CHAPTER 7

Extreme Values, Quantiles, and Value at Risk

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, for example, the volatility index (VIX) of the Chicago Board Options Exchange 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 chapter, we discuss various methods for calculating VaR and the statistical theories behind these methods. In particular, 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.

Other approaches to VaR calculation discussed in the chapter are RiskMetrics, econometric modeling using volatility models, and empirical quantile. We use daily log returns of IBM stock to illustrate the actual calculation of all the methods discussed. The results obtained can therefore be used to compare the performance of different methods. Figure 7.1 shows the time plot of daily log returns of IBM stock from July 3, 1962, to December 31, 1998, for 9190 observations.

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

Analysis of Financial Time Series, Third Edition, By Ruey S. Tsay

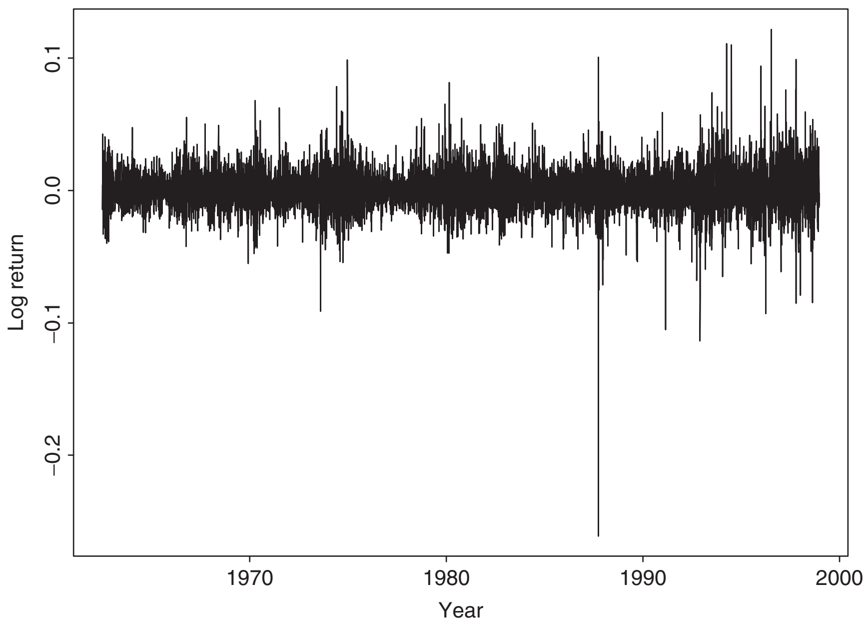

Figure 7.1 Time plot of daily log returns of IBM stock from July 3, 1962, to December 31, 1998.

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 chapter.

7.1 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 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 ℓ 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 ℓ with tail probability p as

$$p = \Pr[L(\ell) \geq \text{VaR}] = 1 - \Pr[L(\ell) < \text{VaR}]. \tag{7.1}$$

From the definition, the probability that the position holder would encounter a loss greater than or equal to VaR over the time horizon ℓ 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 ℓ 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

$$x_q = \inf\{x | F_\ell(x) \geq q\}$$

is called the qth 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. (7.1) 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 pth 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 ℓ. 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 ℓ. 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. (7.1) 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. (7.1) 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, VaR = Value × VaR(of log returns). If necessary, one can also use the approximation VaR = Value × [exp(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 just a quantile of the loss function. It does not fully describe the upper tail behavior of the loss function. In practice, two assets may have the same VaR yet encounter different losses when the VaR is exceeded. Furthermore, the VaR does not satisfy the sub-additivity property, which states that a risk measure for two portfolios after they have been merged should be no greater than the sum of their risk measures before they were merged. 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. The expected shortfall is also known as the *conditional* value at risk (CVaR). □

7.2 RISKMETRICS

J. P. Morgan developed the RiskMetrics methodology to VaR calculation; see Longerstaey and More (1995). In its simple form, RiskMetrics assumes that the continuously compounded daily return of a portfolio follows a conditional normal distribution. Denote the daily log return by r_t and the information set available at time $t-1$ by F_{t-1}. RiskMetrics assumes that $r_t|F_{t-1} \sim N(\mu_t, \sigma_t^2)$, where μ_t is the conditional mean and σ_t^2 is the conditional variance of r_t. In addition, the method assumes that the two quantities evolve over time according to the simple model:

$$\mu_t = 0, \qquad \sigma_t^2 = \alpha\sigma_{t-1}^2 + (1-\alpha)r_{t-1}^2, \qquad 1 > \alpha > 0. \tag{7.2}$$

Therefore, the method assumes that the logarithm of the daily price, $p_t = \ln(P_t)$, of the portfolio satisfies the difference equation $p_t - p_{t-1} = a_t$, where $a_t = \sigma_t\epsilon_t$ is an IGARCH(1,1) process without drift. The value of α is often in the interval (0.9, 1) with a typical value of 0.94.

A nice property of such a special random-walk IGARCH model is that the conditional distribution of a multiperiod return is easily available. Specifically, for a k-period horizon, the log return from time $t+1$ to time $t+k$ (inclusive) is $r_t[k] = r_{t+1} + \cdots + r_{t+k-1} + r_{t+k}$. We use the square bracket $[k]$ to denote a k-horizon return. Under the special IGARCH(1,1) model in Eq. (7.2), the conditional distribution $r_t[k]|F_t$ is normal with mean zero and variance $\sigma_t^2[k]$, where $\sigma_t^2[k]$ can be computed using the forecasting method discussed in Chapter 3. Using the independence assumption of ϵ_t and model (7.2), we have

$$\sigma_t^2[k] = \mathrm{Var}(r_t[k]|F_t) = \sum_{i=1}^{k} \mathrm{Var}(a_{t+i}|F_t),$$

where $\mathrm{Var}(a_{t+i}|F_t) = E(\sigma_{t+i}^2|F_t)$ can be obtained recursively. Using $r_{t-1} = a_{t-1} = \sigma_{t-1}\epsilon_{t-1}$, we can rewrite the volatility equation of the IGARCH(1,1) model in Eq. (7.2) as

$$\sigma_t^2 = \sigma_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2(\epsilon_{t-1}^2 - 1) \quad \text{for all } t.$$

In particular, we have

$$\sigma_{t+i}^2 = \sigma_{t+i-1}^2 + (1-\alpha)\sigma_{t+i-1}^2(\epsilon_{t+i-1}^2 - 1) \quad \text{for} \quad i = 2, \ldots, k.$$

Since $E(\epsilon_{t+i-1}^2 - 1|F_t) = 0$ for $i \geq 2$, the prior equation shows that

$$E(\sigma_{t+i}^2|F_t) = E(\sigma_{t+i-1}^2|F_t) \quad \text{for} \quad i = 2, \ldots, k. \tag{7.3}$$

For the 1-step-ahead volatility forecast, Eq. (7.2) shows that $\sigma_{t+1}^2 = \alpha\sigma_t^2 + (1-\alpha)r_t^2$. Therefore, Eq. (7.3) shows that $\mathrm{Var}(r_{t+i}|F_t) = \sigma_{t+1}^2$ for $i \geq 1$ and, hence, $\sigma_t^2[k] = k\sigma_{t+1}^2$. The results show that $r_t[k]|F_t \sim N(0, k\sigma_{t+1}^2)$. Consequently, under the special IGARCH(1,1) model in Eq. (7.2) the conditional variance of $r_t[k]$ is proportional to the time horizon k. The conditional standard deviation of a k-period horizon log return is then $\sqrt{k}\sigma_{t+1}$, which is $\sqrt{k}$ times σ_{t+1}.

Given a tail probability, RiskMetrics uses the result $r_t[k]|F_t \sim N(0, k\sigma_{t+1}^2)$ to calculate VaR for the log return. If the tail probability is set to 5%, then $\mathrm{VaR} = 1.65\sigma_{t+1}$ for the next trading day. This is the upper 5% quantile (or the 95th percentile) of a normal distribution with mean zero and standard deviation σ_{t+1}. For the next k trading days, $\mathrm{VaR}[k] = 1.65\sqrt{k}\sigma_{t+1}$, which is the 95th percentile of $N(0, k\sigma_{t+1}^2)$. Similarly, if the tail probability is 1%, then VaR $= 2.326\sigma_{t+1}$ for the next trading day and VaR$[k] = 2.326\sqrt{k}\sigma_{t+1}$ for the next k trading days.

Consider the case of 1% tail probability. The VaR for the portfolio under RiskMetrics is then

$$\text{VaR} = \text{Amount of position} \times 2.326\sigma_{t+1},$$

for the next trading day and that of a k-day horizon is

$$\text{VaR}(k) = \text{Amount of position} \times 2.326\sqrt{k}\sigma_{t+1},$$

where the argument (k) of VaR is used to denote the time horizon and the portfolio value is measured in dollars. Consequently, under RiskMetrics, we have

$$\text{VaR}(k) = \sqrt{k} \times \text{VaR}.$$

This is referred to as the *square root of time rule* in VaR calculation under RiskMetrics.

If the log returns are in percentages, then the 1% VaR for the next trading day is VaR $=$ Amount of position $\times\ 2.326\sigma_{t+1}/100$, where σ_{t+1} is the volatility of the percentage log returns.

Note that because RiskMetrics assumes log returns are normally distributed with mean zero, the loss function is symmetric and VaR are the same for long and short financial positions.

Example 7.1. The sample standard deviation of the continuously compounded daily return of the German mark/U.S. dollar exchange rate was about 0.53% in June 1997. Suppose that an investor was long in \$10 million worth of mark/dollar exchange rate contract. Then the 5% VaR for a 1-day horizon of the investor is

$$\$10{,}000{,}000 \times (1.65 \times 0.0053) = \$87{,}450.$$

The corresponding VaR for 10-day horizon is

$$\$10{,}000{,}000 \times (\sqrt{10} \times 1.65 \times 0.0053) \approx \$276{,}541.$$

Example 7.2. Consider the daily IBM log returns of Figure 7.1. As mentioned in Chapter 1, the sample mean of the returns is significantly different from zero. However, for demonstration of VaR calculation using RiskMetrics, we assume in this example that the conditional mean is zero and the volatility of the returns follows an IGARCH(1,1) model without drift. The fitted model is

$$r_t = a_t, \qquad a_t = \sigma_t\epsilon_t, \qquad \sigma_t^2 = 0.9396\sigma_{t-1}^2 + (1 - 0.9396)a_{t-1}^2, \tag{7.4}$$

where $\{\epsilon_t\}$ is a standard Gaussian white noise series. As expected, this model is rejected by the Q statistics. For instance, we have a highly significant statistic $Q(10) = 56.19$ for the squared standardized residuals.

From the data and the fitted model, we have $r_{9190} = -0.0128$ and $\hat{\sigma}^2_{9190} = 0.0003472$. Therefore, the 1-step-ahead volatility forecast is $\hat{\sigma}^2_{9190}(1) = 0.000336$. The 95% quantile of the conditional distribution $r_{9191}|F_{9190}$ is $1.65 \times \sqrt{0.000336} = 0.03025$. Consequently, the 1-day horizon 5% VaR of a long position of \$10 millions is

$$\text{VaR} = \$10{,}000{,}000 \times 0.03025 = \$302{,}500.$$

The 99% quantile is $2.326 \times \sqrt{0.000336} = 0.04265$, and the corresponding 1% VaR for the same long position is \$426,500.

Remark. To implement RiskMetrics in S-Plus, one can use `ewma1` (exponentially weighted moving average of order 1) under the `mgarch` (multivariate GARCH) command to obtain the estimate of $1 - \alpha$. Then, use the command `predict` to obtain volatility forecasts. For the IBM data used, the estimate of α is $1 - 0.036 = 0.964$ and the 1-step-ahead volatility forecast is $\hat{\sigma}_{9190}(1) = 0.01888$. Please see the demonstration below. This leads to VaR = \$10,000,000 × (1.65 × 0.01888) = \$311,520 and VaR = \$439,187 for $p = 0.05$ and 0.01, respectively. These two values are slightly higher than those of Example 7.2, which are based on estimates of the RATS package. □

S-Plus Demonstration

The following output has been simplified:

```
> ibm.risk=mgarch(ibm~-1, ~ewma1)
> ibm.risk
ALPHA 0.036
> predict(ibm.risk,2)
$sigma.pred 0.01888
```

7.2.1 Discussion

An advantage of RiskMetrics is simplicity. It is easy to understand and apply. Another advantage is that it makes risk more transparent in the financial markets. However, as security returns tend to have heavy tails (or fat tails), the normality assumption used often results in underestimation of VaR. Other approaches to VaR calculation avoid making such an assumption.

The square root of time rule is a consequence of the special model used by RiskMetrics. If either the zero mean assumption or the special IGARCH(1,1) model assumption of the log returns fails, then the rule is invalid. Consider the simple model

$$r_t = \mu + a_t, \qquad a_t = \sigma_t \epsilon_t, \qquad \mu \neq 0,$$
$$\sigma_t^2 = \alpha \sigma_{t-1}^2 + (1 - \alpha) a_{t-1}^2,$$

where $\{\epsilon_t\}$ is a standard Gaussian white noise series. The assumption that $\mu \neq 0$ holds for returns of many heavily traded stocks on the NYSE; see Chapter 1. For this simple model, the distribution of r_{t+1} given F_t is $N(\mu, \sigma_{t+1}^2)$. The 95% quantile used to calculate the 1-period horizon VaR becomes $\mu + 1.65\sigma_{t+1}$. For a k-period horizon, the distribution of $r_t[k]$ given F_t is $N(k\mu, k\sigma_{t+1}^2)$, where as before $r_t[k] = r_{t+1} + \cdots + r_{t+k}$. The 95% quantile used in the k-period horizon VaR calculation is $k\mu + 1.65\sqrt{k}\sigma_{t+1} = \sqrt{k}(\sqrt{k}\mu + 1.65\sigma_{t+1})$. Consequently, $\text{VaR}(k) \neq \sqrt{k} \times \text{VaR}$ when the mean return is not zero. It is also easy to show that the rule fails when the volatility model of the return is not an IGARCH(1,1) model without drift.

7.2.2 Multiple Positions

In some applications, an investor may hold multiple positions and needs to compute the overall VaR of the positions. RiskMetrics adopts a simple approach for doing such a calculation under the assumption that daily log returns of each position follow a random-walk IGARCH(1,1) model. The additional quantities needed are the cross-correlation coefficients between the returns. Consider the case of two positions. Let VaR_1 and VaR_2 be the VaR for the two positions and ρ_{12} be the cross-correlation coefficient between the two returns—that is, $\rho_{12} = \text{Cov}(r_{1t}, r_{2t})/[\text{Var}(r_{1t})\text{Var}(r_{2t})]^{0.5}$. Then the overall VaR of the investor is

$$\text{VaR} = \sqrt{\text{VaR}_1^2 + \text{VaR}_2^2 + 2\rho_{12}\text{VaR}_1\text{VaR}_2}.$$

The generalization of VaR to a position consisting of m instruments is straightforward as

$$\text{VaR} = \sqrt{\sum_{i=1}^{m} \text{VaR}_i^2 + 2\sum_{i<j}^{m} \rho_{ij}\text{VaR}_i\text{VaR}_j},$$

where ρ_{ij} is the cross-correlation coefficient between returns of the ith and jth instruments and VaR_i is the VaR of the ith instrument.

The prior formula is obtained using the assumption that the joint distribution of the log returns of assets involved in the portfolio is multivariate normal with mean zero and covariance matrix $\boldsymbol{\Sigma}_{t+1}$. Under such an assumption, the log return of the portfolio is normal with mean zero and finite variance; see Appendix B of Chapter 8 for properties of multivariate normal variables.

7.2.3 Expected Shortfall

Given a tail probability p, VaR is simply the $(1-p)$th quantile of the loss function. In practice, the actual loss, if it occurs, can be greater than VaR. In this sense, VaR may underestimate the actual loss. To have a better assessment of the potential loss, one can consider the expected value of the loss function if the VaR is exceeded. This consideration leads to the concept of *expected shortfall* (ES).

Under RiskMetrics, the loss function is normally distributed so that the conditional distribution of the loss function given that a VaR is exceeded is a truncated (from below) normal distribution. Properties such as mean and variance of a truncated normal distribution have been well-studied in the statistical literature. We can use the mean of the distribution to calculate expected shortfall. Specifically, consider the standard normal distribution $X \sim N(0, 1)$. For a given upper tail probability p, let $q = 1 - p$ and VaR_q be the associated VaR, that is, VaR_q is the qth quantile of X. Then the expectation of X given $X > \text{VaR}_q$ is $E(X|X > \text{VaR}_q) = f(\text{VaR}_q)/p$, where $f(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$ is the pdf of X. The expected shortfall for a log return r_t with conditional distribution $N(0, \sigma_t^2)$ is then

$$\text{ES}_q = \frac{f(\text{VaR}_q)}{p}\sigma_t \quad \text{or} \quad \text{ES}_{1-p} = \frac{f(\text{VaR}_{1-p})}{p}\sigma_t.$$

For example, if $p = 0.05$, then $\text{VaR}_{0.95} \approx 1.645$ and $f(\text{VaR}_q)/p = f(1.645)/0.05 = 2.0627$ so that the expected shortfall under RiskMetrics is $\text{ES}_{0.95} = 2.0627\sigma_t$. If $p = 0.01$, then $\text{ES}_{0.99} = 2.6652\sigma_t$.

7.3 ECONOMETRIC APPROACH TO VAR CALCULATION

A general approach to VaR calculation is to use the time series econometric models of Chapters 2–4. For a log return series, the time series models of Chapter 2 can be used to model the mean equation, and the conditional heteroscedastic models of Chapter 3 or 4 are used to handle the volatility. For simplicity, we use GARCH models in our discussion and refer to the approach as an *econometric approach* to VaR calculation. Other volatility models, including the nonlinear ones in Chapter 4, can also be used.

Consider the log return r_t of an asset. A general time series model for r_t can be written as

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}, \tag{7.5}$$

$$a_t = \sigma_t \epsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i a_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2. \tag{7.6}$$

Equations (7.5) and (7.6) are the mean and volatility equations for r_t. These two equations can be used to obtain 1-step-ahead forecasts of the conditional mean and conditional variance of r_t assuming that the parameters are known. Specifically, we have

$$\hat{r}_t(1) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t+1-i} - \sum_{j=1}^{q} \theta_j a_{t+1-j},$$

$$\hat{\sigma}_t^2(1) = \alpha_0 + \sum_{i=1}^{u} \alpha_i a_{t+1-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t+1-j}^2.$$

If one further assumes that ϵ_t is Gaussian, then the conditional distribution of r_{t+1} given the information available at time t is $N[\hat{r}_t(1), \hat{\sigma}_t^2(1)]$. Quantiles of this conditional distribution can easily be obtained for VaR calculation. For example, the 95% quantile is $\hat{r}_t(1) + 1.65\hat{\sigma}_t(1)$. If one assumes that ϵ_t is a standardized Student-t distribution with v degrees of freedom, then the quantile is $\hat{r}_t(1) + t_v^*(1-p)\hat{\sigma}_t(1)$, where $t_v^*(1-p)$ is the $(1-p)$th quantile of a standardized Student-t distribution with v degrees of freedom.

The relationship between quantiles of a Student-t distribution with v degrees of freedom, denoted by t_v, and those of its standardized distribution, denoted by t_v^*, is

$$p = \Pr(t_v \le q) = \Pr\left[\frac{t_v}{\sqrt{v/(v-2)}} \le \frac{q}{\sqrt{v/(v-2)}}\right] = \Pr\left[t_v^* \le \frac{q}{\sqrt{v/(v-2)}}\right],$$

where $v > 2$. That is, if q is the pth quantile of a Student-t distribution with v degrees of freedom, then $q/\sqrt{v/(v-2)}$ is the pth quantile of a standardized Student-t distribution with v degrees of freedom. Therefore, if ϵ_t of the GARCH model in Eq. (7.6) is a standardized Student-t distribution with v degrees of freedom and the upper tail probability is p, then the $(1-p)$th quantile used to calculate the 1-period horizon VaR at time index t is

$$\hat{r}_t(1) + \frac{t_v(1-p)\hat{\sigma}_t(1)}{\sqrt{v/(v-2)}},$$

where $t_v(1-p)$ is the $(1-p)$th quantile of a Student-t distribution with v degrees of freedom.

Example 7.3. Consider again the daily IBM log returns of Example 7.2. We use two volatility models to calculate VaR of 1-day horizon at $t = 9190$ for a long position of \$10 million. These econometric models are reasonable based on the modeling techniques of Chapters 2 and 3.

Because the position is long, we use $r_t = -r_t^c$, where r_t^c is the usual log return of IBM stock shown in Figure 7.1.

CASE 1. Assume that ϵ_t is standard normal. The fitted model is

$$r_t = -0.00066 - 0.0247r_{t-2} + a_t, \qquad a_t = \sigma_t\epsilon_t,$$
$$\sigma_t^2 = 0.00000389 + 0.0799a_{t-1}^2 + 0.9073\sigma_{t-1}^2.$$

From the data, we have $r_{9189} = 0.00201$, $r_{9190} = 0.0128$, and $\sigma_{9190}^2 = 0.00033455$. Consequently, the prior AR(2)–GARCH(1,1) model produces 1-step-ahead forecasts as

$$\hat{r}_{9190}(1) = -0.00071 \quad \text{and} \quad \hat{\sigma}_{9190}^2(1) = 0.0003211.$$

The 95% quantile is then

$$-0.00071 + 1.6449 \times \sqrt{0.0003211} = 0.02877.$$

The VaR for a long position of \$10 million with probability 0.05 is VaR $= \$10{,}000{,}000 \times 0.02877 = \$287{,}700$. The result shows that, with probability 95%, the potential loss of holding that position next day is \$287, 200 or less assuming that the AR(2)–GARCH(1,1) model holds. If the tail probability is 0.01, then the 99% quantile is

$$-0.00071 + 2.3262 \times \sqrt{0.0003211} = 0.0409738.$$

The VaR for the position becomes \$409, 738.

CASE 2. Assume that ϵ_t is a standardized Student-t distribution with 5 degrees of freedom. The fitted model is

$$r_t = -0.0003 - 0.0335 r_{t-2} + a_t, \qquad a_t = \sigma_t \epsilon_t,$$
$$\sigma_t^2 = 0.000003 + 0.0559 a_{t-1}^2 + 0.9350 \sigma_{t-1}^2.$$

From the data, we have $r_{9189} = 0.00201$, $r_{9190} = 0.0128$, and $\sigma_{9190}^2 = 0.000349$. Consequently, the prior Student-t AR(2)–GARCH(1,1) model produces 1-step-ahead forecasts

$$\hat{r}_{9190}(1) = -0.000367 \quad \text{and} \quad \hat{\sigma}_{9190}^2(1) = 0.0003386.$$

The 95% quantile of a Student-t distribution with 5 degrees of freedom is 2.015 and that of its standardized distribution is $2.015/\sqrt{5/3} = 1.5608$. Therefore, the 95% quantile of the conditional distribution of r_{9191} given F_{9190} is

$$-0.000367 + 1.5608\sqrt{0.0003386} = 0.028354.$$

The VaR for a long position of \$10 million is

$$\text{VaR} = \$10{,}000{,}000 \times 0.028352 = \$283{,}520,$$

which is essentially the same as that obtained under the normality assumption. The 99% quantile of the conditional distribution is

$$-0.000367 + (3.3649/\sqrt{5/3})\sqrt{0.0003386} = 0.0475943.$$

The corresponding VaR is \$475, 943. Comparing with that of Case 1, we see the heavy-tail effect of using a Student-t distribution with 5 degrees of freedom; it increases the VaR when the tail probability becomes smaller. In R and S-Plus, the quantile of a Student-t distribution with m degrees of freedom can be obtained by the command `qt(p,m)`, for example, `xp = qt(0.99,5.23)` for the 99th percentile of a Student-t distribution with 5.23 degrees of freedom.

7.3.1 Multiple Periods

Suppose that at time h we want to compute the k-horizon VaR of an asset whose log return is r_t. The variable of interest is the k-period log return at the forecast origin h (i.e., $r_h[k] = r_{h+1} + \cdots + r_{h+k}$). If the return r_t follows the time series model in Eqs. (7.5) and (7.6), then the conditional mean and variance of $r_h[k]$ given the information set F_h can be obtained by the forecasting methods discussed in Chapters 2 and 3.

Expected Return and Forecast Error

The conditional mean $E(r_h[k]|F_h)$ can be obtained by the forecasting method of ARMA models in Chapter 2. Specifically, we have

$$\hat{r}_h[k] = r_h(1) + \cdots + r_h(k),$$

where $r_h(\ell)$ is the ℓ-step-ahead forecast of the return at the forecast origin h. These forecasts can be computed recursively as discussed in Section 2.6.4. Using the MA representation

$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

of the ARMA model in Eq. (7.5), we can write the ℓ-step-ahead forecast error at the forecast origin h as

$$e_h(\ell) = r_{h+\ell} - r_h(\ell) = a_{h+\ell} + \psi_1 a_{h+\ell-1} + \cdots + \psi_{\ell-1} a_{h+1};$$

see Eq. (2.33) and the associated forecast error. The forecast error of the expected k-period return $\hat{r}_h[k]$ is the sum of 1-step to k-step forecast errors of r_t at the forecast origin h and can be written as

$$\begin{aligned} e_h[k] &= e_h(1) + e_h(2) + \cdots + e_h(k) \\ &= a_{h+1} + (a_{h+2} + \psi_1 a_{h+1}) + \cdots + \sum_{i=0}^{k-1} \psi_i a_{h+k-i} \\ &= a_{h+k} + (1 + \psi_1) a_{h+k-1} + \cdots + \left(\sum_{i=0}^{k-1} \psi_i \right) a_{h+1}, \end{aligned} \tag{7.7}$$

where $\psi_0 = 1$.

Expected Volatility

The volatility forecast of the k-period return at the forecast origin h is the conditional variance of $e_h[k]$ given F_h. Using the independent assumption of ϵ_{t+i} for

$i = 1, \ldots, k$, where $a_{t+i} = \sigma_{t+i}\epsilon_{t+i}$, we have

$$\begin{aligned} V_h(e_h[k]) &= V_h(a_{h+k}) + (1+\psi_1)^2 V_h(a_{h+k-1}) + \cdots + \left(\sum_{i=0}^{k-1}\psi_i\right)^2 V_h(a_{h+1}) \\ &= \sigma_h^2(k) + (1+\psi_1)^2\sigma_h^2(k-1) + \cdots + \left(\sum_{i=0}^{k-1}\psi_i\right)^2 \sigma_h^2(1), \end{aligned} \tag{7.8}$$

where $V_h(z)$ denotes the conditional variance of z given F_h and $\sigma_h^2(\ell)$ is the ℓ-step-ahead volatility forecast at the forecast origin h. If the volatility model is the GARCH model in Eq. (7.6), then these volatility forecasts can be obtained recursively by the methods discussed in Chapter 3.

As an illustration, consider the special time series model

$$\begin{aligned} r_t &= \mu + a_t, \qquad a_t = \sigma_t\epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1\sigma_{t-1}^2. \end{aligned}$$

Then we have $\psi_i = 0$ for all $i > 0$. The point forecast of the k-period return at the forecast origin h is $\hat{r}_h[k] = k\mu$ and the associated forecast error is

$$e_h[k] = a_{h+k} + a_{h+k-1} + \cdots + a_{h+1}.$$

Consequently, the volatility forecast for the k-period return at the forecast origin h is

$$\mathrm{Var}(e_h[k]|F_h) = \sum_{\ell=1}^{k}\sigma_h^2(\ell).$$

Using the forecasting method of GARCH(1,1) models in Section 3.5, we have

$$\begin{aligned} \sigma_h^2(1) &= \alpha_0 + \alpha_1 a_h^2 + \beta_1\sigma_h^2, \\ \sigma_h^2(\ell) &= \alpha_0 + (\alpha_1+\beta_1)\sigma_h^2(\ell-1), \qquad \ell = 2, \ldots, k. \end{aligned} \tag{7.9}$$

Using Eq. (7.9), we obtain that for the case of $\psi_i = 0$ for $i > 0$,

$$\mathrm{Var}(e_h[k]|F_h) = \frac{\alpha_0}{1-\phi}\left(k - \frac{1-\phi^k}{1-\phi}\right) + \frac{1-\phi^k}{1-\phi}\sigma_h^2(1), \tag{7.10}$$

where $\phi = \alpha_1 + \beta_1 < 1$. If $\psi_i \neq 0$ for some $i > 0$, then one should use the general formula of $\mathrm{Var}(e_h[k]|F_h)$ in Eq. (7.8). If ϵ_t is Gaussian, then the conditional distribution of $r_h[k]$ given F_h is normal with mean $k\mu$ and variance $\mathrm{Var}(e_h[k]|F_h)$. The quantiles needed in VaR calculations are readily available. If the conditional distribution of a_t is not Gaussian (e.g., a Student-t or generalized error distribution), simulation can be used to obtain the multiperiod VaR.

Example 7.3 (Continued). Consider the Gaussian AR(2)–GARCH(1,1) model of Example 7.3 for the daily log returns of IBM stock. Suppose that we are interested in the VaR of a 15-day horizon starting at the forecast origin 9190 (i.e., December 31, 1998). We can use the fitted model to compute the conditional mean and variance for the 15-day log return via $r_{9190}[15] = \sum_{i=1}^{15} r_{9190+i}$ given F_{9190}. The conditional mean is -0.00998 and the conditional variance is 0.0047948, which is obtained by the recursion in Eq. (7.9). The 95% quantile of the conditional distribution is then $-0.00998 + 1.6449\sqrt{0.0047948} = 0.1039191$. Consequently, the 5% 15-day horizon VaR for a long position of \$10 million is VaR $= \$10{,}000{,}000 \times 0.1039191 = \$1{,}039{,}191$. This amount is smaller than $\$287{,}700 \times \sqrt{15} = \$1{,}114{,}257$. This example further demonstrates that the square root of time rule used by RiskMetrics holds only for the special white noise IGARCH(1,1) model used. When the conditional mean is not zero, proper steps must be taken to compute the k-horizon VaR.

7.3.2 Expected Shortfall under Conditional Normality

We can use the result of Section 7.2.3 to calculate the ES when the conditional distribution of the log return is $N(\mu_t, \sigma_t^2)$. The result is

$$\mathrm{ES}_q = \mu_t + \frac{f(x_q)}{p}\sigma_t,$$

where $q = 1 - p$ and x_q is the qth quantile of the standard normal distribution. For instance, if $p = 0.01$, then $\mathrm{ES}_{0.99} = \mu_t + 2.6652\sigma_t$.

7.4 QUANTILE ESTIMATION

Quantile estimation provides a nonparametric approach to VaR calculation. It makes no specific distributional assumption on the return of a portfolio except that the distribution continues to hold within the prediction period. There are two types of quantile methods. The first method is to use empirical quantile directly, and the second method uses quantile regression.

7.4.1 Quantile and Order Statistics

Assuming that the distribution of return in the prediction period is the same as that in the sample period, one can use the empirical quantile of the return r_t to calculate VaR. Let $r_1, \ldots, r_n$ be the returns of a portfolio in the sample period. The *order* *statistics* of the sample are these values arranged in increasing order. We use the notation

$$r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(n)}$$

to denote the arrangement and refer to $r_{(i)}$ as the ith order statistic of the sample. In particular, $r_{(1)}$ is the sample minimum and $r_{(n)}$ the sample maximum.

Assume that the returns are independent and identically distributed random variables that have a continuous distribution with probability density function (pdf) $f(x)$ and CDF $F(x)$. Then we have the following asymptotic result from the statistical literature [e.g., Cox and Hinkley (1974), Appendix 2], for the order statistic $r_{(\ell)}$, where $\ell = np$ with $0 < p < 1$.

Result. Let x_p be the pth quantile of $F(x)$, that is, $x_p = F^{-1}(p)$. Assume that the pdf $f(x)$ is not zero at x_p [i.e., $f(x_p) \neq 0$]. Then the order statistic $r_{(\ell)}$ is asymptotically normal with mean x_p and variance $p(1-p)/[nf^2(x_p)]$. That is,

$$r_{(\ell)} \sim N\left\{x_p, \frac{p(1-p)}{n[f(x_p)]^2}\right\}, \qquad \ell = np. \tag{7.11}$$

Based on the prior result, one can use $r_{(\ell)}$ to estimate the quantile x_p, where $\ell = np$. In practice, the probability of interest p may not satisfy that np is a positive integer. In this case, one can use simple interpolation to obtain quantile estimates. More specifically, for noninteger np, let ℓ_1 and ℓ_2 be the two neighboring positive integers such that $\ell_1 < np < \ell_2$. Define $p_i = \ell_i/n$. The previous result shows that $r_{(\ell_i)}$ is a consistent estimate of the quantile x_{p_i}. From the definition, $p_1 < p < p_2$. Therefore, the quantile x_p can be estimated by

$$\hat{x}_p = \frac{p_2 - p}{p_2 - p_1} r_{(\ell_1)} + \frac{p - p_1}{p_2 - p_1} r_{(\ell_2)}. \tag{7.12}$$

In practice, sample quantiles can easily be obtained from most statistical packages, including R and S-Plus. A demonstration is given after the examples.

Example 7.4. Consider the daily log returns of Intel stock from December 15, 1972, to December 31, 2008. There are 9096 observations. For a long position in the Intel stock, we consider the negative log returns. Since $9096 \times 0.95 = 8641.2$, we have $\ell_1 = 8641$, $\ell_2 = 8642$, $p_1 = 8641/9096$, and $p_2 = 8642/9096$. The empirical 95% quantile of the negative log returns can be obtained as

$$\hat{x}_{0.95} = 0.8 r_{(8641)} + 0.2 r_{(8642)} = 4.2952\%,$$

$r_{(i)}$ is the ith order statistic of the negative log returns. In this particular instance, $r_{(8641)} = 4.2951\%$ and $r_{(8642)} = 4.2954\%$.

R Demonstration

```
> da=read.table("d-intc7208.txt",header=T)
> intc=log(da[,2]+1)
> nintc=-intc
> quantile(nintc,0.95)
       95%
0.04295213
> quantile(rtn,.05) % An alternative
       5%
-0.04295213
```

Example 7.5. Consider again the daily log returns of IBM stock from July 3, 1962, to December 31, 1998. Using all 9190 observations, the empirical 95% quantile of the negative log returns can be obtained as $(r_{(8730)} + r_{(8731)})/2 = 0.021603$, where $r_{(i)}$ is the ith order statistic and $np = 9190 \times 0.95 = 8730.5$. The VaR of a long position of \$10 million is \$216,030, which is much smaller than those obtained by the econometric approach discussed before. Because the sample size is 9190, we have $9098 < 9190 \times 0.99 < 9099$. Let $p_1 = 9198/9190 = 0.98999$ and $p_2 = 9099/9190 = 0.9901$. The empirical 99% quantile can be obtained as

$$\begin{aligned}\hat{x}_{0.99} &= \frac{p_2 - 0.99}{p_2 - p_1} r_{(9098)} + \frac{0.99 - p_1}{p_2 - p_1} r_{(9099)} \\ &= \frac{0.0001}{0.00011}(3.627) + \frac{0.00001}{0.00011}(3.657) \\ &\approx 3.630.\end{aligned}$$

The 1% 1-day horizon VaR of the long position is \$363, 000. Again this amount is lower than those obtained before by other methods.

Discussion. Advantages of using the empirical quantile method to VaR calculation include (a) simplicity and (b) using no specific distributional assumption. However, the approach has several drawbacks. First, it assumes that the distribution of the return r_t remains unchanged from the sample period to the prediction period. Given that VaR is concerned mainly with tail probability, this assumption implies that the predicted loss cannot be greater than that of the historical loss. It is definitely not so in practice. Second, when the tail probability p is small, the empirical quantile is not an efficient estimate of the theoretical quantile. Third, the direct quantile estimation fails to take into account the effect of explanatory variables that are relevant to the portfolio under study. In real application, VaR obtained by the empirical quantile can serve as a lower bound for the actual VaR. □

The expected shortfall can also be estimated directly from the sample returns. Let $\hat{x}_q$ be the empirical qth quantile, where $q = 1 - p$ with p being the upper tail probability. We have

$$\mathrm{ES}_q = \frac{1}{N_q}\sum_{i=1}^{n} x_{(i)} I[x_{(i)} > \hat{x}_q],$$

where $I[\cdot] = 1$ if $x_{(i)} > \hat{x}_q$ and $= 0$, otherwise, and N_q denotes the number of x_i greater than $\hat{x}_q$. For illustration, consider the negative IBM daily log returns. If $p = 0.01$, we have $\hat{x}_{0.99} = 3.630$. Therefore, $\mathrm{ES}_{0.99} = 5.097$.

R Demonstration

```
> da=read.table("d-ibm6298.txt",header=T)
> ibm=log(da[,2]+1)*100
> nibm=-ibm
> q99=quantile(nibm,0.99)
> q99
     99%
 [1] 3.630295
> idx=c(1:length(nibm))[nibm>q99] % locate the exceedances
> es=mean(nibm[idx])
> es
[1] 5.097222
```

7.4.2 Quantile Regression

In real application, one often has explanatory variables available that are important to the problem under study. For example, the action taken by Federal Reserve Banks on interest rates could have important impacts on the returns of U.S. stocks. It is then more appropriate to consider the distribution function $r_{t+1}|F_t$, where F_t includes the explanatory variables. In other words, we are interested in the quantiles of the distribution function of r_{t+1} given F_t. Such a quantile is referred to as a *regression quantile* in the literature; see Koenker and Bassett (1978).

To understand regression quantile, it is helpful to cast the empirical quantile of the previous subsection as an estimation problem. For a given probability p, the pth quantile of $\{r_t\}$ is obtained by

$$\hat{x}_p = \mathrm{argmin}_\beta \sum_{i=1}^{n} w_p(r_i - \beta),$$

where $w_p(z)$ is defined by

$$w_p(z) = \begin{cases} pz & \text{if } z \geq 0, \\ (p-1)z & \text{if } z < 0. \end{cases}$$

Regression quantile is a generalization of such an estimate.

To see the generalization, suppose that we have the linear regression

$$r_t = \boldsymbol{\beta}'\boldsymbol{x}_t + a_t, \tag{7.13}$$

where $\boldsymbol{\beta}$ is a k-dimensional vector of parameters and $\boldsymbol{x}_t$ is a vector of predictors that are elements of F_{t-1}. The conditional distribution of r_t given F_{t-1} is a translation of the distribution of a_t because $\boldsymbol{\beta}'\boldsymbol{x}_t$ is known. Viewing the problem this way, Koenker and Bassett (1978) suggest estimating the conditional quantile $x_p|F_{t-1}$ of r_t given F_{t-1} as

$$\hat{x}_p|F_{t-1} \equiv \inf\{\boldsymbol{\beta}_o'\boldsymbol{x}|R_p(\boldsymbol{\beta}_o) = \min\}, \tag{7.14}$$

where "$R_p(\boldsymbol{\beta}_o) = \min$" means that $\boldsymbol{\beta}_o$ is obtained by

$$\boldsymbol{\beta}_o = \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{t=1}^{n} w_p(r_t - \boldsymbol{\beta}'\boldsymbol{x}_t),$$

where $w_p(\cdot)$ is defined as before. A computer program to obtain such an estimated quantile can be found in Koenker and D'Orey (1987). The package `quantreg` of R performs quantile regression analysis.

7.5 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, $r_{(1)} = \min_{1\le j\le n}\{r_j\}$ and $r_{(n)} = \max_{1\le j\le n}\{r_j\}$. 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 $r_{(1)} = -\max_{1\le j\le n}\{-r_j\} = -r^c_{(n)}$, where $r^c_t = -r_t$ 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.

7.5.1 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

$$\begin{aligned}
F_{n,n}(x) &= \Pr[r_{(n)} \le x] \\
&= \Pr(r_1 \le x, r_2 \le x, \ldots, r_n \le x) \quad \text{(by definition of maximum)} \\
&= \prod_{j=1}^{n} \Pr(r_j \le x) \quad \text{(by independence)} \\
&= \prod_{j=1}^{n} F(x) = [F(x)]^n. \qquad (7.15)
\end{aligned}$$

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 $x \ge u$ 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

$$F_*(x) = \begin{cases} \exp[-(1+\xi x)^{-1/\xi}] & \text{if } \xi \ne 0, \\ \exp[-\exp(-x)] & \text{if } \xi = 0, \end{cases} \qquad (7.16)$$

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 ξ 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. (7.16) 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

$$F_*(x) = \exp[-\exp(-x)], \qquad -\infty < x < \infty. \qquad (7.17)$$

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

$$F_*(x) = \begin{cases} \exp[-(1+\xi x)^{-1/\xi}] & \text{if } x > -1/\xi, \\ 0 & \text{otherwise.} \end{cases} \qquad (7.18)$$

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

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

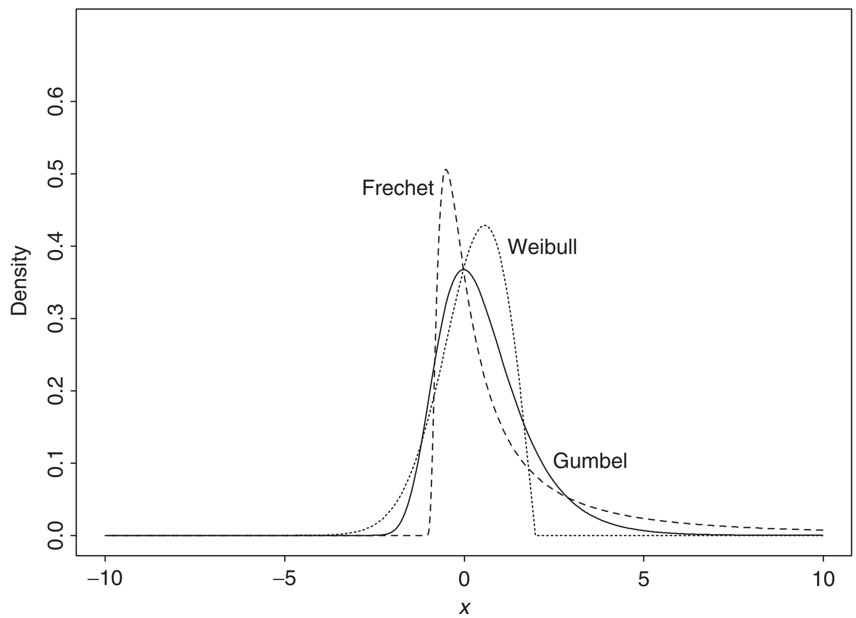

Figure 7.2 Probability density functions of extreme value distributions for maximum. Solid line is for Gumbel distribution, dotted line is for Weibull distribution with $\xi = -0.5$, and dashed line is for Fréchet distribution with $\xi = 0.9$.

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 7.2). 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, which 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. (7.16) can be obtained easily by differentiation:

$$f_*(x) = \begin{cases} (1+\xi x)^{-1/\xi-1}\exp[-(1+\xi x)^{-1/\xi}] & \text{if } \xi \neq 0, \\ \exp[-x-\exp(-x)] & \text{if } \xi = 0, \end{cases} \tag{7.19}$$

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 ξ 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 ρ_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 7.8.

7.5.2 Empirical Estimation

The extreme value distribution contains three parameters—ξ, β_n, and α_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.

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 $\{r_j\}_{j=1}^T$ available. We divide the sample into g nonoverlapping 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 $1 \le j \le n$ 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 ith subsample (i.e., $r_{n,i}$ is the largest return of the ith 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

$$r_{n,i} = \max_{1 \le j \le n} \{r_{(i-1)n+j}\}, \qquad i = 1, \ldots, g. \tag{7.20}$$

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. (7.19), we can obtain the pdf of $r_{n,i}$ by a simple transformation as

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

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 ξ 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 ξ_n, β_n, and α_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. (7.16) 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

$$E\{F_*[r_{n(i)}]\} = \frac{i}{g+1}, \qquad i = 1, \ldots, g. \tag{7.21}$$

For simplicity, we separate the discussion into two cases depending on the value of ξ. First, consider the case of $\xi \neq 0$. From Eq. (7.16), we have

$$F_*[r_{n(i)}] = \exp\left[-\left(1 + \xi_n \frac{r_{n(i)} - \beta_n}{\alpha_n}\right)^{-1/\xi_n}\right]. \tag{7.22}$$

Consequently, using Eqs. (7.21) and (7.22) and approximating expectation by an observed value, we have

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

Taking natural logarithm twice, the prior equation gives

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

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

$$\ln\left[-\ln\left(\frac{i}{g+1}\right)\right] = \frac{-1}{\xi_n} \ln\left(1 + \xi_n \frac{r_{n(i)} - \beta_n}{\alpha_n}\right) + e_i, \qquad i = 1, \ldots, g. \tag{7.23}$$

The least-squares estimates of ξ_n, β_n, and α_n can be obtained by minimizing the sum of squares of e_i.

When $\xi_n = 0$, the regression setup reduces to

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

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 ξ 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

$$r_{(1)} \le r_{(2)} \le \cdots \le r_{(T)}.$$

Let q be a positive integer. The two estimators of ξ are defined as

$$\xi_p(q) = \frac{1}{\ln(2)} \ln\left(\frac{r_{(T-q+1)} - r_{(T-2q+1)}}{r_{(T-2q+1)} - r_{(T-4q+1)}}\right), \quad q \le T/4, \tag{7.24}$$

$$\xi_h(q) = \frac{1}{q} \sum_{i=1}^{q} \left[\ln(r_{(T-i+1)}) - \ln(r_{(T-q)})\right], \tag{7.25}$$

where the argument (q) is used to emphasize that the estimators depend on q and the subscripts p and h denote Pickands and Hill estimators, 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 $\xi^2(2^{2\xi+1} + 1)/[2(2^{\xi} - 1)\ln(2)]^2$. 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 ξ^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).

7.5.3 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 7.3 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 7.1 summarizes some estimation results of the shape parameter ξ 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 7.4 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

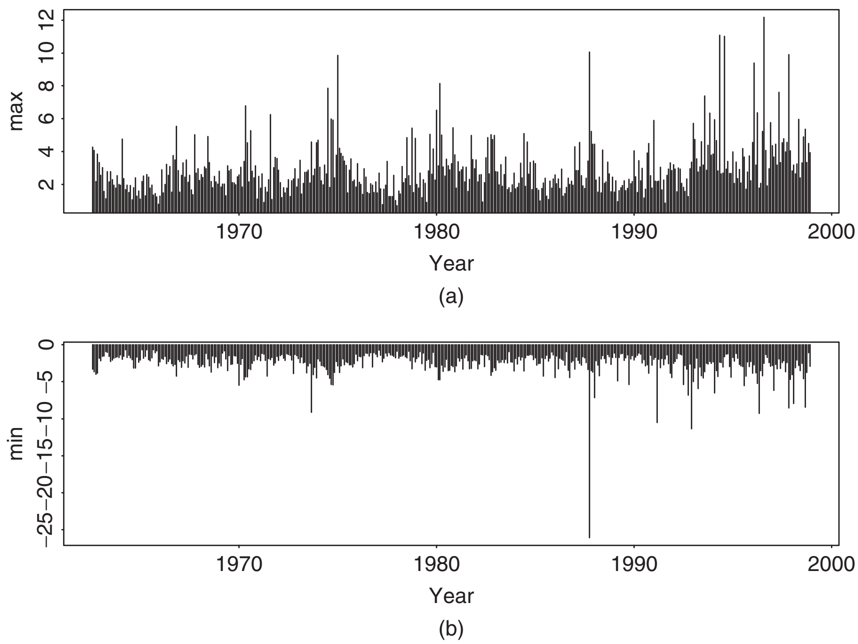

Figure 7.3 Maximum and minimum daily log returns of IBM stock when subperiod is 21 trading days. Data span is from July 3, 1962, to December 31, 1998: (a) positive returns and (b) negative returns.

TABLE 7.1 Results of Hill Estimator for Daily Log Returns of IBM Stock from July 3, 1962, to December 31, 1998[a]

q	190	200	210
r_t	0.300(0.022)	0.299(0.021)	0.305(0.021)
$-r_t$	0.290(0.021)	0.292(0.021)	0.289(0.020)

[a] Standard errors are in parentheses.

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 ξ 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 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 7.2 summarizes the estimation results for different choices of the length of subperiods ranging

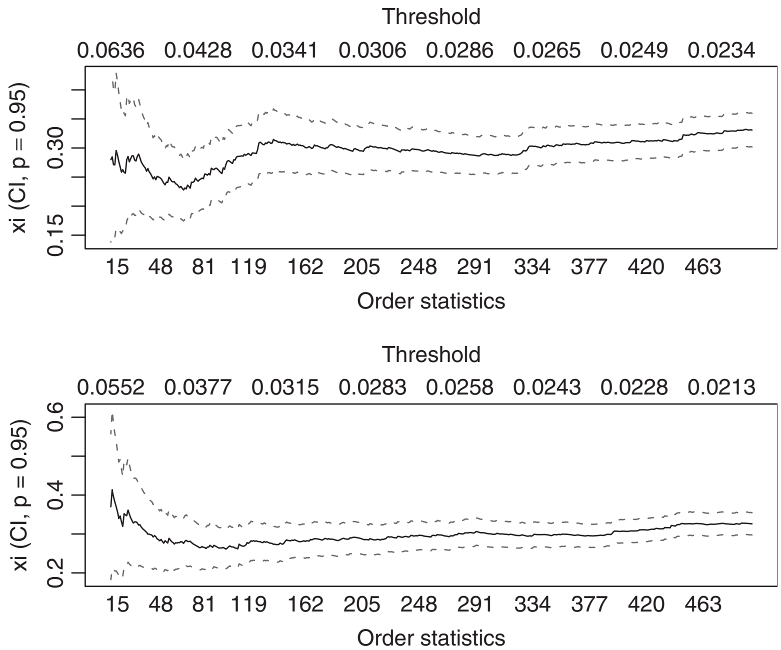

Figure 7.4 Scatterplots of Hill estimator for daily log returns of IBM stock. Sample period is from July 3, 1962, to December 31, 1998: upper plot is for positive returns and lower one for negative returns.

TABLE 7.2 Maximum-Likelihood Estimates of Extreme Value Distribution for Daily Log Returns of IBM Stock from July 3, 1962 to December 31, 1998[a]

Length of Subperiod	Scale α_n	Location β_n	Shape Par. ξ_n
	Minimal Returns		
1 mon. ($n = 21$, $g = 437$)	0.823(0.035)	1.902(0.044)	0.197(0.036)
1 qur ($n = 63$, $g = 145$)	0.945(0.077)	2.583(0.090)	0.335(0.076)
6 mon. ($n = 126$, $g = 72$)	1.147(0.131)	3.141(0.153)	0.330(0.101)
1 year ($n = 252$, $g = 36$)	1.542(0.242)	3.761(0.285)	0.322(0.127)
	Maximal Returns		
1 mon. ($n = 21$, $g = 437$)	0.931(0.039)	2.184(0.050)	0.168(0.036)
1 qur ($n = 63$, $g = 145$)	1.157(0.087)	3.012(0.108)	0.217(0.066)
6 mon. ($n = 126$, $g = 72$)	1.292(0.158)	3.471(0.181)	0.349(0.130)
1 year ($n = 252$, $g = 36$)	1.624(0.271)	4.475(0.325)	0.264(0.186)

[a] Standard errors are in parentheses.

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 β_n and α_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.

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 7.2 were obtained using a Fortran program developed by 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, sigma, mu 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 7.2. A source of minor difference is that in Table 7.2 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 7.2.

R Demonstration for Extreme Value Analysis

The series is daily IBM log returns from 1962 to 1998. The following output was edited:

```
> 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.
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),"\n")
}
> 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
```

Define the residuals of a GEV distribution fit as

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

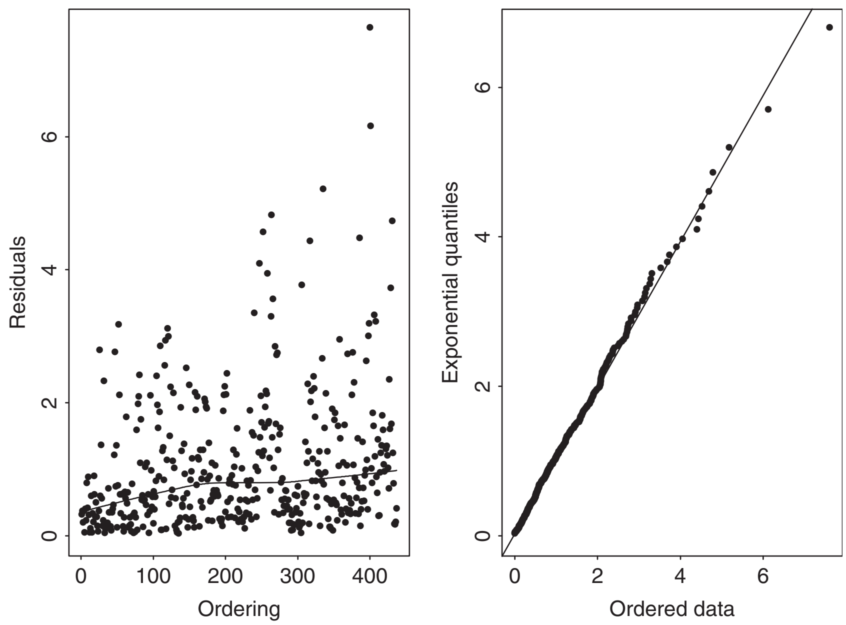

Figure 7.5 Residual plots from fitting GEV distribution to daily negative IBM log returns, in percentage, for data from July 3, 1962, to December 31, 1998, with subperiod length of 21 days.

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 7.5 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`. □

7.6 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 $1 \le m < n$, 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 β_n, α_n, and ξ_n for the subperiod maxima $\{r_{n,i}\}$. Plugging the maximum-likelihood estimates into the CDF in Eq. (7.16) 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

$$r_n^* = \left\{ \begin{array}{ll} \beta_n - \frac{\alpha_n}{\xi_n}\left\{1 - \left[-\ln(1 - p^*)\right]^{-\xi_n}\right\} & \text{if } \xi_n \neq 0, \\ \beta_n - \alpha_n \ln[-\ln(1 - p^*)] & \text{if } \xi_n = 0. \end{array} \right\} \tag{7.26}$$

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. (7.26) 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. (7.15) and obtain

$$1 - p^* = P(r_{n,i} \le r_n^*) = [P(r_t \le r_n^*)]^n. \tag{7.27}$$

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. (7.27), where $P(r_t \le r_n^*) = 1 - p$. Consequently, for a given small upper tail probability p, the VaR of a financial position with log return r_t is

$$\text{VaR} = \left\{ \begin{array}{ll} \beta_n - \frac{\alpha_n}{\xi_n}\left\{1 - \left[-n\ \ln(1-p)\right]^{-\xi_n}\right\} & \text{if } \xi_n \neq 0 \\ \beta_n - \alpha_n \ln[-n\ \ln(1-p)] & \text{if } \xi_n = 0, \end{array} \right\} \tag{7.28}$$

where n is the length of the 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 $g = [T/n]$.
2. Obtain the maximum-likelihood estimates of β_n, α_n, and ξ_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. (7.28) 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 7.2, 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

$$\begin{aligned} \text{VaR} &= 2.583 - \frac{0.945}{0.335}\left\{1 - [-63\ln(1-0.01)]^{-0.335}\right\} \\ &= 3.04969. \end{aligned}$$

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 × 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

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

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 VaR = \$546,641 for the Gaussian AR(2)–GARCH(1,1) model and 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 α_n, β_n, and ξ_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. □

7.6.1 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
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
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
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%.

7.6.2 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 ℓ-day and

1-day horizons is

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

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

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 α root of time rule produces lower ℓ-day horizon VaR than the square root of time rule does.

7.6.3 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

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

provided that $\xi_n \neq 0$. Note that this is precisely the quantile of extreme value distribution given in Eq. (7.26) 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

```
> 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.

7.7 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 η be a prespecified high threshold. We may choose $\eta = 2.5\%$. Suppose that the ith 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 η and t_i is the time at which the ith 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 η 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 η. Different choices of the threshold η 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 η 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 η.

The choice of threshold η 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), η may be as high as 10%. Limited experience shows that η 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 η, see Danielsson and de Vries (1997b).

7.7.1 Statistical Theory

Again consider the log return r_t of an asset. Suppose that the ith exceedance occurs at t_i. Focusing on the exceedance $r_t - \eta$ 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 $x = r_t - \eta$ given $r_t \le \eta$ for a long position. Occurrence of the event $\{r_t \le \eta\}$ follows a point process (e.g., a Poisson process). See Section 6.9 for the definition of a Poisson process. In particular, if the intensity parameter λ of the process is time invariant, then the Poisson process is homogeneous. If λ 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. (7.16). 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 $r \le x + \eta$ given $r > \eta$ is

$$\Pr(r \le x + \eta | r > \eta) = \frac{\Pr(\eta \le r \le x + \eta)}{\Pr(r > \eta)} = \frac{\Pr(r \le x + \eta) - \Pr(r \le \eta)}{1 - \Pr(r \le \eta)}. \quad (7.29)$$

Using the CDF $F_*(\cdot)$ of Eq. (7.16) and the approximation $e^{-y} \approx 1 - y$ and after some algebra, we obtain that

$$\begin{aligned}\Pr(r \le x + \eta | r > \eta) &= \frac{F_*(x+\eta) - F_*(\eta)}{1 - F_*(\eta)} \\ &= \frac{\exp\left\{-\left[1 + \frac{\xi(x+\eta-\beta)}{\alpha}\right]^{-1/\xi}\right\} - \exp\left\{-\left[1 + \frac{\xi(\eta-\beta)}{\alpha}\right]^{-1/\xi}\right\}}{1 - \exp\left\{-\left[1 + \frac{\xi(\eta-\beta)}{\alpha}\right]^{-1/\xi}\right\}} \\ &\approx 1 - \left[1 + \frac{\xi x}{\alpha + \xi(\eta - \beta)}\right]^{-1/\xi}, \qquad (7.30)\end{aligned}$$

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

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

The distribution with cumulative distribution function

$$G_{\xi,\psi(\eta)}(x) = \begin{cases} 1 - \left[1 + \frac{\xi x}{\psi(\eta)}\right]^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp[-x/\psi(\eta)] & \text{for } \xi = 0, \end{cases} \tag{7.31}$$

where $\psi(\eta) > 0$, $x \ge 0$ when $\xi \ge 0$, and $0 \le x \le -\psi(\eta)/\xi$ when $\xi < 0$, is called the *generalized Pareto distribution* (GPD). Thus, the result of Eq. (7.30) shows that the conditional distribution of r given $r > \eta$ is well approximated by a GPD with parameters ξ and $\psi(\eta) = \alpha + \xi(\eta - \beta)$. See Embrechts et al. (1997) for further information. An important property of the GPD is as follows. Suppose that the excess distribution of r given a threshold η_o is a GPD with shape parameter ξ and scale parameter $\psi(\eta_o)$. Then, for an arbitrary threshold $\eta > \eta_o$, the excess distribution over the threshold η is also a GPD with shape parameter ξ and scale parameter $\psi(\eta) = \psi(\eta_o) + \xi(\eta - \eta_o)$.

When $\xi = 0$, the GPD in Eq. (7.31) 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 7.6(a) 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 7.6

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

7.7.2 Mean Excess Function

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

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

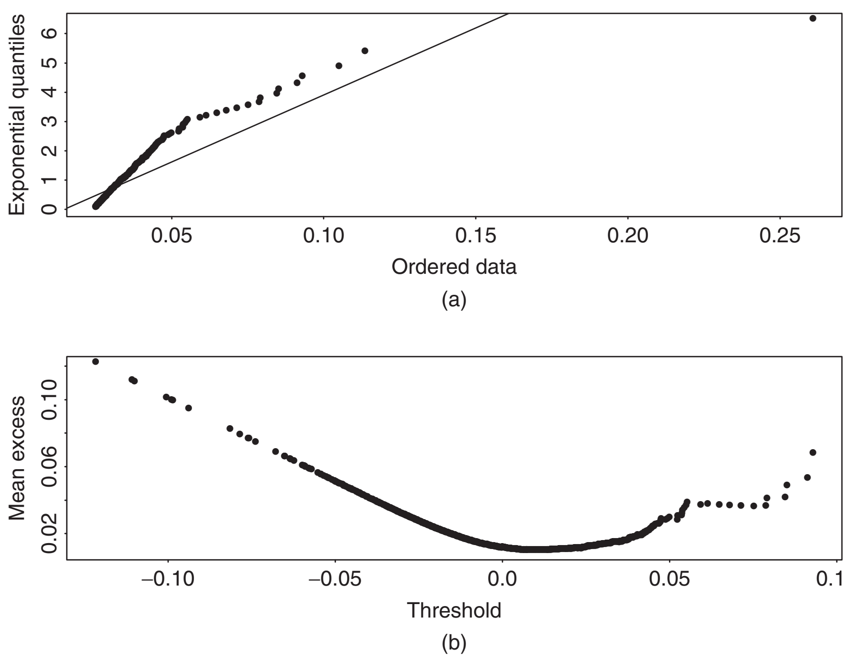

Figure 7.6 Plots for daily negative IBM log returns from July 3, 1962, to December 31, 1998. (a) QQ plot of excess returns over threshold 2.5% and (b) mean excess plot.

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 ξ, 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 η_o for the GPD. Define the *empirical mean excess function* as

$$e_T(\eta) = \frac{1}{N_\eta} \sum_{i=1}^{N_\eta} (r_{t_i} - \eta), \tag{7.32}$$

where N_η is the number of returns that exceed η 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 η is called the *mean excess plot*, which should be linear in η for $\eta > \eta_o$ under the GPD. The plot is also called *mean residual life plot*.

Figure 7.6(b) 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`.

7.7.3 New Approach to Modeling Extreme Values

Using the statistical result in Eq. (7.30) 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 η, 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_η depends on the threshold η.

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

$$\Lambda[(D_2, D_1) \times (r, \infty)] = \frac{D_2 - D_1}{D} S(r; \xi, \alpha, \beta), \tag{7.33}$$

where

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

$0 \le D_1 \le D_2 \le T$, $r > \eta$, $\alpha > 0$, β, and ξ are parameters, and the notation $[x]_+$ is defined as $[x]_+ = \max(x, 0)$. This intensity measure says that the occurrence of exceeding the threshold is proportional to the length of the time interval $[D_1, D_2]$ and the probability is governed by a survival function similar to the exponent of the CDF $F_*(r)$ in Eq. (7.16). A survival function of a random variable X is defined as $S(x) = \Pr(X > x) = 1 - \Pr(X \le x) = 1 - \text{CDF}(x)$. When $\xi = 0$, the intensity measure is taken as the limit of $\xi \to 0$; 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. (7.33), the length of time interval is measured with respect to the baseline interval D.

The idea of using the intensity measure in Eq. (7.33) 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. (7.30). This survival function is obtained from the extreme limiting distribution for maximum in Eq. (7.16). We use survival function here because it denotes the probability of exceedance.

The relationship between the limiting extreme value distribution in Eq. (7.16) and the intensity measure in Eq. (7.33) directly connects the new approach of extreme value theory to the traditional one.

Mathematically, the intensity measure in Eq. (7.33) can be written as an integral of an intensity function:

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

where the intensity function $\lambda(t, z; \xi, \alpha, \beta)$ is defined as

$$\lambda(t, z; \xi, \alpha, \beta) = \frac{1}{D} g(z; \xi, \alpha, \beta), \tag{7.34}$$

where

$$g(z; \xi, \alpha, \beta) = \begin{cases} \frac{1}{\alpha}\left[1 + \frac{\xi(z-\beta)}{\alpha}\right]^{-(1+\xi)/\xi} & \text{if } \xi \neq 0, \\ \frac{1}{\alpha}\exp\left[\frac{-(z-\beta)}{\alpha}\right] & \text{if } \xi = 0. \end{cases}$$

Using the results of a Poisson process, we can write down the likelihood function for the observed exceeding times and their corresponding returns $\{(t_i, r_{t_i})\}$ over the two-dimensional space $[0, T] \times (\eta, \infty)$ as

$$L(\xi, \alpha, \beta) = \left[\prod_{i=1}^{N_\eta} \frac{1}{D} g(r_{t_i}; \xi, \alpha, \beta)\right] \exp\left[-\frac{T}{D} S(\eta; \xi, \alpha, \beta)\right]. \tag{7.35}$$

The parameters ξ, α, and β can then be estimated by maximizing the logarithm of this likelihood function. Since the scale parameter α is nonnegative, 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 7.3 gives some

TABLE 7.3 Estimation Results of a Two-Dimensional Homogeneous Poisson Model for Daily Negative Log Returns of IBM Stock from July 3, 1962 to December 31, 1998[a]

Thr.	Exc.	Shape Parameter ξ	Log(Scale) $\ln(\alpha)$	Location β
		Original Log Returns		
3.0%	175	0.30697(0.09015)	0.30699(0.12380)	4.69204(0.19058)
2.5%	310	0.26418(0.06501)	0.31529(0.11277)	4.74062(0.18041)
2.0%	554	0.18751(0.04394)	0.27655(0.09867)	4.81003(0.17209)
		Removing the Sample Mean		
3.0%	184	0.30516(0.08824)	0.30807(0.12395)	4.73804(0.19151)
2.5%	334	0.28179(0.06737)	0.31968(0.12065)	4.76808(0.18533)
2.0%	590	0.19260(0.04357)	0.27917(0.09913)	4.84859(0.17255)

[a]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.

estimation results of the parameters ξ, α, and β for three choices of the threshold when the negative series $\{-r_t\}$ is used. As mentioned before, we use the negative series $\{-r_t\}$, instead of $\{r_t\}$ 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 section.

7.7.4 VaR Calculation Based on the New Approach

As shown in Eq. (7.30), the two-dimensional Poisson process model used, which employs the intensity measure in Eq. (7.33), has the same parameters as those of the extreme value distribution in Eq. (7.16). Therefore, one can use the same formula as that of Eq. (7.28) 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

$$\text{VaR} = \begin{cases} \beta - \frac{\alpha}{\xi}\left\{1 - \left[-D\,\ln(1-p)\right]^{-\xi}\right\} & \text{if } \xi \neq 0, \\ \beta - \alpha\ln[-D\,\ln(1-p)] & \text{if } \xi = 0, \end{cases} \tag{7.36}$$

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 7.3 to calculate 1-day horizon VaR for the tail probabilities of 0.05 and 0.01.

- Case I: Use the original daily log returns. The three choices of threshold η 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 η 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. □

The command `pot` of the R package `evir` can be used to perform the estimation of the 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

```
> 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.ests
         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
[3,] 0.999 0.07019419 0.09699513
```

7.7.5 Alternative Parameterization

As mentioned before, for a given threshold η, the GPD can also be parameterized by the shape parameter ξ 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, `(xi,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. The following output was edited:

```
> 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 7.3, where percentage log returns are used. The estimates of ξ 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 7.3. In terms of log returns, the estimate of $\psi(\eta)$ is 0.007787, which is the same as the R and S-Plus estimate.

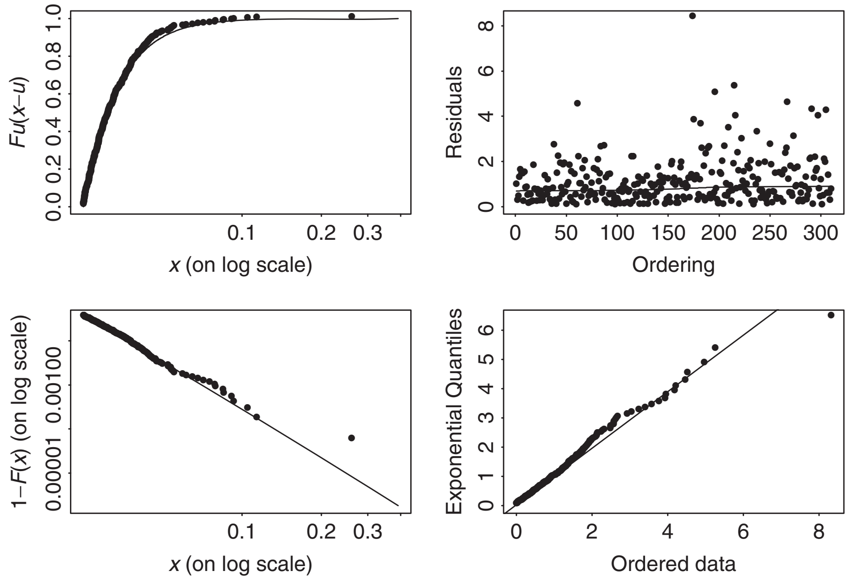

Figure 7.7 Diagnostic plots for GPD fit to daily negative log returns of IBM stock from July 3, 1962, to December 31, 1998.

Figure 7.7 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. (7.29) and (7.30) and the GPD in Eq. (7.31), 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_η is the number of exceedances of the threshold η and T is the sample size. Consequently, by Eq. (7.31),

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

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 qth quantile of $F(y)$, denoted by VaR_q, by

$$\text{VaR}_q = \eta - \frac{\psi(\eta)}{\xi}\left\{1 - \left[\frac{T}{N_\eta}(1-q)\right]^{-\xi}\right\}, \tag{7.37}$$

where, as before, η is the threshold, T is the sample size, N_η is the number of exceedances, and $\psi(\eta)$ and ξ 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 in Section 7.2.3, *expected shortfall* (ES) associated with a given VaR is a useful risk measure. It is defined as the expected loss given that the VaR is exceeded. For generalized Pareto distribution, ES assumes a simple form. Specifically, for a given tail probability p, let $q = 1 - p$ and denote the value at risk by VaR_q. Then, the expected shortfall is defined by

$$\text{ES}_q = E(r|r > \text{VaR}_q) = \text{VaR}_q + E(r - \text{VaR}_q|r > \text{VaR}_q). \tag{7.38}$$

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

$$\text{ES}_q = \frac{\text{VaR}_q}{1 - \xi} + \frac{\psi(\eta) - \xi\eta}{1 - \xi}.$$

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 `evir` package of R and S-Plus, the command to compute VaR and ES via the peak over threshold method is `riskmeasures`:

```
> 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 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 section. The expected shortfalls for the financial position are \$316, 272 and \$507, 576, respectively, for tail probability of 0.05 and 0.01.

7.7.6 Use of Explanatory Variables

The two-dimensional Poisson process model discussed earlier is *homogeneous* because the three parameters ξ, α, and β 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 section. In addition, we also discuss methods that can be used to check the adequacy of a fitted two-dimensional Poisson process model.

Suppose that $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{vt})'$ is a vector of v explanatory variables that are available *prior to* time t. For asset returns, the volatility σ_t^2 of r_t discussed in Chapter 3 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 ξ, α, and β are time varying and are linear functions of the explanatory variables. Specifically, when explanatory variables $\boldsymbol{x}_t$ are available, we assume that

$$\begin{aligned} \xi_t &= \gamma_0 + \gamma_1 x_{1t} + \cdots + \gamma_v x_{vt} \equiv \gamma_0 + \boldsymbol{\gamma}'\boldsymbol{x}_t, \\ \ln(\alpha_t) &= \delta_0 + \delta_1 x_{1t} + \cdots + \delta_v x_{vt} \equiv \delta_0 + \boldsymbol{\delta}'\boldsymbol{x}_t, \\ \beta_t &= \theta_0 + \theta_1 x_{1t} + \cdots + \theta_v x_{vt} \equiv \theta_0 + \boldsymbol{\theta}'\boldsymbol{x}_t. \end{aligned} \tag{7.39}$$

If $\boldsymbol{\gamma} = \mathbf{0}$, then the shape parameter $\xi_t = \gamma_0$, which is time invariant. Thus, testing the significance of $\boldsymbol{\gamma}$ 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. (7.39), we use the same explanatory variables for all three parameters ξ_t, $\ln(\alpha_t)$, and β_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

$$\Lambda[(D_1, D_2) \times (r, \infty)] = \frac{D_2 - D_1}{D}\left[1 + \frac{\xi_t(r - \beta_t)}{\alpha_t}\right]_+^{-1/\xi_t}, \qquad r > \eta. \tag{7.40}$$

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

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

which reduces to

$$L = \left[\prod_{i=1}^{N_\eta} \frac{1}{D} g(r_{t_i}; \xi_{t_i}, \alpha_{t_i}, \beta_{t_i})\right] \exp\left[-\frac{1}{D}\sum_{t=1}^{T} S(\eta; \xi_t, \alpha_t, \beta_t)\right] \tag{7.41}$$

if one assumes that the parameters ξ_t, α_t, and β_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. (7.34) and (7.33), respectively. For given observations $\{r_t, \boldsymbol{x}_t | t = 1, \ldots, T\}$, the baseline time interval D, and the threshold η, the parameters in Eq. (7.39) can be estimated by maximizing the logarithm of the likelihood function in Eq. (7.41). Again we use $\ln(\alpha_t)$ to satisfy the positive constraint of α_t.

Remark. The parameterization in Eq. (7.39) is similar to that of the volatility models of Chapter 3 in the sense that the three parameters are exact functions of the available information at time t. Other functions can be used if necessary. □

7.7.7 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 ith and $(i-1)$th exceedances should follow an exponential distribution. More specifically, letting $t_0 = 0$, we expect that

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

are iid as a standard exponential distribution. Because daily returns are discrete-time observations, we employ the time durations

$$z_{t_i} = \frac{1}{D} \sum_{t=t_{i-1}+1}^{t_i} S(\eta; \xi_t, \alpha_t, \beta_t) \tag{7.42}$$

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 η is a GPD with shape parameter ξ_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

$$w_{t_i} = \begin{cases} \frac{1}{\xi_{t_i}} \ln\left(1 + \xi_{t_i} \frac{r_{t_i} - \eta}{\psi_{t_i}}\right)_+ & \text{if } \xi_{t_i} \neq 0, \\ \frac{r_{t_i} - \eta}{\psi_{t_i}} & \text{if } \xi_{t_i} = 0. \end{cases} \tag{7.43}$$

If the model is adequate, $\{w_{t_i}\}$ 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_i} and w_{t_i}. Under the independence assumption, we expect that both z_{t_i} and w_{t_i} have no serial correlations.

7.7.8 An Illustration

In this section, 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 section. Figures 7.8(a) and 7.8(b) show the autocorrelation functions of the statistics z_{t_i} and w_{t_i}, defined in Eqs. (7.42) and (7.43), 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_i} and w_{t_i} series have some significant serial correlations. Figures 7.9(a) and 7.9(b) show the QQ plots of the z_{t_i} and w_{t_i} 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_i} shows some discrepancy.

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

$$r_t^o = r_t - \bar{r}, \qquad \bar{r} = \frac{1}{9190} \sum_{t=1}^{9190} r_t,$$

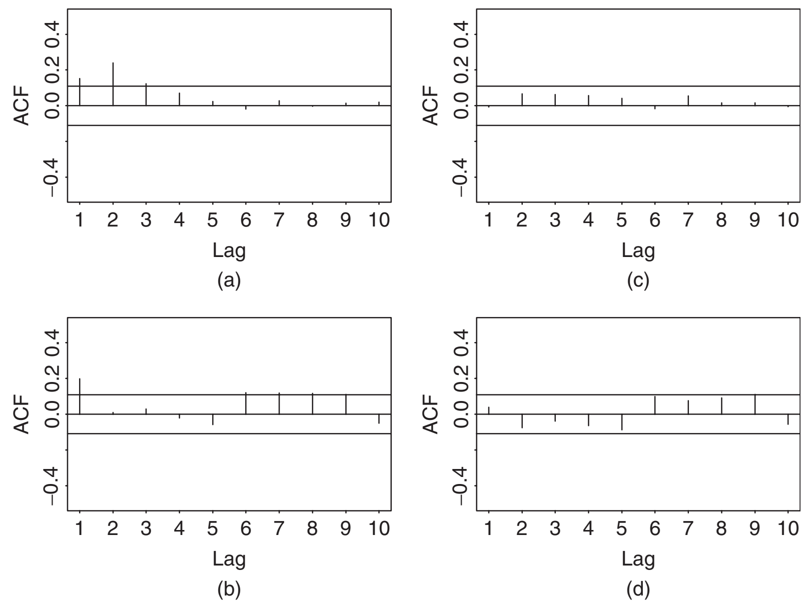

Figure 7.8 Sample autocorrelation functions of the z and w measures for two-dimensional Poisson models. Parts (a) and (b) are for homogeneous model and parts (c) and (d) are for inhomogeneous model. Data are daily mean-corrected log returns, in percentages, of IBM stock from July 3, 1962, to December 31, 1998, and the threshold is 2.5%. A long financial position is used.

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 $r^o_{t-1} \le -2.5\%$. 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 $|r^o_{t-i}| \ge 2.5\%$ 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.

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 σ_t^2 is the conditional variance of the GARCH(1,1) model

$$r_t^o = a_t, \qquad a_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim N(0,1),$$
$$\sigma_t^2 = 0.04565 + 0.0807 a_{t-1}^2 + 0.9031 \sigma_{t-1}^2.$$

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 7.4. Figures 7.8(c) and 7.8(d) and Figures 7.9(c) and 7.9(d) 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_i} and w_{t_i} are within the asymptotic two standard error limits. The QQ plots also show marked improvements as they

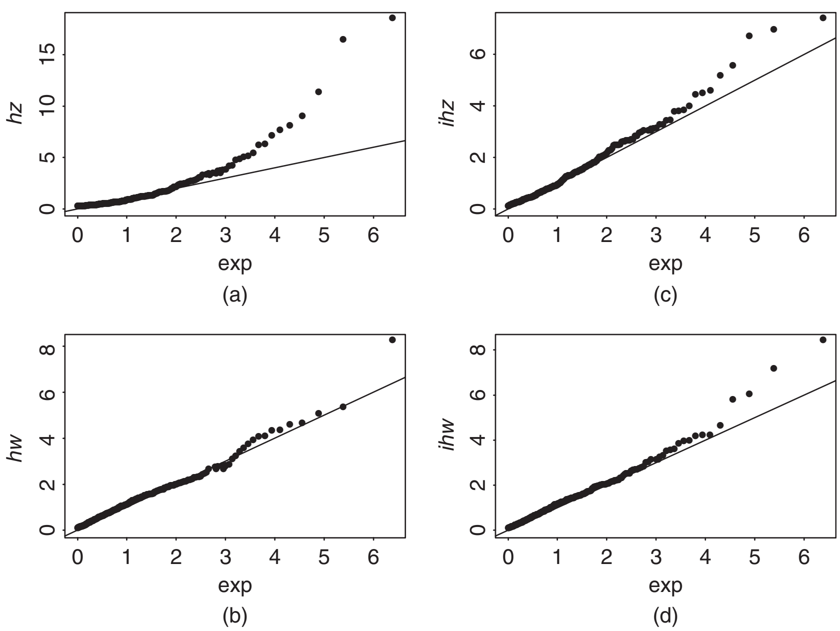

Figure 7.9 Quantile-to-quantile plot of z and w measures for two-dimensional Poisson models. Parts (a) and (b) are for homogeneous model and parts (c) and (d) are for inhomogeneous model. Data are daily mean-corrected log returns, in percentages, of IBM stock from July 3, 1962, to December 31, 1998, and the threshold is 2.5%. A long financial position is used.

TABLE 7.4 Estimation Results of Two-Dimensional Inhomogeneous Poisson Process Model for Daily Log Returns, in Percentages, of IBM Stock from July 3, 1962 to December 31, 1998[a]

Parameter	Constant	Coefficient of x_{3t}	Coefficient of x_{4t}	Coefficient of x_{5t}
		Threshold 2.5% with 334 Exceedances		
β_t	0.3202		1.4772	2.1991
(Std.err)	(0.3387)		(0.3222)	(0.2450)
$\ln(\alpha_t)$	−0.8119	0.3305	1.0324	
(Std.err)	(0.1798)	(0.0826)	(0.2619)	
ξ_t	0.1805	0.2118	0.3551	−0.2602
(Std.err)	(0.1290)	(0.0580)	(0.1503)	(0.0461)
		Threshold 3.0% with 184 Exceedances		
β_t	1.1569			2.1918
(Std.err)	(0.4082)			(0.2909)
$\ln(\alpha_t)$	−0.0316	0.3336		
(Std.err)	(0.1201)	(0.0861)		
ξ_t	0.6008	0.2480		−0.3175
(Std.err)	(0.1454)	(0.0731)		(0.0685)

[a]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.

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 7.4, we obtain

$$\xi_{9190} = 0.01195, \qquad \ln(\alpha_{9190}) = 0.19331, \qquad \beta_{9190} = 6.105.$$

Assume that the tail probability is 0.05. The VaR quantile shown in Eq. (7.36) 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.

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

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

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.

7.8 THE EXTREMAL INDEX

So far our discussions of extreme values are based on the assumption that the data are iid random variables. However, in reality extremal events tend to occur in clusters because of the 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 discussion will be brief. Interested readers are referred to Beirlant et al. (2004, Chapter 10) and Embrechts et al. (1997).

Let $x_1, x_2, \ldots$ be a strictly stationary sequence of random variables with marginal distribution function $F(x)$. Consider the case of n observations $\{x_i | i = 1, \ldots, n\}$.

As before, let $x_{(n)}$ be the maximum of the data, that is, $x_{(n)} = \max\{x_i\}$. We seek the limiting distribution of $(x_{(n)} - \beta_n)/\alpha_n$ for some suitably chosen normalizing constants $\alpha_n > 0$ and β_n. If $\{x_i\}$ were iid, Section 7.5 shows that the only possible nondegenerate 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 ℓ 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 $g = [n/k]$ be the largest integer less than or equal to n/k. The ith 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 ith block, that is, $x_{k,i} = \max\{x_j | j = (i-1) * k + 1, \ldots, i * k\}$. The collection of block maxima is $\{x_{k,i} | i = 1, \ldots, g+1\}$. From the definitions, it is easy to see that

$$x_{(n)} = \max_{i=1,\ldots,g+1} x_{k,i}. \tag{7.44}$$

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 ith block, then $x_{k,i}$ and $x_{k,i+1}$ 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 section. 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 $\{x_i\}$.

7.8.1 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 $\limsup n[1 - F(u_n)] < \infty$, where $F(\cdot)$ 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

$$1 \le i_1 < i_2 < \cdots < i_p < j_1 < \cdots < j_q \le n,$$

where $j_1 - i_p \ge \ell_n$, where ℓ_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 ℓ_n time periods. The condition $D(u_n)$ is satisfied if

$$|P(\max_{i \in A_1 \cup A_2} x_i \le u_n) - P(\max_{i \in A_1} x_i \le u_n) P(\max_{i \in A_2} x_i \le u_n)| \le \delta_{n,\ell_n}, \tag{7.45}$$

where $\delta_{n,\ell_n} \to 0$ as $n \to \infty$. This condition says that any two events of the form $\{\max_{i \in A_1} x_i \le u_n\}$ and $\{\max_{i \in A_2} x_i \le u_n\}$ 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 ℓ_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 ρ_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 β_n and a nondegenerate distribution function $F_*(\cdot)$ such that

$$P\left[\frac{x_{(n)} - \beta_n}{\alpha_n} \le x\right] \to_d F_*(x), \qquad n \to \infty,$$

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 $F_*(x) > 0$, then $F_*(x)$ 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 β_n and a nondegenerate distribution function $\tilde{F}_*(x)$ such that

$$P\left[\frac{\tilde{x}_{(n)} - \beta_n}{\alpha_n} \le x\right] \to_d \tilde{F}_*(x), \qquad n \to \infty,$$

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

$$P\left[\frac{x_{(n)} - \beta_n}{\alpha_n} \le x\right] \to_d F_*(x) = \tilde{F}_*^{\theta}(x), \qquad n \to \infty,$$

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

The constant θ is called the *extremal index*. It plays an important role in determining the limiting distribution $F_*(x)$ for the maximum of a strictly stationary time series. To see this, we provide some simple derivations for the case of $\xi \ne 0$. From the result of Eq. (7.16), $\tilde{F}_*(x)$ 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 $\xi \ne 0$ 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 ξ, β and α. Based on Theorem 2 of Leadbetter (1983), we have

$$\begin{aligned}
F_*(x) = \tilde{F}_*^{\theta}(x) &= \exp\left[-\theta\left(1 + \xi\frac{x - \beta}{\alpha}\right)^{-1/\xi}\right] \\
&= \exp\left[-\left(\frac{1}{\theta^{\xi}} + \xi\frac{x - \beta}{\alpha\theta^{\xi}}\right)^{-1/\xi}\right] = \exp\left[-\left(\xi\frac{\alpha/\xi + x - \beta}{\alpha\theta^{\xi}}\right)^{-1/\xi}\right] \\
&= \exp\left[-\left(1 + \xi\frac{x - \beta + \alpha/\xi - \alpha\theta^{\xi}/\xi}{\alpha\theta^{\xi}}\right)^{-1/\xi}\right] \\
&= \exp\left[-\left(1 + \xi\frac{x - [\beta - \frac{\alpha}{\xi}(1 - \theta^{\xi})]}{\alpha\theta^{\xi}}\right)^{-1/\xi}\right] \\
&= \exp\left[-\left(1 + \xi_*\frac{x - \beta_*}{\alpha_*}\right)^{-1/\xi_*}\right], \qquad (7.46)
\end{aligned}$$

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 ξ, 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 θ. 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 θ a nonnegative number. Assume that for every $\tau > 0$ there exists a sequence of thresholds u_n such that

$$\lim_{n\to\infty} n[1 - F(u_n)] = \tau, \tag{7.47}$$

$$\lim_{n\to\infty} P(x_{(n)} \le u_n) = \exp(-\theta\tau). \tag{7.48}$$

Then θ 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. (7.47) holds, we have

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

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 θ.

7.8.2 Estimation of the Extremal Index

There are several ways to estimate the extremal index θ of a strictly stationary time series $\{x_i\}$. 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 θ, we have, for a large n, that

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

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

$$\lim_{n\to\infty} \frac{\ln P(x_{(n)} \le u_n)}{n \ln F(u_n)} = \theta. \tag{7.49}$$

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

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

where $I(C) = 1$ if the augment C holds and $= 0$ otherwise, that is, $I(C)$ 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 $P(x_{(n)} \le u_n)$ is harder to estimate. One possibility is to use the block maxima. Specifically, let $k = k(n)$ be a properly chosen block size that depends on the sample size n and, as before, let $g = [n/k]$ be the integer part of n/k. For simplicity, assume that $n = gk$. The ith block consists of $\{x_j | j = (i-1)*k+1, \ldots, i*k\}$ and let $x_{k,i}$ be the maximum of the ith block. Using Eq. (7.44) and the approximate independence of block maxima, we have

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

The probability $P(x_{k,i} \le u_n)$ can be estimated from the block maxima, that is,

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

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

$$\hat{\theta}_b^{(1)} = \frac{g}{n}\frac{\ln[1 - G(u_n)/g]}{\ln[1 - N(u_n)/n]} = \frac{1}{k}\frac{\ln[1 - G(u_n)/g]}{\ln[1 - N(u_n)/n]}, \tag{7.50}$$

where the subscript b signifies the blocks method. Note that $N(u_n)$ is the number of exceedances of the sample $\{x_i\}$ 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:

$$\hat{\theta}_b^{(2)} = \frac{1}{k}\frac{G(u_n)/g}{N(u_n)/n} = \frac{G(u_n)}{N(u_n)}.$$

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

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

where $x_{(n)}^* = \max_{2 \le i \le s} x_i$, 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 θ based on *runs*:

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

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 $A_{i,n} = \{x_i > u_n, x_{i+1} \le u_n, \ldots, x_{i+k} \le u_n\}$. 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 θ 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 7.10 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, that is, $\hat{\theta}_b^{(1)}$. From the plot, we 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.

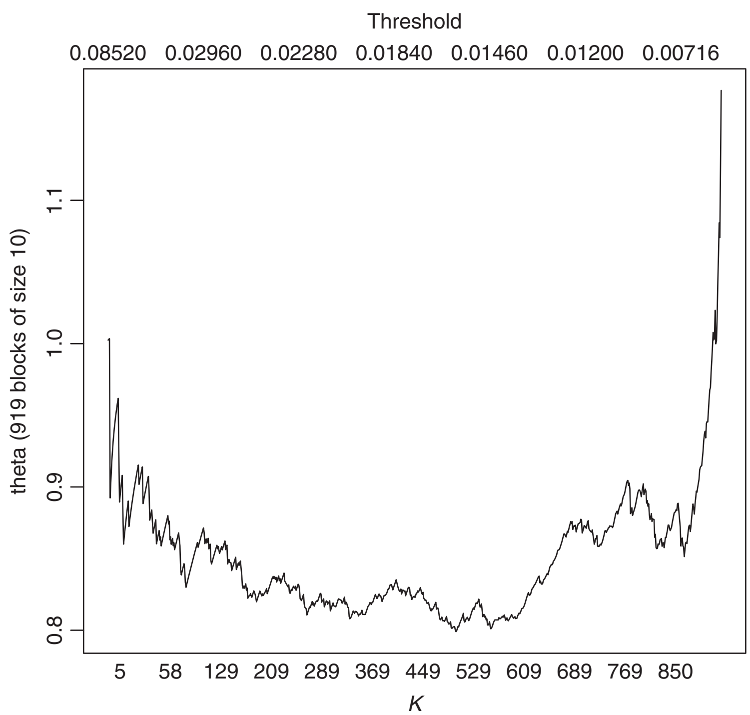

Figure 7.10 Estimates of extremal index for negative daily log returns of IBM stock from July 3, 1962, to December 31, 1998. Block size is $k = 10$ and lower horizontal axis of plot K denotes number of blocks whose maximum exceeds threshold.

7.8.3 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 (1983) can be used to calculate the VaR of a financial position when the associated log returns form a stationary time series. Specifically, from $P(x_{(n)} \leq u_n) \approx [F(x)]^{n\theta}$, the $(1-p)$th quantile of $F(x)$ is the $(1-p)^{n\theta}$th quantile of the limiting extreme value distribution of $x_{(n)}$. Consequently, the VaR of Eq. (7.28) based on the extreme value theory becomes

$$\text{VaR} = \begin{cases} \beta_n - \frac{\alpha_n}{\xi_n}\left\{1 - \left[-n\theta\, \ln(1-p)\right]^{-\xi_n}\right\} & \text{if } \xi_n \neq 0 \\ \beta_n - \alpha_n \ln[-n\theta\, \ln(1-p)] & \text{if } \xi_n = 0, \end{cases} \tag{7.51}$$

where n is the length of the subperiod. From the formula, we risk underestimating the VaR if the extremal index is overlooked.

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 millions 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 the 3.0497 of Example 7.6 when the extremal index is neglected.

R Demonstration

```
> 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
```

EXERCISES

7.1. Consider the daily returns of GE stock from January 2, 1998, to December 31, 2008. The data can be obtained from CRSP or the file `d-ge9808.txt`. Convert the simple returns into log returns. Suppose that you hold a long position on the stock valued at \$1 million. Use the tail probability 0.01. Compute the value at risk of your position for 1-day horizon and 15-day horizon using the following methods:

(a) The RiskMetrics method.

(b) A Gaussian ARMA–GARCH model.

(c) An ARMA–GARCH model with a Student-t distribution. You should also estimate the degrees of freedom.

(d) The traditional extreme value theory with subperiod length $n = 21$.

7.2. The file `d-csco9808.txt` contains the daily simple returns of Cisco Systems stock from 1998 to 2008 with 2767 observations. Transform the simple returns to log returns. Suppose that you hold a long position of Cisco stock valued at \$1 million. Compute the value at risk of your position for the next trading day using probability $p = 0.01$.

(a) Use the RiskMetrics method.

(b) Use a GARCH model with a conditional Gaussian distribution.

(c) Use a GARCH model with a Student-t distribution. You may also estimate the degrees of freedom.

(d) Use the unconditional sample quantile.

(e) Use a two-dimensional homogeneous Poisson process with threshold 2%, that is, focusing on the exceeding times and exceedances that the daily stock price drops 2% or more. Check the fitted model.

(f) Use a two-dimensional nonhomogeneous Poisson process with threshold 2%. The explanatory variables are (1) an annual time trend, (2) a dummy variable for October, November, and December, and (3) a fitted volatility based on a Gaussian GARCH(1,1) model. Perform a diagnostic check on the fitted model.

(g) Repeat the prior two-dimensional nonhomogeneous Poisson process with threshold 2.5 or 3%. Comment on the selection of threshold.

7.3. Use Hill's estimator and the data `d-csco9808.txt` to estimate the tail index for daily log returns of Cisco stock.

7.4. The file `d-hpq3dx9808.txt` contains dates and the daily simple returns of Hewlett-Packard, the CRSP value-weighted index, equal-weighted index, and the S&P 500 index from 1998 to 2008. The returns include dividend distributions. Transform the simple returns to log returns. Assume that the tail probability of interest is 0.01. Calculate value at risk for the following financial positions for the first trading day of year 2009.

(a) Long on Hewlett-Packard stock of \$1 million and S&P 500 index of \$1 million using RiskMetrics. The α coefficient of the IGARCH(1,1) model for each series should be estimated.

(b) The same position as part (a) but using a univariate ARMA–GARCH model for each return series.

(c) A long position on Hewlett-Packard stock of \$1 million using a two-dimensional nonhomogeneous Poisson model with the following explanatory variables: (1) an annual time trend, (2) a fitted volatility based on a Gaussian GARCH model for Hewlett-Packard stock, (3) a fitted volatility based on a Gaussian GARCH model for the S&P 500 index returns, and (4) a fitted volatility based on a Gaussian GARCH model for the value-weighted index return. Perform a diagnostic check for the fitted models. Are the market volatility as measured by the S&P 500 index and value-weighted index returns helpful in determining the tail behavior of stock returns of Hewlett-Packard? You may choose several thresholds.

7.5. Consider the daily returns of Alcoa (AA) stock and the S&P 500 composite index (SPX) from 1998 to 2008. The simple returns and dates are in the file `d-aaspx9808.txt`. Transform the simple returns to log returns and focus on the daily negative log returns of AA stock.

(a) Fit the generalized extreme value distribution to the negative AA log returns, in percentages, with subperiods of 21 trading days. Write down the parameter estimates and their standard errors. Obtain a scatterplot and a QQ plot of the residuals.

(b) What is the return level of the prior fitted model when 24 subperiods of 21 days are used?

(c) Obtain a QQ plot (against exponential distribution) of the negative log returns with threshold 2.5% and a mean excess plot of the returns.

(d) Fit a generalize Pareto distribution to the negative log returns with threshold 3.5%. Write down the parameter estimates and their standard errors.

(e) Obtain (i) a plot of excess distribution, (ii) a plot of the tail of the underlying distribution, (iii) a scatterplot of residuals, and (iv) a QQ plot of the residuals for the fitted GPD.

(f) Based on the fitted GPD model, compute the VaR and expected shortfall for probabilities $q = 0.99$ and 0.999.

7.6. Consider, again, the daily log returns of Alcoa (AA) stock in Exercise 7.5. Focus now on the daily positive log returns. Answer the same questions as in Exercise 7.5. However, use threshold 3% in fitting the GPD model.

7.7. Consider the daily returns of SPX in `d-aaspx9808.txt`. Transform the returns into log returns and focus on the daily negative log returns.

(a) Fit the generalized extreme value distribution to the negative SPX log returns, in percentage, with subperiods of 21 trading days. Write down the parameter estimates and their standard errors. Obtain a scatterplot and a QQ plot of the residuals.

(b) What is the return level of the prior fitted model when 24 subperiods of 21 days are used?

(c) Obtain a QQ plot (against exponential distribution) of the negative log returns with threshold 2.5% and a mean excess plot of the returns.

(d) Fit a generalize Pareto distribution to the negative log returns with threshold 2.5%. Write down the parameter estimates and their standard errors.

(e) Obtain (i) a plot of excess distribution, (ii) a plot of the tail of the underlying distribution, (iii) a scatterplot of residuals, and (iv) a QQ plot of the residuals for the fitted GPD.

(f) Based on the fitted GPD model, compute the VaR and expected shortfall for probabilities $q = 0.99$ and 0.999.

7.8. Consider the daily log returns of the GE stock of Exercise 7.1. Obtain estimates $\hat{\theta}_b^{(1)}$ and $\hat{\theta}_r^{(3)}$ of the extremal index of (a) the positive return series and (b) the negative return series, using block sizes $k = 5$ and 10 and threshold 2.5%.

REFERENCES

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, Hoboken, NJ.

Berman, S. M. (1964). Limiting theorems for the maximum term in stationary sequences. *Annals of Mathematical Statistics* **35**: 502–516.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Danielsson, J. and De Vries, C. G. (1997a). Value at risk and extreme returns. Working paper, London School of Economics, London, UK.

Danielsson, J. and De Vries, C. G. (1997b). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance* **4**: 241–257.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society Series B* **52**: 393–442.

Dekkers, A. L. M. and De Haan, L. (1989). On the estimation of extreme value index and large quantile estimation. *Annals of Statistics* **17**: 1795–1832.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *Journal of Derivatives* **Spring**: 7–48.

Embrechts, P., Kuppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Berlin.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2. Wiley, New York.

Goldie, C. M. and Smith, R. L. (1987). Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics* **38**: 45–71.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum of d'une série Aléatorie. *Annals of Mathematics* **44**: 423–453.

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**: 1163–1173.

Hsing, T., Hüsler, J. and Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* **78**: 97–112.

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

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. McGraw-Hill, Chicago.

Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles. *Econometrica* **46**: 33–50.

Koenker, R. W. and D'Orey, V. (1987). Computing regression quantiles. *Applied Statistics* **36**: 383–393.

Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitsthorie und Verwandte Gebiete* **28**: 289–303.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitsthorie und Verwandte Gebiete* **65**: 291–306.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

Longerstaey, J. and More, L. (1995). *Introduction to RiskMetrics*™, 4th ed. Morgan Guaranty Trust Company, New York.

Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns. *Journal of Business* **69**: 383–408.

Longin, F. M. (1999a). Optimal margin level in futures markets: Extreme price movements. *Journal of Futures Markets* **19**: 127–152.

Longin, F. M. (1999b). From value at risk to stress testing: The extreme value approach. Working paper, Centre for Economic Policy Research, London.

O'Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *Annals of Probability* **15**: 281–291.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**: 119–131.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science* **4**: 367–393.

Smith, R. L. (1999). Measuring risk with extreme value theory. Working paper, Department of Statistics, University of North Carolina at Chapel Hill.

Smith, R. L. and Shively, T. S. (1995). A point process approach to modeling trends in troposheric ozone. *Atmospheric Environment* **29**: 3489–3499.

Tsay, R. S. (1999). Extreme value analysis of financial data. Working paper, Graduate School of Business, University of Chicago.

Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus*. Springer, New York.