Mixed methods for fitting the GEV distribution

Pierre Ailliot
Université de Brest
France

Craig Thompson
National Institute of Water and Atmospheric Research
New Zealand

Peter Thomson
Statistics Research Associates Ltd
New Zealand

30 May 2008

Abstract

The generalised extreme-value (GEV) distribution is widely used for modelling and characterising extremes. It is a flexible 3-parameter distribution that combines three extreme-value distributions within a single framework: the Gumbel, Frechet and Weibull. Common methods used for estimating the GEV parameters are the method of maximum likelihood and the method of L-moments.

In this paper we generalise the mixed maximum likelihood and L-moments estimation procedures proposed by Morrison and Smith (2002) and derive the asymptotic properties of the resulting estimates. Analytic expressions are given for the asymptotic covariance matrices in a number of important cases, including the estimators proposed by Morrison and Smith (2002). These expressions are verified by simulation and the efficiencies of the various estimators established.

The asymptotic results are compared to those obtained for small samples, and the properties of the various estimators, including full maximum likelihood estimators and L-moment estimators, are considered. Finally, these methods are applied to an analysis of Wellington maximum daily rainfall data and graphical tools are developed, using simplified constraints for the support of the log-likelihood, which assist with the determination of the estimates in practice.

Keywords: GEV distribution; mixed estimation methods; asymptotic properties; small samples; quantile estimation.

1 Introduction

The distribution of extremes, such as maximum daily temperatures or minimum daily returns, is of common interest to many disciplines including the natural and social scientists. Katz et al. (2002) provides a comprehensive review of the analysis of hydrological extremes, and Coles (2001) provides a general introduction to the analysis of extreme values. Embrechts et al. (1997) and McNeil et al. (2005) consider the modelling of extremes in insurance, finance and quantitative risk management. However, for many analyses, there is often limited data available on extremes and so fitting an extreme value distribution can lead to difficulties, particularly for the estimation of extreme quantiles. In such cases there is a need to better understand the asymptotic and small sample properties of the parameters of the extreme value distribution chosen, and its quantiles.

This paper is concerned with the fitting of the generalised extreme-value (GEV) distribution, introduced by Jenkinson (1955), which is widely used for modelling and characterising extremes. The GEV distribution is a flexible 3-parameter distribution that combines three extreme-value distributions within a single framework: the Gumbel (EV1), Frechet (EV2) and Weibull (EV3). Common methods of estimating the GEV parameters are the method of maximum likelihood (Prescott and Walden, 1980) and the method of L-moments (Hosking, 1990). While L-moment estimators produce biased estimates, Hosking et al. (1985) found them to be preferable to maximum likelihood estimators in small samples because they resulted in estimated quantiles with smaller variances. For a range of hydrologically important cases, Madsen et al. (1997) showed that GEV quantiles estimated from small samples using conventional method of moments estimates were more accurate than those based on either maximum likelihood or L-moments. Martins and Stedinger (2000) considered restricting the shape of the GEV distribution using a suitable Bayesian prior. They showed that, for heavy tailed GEV distributions, the maximum a posteriori (MAP) estimators (penalised maximum likelihood estimators) again provided better estimates of GEV quantiles than maximum likelihood, method of moments and L-moments, in the case of small samples.

Morrison and Smith (2002) proposed two methods for estimating the GEV parameters that combine both maximum likelihood and L-moment methods. The resulting estimators of the shape parameter displayed reduced variance compared to the maximum likelihood estimator, and reduced bias compared to the L-moment estimator. Furthermore, their study showed the root mean square errors of the mixed method estimators were superior to those of L-moment estimators for flood size samples, although the root mean square errors of the corresponding quantiles were slightly inferior. However, while Morrison and Smith (2002) established that their methods produce estimates that are consistent and asymptotically Gaussian, they did not provide analytical expressions for the asymptotic covariance matrix.

In Section 2 we generalise the estimation procedure proposed by Morrison and Smith (2002) to moments other than L-moments, and derive the asymptotic properties of the resulting estimates. In particular, analytic expressions are given for the asymptotic covariance matrices in a number of important cases, including the estimators proposed by Morrison and Smith (2002). Using simplified constraints for the support of the log-likelihood, graphical tools are developed in Section 3 which assist with the determination of the estimates in practice. In Section 4, the asymptotic results established in Section 2 are verified by simulation and the efficiencies of the various estimators established. The asymptotic results are contrasted and compared to those obtained for small samples, and the properties of the various estimators, including full maximum likelihood estimators and the L-moment estimators, are considered. The corresponding quantile estimators are also assessed for accuracy and bias. The various methods are applied to an analysis of Wellington maximum daily rainfall data in Section 5 and conclusions are drawn in Section 6.

2 GEV estimation using mixed methods

Consider estimating the parameters $\alpha > 0$, β, κ of the GEV distribution with cumulative distribution function

$$F(x|\boldsymbol{\theta}) = \begin{cases} \exp\{-[1-\kappa\frac{x-\beta}{\alpha}]^{\frac{1}{\kappa}}\} & (\kappa \neq 0) \\ \exp\{-\exp[-\frac{x-\beta}{\alpha}]\} & (\kappa = 0) \end{cases} \tag{1}$$

where $\boldsymbol{\theta} = (\beta, \alpha, \kappa)^T$ and x is bounded above by $\beta + \alpha/\kappa$ when $\kappa > 0$, below by $\beta + \alpha/\kappa$ when $\kappa < 0$. Here β, α and κ are location, scale and shape parameters respectively, and the constraints on x are equivalent to requiring $\kappa(x-\beta) < \alpha$. Note that $\kappa < 0$ corresponds to the Frechet distribution, $\kappa > 0$ corresponds to the Weibull distribution, and the case $\kappa = 0$ (the Gumbel distribution) is the limit of the $\kappa \neq 0$ form as κ approaches zero,

The log-likelihood of a random sample $X_1, \ldots, X_n$ drawn from the GEV distribution is given by

$$\ln L(\boldsymbol{\theta}) = -n\ln\alpha - \sum_{i=1}^{n}(1-\kappa\frac{X_i-\beta}{\alpha})^{\frac{1}{\kappa}} + (\frac{1}{\kappa}-1)\sum_{i=1}^{n}\ln(1-\kappa\frac{X_i-\beta}{\alpha}) \tag{2}$$

and the corresponding maximum likelihood estimator is the value of $\boldsymbol{\theta}$ at which $\ln L(\boldsymbol{\theta})$ attains its maximum value. A feasible solution is subject to the parameter constraints

$$\alpha > 0, \qquad \kappa\frac{X_i-\beta}{\alpha} \leq 1 \qquad (i = 1, \ldots, n) \tag{3}$$

and we further assume that $-0.5 < \kappa < 0.5$. The latter ensures that the X_i have finite second moments and satisfy the maximum likelihood regularity conditions given in Smith (1985). In particular, these conditions on κ ensure that both L-moment and the maximum likelihood estimates are consistent with asymptotic (large sample) Gaussian distributions (see Prescott and Walden 1980, Hosking 1990).

Suppose now that $\boldsymbol{\theta}$ is estimated by a mixture of two methods; maximum likelihood and method of moments. A central case of interest here is where α and β are estimated as functions of the remaining parameter κ by the method of moments using the first two L-moments

$$\lambda_1 = E(X_1) = \beta + \frac{\alpha}{\kappa}(1-\Gamma(1+\kappa)), \quad \lambda_2 = \tfrac{1}{2}E(|X_1 - X_2|) = \frac{\alpha}{\kappa}(1-2^{-\kappa})\Gamma(1+\kappa) \tag{4}$$

with λ_1, λ_2 estimated by

$$\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{\lambda}_2 = \frac{1}{n(n-1)}\sum_{i<j} |X_i - X_j|. \tag{5}$$

Then κ is estimated by maximising the log-likelihood $\ln L(\boldsymbol{\theta})$ with α, β constrained to be the solution of (4) and (5). These estimators, denoted by $\hat{\boldsymbol{\theta}}$, are the MIX2 estimators proposed by Morrison and Smith (2002).

Note that $\lambda_3 = \kappa$ and (4) yield a continuously differentiable 1-1 mapping between the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^T$ where

$$\beta = \lambda_1 - \frac{\lambda_2(1-\Gamma(1+\lambda_3))}{(1-2^{-\lambda_3})\Gamma(1+\lambda_3)}, \quad \alpha = \frac{\lambda_2\lambda_3}{(1-2^{-\lambda_3})\Gamma(1+\lambda_3)}, \quad \kappa = \lambda_3 \tag{6}$$

so that $\hat{\boldsymbol{\theta}}$ is now given by (6) with $\boldsymbol{\lambda}$ replaced by $\hat{\boldsymbol{\lambda}}$ where $\hat{\lambda}_1$, $\hat{\lambda}_2$ are given by (5) and

$$\hat{\lambda}_3 = \arg\max_{\lambda_3} \ln L(\hat{\lambda}_1, \hat{\lambda}_2, \lambda_3). \tag{7}$$

This mixed estimation estimation procedure will be referred to as **method M1**.

From Results 1 and 2 in the Appendix, $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ has an asymptotic Gaussian distribution with mean zero and covariance matrix $J^T V J$ where

$$V = BCB^T, \quad C = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & 0 \\ 0 & 0 & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ I_{31}^{(\boldsymbol{\lambda})} & I_{32}^{(\boldsymbol{\lambda})} & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}$$

and $C_{ij} = \lim_{n\to\infty} \mathrm{cov}(\hat{\lambda}_i, \hat{\lambda}_j)$ for $i, j = 1, 2$. The Jacobean of the transformation (6) is given by

$$J = \frac{\partial \boldsymbol{\theta}^T}{\partial \boldsymbol{\lambda}} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{\beta-\lambda_1}{\lambda_2} & \frac{\alpha}{\lambda_2} & 0 \\ \frac{\alpha}{\kappa}\psi(1+\kappa) + (\beta-\lambda_1)\frac{\ln 2}{1-2^{\kappa}} & \frac{\alpha}{\kappa}(1-\kappa\psi(1+\kappa)+\kappa\frac{\ln 2}{1-2^{\kappa}}) & 1 \end{bmatrix} \tag{8}$$

where $\psi(x) = d\ln\Gamma(x)/dx$ is the digamma function and $\Gamma(x)$ is the gamma function. The matrix $\boldsymbol{I}^{(\boldsymbol{\lambda})} = J\boldsymbol{I}^{(\boldsymbol{\theta})}J^T$ where $\boldsymbol{I}^{(\boldsymbol{\theta})}$, the information matrix of the GEV distribution with parameters $\boldsymbol{\theta}$, is given by Result 2 in the Appendix. Since

$$\begin{aligned} n\ \mathrm{var}(\hat{\lambda}_1) &= \mathrm{var}(X_1), \qquad n\ \mathrm{cov}(\hat{\lambda}_1, \hat{\lambda}_2) = \mathrm{cov}(X_1, |X_1 - X_2|), \\ n\ \mathrm{var}(\hat{\lambda}_2) &= \frac{n-1}{n-2}\mathrm{cov}(|X_1-X_2|, |X_2-X_3|) + \frac{\mathrm{var}(|X_1-X_2|)}{2(n-1)} \end{aligned}$$

it follows that

$$C_{11} = \mathrm{var}(X_1), \;\; C_{12} = C_{21} = \mathrm{cov}(X_1, |X_1-X_2|), \;\; C_{22} = \mathrm{cov}(|X_1-X_2|, |X_2-X_3|)$$

and straightforward, but demanding, integration now yields

$$\begin{aligned} C_{11} &= \frac{\alpha^2}{\kappa^2}[\Gamma(1+2\kappa) - \Gamma(1+\kappa)^2] \\ C_{12} = C_{21} &= -\frac{\alpha^2}{\kappa^2}[(1-2^{-2\kappa})\Gamma(1+2\kappa) - 2(1-2^{-\kappa})\Gamma(1+\kappa)^2] \\ C_{22} &= \frac{\alpha^2}{\kappa^2}[\Gamma(1+2\kappa)(1+2^{-2\kappa+2}(H_\kappa(\tfrac{1}{2}) - \tfrac{1}{2})) - \Gamma(1+\kappa)^2(3 - 2^{-\kappa+2} + 2^{-2\kappa+2})] \end{aligned} \tag{9}$$

where

$$H_\kappa(x) = {}_2F_1(\kappa, 2\kappa; 1+\kappa; -x) = 2\frac{1+2\kappa}{x^\kappa}\frac{\Gamma(1+\kappa)^2}{\Gamma(2+2\kappa)}B_{1+\kappa}(\frac{x}{1+x}) + \frac{1-x}{(1+x)^{1+2\kappa}}$$

for $x > 0$, $\kappa > -1$. Here ${}_2F_1(.)$ is the hypergeometric function, and $B_\phi(.)$ the symmetric beta distribution function with parameter $\phi > 0$.

The formulae for the C_{ij} $(i, j = 1, 2)$ agree with those of Hosking et al. (1985). Note that we have given an alternative formula for $H_\kappa(x)$ involving the symmetric beta distribution function. This has the advantage of ready implementation in standard statistical computing packages where the beta distribution function is typically available.

Other mixed GEV estimation methods

The same arguments can be used to establish the asymptotic distribution of the estimators that result when only one moment estimate is used with the remaining two parameters estimated by maximum likelihood. Consider, for example, the first L-moment constraint for β and estimating α, κ by maximum likelihood subject to this constraint (this is the MIX1 method proposed in Morrison and Smith, 2002). In this case the mapping between $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ is given by

$$\lambda_1 = \beta + \frac{\alpha}{\kappa}(1 - \Gamma(1+\kappa)), \qquad \lambda_2 = \alpha, \qquad \lambda_3 = \kappa$$

with λ_1 estimated by $\hat{\lambda}_1$ as before and λ_2, λ_3 estimated by constrained maximum likelihood. Calling this estimation procedure **method M2**, the asymptotic distribution of the M2 estimators is again given by Results 1 and 2 in the Appendix, but now

$$C = \begin{bmatrix} C_{11} & 0 & 0 \\ 0 & I_{22}^{(\boldsymbol{\lambda})} & I_{23}^{(\boldsymbol{\lambda})} \\ 0 & I_{32}^{(\boldsymbol{\lambda})} & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ I_{21}^{(\boldsymbol{\lambda})} & I_{22}^{(\boldsymbol{\lambda})} & I_{23}^{(\boldsymbol{\lambda})} \\ I_{31}^{(\boldsymbol{\lambda})} & I_{32}^{(\boldsymbol{\lambda})} & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}, \quad vJ = \begin{bmatrix} 1 & 0 & 0 \\ J_{21} & 1 & 0 \\ J_{31} & 0 & 1 \end{bmatrix}$$

where

$$J_{21} = \frac{\beta - \lambda_1}{\alpha}, \qquad J_{31} = \frac{\alpha}{\kappa}\psi(1+\kappa)\Gamma(1+\kappa) - \frac{\beta - \lambda_1}{\kappa} \qquad \boldsymbol{I}^{(\boldsymbol{\lambda})} = J\boldsymbol{I}^{(\boldsymbol{\theta})}J^T$$

with C_{11} given in (9) and $\boldsymbol{I}^{(\boldsymbol{\theta})}$ given by Result 2 in the Appendix.

Finally, it is noted that constrained maximum likelihood with other forms of moment estimates could be chosen including, for example, robust estimates of location and scale. In the latter case, one might use the median and interquartile range for λ_1 and λ_2 respectively. Such estimates are not only robust to outliers, but also allow consideration of a wider range of values of κ including the infinite variance case when $\kappa \leq -0.5$. Since the second L-moment is proportional to Gini's mean difference, it is already a reasonably robust estimate of scale, but the first L-moment, the sample mean, is not. Thus, another possibility would be to replace the sample mean by the sample median or trimmed mean which will yield a more robust estimate of location. Once again, the same approach can be used in either case to establish the asymptotic distribution of the resulting estimators.

Consider the case of robust moment estimates of location and scale given by the sample median and the second L-moment $\hat{\lambda}_2$ respectively, where $\hat{\lambda}_2$ is given by (5). Here the mapping between $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ is given by

$$\lambda_1 = \beta + \frac{\alpha}{\kappa}(1 - (\ln 2)^\kappa), \qquad \lambda_2 = \frac{\alpha}{\kappa}(1 - 2^{-\kappa})\Gamma(1+\kappa), \qquad \lambda_3 = \kappa$$

with λ_1 estimated by the sample median, λ_2 estimated by (5), and λ_3 estimated by constrained maximum likelihood. Calling this estimation procedure **method M3**, the asymptotic distribution of the M3 estimators is the same as the M1 estimators, but with C_{11}, $C_{12} = C_{21}$ and J_{31} replaced by

$$\begin{aligned}
C_{11} &= \alpha^2 (\ln 2)^{2\kappa-2} \\
C_{12} &= \frac{\alpha^2}{\kappa}(\ln 2)^{\kappa-1}\Gamma(1+\kappa)(3 - 2^{-\kappa+1} - 4G_{1+\kappa}(\ln 2) + 2^{-\kappa+2}G_{1+\kappa}(2\ln 2)) \\
J_{31} &= \frac{\alpha}{\kappa}(\ln 2)^\kappa \ln\ln 2 + (\beta - \lambda_1)(\frac{\ln 2}{1-2^\kappa} - \psi(1+\kappa))
\end{aligned}$$

where $G_\kappa(.)$ denotes the survivor function of the Gamma distribution with shape parameter κ and unit scale parameter. The formula for C_{12} was derived using results in Ferguson (1996) and, in particular, a generalisation of the argument given in Ferguson (1999).

Using these results, Figure 1 plots the square-root of the asymptotic relative efficiency of the M1, M2 and M3 estimators of β, α and κ for $\beta = 0$, $\alpha = 1$ and $-0.5 < \kappa < 0.5$. Here the asymptotic root-efficiency, expressed as a percentage, gives the ratio of the large-sample standard deviation of the maximum likelihood estimator of the given parameter to the large-sample standard deviation of its estimator.

For $\kappa > -0.35$ all asymptotic root-efficiencies exceed 80% and for $\kappa > -0.4$ all root-efficiencies exceed 70%. Although method M3 is generally less efficient than Methods M1 and M2 (particularly for the location parameter β), it does offer better estimates for values of κ near -0.5. This is expected since the median can be shown to be a better location estimator than the mean when $\kappa < -0.1$, and is considerably better when $\kappa < -0.4$ (the GEV distribution is a heavy-tailed Frechet distribution for $\kappa < 0$). With the exception of the location estimate, methods M1 and M3 are comparable so that the choice of using a robust location estimator is not too costly in terms of asymptotic efficiency.

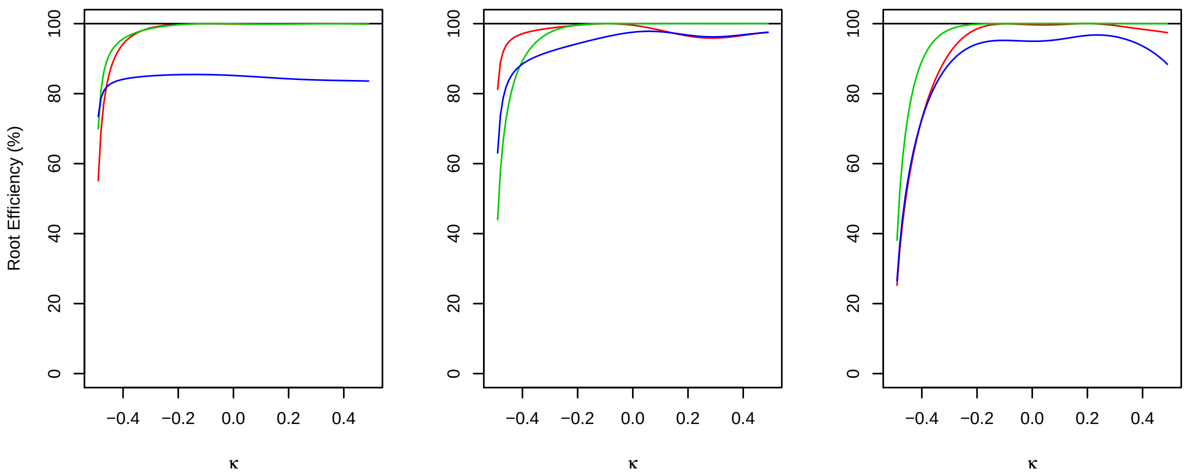

Figure 1: Percentage root-efficiency of the M1 (red), M2 (green) and M3 (blue) estimators of β (left panel), α (middle panel) and κ (right panel) for $\beta = 0$, $\alpha = 1$ and $-0.5 < \kappa < 0.5$.

3 Computational issues

All the methods considered, including maximum likelihood, involve constrained optimisation of the log-likelihood (2). The constraints fall into three groups; constraints (3) related to the support of the GEV density, moment constraints such as (4) with λ_1, λ_2 estimated by (5), and the technical constraint $-0.5 < \kappa < 0.5$. The latter condition ensures that the family of GEV distributions under consideration have finite variance and satisfy maximum likelihood regularity conditions (see Smith 1985). This restriction is unlikely to be a major problem in practice. In hydrology it is commonly assumed that κ lies in a more restrictive range (see Hosking et al. 1985, Martins and Stedinger 2000, for example). However, and as noted in Katz et al. (2002), bound constraints for κ are commonly adopted for technical convenience at the possible expense, in some cases, of physical interpretation.

Imposing constraints on κ is known to improve the performance of likelihood-based GEV estimation techniques in small samples. Martins and Stedinger (2000) report implausibly large estimates of κ for unconstrained maximum likelihood using small samples and show that this problem is eliminated when the range of κ is restricted using a suitable Bayesian prior. Coles and Dixon (1999) use penalised maximum likelihood to restrict the range of κ, resulting in maximum likelihood estimates with improved small sample properties. Morrison and Smith (2002) impose moment constraints which have a similar effect. This is similarly the case here as shown in Section 4.

The estimation strategy we have adopted is to optimise over all parameters other than κ, taking careful account of the constraints, and form a profile log-likelihood that is a function of κ alone. This profile log-likelihood can then be plotted and optimised over κ. The optimisations involved have reduced dimension and take advantage of simplified constraints. As shown in Section 4, this procedure proves to be numerically robust, graphically informative, and computationally efficient.

Consider first the estimation method M1. The constraint (3) is equivalent to

$$\alpha > 0 \qquad \alpha > \kappa(\max_i X_i - \beta) \qquad \alpha > -\kappa(\beta - \min_i X_i) \tag{10}$$

which yields

$$\hat{\lambda}_2 > (\max_i X_i - \hat{\lambda}_1)(1 - 2^{-\kappa}), \qquad \hat{\lambda}_2 > -(\hat{\lambda}_1 - \min_i X_i)(1 - 2^{-\kappa})$$

after substituting for α, β using (4) and (5). Since $\min_i X_i < \hat{\lambda}_1 < \max_i X_i$ this relation can be inverted to give $\kappa_{M1}^- < \kappa < \kappa_{M1}^+$ where

$$\kappa_{M1}^- = -\ln\left(1 + \frac{\hat{\lambda}_2}{\hat{\lambda}_1 - \min_i x_i}\right) / \ln 2, \qquad \kappa_{M1}^+ = -\ln\left(1 - \frac{\hat{\lambda}_2}{\max_i X_i - \hat{\lambda}_1}\right) / \ln 2$$

and $\kappa_{M1}^- < 0$, $\kappa_{M1}^+ > 0$ unless the sample is degenerate. The intersection of this interval with $(-0.5, 0.5)$ yields the simple constraint

$$\max(\kappa_{M1}^-, -0.5) < \kappa < \min(\kappa_{M1}^+, 0.5). \tag{11}$$

Note that these bounds are determined solely from the data. They lead to computationally simple, transparent and numerically robust parameter estimates which can be used in their own right, or as initial estimates to more computationally intensive methods such as full maximum likelihood.

In the case of maximum likelihood, only the constraints (10) and $-0.5 < \kappa < 0.5$ apply. However, we further assume that

$$\min_i X_i \leq \beta \leq \max_i X_i \tag{12}$$

since β is the $e^{-1} = 0.3679$ quantile of the GEV distribution. The probability that the interval $[\min_i X_i, \max_i X_i]$ does not include β is

$$1 - P(\min_i X_i \leq \beta \leq \max_i X_i) = e^{-n} + (1 - e^{-1})^n$$

which is less than 10^{-4} for $n > 20$ and less than 10^{-5} for $n > 25$. Thus, provided n is not very small, it will be reasonable in practice to assume that (12) holds so that β always lies within the range of the data. Note that this constraint will only impact on small sample performance and vanishes for large samples. Now constraint (10) becomes

$$\alpha > \alpha_{ML} = \begin{cases} \kappa(\max_i X_i - \beta) & (\kappa \geq 0) \\ -\kappa(\beta - \min_i X_i) & (\kappa < 0) \end{cases} \tag{13}$$

where α_{ML} is a simple piecewise linear function of β given κ and the data. The simple constraints (12) and (13) can now be used to determine the profile log-likelihood $\max_{\beta,\alpha} \ln L(\boldsymbol{\theta})$ which is a function of κ alone and the data.

Similar considerations apply to methods M2 and M3. For method M2 the constraints are

$$\alpha > \alpha_{M2} = \begin{cases} \kappa(\max_i X_i - \hat{\lambda}_1)/\Gamma(1+\kappa) & (\kappa \geq 0) \\ -\kappa(\hat{\lambda}_1 - \min_i X_i)/\Gamma(1+\kappa) & (\kappa < 0) \end{cases} \tag{14}$$

and $-0.5 < \kappa < 0.5$. For method M3 there is the single constraint

$$\max(\kappa_{M3}^{-}, -0.5) < \kappa < \min(\kappa_{M3}^{+}, 0.5) \tag{15}$$

where $\kappa_{M3}^{-} < 0$ and $\kappa_{M3}^{+} > 0$ are the solutions to

$$\frac{1-2^{-\kappa}}{(\ln 2)^{\kappa}}\Gamma(1+\kappa) = -\frac{\hat{\lambda}_2}{\hat{\lambda}_1 - \min_i X_i}, \qquad \frac{1-2^{-\kappa}}{(\ln 2)^{\kappa}}\Gamma(1+\kappa) = \frac{\hat{\lambda}_2}{\max_i X_i - \hat{\lambda}_1}$$

respectively. Method M3 shares the same properties of transparency and computational simplicity as method M1.

For each of the four methods considered, it is readily shown that the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ does not depend on β and takes the form

$$\begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \Omega \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{16}$$

where the 3-dimensional matrix Ω has elements that depend only on κ. This allows for the simple tabulation and evaluation of the asymptotic covariance matrices for the case $\beta = 0$, $\alpha = 1$, with the more general cases obtained by scaling the results using (16).

4 Numerical study and an application

A simulation study was undertaken to check how well the asymptotic results given in Section 2 applied in practice. Three sample sizes were considered, $n = 30$, 60 and 120, for random samples from a GEV distribution with $\beta = 0$, $\alpha = 1$ and a range of values of κ between -0.5 and 0.5. For each choice of κ, 1000 independent random samples were generated and the M1, M2, M3 and maximum likelihood estimates determined. Using these simulations, the estimated bias and root-mean-squared error for each method and choice of κ was then computed as a function of κ. The result for $n = 30$ is plotted in Figure 2 together with the asymptotic root-mean-squared errors (standard errors in this case since the estimates are asymptotically unbiased).

For $n = 30$ all methods show reasonable agreement with the asymptotic results when $\kappa > -0.3$, both in terms of the estimated bias (which should be close to zero) and the estimated root-mean-squared-error. Note, however, the small underestimation of the scale parameter α in the case of maximum likelihood and the closely allied method M2. When $\kappa < -0.3$ biases become more evident, particularly for the methods based on the L-moments whose variances become infinite as κ approaches -0.5. For the estimates of κ, the estimated root-mean-squared-error was generally less that the asymptotic value as might be expected given that the κ estimates were restricted to the interval (-0.5,0.5). In such cases a censored form of the asymptotic Gaussian distribution may prove to be a more appropriate approximation. Although all methods performed comparably for $\kappa > -0.3$, maximum likelihood generally performed well over all the values of κ considered.

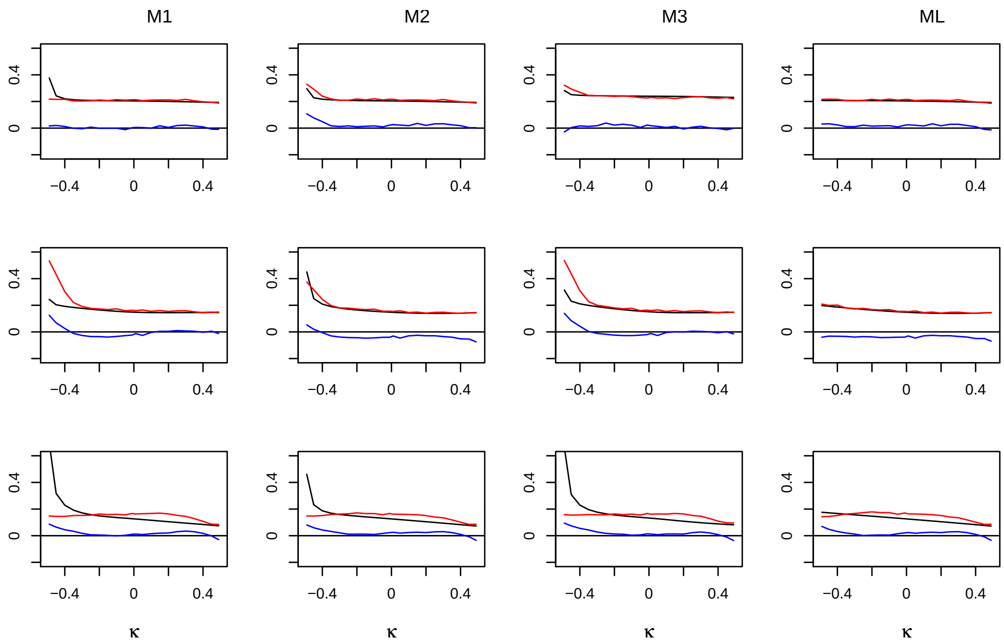

Figure 2: Plots of the asymptotic root-mean-squared-error (black), simulated root-mean-squared-error (red) and simulated bias (blue) of the M1 (column 1), M2 (column 2), M3 (column 3) and maximum likelihood (ML, column 4) estimators of $\beta = 0$ (row 1), $\alpha = 1$ (row 2) and κ (row 3) for $n = 30$ and $-0.5 < \kappa < 0.5$. The simulated results are based on 1000 replications.

As expected, the results for $n = 60$ and $n = 120$ showed improvements over those for $n = 30$. Nevertheless, although considerably reduced, residual biases for $\kappa < -0.35$ remained even for $n = 120$. As before, maximum likelihood performed well over all κ in the interval (-0.5, 0.5). Simulation of very large samples ($n = 1000$) at selected values of $\kappa < -0.3$ served to validate the formulae for the asymptotic variances given in Section 2.

We now consider an application of these mixed methods for fitting the GEV distribution to a sample of 24-hour annual maximum rainfall from Wellington, New Zealand (41.286 °S, 174.767 °E) over the period 1940 to 1999. Summary statistics of the data are given in Table 1, both for the entire data set and for the data disaggregated by the two phases

	- IPO	+ IPO	All
Sample size	31	29	60
Mean (first L-moment)	82.0	75.5	78.9
Median	73.9	70.1	73.6
Standard deviation	28.7	19.0	24.5
Second L-moment	16.3	10.4	13.6
Maximum	153.2	121.2	153.2
Minimum	47.0	51.4	47.0

Table 1: 24-hour annual maximum rainfall statistics (mm) for Wellington, New Zealand, over the period 1940-1999 and for each phase of the IPO. The negative phase of the IPO spans the period 1946-1976, and the positive phases from 1940-1945, and from 1977-1999.

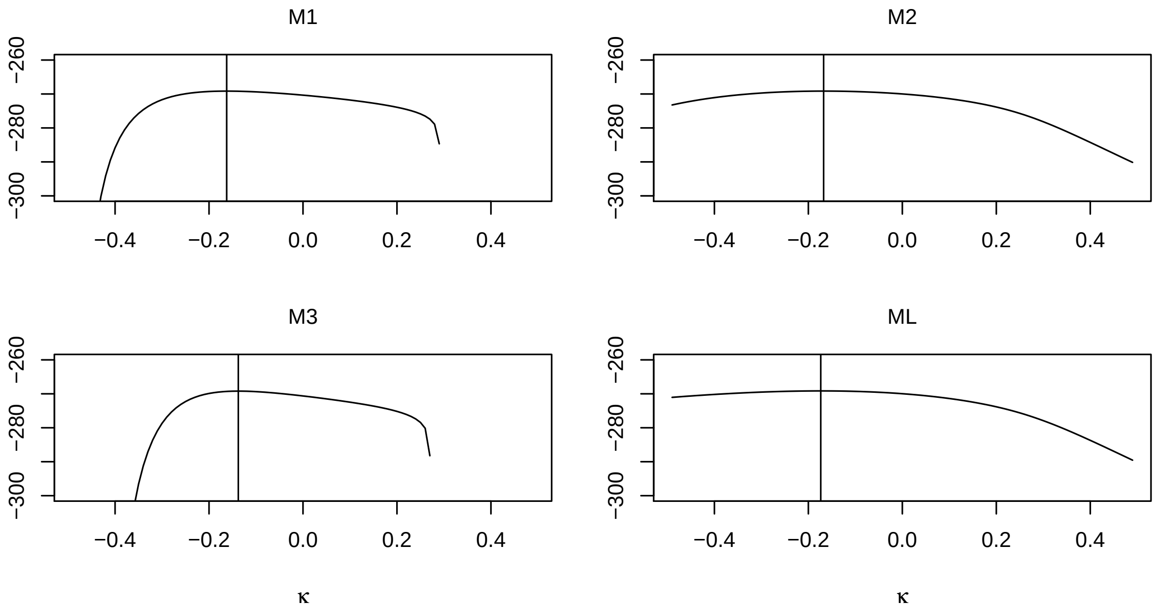

Figure 3: Profile log-likelihoods for the M1 (top left), M2 (top right), M3 (bottom left) and maximum likelihood (bottom right) methods applied to Wellington 24-hour annual maximum rainfall over the period 1940-1999 using L-moments,. The vertical lines mark the respective estimates of κ.

(positive and negative) of the Interdecadal Pacific Oscillation (IPO). The IPO is associated with large-scale atmospheric and oceanic patterns over the Pacific Ocean, and has been shown to have an impact on annual maximum rainfall series in some regions of New Zealand (see Thompson, 2006).

The estimation methods described in Section 2 were first applied to the entire data set. For comparison purposes, a corresponding analysis using L-moments (see Hoskings et al., 1985) was also undertaken. Plots of the resulting profile likelihoods for methods M1, M2, M3 and maximum likelihood are given in Figure 3. Note that, in this case, the limits (11) and (15) are -0.50, 0.29 and -0.45, 0.27 respectively, while methods M2 and maximum likelihood have profile likelihoods defined over the entire range considered ($-0.5 < \kappa < 0.5$). The curvature of any of these profile likelihoods at its point of maxima is proportional to the standard error of the corresponding estimate.

Table 2 shows the various parameter estimates and their (asymptotic) standard errors where the latter are given in Section 2 for methods M1, M2, M3 and maximum likelihood. In the case of L-moments the standard errors were obtained from tables provided in Hosking et al. (1985) and Madsen et al. (1997). The parameter estimates from methods M1, M2, M3 and maximum likelihood are all very similar, and slightly different from those obtained from the method of L-moments. However the latter differences are not statistically significant, being approximately one standard error apart. Nevertheless it is interesting to note that the likelihood-based methods all preferred a slightly heavier tail which may have ramifications for quantile estimation which is sensitive to the estimates of the shape parameter κ.

Method	β	α	κ
L	66.98 (2.67)	18.22 (2.12)	-0.07 (0.10)
M1	66.29 (2.40)	16.45 (1.87)	-0.16 (0.10)
M2	66.19 (2.40)	16.43 (1.87)	-0.17 (0.10)
M3	67.18 (2.89)	16.94 (1.98)	-0.14 (0.10)
ML	66.32 (2.41)	16.54 (1.89)	-0.17 (0.10)

Table 2: Parameter estimates obtained by fitting the GEV distribution to Wellington 24-hour annual maximum rainfall over the period 1940-1999 using L-moments, methods M1, M2, M3 and maximum likelihood (ML). Asymptotic standard errors are given in parentheses.

This analysis was repeated for the observations falling in each phase of the IPO and the results are reported in Table 3. For each IPO phase the likelihood-based estimates again prefer a heavier tail than the method of L-moments, with the differences in the various parameter estimates within each phase being much smaller than those between the two IPO phases. In particular, the differences between the scale parameters for each IPO phase suggest real differences, although these are still not statistically significant, with the annual extreme rainfalls being more variable and heavier during the negative phase of the IPO than in the positive phase (see also Table 1). These conclusions are further reinforced by a more formal likelihood ratio test which gives a p-value of 15% and retains the null hypothesis that thee is no difference between the 24-hour annual maximum rainfall within each phase of the IPO. This result is also consistent with the findings of Thompson (2006).

	- IPO			+ IPO		
Method	β	α	κ	β	α	κ
L	68.30 (4.71)	23.24 (3.62)	-0.01 (0.14)	65.99 (2.77)	13.06 (2.30)	-0.14 (0.16)
M1	66.96 (4.07)	20.05 (3.15)	-0.15 (0.14)	65.66 (2.55)	12.10 (2.02)	-0.20 (0.15)
M2	66.98 (4.07)	20.05 (3.14)	-0.15 (0.14)	65.87 (2.55)	12.12 (2.01)	-0.18 (0.15)
M3	66.47 (4.66)	19.65 (3.27)	-0.17 (0.15)	65.54 (2.94)	11.98 (2.13)	-0.21 (0.16)
ML	67.13 (4.09)	20.16 (3.17)	-0.16 (0.14)	65.97 (2.56)	12.21 (2.02)	-0.19 (0.15)

Table 3: Parameter estimates obtained by fitting the GEV distribution to Wellington 24-hour annual maximum rainfall over the period 1940-1999 using methods M1, M2, M3 and maximum likelihood (ML). Asymptotic standard errors are given in parentheses. The first 4 columns are for negative IPO; the second four rows are for positive IPO

5 Conclusions

Mixed maximum likelihood and L-moments estimation procedures proposed by Morrison and Smith (2002) have been generalised to include other forms of moment estimators. The asymptotic properties of the resulting estimators have been derived and analytic expressions given for the asymptotic covariance matrices in a number of important cases. These expressions were verified by simulation and the efficiencies of the various estimators established. Furthermore, the estimation procedures developed take careful account of the various parameter constraints and use profile likelihoods for estimating κ. An advantage of the latter is that they are amenable to simple graphical analysis and so, in practice, have the potential to provide a better understanding of the the strengths and weaknesses of the data being modelled.

The asymptotic properties of these likelihood-based methods compare favourably with those obtained from finite samples with sample sizes as small as $n = 30$ and for values of $\kappa > -0.3$. In particular, for $n \geq 30$, the method of maximum likelihood generally performed well over all values of κ in the range $-0.5 < \kappa < 0.5$ considered. Finally, the methods were applied to 24-hour annual maximum rainfall in Wellington, New Zealand, to try and detect any differences between the GEV distributions fitted to the positive and negative phases of the IPO. No significant differences were found.

The impact of these estimation methods on the corresponding GEV quantile estimators has yet to be assessed and remains the focus of further research. Bias correction based on the censored asymptotic Gaussian distributions of the estimates of κ will also be investigated.

Appendix

Consider estimating the parameters $\boldsymbol{\theta} = (\beta, \alpha, \kappa)^T$ of the GEV distribution (1) from a random sample $\mathbf{X} = (X_1, \ldots, X_n)^T$ drawn from that distribution. Let

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta}) \tag{17}$$

be a reparameterisation of $\boldsymbol{\theta}$ where $\lambda_3 = \kappa$ and $\boldsymbol{\lambda}(\boldsymbol{\theta})$ is a continuously differentiable 1-1 mapping between $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ so that the log-likelihood $\ln L(\boldsymbol{\lambda})$ based on $\boldsymbol{\lambda}$ is formally equivalent to $\ln L(\boldsymbol{\theta})$ given by (2). Assume that $\boldsymbol{\lambda}$ is estimated by $\hat{\boldsymbol{\lambda}}$ where

$$\hat{\lambda}_1 = \hat{\lambda}_1(\mathbf{X}), \quad \hat{\lambda}_2 = \hat{\lambda}_2(\mathbf{X}), \quad \hat{\lambda}_3 = \arg\max_{\lambda_3} \ln L(\hat{\lambda}_1, \hat{\lambda}_2, \lambda_3) \tag{18}$$

and $\hat{\lambda}_1$, $\hat{\lambda}_2$ are known unbiased estimators of λ_1, λ_2. Further assume that the joint distribution of $\sqrt{n}(\hat{\lambda}_1 - \lambda_1)$, $\sqrt{n}(\hat{\lambda}_2 - \lambda_2)$ is asymptotically Gaussian and known. Now estimate $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\hat{\boldsymbol{\lambda}})$$

where $\boldsymbol{\theta}(\boldsymbol{\lambda})$ denotes the inverse of the mapping (17).

We now determine the asymptotic properties of $\hat{\boldsymbol{\lambda}}$ and then, using these, the required asymptotic properties of $\hat{\boldsymbol{\theta}}$.

Result 1 *Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with GEV distribution (1) and parameters $\boldsymbol{\lambda}$ given by (17) with $-\frac{1}{2} < \lambda_3 = \kappa < \frac{1}{2}$. If $\boldsymbol{\lambda}$ is estimated by $\hat{\boldsymbol{\lambda}}$ given by (18), then $\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ is asymptotically Gaussian with zero mean* *and covariance matrix* $V = BCB^T$ *where*

$$C = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & 0 \\ 0 & 0 & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ I_{31}^{(\boldsymbol{\lambda})} & I_{32}^{(\boldsymbol{\lambda})} & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}$$

with

$$C_{ij} = \lim_{n\to\infty} n \operatorname{cov}(\hat{\lambda}_i, \hat{\lambda}_j) \qquad (i, j = 1, 2)$$

and $\boldsymbol{I}^{(\boldsymbol{\lambda})}$ *is the information matrix of the GEV distribution with parameters* $\boldsymbol{\lambda}$*.*

Proof

The consistency of $\hat{\boldsymbol{\lambda}}$ follows by assumption and from the asymptotic results given in Smith (1985). To establish the asymptotic distribution of $\hat{\boldsymbol{\lambda}}$ we expand $h(\boldsymbol{\lambda}) = n^{-1}\partial \ln L(\boldsymbol{\lambda})/\partial\lambda_3$ in a first order Taylor series about $\boldsymbol{\lambda}^{(0)}$, the true value of $\boldsymbol{\lambda}$. This yields

$$0 = h(\hat{\boldsymbol{\lambda}}) = h(\boldsymbol{\lambda}^{(0)}) + \frac{\partial h(\boldsymbol{\lambda}^{(0)})}{\partial \boldsymbol{\lambda}^T}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^{(0)}) + o(|\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^{(0)}|)$$

so that

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ & -\partial h(\boldsymbol{\lambda}^{(0)})/\partial\boldsymbol{\lambda}^T & \end{bmatrix} \sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^{(0)}) = \sqrt{n} \begin{bmatrix} \hat{\lambda}_1 - \lambda_1^{(0)} \\ \hat{\lambda}_2 - \lambda_2^{(0)} \\ h(\boldsymbol{\lambda}_0) \end{bmatrix} + o(\sqrt{n}|\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^{(0)}|).$$

The matrix on the left hand side of the above converges to B^{-1} and, using standard central limit theorems (see Hosking, 1990, and Lee, 1990, for example) the first term on the right hand side is asymptotically Gaussian with zero mean. Note that $\sqrt{n}h(\boldsymbol{\lambda}^{(0)})$ is asymptotically independent of $\hat{\lambda}_1$ and $\hat{\lambda}_2$. This follows since $\hat{\lambda}_1$, $\hat{\lambda}_2$ are unbiased and

$$\operatorname{cov}(g(\mathbf{X}), \frac{\partial \ln L(\boldsymbol{\lambda})}{\partial\lambda_3}) = E(g(\mathbf{X})\frac{\partial \ln L(\boldsymbol{\lambda})}{\partial\lambda_3}) = \frac{\partial}{\partial\lambda_3}E(g(\mathbf{X}))$$

for any statistic $g(\mathbf{X}) = g(X_1, \ldots, X_n)$ whose expectation is finite and continuously differentiable with respect to $\boldsymbol{\lambda}$. Since the variance of $\sqrt{n}h(\boldsymbol{\lambda}^{(0)})$ converges to $I_{33}^{(\boldsymbol{\lambda})}$ the result follows.

Result 2 *Under the conditions of Result 1, if* $\boldsymbol{\theta} = (\beta, \alpha, \kappa)^T$ *is estimated by* $\hat{\boldsymbol{\theta}}$ *given by (17) with* $\boldsymbol{\lambda}$ *replaced by* $\hat{\boldsymbol{\lambda}}$*, then* $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ *is asymptotically Gaussian with zero mean and covariance matrix* $\Sigma = J^T V J$ *where* J *is the Jacobean of the transformation given by* $J = \partial\boldsymbol{\theta}^T/\partial\boldsymbol{\lambda}$*. Moreover, the information matrix* $\boldsymbol{I}^{(\boldsymbol{\lambda})}$ *of the GEV distribution with parameters* $\boldsymbol{\lambda}$ *is given by*

$$\boldsymbol{I}^{(\boldsymbol{\lambda})} = J\boldsymbol{I}^{(\boldsymbol{\theta})}J^T$$

where the information matrix $\boldsymbol{I}^{(\boldsymbol{\theta})}$ *of the GEV distribution with parameters* $\boldsymbol{\theta}$ *is given by*

$$\frac{1}{\alpha^2\kappa^2}\begin{bmatrix} \kappa^2(\delta_2 - 1) & \kappa(\delta_2 - 1 - \Gamma(2-\kappa)) & -\alpha(\delta_2 - 1 + \kappa\delta_3) \\ \kappa(\delta_2 - 1 - \Gamma(2-\kappa)) & \delta_2 - 2\Gamma(2-\kappa) & -\frac{\alpha}{\kappa}(\delta_2 - \Gamma(2-\kappa) + \kappa(\delta_3 - \delta_1)) \\ -\alpha(\delta_2 - 1 + \kappa\delta_3) & -\frac{\alpha}{\kappa}(\delta_2 - \Gamma(2-\kappa) + \kappa(\delta_3 - \delta_1)) & \frac{\alpha^2}{\kappa^2}(\delta_2 + 2\kappa(\delta_3 - \delta_1) + \kappa^2(\delta_1^2 + \frac{\pi^2}{6})) \end{bmatrix}$$

with

$$\delta_1 = 1 - \gamma, \quad \delta_2 = 1 + (1-\kappa)^2\Gamma(1-2\kappa), \quad \delta_3 = \Gamma(2-\kappa)(\psi(1-\kappa) - \frac{1-\kappa}{\kappa})$$

and γ *denotes Euler's constant.*

Result 2 follows from Result 1 and the transformation (17). The information matrix $\boldsymbol{I}^{(\boldsymbol{\theta})}$ is given in Prescott and Walden (1980).

The same arguments can be used to establish the asymptotic distribution of the estimators that result when only one moment estimate is used with the remaining two estimated by maximum likelihood. Consider, for example, using a moment constraint for the location parameter β and estimating α, κ by maximum likelihood subject to this constraint (if the first L-moment is used then this gives the MIX1 method proposed in Morrison and Smith, 2002). Here the mapping between $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ is given by (17) with $\lambda_2 = \alpha$, $\lambda_3 = \kappa$ and now $\boldsymbol{\lambda}$ is estimated by $\hat{\boldsymbol{\lambda}}$ with λ_1 estimated by $\hat{\lambda}_1(\mathbf{X})$ as before, and λ_2, λ_3 are estimated by constrained maximum likelihood. With these changes Results 1 and 2 continue to apply with

$$C = \begin{bmatrix} C_{11} & 0 & 0 \\ 0 & I_{22}^{(\boldsymbol{\lambda})} & I_{23}^{(\boldsymbol{\lambda})} \\ 0 & I_{32}^{(\boldsymbol{\lambda})} & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ I_{21}^{(\boldsymbol{\lambda})} & I_{22}^{(\boldsymbol{\lambda})} & I_{23}^{(\boldsymbol{\lambda})} \\ I_{31}^{(\boldsymbol{\lambda})} & I_{32}^{(\boldsymbol{\lambda})} & I_{33}^{(\boldsymbol{\lambda})} \end{bmatrix}, \quad J = \begin{bmatrix} 1 & 0 & 0 \\ J_{21} & 1 & 0 \\ J_{31} & 0 & 1 \end{bmatrix}.$$

Other possible subsets of constraints can also be handled in a similar way.

References

Coles, S.G. (2001) *An Introduction to the Statistical Modeling of Extreme Values.* Springer Series in Statistics.

Coles, S.G. and Dixon, M.J. (1999) Likelihood-based inference for extreme value models. *Extremes* **2**, 5-23.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer.

Ferguson, T.S. (1996) *A Course in Large Sample Theory.* Chapman & Hall.

Ferguson, T.S. (1999) Asymptotic joint distribution of sample mean and a sample quantile. Unpublished paper. http://www.math.ucla.edu/ tom/papers/unpublished.html

Jenkinson, A.F. (1955) The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* **81**, 158-171.

Hosking, J.R.M. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B* **52**, 105-124.

Hosking, J.R.M., Wallis J.R. and Wood E.F. (1985) Estimation of the generalised extreme value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251-261.

Katz, R.W., Parlange, M.B. and Naveau, P. (2002) Statistics of extremes in hydrology. *Advances in Water Resources* **25**, 1287-1304.

McNeil, A.J., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press.

Madsen, H., Rasmussen, P.F. and Rosbjerg, D. (1997) Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events 1. At-site modeling. *Water Resources Research* **33**, 747-757.

Martins, E.S. and Stedinger, J.R. (2000) Generalised maximum-likelihood generalised extreme-value quantile estimators for hydrologic data. *Water Resources Research* **36**, 737-744.

Morrison J.E. and Smith J.A. (2002) Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research* **38**, 1305, doi:10.1029/2001WR000502.

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.

Thompson, C.S. (2006) Decadal climate variability of extreme rainfalls in New Zealand. *Weather and Climate* **26**, 3-20.