CHAPTER 3

PARAMETRIC APPROACH TO MODELLING EXTREMES

3.1 Introduction

By almost universal consent, a starting point for modelling the extremes of a process is based on distributional models derived from asymptotic theory. The parametric approach to modelling extremes is based on the assumption that the data in hand (X_1, X_2, ..., X_n) form an i.i.d. sample, having an exact generalized extreme-value d.f.

$$H_\theta(x) = H_{\gamma;\mu,\sigma}(x) = \exp\left\{-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\},$$

where $1+\gamma\frac{x-\mu}{\sigma} > 0$, and $\theta = (\gamma,\mu,\sigma) \in \Re\times\Re\times\Re_+$.

Of less common agreement is the method of inference by which such asymptotic models are fitted to the data. As both fashions and technologies have developed, so procedures for parameter estimation in extreme value models have sprouted. Actually, under the assumption of exact distributional form, standard statistical methodology from parametric estimation theory can be utilized in order to derive estimates of the parameters θ.

In general, this approach may seem restrictive and not very realistic, still it is rather valid (and has been extensively used) in the following setting:
As we have shown, the generalized extreme-value d.f. can be derived as the limiting d.f. of the maximum X of an i.i.d. sample Y_1, Y_2, ..., Y_m ($m \to \infty$). Consequently, a sample of such maxima X_1, X_2, ..., X_n, each of which comes from an independent sample Y_{i1}, Y_{i2}, ..., Y_{im} ($m_i \to \infty$, i=1, ...n), can be reasonably assumed to have a generalized extreme-value d.f.

So, in practice, the parametric approach to modelling extremes is adopted whenever our dataset is consisted of maxima of independent samples (e.g. in hydrology we have disjoint time periods). This method is often called method of *block maxima*. In the sequel, we present the most common estimation methods for the generalized extreme-value distribution.

3.2 Maximum Likelihood Estimation

Since we are in the realm of traditional parametric estimation, the most prominent estimation method is obviously the maximum likelihood estimation method.

If a r.v. X has generalized extreme-value d.f.

$$H_\theta(x) = H_{\gamma;\mu,\sigma}(x) = \exp\left\{-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\}$$

(the case $\gamma=0$ is derived by taking the limit in 0),

the corresponding p.d.f. is given by the formula

$$h_\theta(x) = h_{\gamma;\mu,\sigma}(x) = \begin{cases} \dfrac{1}{\sigma}\left(1+\gamma\dfrac{x-\mu}{\sigma}\right)^{-1/\gamma-1} \exp\left\{-\left(1+\gamma\dfrac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\}, & \text{for } 1+\gamma\dfrac{x-\mu}{\sigma}>0 \\ 0, & \text{elsewhere.} \end{cases}$$

Consequently, the likelihood function based on data $X=(X_1, ..., X_n)$ is given by

$$L(\theta;X) = \prod_{i=1}^{n} h_\theta(x_i) = \prod_{i=1}^{n}\frac{1}{\sigma}\left(1+\gamma\frac{x_i-\mu}{\sigma}\right)^{-1/\gamma-1} \exp\left\{-\left(1+\gamma\frac{x_i-\mu}{\sigma}\right)^{-1/\gamma}\right\}$$

$$= \sigma^{-n}\left[\prod_{i=1}^{n}\left(1+\gamma\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\gamma-1} \exp\left\{-\sum_{i=1}^{n}\left(1+\gamma\frac{x_i-\mu}{\sigma}\right)^{-1/\gamma}\right\},$$

for $1+\gamma\frac{x_i-\mu}{\sigma}>0$ for all x_i, and 0 elsewhere.

The corresponding log-likelihood would be

$$l(\theta;X) = -n\ln\sigma-(1+\gamma)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} e^{-y_i}\text{, where } y_i = \gamma^{-1}\ln\left(1+\gamma\frac{x_i-\mu}{\sigma}\right) \qquad (3.1)$$

Then, by definition, the maximum likelihood estimator (MLE) $\hat{\theta}=(\hat{\gamma},\hat{\mu},\hat{\sigma})$ for the unknown parameters $\theta=(\gamma,\mu,\sigma)$ equals

$$\hat{\theta} = \arg\max_{\theta\in\Theta} l(\theta;X).$$

As long as the estimation of large quantiles x_p is concerned, the equi-variance property of maximum likelihood implies that the MLE of a quantile is obtained by substitution of the MLE's $\hat{\theta} = (\hat{\gamma}, \hat{\mu}, \hat{\sigma})$ into the quantile function (generalized inverse d.f.) of the generalized extreme-value distribution

$$x_p = \mu - \frac{\sigma}{\gamma}\left[1 - (-\ln p)^{-\gamma}\right].$$

Differentiating equation (3.1) with respect to the parameters $\theta = (\gamma, \mu, \sigma)$ yields the likelihood system of equations. Clearly, no explicit solution exists for these equations, so the likelihood equations must be solved iteratively. Numerical procedures such as variants of Newton-Raphson algorithm are required.

Hosking (1985) published a FORTRAN algorithm for MLE of the parameters of generalized extreme-value d.f., based on Newton-Raphson iteration method, with some modifications designed to improve the speed and probability of convergence. This numerical method can be described as follows.

The Newton-Raphson method solves the likelihood equations $\partial l / \partial \theta$ by the iteration

$$\theta_{j+1} = \theta_j + \delta\theta \qquad (3.2a)$$

$$\delta\theta = H^{-1} u \qquad (3.2b)$$

$$u = \left.\frac{\partial l}{\partial \theta}\right|_{\theta=\theta_j}, \text{ and } H = \left.\frac{-\partial^2 l}{\partial\theta\partial\theta'}\right|_{\theta=\theta_j} \qquad (3.2c).$$

The derivatives u and H are calculated analytically. The iterations are assumed to have converged when the step-length $\partial\theta = (\delta\gamma, \delta\mu, \delta\sigma)'$ satisfies

$$|\delta\gamma| \le ACCU \cdot \sigma, \; |\delta\mu| \le ACCG, \; |\delta\sigma| \le ACCA \cdot \sigma,$$

for suitable constants ACCU, ACCG, ACCA.

This basic Newton-Raphson algorithm is modified as follows

(a) A maximum step-length is imposed : $|\delta\theta|$ is scaled so that it satisfies

$$|\delta\gamma| \le STEPU \cdot \sigma, \; |\delta\mu| \le STEPG, \; |\delta\sigma| \le STEPA \cdot \sigma, \qquad (3.3)$$

for suitable constants STEPU, STEPG, STEPA.

(b) If the Hessian matrix H is not positive definite, then the current step is made in the direction of steepest ascent, i.e. (3.2b) is replaced by $\delta\theta = \beta u$. The scalar β is chosen so that the step-length is the greatest length which satisfies (3.3).

(c) If a step leads to a value of θ which is infeasible, i.e. $1+\gamma\frac{x_i-\mu}{\sigma}\leq 0$ for some i, or if a step does not increase the log-likelihood, then the step-length is reduced by a factor ρ, i.e. step (3.2a) is replaced by $\theta_{j+1} = \theta_j + \rho\delta\theta$. This procedure is repeated until a value of θ is obtained which is feasible and which gives an increase in the log-likelihood, or until MAXSR step-length reductions have been made. In the latter case, if $\delta\theta$ was obtained by a Newton-Raphson step, then steepest ascent is tried instead. If , if $\delta\theta$ was obtained by a steepest-ascent step, then the routine is terminated.

(d) The routine is terminated if either the number of iterations or the number of evaluations of the log-likelihood exceeds the limit MAXIT or MAXEV respectively.

Hosking (1985) also remarks that the use of good initial parameters, such as probability-weighted moments estimators, can greatly increase the speed of the convergence.

So, under the known regularity conditions of ML theory (Cox and Hinkley, 1974), we are equipped with a very good set of estimators for inference on generalized extreme-value distribution. The key-point here is the satisfaction of regularity conditions. It is well known that if the support of a probability density depends on the unknown parameter, then the classical regularity conditions for MLE are not satisfied. Since, in our case, the support of the generalized extreme-value p.d.f. is dependent on the parameters, the regularity of the model is not immediately obvious. Therefore, although we have reliable numerical procedures for finding MLE, we are less certain about the properties of these estimators.

This issue was, somewhat, clarified by Smith (1985). He showed that the essentially regularity of MLE for the generalized extreme-value d.f. (i.e. the holding of the classical, good properties, asymptotic consistency, asymptotic efficiency and asymptotic normality) actually depends on the value of the unknown shape parameter γ. Particularly, he proved that MLE exists for $\gamma > -1$, but the classical asymptotic properties of maximum likelihood estimators hold only for $\gamma > -0.5$ (because only then the information matrix is finite). For $\gamma \leq -1$ the log-likelihood is J-shaped, i.e. it is globally maximized at sample maximum, which is a consistent estimator itself. For $-1 < \gamma < -0.5$ MLE has a non-normal limiting distribution with rate of convergence $O(n^{-\gamma})$. For $\gamma = -0.5$, ML estimators are asymptotically normal and efficient but the rate of convergence is $O\left\{(n \ln n)^{0.5}\right\}$ instead of the usual $O(n^{0.5})$. For more details and proofs of the above results see Smith (1985). DuMouchel (1983) explores more analytically the behaviour of ML estimator in case of α-stable distributions (i.e. cases where $0 < \gamma < 1/2$).

Concluding, we may say that, as in all classical statistical areas, MLE method is an attractive one, since it holds very desirable (and well studied) asymptotic properties. However, in the context of generalized extreme-value theory, this may not be the main attraction of the method. Indeed, up to now, we have been referring to the simple case where we have an i.i.d. sample from a generalized extreme-value distribution. Still, rarely, in practice, is an extreme value analysis as straightforward as simply fitting the generalized extreme-value model to a series of block maxima. At this point emerges another substantial advantage of MLE. That is, likelihood-based techniques, by small modifications, can handle features such as missing data, non-stationarity, temporal dependence and covariate effects, while this task is very tedious, if possible at all, for other estimation methods. Scarf (1992), for example, provides ML and PWM estimators for the generalized extreme-value distribution with location and scale parameters having a power law dependence on time (the latter estimators are described in the next section). He, too, stresses the fact that, though in some cases PWM appear to be superior, still they could not have been applied in more complex structures. Still, the main disadvantage of MLE for generalized extreme value distribution, remains the fact that its good asymptotic properties depend on the parameters value, in particular, where γ falls in the range (-∞, +∞). Of course many authors (such as Embrechts et al., 1997, and Coles and Dixon, 1999), defending ML method, state that the apparent restrictions are of little practical importance , since distributions with shape parameter $\gamma \leq -0.5$ are distributions with very short upper tails (actually, they have a finite upper end-point), a situation which is rarely encountered in applications for which extreme-value modelling is called for. Another disadvantage of MLE which will be made more clear in a subsequent section, is the fact that its small-sample properties are not as good as its asymptotic properties. Actually in small-sample cases MLEs can be outperformed by other methods, as we will see in the sequel. Another, known, general problem of MLE is its lack of robustness.

3.3 Method of Probability Weighted Moments

Another very popular estimation method, for the generalized extreme-value d.f., is the probability-weighted moments (PWM) estimator. Probability-weighted moments are generalization of the usual moments of a probability distribution, which give increasing weight to the tail information. There are several distributions, such as the Gumbel, logistic and Weibull d.f., whose parameters can be conveniently estimated from the probability-weighted moments. Notice that the Gumbel distribution is a special case of the generalized extreme-value distribution, thus implying that the PWM may be useful in the generalized extreme-value d.f., too. The estimation of the parameters of the generalized extreme-value distribution by the method of PWM is described in details by Hosking et al. (1985). They summarize some theoretical results for probability-weighted moments and show that they can be used to obtain estimates of parameters and quantiles of the generalized extreme-value d.f. Asymptotic as well as small-sample properties of these estimators are derived, and they are compared with other estimators.

Definition: ***Probability-Weighted Moments***

The probability-weighted moments of a random variable X with distribution function F are the quantities

$$M_{p,r,s} = E\left[X^p \{F(x)\}^r \{1-F(x)\}^s\right], \qquad p, r, s \in \Re .$$

Probability-weighted moments are likely to be most useful when the inverse d.f. $F^{\leftarrow}(.)$ can be written in closed form. Then we may write

$$M_{p,r,s} = \int_0^1 \left\{F^{\leftarrow}(u)\right\}^p u^r (1-u)^s \, du$$

which is often the most convenient way of evaluating these moments.

Specifically in the context of generalized extreme-value estimation, we use probability weighted moments of the form

$$\beta_r = M_{1,r,0} = E\left[X\left\{F(x)\right\}^r\right], \text{ r=0,1,2,...}$$

Given a random sample of size n from the distribution F, estimation of β_r is most conveniently based on the ordered sample $X_{1:n} \geq X_{2:n} \geq \ldots \geq X_{n:n}$. An unbiased estimator of β_r is provided by the statistic

$$b_r = \frac{1}{n}\sum_{j=1}^{n}\left[\frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} X_{n-j+1:n}\right].$$

Other asymptotically equivalent and consistent estimators are provided by the formulae

$$\hat{\beta}_r[p_{j,n}] = \frac{1}{n}\sum p_{j,n}^r X_{n-j+1:n}$$

where $p_{j,n}$ is a plotting position, such as $p_{j,n} = \frac{j-a}{n}$, with $0 < a < 1$, or $p_{j,n} = \frac{j-a}{n+1-2a}$ with $-0.5 < a < 0.5$.

The estimators b_r are closely related to U-statistics, which are widely used in non-parametric statistics. Their desirable properties of robustness to outliers in the sample, high efficiency, and asymptotic normality may be expected to extend to the PWMs estimators b_r and other quantities calculated from them.

Using straightforward calculations, one can derive the PWMs for the generalized extreme-value d.f. $H_{\gamma;\mu.\sigma}$ (for a proof of the formula see Hosking et al., 1985):

$$\beta_r = \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\gamma}\left[1-(r+1)^{\gamma}\Gamma(1-\gamma)\right]\right\}, \quad \text{for } \gamma < 1 \text{ and } \gamma \neq 0.$$

Note that the PWMs for generalized-extreme value distribution exist only for $\gamma < 1$. Hosking et al. (1985) do not consider that to be a problem, since, as they say, in most hydrological applications (which is the field they work on) the shape parameter usually lies in the interval (-1/2 , 1/2). As far as the case $\gamma = 0$ is concerned they develop a test for H_0: $\gamma = 0$, which turns out to perform very well.

For r=0,1,2 we have the following system of equations:

$$\beta_0 = \mu - \frac{\sigma}{\gamma}\left[1 - \Gamma(1-\gamma)\right],$$

$$2\beta_1 - \beta_0 = \frac{\sigma}{\gamma}\Gamma(1-\gamma)(2^{\gamma} - 1), \text{ and}$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1 - 3^{\gamma}}{1 - 2^{\gamma}}.$$

The PWM estimators $\hat{\theta} = (\hat{\gamma}, \hat{\mu}, \hat{\sigma})$ of the parameters are the solutions of the above equations when the moments β_r are replaced by their empirical counterparts (estimators b_r or $\hat{\beta}_r\left[p_{j,n}\right]$).

The exact solution requires iterative methods, but, as Hosking et al. (1985) propose, we can use an accurate low-order polynomial approximation for $\hat{\gamma}$. According to their suggestion, the (approximate) PWM estimators for the parameters of the generalized extreme-value distribution are

$$\hat{\gamma} = 7.8590c - 2.9554c^2, \text{ where } c = \frac{\ln 2}{\ln 3} - \frac{2b_1 - b_0}{3b_2 - b_0},$$

$$\hat{\sigma} = \frac{(2b_1 - b_0)\hat{\gamma}}{\Gamma(1-\hat{\gamma})(2^{\hat{\gamma}} - 1)}, \text{ and}$$

$$\hat{\mu} = b_0 - \frac{\hat{\sigma}}{\hat{\gamma}}\left\{\Gamma(1-\hat{\gamma}) - 1\right\}.$$

Asymptotic Properties

Using well known results from asymptotic theory, Hosking et al. (1985) derived some asymptotic properties (i.e. properties that hold only when the sample size is very large, tending to infinity) of the PWM estimators of the generalized extreme-value d.f. These are

- ***Asymptotic Normality*** : $\hat{\theta} \overset{n\to\infty}{\sim} MVN(M, C)$

(MVN stands for multivariate normal d.f.)

where the mean vector is $M = \theta$, and

the covariance matrix is of the form $C = \frac{1}{n}\begin{pmatrix} \sigma^2 w_{11} & \sigma^2 w_{12} & \sigma w_{13} \\ \sigma^2 w_{12} & \sigma^2 w_{22} & \sigma w_{23} \\ \sigma w_{13} & \sigma w_{23} & w_{33} \end{pmatrix}$.

The w_{ij} are functions of γ and have a complicated algebraic form. As γ approaches ½ the variance of the generalized extreme value d.f. becomes infinite and the variances of the parameter estimates are no longer of order 1/n asymptotically.

- The ***asymptotic biases*** of the parameter estimators are of order 1/n.
- The ***overall asymptotic efficiency*** (with respect to MLE's) of the PWM parameter estimators tends to zero at $\gamma = \pm 0.5$, but for values of γ not too far from zero the PWM is reasonably efficient.

3.4 Comparison of ML and PWM Estimation Methods for GEV Distribution

Hosking et al. (1985) performed a simulation study of the small-sample properties (in particular their study was concentrated on sample sizes n= 15, 25, 50, 100) of the PWM estimators of the generalized extreme-value distribution, in comparison to ML and sextiles estimators. Sextiles estimators had been originally also used for the estimation of the parameters of generalized extreme-value d.f., still they have been proven to be inferior of the other methods, so they are not used anymore. As far as estimators of the parameter γ are concerned, all three estimation methods are equivalent for n=100, but for smaller sample sizes the PWM estimator has lower variance. Still, PWM estimator has in general larger bias that the other estimators, though the bias is small near the critical value γ=0. Similar results are obtained for the estimators of the location and scale parameters, though the differences in the variances are less pronounced.

Another quantity of interest in extreme-value analysis are (large) quantiles. Estimators of quantiles can be simply obtaining by substituting the estimators of the parameters θ to the quantile function of the generalized extreme-value distribution. So, when comparing (via simulation) the quantile estimators obtained by the three different estimation methods, Hosking et al. (1985) get the following findings. For n=100, all methods are comparable. For small samples the upper quantiles obtained by PWM method are rather biased, but they are still preferable to the ML estimators, which have very large biases and variances. All the methods are very inaccurate when estimating extreme quantiles in small samples with $\gamma>0$.

Another comparison of ML and PWM estimation methods was performed by Coles and Dixon (1999). They also compiled a simulation study to explore the small-sample properties of ML and PWM. Their results confirmed the results of Hosking et al. (1985), that is, for small sample sizes, ML estimator is seen to be a poor competitor to the PWM estimator, in terms of both bias and mean square error. Still, while Hosking et al. (1985) were obviously in favour of the PWM estimators, Coles and Dixon (1999) share the view that as a general, all-round procedure for extreme value modelling, likelihood-based methods are preferable to any other. Still, despite the many advantages of maximum likelihood mentioned in the previous section, poor performance in small samples remains a serious criticism, since it is not uncommon in practice to need to make inferences about extremes with very few data – the rarity of extreme events means that even long observational periods may lead to very few data that can be incorporated into an extreme-value model. So, Coles and Dixon (1999) try to examine more closely the comparison between ML and PWM for estimating parameters of the generalized extreme-value d.f. with small datasets, by exploiting the simulated distributions of the parameter estimates. Understanding more clearly the failings of ML, they explore the possibility of correcting for such deficiencies. Some considerations of how to modify ML in order to improve its performance in small sample problems, whilst retaining its flexibility and asymptotic optimality, are presented in the following section.

3.5 Modified ML Estimators

The relatively poor estimation of the shape parameter γ by maximum likelihood can be attributed to the different assumptions made by the estimating equations. In particular, the PWM estimator assumes a priori that $\gamma < 1$, equivalent to specifying that the distribution has finite mean. Thus, the parameter space of (-∞ , ∞) is mapped to (-∞ , 1) in the estimated parameter space, leading to a reduction in the sampling variation of the estimate of γ. The price of this is a negative bias in the PWM estimate of γ which increases with increasing values of γ. However, the non-linearity in estimation formula of quantiles $\left(x_p = \mu - \frac{\sigma}{\gamma}\left[1 - \left(-\ln p\right)^{-\gamma}\right]\right)$ means that, judged in terms of mean square error of large quantiles, underestimation of γ is penalized much less heavily than overestimation. Thus the PWM estimator admits a certain amount of bias-variance trade-off in the estimate of γ, which after transformation to large quantiles avoids the heavy upper tail of the sampling distribution of the ML estimator of x_p.

So, the restriction $\gamma < 1$ can be viewed as prior information, which in a fair comparison of methods, should be available also to the likelihood-based analysis. Simply restricting the parameter space for γ to (-∞ , 1) leads to ML estimates on the boundary $\gamma = 1$ for samples which would have had ML estimate $\hat{\gamma} > 1$ in the unrestricted case. A more satisfactory adjustment would apply a bijective mapping of the parameter space for γ from (-∞ , ∞), into (-∞ , 1).

One possibility is to adopt a prior distribution for γ with support restricted to (-∞ , 1), and to apply a fully Bayesian analysis, adopting the mean or some other average of the corresponding posterior distribution as the estimate of γ. Some such efforts have indicated, that Bayesian techniques can give a more informative likelihood-based analysis of extreme-value data than ML. The extra computational burden of a fully Bayesian analysis is not restrictive for normal applications.

As an alternative, with the same aim of restricting the parameter space, a simpler estimate can be based on penalized likelihood. According to the general philosophy of penalized likelihood methods, we apply a penalty function to the likelihood which penalizes estimates of γ that are close to, or greater than, 1. Coles and Dixon (1999), after some experimentation, restrict their attention to penalty functions of the form

$$P(\gamma) = \begin{cases} 1 & \text{if } \gamma \le 0 \\ \exp\left[-\lambda \left(\dfrac{1}{1-\gamma} - 1 \right)^{\alpha} \right] & \text{if } 0 < \gamma < 1 \\ 0 & \text{if } \gamma \ge 1 \end{cases}$$

for different non-negative values of α and λ, with an associated penalized likelihood function

$$L_{pen}(\gamma;\mu,\sigma) = L(\gamma;\mu,\sigma) \times P(\gamma),$$

where L is the well-known likelihood function of the generalized extreme-value d.f. $H_{\gamma;\mu,\sigma}$. Large values of α in the penalty function correspond to a more severe relative penalty for values of γ which are large, but less than 1, while λ determines the overall weighting attached to the penalty.

Coles and Dixon (1999) adopted this approach, and after further experimentation they found that the combination α=λ=1 led to a reasonable performance across a range of values for γ and sample sizes (the results reported in the sequel refer to this combination of values). They, again, performed simulations in order to compare the small sample behaviour of these modified ML estimators with the classical ML and PWM estimators. The main findings of their simulation study are summarized in the following lines.

For negative values of γ, the penalized estimator is almost indistinguishable from the ML estimator. However, for positive values of γ, the penalized estimator has a behaviour much closer to that of the PWM estimator, thus inheriting the characteristics of smaller variance at the expense of negative bias. In terms of both bias and variance, the penalized estimator appears to be slightly better than, or at least as good as, the PWM estimator. As far as the quantile estimates are concerned, the penalized estimator is seen to inherit the usual large sample properties of ML estimator, without being affected by particularly poor performance with smaller samples. With respect to PWM estimator, the penalized estimator is almost uniformly better in terms of both bias and mean square error.

Concluding, we may say that likelihood-based estimation methods are advantageous enough, and under appropriate modifications can hold a prominent role in parametric estimation of extremes.

3.6 Other Estimation Techniques

Other estimation principles that could be used to estimate the parameters of generalized extreme-value distribution, though not so popular as ML and PWM estimators, are presented in the list below (taken from Reiss and Thomas, 1997).

- *Minimum Distance Estimators (MDE)*

Let d be a distance function on the family of d.f.'s. Then $\theta_n = (\gamma_n, \mu_n, \sigma_n)$ is a MDE if

$$d\left(\hat{F}_n, H_{\gamma_n;\mu_n,\sigma_n}\right) = \inf_{\gamma,\mu,\sigma} d\left(\hat{F}_n, H_{\gamma\ ;\mu,\sigma}\right),$$

where $\hat{F}_n$ is the empirical d.f.

Likewise, an MDE may be based on distances between densities. Dietrich and Hüsler (1996) deal with the MDE location and scale parameters for the Gumbel case (γ=0) based on the Cramer von Mises distance. They prove that these estimators are asymptotically consistent and normal, rather efficient (though less efficient than MLE) and robust with respect to contamination.

- *'Linear Combinations of Ratios of Spacings' Estimators (LRSE)*

This is a class of estimators that are based on a statistic of the form

$$\hat{r} = \frac{X_{[nq_0]:n} - X_{[nq_1]:n}}{X_{[nq_1]:n} - X_{[nq_2]:n}},$$

where $q_0 < q_1 < q_2$. This statistic is independent of the location and scale parameters in distribution (i.e. is invariant under affine transformations of the data) and several estimators can be derived depending on the choice of q_0, q_1, q_2. Reiss and Thomas (1997) suggest two particular choices:

- For q_0, q_1, q_2 such that $(-\ln q_1)^2 = (-\ln q_2)(-\ln q_0)$, $\hat{\gamma} = \dfrac{2\ln \hat{r}}{\ln[\ln q_0 / \ln q_1]}$.
- For q_0, q_1, q_2 such that, $q_i = q^{a^i}$, i=0,1,2, $q > 0$, $a < 1$, $\hat{\gamma} = \dfrac{\ln \hat{r}}{\ln[1/a]}$.

The Least Squares method may be then applied to obtain the additionally required estimators of the parameters μ, σ.

- *L-Moments Estimators*

L-Moments estimators (for simplicity, L-estimators) of the generalized extreme-value d.f. are based on the L-moments and they have been elaborated by Hosking (1990).
In general, L-moments are defined as

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}E\left(X_{k+1:r}\right),\quad \text{r=1, 2, ...,}$$

while the L-estimators of the parameters of the generalized extreme-value d.f. are given by the following formulae :

$\hat{\gamma} = 7.8590\text{z} - 2.9554\text{z}^2$, $\hat{\sigma} = \dfrac{l_2\hat{\gamma}}{\Gamma(1-\hat{\gamma})(2^{\hat{\gamma}}-1)}$, and $\hat{\mu} = l_1 - \dfrac{\hat{\sigma}}{\hat{\gamma}}\left\{\Gamma(1-\hat{\gamma})-1\right\}$,

where $l_1 = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ is the 1st sample L-moment

$l_2 = \dfrac{1}{2}\binom{n}{2}^{-1}\sum_{i>j}\sum(X_{n-i+1:n} - X_{n-j+1:n})$ is the 2nd sample L-moment,

$t_3 = \dfrac{l_3}{l_2} = \dfrac{1}{l_2}\dfrac{1}{3}\binom{n}{3}^{-1}\sum_{i>j>k}\sum\sum\left(X_{n-i+1:n} - 2X_{n-j+1:n} + X_{n-k+1:n}\right)$ is the 3rd sample L-moment ratio, and

$$z = \frac{\ln 2}{\ln 3} - \frac{2}{3+t_3}.$$

L-moments are analogous to the conventional moments but can be estimated by linear combinations of order statistics. This property provides them with several advantages over the conventional moments, i.e. they are less sensitive to sampling variability or measurement errors in the extreme data values, therefore giving more robust and accurate estimates. In addition, they are less subject to bias, and approximate normality more closely in finite samples. In comparison to MLE, they turn out to be more accurate in the small sample case. One could remark here, that the estimation formulae as well as the behaviour of L-estimators seems quite similar to the ones of PWM estimators. Actually, PWM can be expressed as linear combinations of L-moments, so procedures based on PWMs and on L-moments are essentially equivalent.

A detailed treatment of the generalized extreme-value d.f. as an exact distribution of a data-set can be found in the book of Reiss and Thomas (1997). Johnson *et al.* (1995) provide a long list and quite detailed description of many other estimation methods of location and scale parameters for the Gumbel case (i.e. for γ=0).