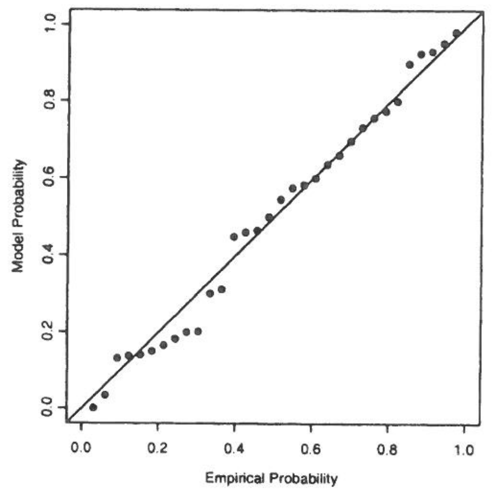

FIGURE 2.4. Probability plot for fitted model in the engine component failure time example.

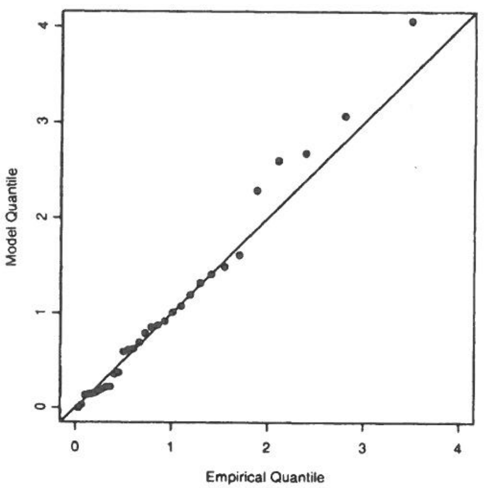

FIGURE 2.5. Quantile plot for fitted model in the engine component failure time example.

3 Classical Extreme Value Theory and Models

3.1 Asymptotic Models

3.1.1 *Model Formulation*

In this chapter we develop the model which represents the cornerstone of extreme value theory. The model focuses on the statistical behavior of

$$M_n = \max\{X_1, \ldots, X_n\},$$

where $X_1, \ldots, X_n$, is a sequence of independent random variables having a common distribution function F. In applications, the X_i usually represent values of a process measured on a regular time-scale – perhaps hourly measurements of sea-level, or daily mean temperatures – so that M_n represents the maximum of the process over n time units of observation. If n is the number of observations in a year, then M_n corresponds to the annual maximum.

In theory the distribution of M_n can be derived exactly for all values of n:

$$\begin{aligned} \Pr\{M_n \le z\} &= \Pr\{X_1 \le z, \ldots, X_n \le z\} \\ &= \Pr\{X_1 \le z\} \times \cdots \times \Pr\{X_n \le z\} \\ &= \{F(z)\}^n. \end{aligned} \tag{3.1}$$

However, this is not immediately helpful in practice, since the distribution function F is unknown. One possibility is to use standard statistical techniques to estimate F from observed data, and then to substitute this estimate into (3.1). Unfortunately, very small discrepancies in the estimate of F can lead to substantial discrepancies for F^n.

An alternative approach is to accept that F is unknown, and to look for approximate families of models for F^n, which can be estimated on the basis of the extreme data only. This is similar to the usual practice of approximating the distribution of sample means by the normal distribution, as justified by the central limit theorem. The arguments in this chapter are essentially an extreme value analog of the central limit theory.

We proceed by looking at the behavior of F^n as $n \to \infty$. But this alone is not enough: for any $z < z_+$, where z_+ is the upper end-point of F,[1] $F^n(z) \to 0$ as $n \to \infty$, so that the distribution of M_n degenerates to a point mass on z_+. This difficulty is avoided by allowing a linear renormalization of the variable M_n:

$$M_n^* = \frac{M_n - b_n}{a_n},$$

for sequences of constants $\{a_n > 0\}$ and $\{b_n\}$. Appropriate choices of the $\{a_n\}$ and $\{b_n\}$ stabilize the location and scale of M_n^* as n increases, avoiding the difficulties that arise with the variable M_n. We therefore seek limit distributions for M_n^*, with appropriate choices of $\{a_n\}$ and $\{b_n\}$, rather than M_n.

3.1.2 *Extremal Types Theorem*

The entire range of possible limit distributions for M_n^* is given by Theorem 3.1, the extremal types theorem.

Theorem 3.1 If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\{(M_n - b_n)/a_n \le z\} \to G(z) \quad \text{as } n \to \infty,$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\begin{aligned}
\mathrm{I}: G(z) &= \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty; \\
\mathrm{II}: G(z) &= \begin{cases} 0, & z \le b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b; \end{cases} \\
\mathrm{III}: G(z) &= \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b, \\ 1, & z \ge b, \end{cases}
\end{aligned}$$

for parameters $a > 0, b$ and, in the case of families II and III, $\alpha > 0$. □

[1] z_+ is the smallest value of z such that $F(z) = 1$.

In words, Theorem 3.1 states that the rescaled sample maxima $(M_n - b_n)/a_n$ converge in distribution to a variable having a distribution within one of the families labeled I, II and III. Collectively, these three classes of distribution are termed the **extreme value distributions**, with types I, II and III widely known as the **Gumbel**, **Fréchet** and **Weibull** families respectively. Each family has a location and scale parameter, b and a respectively; additionally, the Fréchet and Weibull families have a shape parameter α.

Theorem 3.1 implies that, when M_n can be stabilized with suitable sequences $\{a_n\}$ and $\{b_n\}$, the corresponding normalized variable M_n^* has a limiting distribution that must be one of the three types of extreme value distribution. The remarkable feature of this result is that the three types of extreme value distributions are the only possible limits for the distributions of the M_n^*, regardless of the distribution F for the population. It is in this sense that the theorem provides an extreme value analog of the central limit theorem.

3.1.3 *The Generalized Extreme Value Distribution*

The three types of limits that arise in Theorem 3.1 have distinct forms of behavior, corresponding to the different forms of tail behavior for the distribution function F of the X_i. This can be made precise by considering the behavior of the limit distribution G at z_+, its upper end-point. For the Weibull distribution z_+ is finite, while for both the Fréchet and Gumbel distributions $z_+ = \infty$. However, the density of G decays exponentially for the Gumbel distribution and polynomially for the Fréchet distribution, corresponding to relatively different rates of decay in the tail of F. It follows that in applications the three different families give quite different representations of extreme value behavior. In early applications of extreme value theory, it was usual to adopt one of the three families, and then to estimate the relevant parameters of that distribution. But there are two weaknesses: first, a technique is required to choose which of the three families is most appropriate for the data at hand; second, once such a decision is made, subsequent inferences presume this choice to be correct, and do not allow for the uncertainty such a selection involves, even though this uncertainty may be substantial.

A better analysis is offered by a reformulation of the models in Theorem 3.1. It is straightforward to check that the Gumbel, Fréchet and Weibull families can be combined into a single family of models having distribution functions of the form

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{3.2}$$

defined on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. This is the **generalized extreme value** (GEV) family of distributions. The model has three parameters: a location parameter, μ; a scale parameter, σ; and a shape parameter, ξ. The type II and type III classes of extreme value distribution correspond respectively to the cases $\xi > 0$ and $\xi < 0$ in this parameterization. The subset of the GEV family with $\xi = 0$ is interpreted as the limit of (3.2) as $\xi \to 0$, leading to the **Gumbel family** with distribution function

$$G(z) = \exp\left[-\exp\left\{-\left(\frac{z-\mu}{\sigma}\right)\right\}\right], \quad -\infty < z < \infty.$$

The unification of the original three families of extreme value distribution into a single family greatly simplifies statistical implementation. Through inference on ξ, the data themselves determine the most appropriate type of tail behavior, and there is no necessity to make subjective a priori judgements about which individual extreme value family to adopt. Moreover, uncertainty in the inferred value of ξ measures the lack of certainty as to which of the original three types is most appropriate for a given dataset.

For convenience we re-state Theorem 3.1 in modified form.

Theorem 3.1.1 If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\{(M_n - b_n)/a_n \le z\} \to G(z) \quad \text{as } n \to \infty \tag{3.3}$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

defined on $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. □

Interpreting the limit in Theorem 3.1.1 as an approximation for large values of n suggests the use of the GEV family for modeling the distribution of maxima of long sequences. The apparent difficulty that the normalizing constants will be unknown in practice is easily resolved. Assuming (3.3),

$$\Pr\{(M_n - b_n)/a_n \le z\} \approx G(z)$$

for large enough n. Equivalently,

$$\begin{aligned}\Pr\{M_n \le z\} &\approx G\{(z - b_n)/a_n\} \\ &= G^*(z),\end{aligned}$$

where G^* is another member of the GEV family. In other words, if Theorem 3.1.1 enables approximation of the distribution of M_n^* by a member of the GEV family for large n, the distribution of M_n itself can also be approximated by a different member of the same family. Since the parameters of the distribution have to be estimated anyway, it is irrelevant in practice that the parameters of the distribution G are different from those of G^*.

This argument leads to the following approach for modeling extremes of a series of independent observations $X_1, X_2, \ldots$. Data are blocked into sequences of observations of length n, for some large value of n, generating a series of block maxima, $M_{n,1}, \ldots, M_{n,m}$, say, to which the GEV distribution can be fitted. Often the blocks are chosen to correspond to a time period of length one year, in which case n is the number of observations in a year and the block maxima are annual maxima. Estimates of extreme quantiles of the annual maximum distribution are then obtained by inverting Eq. (3.2):

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \ne 0, \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{for } \xi = 0, \end{cases} \tag{3.4}$$

where $G(z_p) = 1 - p$. In common terminology, z_p is the **return level** associated with the **return period** $1/p$, since to a reasonable degree of accuracy, the level z_p is expected to be exceeded on average once every $1/p$ years. More precisely, z_p is exceeded by the annual maximum in any particular year with probability p.

Since quantiles enable probability models to be expressed on the scale of data, the relationship of the GEV model to its parameters is most easily interpreted in terms of the quantile expressions (3.4). In particular, defining $y_p = -\log(1-p)$, so that

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - y_p^{-\xi}\right], & \text{for } \xi \ne 0, \\ \mu - \sigma \log y_p, & \text{for } \xi = 0, \end{cases}$$

it follows that, if z_p is plotted against y_p on a logarithmic scale – or equivalently, if z_p is plotted against $\log y_p$ – the plot is linear in the case $\xi = 0$. If $\xi < 0$ the plot is convex with asymptotic limit as $p \to 0$ at $\mu - \sigma/\xi$; if $\xi > 0$ the plot is concave and has no finite bound. This graph is a **return level plot**. Because of the simplicity of interpretation, and because the choice of scale compresses the tail of the distribution so that the effect of extrapolation is highlighted, return level plots are particularly convenient for both model presentation and validation. Fig. 3.1 shows return level plots for a range of shape parameters.

3.1.4 Outline Proof of the Extremal Types Theorem

Formal justification of the extremal types theorem is technical, though not especially complicated – see Leadbetter et al. (1983), for example. In this section we give an informal proof. First, it is convenient to make the following definition.

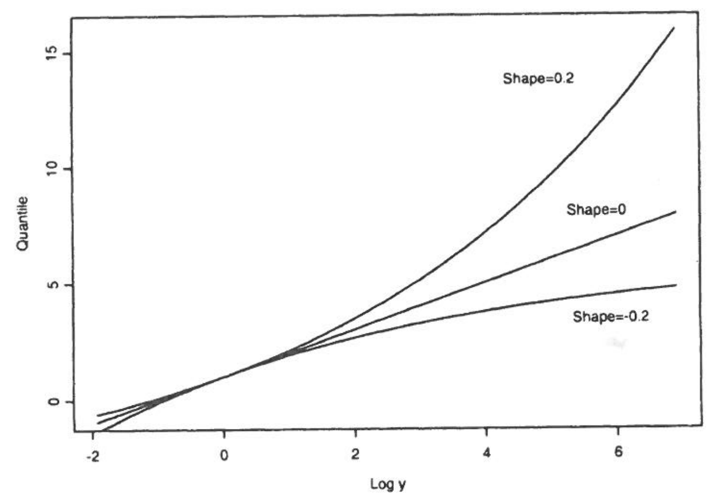

FIGURE 3.1. Return level plots of the GEV distribution with shape parameters $\xi = -0.2$, $\xi = 0$ and $\xi = 0.2$ respectively.

Definition 3.1 A distribution G is said to be **max-stable** if, for every $n = 2, 3, \ldots$, there are constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n z + \beta_n) = G(z).$$

△

Since G^n is the distribution function of $M_n = \max\{X_1, \ldots, X_n\}$, where the X_i are independent variables each having distribution function G, max-stability is a property satisfied by distributions for which the operation of taking sample maxima leads to an identical distribution, apart from a change of scale and location. The connection with the extreme value limit laws is made by the following result.

Theorem 3.2 A distribution is max-stable if, and only if, it is a generalized extreme value distribution. □

It requires only simple algebra to check that all members of the GEV family are indeed max-stable. The converse requires ideas from functional analysis that are beyond the scope of this book.

Theorem 3.2 is used directly in the proof of the extremal types theorem. The idea is to consider M_{nk}, the maximum random variable in a sequence of $n \times k$ variables for some large value of n. This can be regarded as the maximum of a single sequence of length $n \times k$, or as the maximum of k

maxima, each of which is the maximum of n observations. More precisely, suppose the limit distribution of $(M_n - b_n)/a_n$ is G. So, for large enough n,

$$\Pr\{(M_n - b_n)/a_n \le z\} \approx G(z)$$

by Theorem 3.1.1. Hence, for any integer k, since nk is large,

$$\Pr\{(M_{nk} - b_{nk})/a_{nk} \le z\} \approx G(z). \tag{3.5}$$

But, since M_{nk} is the maximum of k variables having the same distribution as M_n,

$$\Pr\{(M_{nk} - b_n)/a_n \le z\} = \left[\Pr\{(M_n - b_n)/a_n \le z\}\right]^k. \tag{3.6}$$

Hence, by (3.5) and (3.6) respectively,

$$\Pr\{M_{nk} \le z\} \approx G\left(\frac{z - b_{nk}}{a_{nk}}\right)$$

and

$$\Pr\{M_{nk} \le z\} \approx G^k\left(\frac{z - b_n}{a_n}\right).$$

Therefore, G and G^k are identical apart from location and scale coefficients. It follows that G is max-stable and therefore a member of the GEV family by Theorem 3.2.

3.1.5 *Examples*

One issue we have not discussed in connection with Theorem 3.1 is, given a distribution function F, how to establish whether convergence of the distribution of the normalized M_n can actually be achieved. If it can, there are additional questions: what choices of normalizing sequences $\{a_n\}$ and $\{b_n\}$ are necessary and which member of the GEV family is obtained as a limit? Each of the books on extreme value theory referenced in Section 1.4 gives extensive details on these aspects. Since our primary consideration is the statistical inference of real data for which the underlying distribution F is unknown, we will give only a few examples that illustrate how careful choice of normalizing sequences does lead to a limit distribution within the GEV family, as implied by Theorem 3.1. These examples will also be useful for illustrating other limit results in subsequent chapters.

Example 3.1 If $X_1, X_2, \ldots$ is a sequence of independent standard exponential Exp(1) variables, $F(x) = 1 - e^{-x}$ for $x > 0$. In this case, letting

$a_n = 1$ and $b_n = n$,

$$\begin{aligned}\Pr\{(M_n - b_n)/a_n \le z\} &= F^n(z + \log n)\\ &= \left[1 - e^{-(z+\log n)}\right]^n\\ &= \left[1 - n^{-1}e^{-z}\right]^n\\ &\to \exp(-e^{-z})\end{aligned}$$

as $n \to \infty$, for each fixed $z \in \mathbb{R}$. Hence, with the chosen a_n and b_n, the limit distribution of M_n as $n \to \infty$ is the Gumbel distribution, corresponding to $\xi = 0$ in the GEV family. ▲

Example 3.2 If $X_1, X_2, \dots$ is a sequence of independent standard Fréchet variables, $F(x) = \exp(-1/x)$ for $x > 0$. Letting $a_n = n$ and $b_n = 0$,

$$\begin{aligned}\Pr\{(M_n - b_n)/a_n \le z\} &= F^n(nz)\\ &= [\exp\{-1/(nz)\}]^n\\ &= \exp(-1/z)\end{aligned}$$

as $n \to \infty$, for each fixed $z > 0$. Hence, the limit in this case – which is an exact result for all n, because of the max-stability of F – is also the standard Fréchet distribution: $\xi = 1$ in the GEV family. ▲

Example 3.3 If $X_1, X_2, \dots$ are a sequence of independent uniform U(0, 1) variables, $F(x) = x$ for $0 \le x \le 1$. For fixed $z < 0$, suppose $n > -z$ and let $a_n = 1/n$ and $b_n = 1$. Then,

$$\begin{aligned}\Pr\{(M_n - b_n)/a_n \le z\} &= F^n(n^{-1}z + 1)\\ &= \left(1 + \frac{z}{n}\right)^n\\ &\to e^z\end{aligned}$$

as $n \to \infty$. Hence, the limit distribution is of Weibull type, with $\xi = -1$ in the GEV family. ▲

There is some latitude in the choice of $\{a_n\}$ and $\{b_n\}$ in such examples. However, different choices that lead to a non-degenerate limit always yield a limit distribution in the GEV family with the same value of ξ, though possibly with other values of the location and scale parameters.

3.2 Asymptotic Models for Minima

Some applications require models for extremely small, rather than extremely large, observations. This is not usually the case for problems involving environmental data, but system failure models, as discussed in Example 1.2, can often be constructed such that the lifetime of a system is equal to the minimum lifetime of a number, n, of individual components. The overall system lifetime is then $\tilde{M}_n = \min\{X_1, \dots, X_n\}$, where the X_i denote the individual component lifetimes. Assuming the X_i to be independent and identically distributed, analogous arguments apply to $\tilde{M}_n$ as were applied to M_n, leading to a limiting distribution of a suitably re-scaled variable.

The results are also immediate from the corresponding results for M_n. Letting $Y_i = -X_i$ for $i = 1, \dots, n$, the change of sign means that small values of X_i correspond to large values of Y_i. So if $\tilde{M}_n = \min\{X_1, \dots, X_n\}$ and $M_n = \max\{Y_1, \dots, Y_n\}$, then $\tilde{M}_n = -M_n$. Hence, for large n,

$$\begin{aligned}\Pr\{\tilde{M}_n \le z\} &= \Pr\{-M_n \le z\}\\ &= \Pr\{M_n \ge -z\}\\ &= 1 - \Pr\{M_n \le -z\}\\ &\approx 1 - \exp\left\{-\left[1 + \xi\left(\frac{-z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}\\ &= 1 - \exp\left\{-\left[1 - \xi\left(\frac{z-\tilde{\mu}}{\sigma}\right)\right]^{-1/\xi}\right\},\end{aligned}$$

on $\{z : 1 - \xi(z - \tilde{\mu})/\sigma > 0\}$, where $\tilde{\mu} = -\mu$. This distribution is the **GEV distribution for minima**. We can state the result formally as a theorem analogous to Theorem 3.1.1 for maxima.

Theorem 3.3 If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\{(\tilde{M}_n - b_n)/a_n \le z\} \to \tilde{G}(z) \quad \text{as } n \to \infty$$

for a non-degenerate distribution function $\tilde{G}$, then $\tilde{G}$ is a member of the GEV family of distributions for minima:

$$\tilde{G}(z) = 1 - \exp\left\{-\left[1 - \xi\left(\frac{z-\tilde{\mu}}{\sigma}\right)\right]^{-1/\xi}\right\},$$

defined on $\{z : 1 - \xi(z - \tilde{\mu})/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\tilde{\sigma} > 0$ and $-\infty < \xi < \infty$. □

In situations where it is appropriate to model block minima, the GEV distribution for minima can be applied directly. An alternative is to exploit the duality between the distributions for maxima and minima. Given data $z_1, \dots, z_m$ that are realizations from the GEV distribution for minima, with parameters $(\tilde{\mu}, \sigma, \xi)$, this implies fitting the GEV distribution for maxima to the data $-z_1, \dots, -z_m$. The maximum likelihood estimate of the parameters of this distribution corresponds exactly to that of the required GEV distribution for minima, apart from the sign correction $\hat{\tilde{\mu}} = -\hat{\mu}$. This approach is used in Section 3.4.2 to model the glass fiber data described in Example 1.2.

3.3 Inference for the GEV Distribution

3.3.1 General Considerations

Motivated by Theorem 3.1.1, the GEV provides a model for the distribution of block maxima. Its application consists of blocking the data into blocks of equal length, and fitting the GEV to the set of block maxima. But in implementing this model for any particular dataset, the choice of block size can be critical. The choice amounts to a trade-off between bias and variance: blocks that are too small mean that approximation by the limit model in Theorem 3.1.1 is likely to be poor, leading to bias in estimation and extrapolation; large blocks generate few block maxima, leading to large estimation variance. Pragmatic considerations often lead to the adoption of blocks of length one year. For example, only the annual maximum data may have been recorded, so that the use of shorter blocks is not an option. Even when this is not the case, an analysis of annual maximum data is likely to be more robust than an analysis based on shorter blocks if the conditions of Theorem 3.1.1 are violated. For example, daily temperatures are likely to vary according to season, violating the assumption that the X_i have a common distribution. If the data were blocked into block lengths of around 3 months, the maximum of the summer block is likely to be much greater than that of the winter block, and an inference that failed to take this non-homogeneity into account would be likely to give inaccurate results. Taking, instead, blocks of length one year means the assumption that individual block maxima have a common distribution is plausible, though the formal justification for the GEV approximation remains invalid.

We now simplify notation by denoting the block maxima $Z_1, \ldots, Z_m$. These are assumed to be independent variables from a GEV distribution whose parameters are to be estimated. If the X_i are independent then the Z_i are also independent. However, independence of the Z_i is likely to be a reasonable approximation even if the X_i constitute a dependent series. In this case, although not covered by Theorem 3.1.1, the conclusion that the Z_i have a GEV distribution may still be reasonable; see Chapter 5.

Many techniques have been proposed for parameter estimation in extreme value models. These include graphical techniques based on versions of probability plots; moment-based techniques in which functions of model moments are equated with their empirical equivalents; procedures in which the parameters are estimated as specified functions of order statistics; and likelihood-based methods. Each technique has its pros and cons, but the all-round utility and adaptability to complex model-building of likelihood-based techniques make this approach particularly attractive.

A potential difficulty with the use of likelihood methods for the GEV concerns the regularity conditions that are required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid. Such conditions are not satisfied by the GEV model because the end-points

of the GEV distribution are functions of the parameter values: $\mu - \sigma/\xi$ is an upper end-point of the distribution when $\xi < 0$, and a lower end-point when $\xi > 0$. This violation of the usual regularity conditions means that the standard asymptotic likelihood results are not automatically applicable. Smith (1985) studied this problem in detail and obtained the following results:

- when $\xi > -0.5$, maximum likelihood estimators are regular, in the sense of having the usual asymptotic properties;
- when $-1 < \xi < -0.5$, maximum likelihood estimators are generally obtainable, but do not have the standard asymptotic properties;
- when $\xi < -1$, maximum likelihood estimators are unlikely to be obtainable.

The case $\xi \leq -0.5$ corresponds to distributions with a very short bounded upper tail. This situation is rarely encountered in applications of extreme value modeling, so the theoretical limitations of the maximum likelihood approach are usually no obstacle in practice.

3.3.2 Maximum Likelihood Estimation

Under the assumption that $Z_1, \ldots, Z_m$ are independent variables having the GEV distribution, the log-likelihood for the GEV parameters when $\xi \neq 0$ is

$$\begin{aligned}\ell(\mu, \sigma, \xi) &= -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^{m} \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right] \\ &\quad - \sum_{i=1}^{m} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi}, \qquad (3.7)\end{aligned}$$

provided that

$$1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) > 0, \text{ for } i = 1, \ldots, m. \qquad (3.8)$$

At parameter combinations for which (3.8) is violated, corresponding to a configuration for which at least one of the observed data falls beyond an end-point of the distribution, the likelihood is zero and the log-likelihood equals $-\infty$.

The case $\xi = 0$ requires separate treatment using the Gumbel limit of the GEV distribution. This leads to the log-likelihood

$$\ell(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^{m} \left(\frac{z_i - \mu}{\sigma} \right) - \sum_{i=1}^{m} \exp \left\{ - \left(\frac{z_i - \mu}{\sigma} \right) \right\}. \qquad (3.9)$$

Maximization of the pair of Eqs. (3.7) and (3.9) with respect to the parameter vector (μ, σ, ξ) leads to the maximum likelihood estimate with respect to the entire GEV family. There is no analytical solution, but for any given dataset the maximization is straightforward using standard numerical optimization algorithms. Some care is needed to ensure that such algorithms do not move to parameter combinations violating (3.8), and also that numerical difficulties that would arise from the evaluation of (3.7) in the vicinity of $\xi = 0$ are avoided. This latter problem is easily solved by using (3.9) in place of (3.7) for values of ξ falling within a small window around zero.

Subject to the limitations on ξ discussed in Section 3.3.1, the approximate distribution of $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ is multivariate normal with mean (μ, σ, ξ) and variance-covariance matrix equal to the inverse of the observed information matrix evaluated at the maximum likelihood estimate. Though this matrix can be calculated analytically, it is easier to use numerical differencing techniques to evaluate the second derivatives, and standard numerical routines to carry out the inversion. Confidence intervals and other forms of inference follow immediately from the approximate normality of the estimator.

3.3.3 *Inference for Return Levels*

By substitution of the maximum likelihood estimates of the GEV parameters into (3.4), the maximum likelihood estimate of z_p for $0 < p < 1$, the $1/p$ return level, is obtained as

$$\hat{z}_p = \begin{cases} \mu - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - y_p^{-\hat{\xi}}\right], & \text{for } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{for } \hat{\xi} = 0, \end{cases} \tag{3.10}$$

where $y_p = -\log(1-p)$. Furthermore, by the delta method,

$$\mathrm{Var}(\hat{z}_p) \approx \nabla z_p^T V \nabla z_p, \tag{3.11}$$

where V is the variance-covariance matrix of $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and

$$\begin{aligned} \nabla z_p^T &= \left[\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \xi}\right] \\ &= \left[1, \ -\xi^{-1}(1 - y_p^{-\xi}), \ \sigma\xi^{-2}(1 - y_p^{-\xi}) - \sigma\xi^{-1} y_p^{-\xi} \log y_p\right] \end{aligned}$$

evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.

It is usually long return periods, corresponding to small values of p, that are of greatest interest. If $\hat{\xi} < 0$ it is also possible to make inferences on the upper end-point of the distribution, which is effectively the 'infinite-observation return period', corresponding to z_p with $p = 0$. The maximum likelihood estimate is

$$\hat{z}_0 = \hat{\mu} - \hat{\sigma}/\hat{\xi},$$

and (3.11) is still valid with

$$\nabla z_0^T = \left[1, \ -\xi^{-1}, \ \sigma\xi^{-2}\right],$$

again evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$. When $\hat{\xi} \geq 0$ the maximum likelihood estimate of the upper end-point is infinity.

Caution is required in the interpretation of return level inferences, especially for return levels corresponding to long return periods. First, the normal approximation to the distribution of the maximum likelihood estimator may be poor. Better approximations are generally obtained from the appropriate profile likelihood function; see Section 2.6.6. More fundamentally, estimates and their measures of precision are based on an assumption that the model is correct. Though the GEV model is supported by mathematical argument, its use in extrapolation is based on unverifiable assumptions, and measures of uncertainty on return levels should properly be regarded as lower bounds that could be much greater if uncertainty due to model correctness were taken into account.

3.3.4 *Profile Likelihood*

Numerical evaluation of the profile likelihood for any of the individual parameters μ, σ or ξ is straightforward. For example, to obtain the profile likelihood for ξ, we fix $\xi = \xi_0$, and maximize the log-likelihood (3.7) with respect to the remaining parameters, μ and σ. This is repeated for a range of values of ξ_0. The corresponding maximized values of the log-likelihood constitute the profile log-likelihood for ξ, from which Theorem 2.6 leads to obtain approximate confidence intervals.

This methodology can also be applied when inference is required on some combination of parameters. In particular, we can obtain confidence intervals for any specified return level z_p. This requires a reparameterization of the GEV model, so that z_p is one of the model parameters, after which the profile log-likelihood is obtained by maximization with respect to the remaining parameters in the usual way. Reparameterization is straightforward:

$$\mu = z_p + \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], \tag{3.12}$$

so that replacement of μ in (3.7) with (3.12) has the desired effect of expressing the GEV model in terms of the parameters (z_p, σ, ξ).

3.3.5 *Model Checking*

Though it is impossible to check the validity of an extrapolation based on a GEV model, assessment can be made with reference to the observed data. This is not sufficient to justify extrapolation, but is a reasonable prerequisite. In Chapter 2 we discussed the use of probability plots and quantile plots for model checking; we now describe these in more detail for checking the validity of a GEV model, and describe two further graphical goodness-of-fit checks.

As described in Chapter 2, a probability plot is a comparison of the empirical and fitted distribution functions. With ordered block maximum data $z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(m)}$, the empirical distribution function evaluated at $z_{(i)}$ is given by

$$\tilde{G}(z_{(i)}) = i/(m+1).$$

By substitution of parameter estimates into (3.2), the corresponding model-based estimates are

$$\hat{G}(z_{(i)}) = \exp\left\{-\left[1+\hat{\xi}\left(\frac{z_{(i)}-\hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\}.$$

If the GEV model is working well,

$$\hat{G}(z_{(i)}) \approx \tilde{G}(z_{(i)})$$

for each i, so a probability plot, consisting of the points

$$\left\{\left(\tilde{G}(z_{(i)}), \hat{G}(z_{(i)})\right),\ i = 1,\ldots,m\right\},$$

should lie close to the unit diagonal. Any substantial departures from linearity are indicative of some failing in the GEV model.

A weakness of the probability plot for extreme value models is that both $\hat{G}(z_{(i)})$ and $\tilde{G}(z_{(i)})$ are bound to approach 1 as $z_{(i)}$ increases, while it is usually the accuracy of the model for large values of z that is of greatest concern. That is, the probability plot provides the least information in the region of most interest. This deficiency is avoided by the quantile plot, consisting of the points

$$\left\{\left(\hat{G}^{-1}(i/(m+1)), z_{(i)}\right),\ i = 1,\ldots,m\right\}, \tag{3.13}$$

where, from (3.10),

$$\hat{G}^{-1}\left(\frac{i}{m+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \left\{-\log\left(\frac{i}{m+1}\right)\right\}^{-\hat{\xi}}\right].$$

Departures from linearity in the quantile plot also indicate model failure.

As discussed in Section 3.1.3, the return level plot, comprising a graph of

$$z_p = \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right]$$

against $y_p = -\log(1-p)$ on a logarithmic scale, is particularly convenient for interpreting extreme value models. The tail of the distribution is compressed, so that return level estimates for long return periods are displayed,

while the linearity of the plot in the case $\xi = 0$ provides a baseline against which to judge the effect of the estimated shape parameter.

As a summary of a fitted model the return level plot consists of the locus of points

$$\{(\log y_p, \hat{z}_p) : 0 < p < 1\},$$

where $\hat{z}_p$ is the maximum likelihood estimate of z_p. Confidence intervals can be added to the plot to increase its informativeness. Empirical estimates of the return level function, obtained from the points (3.13), can also be added, enabling the return level plot to be used as a model diagnostic. If the GEV model is suitable for the data, the model-based curve and empirical estimates should be in reasonable agreement. Any substantial or systematic disagreement, after allowance for sampling error, suggests an inadequacy of the GEV model.

The probability, quantile and return level plots are each based on a comparison of model-based and empirical estimates of the distribution function. For completeness, an equivalent diagnostic based on the density function is a comparison of the probability density function of a fitted model with a histogram of the data. This is generally less informative than the previous plots, since the form of a histogram can vary substantially with the choice of grouping intervals. That is, in contrast with the empirical distribution function, there is no unique empirical estimator of a density function, making comparisons with a model-based estimator difficult and subjective.

3.4 Examples

3.4.1 Annual Maximum Sea-levels at Port Pirie

This analysis is based on the series of annual maximum sea-levels recorded at Port Pirie, South Australia, over the period 1923–1987, as described in Example 1.1. From Fig. 1.1 it seems reasonable to assume that the pattern of variation has stayed constant over the observation period, so we model the data as independent observations from the GEV distribution.

Maximization of the GEV log-likelihood for these data leads to the estimate

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (3.87, 0.198, -0.050),$$

for which the log-likelihood is 4.34. The approximate variance-covariance matrix of the parameter estimates is

$$V = \begin{bmatrix} 0.000780 & 0.000197 & -0.00107 \\ 0.000197 & 0.000410 & -0.000778 \\ -0.00107 & -0.000778 & 0.00965 \end{bmatrix}.$$

The diagonals of the variance-covariance matrix correspond to the variances of the individual parameters of (μ, σ, ξ) . Taking square roots, the standard

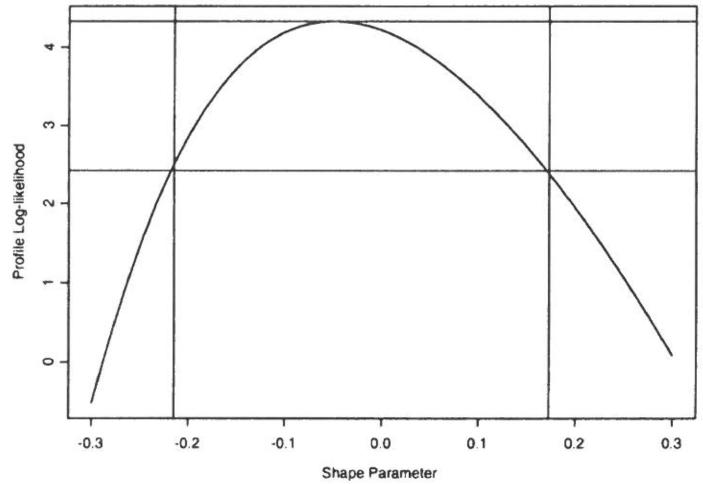

FIGURE 3.2. Profile likelihood for ξ in the Port Pirie sea-level example.

errors are 0.028, 0.020 and 0.098 for $\hat{\mu}, \hat{\sigma}$ and $\hat{\xi}$ respectively. Combining estimates and standard errors, approximate 95% confidence intervals for each parameter are $[3.82, 3.93]$ for μ, $[0.158, 0.238]$ for σ, and $[-0.242, 0.142]$ for ξ. In particular, although the maximum likelihood estimate for ξ is negative, corresponding to a bounded distribution, the 95% confidence interval extends well above zero, so that the strength of evidence from the data for a bounded distribution is not strong. Greater accuracy of confidence intervals can usually be achieved by the use of profile likelihood. Fig. 3.2 shows the profile log-likelihood for ξ, from which a 95% confidence interval for ξ is obtained as $[-0.21, 0.17]$, which is only slightly different to the earlier calculation.

Estimates and confidence intervals for return levels are obtained by substitution into (3.10) and (3.11). For example, to estimate the 10-year return level, we set $p = 1/10$ and find $\hat{z}_{0.1} = 4.30$ and $\mathrm{Var}(\hat{z}_{0.1}) = 0.00303$. Hence, a 95% confidence interval for $z_{0.1}$ is evaluated as $4.30 \pm 1.96 \times \sqrt{0.00303} = [4.19, 4.41]$. The corresponding estimate for the 100-year return level is $\hat{z}_{0.01} = 4.69$, with a 95% confidence interval of $[4.38, 5.00]$.

Better accuracy again comes from the profile likelihood. Figs. 3.3 and 3.4 show the profile log-likelihood for the 10- and 100-year return levels respectively. By Theorem 2.6 we obtain confidence intervals for $z_{0.1}$ and $z_{0.01}$ of $[4.21, 4.45]$ and $[4.50, 5.27]$ respectively. The first of these is similar to that obtained from the delta method, while the second is not. The latter discrepancy arises because of asymmetry in the profile log-likelihood surface, the extent of which increases with increasing return period. Such

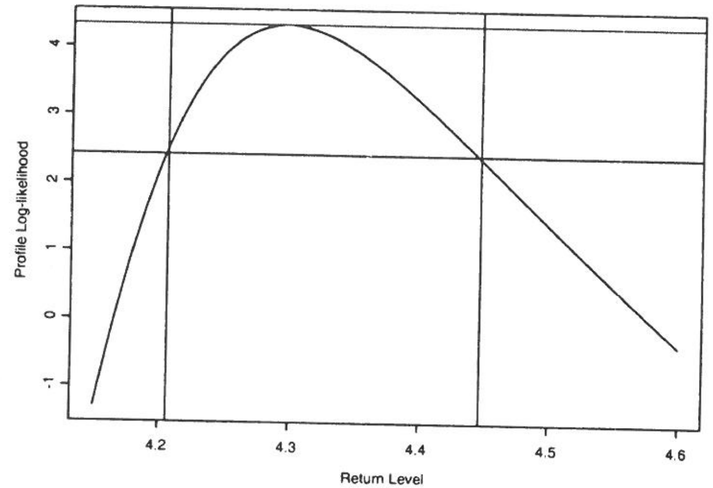

FIGURE 3.3. Profile likelihood for 10-year return level in the Port Pirie sea-level example.

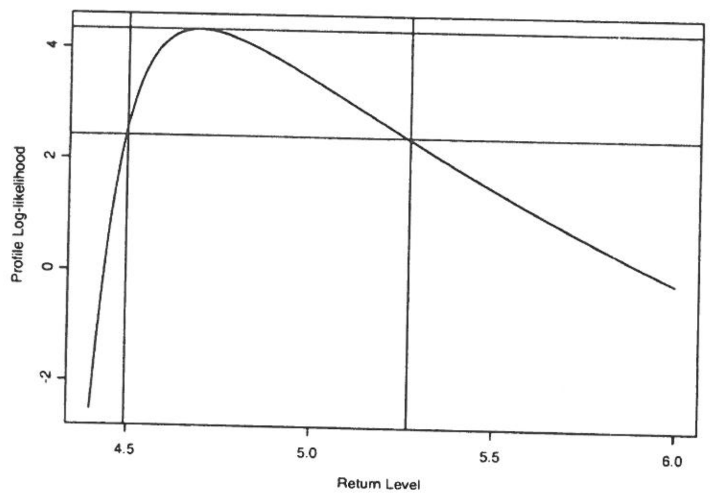

FIGURE 3.4. Profile likelihood for 100-year return level in the Port Pirie sea-level example.

asymmetries are to be expected, since the data provide increasingly weaker information about high levels of the process.

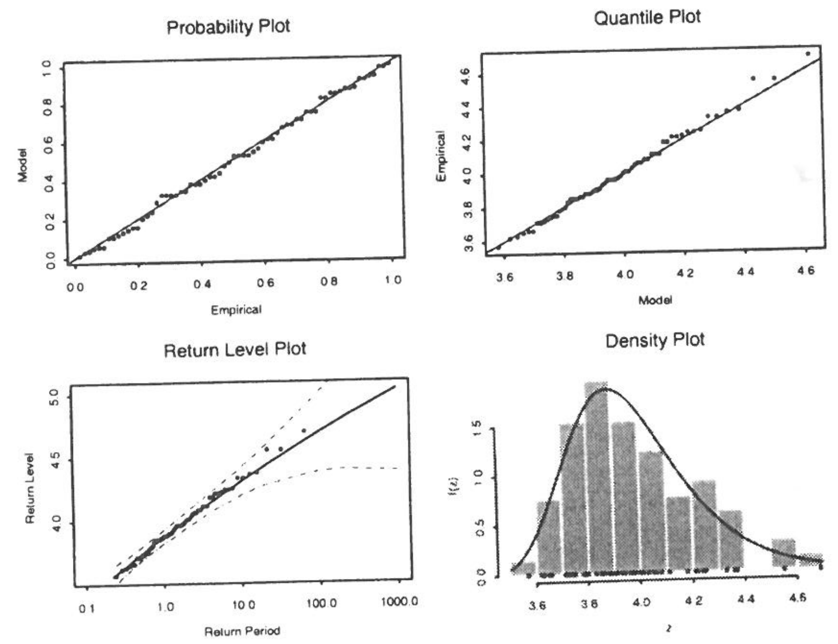

FIGURE 3.5. Diagnostic plots for GEV fit to the Port Pirie sea-level data.

The various diagnostic plots for assessing the accuracy of the GEV model fitted to the Port Pirie data are shown in Fig. 3.5. Neither the probability plot nor the quantile plot give cause to doubt the validity of the fitted model: each set of plotted points is near-linear. The return level curve asymptotes to a finite level as a consequence of the negative estimate of ξ, though since the estimate is close to zero, the estimated curve is close to linear. The curve also provides a satisfactory representation of the empirical estimates, especially once sampling variability is taken into account. Finally, the corresponding density estimate seems consistent with the histogram of the data. Consequently, all four diagnostic plots lend support to the fitted GEV model.

The original version of the extremal types theorem, as given in Theorem 3.1, identifies three possible families of limit distributions for maxima. Before the unification of the three distributions into the single GEV family, it was natural to make a preliminary choice of model type prior to parameter estimation. This approach now has little merit, given the alternative option of modeling within the entire GEV family. However, the suitability of any particular member of the GEV family can be assessed by comparison with

the maximum likelihood estimate within the entire family. For example, the appropriateness of replacing the GEV family with the Gumbel family, corresponding to the $\xi = 0$ subset of the GEV family, can be assessed.

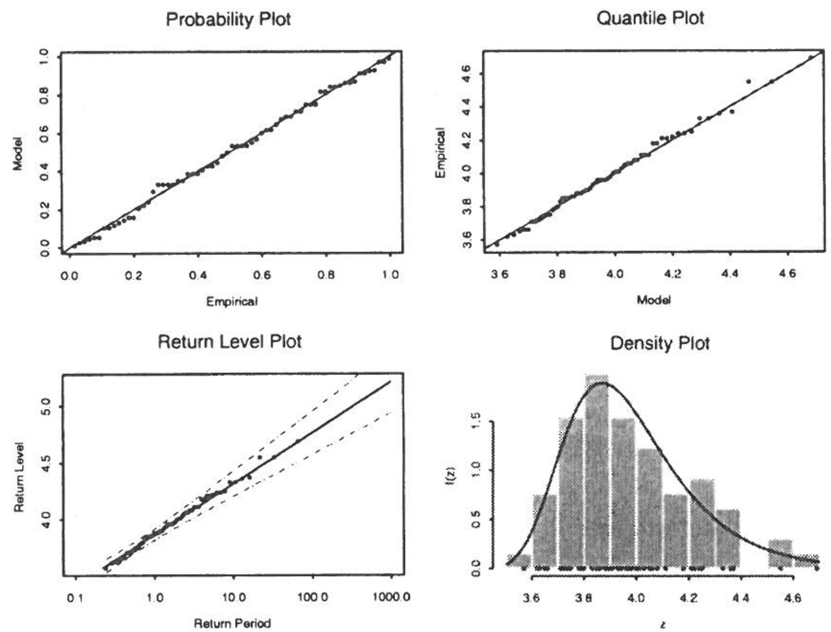

FIGURE 3.6. Diagnostic plots for Gumbel fit to the Port Pirie sea-level data.

Maximum likelihood in the Gumbel case corresponds to maximization of (3.9), followed by standard treatment to obtain standard errors etc. For the Port Pirie sea-level data this leads to $(\hat{\mu}, \hat{\sigma}) = (3.87, 0.195)$, with standard errors 0.03 and 0.019 respectively. The maximized log-likelihood is -4.22. The likelihood ratio test statistic for the reduction to the Gumbel model is therefore $D = 2\{-4.22 - (-4.34)\} = 0.24$. This value is small when compared to the χ_1^2 distribution, suggesting that the Gumbel model is adequate for these data. This is confirmed by the standard diagnostic graphical checks in Fig. 3.6, which imply that the goodness-of-fit is comparable with that of the GEV model. This is not surprising since the estimated parameters in the two models are so similar, which also means that (short-term) model extrapolation on the basis of either model leads to similar answers. The greatest difference between the two models is in terms of the precision of estimation: the model parameters and return levels have estimates with considerably shorter confidence intervals in the Gumbel model, compared with the GEV model.

The issue of choice between the Gumbel and GEV models is starkly illustrated by the respective return level plots of Figs. 3.5 and 3.6. The estimated return level curves are similar, but the confidence intervals are much wider for the GEV model, especially for long return periods. Reduction of uncertainty is desirable, so that if the Gumbel model could be trusted, its inferences would be preferred. But can the model be trusted? The extremal types theorem provides support for modeling block maxima with the GEV family, of which the Gumbel family is a subset. The data suggest that a Gumbel model is plausible, but this does not imply that other models are not. Indeed, the maximum likelihood estimate within the GEV family is not in the Gumbel family (although, in the sense that $\hat{\xi} \approx 0$, it is 'close'). There is no common agreement about this issue, but the safest option is to accept there is uncertainty about the value of the shape parameter – and hence whether the Gumbel model is correct or not – and to prefer the inference based on the GEV model. The larger measures of uncertainty generated by the GEV model then provide more a realistic quantification of the genuine uncertainties involved in model extrapolation.

3.4.2 *Glass Fiber Strength Example*

We now consider the glass fiber breaking strength data of Example 1.2. For the reasons discussed in Section 3.2, the GEV model for minima is an appropriate starting point for data of this type. There are two equivalent approaches to the modeling. Either the GEV distribution for minima can be fitted directly to the data, or the data can be negated and the GEV distribution for maxima fitted to these data. The equivalence of these operations is justified in Section 3.2. To economize on the writing of model-fitting routines, we take the approach of fitting the GEV distribution to the negated data. This leads to the maximum likelihood estimate

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (-1.64, 0.27, -0.084),$$

with a maximized value of the log-likelihood equal to -14.3. The corresponding estimated variance-covariance matrix is

$$V = \begin{bmatrix} 0.00141 & 0.000214 & -0.000795 \\ 0.000214 & 0.000652 & -0.000441 \\ -0.000795 & -0.0000441 & 0.00489 \end{bmatrix}.$$

Taking square roots of the diagonals of this matrix leads to standard errors of μ, σ and ξ as 0.038, 0.026 and 0.070 respectively. The estimates and standard errors combine to give approximate confidence intervals. In particular, a 95% confidence interval for ξ is obtained as $-0.084 \pm 1.96 \times 0.07 = [-0.22, 0.053]$. So, as in the previous example, the maximum likelihood estimate of the shape parameter is negative, but both negative and positive values are plausible once sampling uncertainty is accounted for.

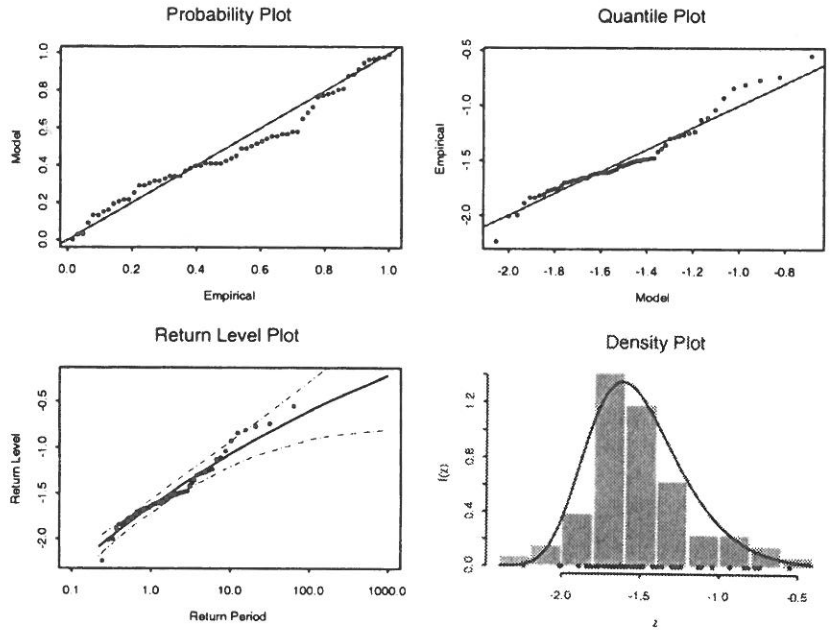

FIGURE 3.7. Diagnostic plots of GEV fit to negative breaking strengths of glass fibers.

From Section 3.2 the estimates for the parameters $(\tilde{\mu}, \sigma, \xi)$ of the corresponding GEV distribution for minima applied directly to the original data are

$$(\hat{\tilde{\mu}}, \hat{\sigma}, \hat{\xi}) = (1.64, 0.27, -0.084).$$

The change of sign of the location parameter induces a change to the sign of appropriate components of the variance-covariance matrix, which now becomes

$$V = \begin{bmatrix} 0.00141 & -0.000214 & 0.000795 \\ -0.000214 & 0.000652 & -0.000441 \\ 0.000795 & -0.0000441 & 0.00489 \end{bmatrix}.$$

Returning to the GEV analysis, the diagnostic plots for the fitted model are shown in Fig. 3.7. The probability and quantile plots are less convincing than in the previous example, but there is less doubt about the quality of fit once confidence intervals are added to the return level curve.

Interpretation of return levels in this example needs some explanation. The "return period" of 1000 has a "return level" of around -0.4. Such terminology doesn't work so well here; the values imply that, given 1000 such glass fibers, just one would be expected to have a breaking strength below

0.4 units. This point also raises an issue that is often used as an objection to the use of the GEV family. Looking at the return level plot in Fig. 3.7, it is clear that the model extrapolates to positive values, or equivalently, to negative values of breaking strength. This is incompatible with the physical process under study. In fact, the estimated upper end-point of the fitted distribution is $\hat{z}_0 = \hat{\mu} - \hat{\sigma}/\hat{\xi} = 1.59$, corresponding to a breaking strength of -1.59 units. The situation would be worse had the estimate of ξ been non-negative, since the estimated upper end-point of the distribution would have been infinite. This situation is not uncommon. For example, GEV estimates of annual maxima of daily rainfall often lead to positive estimates of ξ, though it is unreasonable on physical grounds to believe that daily rainfall levels are truly without limit. What these arguments really illustrate is the danger of relying on the arguments leading to the GEV distribution as a basis for very long-term extrapolation. Although the arguments for fitting the GEV distribution to block maxima are compelling, the temptation to extrapolate to extremely high levels should be tempered by caution and physical knowledge.

3.5 Model Generalization: the r Largest Order Statistic Model

3.5.1 *Model Formulation*

An implicit difficulty in any extreme value analysis is the limited amount of data for model estimation. Extremes are scarce, by definition, so that model estimates, especially of extreme return levels, have a large variance. This issue has motivated the search for characterizations of extreme value behavior that enable the modeling of data other than just block maxima.

There are two well-known general characterizations. One is based on exceedances of a high threshold, the other is based on the behavior of the r largest order statistics within a block, for small values of r. Both characterizations can be unified using a point process representation discussed in Chapter 7. In this section we focus on a model for the r largest order statistics.

As in previous sections, we suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables, and aim to characterize the extremal behavior of the X_i. In Section 3.1.3 we obtained that the limiting distribution as $n \to \infty$ of M_n, suitably re-scaled, is GEV. We first extend this result to other extreme order statistics, by defining

$$M_n^{(k)} = k\text{th largest of } \{X_1, \ldots, X_n\},$$

and identifying the limiting behavior of this variable, for fixed k, as $n \to \infty$. The following result generalizes Theorem 3.1.

Theorem 3.4 If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\{(M_n - b_n)/a_n \le z\} \to G(z) \quad \text{as } n \to \infty$$

for some non-degenerate distribution function G, so that G is the GEV distribution function given by (3.2), then, for fixed k,

$$\Pr\{(M_n^{(k)} - b_n)/a_n \le z\} \to G_k(z)$$

on $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where

$$G_k(z) = \exp\{-\tau(z)\} \sum_{s=0}^{k-1} \frac{\tau(z)^s}{s!} \tag{3.14}$$

with

$$\tau(z) = \left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}.$$

□

Theorem 3.4 implies that, if the kth largest order statistic in a block is normalized in exactly the same way as the maximum, then its limiting distribution is of the form given by (3.14), the parameters of which correspond to the parameters of the limiting GEV distribution of the block maximum. Again, by absorbing the unknown scaling constants into the model location and scale parameters, it follows that, for large n, the approximate distribution of $M_n^{(k)}$ is within the family (3.14).

There is, however, a difficulty in using (3.14) as a model. The situation we are often faced with, as with the Venice sea-level example, is of having each of the largest r order statistics within each of several blocks, for some value of r. That is, we usually have the complete vector

$$\boldsymbol{M}_n^{(r)} = (M_n^{(1)}, \ldots, M_n^{(r)})$$

for each of several blocks. In the case of the Venice sea-level data, $r = 10$ for most of the blocks, each of which corresponds to one year of observations. Whilst Theorem 3.4 gives a family for the approximate distribution of each of the components of $\boldsymbol{M}_n^{(r)}$, it does not give the joint distribution of $\boldsymbol{M}_n^{(r)}$. Moreover, the components cannot be independent: $M_n^{(2)}$ can be no greater than $M_n^{(1)}$, for example, so the outcome of each component influences the distribution of the other. Consequently, the result of Theorem 3.4 does not in itself lead to a model for $\boldsymbol{M}_n^{(r)}$. Instead, we require a characterization of the limiting joint distribution of the entire vector $\boldsymbol{M}_n^{(r)}$. With appropriate re-scaling this can be achieved, but the limiting joint distribution it leads to is intractable. However, the following theorem gives the joint density function of the limit distribution.

Theorem 3.5 If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\{(M_n - b_n)/a_n \le z\} \to G(z) \quad \text{as } n \to \infty$$

for some non-degenerate distribution function G, then, for fixed r, the limiting distribution as $n \to \infty$ of

$$\tilde{M}_n^{(r)} = \left(\frac{M_n^{(1)} - b_n}{a_n}, \ldots, \frac{M_n^{(r)} - b_n}{a_n} \right)$$

falls within the family having joint probability density function

$$f(z^{(1)}, \ldots, z^{(r)}) = \exp\left\{ -\left[1 + \xi\left(\frac{z^{(r)} - \mu}{\sigma}\right)\right]^{-1/\xi} \right\} \times \prod_{k=1}^{r} \sigma^{-1} \left[1 + \xi\left(\frac{z^{(k)} - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}, \quad (3.15)$$

where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$; $z^{(r)} \le z^{(r-1)} \le \cdots \le z^{(1)}$; and $z^{(k)} : 1 + \xi(z^{(k)} - \mu)/\sigma > 0$ for $k = 1, \ldots, r$. □

Proofs of Theorems 3.4 and 3.5 are given in Chapter 7. In the case $r = 1$, (3.15) reduces to the GEV family of density functions. The case $\xi = 0$ in (3.15) is interpreted as the limiting form as $\xi \to 0$, leading to the family of density functions

$$f(z^{(1)}, \ldots, z^{(r)}) = \exp\left\{ -\exp\left[-\left(\frac{z^{(r)} - \mu}{\sigma}\right)\right]\right\} \times \prod_{k=1}^{r} \sigma^{-1} \exp\left[-\left(\frac{z^{(k)} - \mu}{\sigma}\right)\right], \quad (3.16)$$

for which the case $r = 1$ reduces to the density of the Gumbel family.

3.5.2 Modeling the r Largest Order Statistics

Starting with a series of independent and identically distributed variables, data are grouped into m blocks. In block i the largest r_i observations are recorded, leading to the series $M_i^{(r_i)} = (z_i^{(1)}, \ldots, z_i^{(r_i)})$, for $i = 1, \ldots, m$. It is usual to set $r_1 = \cdots = r_m = r$ for some specified value of r, unless fewer data are available in some blocks.

As with the GEV model, the issue of block size amounts to a trade-off between bias and variance that is usually resolved by making a pragmatic choice, such as a block size of length one year. The number of order statistics used in each block also comprises a bias-variance trade-off: small values of r generate few data leading to high variance; large values of r are likely to violate the asymptotic support for the model, leading to bias. In practice it is usual to select the r_i as large as possible, subject to adequate model diagnostics.

The likelihood for this model is obtained from (3.15) and (3.16), by absorbing the unknown scaling coefficients into location and scale parameters in the usual way, and by taking products across blocks. So, when $\xi \ne 0$,

$$L(\mu, \sigma, \xi) = \prod_{i=1}^{m} \left(\exp\left\{ -\left[1 + \xi\left(\frac{z_i^{(r_i)} - \mu}{\sigma}\right)\right]^{-1/\xi} \right\} \times \prod_{k=1}^{r_i} \sigma^{-1} \left[1 + \xi\left(\frac{z_i^{(k)} - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \right), \quad (3.17)$$

provided $1 + \xi(z^{(k)} - \mu)/\sigma > 0$, $k = 1, \ldots, r_i, i = 1, \ldots, m$; otherwise the likelihood is zero. When $\xi = 0$,

$$L(\mu, \sigma, \xi) = \prod_{i=1}^{m} \left(\exp\left\{ -\exp\left[-\left(\frac{z^{(r_i)} - \mu}{\sigma}\right)\right]\right\} \times \prod_{k=1}^{r_i} \sigma^{-1} \exp\left[-\left(\frac{z_i^{(k)} - \mu}{\sigma}\right)\right] \right). \quad (3.18)$$

The likelihood (3.17) and (3.18) or, more commonly, the corresponding log-likelihood, can be maximized numerically to obtain maximum likelihood estimates. Standard asymptotic likelihood theory also gives approximate standard errors and confidence intervals. In the special case of $r_i = 1$ for each i, the likelihood function reduces to the likelihood of the GEV model for block maxima. More generally, the r largest order statistic model gives a likelihood whose parameters correspond to those of the GEV distribution of block maxima, but which incorporates more of the observed extreme data. So, relative to a standard block maxima analysis, the interpretation of the parameters is unaltered, but precision should be improved due to the inclusion of extra information.

3.5.3 Venice Sea-level Data

These data, discussed in Example 1.5, consist of the 10 largest sea-levels in Venice over the period 1931–1981, except for the year 1935, for which only the largest 6 observations are available. So, with due allowance for the exceptional year, model (3.15) can be applied for any value of $r = 1, \ldots, 10$. Maximum likelihood estimates and standard errors are given in Table 3.1 for inferences based on selected values of r. As anticipated, with increasing

TABLE 3.1. Maximized log-likelihoods ℓ, parameter estimates and standard errors (in parentheses) of r largest order statistic model fitted to the Venice sea-level data with different values of r.

r	ℓ	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
1	−222.7	111.1 (2.6)	17.2 (1.8)	−0.077 (0.074)
5	−732.0	118.6 (1.6)	13.7 (0.8)	−0.088 (0.033)
10	−1149.3	120.4 (1.3)	12.7 (0.5)	−0.115 (0.019)

values of r, the standard errors decrease, corresponding to increased model precision. However, if the asymptotic approximation is valid for a particular choice of r, then parameter estimates should be stable when the model is fitted with fewer order statistics. But from Table 3.1, there is little evidence of stability in the location and scale parameter estimates, even once sampling variability is accounted for. This brings into doubt the validity of the model, at least for values of $r \geq 5$.

Since the parameters μ, σ and ξ correspond exactly to the GEV parameters of the annual maxima distribution, a more detailed assessment of model fit is derived from return level curves of the annual maximum distribution. These are constructed in exactly the same way as for the block maximum model, but this time using the maximum likelihood estimates and variance-covariance matrix from the r largest order statistic model. Fig. 3.8 shows these plots for each value of r from 2 to 10. Even in the case of $r = 2$, the fit is not particularly good; the reasons for this will be discussed in Chapter 6. Notwithstanding general concerns about the lack of fit, Fig. 3.8 also illustrates that the agreement between model and data diminishes as r increases, although the confidence intervals become less wide. This is a graphic illustration of the bias-variance trade-off determined by the choice of r.

For any particular choice of r, the accuracy of the fit can also be examined in greater detail. First, the complete range of diagnostics for the block maximum can be examined. As an example, with $r = 5$, the usual suite of annual maximum diagnostics is shown in Fig. 3.9. Like the return level plots, these are obtained in exactly the same way as for the block maximum model, substituting the parameter estimates and variance-covariance matrix with those obtained by the maximization of (3.17). For the Venice data, the concern for lack of fit is reinforced by these diagnostics. Checks can also be made on the quality of fit for each of the order statistics by plotting probability and quantile plots. These are obtained by comparing the distribution of the kth order statistic – model (3.14), with parameter values replaced by their estimates – with the corresponding empirical estimates. For the probability plot this is straightforward. The quantile plot is more complicated, since (3.14) cannot be analytically inverted and it is

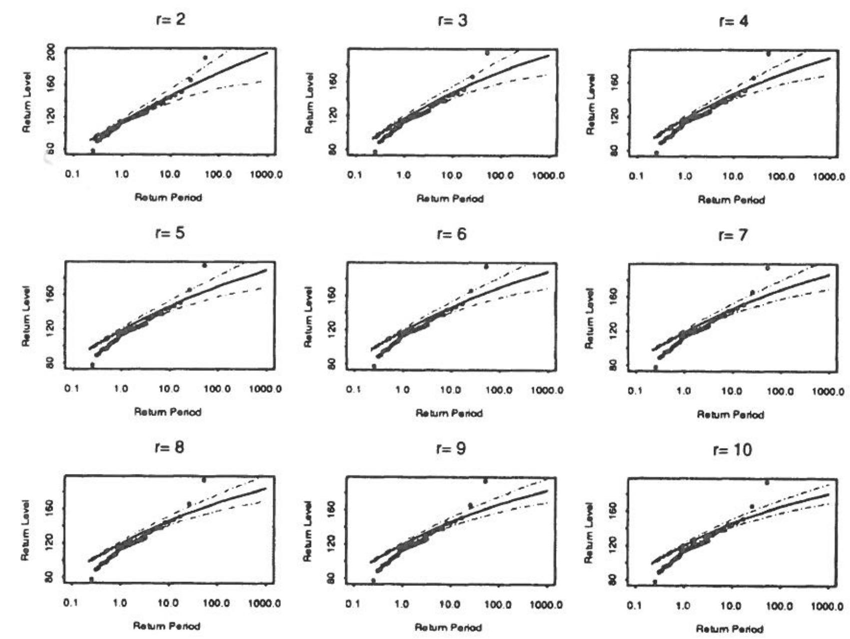

FIGURE 3.8. Return level estimates with 95% confidence intervals for annual maxima distribution based on r largest order statistic model fitted to the Venice sea-level data.

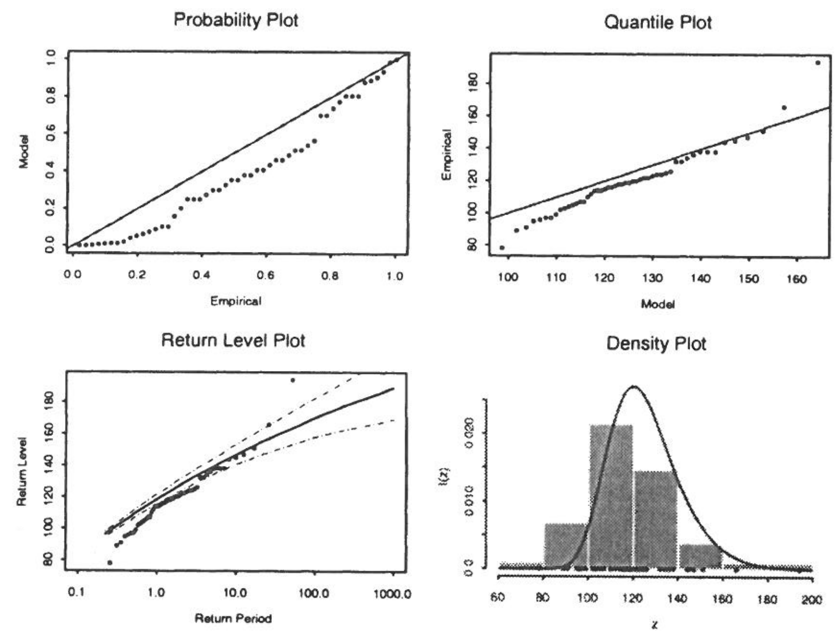

FIGURE 3.9. Annual maximum GEV diagnostics for the Venice sea-level data

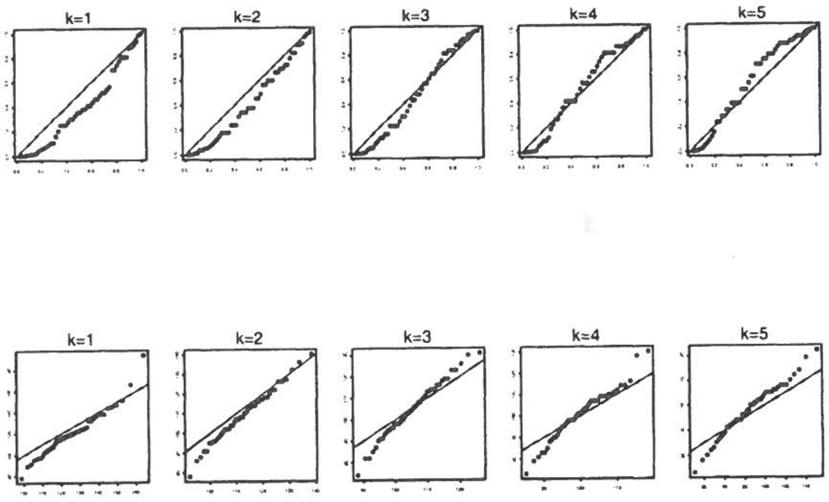

FIGURE 3.10. Model diagnostics for the Venice sea-level data on basis of fitted r largest order statistic model with $r = 5$. Plots shown are probability plots (top row) and quantile plots (bottom row) for kth largest order statistic, $k = 1, \ldots, 5$.

necessary to solve numerically the equation

$$G_k(z_p) = 1 - p$$

to obtain the model estimate of the $1 - p$ quantile. Nonetheless, this is straightforward using standard numerical techniques. For the Venice data, with the fitted model corresponding to $r = 5$, probability and quantile plots for each of the four largest order statistics are given in Fig. 3.10. These plots again indicate a fundamental lack of fit for the model.

3.6 Further Reading

The origins of the asymptotic sample maximum characterization can be traced back to Fisher & Tippett (1928). Their arguments were completed and formalized by Gnedenko (1943). Serious use of the block maximum model for statistical applications seems only to have started in the 1950's. Gumbel (1958) was influential in promoting the methodology, and this book is still relevant today. The GEV parameterization of the extreme

value limit models was independently proposed by von Mises (1954) and Jenkinson (1955).

Aspects of likelihood inference for the GEV model, and in particular the calculation of the expected information matrix, were considered by Prescott & Walden (1980); this was subsequently generalized to the case of censored data by Prescott & Walden (1983). An explicit algorithm for estimating maximum likelihood parameters was given by Hosking (1985). Smith (1985) contains important calculations for establishing the asymptotic properties of the maximum likelihood estimator for a class of models that includes the GEV. The issue of testing for the Gumbel model as a special case of the GEV distribution is discussed by Hosking (1984).

Competitor methods to maximum likelihood for estimating the parameters of the GEV distribution include the technique of probability weighted moments (Hosking et al., 1985) and methods based on order statistics (de Haan, 1990). Modified moment and likelihood techniques have also been proposed in a series of articles by JP Cohen: see Cohen (1988) for an overview and Smith (1995) for a general discussion and comparison.

There are numerous published applications of the GEV model in a variety of disciplines. A number of recent publications were listed in Section 1.1. Other influential examples include Buishand (1989) and Carter & Challenor (1981) for climatology; de Haan (1990), Tawn (1992) and Robinson & Tawn (1997) for oceanography; Zwiers (1987) and Walshaw & Anderson (2000) for wind field modeling; Henery (1984) and Robinson & Tawn (1995) for sports data modeling. The connections between extreme value models and reliability models are discussed in detail by Crowder et al. (1991). Applications in the context of corrosion engineering are described by Scarf & Laycock (1996).

Examples based on the r largest order statistic model are less common. As a modeling tool, the technique was first developed in the Gumbel case of $\xi = 0$ by Smith (1986), building on theoretical developments in Weissman (1978). The general case, having arbitrary ξ, was developed by Tawn (1988b).