\alpha\xi)^2}{\alpha^2}\Gamma'(1-\xi), \\ C_2(\xi) &= \frac{1-\alpha\xi}{\alpha}\{\gamma+\Psi(1-\alpha\xi)-1\}\Gamma(1-\xi) - \frac{1-\alpha\xi}{\alpha^2}\Gamma'(1-\xi) \\ &\quad + \frac{H_2(\xi)}{(2-\alpha\xi)(3-\alpha\xi)} + \frac{\alpha H_1(\xi)}{(2-\alpha\xi)(3-\alpha\xi)}\{1-\gamma-\Psi(4-\alpha\xi)\}, \\ D_1(\xi) &= -\frac{(1-\alpha\xi)^2}{\alpha^2}\Gamma(1-\xi) + \frac{H_1(\xi)}{3-\alpha\xi}, \\ D_2(\xi) &= -\frac{1-\alpha\xi}{\alpha^2}\Gamma(1-\xi) + \frac{H_1(\xi)}{(2-\alpha\xi)(3-\alpha\xi)}. \end{aligned}$$

Also let $a_1 = A_1(0)$, $b_1 = B_1(0) = D_1(0)$, $b_2 = B_2(0,0) = D_2(0)$, $c_1 = C_1(0)$, $c_2 = C_2(0)$, and let a_0, a_2, d_1, d_2 respectively denote the expected values of $(\partial \log f/\partial\alpha)^2$, $\sigma(\partial \log f/\partial\sigma)(\partial \log f/\partial\alpha)$, $\sigma^2(\partial \log f/\partial\sigma)^2$ and $\sigma\tau(\partial \log f/\partial\sigma)(\partial \log f/\partial\tau)$ in the Gumbel case. These expressions must be read off from Oakes and Manatunga (1992) via the transformation described in Section 2.

We then have:

$$E\left\{\left(\frac{\partial \log f}{\partial\alpha}\right)^2\right\} = a_0,$$

$$E\left\{\frac{\partial \log f}{\partial\mu}\frac{\partial \log f}{\partial\alpha}\right\} = -\frac{A_1(\xi)}{\sigma},$$

$$E\left\{\frac{\partial \log f}{\partial\sigma}\frac{\partial \log f}{\partial\alpha}\right\} = \frac{a_1 - A_1(\xi)}{\sigma\xi},$$

$$E\left\{\frac{\partial \log f}{\partial\xi}\frac{\partial \log f}{\partial\alpha}\right\} = \frac{a_2}{\xi} + \frac{A_1(\xi) - a_1}{\xi^2},$$

$$E\left\{\left(\frac{\partial \log f}{\partial\mu}\right)^2\right\} = \frac{B_1(\xi)}{\sigma^2},$$

$$E\left\{\frac{\partial \log f}{\partial\mu}\frac{\partial \log f}{\partial\nu}\right\} = \frac{B_2(\xi,\eta)}{\sigma\tau},$$

$$E\left\{\frac{\partial \log f}{\partial\mu}\frac{\partial \log f}{\partial\sigma}\right\} = \frac{B_1(\xi) - D_1(\xi)}{\sigma^2\xi},$$

$$E\left\{\frac{\partial \log f}{\partial\mu}\frac{\partial \log f}{\partial\tau}\right\} = \frac{B_2(\xi,\eta) - D_2(\xi)}{\sigma\tau\eta},$$

$$E\left\{\frac{\partial \log f}{\partial\mu}\frac{\partial \log f}{\partial\xi}\right\} = \frac{D_1(\xi) - B_1(\xi)}{\sigma\xi^2} - \frac{C_1(\xi)}{\sigma\xi},$$

$$E\left\{\frac{\partial \log f}{\partial\mu}\frac{\partial \log f}{\partial\eta}\right\} = \frac{D_2(\xi) - B_2(\xi,\eta)}{\sigma\eta^2} - \frac{C_2(\xi)}{\sigma\eta},$$

$$E\left\{\left(\frac{\partial \log f}{\partial\sigma}\right)^2\right\} = \frac{B_1(\xi) - 2D_1(\xi) + b_1}{\sigma^2\xi^2},$$

$$E\left\{\frac{\partial \log f}{\partial\sigma}\frac{\partial \log f}{\partial\tau}\right\} = \frac{B_2(\xi,\eta) + b_2 - D_2(\xi) - D_2(\eta)}{\sigma\tau\xi\eta},$$

$$E\left\{\frac{\partial \log f}{\partial\sigma}\frac{\partial \log f}{\partial\xi}\right\} = \frac{c_1 - C_1(\xi)}{\sigma\xi^2} - \frac{B_1(\xi) - 2D_1(\xi) + b_1}{\sigma\xi^3},$$

$$E\left\{\frac{\partial \log f}{\partial\sigma}\frac{\partial \log f}{\partial\eta}\right\} = \frac{c_2 - C_2(\xi)}{\sigma\xi\eta} + \frac{D_2(\xi) + D_2(\eta) - B_2(\xi,\eta) - b_2}{\sigma\xi\eta^2},$$

$$E\left\{\left(\frac{\partial \log f}{\partial\xi}\right)^2\right\} = \frac{d_1}{\xi^2} - \frac{2(c_1 - C_1(\xi))}{\xi^3} + \frac{B_1(\xi) + b_1 - 2D_1(\xi)}{\xi^4},$$

$$E\left\{\frac{\partial \log f}{\partial \xi}\frac{\partial \log f}{\partial \eta}\right\} = \frac{1}{\xi\eta}\left\{d_2 - c_2\left(\frac{1}{\xi}+\frac{1}{\eta}\right) + \frac{C_2(\xi)}{\xi} + \frac{C_2(\eta)}{\eta}\right\}$$
$$+\frac{B_2(\xi,\eta)+b_2-D_2(\xi)-D_2(\eta)}{\xi^2\eta^2}.$$

The remaining elements of the Fisher information matrix are computed using obvious symmetries. For example,

$$E\left\{\frac{\partial \log f}{\partial \nu}\frac{\partial \log f}{\partial \alpha}\right\} = -\frac{A_1(\eta)}{\tau}.$$

Now let us turn to the analogues of (3.4)-(3.6). With S, T as in Section 2 and

$$P = \frac{1-\xi-S}{S^{\xi}}, Q = \frac{1-\eta-T}{T^{\eta}}$$

we have

$$\frac{\partial \log f^*}{\partial \mu} = \frac{P}{\sigma}, \frac{\partial \log f^*}{\partial \sigma} = \frac{P-1+S}{\sigma\xi}, \quad \frac{\partial \log f^*}{\partial \xi} = -\frac{P-1+S}{\xi^2} - \frac{1+(1-S)\log S}{\xi}$$

with analogous expressions for differentiation with respect to ν, τ,η. Formulae such as

$$-\sigma\xi\eta^2 E\left\{\frac{\partial \log f^*}{\partial \sigma}\frac{\partial \log f^*}{\partial \eta}\right\} = \ \mathrm{cov}(P+S,Q+T)+\eta\ \mathrm{cov}\{P+S,(1-T)\log T\}$$

follow at once. To complete the picture, in addition to (2.8)-(2.10), we have from (2.7) that

$$\begin{aligned}
\mathrm{cov}(P,Q) &= (1-\xi)(1-\eta)(1-\alpha\xi-\alpha\eta)\Gamma(1-\xi-\eta)B(1-\alpha\xi,1-\alpha\eta)\\
&\quad + (1+2\alpha-\alpha\xi-\alpha\eta)\Gamma(3-\xi-\eta)B(1+\alpha-\alpha\xi,1+\alpha-\alpha\eta)\\
&\quad - (1+\alpha-\alpha\xi-\alpha\eta)\Gamma(2-\xi-\eta).\\
&\quad .\left\{(1-\xi)B(1-\alpha\xi,1+\alpha-\alpha\eta)+(1-\eta)B(1-\alpha\eta,1+\alpha-\alpha\xi)\right\},\\
\mathrm{cov}(P,T) &= \Gamma(2-\xi)\big\{(1-\xi)(1+\alpha-\alpha\xi)B(1+\alpha,1-\alpha\xi)\\
&\quad - (2-\xi)(1+2\alpha-\alpha\xi)B(1+\alpha,1+\alpha-\alpha\xi)\big\},
\end{aligned}$$
$$\mathrm{cov}\left\{P,(1-T)\log T\right\} = \Gamma(2-\xi)\left\{\alpha\Psi(1+\alpha-\alpha\xi)-\alpha\Psi(1-\alpha\xi)-\frac{1}{1-\xi}\right\}$$
$$-(1-\xi)\Gamma(2-\xi)\frac{\Gamma(1-\alpha\xi)\Gamma(1+\alpha)}{\Gamma(1-\alpha\xi+\alpha)}\left\{\alpha\Psi(1+\alpha)+\Psi(2-\xi)-\alpha\Psi(1+\alpha-\alpha\xi)\right\}$$
$$+\Gamma(3-\xi)\frac{\Gamma(1+\alpha-\alpha\xi)\Gamma(1+\alpha)}{\Gamma(1+2\alpha-\alpha\xi)}\left\{\alpha\Psi(1+\alpha)+\Psi(3-\xi)-\alpha\Psi(1+2\alpha-\alpha\xi)\right\}$$

with analogous formulae for cov (Q,S) and cov $\{Q,(1-S)\log S\}$.