Extreme Values and Their Applications in Finance

by Ruey S. Tsay
Booth School of Business, The University of Chicago
and
Risk Management Institute, National University of Singapore

Extreme price movements in the financial markets are rare, but important. The stock market crash on Wall Street in October 1987 and other big financial crises such as the Long Term Capital Management and the bankruptcy of Lehman Brothers have attracted a great deal of attention among investors, practitioners and researchers. The recent worldwide financial crisis characterized by the substantial increase in market volatility, e.g., the VIX index, and the big drops in market indices has further generated discussions on market risk and margin setting for financial institutions. As a result, value at risk (VaR) has become the standard measure of market risk in risk management. Its usefulness and weaknesses are widely discussed.

In this lecture, we consider the extreme value theory developed in the statistical literature for studying rare (or extraordinary) events and its application to VaR. Both unconditional and conditional concepts of extreme values are discussed. The unconditional approach to VaR calculation for a financial position uses the historical returns of the instruments involved to compute VaR. On the other hand, a conditional approach uses the historical data and explanatory variables to calculate VaR. The explanatory variables may include macroeconomic variables of an economy and accounting variables of companies involved.

VaR is a point estimate of potential financial loss. It contains a certain degree of uncertainty. It also has a tendency to underestimate the actual loss if an extreme event actually occurs. To overcome the weaknesses of VaR, we discuss other risk measures such as expected shortfalls and the loss distribution of a financial position in the lecture.

We use daily log returns of IBM stock to illustrate the applications of the methods discussed. Figure 1 shows the time plot of daily log returns of IBM stock from July 3, 1962 to December 31, 1998 for 9190 observations.

1 Value at Risk

There are several types of risk in financial markets. Credit risk, operational risk, and market risk are the three main categories of financial risk. Value at risk (VaR) is mainly concerned with market risk, but the concept is also applicable to other types of risk. VaR is a single estimate of the amount by which an institution’s position in a risk category could decline due to general market movements during a given holding period; see Duffie and Pan (1997) and Jorion (2006) for a general exposition of VaR. The measure can be used by financial institutions to assess their risks or by a regulatory committee to set margin requirements. In either case, VaR is used to ensure that the financial institutions can still be in business after a catastrophic event. From the viewpoint of a financial institution, VaR can be defined as the maximal loss of a financial position during a given time period for a given probability. In this view, one treats VaR as a measure of loss associated with a rare (or extraordinary) event under normal market conditions. Alternatively, from the viewpoint of a regulatory committee, VaR can be defined as the minimal loss under extraordinary market conditions.
circumstances. Both definitions will lead to the same VaR measure, even though the concepts appear to be different.

In what follows, we define VaR under a probabilistic framework. Suppose that at the time index \( t \) we are interested in the risk of a financial position for the next \( \ell \) periods. Let \( \Delta V(\ell) \) be the change in value of the underlying assets of the financial position from time \( t \) to \( t + \ell \) and \( L(\ell) \) be the associated loss function. These two quantities are measured in dollars and are random variables at the time index \( t \). \( L(\ell) \) is a positive or negative function of \( \Delta V(\ell) \) depending on the position being short or long. Denote the cumulative distribution function (CDF) of \( L(\ell) \) by \( F_\ell(x) \). We define the VaR of a financial position over the time horizon \( \ell \) with tail probability \( p \) as

\[
p = \Pr[L(\ell) \geq \text{VaR}] = 1 - \Pr[L(\ell) < \text{VaR}].
\]

(1)

From the definition, the probability that the position holder would encounter a loss greater than or equal to VaR over the time horizon \( \ell \) is \( p \). Alternatively, VaR can be interpreted as follows. With probability \( (1 - p) \), the potential loss encountered by the holder of the financial position over the time horizon \( \ell \) is less than VaR.

The previous definition shows that VaR is concerned with the upper tail behavior of the loss CDF \( F_\ell(x) \). For any univariate CDF \( F_\ell(x) \) and probability \( q \), such that \( 0 < q < 1 \), the quantity

\[
x_q = \inf \{x \mid F_\ell(x) \geq q\}
\]

is called the \( q \)th quantile of \( F_\ell(x) \), where \( \inf \) denotes the smallest real number \( x \) satisfying \( F_\ell(x) \geq q \). If the random variable \( L(\ell) \) of \( F_\ell(x) \) is continuous, then \( q = \Pr[L(\ell) \leq x_q] \).

If the CDF \( F_\ell(x) \) of Eq. (1) is known, then \( 1 - p = \Pr[L(\ell) < \text{VaR}] \) so that VaR is simply the \( (1 - p) \)th quantile of the CDF of the loss function \( L(\ell) \) (i.e., \( \text{VaR} = x_{1-p} \)). Sometimes, VaR is referred to as the upper \( p \)th quantile because \( p \) is the upper tail probability of the loss distribution. The CDF is unknown in practice, however. Studies of VaR are essentially concerned with estimation of the CDF and/or its quantile, especially the upper tail behavior of the loss CDF.
In real applications, calculation of VaR involves several factors:

1. The probability of interest \( p \), such as \( p = 0.01 \) for risk management and \( p = 0.001 \) in stress testing.

2. The time horizon \( \ell \). It might be set by a regulatory committee, such as 1 day or 10 days for market risk and 1 year or 5 years for credit risk.

3. The frequency of the data, which might not be the same as the time horizon \( \ell \). Daily observations are often used in market risk analysis.

4. The CDF \( F_\ell(x) \) or its quantiles.

5. The amount of the financial position or the mark-to-market value of the portfolio.

Among these factors, the CDF \( F_\ell(x) \) is the focus of econometric modeling. Different methods for estimating the CDF give rise to different approaches to VaR calculation.

**Remark.** The definition of VaR in Eq. (1) is based on the upper tail of a loss function. For a long financial position, loss occurs when the returns are negative. Therefore, we shall use negative returns in data analysis for a long financial position. Furthermore, the VaR defined in Eq. (1) is in dollar amount. Since log returns correspond approximately to percentage changes in value of a financial asset, we use log returns \( r_t \) in data analysis. The VaR calculated from the upper quantile of the distribution of \( r_{t+1} \) given information available at time \( t \) is therefore in percentage. The dollar amount of VaR is then the cash value of the financial position times the VaR of the log return series. That is, \( \text{VaR} = \text{Value} \times \text{VaR of log returns} \). If necessary, one can also use the approximation \( \text{VaR} = \text{Value} \times [\exp(\text{VaR of log returns}) - 1] \). □

**Remark.** VaR is a prediction concerning possible loss of a portfolio in a given time horizon. It should be computed using the predictive distribution of future returns of the financial position. For example, the VaR for a 1-day horizon of a portfolio using daily returns \( r_t \) should be calculated using the predictive distribution of \( r_{t+1} \) given information available at time \( t \). From a statistical viewpoint, predictive distribution takes into account the parameter uncertainty in a properly specified model. However, predictive distribution is hard to obtain, and most of the available methods for VaR calculation ignore the effects of parameter uncertainty. □

**Remark.** From the prior discussion, VaR is a point estimate of potential loss. It tends to underestimate the actual loss when an extreme event occurs resulting in a loss belongs to the upper tail of the loss distribution. Ideally, we prefer knowing the loss distribution over a single upper quantile. Therefore, care must be exercised in using VaR to measure risk. We discuss the concept of expected shortfall later as an alternative to measuring risk. □

### 2 Extreme Value Theory

In this section, we review some extreme value theory in the statistical literature. Denote the return of an asset, measured in a fixed time interval such as daily, by \( r_t \). Consider the collection of \( n \) returns, \( \{r_1, \ldots, r_n\} \). The minimum return of the collection is \( r_{(1)} \), that is, the smallest order
statistic, whereas the maximum return is $r_{(n)}$, the maximum order statistic. Specifically, $r_{(1)} = \min_{1 \leq j \leq n} \{r_j\}$ and $r_{(n)} = \max_{1 \leq j \leq n} \{r_j\}$. Following the literature and using the loss function in VaR calculation, we focus on properties of the maximum return $r_{(n)}$. However, the theory discussed also applies to the minimum return of an asset over a given time period because properties of the minimum return can be obtained from those of the maximum by a simple sign change. Specifically, we have $r_{(1)} = -\max_{1 \leq j \leq n} \{-r_j\} = -r_{(n)}^c$, where $r_t^c = -r_t$ with the superscript $c$ denoting sign change. The minimum return is relevant to holding a long financial position. As before, we shall use negative log returns, instead of the log returns, to perform VaR calculation for a long position.

2.1 Review of Extreme Value Theory

Assume that the returns $r_t$ are serially independent with a common cumulative distribution function $F(x)$ and that the range of the return $r_t$ is $[l, u]$. For log returns, we have $l = -\infty$ and $u = \infty$. Then the CDF of $r_{(n)}$, denoted by $F_{n,n}(x)$, is given by

$$F_{n,n}(x) = \Pr[r_{(n)} \leq x] = \Pr(r_1 \leq x, r_2 \leq x, \ldots, r_n \leq x) \quad \text{(by definition of maximum)}$$

$$= \prod_{j=1}^{n} \Pr(r_j \leq x) \quad \text{(by independence)}$$

$$= \prod_{j=1}^{n} F(x) = [F(x)]^n. \quad (2)$$

In practice, the CDF $F(x)$ of $r_t$ is unknown and, hence, $F_{n,n}(x)$ of $r_{(n)}$ is unknown. However, as $n$ increases to infinity, $F_{n,n}(x)$ becomes degenerated — namely, $F_{n,n}(x) \to 0$ if $x < u$ and $F_{n,n}(x) \to 1$ if $x \geq u$ as $n$ goes to infinity. This degenerated CDF has no practical value. Therefore, the extreme value theory is concerned with finding two sequences $\{\beta_n\}$ and $\{\alpha_n\}$, where $\alpha_n > 0$, such that the distribution of $r_{(n^*)} \equiv (r_{(n)} - \beta_n)/\alpha_n$ converges to a nondegenerate distribution as $n$ goes to infinity. The sequence $\{\beta_n\}$ is a location series and $\{\alpha_n\}$ is a series of scaling factors. Under the independent assumption, the limiting distribution of the normalized minimum $r_{(n^*)}$ is given by

$$F_*(x) = \begin{cases} 
\exp[-(1 + \xi x)^{-1/\xi}] & \text{if } \xi \neq 0 \\
\exp[-\exp(-x)] & \text{if } \xi = 0
\end{cases} \quad (3)$$

for $x < -1/\xi$ if $\xi < 0$ and for $x > -1/\xi$ if $\xi > 0$, where the subscript * signifies the maximum. The case of $\xi = 0$ is taken as the limit when $\xi \to 0$. The parameter $\xi$ is referred to as the shape parameter that governs the tail behavior of the limiting distribution. The parameter $\alpha = 1/\xi$ is called the tail index of the distribution.

The limiting distribution in Eq. (3) is the generalized extreme value (GEV) distribution of Jenkinson (1955) for the maximum. It encompasses the three types of limiting distribution of Gnedenko (1943):

- Type I: $\xi = 0$, the Gumbel family. The CDF is
  $$F_*(x) = \exp[-\exp(-x)], \quad -\infty < x < \infty. \quad (4)$$

4
• Type II: $\xi > 0$, the Fréchet family. The CDF is

$$F_*(x) = \begin{cases} 
\exp[-(1 + \xi x)^{-1/\xi}] & \text{if } x > -1/\xi, \\
0 & \text{otherwise.} 
\end{cases} \quad (5)$$

• Type III: $\xi < 0$, the Weibull family. The CDF here is

$$F_*(x) = \begin{cases} 
\exp[-(1 + \xi x)^{-1/\xi}] & \text{if } x < -1/\xi, \\
1 & \text{otherwise.} 
\end{cases}$$

Gnedenko (1943) gave necessary and sufficient conditions for the CDF $F(x)$ of $r_t$ to be associated with one of the three types of limiting distribution. Briefly speaking, the tail behavior of $F(x)$ determines the limiting distribution $F_*(x)$ of the maximum. The right tail of the distribution declines exponentially for the Gumbel family, by a power function for the Fréchet family, and is finite for the Weibull family (Figure 2). Readers are referred to Embrechts, Kuppelberg, and Mikosch (1997) for a comprehensive treatment of the extreme value theory. For risk management, we are mainly interested in the Fréchet family that includes stable and Student-$t$ distributions. The Gumbel family consists of thin-tailed distributions such as normal and lognormal distributions. The probability density function (pdf) of the generalized limiting distribution in Eq. (3) can be obtained easily by differentiation:

$$f_*(x) = \begin{cases} 
(1 + \xi x)^{-1/\xi - 1} \exp[-(1 + \xi x)^{-1/\xi}] & \text{if } \xi \neq 0, \\
\exp[-x - \exp(-x)] & \text{if } \xi = 0, 
\end{cases} \quad (6)$$

where $-\infty < x < \infty$ for $\xi = 0$, and $x < -1/\xi$ for $\xi < 0$, and $x > -1/\xi$ for $\xi > 0$.

The aforementioned extreme value theory has two important implications. First, the tail behavior of the CDF $F(x)$ of $r_t$, not the specific distribution, determines the limiting distribution $F_*(x)$ of the (normalized) maximum. Thus, the theory is generally applicable to a wide range of distributions for the return $r_t$. The sequences $\{\beta_n\}$ and $\{\alpha_n\}$, however, may depend on the CDF $F(x)$. Second, Feller (1971, p. 279) shows that the tail index $\xi$ does not depend on the time interval of $r_t$. That is, the tail index (or equivalently the shape parameter) is invariant under time aggregation. This second feature of the limiting distribution becomes handy in the VaR calculation.

The extreme value theory has been extended to serially dependent observations $\{r_t\}_{t=1}^n$ provided that the dependence is weak. Berman (1964) shows that the same form of the limiting extreme value distribution holds for stationary normal sequences provided that the autocorrelation function of $r_t$ is squared summable (i.e., $\sum_{i=1}^{\infty} \rho_i^2 < \infty$), where $\rho_i$ is the lag-$i$ autocorrelation function of $r_t$. For further results concerning the effect of serial dependence on the extreme value theory, readers are referred to Leadbetter, Lindgren, and Rootzén (1983, Chapter 3). We shall discuss extremal index for a strictly stationary time series later in Section 5.

### 2.2 Empirical Estimation

The extreme value distribution contains three parameters $-\xi, \beta_n$, and $\alpha_n$. These parameters are referred to as the shape, location, and scale parameters, respectively. They can be estimated by using either parametric or nonparametric methods. We review some of the estimation methods.
Figure 2: Probability density functions of extreme value distributions for maximum. The solid line is for a Gumbel distribution, the dotted line is for the Weibull distribution with $\xi = -0.5$, and the dashed line is for the Fréchet distribution with $\xi = 0.9$.

For a given sample, there is only a single minimum or maximum, and we cannot estimate the three parameters with only an extreme observation. Alternative ideas must be used. One of the ideas used in the literature is to divide the sample into subsamples and apply the extreme value theory to the subsamples. Assume that there are $T$ returns $\{r_j\}^T_{j=1}$ available. We divide the sample into $g$ non-overlapping subsamples each with $n$ observations, assuming for simplicity that $T = ng$. In other words, we divide the data as

$$\{r_1, \ldots, r_n | r_{n+1}, \ldots, r_{2n} | r_{2n+1}, \ldots, r_{3n} | \cdots | r_{(g-1)n+1}, \ldots, r_{ng}\}$$

and write the observed returns as $r_{in+j}$, where $1 \leq j \leq n$ and $i = 0, \ldots, g - 1$. Note that each subsample corresponds to a subperiod of the data span. When $n$ is sufficiently large, we hope that the extreme value theory applies to each subsample. In application, the choice of $n$ can be guided by practical considerations. For example, for daily returns, $n = 21$ corresponds approximately to the number of trading days in a month and $n = 63$ denotes the number of trading days in a quarter. Let $r_{n,i}$ be the maximum of the $i$th subsample (i.e., $r_{n,i}$ is the largest return of the $i$th subsample), where the subscript $n$ is used to denote the size of the subsample. When $n$ is sufficiently large, $x_{n,i} = (r_{n,i} - \beta_n)/\alpha_n$ should follow an extreme value distribution, and the collection of subsample maxima $\{r_{n,i} | i = 1, \ldots, g\}$ can then be regarded as a sample of $g$ observations from that extreme value distribution. Specifically, we define

$$r_{n,i} = \max_{1 \leq j \leq n} \{r_{(i-1)n+j}\}, \quad i = 1, \ldots, g. \quad (7)$$

The collection of subsample maxima $\{r_{n,i}\}$ is the data we use to estimate the unknown parameters of the extreme value distribution. Clearly, the estimates obtained may depend on the choice of subperiod length $n$. 
Remark. When $T$ is not a multiple of the subsample size $n$, several methods have been used to deal with this issue. First, one can allow the last subsample to have a smaller size. Second, one can ignore the first few observations so that each subsample has size $n$. □.

The Parametric Approach
Two parametric approaches are available. They are the maximum likelihood and regression methods.

Maximum Likelihood Method
Assuming that the subperiod maxima $\{r_{n,i}\}$ follow a generalized extreme value distribution such that the pdf of $x_i = (r_{n,i} - \beta_n)/\alpha_n$ is given in Eq. (6), we can obtain the pdf of $r_{n,i}$ by a simple transformation as

$$f(r_{n,i}) = \begin{cases} \frac{1}{\alpha_n} (1 + \xi_n \frac{r_{n,i} - \beta_n}{\alpha_n})^{-(1+\xi_n)/\xi_n} \exp \left[ - \left( 1 + \xi_n \frac{r_{n,i} - \beta_n}{\alpha_n} \right)^{-1/\xi_n} \right] & \text{if } \xi_n \neq 0, \\ \frac{1}{\alpha_n} \exp \left[ - \frac{r_{n,i} - \beta_n}{\alpha_n} \right] - \exp \left( - \frac{r_{n,i} - \beta_n}{\alpha_n} \right) & \text{if } \xi_n = 0, \end{cases}$$

where it is understood that $1 + \xi_n (r_{n,i} - \beta_n)/\alpha_n > 0$ if $\xi_n \neq 0$. The subscript $n$ is added to the shape parameter $\xi$ to signify that its estimate depends on the choice of $n$. Under the independence assumption, the likelihood function of the subperiod maxima is

$$\ell(r_{n,1}, \ldots, r_{n,g} | \xi_n, \alpha_n, \beta_n) = \prod_{i=1}^{g} f(r_{n,i}).$$

Nonlinear estimation procedures can then be used to obtain maximum likelihood estimates of $\xi_n$, $\beta_n$, and $\alpha_n$. These estimates are unbiased, asymptotically normal, and of minimum variance under proper assumptions. See Embrechts et al. (1997) and Coles (2001) for details. We apply this approach to some stock return series later.

Regression method
This method assumes that $\{r_{n,i}\}_{i=1}^{g}$ is a random sample from the generalized extreme value distribution in Eq. (3) and makes use of properties of order statistics; see Gumbel (1958). Denote the order statistics of the subperiod maxima $\{r_{n,i}\}_{i=1}^{g}$ as $r_{n(1)} \leq r_{n(2)} \leq \cdots \leq r_{n(g)}$.

Using properties of order statistics (e.g., Cox and Hinkley, 1974, p. 467), we have

$$E\{F^*_i[r_{n(i)}]\} = \frac{i}{g+1}, \quad i = 1, \ldots, g. \quad \text{(8)}$$

For simplicity, we separate the discussion into two cases depending on the value of $\xi$. First, consider the case of $\xi \neq 0$. From Eq. (3), we have

$$F^*_i[r_{n(i)}] = \exp \left[ - \left( 1 + \xi_n \frac{r_{n(i)} - \beta_n}{\alpha_n} \right)^{-1/\xi_n} \right]. \quad \text{(9)}$$
Consequently, using Eqs. (8) and (9) and approximating expectation by an observed value, we have

\[
\frac{i}{g + 1} = \exp \left[ - \left( 1 + \xi_n \frac{r_{n(i)} - \beta_n}{\alpha_n} \right)^{-1/\xi_n} \right], \quad i = 1, \ldots, g.
\]

Taking natural logarithm twice, the prior equation gives

\[
\ln \left[ - \ln \left( \frac{i}{g + 1} \right) \right] = -\frac{1}{\xi_n} \ln \left( 1 + \xi_n \frac{r_{n(i)} - \beta_n}{\alpha_n} \right), \quad i = 1, \ldots, g.
\]

In practice, letting \( e_i \) be the deviation between the previous two quantities and assuming that the series \( \{e_i\} \) is not serially correlated, we have a regression setup

\[
\ln \left[ - \ln \left( \frac{i}{g + 1} \right) \right] = -\frac{1}{\xi_n} \ln \left( 1 + \xi_n \frac{r_{n(i)} - \beta_n}{\alpha_n} \right) + e_i, \quad i = 1, \ldots, g.
\] (10)

The least squares estimates of \( \xi_n, \beta_n, \) and \( \alpha_n \) can be obtained by minimizing the sum of squares of \( e_i \). When \( \xi_n = 0 \), the regression setup reduces to

\[
\ln \left[ - \ln \left( \frac{i}{g + 1} \right) \right] = -\frac{1}{\alpha_n} r_{n(i)} + \frac{\beta_n}{\alpha_n} + e_i, \quad i = 1, \cdots, g.
\]

The least squares estimates are consistent but less efficient than the likelihood estimates. We use the likelihood estimates in this chapter.

The Nonparametric Approach

The shape parameter \( \xi \) can be estimated using some nonparametric methods. We mention two such methods here. These two methods are proposed by Hill (1975) and Pickands (1975) and are referred to as the Hill estimator and Pickands estimator, respectively. Both estimators apply directly to the returns \( \{r_t\}_{t=1}^T \). Thus, there is no need to consider subsamples. Denote the order statistics of the sample as

\[
r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(T)}.
\]

Let \( q \) be a positive integer. The two estimators of \( \xi \) are defined as

\[
\xi_p(q) = \frac{1}{\ln(2)} \ln \left( \frac{r_{(T-q+1)} - r_{(T-2q+1)}}{r_{(T-2q+1)} - r_{(T-4q+1)}} \right), \quad q \leq T/4,
\] (11)

\[
\xi_h(q) = \frac{1}{q} \sum_{i=1}^{q} \left[ \ln(r_{(T-i+1)}) - \ln(r_{(T-q)}) \right],
\] (12)

where the argument \( (q) \) is used to emphasize that the estimators depend on \( q \) and the subscripts \( p \) and \( h \) denote Pickands and Hill estimator, respectively. The choice of \( q \) differs between Hill and Pickands estimators. It has been investigated by several researchers, but there is no general consensus on the best choice available. Dekkers and De Haan (1989) show that \( \xi_p(q) \) is consistent if \( q \) increases at a properly chosen pace with the sample size \( T \). In addition, \( \sqrt{q} |\xi_p(q) - \xi| \) is asymptotically normal with mean zero and variance \( \frac{\xi^2(2^{2\xi+1} + 1)}{2(2^\xi - 1) \ln(2)} \). The Hill estimator is applicable to the Fréchet distribution only, but it is more efficient than the Pickands
estimator when applicable. Goldie and Smith (1987) show that $\sqrt{q}[\xi_h(q) - \xi]$ is asymptotically normal with mean zero and variance $\xi^2$. In practice, one may plot the Hill estimator $\xi_h(q)$ against $q$ and find a proper $q$ such that the estimate appears to be stable. The estimated tail index $\alpha = 1/\xi_h(q)$ can then be used to obtain extreme quantiles of the return series; see Zivot and Wang (2003).

### 2.3 Application to Stock Returns

We apply the extreme value theory to the daily log returns of IBM stock from July 3, 1962 to December 31, 1998. The returns are measured in percentages, and the sample size is 9190 (i.e., $T = 9190$). Figure 3 shows the time plots of extreme daily log returns when the length of the subperiod is 21 days, which corresponds approximately to a month. The October 1987 crash is clearly seen from the plot. Excluding the 1987 crash, the range of extreme daily log returns is between 0.5% and 13%.

Table 1 summarizes some estimation results of the shape parameter $\xi$ via the Hill estimator. Three choices of $q$ are reported in the table, and the results are stable. To provide an overall picture of the performance of the Hill estimator, Figure 4 shows the scatterplots of the Hill estimator $\xi_h(q)$ and its pointwise 95% confidence interval against $q$. For both positive and negative extreme daily log returns, the estimator is stable except for cases when $q$ is small. The estimated shape parameters are about 0.30 and are significantly different from zero at the asymptotic 5% level. The plots also indicate that the shape parameter $\xi$ appears to be larger for the negative extremes, indicating that the daily log return may have a heavier left tail. Overall, the result indicates that the distribution of daily log returns of IBM stock belongs to the Fréchet family. The analysis thus rejects the normality assumption commonly used in practice. Such a conclusion is in agreement with that of Longin (1996), who used a U.S. stock market index series. R and S-Plus commands used to perform
Figure 4: Scatterplots of the Hill estimator for the daily log returns of IBM stock. The sample period is from July 3, 1962 to December 31, 1998: the upper plot is for positive returns and the lower one for negative returns.

Table 1: Results of the Hill Estimator for Daily Log Returns of IBM Stock from July 3, 1962 to December 31, 1998. Standard errors are in parentheses.

<table>
<thead>
<tr>
<th>q</th>
<th>190</th>
<th>200</th>
<th>210</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_t$</td>
<td>0.300(0.022)</td>
<td>0.299(0.021)</td>
<td>0.305(0.021)</td>
</tr>
<tr>
<td>$-r_t$</td>
<td>0.290(0.021)</td>
<td>0.292(0.021)</td>
<td>0.289(0.020)</td>
</tr>
</tbody>
</table>

the analysis are given in the demonstration below.

Next, we apply the maximum likelihood method to estimate parameters of the generalized extreme value distribution for IBM daily log returns. Table 2 summarizes the estimation results for different choices of the length of subperiods ranging from 1 month ($n = 21$) to 1 year ($n = 252$). From the table, we make the following observations:

- Estimates of the location and scale parameters $\beta_n$ and $\alpha_n$ increase in modulus as $n$ increases. This is expected as magnitudes of the subperiod minimum and maximum are nondecreasing functions of $n$.
- Estimates of the shape parameter (or equivalently the tail index) are stable for the negative extremes when $n \geq 63$ and are approximately 0.33.
- Estimates of the shape parameter are less stable for the positive extremes. The estimates are smaller in magnitude but remain significantly different from zero.
- The results for $n = 252$ have higher variabilities as the number of subperiods $g$ is relatively small.

<table>
<thead>
<tr>
<th>Length of subperiod</th>
<th>Scale $\alpha_n$</th>
<th>Location $\beta_n$</th>
<th>Shape Par. $\xi_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimal returns</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 mon. ($n = 21, g = 437$)</td>
<td>0.823(0.035)</td>
<td>1.902(0.044)</td>
<td>0.197(0.036)</td>
</tr>
<tr>
<td>1 qur ($n = 63, g = 145$)</td>
<td>0.945(0.077)</td>
<td>2.583(0.090)</td>
<td>0.335(0.076)</td>
</tr>
<tr>
<td>6 mon. ($n = 126, g = 72$)</td>
<td>1.147(0.131)</td>
<td>3.141(0.153)</td>
<td>0.330(0.101)</td>
</tr>
<tr>
<td>1 year ($n = 252, g = 36$)</td>
<td>1.542(0.242)</td>
<td>3.761(0.285)</td>
<td>0.322(0.127)</td>
</tr>
<tr>
<td><strong>Maximal returns</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 mon. ($n = 21, g = 437$)</td>
<td>0.931(0.039)</td>
<td>2.184(0.050)</td>
<td>0.168(0.036)</td>
</tr>
<tr>
<td>1 qur ($n = 63, g = 145$)</td>
<td>1.157(0.087)</td>
<td>3.012(0.108)</td>
<td>0.217(0.066)</td>
</tr>
<tr>
<td>6 mon. ($n = 126, g = 72$)</td>
<td>1.292(0.158)</td>
<td>3.471(0.181)</td>
<td>0.349(0.130)</td>
</tr>
<tr>
<td>1 year ($n = 252, g = 36$)</td>
<td>1.624(0.271)</td>
<td>4.475(0.325)</td>
<td>0.264(0.186)</td>
</tr>
</tbody>
</table>

Again the conclusion obtained is similar to that of Longin (1996), who provided a good illustration of applying the extreme value theory to stock market returns.

The results of Table 2 were obtained using a Fortran program developed by Professor Richard Smith and modified by the author. The package `evir` of R performs similar estimation. S-Plus is also based on the `evir` package. I demonstrate below the commands used. Note that the package uses subgroup maxima in the estimation so that negative log returns are used for holding long financial positions. Furthermore, $(\xi_n, \alpha_n, \beta_n)$ in the package corresponds to $(\xi_n, \alpha_n, \beta_n)$ of the table. The estimates obtained by R and S-Plus are close to those in Table 2. A source of the minor difference is that in Table 2, I dropped some data points at the beginning when the sample size $T$ is not a multiple of the subgroup size $n$. Consequently, Results of the R package have one more subgroup than that of Table 2.

**R Demonstration for Extreme Value Analysis**

The series is daily ibm log returns from 1962 to 1998. Output edited.

```r
> library(evir)
> help(hill)
> da=read.table("d-ibm6298.txt",header=T)
> ibm=log(da[,2]+1)*100
> nibm=-ibm
> par(mfcol=c(2,1)) # Obtain plots
> hill(ibm,option=c("xi"),end=500)
> hill(nibm,option=c("xi"),end=500)
# A simple R program to compute Hill estimate
> source("Hill.R")
> Hill
function(x,q){
# Compute the Hill estimate of the shape parameter.
# x: data and q: the number of order statistics used.
```

11
sx=sort(x)
T=length(x)
ist=T-q
y=log(sx[ist:T])
hill=sum(y[2:length(y)])/q
hill=hill-y[1]
sd=sqrt(hill^2/q)
cat("Hill estimate & std-err:",c(hill,sd),"
")

> m1=Hill(ibm,190)
Hill estimate & std-err: 0.3000144 0.02176533

> m1=Hill(nibm,190)
Hill estimate & std-err: 0.2903796 0.02106635

> m1=gev(nibm,block=21)
> m1
$n.all
[1] 9190
$n
[1] 438
$data
 [1] 3.2884827 3.6186920 3.9936970 ...$block
 [1] 21
$par.ests
   xi   sigma    mu
0.1954537 0.8240286 1.9033817
$par.ses
   xi   sigma    mu
0.03553259 0.03477151 0.04413856
$varcov
 [,1]       [,2]       [,3]
[1,] 1.262565e-03 -2.831235e-05 -0.0004336771
[2,] -2.831235e-05 1.209058e-03 0.0008477562
[3,] -4.336771e-04 8.477562e-04 0.0019482125

> names(m1)
[1] "n.all" "n" "data" "block" "par.ests"
[6] "par.ses" "varcov" "converged" "nllh.final"

> plot(m1)
Make a plot selection (or 0 to exit):
1: plot: Scatterplot of Residuals
2: plot: QQplot of Residuals
Selection: 1
Figure 5: Residual plots from fitting a GEV distribution to daily negative IBM log returns, in percentage, for data from July 3, 1962 to December 31, 1998 with a subperiod length of 21 days.

Define the residuals of a GEV distribution fit as

\[ w_i = \left( 1 + \xi_n \frac{r_{n,i} - \beta_n}{\alpha_n} \right)^{-1/\xi_n}. \]

Using the pdf of the GEV distribution and transformation of variables, one can easily show that \( \{w_i\} \) should form an iid random sample of exponentially distributed random variables if the fitted model is correctly specified. Figure 5 shows the residual plots of the GEV distribution fit to the daily negative IBM log returns with subperiod length of 21 days. The left panel gives the residuals and the right panel shows a quantile-to-quantile (QQ) plot against an exponential distribution. The plots indicate that the fit is reasonable.

**Remark.** Besides `evir`, several other packages are also available in R to perform extreme value analysis. They are `evd`, `POT` and `extRemes`.

### 3 Extreme Value Approach to VaR

In this section, we discuss an approach to VaR calculation using the extreme value theory. The approach is similar to that of Longin (1999a,b), who proposed an eight-step procedure for the same purpose. We divide the discussion into two parts. The first part is concerned with parameter estimation using the method discussed in the previous subsections. The second part focuses on VaR calculation by relating the probabilities of interest associated with different time intervals.

**Part I**

Assume that there are \( T \) observations of an asset return available in the sample period. We partition the sample period into \( g \) nonoverlapping subperiods of length \( n \) such that \( T = ng \). If \( T = ng + m \)
with $1 \leq m < n$, then we delete the first $m$ observations from the sample. The extreme value theory discussed in the previous section enables us to obtain estimates of the location, scale, and shape parameters $\beta_n, \alpha_n$, and $\xi_n$ for the subperiod maxima $\{r_{n,i}\}$. Plugging the maximum likelihood estimates into the CDF in Eq. (3) with $x = (r - \beta_n)/\alpha_n$, we can obtain the quantile of a given probability of the generalized extreme value distribution. Let $p^*$ be a small upper tail probability that indicates the potential loss and $r_n^*$ be the $(1 - p^*)^{th}$ quantile of the subperiod maxima under the limiting generalized extreme value distribution. Then we have

$$1 - p^* = \begin{cases} \exp \left[ -\left( 1 + \frac{\xi_n (r_n^* - \beta_n)}{\alpha_n} \right)^{-1/\xi_n} \right] & \text{if } \xi_n \neq 0, \\ \exp \left[ -\exp \left( -\frac{r_n^* - \beta_n}{\alpha_n} \right) \right] & \text{if } \xi_n = 0, \end{cases}$$

where it is understood that $1 + \xi_n (r_n^* - \beta_n)/\alpha_n > 0$ for $\xi_n \neq 0$. Rewriting this equation as

$$\ln(1 - p^*) = \begin{cases} -\left[ 1 + \frac{\xi_n (r_n^* - \beta_n)}{\alpha_n} \right]^{-1/\xi_n} & \text{if } \xi_n \neq 0, \\ -\exp \left( -\frac{r_n^* - \beta_n}{\alpha_n} \right) & \text{if } \xi_n = 0, \end{cases}$$

we obtain the quantile as

$$r_n^* = \begin{cases} \beta_n - \frac{\alpha_n}{\xi_n} \left[ 1 - \left[ -\ln(1 - p^*) \right]^{-\xi_n} \right] & \text{if } \xi_n \neq 0, \\ \beta_n - \alpha_n \ln[-\ln(1 - p^*)] & \text{if } \xi_n = 0. \end{cases} \quad (13)$$

In financial applications, the case of $\xi_n \neq 0$ is of major interest.

**Part II**

For a given upper tail probability $p^*$, the quantile $r_n^*$ of Eq. (13) is the VaR based on the extreme value theory for the subperiod maximum. The next step is to make explicit the relationship between subperiod maxima and the observed return $r_t$ series. Because most asset returns are either serially uncorrelated or have weak serial correlations, we may use the relationship in Eq. (2) and obtain

$$1 - p^* = P(r_{n,i} \leq r_n^*) = \left[ P(r_t \leq r_n^*) \right]^n. \quad (14)$$

This relationship between probabilities allows us to obtain VaR for the original asset return series $r_t$. More precisely, for a specified small upper probability $p$, the $(1 - p)^{th}$ quantile of $r_t$ is $r_n^*$ if the upper tail probability $p^*$ of the subperiod maximum is chosen based on Eq. (14), where $P(r_t \leq r_n^*) = 1 - p$. Consequently, for a given small upper tail probability $p$, the VaR of a financial position with log return $r_t$ is

$$\text{VaR} = \begin{cases} \beta_n - \frac{\alpha_n}{\xi_n} \left[ 1 - \left[ -n \ln(1 - p) \right]^{-\xi_n} \right] & \text{if } \xi_n \neq 0, \\ \beta_n - \alpha_n \ln[-n \ln(1 - p)] & \text{if } \xi_n = 0, \end{cases} \quad (15)$$
where \( n \) is the length of subperiod.

**Summary**

We summarize the approach of applying the traditional extreme value theory to VaR calculation as follows:

1. Select the length of the subperiod \( n \) and obtain subperiod maxima \( \{r_{n,i}\} \), \( i = 1, \ldots, g \), where \( g = \lfloor T/n \rfloor \).
2. Obtain the maximum likelihood estimates of \( \beta_n, \alpha_n, \) and \( \xi_n \).
3. Check the adequacy of the fitted extreme value model; see the next section for some methods of model checking.
4. If the extreme value model is adequate, apply Eq. (15) to calculate VaR.

**Remark.** Since we focus on loss function so that maxima of log returns are used in the derivation. Keep in mind that for a long financial position, the return series used in loss function is the negative log returns, not the traditional log returns. □

**Example 7.6.** Consider the daily log return, in percentage, of IBM stock from July 3, 1962 to December 31, 1998. From Table 2, we have \( \hat{\alpha}_n = 0.945 \), \( \hat{\beta}_n = 2.583 \), and \( \hat{\xi}_n = 0.335 \) for \( n = 63 \). Therefore, for the left-tail probability \( p = 0.01 \), the corresponding VaR is

\[
\text{VaR} = 2.583 - \frac{0.945}{0.335} \left(1 - \left[-63 \ln(1 - 0.01)\right]^{-0.335}\right)
\]

\[
= 3.04969.
\]

Thus, for daily negative log returns of the stock, the upper 1% quantile is 3.04969. If one holds a long position on the stock worth $10 million, then the estimated VaR with probability 1% is $10,000,000 \times 0.0304969 = $304,969. If the probability is 0.05, then the corresponding VaR is $166,641.

If we chose \( n = 21 \) (i.e., approximately 1 month), then \( \hat{\alpha}_n = 0.823 \), \( \hat{\beta}_n = 1.902 \), and \( \hat{\xi}_n = 0.197 \). The upper 1% quantile of the negative log returns based on the extreme value distribution is

\[
\text{VaR} = 1.902 - \frac{0.823}{0.197} \left(1 - \left[-21 \ln(1 - 0.01)\right]^{-0.197}\right) = 3.40013.
\]

Therefore, for a long position of $10,000,000, the corresponding 1-day horizon VaR is $340,013 at the 1% risk level. If the probability is 0.05, then the corresponding VaR is $184,127. In this particular case, the choice of \( n = 21 \) gives higher VaR values.

It is somewhat surprising to see that the VaR values obtained in Example 7.6 using the extreme value theory are smaller than those of Example 7.3 that uses a GARCH(1,1) model. In fact, the VaR values of Example 7.6 are even smaller than those based on the empirical quantile in Example 7.5. This is due in part to the choice of probability 0.05. If one chooses probability 0.001 = 0.1% and considers the same financial position, then we have \( \text{VaR} = 546,641 \) for the Gaussian AR(2)–GARCH(1,1) model and \( \text{VaR} = 666,590 \) for the extreme value theory with \( n = 21 \). Furthermore, the VaR obtained here via the traditional extreme value theory may not be adequate because the independent assumption of daily log returns is often rejected by statistical
testings. Finally, the use of subperiod maxima overlooks the fact of volatility clustering in the daily log returns. The new approach of extreme value theory discussed in the next section overcomes these weaknesses.

**Remark.** As shown by the results of Example 7.6, the VaR calculation based on the traditional extreme value theory depends on the choice of \(n\), which is the length of subperiods. For the limiting extreme value distribution to hold, one would prefer a large \(n\). But a larger \(n\) means a smaller \(g\) when the sample size \(T\) is fixed, where \(g\) is the effective sample size used in estimating the three parameters \(\alpha_n, \beta_n,\) and \(\xi_n\). Therefore, some compromise between the choices of \(n\) and \(g\) is needed. A proper choice may depend on the returns of the asset under study. We recommend that one should check the stability of the resulting VaR in applying the traditional extreme value theory.

### 3.1 Discussion

We have applied various methods of VaR calculation to the daily log returns of IBM stock for a long position of $10 million. Consider the VaR of the position for the next trading day. If the probability is 5%, which means that with probability 0.95 the loss will be less than or equal to the VaR for the next trading day, then the results obtained are

1. $302,500 for the RiskMetrics,
2. $287,200 for a Gaussian AR(2)\(−\)GARCH(1,1) model,
3. $283,520 for an AR(2)\(−\)GARCH(1,1) model with a standardized Student-\(t\) distribution with 5 degrees of freedom,
4. $216,030 for using the empirical quantile, and
5. $184,127 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length \(n = 21\)) of the log returns (or maxima of the negative log returns).

If the probability is 1%, then the VaR is

1. $426,500 for the RiskMetrics,
2. $409,738 for a Gaussian AR(2)\(−\)GARCH(1,1) model,
3. $475,943 for an AR(2)\(−\)GARCH(1,1) model with a standardized Student-\(t\) distribution with 5 degrees of freedom,
4. $365,709 for using the empirical quantile, and
5. $340,013 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length \(n = 21\)).

If the probability is 0.1%, then the VaR becomes

1. $566,443 for the RiskMetrics,
2. $546,641 for a Gaussian AR(2)\(−\)GARCH(1,1) model,
3. $836,341 for an AR(2)−GARCH(1,1) model with a standardized Student-\( t \) distribution with 5 degrees of freedom,

4. $780,712 for using the empirical quantile, and

5. $666,590 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length \( n = 21 \)).

There are substantial differences among different approaches. This is not surprising because there exists substantial uncertainty in estimating tail behavior of a statistical distribution. Since there is no true VaR available to compare the accuracy of different approaches, we recommend that one applies several methods to gain insight into the range of VaR.

The choice of tail probability also plays an important role in VaR calculation. For the daily IBM stock returns, the sample size is 9190 so that the empirical quantiles of 5% and 1% are decent estimates of the quantiles of the return distribution. In this case, we can treat the results based on empirical quantiles as conservative estimates of the true VaR (i.e., lower bounds). In this view, the approach based on the traditional extreme value theory seems to underestimate the VaR for the daily log returns of IBM stock. The conditional approach of extreme value theory discussed in the next section overcomes this weakness.

When the tail probability is small (e.g., 0.1%), the empirical quantile is a less reliable estimate of the true quantile. The VaR based on empirical quantiles can no longer serve as a lower bound of the true VaR. Finally, the earlier results show clearly the effects of using a heavy-tail distribution in VaR calculation when the tail probability is small. The VaR based on either a Student-\( t \) distribution with 5 degrees of freedom or the extreme value distribution is greater than that based on the normal assumption when the probability is 0.1%.

### 3.2 Multiperiod VaR

The square root of time rule of the RiskMetrics methodology becomes a special case under the extreme value theory. The proper relationship between \( \ell\)-day and 1-day horizons is

\[
\text{VaR}(\ell) = \ell^{1/\alpha}\text{VaR} = \ell^{\xi}\text{VaR},
\]

where \( \alpha \) is the tail index and \( \xi \) is the shape parameter of the extreme value distribution; see Danielsson and de Vries (1997a). This relationship is referred to as the \( \alpha \)-root of time rule. Here \( \alpha = 1/\xi \), not the scale parameter \( \alpha_q \).

For illustration, consider the daily log returns of IBM stock in Example 7.6. If we use \( p = 0.01 \) and the results of \( n = 63 \), then for a 30-day horizon we have

\[
\text{VaR}(30) = (30)^{0.335}\text{VaR} = 3.125 \times 304,969 = 952,997.
\]

Because \( \ell^{0.335} < \ell^{0.5} \), the \( \alpha \)-root of time rule produces lower \( \ell \)-day horizon VaR than the square root of time rule does.

### 3.3 Return Level

Another risk measure based on the extreme values of subperiods is the return level. The \( g \) \( n \)-subperiod return level, \( L_{n,g} \), is defined as the level that is exceeded in one out of every \( g \) subperiods.
of length $n$. That is,

$$P(r_{n,i} > L_{n,g}) = \frac{1}{g},$$

where $r_{n,i}$ denotes subperiod maximum. The subperiod in which the return level is exceeded is called a stress period. If the subperiod length $n$ is sufficiently large so that normalized $r_{n,i}$ follows the GEV distribution, then the return level is

$$L_{n,g} = \beta_n - \frac{\alpha_n}{\xi_n} \left\{ 1 - [-\ln(1 - 1/g)]^{-\xi_n} \right\},$$

provided that $\xi_n \neq 0$. Note that this is precisely the quantile of extreme value distribution given in Eq. (13) with tail probability $p^* = 1/g$, even though we write it in a slightly different way. Thus, return level applies to the subperiod maximum, not to the underlying returns. This marks the difference between VaR and return level.

For the daily negative IBM log returns with subperiod length of 21 days, we can use the fitted model to obtain the return level for 12 such subperiods (i.e., $g = 12$). The return level is 4.4835%.

### R and S-Plus Commands for Obtaining Return Level

R and S-Plus Commands for Obtaining Return Level

```r
> m1 = gev(nibm, block=21)
# S-Plus output
> rl.21.12 = rlevel.gev(m1, k.blocks=12, type='profile')
> class(rl.21.12)
[1] "list"
> names(rl.21.12)
[1] "Range" "rlevel"
> rl.21.12$rlevel
[1] 4.483506
# R output
> rl.21.12 = rlevel.gev(m1, k.blocks=12)
> rl.21.12
[1] 4.177923 4.481976 4.858102
```

In the prior demonstration, the number of subperiods is denoted by `k.blocks` and the subcommand, `type='profile'`, produces a plot of the profile log-likelihood confidence interval for the return level. The plot is not shown here.

### 4 A New Approach Based on the Extreme Value Theory

4 A New Approach Based on the Extreme Value Theory

The aforementioned approach to VaR calculation using the extreme value theory encounters some difficulties. First, the choice of subperiod length $n$ is not clearly defined. Second, the approach is unconditional and, hence, does not take into consideration effects of other explanatory variables. To overcome these difficulties, a modern approach to extreme value theory has been proposed in the statistical literature; see Davison and Smith (1990) and Smith (1989). Instead of focusing on the extremes (maximum or minimum), the new approach focuses on exceedances of the measurement over some high threshold and the times at which the exceedances occur. Thus, this new approach is
also referred to as *peaks over thresholds* (POT). For illustration, consider the daily returns of IBM stock used in this chapter and a long position on the stock. Denote the negative daily log return by \( r_t \). Let \( \eta \) be a prespecified high threshold. We may choose \( \eta = 2.5\% \). Suppose that the \( i \)th exceedance occurs at day \( t_i \) (i.e., \( r_{t_i} \leq \eta \)). Then the new approach focuses on the data \((t_i, r_{t_i} - \eta)\). Here \( r_{t_i} - \eta \) is the exceedance over the threshold \( \eta \) and \( t_i \) is the time at which the \( i \)th exceedance occurs. Similarly, for a short position, we may choose \( \eta = 2\% \) and focus on the data \((t_i, r_{t_i} - \eta)\) for which \( r_{t_i} \geq \eta \).

In practice, the occurrence times \( \{t_i\} \) provide useful information about the intensity of the occurrence of important “rare events” (e.g., less than the threshold \( \eta \) for a long position). A cluster of \( t_i \) indicates a period of large market declines. The exceeding amount (or exceedance) \( r_{t_i} - \eta \) is also of importance as it provides the actual quantity of interest.

Based on the prior introduction, the new approach does not require the choice of a subperiod length \( n \), but it requires the specification of threshold \( \eta \). Different choices of the threshold \( \eta \) lead to different estimates of the shape parameter \( k \) (and hence the tail index \( 1/\xi \)). In the literature, some researchers believe that the choice of \( \eta \) is a statistical problem as well as a financial one, and it cannot be determined based purely on statistical theory. For example, different financial institutions (or investors) have different risk tolerances. As such, they may select different thresholds even for an identical financial position. For the daily log returns of IBM stock considered in this chapter, the calculated VaR is not sensitive to the choice of \( \eta \).

The choice of threshold \( \eta \) also depends on the observed log returns. For a stable return series, \( \eta = 2.5\% \) may fare well for a long position. For a volatile return series (e.g., daily returns of a dot-com stock), \( \eta \) may be as high as 10\%. Limited experience shows that \( \eta \) can be chosen so that the number of exceedances is sufficiently large (e.g., about 5\% of the sample). For a more formal study on the choice of \( \eta \), see Danielsson and de Vries (1997b).

### 4.1 Statistical Theory

Again consider the log return \( r_t \) of an asset. Suppose that the \( i \)th exceedance occurs at \( t_i \). Focusing on the exceedance \( r_{t_i} - \eta \) and exceeding time \( t_i \) results in a fundamental change in statistical thinking. Instead of using the marginal distribution (e.g., the limiting distribution of the minimum or maximum), the new approach employs a conditional distribution to handle the magnitude of exceedance given that the measurement exceeds a threshold. The chance of exceeding the threshold is governed by a probability law. In other words, the new approach considers the conditional distribution of \( x = r_{t_i} - \eta \) given \( r_{t_i} \leq \eta \) for a long position. Occurrence of the event \( \{r_{t_i} \leq \eta\} \) follows a point process (e.g., a Poisson process). See Section ?? for the definition of a Poisson process. In particular, if the intensity parameter \( \lambda \) of the process is time-invariant, then the Poisson process is homogeneous. If \( \lambda \) is time-variant, then the process is nonhomogeneous. The concept of Poisson process can be generalized to the multivariate case.

The basic theory of the new approach is to consider the conditional distribution of \( r = x + \eta \) given \( r > \eta \) for the limiting distribution of the maximum given in Eq. (3). Since there is no need to choose the subperiod length \( n \), we do not use it as a subscript of the parameters. Then the conditional distribution of \( r \leq x + \eta \) given \( r > \eta \) is

\[
\Pr(r \leq x + \eta | r > \eta) = \frac{\Pr(\eta \leq r \leq x + \eta)}{\Pr(r > \eta)} = \frac{\Pr(r \leq x + \eta) - \Pr(r \leq \eta)}{1 - \Pr(r \leq \eta)}.
\]  

(16)
Using the CDF $F_\ast(.)$ of Eq. (3) and the approximation $e^{-y} \approx 1 - y$ and after some algebra, we obtain that

$$\Pr(r \leq x + \eta | r > \eta) = \frac{F_\ast(x + \eta) - F_\ast(\eta)}{1 - F_\ast(\eta)}$$

$$= \exp[-(1 + \frac{\xi(x+\eta-\beta)}{\alpha})^{-1/\xi}] - \exp[-(1 + \frac{\xi(\eta-\beta)}{\alpha})^{-1/\xi}]$$

$$1 - \exp[-(1 + \frac{\xi(\eta-\beta)}{\alpha})^{-1/\xi}]$$

$$\approx 1 - \left(1 + \frac{\xi x}{\alpha + \xi(\eta - \beta)}\right)^{-1/\xi},$$  \hspace{1cm} (17)

where $x > 0$ and $1 + \xi(\eta - \beta)/\alpha > 0$. As is seen later, this approximation makes explicit the connection of the new approach to the traditional extreme value theory. The case of $\xi = 0$ is taken as the limit of $\xi \to 0$ so that

$$\Pr(r \leq x + \eta | r > \eta) \approx 1 - \exp(-x/\alpha).$$

The distribution with cumulative distribution function

$$G_{\xi,\psi(\eta)}(x) = \begin{cases} 1 - \left[1 + \frac{\xi x}{\psi(\eta)}\right]^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp[-x/\psi(\eta)] & \text{for } \xi = 0, \end{cases}$$  \hspace{1cm} (18)

where $\psi(\eta) > 0$, $x \geq 0$ when $\xi \geq 0$, and $0 \leq x \leq -\psi(\eta)/\xi$ when $\xi < 0$, is called the generalized Pareto distribution (GPD). Thus, the result of Eq. (17) shows that the conditional distribution of $r$ given $r > \eta$ is well approximated by a GPD with parameters $\xi$ and $\psi(\eta) = \alpha + \xi(\eta - \beta)$. See Embrechts et al. (1997) for further information. Suppose that the excess distribution of $r$ given a threshold $\eta_o$ is a GPD with shape parameter $\xi$ and scale parameter $\psi(\eta_o)$. Then, for an arbitrary threshold $\eta > \eta_o$, the excess distribution over the threshold $\eta$ is also a GPD with shape parameter $\xi$ and scale parameter $\psi(\eta) = \psi(\eta_o) + \xi(\eta - \eta_o)$. When $\xi = 0$, the GPD in Eq. (18) reduces to an exponential distribution. This result motivates the use of a QQ-plot of excess returns over a threshold against exponential distribution to infer the tail behavior of the returns. If $\xi = 0$, then the QQ-plot should be linear. Figure 6a shows the QQ-plot of daily negative IBM log returns used in this chapter with threshold 0.025. The nonlinear feature of the plot clearly shows that the left-tail of the daily IBM log returns is heavier than that of a normal distribution, that is, $\xi \neq 0$.

**R and S-Plus Commands Used to Produce Figure 6**

```r
> par(mfcol=c(2,1))
> qplot(-ibm, threshold=0.025, main='Negative daily IBM log returns')
> meplot(-ibm)
> title(main='Mean excess plot')
```
4.2 Mean Excess Function

Given a high threshold \( \eta_o \), suppose that the excess \( r - \eta_o \) follows a GPD with parameter \( \xi \) and \( \psi(\eta_o) \), where \( 0 < \xi < 1 \). Then the mean excess over the threshold \( \eta_o \) is

\[
E(r - \eta_o | r > \eta_o) = \frac{\psi(\eta_o)}{1 - \xi}.
\]

For any \( \eta > \eta_o \), define the mean excess function \( e(\eta) \) as

\[
e(\eta) = E(r - \eta | r > \eta) = \frac{\psi(\eta_o) + \xi (\eta - \eta_o)}{1 - \xi}.
\]

In other words, for any \( y > 0 \),

\[
e(\eta_o + y) = E[r - (\eta_o + y) | r > \eta_o + y] = \frac{\psi(\eta_o) + \xi y}{1 - \xi}.
\]

Thus, for a fixed \( \xi \), the mean excess function is a linear function of \( y = \eta - \eta_o \). This result leads to a simple graphical method to infer the appropriate threshold value \( \eta_o \) for the GPD. Define the empirical mean excess function as

\[
e_T(\eta) = \frac{1}{N_\eta} \sum_{i=1}^{N_\eta} (r_{t_i} - \eta),
\]

where \( N_\eta \) is the number of returns that exceed \( \eta \) and \( r_{t_i} \) are the values of the corresponding returns.

See the next subsection for more information on the notation. The scatterplot of \( e_T(\eta) \) against \( \eta \) is called the mean excess plot, which should be linear in \( \eta \) for \( \eta > \eta_o \) under the GPD. The plot is also called mean residual life plot. Figure 6b shows the mean excess plot of the daily negative IBM log returns. It shows that, among others, a threshold of about 3% is reasonable for the negative return series. In the evir package of R and S-Plus, the command for mean excess plot is meplot.
4.3 A New Approach to Modeling Extreme Values

Using the statistical result in Eq. (17) and considering jointly the exceedances and exceeding times, Smith (1989) proposes a two-dimensional Poisson process to model \((t_i, r_{t_i})\). This approach was used by Tsay (1999) to study VaR in risk management. We follow the same approach.

Assume that the baseline time interval is \(D\), which is typically a year. In the United States, \(D = 252\) is used as there are typically 252 trading days in a year. Let \(t\) be the time interval of the data points (e.g., daily) and denote the data span by \(t = 1, 2, \ldots, T\), where \(T\) is the total number of data points. For a given threshold \(\eta\), the exceeding times over the threshold are denoted by \(\{t_i, i = 1, \ldots, N_\eta\}\) and the observed log return at \(t_i\) is \(r_{t_i}\). Consequently, we focus on modeling \(\{(t_i, r_{t_i})\}\) for \(i = 1, \ldots, N_\eta\), where \(N_\eta\) depends on the threshold \(\eta\).

The new approach to applying the extreme value theory is to postulate that the exceeding times and the associated returns (i.e., \((t_i, r_{t_i})\)) jointly form a two-dimensional Poisson process with intensity measure given by

\[
\Lambda[(D_2, D_1) \times (r, \infty)] = \frac{D_2 - D_1}{D} S(r; \xi, \alpha, \beta),
\]

(20)

where

\[
S(r; \xi, \alpha, \beta) = \left[1 + \frac{\xi(r - \beta)}{\alpha}\right]^{-1/\xi} +
\]

0 \leq D_1 \leq D_2 \leq T, \ r > \eta, \ \alpha > 0, \ \beta, \ \text{and} \ \xi \ \text{are parameters, and the notation} \ [x]_+ \ \text{is defined as} \ [x]_+ = \max(x, 0). \ \text{This intensity measure says that the occurrence of exceeding the threshold is proportional to the length of the time interval} \ [D_1, D_2] \ \text{and the probability is governed by a survival function similar to the exponent of the CDF} \ F_*(r) \ \text{in Eq. (3)}. \ A \ \text{survival function of a random variable} \ X \ \text{is defined as} \ S(x) = \Pr(X > x) = 1 - \Pr(X \leq x) = 1 - \text{CDF}(x). \ \text{When} \ \xi = 0, \ \text{the intensity measure is taken as the limit of} \ \xi \rightarrow 0; \ \text{that is,}

\[
\Lambda[(D_2, D_1) \times (r, \infty)] = \frac{D_2 - D_1}{D} \exp \left[\frac{-(r - \beta)}{\alpha}\right].
\]

In Eq. (20), the length of time interval is measured with respect to the baseline interval \(D\).

The idea of using the intensity measure in Eq. (20) becomes clear when one considers its implied conditional probability of \(r = x + \eta\) given \(r > \eta\) over the time interval \([0, D]\), where \(x > 0,\)

\[
\frac{\Lambda[(0, D) \times (x + \eta, \infty)]}{\Lambda[(0, D) \times (\eta, \infty)]} = \left[\frac{1 + \xi(x + \eta - \beta)/\alpha}{1 + \xi(\eta - \beta)/\alpha}\right]^{-1/\xi} = \left[1 + \frac{\xi x}{\alpha + \xi(\eta - \beta)}\right]^{-1/\xi},
\]

which is precisely the survival function of the conditional distribution given in Eq. (17). This survival function is obtained from the extreme limiting distribution for maximum in Eq. (3). We use survival function here because it denotes the probability of exceedance.

The relationship between the limiting extreme value distribution in Eq. (3) and the intensity measure in Eq. (20) directly connects the new approach of extreme value theory to the traditional one.

Mathematically, the intensity measure in Eq. (20) can be written as an integral of an intensity function:

\[
\Lambda[(D_2, D_1) \times (r, \infty)] = \int_{D_1}^{D_2} \int_r^\infty \lambda(t, z; \xi, \alpha, \beta) \, dz \, dt,
\]

(22)
where the intensity function \( \lambda(t, z; \xi, \alpha, \beta) \) is defined as
\[
\lambda(t, z; \xi, \alpha, \beta) = \frac{1}{D} g(z; \xi, \alpha, \beta),
\]
(21)
where
\[
g(z; \xi, \alpha, \beta) = \begin{cases} 
\frac{1}{\alpha} \left[ 1 + \frac{\xi(z-\beta)}{\alpha} \right]^{-1(1+\xi)/\xi} & \text{if } \xi \neq 0, \\
\frac{1}{\alpha} \exp \left[ -\frac{(z-\beta)}{\alpha} \right] & \text{if } \xi = 0.
\end{cases}
\]
Using the results of a Poisson process, we can write down the likelihood function for the observed exceeding times and their corresponding returns \( \{(t_i, r_{ti})\} \) over the two-dimensional space \([0, T] \times (\eta, \infty)\) as
\[
L(\xi, \alpha, \beta) = \left( \prod_{i=1}^{N_\eta} \frac{1}{D} g(r_{ti}; \xi, \alpha, \beta) \right) \times \exp \left[ -\frac{T}{D} S(\eta; \xi, \alpha, \beta) \right].
\]
(22)
The parameters \( \xi, \alpha, \) and \( \beta \) can then be estimated by maximizing the logarithm of this likelihood function. Since the scale parameter \( \alpha \) is non-negative, we use \( \ln(\alpha) \) in the estimation.

**Example 7.7.** Consider again the daily log returns of IBM stock from July 3, 1962 to December 31, 1998. There are 9190 daily returns. Table 3 gives some estimation results of the parameters \( \xi, \alpha, \) and \( \beta \) for three choices of the threshold when the negative series \( \{-r_{ti}\} \) is used. As mentioned before, we use the negative series \( \{-r_{ti}\} \), instead of \( \{r_{ti}\} \), because we focus on holding a long financial position. The table also shows the number of exceeding times for a given threshold. It is seen that the chance of dropping 2.5% or more in a day for IBM stock occurred with probability \( 310/9190 \approx 3.4\% \). Because the sample mean of IBM stock returns is not zero, we also consider the case when the sample mean is removed from the original daily log returns. From the table, removing the sample mean has little impact on the parameter estimates. These parameter estimates are used next to calculate VaR, keeping in mind that in a real application one needs to check carefully the adequacy of a fitted Poisson model. We discuss methods of model checking in the next subsection.

### 4.4 VaR Calculation Based on the New Approach

As shown in Eq. (17), the two-dimensional Poisson process model used, which employs the intensity measure in Eq. (20), has the same parameters as those of the extreme value distribution in Eq. (3). Therefore, one can use the same formula as that of Eq. (15) to calculate VaR of the new approach. More specifically, for a given upper tail probability \( p \), the \((1-p)\)th quantile of the log return \( r_t \) is
\[
\text{VaR} = \begin{cases} 
\beta - \frac{\alpha}{\xi} \left( 1 - \left[ -D \ln(1-p) \right]^{-\xi} \right) & \text{if } \xi \neq 0, \\
\beta - \alpha \ln[-D \ln(1-p)] & \text{if } \xi = 0,
\end{cases}
\]
(23)
where \( D \) is the baseline time interval used in estimation. In the United States, one typically uses \( D = 252 \), which is approximately the number of trading days in a year.

**Example 7.8.** Consider again the case of holding a long position of IBM stock valued at $10 million. We use the estimation results of Table 3 to calculate 1-day horizon VaR for the tail probabilities of 0.05 and 0.01.
Table 3: Estimation Results of a Two-Dimensional Homogeneous Poisson Model for the Daily Negative Log Returns of IBM Stock from July 3, 1962 to December 31, 1998. The baseline time interval is 252 (i.e., 1 year). The numbers in parentheses are standard errors, where “Thr.” and “Exc.” stand for threshold and the number of exceedings.

<table>
<thead>
<tr>
<th>Thr.</th>
<th>Exc.</th>
<th>Shape Par. $\xi$</th>
<th>Log(Scale) $\ln(\alpha)$</th>
<th>Location $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0%</td>
<td>175</td>
<td>0.30697(0.09015)</td>
<td>0.30699(0.12380)</td>
<td>4.69204(0.19058)</td>
</tr>
<tr>
<td>2.5%</td>
<td>310</td>
<td>0.26418(0.06501)</td>
<td>0.31529(0.11277)</td>
<td>4.74062(0.18041)</td>
</tr>
<tr>
<td>2.0%</td>
<td>554</td>
<td>0.18751(0.04394)</td>
<td>0.27655(0.09867)</td>
<td>4.81003(0.17209)</td>
</tr>
</tbody>
</table>

(b) Removing the sample mean

<table>
<thead>
<tr>
<th>Thr.</th>
<th>Exc.</th>
<th>Shape Par. $\xi$</th>
<th>Log(Scale) $\ln(\alpha)$</th>
<th>Location $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0%</td>
<td>184</td>
<td>0.30516(0.08824)</td>
<td>0.30807(0.12395)</td>
<td>4.73804(0.19151)</td>
</tr>
<tr>
<td>2.5%</td>
<td>334</td>
<td>0.28179(0.06737)</td>
<td>0.31968(0.12065)</td>
<td>4.76808(0.18533)</td>
</tr>
<tr>
<td>2.0%</td>
<td>590</td>
<td>0.19260(0.04357)</td>
<td>0.27917(0.09913)</td>
<td>4.84859(0.17255)</td>
</tr>
</tbody>
</table>

- Case I: Use the original daily log returns. The three choices of threshold $\eta$ result in the following VaR values:
  1. $\eta = 3.0\%$: VaR(5\%) = $228,239, VaR(1\%) = $359,303.
  2. $\eta = 2.5\%$: VaR(5\%) = $219,106, VaR(1\%) = $361,119.
  3. $\eta = 2.0\%$: VaR(5\%) = $212,981, VaR(1\%) = $368,552.

- Case II: The sample mean of the daily log returns is removed. The three choices of threshold $\eta$ result in the following VaR values:
  1. $\eta = 3.0\%$: VaR(5\%) = $232,094, VaR(1\%) = $363,697.
  2. $\eta = 2.5\%$: VaR(5\%) = $225,782, VaR(1\%) = $364,254.
  3. $\eta = 2.0\%$: VaR(5\%) = $217,740, VaR(1\%) = $372,372.

As expected, removing the sample mean, which is positive, slightly increases the VaR. However, the VaR is rather stable among the three threshold values used. In practice, we recommend that one removes the sample mean first before applying this new approach to VaR calculation.

**Discussion.** Compared with the VaR of Example 7.6 that uses the traditional extreme value theory, the new approach provides a more stable VaR calculation. The traditional approach is rather sensitive to the choice of the subperiod length $n$.

### 4.5 An Alternative Parameterization

As mentioned before, for a given threshold $\eta$, the GPD can also be parameterized by the shape parameter $\xi$ and the scale parameter $\psi(\eta) = \alpha + \xi(\eta - \beta)$. This is the parameterization used in the *evir* package of R and S-Plus. Specifically, $(x, \beta)$ of R and S-Plus corresponds to $(\xi, \psi(\eta))$ of this chapter. The command for estimating a GPD model in R and S-Plus is `gpd`. The output
format for S-Plus is slightly different from that of R. For illustration, consider the daily negative IBM log return series from 1962 to 1998. The results of R are given below.

**R Demonstration**

Data are negative IBM log returns. Output edited.

```r
> library(evir)
> mgpd=gpd(nibm,threshold=0.025)
> names(mgpd)
 [1] "n" "data" "threshold" "p.less.thresh"
 [5] "n.exceed" "method" "par.ests" "par.ses"
 [9] "varcov" "information" "converged" "nllh.final"
> mgpd
$n
[1] 9190
$data
 [1] 0.03288483 0.02648772 0.02817316 0.03618692 ....
$threshold
[1] 0.025
$p.less.thresh %Percentage of data below the threshold.
[1] 0.9662677
$n.exceed % Number of exceedances
[1] 310
$method
[1] "ml"
$par.ests
   xi     beta
0.264184649 0.007786063
$par.ses
   xi     beta
0.0662137508 0.0006427826
$varcov
     [,1]      [,2]
[1,] 4.384261e-03 -2.461142e-05
[2,] -2.461142e-05 4.131694e-07
> par(mfcol=c(2,2)) %Plots for residual analysis
> plot(mgpd)
```

Make a plot selection (or 0 to exit):
1: plot: Excess Distribution
2: plot: Tail of Underlying Distribution
3: plot: Scatterplot of Residuals
4: plot: QQplot of Residuals
Selection:

Note that the results are very close to those in Table 3, where percentage log returns are used. The
estimates of $\xi$ and $\psi(\eta)$ are 0.26418 and $\alpha + \xi(\eta - \beta) = \exp(0.31529) + (0.26418) (2.5 - 4.7406) = 0.77873$, respectively, in Table 3. In terms of log returns, the estimate of $\psi(\eta)$ is 0.007787, which is the same as the R and S-Plus estimate.

Figure 7 shows the diagnostic plots for the GPD fit to the daily negative log returns of IBM stock. The QQ-plot (lower-right panel) and the tail probability estimate (in log scale and in the lower-left panel) show some minor deviation from a straight line, indicating further improvement is possible. From the conditional distributions in Eqs. (16) and (17) and the GPD in Eq. (18), we have

$$ \frac{F(y) - F(\eta)}{1 - F(\eta)} \approx G_{\eta, \psi(\eta)}(x), $$

where $y = x + \eta$ with $x > 0$. If we estimate the CDF $F(\eta)$ of the returns by the empirical CDF, then

$$ \hat{F}(\eta) = \frac{T - N_\eta}{T}, $$

where $N_\eta$ is the number of exceedances of the threshold $\eta$ and $T$ is the sample size. Consequently, by Eq. (18),

$$ F(y) = F(\eta) + G(x)[1 - F(\eta)] 
\approx 1 - \frac{N_\eta}{T} \left[ 1 + \frac{\xi(y - \eta)}{\psi(\eta)} \right]^{-1/\xi}. $$

This leads to an alternative estimate of the quantile of $F(y)$ for use in VaR calculation. Specifically, for a small upper tail probability $p$, let $q = 1 - p$. Then, by solving for $y$, we can estimate the $q$th quantile of $F(y)$, denoted by VaR$_q$, by

$$ \text{VaR}_q = \eta - \frac{\psi(\eta)}{\xi} \left\{ 1 - \left[ \frac{T}{N_\eta} (1 - q) \right]^{-1/\xi} \right\}, \quad (24) $$

26
where, as before, \( \eta \) is the threshold, \( T \) is the sample size, \( N_\eta \) is the number of exceedances, and \( \psi(\eta) \) and \( \xi \) are the scale and shape parameters of the GPD distribution. This method to VaR calculation is used in R and S-Plus.

As mentioned before, VaR is a point estimate of potential loss for a given tail probability under a loss function. The estimate necessarily contains an element of uncertainty. On the other hand, it is often hard to obtain a good estimate of the loss function itself. A compromise is to consider the expected loss when the tail event occurs. This leads to the development of expected shortfall (ES) associated with a given VaR. Specifically, ES is defined as the expected loss given that the VaR is exceeded. Specifically, for a given tail probability \( p \), let \( q = 1 - p \) and denote the value at risk by \( \text{VaR}_q \). Then, the expected shortfall is defined by

\[
ES_q = E(r | r > \text{VaR}_q) = \text{VaR}_q + E(r - \text{VaR}_q | r > \text{VaR}_q).
\]

Using properties of the GPD, it can be shown that

\[
E(r - \text{VaR}_q | r > \text{VaR}_q) = \frac{\psi(\eta) + \xi(\text{VaR}_q - \eta)}{1 - \xi},
\]

provided that \( 0 < \xi < 1 \). Consequently, we have

\[
ES_q = \frac{\text{VaR}_q}{1 - \xi} + \frac{\psi(\eta) - \xi \eta}{1 - \xi}.
\]

To illustrate the new method to VaR and ES calculations, we again use the daily negative log returns of IBM stock with threshold 2.5%. In the \texttt{evir} package of R and S-Plus, the command to compute VaR and ES via the peak over threshold method is \texttt{riskmeasures}:

```r
> riskmeasures(mgpd,c(0.95,0.99,0.999))
   p quantile  sfall
[1,] 0.950 0.02208959 0.03162619
[2,] 0.990 0.03616405 0.05075390
[3,] 0.999 0.07018944 0.09699565
```

From the output, the VaR values for the financial position of \$10 million dollars are \$220,889 and \$361,661, respectively, for tail probability of 0.05 and 0.01. These two values are rather close to those given in Example 7.8 that are based on the method of the previous subsection. The expected shortfalls for the financial position are \$316,272 and \$507,576, respectively, for tail probability of 0.05 and 0.01.

Finally, the command \texttt{pot} of the R package \texttt{evir} can also be used to perform the estimation of the peaks over threshold (POT) model. We demonstrate it below using the negative log returns of IBM stock. As expected, the results are very close to those obtained before.

**R Demonstration Using POT Command**

```r
> library(evir)
> m3=pot(nibm,0.025)
```
> m3
$n
[1] 9190
$period
[1] 1 9190
$data
 [1] 0.03288483 0.02648772 0.02817316 ..... 
$span
[1] 9189
$threshold
[1] 0.025
$p.less.thresh
[1] 0.9662677
$n.exceed
[1] 310
$par.est

          xi    sigma       mu      beta
0.264078835 0.003182365 0.007557534 0.007788551
$par.ses

          xi    sigma       mu
0.0229175739 0.0001808472 0.0007675515
$varcov

[,1]       [,2]       [,3]
[1,] 5.252152e-04 -2.873160e-06 -6.970497e-07
[2,] -2.873160e-06  3.270571e-08 -7.907532e-08
[3,] -6.970497e-07 -7.907532e-08  5.891353e-07
$intensity %intensity function of exceeding the threshold
[1] 0.03373599
> plot(m3)  % model checking
Make a plot selection (or 0 to exit):

1: plot: Point Process of Exceedances
2: plot: Scatterplot of Gaps
3: plot: Qplot of Gaps
4: plot: ACF of Gaps
5: plot: Scatterplot of Residuals
6: plot: Qplot of Residuals
7: plot: ACF of Residuals
8: plot: Go to GPD Plots
Selection:

> riskmeasures(m3,c(0.95,0.99,0.999))

       p    quantile    sfall
[1,] 0.950 0.02208860 0.03162728
[2,] 0.990 0.03616686 0.05075740
4.6 Use of Explanatory Variables

The two-dimensional Poisson process model discussed earlier is homogeneous because the three parameters $\xi, \alpha$, and $\beta$ are constant over time. In practice, such a model may not be adequate. Furthermore, some explanatory variables are often available that may influence the behavior of the log returns $r_t$. A nice feature of the new extreme value theory approach to VaR calculation is that it can easily take explanatory variables into consideration. We discuss such a framework in this subsection. In addition, we also discuss methods that can be used to check the adequacy of a fitted two-dimensional Poisson process model.

Suppose that $x_t = (x_{1t}, \ldots, x_{vt})'$ is a vector of $v$ explanatory variables that are available prior to time $t$. For asset returns, the volatility $\sigma^2_t$ of $r_t$ is an example of explanatory variables. Another example of explanatory variables in the U.S. equity markets is an indicator variable denoting the meetings of the Federal Open Market Committee. A simple way to make use of explanatory variables is to postulate that the three parameters $\xi, \alpha$, and $\beta$ are time-varying and are linear functions of the explanatory variables. Specifically, when explanatory variables $x_t$ are available, we assume that

$$
\begin{align*}
\xi_t &= \gamma_0 + \gamma_1 x_{1t} + \cdots + \gamma_v x_{vt} \equiv \gamma' x_t, \\
\ln(\alpha_t) &= \delta_0 + \delta_1 x_{1t} + \cdots + \delta_v x_{vt} \equiv \delta' x_t, \\
\beta_t &= \theta_0 + \theta_1 x_{1t} + \cdots + \theta_v x_{vt} \equiv \theta' x_t.
\end{align*}
$$

(26)

If $\gamma = 0$, then the shape parameter $\xi_t = \gamma_0$, which is time-invariant. Thus, testing the significance of $\gamma$ can provide information about the contribution of the explanatory variables to the shape parameter. Similar methods apply to the scale and location parameters. In Eq. (26), we use the same explanatory variables for all three parameters $\xi_t, \ln(\alpha_t)$, and $\beta_t$. In an application, different explanatory variables may be used for different parameters.

When the three parameters of the extreme value distribution are time-varying, we have an inhomogeneous Poisson process. The intensity measure becomes

$$
\Lambda[(D_1, D_2) \times (r, \infty)] = \frac{D_2 - D_1}{D} \left[ 1 + \frac{\xi_t(r - \beta_t)}{\alpha_t} \right]^{-1/\xi_t}, \quad r > \eta.
$$

(27)

The likelihood function of the exceeding times and returns $\{(t_i, r_{t_i})\}$ becomes

$$
L = \left( \prod_{i=1}^{N_\eta} \frac{1}{D} g(r_{t_i}; \xi_{t_i}, \alpha_{t_i}, \beta_{t_i}) \right) \times \exp \left[ -\frac{1}{D} \int_0^T S(\eta; \xi_t, \alpha_t, \beta_t) dt \right],
$$

which reduces to

$$
L = \left( \prod_{i=1}^{N_\eta} \frac{1}{D} g(r_{t_i}; \xi_{t_i}, \alpha_{t_i}, \beta_{t_i}) \right) \times \exp \left[ -\frac{1}{D} \sum_{t=1}^T S(\eta; \xi_t, \alpha_t, \beta_t) \right]
$$

(28)

if one assumes that the parameters $\xi_t$, $\alpha_t$, and $\beta_t$ are constant within each trading day, where $g(z; \xi_t, \alpha_t, \beta_t)$ and $S(\eta; \xi_t, \alpha_t, \beta_t)$ are given in Eqs. (21) and (20), respectively. For given observations
The set \( \{r_t, x_t | t = 1, \cdots, T \} \), the baseline time interval \( D \), and the threshold \( \eta \), the parameters in Eq. (26) can be estimated by maximizing the logarithm of the likelihood function in Eq. (28). Again we use \( \ln(\alpha_t) \) to satisfy the positive constraint of \( \alpha_t \).

**Remark.** The parameterization in Eq. (26) is similar to that of the volatility models of Chapter 2 in the sense that the three parameters are exact functions of the available information at time \( t \). Other functions can be used if necessary.

### 4.7 Model Checking

Checking an entertained two-dimensional Poisson process model for exceedance times and excesses involves examining three key features of the model. The first feature is to verify the adequacy of the exceedance rate, the second feature is to examine the distribution of exceedances, and the final feature is to check the independence assumption of the model. We discuss briefly some statistics that are useful for checking these three features. These statistics are based on some basic statistical theory concerning distributions and stochastic processes.

**Exceedance Rate**

A fundamental property of univariate Poisson processes is that the time durations between two consecutive events are independent and exponentially distributed. To exploit a similar property for checking a two-dimensional process model, Smith and Shively (1995) propose examining the time durations between consecutive exceedances. If the two-dimensional Poisson process model is appropriate for the exceedance times and excesses, the time duration between the \( i \)th and \( (i-1) \)th exceedances should follow an exponential distribution. More specifically, letting \( t_0 = 0 \), we expect that

\[
z_{t_i} = \int_{t_{i-1}}^{t_i} \frac{1}{D} g(\eta; \xi_s, \alpha_s, \beta_s) ds, \quad i = 1, 2, \ldots
\]

are independent and identically distributed (iid) as a standard exponential distribution. Because daily returns are discrete-time observations, we employ the time durations

\[
z_{t_i} = \frac{1}{D} \sum_{t=t_{i-1}+1}^{t_i} S(\eta; \xi_t, \alpha_t, \beta_t)
\]

and use the QQ-plot to check the validity of the iid standard exponential distribution. If the model is adequate, the QQ-plot should show a straight line through the origin with unit slope.

**Distribution of Excesses**

Under the two-dimensional Poisson process model considered, the conditional distribution of the excess \( x_t = r_t - \eta \) over the threshold \( \eta \) is a GPD with shape parameter \( \xi_t \) and scale parameter \( \psi_t = \alpha_t + \xi_t(\eta - \beta_t) \). Therefore, we can make use of the relationship between a standard exponential distribution and GPD, and define

\[
w_{t_i} = \begin{cases} 
\frac{1}{\xi_t} \ln \left( 1 + \frac{r_{t_i} - \eta}{\psi_{t_i}} \right) & \text{if } \xi_t \neq 0, \\
\frac{r_{t_i} - \eta}{\psi_{t_i}} & \text{if } \xi_t = 0.
\end{cases}
\]
If the model is adequate, \( \{w_t\} \) are independent and exponentially distributed with mean 1; see also Smith (1999). We can then apply the QQ-plot to check the validity of the GPD assumption for excesses.

**Independence**

A simple way to check the independence assumption, after adjusting for the effects of explanatory variables, is to examine the sample autocorrelation functions of \( z_t \) and \( w_t \). Under the independence assumption, we expect that both \( z_t \) and \( w_t \) have no serial correlations.

### 4.8 An Illustration

In this subsection, we apply a two-dimensional inhomogeneous Poisson process model to the daily log returns, in percentages, of IBM stock from July 3, 1962 to December 31, 1998. We focus on holding a long position of $10 million. The analysis enables us to compare the results with those obtained before by using other approaches to calculating VaR.

We begin by pointing out that the two-dimensional homogeneous model of Example 7.7 needs further refinements because the fitted model fails to pass the model checking statistics of the previous subsection. Figures 8a and 8b show the autocorrelation functions of the statistics \( z_t \) and \( w_t \), defined in Eqs. (29) and (30), of the homogeneous model when the threshold is \( \eta = 2.5\% \). The horizontal lines in the plots denote asymptotic limits of two standard errors. It is seen that both \( z_t \) and \( w_t \) series have some significant serial correlations. Figures 9a and 9b show the QQ-plots of \( z_t \) and \( w_t \) series. The straight line in each plot is the theoretical line, which passes through the origin and has a unit slope under the assumption of a standard exponential distribution. The QQ-plot of \( z_t \) shows some discrepancy.

To refine the model, we use the mean-corrected log return series

\[
 r^o_t = r_t - \bar{r}, \quad \bar{r} = \frac{1}{9190} \sum_{t=1}^{9190} r_t,
\]

where \( r_t \) is the daily log return in percentages, and employ the following explanatory variables:

1. \( x_{1t} \): an indicator variable for October, November, and December. That is, \( x_{1t} = 1 \) if \( t \) is in October, November, or December. This variable is chosen to take care of the fourth-quarter effect (or year-end effect), if any, on the daily IBM stock returns.

2. \( x_{2t} \): an indicator variable for the behavior of the previous trading day. Specifically, \( x_{2t} = 1 \) if and only if the log return \( r^o_{t-1} \leq -2.5\% \). Since we focus on holding a long position with threshold 2.5%, an exceedance occurs when the daily price drops over 2.5%. Therefore, \( x_{2t} \) is used to capture the possibility of panic selling when the price of IBM stock dropped 2.5% or more on the previous trading day.

3. \( x_{3t} \): a qualitative measurement of volatility, which is the number of days between \( t - 1 \) and \( t - 5 \) (inclusive) that has a log return with magnitude exceeding the threshold. In our case, \( x_{3t} \) is the number of \( r^o_{t-i} \) satisfying \( |r^o_{t-i}| \geq 2.5\% \) for \( i = 1, \ldots, 5 \).

4. \( x_{4t} \): an annual trend defined as \( x_{4t} = (\text{year of time } t - 1961)/38 \). This variable is used to detect any trend in the behavior of extreme returns of IBM stock.
Figure 8: Sample autocorrelation functions of the $z$ and $w$ measures for two-dimensional Poisson models. Parts (a) and (b) are for the homogeneous model and parts (c) and (d) are for the inhomogeneous model. The data are daily mean-corrected log returns, in percentages, of IBM stock from July 3, 1962 to December 31, 1998, and the threshold is 2.5%. A long financial position is used.

Figure 9: Quantile-to-quantile plot of the $z$ and $w$ measures for two-dimensional Poisson models. Parts (a) and (b) are for the homogeneous model and parts (c) and (d) are for the inhomogeneous model. The data are daily mean-corrected log returns, in percentages, of IBM stock from July 3, 1962 to December 31, 1998, and the threshold is 2.5%. A long financial position is used.
5. \(x_{5t}\): a volatility series based on a Gaussian GARCH(1,1) model for the mean-corrected series \(r_t^o\). Specifically, \(x_{5t} = \sigma_t\), where \(\sigma_t^2\) is the conditional variance of the GARCH(1,1) model

\[
r_t^o = a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1),
\]

\[
\sigma_t^2 = 0.04565 + 0.0807a_{t-1}^2 + 0.9031\sigma_{t-1}^2.
\]

These five explanatory variables are all available at time \(t - 1\). We use two volatility measures (\(x_{3t}\) and \(x_{5t}\)) to study the effect of market volatility on VaR. As shown in Example 7.3 by the fitted AR(2)–GARCH(1,1) model, the serial correlations in \(r_t\) are weak so that we do not entertain any ARMA model for the mean equation.

Using the prior five explanatory variables and deleting insignificant parameters, we obtain the estimation results shown in Table 4. Figures 8c and 8d and Figures 9c and 9d show the model checking statistics for the fitted two-dimensional inhomogeneous Poisson process model when the threshold is \(\eta = 2.5\%\). All autocorrelation functions of \(z_t\) and \(w_t\) are within the asymptotic two standard-error limits. The QQ-plots also show marked improvements as they indicate no model inadequacy. Based on these checking results, the inhomogeneous model seems adequate.

Consider the case of threshold 2.5%. The estimation results show the following:

1. All three parameters of the intensity function depend significantly on the annual time trend. In particular, the shape parameter has a negative annual trend, indicating that the log returns of IBM stock are moving farther away from normality as time passes. Both the location and scale parameters increase over time.

2. Indicators for the fourth quarter, \(x_{1t}\), and for panic selling, \(x_{2t}\), are not significant for all three parameters.

3. The location and shape parameters are positively affected by the volatility of the GARCH(1,1) model; see the coefficients of \(x_{5t}\). This is understandable because the variability of log returns increases when the volatility is high. Consequently, the dependence of log returns on the tail index is reduced.

4. The scale and shape parameters depend significantly on the qualitative measure of volatility. Signs of the estimates are also plausible.

The explanatory variables for December 31, 1998 assumed the values \(x_{3,9190} = 0\), \(x_{4,9190} = 0.9737\), and \(x_{5,9190} = 1.9766\). Using these values and the fitted model in Table 4, we obtain

\[
\xi_{9190} = 0.01195, \quad \ln(\alpha_{9190}) = 0.19331, \quad \beta_{9190} = 6.105.
\]

Assume that the tail probability is 0.05. The VaR quantile shown in Eq. (23) gives VaR = 3.03756\%. Consequently, for a long position of $10 million, we have

\[
\text{VaR} = 10,000,000 \times 0.0303756 = 303,756.
\]

If the tail probability is 0.01, the VaR is $497,425. The 5\% VaR is slightly larger than that of Example 7.3, which uses a Gaussian AR(2)–GARCH(1,1) model. The 1\% VaR is larger than that of Case 1 of Example 7.3. Again, as expected, the effect of extreme values (i.e., heavy tails) on VaR is more pronounced when the tail probability used is small.
Table 4: Estimation Results of a Two-Dimensional Inhomogeneous Poisson Process Model for Daily Log Returns, in Percentages, of IBM Stock from July 3, 1962 to December 31, 1998. Four explanatory variables defined in the text are used. The model is for holding a long position on IBM stock. The sample mean of the log returns is removed from the data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>constant</th>
<th>Coef. of $x_{3t}$</th>
<th>Coef. of $x_{4t}$</th>
<th>Coef. of $x_{5t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_t$</td>
<td>0.3202</td>
<td>1.4772</td>
<td>2.1991</td>
<td></td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.3387)</td>
<td>(0.3222)</td>
<td>(0.2450)</td>
<td></td>
</tr>
<tr>
<td>$\ln(\alpha_t)$</td>
<td>-0.8119</td>
<td>0.3305</td>
<td>1.0324</td>
<td></td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.1798)</td>
<td>(0.0826)</td>
<td>(0.2619)</td>
<td></td>
</tr>
<tr>
<td>$\xi_t$</td>
<td>0.1805</td>
<td>0.2118</td>
<td>0.3551</td>
<td>-0.2602</td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.1290)</td>
<td>(0.0580)</td>
<td>(0.1503)</td>
<td>(0.0461)</td>
</tr>
</tbody>
</table>

Threshold 2.5% with 334 Exceedances

| $\beta_t$      | 1.1569   |                  |                  | 2.1918           |
| (Std.err)       | (0.4082) |                  |                  | (0.2909)         |
| $\ln(\alpha_t)$| -0.0316  | 0.3336           |                  |                  |
| (Std.err)       | (0.1201) | (0.0861)         |                  |                  |
| $\xi_t$         | 0.6008   | 0.2480           |                  | -0.3175          |
| (Std.err)       | (0.1454) | (0.0731)         |                  | (0.0685)         |

Threshold 3.0% with 184 Exceedances

An advantage of using explanatory variables is that the parameters are adaptive to the change in market conditions. For example, the explanatory variables for December 30, 1998 assumed the values $x_{3,9189} = 1$, $x_{4,9189} = 0.9737$, and $x_{5,9189} = 1.8757$. In this case, we have

$$\xi_{9189} = 0.2500, \quad \ln(\alpha_{9189}) = 0.52385, \quad \beta_{9189} = 5.8834.$$ 

The 95% quantile (i.e., the tail probability is 5%) then becomes 2.69139%. Consequently, the VaR is

$$\text{VaR} = 10,000,000 \times 0.0269139 = 269,139.$$ 

If the tail probability is 0.01, then VaR becomes $448,323$. Based on this example, the homogeneous Poisson model shown in Example 7.8 seems to underestimate the VaR.

5 The Extremal Index

So far our discussions of extreme values are based on the assumption that the data are independently and identically distributed (iid) random variables. However, in reality extremal events tend to occur in clusters caused by serial dependence in the data. For instance, we often observe large returns (both positive and negative) of an asset after some news event. In this section we extend the theory and applications of extreme values to cases in which the data form a strictly stationary time series. The basic concept of the extension is extremal index which allows one to characterize the relationship between the dependence structure of the data and their extremal behavior. Our
function Let $x$ Embrechts et al. (1997).

Discussion will be brief. Interested readers are referred to Beirlant et al. (2004, Chapter 10) and

some suitably chosen normalizing constants $\alpha_n$ for the maximum of the data, i.e.,

$$x(n) = \max\{x_i\}.$$ We seek the limiting distribution of $(x(n) - \beta_n)/\alpha_n$ for

some suitably chosen normalizing constants $\alpha_n > 0$ and $\beta_n$. If $\{x_i\}$ were iid, Section 2 shows that

the only possible non-degenerate limits are the extreme value distributions. What is the limiting

distribution when $\{x_i\}$ are serially dependent?

To answer this question, we start with a heuristic argument. Suppose that the serial dependence of the stationary series $x_i$ decays quickly so that $x_i$ and $x_{i+\ell}$ are essentially independent when $\ell$ is

sufficiently large. In other words, assume that the long-range dependence of $x_i$ vanishes quickly.

Now divide the data into disjoint blocks of size $k$. Specifically, let $g = \lceil n/k \rceil$ be the largest integer

less than or equal to $n/k$. The $i$th block of the data is then $\{x_j| j = (i - 1) * k + 1, \ldots, i * k\}$, where

it is understood that the $(g + 1)$th block may contain less than $k$ observations. Let $x_{k,i}$ be the

maximum of the $i$th block, i.e. $x_{k,i} = \max\{x_j| j = (i - 1) * k + 1, \ldots, i * k\}$. The collection of block

maxima is $\{x_{k,i}| i = 1, \ldots, g + 1\}$. From the definitions, it is easy to see that

$$x(n) = \max_{i=1,\ldots,g+1} x_{k,i}.$$ (31)

That is, the sample maximum is also the maximum of the block maxima. If the block size $k$ is

sufficiently large and the block maximum $x_{k,i}$ does not occur near the end of the $i$th block, then

$x_{k,i}$ and $x_{i+1,k}$ are sufficiently far apart and essentially independent under the assumption of weak

long-range dependence in $\{x_i\}$. Consequently, $\{x_{k,i}| i = 1, \ldots, g + 1\}$ can be regarded as a sample

of iid random variables, and the limiting distribution of its maximum, which is $x(n)$, should be the

extreme value distribution. The prior discussion shows that, under some proper condition, the

limiting distribution of the maximum of a strictly stationary time series is also the extreme value

distribution.

The proper condition needed for the maximum $x(n)$ of a strictly stationary time series to have

the extreme value limiting distribution is obtained by Leadbetter (1974) and known as the $D(u_n)$

condition. Details are given in the next subsection. The prior heuristic argument also suggests that,

even though the limiting distribution of $x(n)$ is also the extreme value distribution, the parameters

associated with the limiting distribution, however, will not be the same as those when $\{x_i\}$ are

iid random samples, because the limiting distribution depends on the marginal distribution of the

underlying sequences. For the iid sequences, the marginal distribution is $F(x)$, but for a stationary

series the underlying sequences are the block maxima $x_{k,i}$ whose marginal distribution is not $F(x)$.

The marginal distribution of $x_{k,i}$ depends on $k$ and the strength of serial dependence in $\{x_i\}$.

5.1 The $D(u_n)$ condition

Consider the sample $x_1, x_2, \ldots, x_n$. To place limits on the long-range dependence of $\{x_i\}$, let $u_n$

be a sequence of thresholds increasing at a rate for which the expected number of exceedances of

$x_i$ over $u_n$ remains bounded. Mathematically, this says that $\lim sup n[1 - F(u_n)] < \infty$, where $F(.)$

is the marginal cumulative distribution function of $x_i$. For any positive integers $p$ and $q$, suppose

that $i_v (v = 1, \ldots, p)$ and $j_t (t = 1, \ldots, q)$ are arbitrary integers satisfying

$$1 \leq i_1 < i_2 < \cdots < i_p < j_1 < \cdots < j_q \leq n,$$
where $j_1 - i_p \geq \ell_n$, where $\ell_n$ is a function of the sample size $n$ such that $\ell_n/n \to 0$ as $n \to \infty$. Let $A_1 = \{i_1, i_2, \ldots, i_p\}$ and $A_2 = \{j_1, j_2, \ldots, j_q\}$ be two sets of time indices. From the prior condition, elements in $A_1$ and $A_2$ are separated by at least $\ell_n$ time periods. The condition $D(u_n)$ is satisfied if

$$|P(\max_{i \in A_1 \cup A_2} x_i \leq u_n) - P(\max_{i \in A_1} x_i \leq u_n) P(\max_{i \in A_2} x_i \leq u_n)| \leq \delta_n \ell_n,$$

(32)

where $\delta_n \ell_n \to 0$ as $n \to \infty$. This condition says that any two events of the form $\{\max_{i \in A_1} x_i \leq u_n\}$ and $\{\max_{i \in A_2} x_i \leq u_n\}$ can become asymptotically independent as the sample size $n$ increases when the index subsets $A_1$ and $A_2$ of $\{1, 2, \ldots, n\}$ are separated by a distance $\ell_n$ which satisfies $\ell_n/n \to 0$ as $n \to \infty$. The $D(u_n)$ condition looks complicated, but it is relatively weak. For instance, consider Gaussian sequences with autocorrelation $\rho_n$ for lag $n$. The $D(u_n)$ condition is satisfied if $\rho_n \ln(n) \to 0$ as $n \to \infty$; see Berman (1964).

**Leadbetter’s Theorem 1**

Suppose that $\{x_i| i = 1, \ldots, n\}$ is a strictly stationary time series for which there exist sequences of constants $\alpha_n > 0$ and $\beta_n$ and a non-degenerate distribution function $F_n(.)$ such that

$$P\left[\frac{x(n) - \beta_n}{\alpha_n} \leq x\right] \to_d F_n(x), \quad n \to \infty,$$

where $\to_d$ denotes convergence in distribution. If $D(u_n)$ holds with $u_n = \alpha_n x + \beta_n$ for each $x$ such that $F_n(x) > 0$, then $F_n(x)$ is an extreme value distribution function. □

The prior theorem shows that the possible limiting distributions for the maxima of strictly stationary time series satisfying the $D(u_n)$ condition are also the extreme value distributions. As noted before, the dependence can affect the limiting distribution, however. The effect of the dependence appears in the marginal distribution of the block maxima $x_{k,i}$. To state the effect more precisely, let $\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\}$ be a sequence of iid random variables such that the marginal distribution of $\tilde{x}_i$ is the same as that of the stationary time series $x_i$. Let $\tilde{x}(n)$ be the maximum of $\{\tilde{x}_i\}$. Leadbetter (1983) establishes the following result.

**Leadbetter’s Theorem 2**

If there exist sequences of constants $\alpha_n > 0$ and $\beta_n$ and a non-degenerate distribution function $\tilde{F}_n(x)$ such that

$$P\left[\frac{\tilde{x}(n) - \beta_n}{\alpha_n} \leq x\right] \to_d \tilde{F}_n(x), \quad n \to \infty,$$

if the condition $D(u_n)$ holds with $u_n = \alpha_n x + \beta_n$ for each $x$ such that $\tilde{F}_n(x) > 0$, and if $p[(x(n) - \beta_n)/\alpha_n \leq x]$ converges for some $x$, then

$$P\left[\frac{x(n) - \beta_n}{\alpha_n} \leq x\right] \to_d F_n(x) = \tilde{F}_n^{\theta}(x), \quad n \to \infty,$$

for some constant $\theta \in (0, 1]$. □

The constant $\theta$ is called the extremal index. It plays an important role in determining the limiting distribution $F_n(x)$ for the maximum of a strictly stationary time series. To see this, we provide some simple derivations for the case of $\xi \neq 0$. From the result of Eq. (3), $\tilde{F}_n(x)$ is the generalized...
extreme value distribution and assumes the form

\[ \tilde{F}_*(x) = \exp \left[ - \left(1 + \xi \frac{x - \beta}{\alpha} \right)^{-1/\xi} \right], \]

where \( \xi \neq 0 \) and \( 1 + \xi (x - \beta)/\alpha > 0 \). In other words, we assume that for the iid sequence \( \{\tilde{x}_i\} \), the limiting extreme distribution of \( \tilde{x}_{(n)} \) has parameters \( \xi, \beta \) and \( \alpha \). Based on Theorem 2 of Leadbetter, we have

\[
F_*(x) = \tilde{F}_*(x) = \exp \left[ - \theta \left(1 + \xi \frac{x - \beta}{\alpha} \right)^{-1/\xi} \right] = \exp \left[ - \left( \frac{\alpha / \xi + x - \beta}{\alpha \theta \xi} \right)^{-1/\xi} \right]
\]

\[
= \exp \left[ - \left(1 + \xi \frac{x - \beta + \alpha / \xi - \alpha \theta \xi / \xi}{\alpha \theta \xi} \right)^{-1/\xi} \right]
\]

\[
= \exp \left[ - \left(1 + \xi \frac{x - \beta - \alpha (1 - \theta \xi)}{\alpha \theta \xi} \right)^{-1/\xi} \right]
\]

\[
= \exp \left[ - \left(1 + \xi \frac{x - \beta_*}{\alpha_*} \right)^{-1/\xi} \right], \tag{33}
\]

where \( \xi_* = \xi, \alpha_* = \alpha \theta \xi \), and \( \beta_* = \beta - \alpha (1 - \theta \xi) / \xi \). Therefore, for a stationary time series \( \{x_i\} \) satisfying the \( D(u_n) \) condition, the limiting distribution of the sample maximum is the generalized extreme value distribution with the shape parameter \( \xi \) which is the same as that of the iid sequences. On the other hand, the location and scale parameters are affected by the extremal index \( \theta \). Specifically, \( \alpha_* = \alpha \theta \xi \) and \( \beta_* = \beta - \alpha (1 - \theta \xi) / \xi \). Results for the case of \( \xi = 0 \) can be derived via the same approach and we have \( \alpha_* = \alpha \) and \( \beta_* = \beta + \alpha \ln(\theta) \).

A formal definition of the extremal index is as follows: Let \( \{x_i\} \) be a strictly stationary time series with marginal cumulative distribution function \( F(x) \) and \( \theta \) a non-negative number. Assume that for every \( \tau > 0 \) there exists a sequence of thresholds \( u_n \) such that

\[
\lim_{n \to \infty} n [1 - F(u_n)] = \tau, \tag{34}
\]

\[
\lim_{n \to \infty} P(x_{(n)} \leq u_n) = \exp(-\theta \tau). \tag{35}
\]

Then \( \theta \) is called the extremal index of the time series \( \{x_i\} \). See Embrechts et al. (1997). Note that, for the corresponding iid sequence \( \{\tilde{x}_i\} \), under the assumption that Eq. (34) holds, we have

\[
\lim_{n \to \infty} P(\tilde{x}_{(n)} \leq u_n) = \lim_{n \to \infty} [F(u_n)]^n = \lim_{n \to \infty} [1 - \frac{1}{n}(1 - F(u_n))]^n \to \exp(-\tau),
\]

where we have used the property \( \lim_{n \to \infty} (1 - y/n)^n = \exp(-y) \). Thus, the definition also highlights the role played by the extremal index \( \theta \).
5.2 Estimation of the extremal index

There are several ways to estimate the extremal index $\theta$ of a strictly stationary time series $\{x_t\}$. Each estimation method is associated with an interpretation of the extremal index. In what follows, we discuss some of the estimation methods.

The Blocks Method

From the definition of the extremal index $\theta$, we have, for a large $n$, that

$$P(x_{(n)} \leq u_n) \approx P^\theta(\bar{x}_{(n)} \leq u_n) = [F(u_n)]^\theta,$$

provided that $n[1 - F(u_n)] \to \tau > 0$. Hence

$$\lim_{n \to \infty} \frac{\ln P(x_{(n)} \leq u_n)}{n \ln F(u_n)} = \theta. \quad (36)$$

This limiting relationship suggests a method to estimate $\theta$. The denominator can be estimated by the sample quantile, namely

$$\hat{F}(u_n) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq u_n) = 1 - \frac{1}{n} \sum_{i=1}^{n} I(x_i > u_n) = 1 - \frac{N(u_n)}{n},$$

where $I(C) = 1$ if the augment $C$ holds and $= 0$ otherwise, i.e., $I(C)$ is the indicator variable for the statement $C$, and $N(u_n)$ denotes the number of exceedances of the sample over the threshold $u_n$. The numerator $P(x_{(n)} \leq u_n)$ is harder to estimate. One possibility is to use the block maxima. Specifically, let $k = k(n)$ be a properly chosen block size which depends on the sample size $n$ and, as before, let $g = \lfloor n/k \rfloor$ be the integer part of $n/k$. For simplicity, assume that $n = gk$. The $i$th block consists of $\{x_j | j = (i - 1)k + 1, \ldots, i \cdot k\}$ and let $x_{k,i}$ be the maximum of the $i$th block. Using Eq. (31) and the approximate independence of block maxima, we have

$$P(x_{(n)} \leq u_n) = P(\max_{1 \leq i \leq g} x_{k,i} \leq u_n) \approx [P(x_{k,i} \leq u_n)]^g.$$

The probability $P(x_{k,i} \leq u_n)$ can be estimated from the block maxima, i.e.,

$$\hat{P}(x_{k,i} \leq u_n) = \frac{1}{g} \sum_{i=1}^{g} I(x_{k,i} \leq u_n) = 1 - \frac{1}{g} \sum_{i=1}^{g} I(x_{k,i} > u_n) = 1 - \frac{G(u_n)}{g},$$

where $G(u_n)$ is the number of blocks such that the block maximum exceeds the threshold $u_n$. Combining the estimators for numerator and denominator, we obtain

$$\hat{\theta}_b^{(1)} = \frac{g \ln(1 - G(u_n)/g)}{n \ln(1 - N(u_n)/n)} = \frac{1 \ln(1 - G(u_n)/g)}{k \ln(1 - N(u_n)/n)}, \quad (37)$$

where the subscript $b$ signifies the blocks method. Note that $N(u_n)$ is the number of exceedances of the sample $\{x_t\}$ over the threshold $u_n$ and $G(u_n)$ is the number of blocks with one or more exceedances. Using approximation based on Taylor expansion of $\ln(1 - x)$, we obtain a second estimator

$$\hat{\theta}_b^{(2)} = \frac{1}{k N(u_n)/n} = \frac{G(u_n)}{N(u_n)}.$$
Based on the results of Hsing et al. (1988), this estimator can also be interpreted as the reciprocal of the mean cluster size of the limiting compound Poisson process $N(u_n)$.

**The Runs Method**

O’Brien (1987) proved, under certain weak mixing condition, that

$$\lim_{n \to \infty} P(x_{(n)}^* \leq u_n|x_1 > u_n) = \theta,$$

where $x_{(n)}^* = \max_{2 \leq i \leq s} x_i$, where $s$ is a function of the sample size $n$ satisfying some growth conditions, including $s \to \infty$ and $s/n \to 0$ as $n \to \infty$. See Beirlant et al. (2004) and Embrechts et al. (1997) for details. This result has been used to construct an estimator of $\theta$ based on runs:

$$\hat{\theta}_r(3) = \frac{\sum_{i=1}^{n-k} I(A_{i,n})}{\sum_{i=1}^{n} I(x_i > u_n)} \frac{1}{N(u_n)},$$

where $N(u_n)$ is the number of exceedances of the sample $\{x_i\}$ over the threshold $u_n$, $k$ is a function of $n$, and $A_{i,n} = \{x_i > u_n, x_{i+1} \leq u_n, \ldots, x_{i+k} \leq u_n\}$ Note that $A_{i,n}$ denotes the event that an exceedance is followed by a run of $k$ observations below the threshold. Since $k/n \to 0$ as $n \to \infty$, we can write the runs estimator as

$$\hat{\theta}_r(3) \approx \frac{(n-k)^{-1} \sum_{i=1}^{n-k} I(A_{i,n})}{n^{-1} N(u_n)}.$$

Finally, other estimators of $\theta$ are available in the literature. See, for instance, the methods discussed in Beirlant et al. (2004). For demonstration, we consider, again, the negative daily log returns of IBM stock from July 3, 1962 to December 31, 1998. Figure 10 shows the estimates of the extremal index for various thresholds when the block size $k = 10$. We chose $k = 10$ because the daily log returns have weak serial dependence. The estimates are based on the blocks method, i.e., $\hat{\theta}_b^{(1)}$.

From the plot, I see that $\hat{\theta}_b^{(1)} \approx 0.82$ for threshold 0.025. Indeed, a simple direct calculation using $k = 10$ and threshold 0.025 gives $\hat{\theta}_b^{(1)} = 0.823$. The plot also shows that the estimate $\hat{\theta}_b^{(1)}$ of the extremal index might be sensitive to the choices of threshold and block size $k$.

### 5.3 Value at Risk for a stationary time series

The relationship between $F_*(x)$ of the maximum of a stationary time series and $\tilde{F}_*(x)$ of its iid counterpart established in Theorem 2 of Leadbetter can be used to calculate the VaR of a financial position when the associated log returns form a stationary time series. Specifically, from $P(x_{(n)} \leq u_n) \approx |F(x)|^{\alpha n}$, the $(1-p)$th quantile of $F(x)$ is the $(1-p)^{\alpha n}$th quantile of the limiting extreme value distribution of $x_{(n)}$. Consequently, the VaR of Eq. (15) based on the extreme value theory becomes

$$\text{VaR} = \begin{cases} 
\beta_n - \frac{\alpha n}{\xi_n} \left\{ 1 - [-n\theta \ln(1-p)]^{-\xi_n} \right\} & \text{if } \xi_n \neq 0 \\
\beta_n - \alpha_n \ln[-n\theta \ln(1-p)] & \text{if } \xi_n = 0
\end{cases},$$

(38)

where $n$ is the length of subperiod. From the formula, we risk underestimating the VaR if the extremal index is overlooked.
Figure 10: Estimates of the extremal index for the negative daily log returns of the IBM stock from July 3, 1962 to December 31, 1998. The block size is \( k = 10 \).

As an illustration, again consider the negative daily log returns of IBM stock from July 3, 1962 to December 31, 1998. Using \( \hat{\theta}_b^{(1)} = 0.823 \), the 1\% VaR for the long position of $10 million dollars on the stock for the next trading day becomes 3.2714 for the case of choosing \( n = 63 \) days in parameter estimation. As expected, this is higher than 3.0497 of Example 7.6 when the extremal index is neglected.

**R Demonstration**

```r
> library(evir)
> help(exindex)
> m1=exindex(nibm,10) % Estimate the extremal index of Figure 7.10.
> % VaR calculation.
> 2.583-(.945/.335)*(1-(-63*.823*log(.99))^-.335)
[1] 3.271388
```

**REFERENCES**


