4
Threshold Models

4.1 Introduction

As discussed in Chapter 3, modeling only block maxima is a wasteful approach to extreme value analysis if other data on extremes are available. Though the r largest order statistic model is a better alternative, it is unusual to have data of this form. Moreover, even this method can be wasteful of data if one block happens to contain more extreme events than another. If an entire time series of, say, hourly or daily observations is available, then better use is made of the data by avoiding altogether the procedure of blocking.

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, having marginal distribution function F. It is natural to regard as extreme events those of the X_i that exceed some high threshold u. Denoting an arbitrary term in the X_i sequence by X, it follows that a description of the stochastic behavior of extreme events is given by the conditional probability

$$\Pr\{X > u + y \mid X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, \quad y > 0. \tag{4.1}$$

If the parent distribution F were known, the distribution of threshold exceedances in (4.1) would also be known. Since, in practical applications, this is not the case, approximations that are broadly applicable for high values of the threshold are sought. This parallels the use of the GEV as an

approximation to the distribution of maxima of long sequences when the parent population is unknown.

4.2 Asymptotic Model Characterization

4.2.1 *The Generalized Pareto Distribution*

The main result is contained in the following theorem.

Theorem 4.1 Let $X_1, X_2, \ldots$ be a sequence of independent random variables with common distribution function F, and let

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

Denote an arbitrary term in the X_i sequence by X, and suppose that F satisfies Theorem 3.1.1, so that for large n,

$$\Pr\{M_n \le z\} \approx G(z),$$

where

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

for some $\mu, \sigma > 0$ and ξ. Then, for large enough u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{4.2}$$

defined on $\{y : y > 0 \text{ and } (1 + \xi y/\tilde{\sigma}) > 0\}$, where

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \tag{4.3}$$

□

Theorem 4.1 can also be made more precise, justifying (4.2) as a limiting distribution as u increases. In Section 4.2.2 we give an outline proof of the theorem as stated here.

The family of distributions defined by Eq. (4.2) is called the **generalized Pareto family**. Theorem 4.1 implies that, if block maxima have approximating distribution G, then threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Moreover, the parameters of the generalized Pareto distribution of threshold excesses are uniquely determined by those of the associated GEV distribution of block maxima. In particular, the parameter ξ in (4.2) is equal to that of the corresponding GEV distribution. Choosing a different, but still large, block size n would affect the values of the GEV parameters, but not those of the corresponding generalized Pareto distribution of threshold excesses: ξ is invariant to block size, while the calculation of $\tilde{\sigma}$ in (4.3) is unperturbed by the changes in μ and σ which are self-compensating.

The duality between the GEV and generalized Pareto families means that the shape parameter ξ is dominant in determining the qualitative behavior of the generalized Pareto distribution, just as it is for the GEV distribution. If $\xi < 0$ the distribution of excesses has an upper bound of $u - \tilde{\sigma}/\xi$; if $\xi > 0$ the distribution has no upper limit. The distribution is also unbounded if $\xi = 0$, which should again be interpreted by taking the limit $\xi \to 0$ in (4.2), leading to

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \quad y > 0, \tag{4.4}$$

corresponding to an exponential distribution with parameter $1/\tilde{\sigma}$.

4.2.2 *Outline Justification for the Generalized Pareto Model*

This section provides an outline proof of Theorem 4.1. A more precise argument is given by Leadbetter et al. (1983).

Let X have distribution function F. By the assumption of Theorem 3.1, for large enough n,

$$F^n(z) \approx \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for some parameters $\mu, \sigma > 0$ and ξ. Hence,

$$n \log F(z) \approx -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}. \tag{4.5}$$

But for large values of z, a Taylor expansion implies that

$$\log F(z) \approx -\{1 - F(z)\}.$$

Substitution into (4.5), followed by rearrangement, gives

$$1 - F(u) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-1/\xi}. \tag{4.6}$$

Hence,

$$\begin{aligned}\Pr\{X > u + y \mid X > u\} &\approx \frac{n^{-1}[1 + \xi(u + y - \mu)/\sigma]^{-1/\xi}}{n^{-1}[1 + \xi(u-\mu)/\sigma]^{-1/\xi}} \\ &= \left[1 + \frac{\xi(u+y-\mu)/\sigma}{1 + \xi(u-\mu)/\sigma}\right]^{-1/\xi} \\ &= \left[1 + \frac{\xi y}{\tilde{\sigma}}\right]^{-1/\xi},\end{aligned} \tag{4.7}$$

where

$$\tilde{\sigma} = \sigma + \xi(u - \mu),$$

as required.

4.2.3 *Examples*

We now reconsider the three theoretical examples of Section 3.1.5 in terms of threshold exceedance models.

Example 4.1 For the exponential model, $F(x) = 1 - e^{-x}$, for $x > 0$. By direct calculation,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}$$

for all $y > 0$. Consequently, the limit distribution of threshold exceedances is the exponential distribution, corresponding to $\xi = 0$ and $\tilde{\sigma} = 1$ in the generalized Pareto family. Furthermore, this is an exact result for all thresholds $u > 0$. ▲

Example 4.2 For the standard Fréchet model, $F(x) = \exp(-1/x)$, for $x > 0$. Hence,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - \exp\{-(u+y)^{-1}\}}{1 - \exp(-u^{-1})} \sim \left(1 + \frac{y}{u}\right)^{-1}$$

as $u \to \infty$, for all $y > 0$. This corresponds to the generalized Pareto distribution with $\xi = 1$ and $\tilde{\sigma} = u$. ▲

Example 4.3 For the uniform distribution model U(0, 1), $F(x) = x$, for $0 \le x \le 1$. Hence,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - (u+y)}{1 - u} = 1 - \frac{y}{1-u}$$

for $0 \le y \le 1 - u$. This corresponds to the generalized Pareto distribution with $\xi = -1$ and $\tilde{\sigma} = 1 - u$. ▲

Comparison of the limit families obtained here for threshold exceedances with the corresponding block maxima limits obtained in Section 3.1.5 confirms the duality of the two limit model formulations implied by Theorem 4.1. In particular, the values of ξ are common across the two models. Furthermore, the value of $\tilde{\sigma}$ is found to be threshold-dependent, except in the case where the limit model has $\xi = 0$, as implied by Eq. (4.3).

Until this point we have used the notation $\tilde{\sigma}$ to denote the scale parameter of the generalized Pareto distribution, so as to distinguish it from the corresponding parameter of the GEV distribution. For notational convenience we now drop this distinction, using σ to denote the scale parameter within either family.

4.3 Modeling Threshold Excesses

4.3.1 *Threshold Selection*

Theorem 4.1 suggests the following framework for extreme value modeling. The raw data consist of a sequence of independent and identically distributed measurements $x_1, \ldots, x_n$. Extreme events are identified by defining a high threshold u, for which the exceedances are $\{x_i : x_i > u\}$. Label these exceedances by $x_{(1)}, \ldots, x_{(k)}$, and define threshold excesses by $y_j = x_{(j)} - u$, for $j = 1, \ldots, k$. By Theorem 4.1, the y_j may be regarded as independent realizations of a random variable whose distribution can be approximated by a member of the generalized Pareto family. Inference consists of fitting the generalized Pareto family to the observed threshold exceedances, followed by model verification and extrapolation.

This approach contrasts with the block maxima approach through the characterization of an observation as extreme if it exceeds a high threshold. But the issue of threshold choice is analogous to the choice of block size in the block maxima approach, implying a balance between bias and variance. In this case, too low a threshold is likely to violate the asymptotic basis of the model, leading to bias; too high a threshold will generate few excesses with which the model can be estimated, leading to high variance. The standard practice is to adopt as low a threshold as possible, subject to the limit model providing a reasonable approximation. Two methods are available for this purpose: one is an exploratory technique carried out prior to model estimation; the other is an assessment of the stability of parameter estimates, based on the fitting of models across a range of different thresholds.

In more detail, the first method is based on the mean of the generalized Pareto distribution. If Y has a generalized Pareto distribution with parameters σ and ξ, then

$$\mathrm{E}(Y) = \frac{\sigma}{1-\xi}, \tag{4.8}$$

provided $\xi < 1$. When $\xi \geq 1$ the mean is infinite. Now, suppose the generalized Pareto distribution is valid as a model for the excesses of a threshold u_0 generated by a series $X_1, \ldots, X_n$, of which an arbitrary term is denoted X. By (4.8),

$$\mathrm{E}(X - u_0 \mid X > u_0) = \frac{\sigma_{u_0}}{1-\xi},$$

provided $\xi < 1$, where we adopt the convention of using σ_{u_0} to denote the scale parameter corresponding to excesses of the threshold u_0. But if the generalized Pareto distribution is valid for excesses of the threshold u_0, it should equally be valid for all thresholds $u > u_0$, subject to the appropriate change of scale parameter to σ_u. Hence, for $u > u_0$,

$$\begin{aligned} \mathrm{E}(X - u \mid X > u) &= \frac{\sigma_u}{1-\xi} \\ &= \frac{\sigma_{u_0} + \xi u}{1-\xi} \end{aligned} \tag{4.9}$$

by virtue of (4.3). So, for $u > u_0$, $\mathrm{E}(X - u \mid X > u)$ is a linear function of u. Furthermore, $\mathrm{E}(X - u \mid X > u)$ is simply the mean of the excesses of the threshold u, for which the sample mean of the threshold excesses of u provides an empirical estimate. According to (4.9), these estimates are expected to change linearly with u, at levels of u for which the generalized Pareto model is appropriate. This leads to the following procedure. The locus of points

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{\max} \right\},$$

where $x_{(1)}, \ldots, x_{(n_u)}$ consist of the n_u observations that exceed u, and $x_{\max}$ is the largest of the X_i, is termed the **mean residual life plot**. Above a threshold u_0 at which the generalized Pareto distribution provides a valid approximation to the excess distribution, the mean residual life plot should be approximately linear in u. Confidence intervals can be added to the plot based on the approximate normality of sample means.

The interpretation of a mean residual life plot is not always simple in practice. Fig. 4.1 shows the mean residual life plot with approximate 95% confidence intervals for the daily rainfall data of Example 1.6. Once the confidence intervals are taken into account, the graph appears to curve from $u = 0$ to $u \approx 30$, beyond which it is approximately linear until $u \approx 60$, whereupon it decays sharply. It is tempting to conclude that there is no stability until $u = 60$, after which there is approximate linearity. This suggests we take $u_0 = 60$. However, there are just 6 exceedances of the threshold $u = 60$, too few to make meaningful inferences. Moreover, the information in the plot for large values of u is unreliable due to the limited amount of data on which the estimate and confidence interval are based.

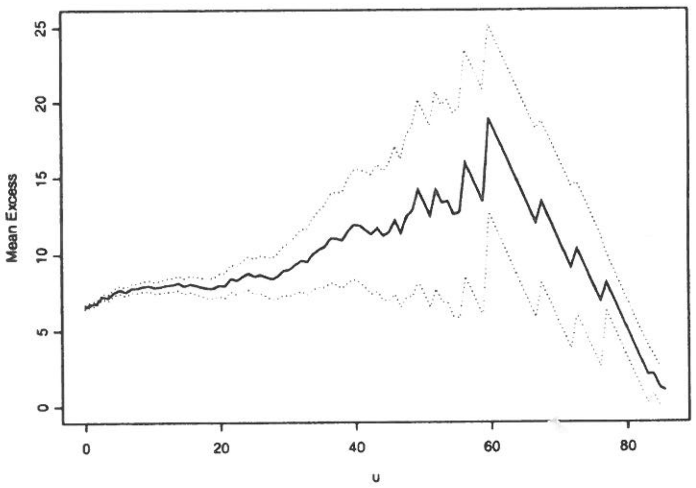

FIGURE 4.1. Mean residual life plot for daily rainfall data.

Accordingly, it is probably better to conclude that there is some evidence for linearity above $u = 30$, and to work initially with a threshold set at $u_0 = 30$.

The second procedure for threshold selection is to estimate the model at a range of thresholds. Above a level u_0 at which the asymptotic motivation for the generalized Pareto distribution is valid, estimates of the shape parameter ξ should be approximately constant, while estimates of σ_u should be linear in u, due to (4.9). We describe this method in greater detail in Section 4.3.4.

4.3.2 *Parameter Estimation*

Having determined a threshold, the parameters of the generalized Pareto distribution can be estimated by maximum likelihood. Suppose that the values $y_1, \ldots, y_k$ are the k excesses of a threshold u. For $\xi \neq 0$ the log-likelihood is derived from (4.2) as

$$\ell(\sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^{k} \log(1 + \xi y_i / \sigma), \tag{4.10}$$

provided $(1 + \sigma^{-1} \xi y_i) > 0$ for $i = 1, \ldots, k$; otherwise, $\ell(\sigma, \xi) = -\infty$. In the case $\xi = 0$ the log-likelihood is obtained from (4.4) as

$$\ell(\sigma) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^{k} y_i.$$

Analytical maximization of the log-likelihood is not possible, so numerical techniques are again required, taking care to avoid numerical instabilities when $\xi \approx 0$ in (4.10), and ensuring that the algorithm does not fail due to evaluation outside of the allowable parameter space. Standard errors and confidence intervals for the generalized Pareto distribution are obtained in the usual way from standard likelihood theory.

4.3.3 *Return Levels*

As discussed in Chapter 3, it is usually more convenient to interpret extreme value models in terms of quantiles or return levels, rather than individual parameter values. So, suppose that a generalized Pareto distribution with parameters σ and ξ is a suitable model for exceedances of a threshold u by a variable X. That is, for $x > u$,

$$\Pr\{X > x \mid X > u\} = \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}.$$

It follows that

$$\Pr\{X > x\} = \zeta_u \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}, \tag{4.11}$$

where $\zeta_u = \Pr\{X > u\}$. Hence, the level x_m that is exceeded on average once every m observations is the solution of

$$\zeta_u \left[1 + \xi \left(\frac{x_m - u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m}. \tag{4.12}$$

Rearranging,

$$x_m = u + \frac{\sigma}{\xi} \left[(m\zeta_u)^{\xi} - 1\right], \tag{4.13}$$

provided m is sufficiently large to ensure that $x_m > u$. This all assumes that $\xi \neq 0$. If $\xi = 0$, working in the same way with (4.4) leads to

$$x_m = u + \sigma \log(m\zeta_u), \tag{4.14}$$

again provided m is sufficiently large.

By construction, x_m is the **m-observation return level**. From (4.13) and (4.14), plotting x_m against m on a logarithmic scale produces the same qualitative features as return level plots based on the GEV model: linearity if $\xi = 0$; concavity if $\xi > 0$; convexity if $\xi < 0$. For presentation, it is often more convenient to give return levels on an annual scale, so that the N-year return level is the level expected to be exceeded once every N years. If there are n_y observations per year, this corresponds to the m-observation return level, where $m = N \times n_y$. Hence, the N-year return level is defined by

$$z_N = u + \frac{\sigma}{\xi}\left[(N n_y \zeta_u)^{\xi} - 1\right],$$

unless $\xi = 0$, in which case

$$z_N = u + \sigma \log(N n_y \zeta_u).$$

Estimation of return levels requires the substitution of parameter values by their estimates. For σ and ξ this corresponds to substitution by the corresponding maximum likelihood estimates, but an estimate of ζ_u, the probability of an individual observation exceeding the threshold u, is also needed. This has a natural estimator of

$$\hat{\zeta}_u = \frac{k}{n},$$

the sample proportion of points exceeding u. Since the number of exceedances of u follows the binomial $\mathrm{Bin}(n, \zeta_u)$ distribution, $\hat{\zeta}_u$ is also the maximum likelihood estimate of ζ_u.

Standard errors or confidence intervals for x_m can be derived by the delta method, but the uncertainty in the estimate of ζ_u should also be included in the calculation. From standard properties of the binomial distribution, $\mathrm{Var}(\hat{\zeta}_u) \approx \hat{\zeta}_u(1 - \hat{\zeta}_u)/n$, so the complete variance-covariance matrix for $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$ is approximately

$$V = \begin{bmatrix} \hat{\zeta}_u(1-\hat{\zeta}_u)/n & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix},$$

where $v_{i,j}$ denotes the (i,j) term of the variance-covariance matrix of $\hat{\sigma}$ and $\hat{\xi}$. Hence, by the delta method,

$$\mathrm{Var}(\hat{x}_m) \approx \nabla x_m^T V \nabla x_m, \tag{4.15}$$

where

$$\begin{aligned} \nabla x_m^T &= \left[\frac{\partial x_m}{\partial \zeta_u}, \frac{\partial x_m}{\partial \sigma}, \frac{\partial x_m}{\partial \xi}\right] \\ &= \left[\sigma m^{\xi} \zeta_u^{\xi-1}, \xi^{-1}\left\{(m\zeta_u)^{\xi} - 1\right\},\right. \\ &\qquad \left. -\sigma\xi^{-2}\left\{(m\zeta_u)^{\xi} - 1\right\} + \sigma\xi^{-1}(m\zeta_u)^{\xi}\log(m\zeta_u)\right], \end{aligned}$$

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

As with previous models, better estimates of precision for parameters and return levels are obtained from the appropriate profile likelihood. For σ or ξ this is straightforward. For return levels, a reparameterization is required. Life is made simpler by ignoring the uncertainty in ζ_u, which is usually small relative to that of the other parameters. From (4.13) and (4.14)

$$\sigma = \begin{cases} \dfrac{(x_m - u)\xi}{(m\zeta_u)^{\xi} - 1}, & \text{if } \xi \neq 0; \\ \dfrac{x_m - u}{\log(m\zeta_u)}, & \text{if } \xi = 0. \end{cases}$$

With fixed x_m, substitution into (4.10) leads to a one-parameter likelihood that can be maximized with respect to ξ. As a function of x_m, this is the profile log-likelihood for the m-observation return level.

4.3.4 Threshold Choice Revisited

We saw in Section 4.3 that mean residual life plots can be difficult to interpret as a method of threshold selection. A complementary technique is to fit the generalized Pareto distribution at a range of thresholds, and to look for stability of parameter estimates. The argument is as follows.

By Theorem 4.1, if a generalized Pareto distribution is a reasonable model for excesses of a threshold u_0, then excesses of a higher threshold u should also follow a generalized Pareto distribution. The shape parameters of the two distributions are identical. However, denoting by σ_u the value of the generalized Pareto scale parameter for a threshold of $u > u_0$, it follows from (4.3) that

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0), \tag{4.16}$$

so that the scale parameter changes with u unless $\xi = 0$. This difficulty can be remedied by reparameterizing the generalized Pareto scale parameter as

$$\sigma^* = \sigma_u - \xi u,$$

which is constant with respect to u by virtue of (4.16). Consequently, estimates of both σ^* and ξ should be constant above u_0, if u_0 is a valid threshold for excesses to follow the generalized Pareto distribution. Sampling variability means that the estimates of these quantities will not be exactly constant, but they should be stable after allowance for their sampling errors.

This argument suggests plotting both $\hat{\sigma}^*$ and $\hat{\xi}$ against u, together with confidence intervals for each of these quantities, and selecting u_0 as the lowest value of u for which the estimates remain near-constant. The confidence intervals for $\hat{\xi}$ are obtained immediately from the variance-covariance matrix V. Confidence intervals for $\hat{\sigma}^*$ require the delta method, using

$$\mathrm{Var}(\sigma^*) \approx \nabla\sigma^{*T} V \nabla\sigma^*,$$

where

$$\nabla\sigma^{*T} = \left[\frac{\partial\sigma^*}{\partial\sigma_u}, \frac{\partial\sigma^*}{\partial\xi}\right] = [1, -u].$$

4.3.5 Model Checking

Probability plots, quantile plots, return level plots and density plots are all useful for assessing the quality of a fitted generalized Pareto model. Assuming a threshold u, threshold excesses $y_{(1)} \le \cdots \le y_{(k)}$ and an estimated model $\hat{H}$, the probability plot consists of the pairs

$$\{(i/(k+1), \hat{H}(y_{(i)}));\ i = 1, \ldots, k\},$$

where

$$\hat{H}(y) = 1 - \left(1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right)^{-1/\hat{\xi}},$$

provided $\hat{\xi} \neq 0$. If $\hat{\xi} = 0$ the plot is constructed using (4.4) in place of (4.2). Again assuming $\hat{\xi} \neq 0$, the quantile plot consists of the pairs

$$\{(\hat{H}^{-1}(i/(k+1)), y_{(i)}),\ i = 1, \ldots, k\},$$

where

$$\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[y^{-\hat{\xi}} - 1\right].$$

If the generalized Pareto model is reasonable for modeling excesses of u, then both the probability and quantile plots should consist of points that are approximately linear.

A return level plot consists of the locus of points $\{(m, \hat{x}_m)\}$ for large values of m, where $\hat{x}_m$ is the estimated m-observation return level:

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[(m\hat{\zeta}_u)^{\hat{\xi}} - 1\right],$$

again modified if $\hat{\xi} = 0$. As with the GEV return level plot, it is usual to plot the return level curve on a logarithmic scale to emphasize the effect of extrapolation, and also to add confidence bounds and empirical estimates of the return levels.

Finally, the density function of the fitted generalized Pareto model can be compared to a histogram of the threshold exceedances.

4.4 Examples

4.4.1 Daily Rainfall Data

This example is based on the daily rainfall series discussed in Example 1.6. In Section 4.3 a mean residual life plot for these data suggested a threshold of $u = 30$. Further support for this choice is provided by the model-based check described in Section 4.3.4. The plots of $\hat{\sigma}^*$ and $\hat{\xi}$ against u are shown

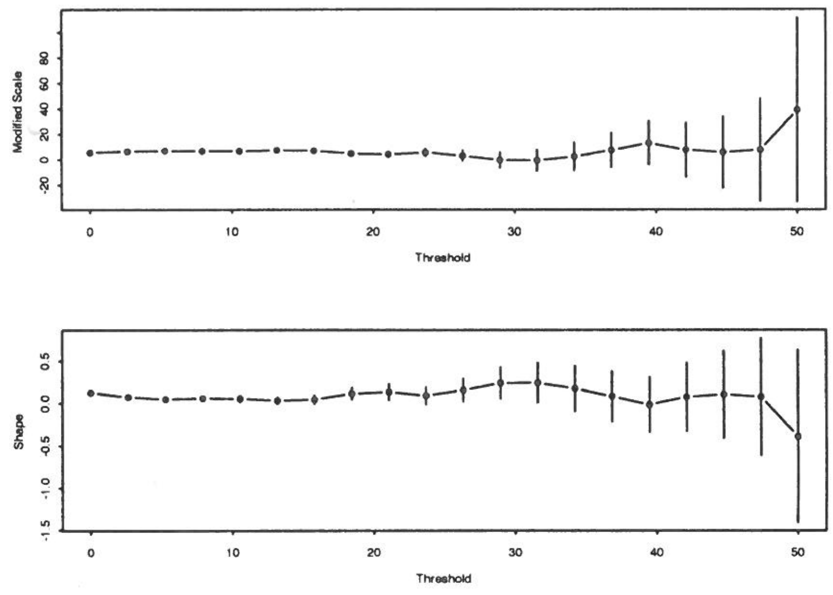

FIGURE 4.2. Parameter estimates against threshold for daily rainfall data.

in Fig. 4.2. The change in pattern for very high thresholds that was observed in the mean residual life plot is also apparent here, but the perturbations are now seen to be small relative to sampling errors. Hence, the selected threshold of $u = 30$ appears reasonable. Maximum likelihood estimates in this case are

$$(\hat{\sigma}, \hat{\xi}) = (7.44, 0.184)$$

with a corresponding maximized log-likelihood of -485.1. The variance-covariance matrix is calculated as

$$\begin{bmatrix} 0.9188 & -0.0655 \\ -0.0655 & 0.0102 \end{bmatrix},$$

leading to standard errors of 0.959 and 0.101 for $\hat{\sigma}$ and $\hat{\xi}$ respectively. In particular, it follows that a 95% confidence interval for ξ is obtained as $0.184 \pm 1.96 \times 0.101 = [-0.014, 0.383]$. The maximum likelihood estimate corresponds, therefore, to an unbounded distribution (since $\hat{\xi} > 0$), and the evidence for this is reasonably strong, since the 95% interval for ξ is almost exclusively in the positive domain.

Since there are 152 exceedances of the threshold $u = 30$ in the complete set of 17531 observations, the maximum likelihood estimate of the exceedance probability is $\hat{\zeta}_u = 152/17531 = 0.00867$, with approximate variance $\text{Var}(\hat{\zeta}_u) = \hat{\zeta}_u(1 - \hat{\zeta}_u)/17531 = 4.9 \times 10^{-7}$. Hence, the complete variance-covariance matrix for $(\hat{\zeta}, \hat{\sigma}, \hat{\xi})$ is

$$V = \begin{bmatrix} 4.9 \times 10^{-7} & 0 & 0 \\ 0 & 0.9188 & -0.0655 \\ 0 & -0.0655 & 0.0102 \end{bmatrix}.$$

Since $\hat{\xi} > 0$, it is not useful to carry out a detailed inference of the upper limit. Instead, we focus on extreme return levels. The data are daily, so the 100-year return level corresponds to the m-observation return level with $m = 365 \times 100$. Substitution into (4.13) and (4.15) gives $\hat{x}_m = 106.3$ and $\text{Var}(\hat{x}_m) = 431.3$, leading to a 95% confidence interval for x_m of $106.3 \pm 1.96\sqrt{431.3} = [65.6, 147.0]$.

Comparison with the observed data casts some doubt on the accuracy of this interval: in the 48 years of observation, the lower interval limit of 65.6 was exceeded 6 times, suggesting that the 100-year return level is almost certainly not as low in value as the confidence interval implies is plausible. Better accuracy is achieved by using profile likelihood intervals. Fig. 4.3 shows the profile log-likelihood for ξ in this example. By Theorem 2.6 an approximate 95% confidence interval for ξ is obtained from this graph as $[0.019, 0.418]$. This is not so different from the previous interval obtained previously, but strengthens slightly the conclusion that $\xi > 0$. The profile log-likelihood for the 100-year return level is plotted in Fig. 4.4. In this case the surface is highly asymmetric, reflecting the greater uncertainty about large values of the process. The 95% confidence interval for the 100-year return level is obtained from the profile log-likelihood as $[81.6, 185.7]$, an interval which now excludes the implausible range that formed part of the interval based on the delta method. Furthermore, the upper limit of 185.7 is very much greater than the delta-method value, more realistically accounting for the genuine uncertainties of extreme model extrapolation. This again illustrates that intervals derived from the delta method are anti-conservative, and should be replaced by profile likelihood intervals whenever a more precise measure of uncertainty is required.

Finally, diagnostic plots for the fitted generalized Pareto model are shown in Fig. 4.5. None of the plots gives any real cause for concern about the quality of the fitted model.

4.4.2 *Dow Jones Index Series*

The Dow Jones Index data discussed in Example 1.8 provide a second example for exploring the utility of the threshold exceedance modeling approach. Because of the strong non-stationarity observed in the original series $X_1, \ldots, X_n$, the data are transformed as $\tilde{X}_i = \log X_i - \log X_{i-1}$. Fig. 1.9 suggests that the transformed series is reasonably close to stationarity.

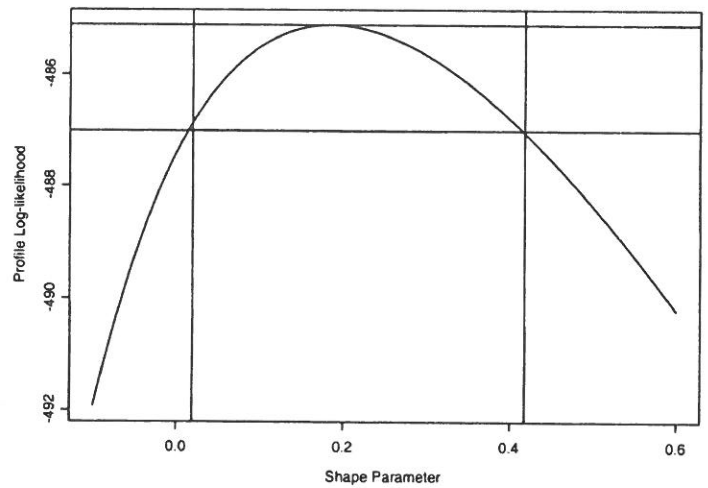

FIGURE 4.3. Profile likelihood for ξ in threshold excess model of daily rainfall data.

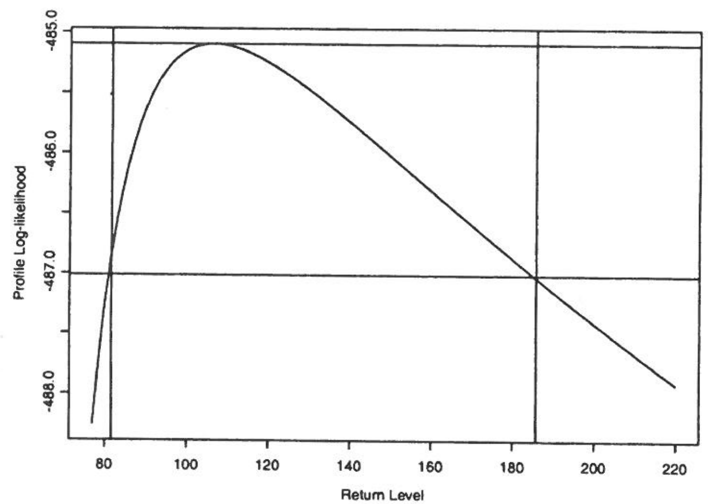

FIGURE 4.4. Profile likelihood for 100-year return level in threshold excess model of daily rainfall data.

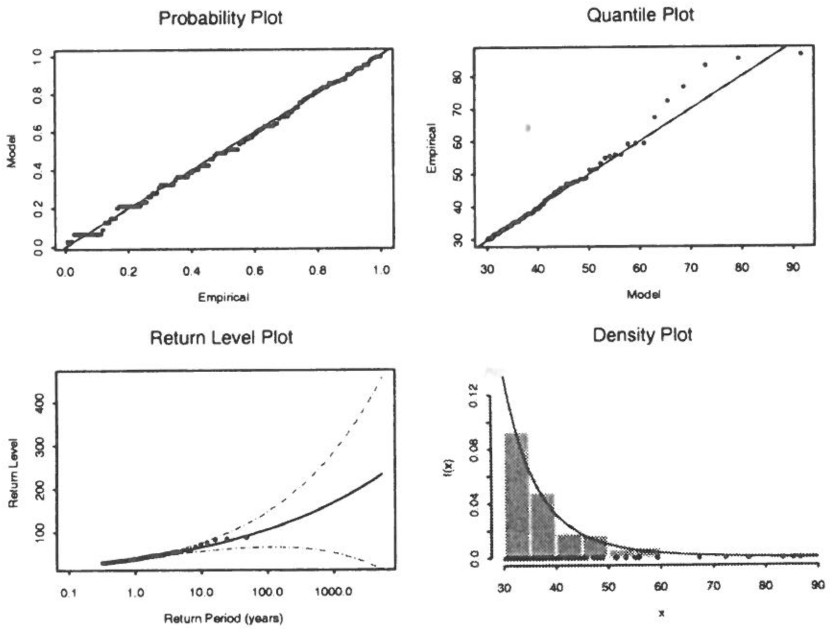

FIGURE 4.5. Diagnostic plots for threshold excess model fitted to daily rainfall data.

For convenience of presentation, the data are now re-scaled as $\tilde{X} \rightarrow 100\tilde{X}$. A mean residual life for the resulting $\tilde{X}_i$ series is shown in Fig. 4.6. The plot is initially linear, but shows substantial curvature in the range $-1 \le u \le 2$. For $u > 2$ the plot is reasonably linear when judged relative to confidence intervals, suggesting we set $u = 2$. This choice leads to 37 exceedances in the series of length 1303, so $\hat{\zeta}_u = 37/1303 = 0.028$. The maximum likelihood estimates of the generalized Pareto distribution parameters are $(\hat{\sigma}, \hat{\xi}) = (0.495, 0.288)$, with standard errors of 0.150 and 0.258 respectively. The maximum likelihood estimate corresponds, therefore, to an unbounded excess distribution, though the evidence for this is not overwhelming: 0 lies comfortably inside a 95% confidence interval for ξ.

Diagnostic plots for the fitted generalized Pareto distribution are shown in Fig. 4.7. The goodness-of-fit in the quantile plot seems unconvincing, but the confidence intervals on the return level plot suggest that the model departures are not large after allowance for sampling. The return level plot also illustrates the very large uncertainties that accrue once the model is extrapolated to higher levels.

In the context of financial modeling, extreme quantiles of the daily returns are generally referred to as the **value-at-risk**. It follows that the

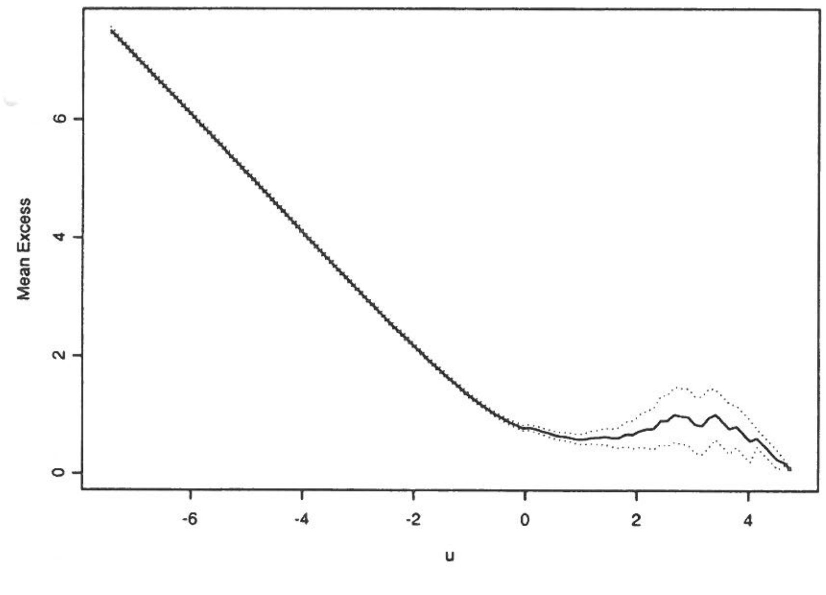

FIGURE 4.6. Mean residual life plot for transformed Dow Jones Index data.

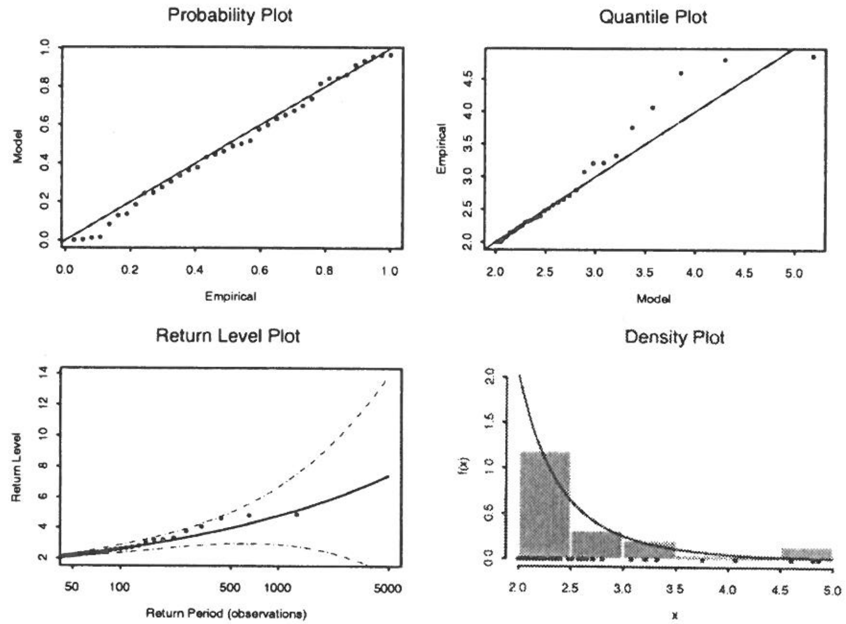

FIGURE 4.7. Diagnostic plots for threshold excess model fitted to transformed Dow Jones Index data.

generalized Pareto threshold model provides a direct method for the estimation of value-at-risk. Furthermore, the return level plot is simply a graph of value-at-risk against risk, on a convenient scale.

Finally, we discussed in the introduction of Example 1.8 that financial series often have a rich structure of temporal dependence. The transformation $X_i \to \tilde{X}_i$ is successful in reducing non-stationarity – the pattern of variation is approximately constant through time – but the induced series is not independent. That is, the distribution of $\tilde{X}_i$ is dependent on the history of the process $\{\tilde{X}_1, \ldots, \tilde{X}_{i-1}\}$. One illustration of this is provided by Fig. 4.8, which shows the threshold exceedances at their times of occurrence. If the series were independent the times of threshold exceedances would be uniform distributed; in actual fact the data of Fig. 4.8 demonstrate a tendency for the extreme events to cluster together. Such violation of the assumptions of Theorem 4.1 brings into doubt the validity of the simple threshold excess model for the Dow Jones Index series. We return to this issue in Chapter 5.

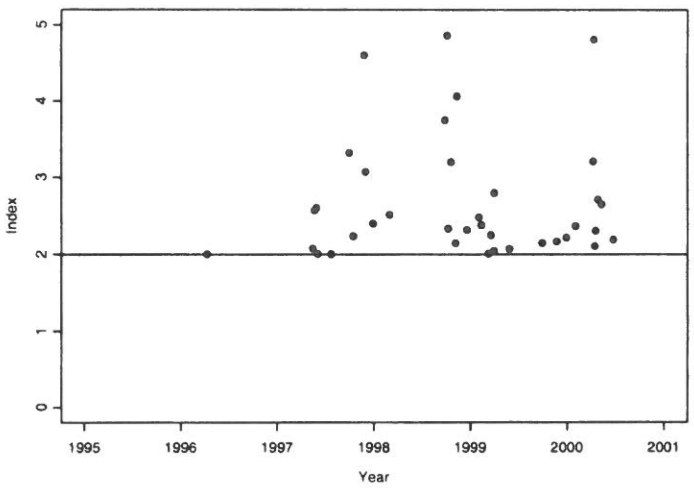

FIGURE 4.8. Threshold exceedances by transformed Dow Jones Index series.

4.5 Further Reading

The basic strategy of modeling threshold excesses has a long history in the hydrological literature, though the excesses were originally assumed to have a distribution belonging to the family of exponential distributions rather

than the complete generalized Pareto family. The arguments leading to the generalized Pareto model are attributable to Pickands (1975).

Statistical properties of the threshold approach were looked at in detail by Davison & Smith (1990), synthesizing earlier work in Davison (1984) and Smith (1984). In particular, they advocated the use of the mean residual life plot for threshold selection, and a likelihood approach to inference. Applications of the threshold approach are now widespread: see, for example, Grady (1992), Walshaw (1994) and Fitzgerald (1989).

A popular alternative to maximum likelihood estimation in the threshold excess model is a class of procedures based on functions of order statistics. These techniques often have an interpretation in terms of properties of quantile plots on appropriate scales, and generally incorporate procedures for threshold selection. Many of the techniques are variants on a proposal by Hill (1975); Dekkers & de Haan (1993), Beirlant et al. (1996) and Drees et al. (2000) make modified suggestions.