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 have attracted a great deal of attention among practitioners and researchers, and some people even called for government regulations on the derivative markets. In recent years, the seemingly large daily price movements in high-tech stocks have further generated discussions on market risk and margin setting for financial institutions. As a result, value at risk (VaR) has become a widely used measure of market risk in risk management.

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. However, a conditional approach uses the historical data and explanatory variables to calculate VaR.

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.

7.1 VALUE AT RISK

There are several types of risk in financial markets. Credit risk, liquidity risk, and market risk are three examples. Value at risk (VaR) is mainly concerned with market risk. It 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 (1997) 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 $\ell$ periods. Let $\Delta V(\ell)$ be the change in value of the assets in the financial position from time $t$ to $t + \ell$. This quantity is measured in dollars and is a random variable at the time index $t$. Denote the cumulative distribution function (CDF) of $\Delta V(\ell)$ by $F_\ell(x)$. We define the VaR of a long position over the time horizon $\ell$ with probability $p$ as

$$ p = \Pr[\Delta V(\ell) \leq \text{VaR}] = F_\ell(\text{VaR}). $$

(7.1)
Since the holder of a long financial position suffers a loss when $\Delta V(\ell) < 0$, the VaR defined in Eq. (7.1) typically assumes a negative value when $p$ is small. The negative sign signifies a loss. From the definition, the probability that the holder would encounter a loss greater than or equal to VaR over the time horizon $\ell$ is $p$. Alternatively, VaR can be interpreted as follows. With probability $(1 - p)$, the potential loss encountered by the holder of the financial position over the time horizon $\ell$ is less than or equal to VaR.

The holder of a short position suffers a loss when the value of the asset increases [i.e., $\Delta V(\ell) > 0$]. The VaR is then defined as

$$p = \Pr[\Delta V(\ell) \geq \text{VaR}] = 1 - \Pr[\Delta V(\ell) \leq \text{VaR}] = 1 - F_\ell(\text{VaR}).$$

For a small $p$, the VaR of a short position typically assumes a positive value. The positive sign signifies a loss.

The previous definitions show that VaR is concerned with tail behavior of the CDF $F_\ell(x)$. For a long position, the left tail of $F_\ell(x)$ is important. Yet a short position focuses on the right tail of $F_\ell(x)$. Notice that the definition of VaR in Eq. (7.1) continues to apply to a short position if one uses the distribution of $-\Delta V(\ell)$. Therefore, it suffices to discuss methods of VaR calculation using a long position.

For any univariate CDF $F_\ell(x)$ and probability $p$, such that $0 < p < 1$, the quantity

$$x_p = \inf\{x \mid F_\ell(x) \geq p\}$$

is called the $p$th quantile of $F_\ell(x)$, where $\inf$ denotes the smallest real number satisfying $F_\ell(x) \geq p$. If the CDF $F_\ell(x)$ of Eq. (7.1) is known, then VaR is simply its $p$th quantile (i.e., VaR = $x_p$). The CDF is unknown in practice, however. Studies of VaR are essentially concerned with estimation of the CDF and/or its quantile, especially the tail behavior of the CDF.

In practical applications, calculation of VaR involves several factors:

1. The probability of interest $p$, such as $p = 0.01$ or $p = 0.05$.
2. The time horizon $\ell$. It might be set by a regulatory committee, such as 1 day or 10 days.
3. The frequency of the data, which might not be the same as the time horizon $\ell$. Daily observations are often used.
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 in dollar amount. Since log returns correspond approximately to percentage changes in value of a financial position,
we use log returns $r_t$ in data analysis. The VaR calculated from the 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.

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.

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 $r_t \mid F_{t-1} \sim N(\mu_t, \sigma^2_t)$, where $\mu_t$ is the conditional mean and $\sigma^2_t$ is the conditional variance of $r_t$. In addition, the method assumes that the two quantities evolve over time according to the simple model:

$$
\mu_t = 0, \quad \sigma^2_t = \alpha \sigma^2_{t-1} + (1 - \alpha)r^2_{t-1}, \quad 1 > \alpha > 0. \quad (7.2)
$$

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 a drift. The value of $\alpha$ is often in the interval $(0.9, 1)$.

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 $r_t[k] \mid F_t$ is normal with mean zero and variance $\sigma^2_t[k]$, where $\sigma^2_t[k]$ can be computed using the forecasting method discussed in Chapter 3. Using the independence assumption of $\epsilon_t$ and model (7.2), we have

$$
\sigma^2_t[k] = \text{Var}(r_t[k] \mid F_t) = \sum_{i=1}^{k} \text{Var}(a_{t+i} \mid F_t),
$$

where $\text{Var}(a_{t+i} \mid F_t) = E(\sigma^2_{t+i} \mid F_t)$ 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
\[ \sigma_{t+i}^2 = \sigma_{t+i-1}^2 + (1 - \alpha)\sigma_{t+i-1}^2(\epsilon_{t+i-1}^2 - 1) \quad \text{for } i = 2, \ldots, k. \]

Since \( E(\epsilon_{t+i-1}^2 - 1 \mid F_t) = 0 \) for \( i \geq 2 \), the prior equation shows that
\[ E(\sigma_{t+i}^2 \mid F_t) = E(\sigma_{t+i-1}^2 \mid F_t) \quad \text{for } i = 2, \ldots, k. \]

For the 1-step ahead volatility forecast, Eq. (7.2) shows that
\[ \sigma_{t+1}^2 = \alpha \sigma_t^2 + (1 - \alpha)\sigma_t^2. \]

Therefore, Eq. (7.3) shows that \( \text{Var}(r_{t+i} \mid F_t) = \sigma_{t+i+1}^2 \) for \( i \geq 1 \) and hence, \( \sigma_t^2[k] = k \sigma_{t+1}^2 \). The results show that \( r_t[k] \mid F_t \sim N(0, k \sigma_{t+1}^2) \). 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 \( \sqrt{k} \sigma_{t+1} \).

Suppose that the financial position is a long position so that loss occurs when there is a big price drop (i.e., a large negative return). If the probability is set to 5%, then RiskMetrics uses \( 1.65 \sigma_{t+1} \) to measure the risk of the portfolio—that is, it uses the one-sided 5% quantile of a normal distribution with mean zero and standard deviation \( \sigma_{t+1} \). The actual 5% quantile is \( -1.65 \sigma_{t+1} \), but the negative sign is ignored with the understanding that it signifies a loss. Consequently, if the standard deviation is measured in percentage, then the daily VaR of the portfolio under RiskMetrics is
\[ \text{VaR} = \text{Amount of Position} \times 1.65 \sigma_{t+1}, \]
and that of a \( k \)-day horizon is
\[ \text{VaR}(k) = \text{Amount of Position} \times 1.65 \sqrt{k} \sigma_{t+1}, \]
where the argument \( (k) \) of VaR is used to denote the time horizon. Consequently, under RiskMetrics, we have
\[ \text{VaR}(k) = \sqrt{k} \times \text{VaR}. \]

This is referred to as the \textit{square root of time rule} in VaR calculation under RiskMetrics.

\textbf{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
The investor is

\[ 10,000,000 \times (1.65 \times 0.0053) = 87,450. \]

The corresponding VaR for 1-month horizon (30 days) is

\[ 10,000,000 \times (\sqrt{30} \times 1.65 \times 0.0053) \approx 478,983. \]

**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 a drift. The fitted model is

\[ r_t = a_t, \quad a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = 0.9396 \sigma_{t-1}^2 + (1 - 0.9396) a_{t-1}^2, \quad (7.4) \]

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 5% quantile of the conditional distribution \(r_{9191} | F_{9190}\) is \(-1.65 \times \sqrt{0.000336} = -0.03025\), where it is understood that the negative sign signifies a loss. Consequently, the 1-day horizon 5% VaR of a long position of $10 million is

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

The 1% quantile is \(-2.3262 \times \sqrt{0.000336} = -0.04265\), and the corresponding 1% VaR for the same long position is $426,500.

### 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:

\[ r_t = \mu + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \mu \neq 0 \]
\[ \sigma_t^2 = \alpha \sigma_{t-1}^2 + (1 - \alpha) a_{t-1}^2, \]
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 \( N(\mu, \sigma^2_{t+1}) \). The 5% quantile used to calculate the 1-period horizon VaR becomes \( \mu - 1.65\sigma_{t+1} \). For a \( k \)-period horizon, the distribution of \( r_k \) given \( F_t \) is \( N(k\mu, k\sigma^2_{t+1}) \), where as before \( r_k = r_t + \cdots + r_{t+k} \). The 5% quantile used in \( k \)-period horizon VaR calculation is\( k\mu - 1.65\sqrt{k}\sigma_{t+1} = \sqrt{k}(\mu - 1.65\sigma_{t+1}) \). 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 a 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 \( \text{VaR}_1 \) and \( \text{VaR}_2 \) be the VaR for the two positions and \( \rho_{12} \) be the cross-correlation coefficient between the two returns—that is,

\[
\rho_{12} = \frac{\text{Cov}(r_{1t}, r_{2t})}{\sqrt{\text{Var}(r_{1t})\text{Var}(r_{2t})}}.
\]

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 \( \rho_{ij} \) is the cross-correlation coefficient between returns of the \( i \)th and \( j \)th instruments and \( \text{VaR}_i \) is the VaR of the \( i \)th instrument.

### 7.3 An Econometric Approach to VaR Calculation

A general approach to VaR calculation is to use the time-series econometric models of Chapters 2 to 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

\[
r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}
\]

(7.5)

\[
a_t = \sigma_t \epsilon_t
\]

\[
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i a_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2
\]

(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

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

\[
\hat{\sigma}_t^2(1) = \alpha_0 + \sum_{i=1}^{u} \alpha_i \hat{a}_{t+1-i}^2 + \sum_{j=1}^{v} \beta_j \hat{\sigma}_{t+1-j}^2
\]

If one further assumes that \( \epsilon_t \) 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 5% quantile is \( \hat{r}_t(1) - 1.65 \hat{\sigma}_t(1) \). If one assumes that \( \epsilon_t \) is a standardized Student-\( t \) distribution with \( v \) degrees of freedom, then the quantile is \( \hat{r}_t(1) - t_v^*(p)\hat{\sigma}_t(1) \), where \( t_v^*(p) \) is the \( p \)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

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

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

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

where \( t_v(p) \) is the \( p \)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.

Case 1
Assume that $\epsilon_t$ is standard normal. The fitted model is

$$r_t = 0.00066 - 0.0247r_{t-2} + a_t, \quad a_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = 0.00000389 + 0.0799a_{t-1}^2 + 0.9073\sigma_{t-1}^2.$$ 

From the data, we have $r_{9189} = -0.00201$, $r_{9190} = -0.0128$, and $\sigma_{9190}^2 = 0.0003455$. 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 5% quantile is then

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

where it is understood that the negative sign denotes left tail of the conditional normal distribution. The VaR for a long position of $10$ million with probability 0.05 is $\text{VaR} = 10,000,000 \times 0.02877 = 287,700$. 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 probability is 0.01, then the 1% quantile is

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

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

Case 2
Assume that $\epsilon_t$ is a standardized Student-$t$ distribution with 5 degrees of freedom. The fitted model is

$$r_t = 0.0003 - 0.0335r_{t-2} + a_t, \quad a_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = 0.000003 + 0.0559a_{t-1}^2 + 0.9350\sigma_{t-1}^2.$$ 

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 5% 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 5% 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 1% 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 I, 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.

### 7.3.1 Multiple Periods

Suppose that at time $h$ we like 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 $E(r_h[k] \mid F_h)$ 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 $\ell$-step ahead forecast of the return at the forecast origin $h$. These forecasts can be computed recursively as discussed in subsection 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 $\ell$-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. (2.30) 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

$$e_h[k] = e_h(1) + e_h(2) + \cdots + e_h(k)$$

$$= a_{h+1} + (a_{h+2} + \psi_1 a_{h+1}) + \cdots + \sum_{i=0}^{k-1} \psi_i a_{h+k-i}$$

$$= a_{h+k} + (1 + \psi_1) a_{h+k-1} + \cdots + \left(\sum_{i=0}^{k-1} \psi_i\right) a_{h+1}$$

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

$$\text{Var}(e_h[k] | F_h) = \text{Var}(a_{h+k} | F_h) + (1 + \psi_1)^2 \text{Var}(a_{h+k-1} | F_h) + \cdots$$

$$+ \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \text{Var}(a_{h+1} | F_h)$$

$$= \sigma_h^2(k) + (1 + \psi_1)^2 \sigma_h^2(k-1) + \cdots + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \sigma_h^2(1),$$

where $\sigma_h^2(\ell)$ is the $\ell$-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

$$r_t = \mu + a_t, \quad a_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

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

$$\text{Var}(e_h[k] | F_h) = \sum_{\ell=1}^{k} \sigma_h^2(\ell).$$
Using the forecasting method of GARCH(1, 1) models in Section 3.4, we have

\[
\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2 \\
\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_h^2(\ell - 1), \quad \ell = 2, \ldots, k.
\]

Thus, \( \text{Var}(e_h[k] | F_h) \) can be obtained by the prior recursion. If \( \epsilon_t \) is Gaussian, then the conditional distribution of \( r_h[k] \) given \( F_h \) is normal with mean \( k\mu \) and variance \( \text{Var}(e_h[k] | F_h) \). The quantiles needed in VaR calculation are readily available.

**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. (7.8). The 5\% quantile of the conditional distribution is then 0.00998 $-$ 1.6449 \( \sqrt{0.0047948} \) $= -0.1039191$. Consequently, the 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.

### 7.4 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 \( i \)th 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 \( p \)th quantile of \( F(x) \) [i.e., \( x_p = F^{-1}(p) \)]. 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,

\[
r(\ell) \sim N \left[x_p, \frac{p(1 - p)}{nf^2(x_p)}\right], \quad \ell = np.
\] (7.9)

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 \( \ell_1 \) and \( \ell_2 \) 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 \( x_{pi} \). From the definition, \( p_1 < p < p_2 \). Therefore, the quantile \( x_p \) can be estimated by

\[
\hat{x}_p = \frac{p_2 - p}{p_2 - p_1}r(\ell_1) + \frac{p - p_1}{p_2 - p_1}r(\ell_2).
\] (7.10)

**Example 7.4.** Consider the daily log returns of Intel stock from December 15, 1972 to December 31, 1997. There are 6329 observations. The empirical 5% quantile of the data can be obtained as

\[
\hat{x}_{0.05} = 0.55r(316) + 0.45r(317) = -4.229\%,
\]

where \( np = 6329 \times 0.05 = 316.45 \) and \( r(i) \) is the \( i \)th order statistic of the sample. In this particular instance, \( r(316) = -4.237\% \) and \( r(317) = -4.220\% \). Here we use the lower tail of the empirical distribution because it is relevant to holding a long position in VaR calculation.

**Example 7.5.** Consider again the daily log returns of IBM stock from July 3, 1962 to December 31, 1998. Using all the 9190 observations, the empirical 5% quantile can be obtained as \( (r(459) + r(460))/2 = -0.021603 \), where \( r(i) \) is the \( i \)th order statistic and \( np = 9190 \times 0.05 = 459.5 \). 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 91 < 9190 × 0.01 < 92. Let \( p_1 = 91/9190 = 0.0099 \) and \( p_2 = 92/9190 = 0.01001 \). The empirical 1% quantile can be obtained as
\[
\hat{x}_{0.01} = \frac{p_2 - 0.01}{p_2 - p_1} r_{(91)} + \frac{0.01 - p_1}{p_2 - p_1} r_{(92)} \\
= \frac{0.00001}{0.00011}(-3.658) + \frac{0.0001}{0.00011}(-3.657) \\
\approx -3.657.
\]

The 1% 1-day horizon VaR of the long position is $365,709. Again, this amount is lower than those obtained before by other methods.

**Discussion:** Advantages of using the prior 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, for extreme quantiles (i.e., when \( p \) is close to zero or unity), the empirical quantiles are not efficient estimates of the theoretical quantiles. 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.

### 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 \( r_{t+1} \mid F_t \), 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 \( p \)th quantile of \( \{r_t\} \) is obtained by

\[
\hat{x}_p = \arg\min_\beta \sum_{i=1}^n w_p(r_i - \beta),
\]

where \( w_p(z) \) is defined by

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

Regression quantile is a generalization of such an estimate.
To see the generalization, suppose that we have the linear regression
\[ r_t = \beta' x_t + a_t, \]
where $\beta$ is a $k$-dimensional vector of parameters and $x_t$ 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 $\beta' x_t$ is known. Viewing the problem this way, Koenker and Bassett (1978) suggest to estimate the conditional quantile $x_p \mid F_{t-1}$ of $r_t$ given $F_{t-1}$ as
\[ \hat{x}_p \mid F_{t-1} = \inf\{\beta' x \mid R_p(\beta_o) = \min\}, \]
where “$R_p(\beta_o) = \min$” means that $\beta_o$ is obtained by
\[ \beta_o = \arg\min_\beta \sum_{t=1}^n w_p(r_t - \beta' x_t), \]
where $w_p(.)$ is defined as before. A computer program to obtain such an estimated quantile can be found in Koenker and D’Orey (1987).

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, $r_{(1)} = \min_{1 \leq j \leq n} \{r_j\}$ and $r_{(n)} = \max_{1 \leq j \leq n} \{r_j\}$. We focus on properties of the minimum return $r_{(1)}$ because this minimum is highly relevant to VaR calculation for a long position. However, the theory discussed also applies to the maximum return of an asset over a given time period because properties of the maximum return can be obtained from those of the minimum by a simple sign change. Specifically, we have $r_{(n)} = -\min_{1 \leq j \leq n} \{-r_j\} = -r_{(1)}$, where $r_{(1)}^c = -r_t$ with the superscript $c$ denoting sign change. The maximum return is relevant to holding a short financial 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 $r_{(1)}$, denoted by $F_{n,1}(x)$, is given by
\[
F_{n,1}(x) = \Pr[r_{(1)} \leq x] = 1 - \Pr[r_{(1)} > x]
= 1 - \Pr(r_1 > x, r_2 > x, \ldots, r_n > x)
\]
\[ 1 - \prod_{j=1}^{n} \Pr(r_j > x), \quad \text{(by independence)} \]

\[ = 1 - \prod_{j=1}^{n} [1 - \Pr(r_j \leq x)] \]

\[ = 1 - \prod_{j=1}^{n} [1 - F(x)] \quad \text{(by common distribution)} \]

\[ = 1 - [1 - F(x)]^{n}. \quad (7.13) \]

In practice, the CDF \( F(x) \) of \( r_i \) is unknown and, hence, \( F_{n,1}(x) \) of \( r_{(1)} \) is unknown. However, as \( n \) increases to infinity, \( F_{n,1}(x) \) becomes degenerated—namely, \( F_{n,1}(x) \to 0 \) if \( x \leq l \) and \( F_{n,1}(x) \to 1 \) if \( x > l \) 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 \( r_{(1^*)} \equiv \left( r_{(1)} - \beta_n \right)/\alpha_n \) converges to a nondegenerated 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 \( r_{(1^*)} \) is given by

\[ F_*(x) = \begin{cases} 
1 - \exp\left(-(1/k)k \right) & \text{if } k \neq 0 \\
1 - \exp[-\exp(x)] & \text{if } k = 0 
\end{cases} \quad (7.14) \]

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

The limiting distribution in Eq. (7.14) is the generalized extreme value distribution of Jenkinson (1955) for the minimum. It encompasses the three types of limiting distribution of Gnedenko (1943):

- **Type I**: \( k = 0 \), the Gumbel family. The CDF is
  \[ F_*(x) = 1 - \exp[-\exp(x)], \quad -\infty < x < \infty. \quad (7.15) \]

- **Type II**: \( k < 0 \), the Fréchet family. The CDF is
  \[ F_*(x) = \begin{cases} 
1 - \exp\left(-(1/k)k \right) & \text{if } x < -1/k \\
1 & \text{otherwise.} 
\end{cases} \quad (7.16) \]

- **Type III**: \( k > 0 \), the Weibull family. The CDF here is
  \[ F_*(x) = \begin{cases} 
1 - \exp\left(-(1/k)k \right) & \text{if } x > -1/k \\
0 & \text{otherwise.} 
\end{cases} \]
Figure 7.2. Probability density functions of extreme value distributions for minimum: The solid line is for a Gumbel distribution, the dotted line is for the Weibull distribution with \( k = 0.5 \), and the dashed line for the Fréchet distribution with \( k = -0.9 \).

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 minimum. The (left) 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. Readers are referred to Embrechts, Kuppelberg, and Mikosch (1997) for a comprehensive treatment of the extreme value theory. For risk management, we are mainly interested in the Fréchet family that includes stable and Student-\( t \) distributions. The Gumbel family consists of thin-tailed distributions such as normal and log-normal distributions. The probability density function (pdf) of the generalized limiting distribution in Eq. (7.14) can be obtained easily by differentiation:

\[
    f_*(x) = \begin{cases} 
        (1 + kx)^{1/k-1} \exp[-(1 + kx)^{1/k}] & \text{if } k \neq 0 \\
        \exp[x - \exp(x)] & \text{if } k = 0,
    \end{cases} \tag{7.17}
\]

where \(-\infty < x < \infty \) for \( k = 0 \), \( x < -1/k \) for \( k < 0 \), and \( x > -1/k \) for \( k > 0 \).

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) minimum. 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 $k$ 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 \( \{r_t\}_{t=1}^\infty \) provided that the dependence is weak. Berman (1964) shows that the same form of the limiting extreme value distribution holds for stationary normal sequences provided that the autocorrelation function of $r_t$ is squared summable (i.e., $\sum_{i=1}^{\infty} \rho_i^2 < \infty$), where $\rho_i$ is the lag-$i$ autocorrelation function of $r_t$. For further results concerning the effect of serial dependence on the extreme value theory, readers are referred to Leadbetter, Lindgren, and Rootzén (1983, Chapter 3).

### 7.5.2 Empirical Estimation

The extreme value distribution contains three parameters—$k$, $\beta_n$, and $\alpha_n$. These parameters are referred to as the shape, location, and scale parameter, 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 sub-samples 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 $g$ non-overlapping sub-samples each with $n$ observations, assuming for simplicity that $T = ng$. In other words, we divide the data as

\[
\{r_1, \ldots, r_n \mid r_{n+1}, \ldots, r_{2n} \mid r_{2n+1}, \ldots, r_{3n} \mid \cdots \mid r_{(g-1)n+1}, \ldots, r_{ng}\}
\]

and write the observed returns as $r_{in+j}$, where $1 \leq j \leq n$ and $i = 0, \ldots, g - 1$.

Notice 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 minimum of the $i$th subsample (i.e., $r_{n,i}$ is the smallest return of the $i$th subsample), where the subscript $n$ is used to denote the size of the subsample. When $n$ is sufficiently large, $x_{n,i} = (r_{n,i} - \beta_n)/\alpha_n$ should follow an extreme value distribution, and the collection of subsample minima $\{r_{n,i} \mid i = 1, \ldots, g\}$ can then be regarded as a sample of $g$ observations from that extreme value distribution. Specifically, we define

\[
r_{n,i} = \min_{1 \leq j \leq n} \{r_{(i-1)n+j}\}, \quad i = 1, \ldots, g.
\]

The collection of subsample minima $\{r_{n,i}\}$ are 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 $n$. 

---

**References**

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

Maximum likelihood method
Assuming that the subperiod minima \( \{ r_{n,i} \} \) follow a generalized extreme value distribution such that the pdf of \( x_i = (r_{n,i} - \beta_n) / \alpha_n \) is given in Eq. (7.17), we can obtain the pdf of \( r_{n,i} \) by a simple transformation as

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

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

\[
\ell(r_{n,1}, \ldots, r_{n,g} \mid k_n, \alpha_n, \beta_n) = \prod_{i=1}^{g} f(r_{n,i}).
\]

Nonlinear estimation procedures can then be used to obtain maximum likelihood estimates of \( k_n, \beta_n, \) and \( \alpha_n \). These estimates are unbiased, asymptotically normal, and of minimum variance under proper assumptions. 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. (7.14) and make uses of properties of order statistics; see Gumbel (1958). Denote the order statistics of the subperiod minima \( \{ r_{n,i} \}_{i=1}^{g} \) 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

\[
E\{ F_\ast [r_{n(i)}] \} = \frac{i}{g + 1}, \quad i = 1, \ldots, g.
\] (7.19)

For simplicity, we separate the discussions into two cases depending on the value of \( k \). First, consider the case of \( k \neq 0 \). From Eq. (7.14), we have
\[
F_n[r_{n(i)}] = 1 - \exp \left\{ - \left[ 1 + k_n \frac{r_{n(i)} - \beta_n}{\alpha_n} \right]^{1/k_n} \right\}
\]  
(7.20)

Consequently, using Eqs. (7.19) and (7.20) and approximating expectation by an observed value, we have

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

Therefore,

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

Taking natural logarithm twice, the prior equation gives

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

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

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

The least squares estimates of \( k_n, \beta_n, \) and \( \alpha_n \) can be obtained by minimizing the sum of squares of \( e_i \).

When \( k_n = 0 \), the regression setup reduces to

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

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

7.5.2.2 The Nonparametric Approach

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

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

Let \( q \) be a positive integer. The two estimators of \( k \) are defined as
\[ k_p(q) = -\frac{1}{\ln(2)} \ln \left( \frac{-r_1 + r_2}{-r_2 + r_4} \right) \]  
(7.22)

\[ k_h(q) = -\frac{1}{q} \sum_{i=1}^{q} \{ \ln[-r_i] - \ln[-r_{i+1}] \}, \]  
(7.23)

where the argument \( q \) is used to emphasize that the estimators depend on \( q \). 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 \( k_p(q) \) is consistent if \( q \) increases at a properly chosen pace with the sample size \( T \). In addition, \( \sqrt{q}[k_p(q) - k] \) is asymptotically normal with mean zero and variance \( k^2(2^{-2k+1} + 1/[2(2^{-k} - 1) \ln(2)])^2 \). 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 \( \sqrt{q}[k_h(q) - k] \) is asymptotically normal with mean zero and variance \( k^2 \). In practice, one may plot the Hill estimator \( k_h(q) \) against \( q \) and find a proper \( q \) such that the estimate appears to be stable.

(a) Monthly maximum log returns

(b) Monthly minimum log returns

Figure 7.3. Maximum and minimum daily log returns of IBM stock when the subperiod is 21 trading days. The data span is from July 3, 1962 to December 31, 1998: (a) positive returns, and (b) negative returns.
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, 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 $k$ 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 Hill estimator, Figure 7.4 shows

<table>
<thead>
<tr>
<th>$q$</th>
<th>190</th>
<th>200</th>
<th>210</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>$-0.300 (0.022)$</td>
<td>$-0.297 (0.021)$</td>
<td>$-0.303 (0.021)$</td>
</tr>
<tr>
<td>Minimum</td>
<td>$-0.290 (0.021)$</td>
<td>$-0.292 (0.021)$</td>
<td>$-0.289 (0.020)$</td>
</tr>
</tbody>
</table>

Figure 7.4. Scatterplots of the Hill estimator for the daily log returns of IBM stock. The sample period is from July 3, 1962 to December 31, 1998: (a) positive returns, and (b) negative returns.
the scatter plots of the Hill estimator $k_h(q)$ against $q$. For both positive and negative extreme daily log returns, the estimator is stable except for cases when $q$ is small. The estimated shape parameters are about $-0.30$ and are significantly different from zero at the asymptotic 5% level. The plots also indicate that the shape parameter $k$ appears to be smaller 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.

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 from 1 month ($n = 21$) to 1 year ($n = 252$). From the table, we make the following observations:

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

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.


<table>
<thead>
<tr>
<th>Length of subperiod</th>
<th>Scale $\alpha_n$</th>
<th>Location $\beta_n$</th>
<th>Shape Par. $k_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Minimal returns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 mon. ($n = 21, g = 437$)</td>
<td>0.823(0.035)</td>
<td>-1.902(0.044)</td>
<td>-0.197(0.036)</td>
</tr>
<tr>
<td>1 qur ($n = 63, g = 145$)</td>
<td>0.945(0.077)</td>
<td>-2.583(0.090)</td>
<td>-0.335(0.076)</td>
</tr>
<tr>
<td>6 mon. ($n = 126, g = 72$)</td>
<td>1.147(0.131)</td>
<td>-3.141(0.153)</td>
<td>-0.330(0.101)</td>
</tr>
<tr>
<td>1 year ($n = 252, g = 36$)</td>
<td>1.542(0.242)</td>
<td>-3.761(0.285)</td>
<td>-0.322(0.127)</td>
</tr>
<tr>
<td>(b) Maximal returns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 mon. ($n = 21, g = 437$)</td>
<td>0.931(0.039)</td>
<td>2.184(0.050)</td>
<td>-0.168(0.036)</td>
</tr>
<tr>
<td>1 qur ($n = 63, g = 145$)</td>
<td>1.157(0.087)</td>
<td>3.012(0.108)</td>
<td>-0.217(0.066)</td>
</tr>
<tr>
<td>6 mon. ($n = 126, g = 72$)</td>
<td>1.292(0.158)</td>
<td>3.471(0.181)</td>
<td>-0.349(0.130)</td>
</tr>
<tr>
<td>1 year ($n = 252, g = 36$)</td>
<td>1.624(0.271)</td>
<td>4.475(0.325)</td>
<td>-0.264(0.186)</td>
</tr>
</tbody>
</table>
7.6 AN 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, 1999b), who proposed an eight-step procedure for the same purpose. We divide the discussion into two parts. The first part is concerned with parameter estimation using the method discussed in the previous subsections. The second part focuses on VaR calculation by relating the probabilities of interest associated with different time intervals.

**Part I**

Assume that there are \( T \) observations of an asset return available in the sample period. We partition the sample period into \( g \) nonoverlapping subperiods of length \( n \) such that \( T = ng \). If \( T = ng + m \) with \( 1 \leq m < n \), then we delete the first \( m \) observations from the sample. The extreme value theory discussed in the previous section enables us to obtain estimates of the location, scale, and shape parameters \( \beta_n, \alpha_n \), and \( k_n \) for the subperiod minima \( \{ r_{n,i} \} \). Plugging the MLE estimates into the CDF in Eq. (7.14) with \( x = (r - \beta_n)/\alpha_n \), we can obtain the quantile of a given probability of the generalized extreme value distribution. Because we focus on holding a long financial position, the lower probability (or left) quantiles are of interest. Let \( p^* \) be a small probability that indicates the potential loss of a long position and \( r_n^* \) be the \( p^* \)th quantile of the subperiod minimum under the limiting generalized extreme value distribution. Then we have

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

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

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

we obtain the quantile as

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

(7.24)

In financial applications, the case of \( k_n \neq 0 \) is of major interest.
Part II
For a given lower (or left tail) probability \( p^* \), the quantile \( r^*_n \) of Eq. (7.24) is the VaR based on the extreme value theory for the subperiod minima. The next step is to make explicit the relationship between subperiod minima 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. (7.13) and obtain

\[
p^* = P(r_{n,i} \leq r^*_n) = 1 - [1 - P(r_t \leq r^*_n)]^n
\]

or, equivalently,

\[
1 - p^* = [1 - P(r_t \leq r^*_n)]^n. \tag{7.25}
\]

This relationship between probabilities allows us to obtain VaR for the original asset return series \( r_t \). More precisely, for a specified small lower probability \( p \), the \( p \)th quantile of \( r_t \) is \( r^*_n \) if the probability \( p^* \) is chosen based on Eq. (7.25), where \( p = P(r_t \leq r^*_n) \). Consequently, for a given small probability \( p \), the VaR of holding a long position in the asset underlying the log return \( r_t \) is

\[
\text{VaR} = \begin{cases} 
\beta_n - \frac{\alpha_n}{k_n} \{1 - [-n \ln(1 - p)]^{k_n}\} & \text{if } k_n \neq 0 \\
\beta_n + \alpha_n \ln[-n \ln(1 - p)] & \text{if } k_n = 0.
\end{cases} \tag{7.26}
\]

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 minima \( \{r_{n,i}\} \), \( i = 1, \ldots, g \), where \( g = T/n \).
2. Obtain the maximum likelihood estimates of \( \beta_n, \alpha_n, \) and \( k_n \).
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. (7.26) to calculate VaR.

Remark: Since we focus on holding a long financial position and, hence, on the quantile in the left tail of a return distribution, the quantile is negative. Yet it is customary in practice to use a positive number for VaR calculation. Thus, in using Eq. (7.26), one should be aware that the negative sign signifies a loss.

Example 7.6. Consider the daily log return, in percentage, of IBM stock from July 7, 1962 to December 31, 1998. From Table 7.2, we have \( \hat{\alpha}_n = 0.945 \), \( \hat{\beta}_n = -2.583 \), and \( \hat{k}_n = -0.335 \) for \( n = 63 \). Therefore, for the left-tail probability \( p = 0.01 \), the corresponding VaR is
\[
\text{VaR} = -2.583 - \frac{0.945}{-0.335} \left\{ 1 - \left[ -63 \ln(1 - 0.01) \right]^{-0.335} \right\}
= -3.04969.
\]

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

If we chose \( n = 21 \) (i.e., approximately 1 month), then \( \hat{\alpha}_n = 0.823 \), \( \hat{\beta}_n = -1.902 \), and \( \hat{k}_n = -0.197 \). The 1% quantile of the extreme value distribution is
\[
\text{VaR} = -1.902 - \frac{0.823}{-0.197} \left\{ 1 - \left[ -21 \ln(1 - 0.01) \right]^{-0.197} \right\} = -3.40013.
\]

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

It is somewhat surprising to see that the VaR values obtained in Example 7.6 using the extreme value theory are smaller than those of Example 7.3 that uses a GARCH(1, 1) model. In fact, the VaR values of Example 7.6 are even smaller than those based on the empirical quantile in Example 7.5. This is due in part to the choice of probability 0.05. If one chooses probability 0.001 = 0.1% and considers the same financial position, then we have 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 minima overlooks the fact of volatility clustering in the daily log returns. The new approach of extreme value theory discussed in the next section overcomes these weaknesses.

**Remark:** As shown by the results of Example 7.6, the VaR calculation based on the traditional extreme value theory depends on the choice of \( n \), which is the length of subperiods. For the limiting extreme value distribution to hold, one would prefer a large \( n \). But a larger \( n \) means a smaller \( g \) when the sample size \( T \) is fixed, where \( g \) is the effective sample size used in estimating the three parameters \( \alpha_n \), \( \beta_n \), and \( k_n \). 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, and
5. $184,127 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length $n = 21$).

If the probability is 1%, then the VaR is

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

If the probability is 0.1%, then the VaR becomes

1. $566,443 for the RiskMetrics,
2. $546,641 for a Gaussian AR(2)-GARCH(1, 1) model,
3. $836,341 for an AR(2)-GARCH(1, 1) model with a standardized Student-$t$ distribution with 5 degrees of freedom,
4. $780,712 for using the empirical quantile, and
5. $666,590 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length $n = 21$).

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

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

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

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 $\ell$-day and 1-day horizons is

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

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

For illustration, consider the daily log returns of IBM stock in Example 7.6. If we use $p = 0.05$ and the results of $n = 21$, then for a 30-day horizon we have

$$\text{VaR}(30) = (30)^{0.335} \text{VaR} = 3.125 \times 184,127 = 575,397.$$ 

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

7.6.3 VaR for a Short Position

In this subsection, we give the formulas of VaR calculation for holding short positions. Here the quantity of interest is the subperiod maximum and the limiting extreme value distribution becomes

$$F^*_n(r) = \begin{cases} 
\exp \left\{ - \left[ 1 - \frac{k_n (r - \beta_n)}{\alpha_n} \right]^{1/k_n} \right\} & \text{if } k_n \neq 0 \\
\exp \left[ - \exp \left( \frac{r - \beta_n}{\alpha_n} \right) \right] & \text{if } k_n = 0,
\end{cases}$$

(7.27)

where $r$ denotes a value of the subperiod maximum and it is understood that $1 - k_n (r - \beta_n)/\alpha_n > 0$ for $k_n \neq 0$. 
Following similar procedures as those of long positions, we obtain the \((1 - p)\)th quantile of the return \(r_t\) as

\[
\text{VaR} = \begin{cases} 
\beta_n + \frac{\alpha_n}{k_n} \left[1 - \left(-n \ln(1 - p)\right)^{k_n}\right] & \text{if } k_n \neq 0 \\
\beta_n + \alpha_n \ln\left(-n \ln(1 - p)\right) & \text{if } k_n = 0.
\end{cases}
\] (7.28)

where \(p\) is a small probability denoting the chance of loss for holding a short position.

### 7.7 A NEW APPROACH BASED ON THE EXTREME VALUE THEORY

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

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

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

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

7.7.1 Statistical Theory

Again consider the log return $r_t$ of an asset. Suppose that the $i$th exceedance occurs at $t_i$. Focusing on the exceedance $r_t - \eta$ 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 $r_t \leq \eta$ for a long position. Occurrence of the event \{r$_t \leq$ \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 $\lambda$ of the process is time-invariant, then the Poisson process is homogeneous. If $\lambda$ is time-variant, then the process is nonhomogeneous. The concept of Poisson process can be generalized to the multivariate case.

For ease in presentation, in what follows we use a positive threshold and the right-hand side of a return distribution to discuss the statistical theory behind the new approach of extreme value theory. This corresponds to holding a short financial position. However, the theory applies equally well to holding a long position if it is applied to the $r^c_t$ series, where $r^c_t = -r_t$. This is easily seen because $r^c_t \geq \eta$ for a positive threshold is equivalent to $r_t \leq -\eta$, where $-\eta$ becomes a negative threshold.

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. (7.27). Since there is no need to choose the subperiod length $n$, we do not use it as a subscript of the parameters. Then the conditional distribution of $r \leq x + \eta$ given $r > \eta$ is

$$
\Pr(r \leq x + \eta \mid r > \eta) = \frac{\Pr(\eta \leq r \leq x + \eta)}{\Pr(r > \eta)} = \frac{\Pr(r \leq x + \eta) - \Pr(r \leq \eta)}{1 - \Pr(r \leq \eta)}.
$$

(7.29)

Using the CDF $F_*(.)$ of Eq. (7.27) and the approximation $e^{-y} \approx 1 - y$ and after some algebra, we obtain that

$$
\Pr(r \leq x + \eta \mid r > \eta) = \frac{F_*(x + \eta) - F_*(\eta)}{1 - F_*(\eta)}
= \frac{\exp \left\{ - \left( 1 - \frac{k(x+\eta-\beta)}{\alpha} \right)^{1/k} \right\} - \exp \left\{ - \left( 1 - \frac{k(\eta-\beta)}{\alpha} \right)^{1/k} \right\}}{1 - \exp \left\{ - \left( 1 - \frac{k(\eta-\beta)}{\alpha} \right)^{1/k} \right\}}
$$
\[ \approx 1 - \left[ 1 - \frac{kx}{\alpha - k(\eta - \beta)} \right]^{1/k}, \] (7.30)

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

\[ \Pr(r \leq x + \eta \mid r > \eta) \approx 1 - \exp\left(-\frac{x}{\alpha}\right). \]

### 7.7.2 A New Approach

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 \((t_i, r_{ti})\). 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 \( T \), which is typically a year. In the United States, \( T = 252 \) 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 \( t = 1, 2, \ldots, N \), where \( N \) is the total number of data points. For a given threshold \( \eta \), the exceeding times over the threshold are denoted by \( \{t_i, i = 1, \ldots, N_\eta\} \) and the observed log return at \( t_i \) is \( r_{ti} \). Consequently, we focus on modeling \( \{(t_i, r_{ti})\} \) for \( i = 1, \ldots, N_\eta \), where \( N_\eta \) depends on the threshold \( \eta \).

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

\[ \Lambda[(T_2, T_1) \times (r, \infty)] = \frac{T_2 - T_1}{T} S(r; k, \alpha, \beta), \] (7.31)

where

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

\( 0 \leq T_1 \leq T_2 \leq N, r > \eta, \alpha > 0, \beta \), and \( k \) are parameters, and the notation \( [x]_+ = \max(x, 0) \). This intensity measure says that the occurrence of exceeding the threshold is proportional to the length of the time interval \([T_1, T_2]\) and the probability is governed by a survival function similar to the exponent of the CDF \( F_\ast(r) \) in Eq. (7.27). A survival function of a random variable \( X \) is defined as \( S(x) = \Pr(X > x) = 1 - \Pr(X \leq x) = 1 - \text{CDF}(x) \). When \( k = 0 \), the intensity measure is taken as the limit of \( k \to 0 \)—that is,

\[ \Lambda[(T_2, T_1) \times (r, \infty)] = \frac{T_2 - T_1}{T} \exp\left[-\frac{(r - \beta)}{\alpha}\right]. \]
In Eq. (7.31), the length of time interval is measured with respect to the baseline interval $T$.

The idea of using the intensity measure in Eq. (7.31) becomes clear when one considers its implied conditional probability of $r = x + \eta$ given $r > \eta$ over the time interval $[0, T]$, where $x > 0$,

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

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. (7.27). We use survival function here because it denotes the probability of exceedance.

The relationship between the limiting extreme value distribution in Eq. (7.27) and the intensity measure in Eq. (7.31) directly connects the new approach of extreme value theory to the traditional one.

Mathematically, the intensity measure in Eq. (7.31) can be written as an integral of an intensity function:

$$\Lambda([T_2, T_1] \times (r, \infty)) = \int_{T_1}^{T_2} \int_r^\infty \lambda(t, z; k, \alpha, \beta) \, dt \, dz,$$

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

$$\lambda(t, z; k, \alpha, \beta) = \frac{1}{T} g(z; k, \alpha, \beta)$$ (7.32)

where

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

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

$$L(k, \alpha, \beta) = \left(\prod_{i=1}^{N_\eta} \frac{1}{T} g(r_i; k, \alpha, \beta)\right) \times \exp\left[-\frac{N}{T} S(\eta; k, \alpha, \beta)\right].$$ (7.33)

The parameters $k, \alpha, \beta$ can then be estimated by maximizing the logarithm of this likelihood function. Since the scale parameter $\alpha$ is nonnegative, we use $\ln(\alpha)$ in the estimation.

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

(b) Removing the sample mean

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

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 estimation results of the parameters $k, \alpha, \beta$ for three choices of the threshold when the negative series $\{-r_t\}$ is used. 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 subsection.

7.7.3 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. (7.31), has the same parameters as those of the extreme value distribution in Eq. (7.27). Therefore, one can use the same formula as that of the 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

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

(7.34)

where $T$ is the baseline time interval used in estimation. Typically, $T = 252$ in the United States for the approximate 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 $\eta$ result in the following VaR values:

  1. $\eta = 3.0\%$: VaR(5%) = $228,239, VaR(1\%) = $359,303.
  2. $\eta = 2.5\%$: VaR(5%) = $219,106, VaR(1\%) = $361,119.
  3. $\eta = 2.0\%$: VaR(5%) = $212,981, VaR(1\%) = $368,552.

- Case II: The sample mean of the daily log returns is removed. The three choices of threshold $\eta$ result in the 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, increases slightly 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 $n$.

7.7.4 Use of Explanatory Variables

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

Suppose that $x_t = (x_{1t}, \ldots, x_{vt})'$ is a vector of $v$ explanatory variables that are available prior to time $t$. For asset returns, the volatility $\sigma_t^2$ 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 Federal Open Market Committee. A simple way to make use of explanatory variables is to postulate that the three parameters $k$, $\alpha$, and $\beta$ are time-varying and are linear functions of the explanatory variables. Specifically, when explanatory variables $x_t$
are available, we assume that
\[ k_t = \gamma_0 + \gamma_1 x_{1t} + \cdots + \gamma_v x_{vt} \equiv \gamma_0 + \gamma' x_t, \]
\[ \ln(\alpha_t) = \delta_0 + \delta_1 x_{1t} + \cdots + \delta_v x_{vt} \equiv \delta_0 + \delta' x_t \] (7.35)
\[ \beta_t = \theta_0 + \theta_1 x_{1t} + \cdots + \theta_v x_{vt} \equiv \theta_0 + \theta' x_t. \]

If \( \gamma = 0 \), then the shape parameter \( k_t = \gamma_0 \), which is time-invariant. Thus, testing the significance of \( \gamma \) can provide information about the contribution of the explanatory variables to the shape parameter. Similar methods apply to the scale and location parameters. In Eq. (7.35), we use the same explanatory variables for all the three parameters \( k_t \), \( \ln(\alpha_t) \), and \( \beta_t \). In an application, different explanatory variables may be used for different parameters.

When the three parameters of the extreme value distribution are time-varying, we have an inhomogeneous Poisson process. The intensity measure becomes
\[ \Lambda((T_1, T_2) \times (r, \infty)) = \frac{T_2 - T_1}{T} \left( 1 - \frac{k_t(r - \beta_t)}{\alpha_t} \right)^{1/k_t}, \quad r > \eta. \] (7.36)

The likelihood function of the exceeding times and returns \( \{(t_i, r_{ti})\} \) becomes
\[ L = \left( \prod_{i=1}^{N} \frac{1}{T} g(r_{ti}; k_{ti}, \alpha_{ti}, \beta_{ti}) \right) \times \exp \left[ -\frac{1}{T} \int_0^T S(\eta; k_t, \alpha_t, \beta_t)dt \right], \]
which reduces to
\[ L = \left( \prod_{i=1}^{N} \frac{1}{T} g(r_{ti}; k_{ti}, \alpha_{ti}, \beta_{ti}) \right) \times \exp \left[ -\frac{1}{T} \sum_{t=1}^{N} S(\eta; k_t, \alpha_t, \beta_t) \right] \] (7.37)
if one assumes that the parameters \( k_t, \alpha_t, \) and \( \beta_t \) are constant within each trading day, where \( g(z; k_t, \alpha_t, \beta_t) \) and \( S(\eta; k_t, \alpha_t, \beta_t) \) are given in Eqs. (7.32) and (7.31), respectively. For given observations \( \{r_t, x_t \mid t = 1, \ldots, N\} \), the baseline time interval \( T \), and the threshold \( \eta \), the parameters in Eq. (7.35) can be estimated by maximizing the logarithm of the likelihood function in Eq. (7.37). Again we use \( \ln(\alpha_t) \) to satisfy the positive constraint of \( \alpha_t \).

**Remark:** The parameterization in Eq. (7.35) 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.

### 7.7.5 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 to examine the time durations between consecutive exceedances. If the two-dimensional Poisson process model is appropriate for the exceedance times and excesses, the time duration between the \(i\)th and \((i - 1)\)th exceedances should follow an exponential distribution. More specifically, letting \(t_0 = 0\), we expect that

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

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

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

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

**Distribution of Excesses**

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

\[
w_{ti} = \begin{cases} 
- \frac{1}{k_t} \ln \left( 1 - k_t \frac{r_{ti} - \eta}{\psi_{ti}} \right) & \text{if } k_t \neq 0 \\
\frac{r_{ti} - \eta}{\psi_{ti}} & \text{if } k_t = 0. 
\end{cases}
\]

(7.39)

If the model is adequate, \(\{w_{ti}\}\) 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 zero serial correlations for both $z_{t_i}$ and $w_{t_i}$.

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

We begin by pointing out that the two-dimensional homogeneous model of Example 7.7 needs further refinements because the fitted model fails to pass the model checking statistics of the previous subsection. Figures 7.5(a) and (b) show the autocorrelation functions for the homogeneous model and Figures 7.5(c) and (d) show the autocorrelation functions for the inhomogeneous model.

**Figure 7.5.** Sample autocorrelation functions of the $z$ and $w$ measures for two-dimensional Poisson models. (a) and (b) are for the homogeneous model, and (c) and (d) are for the inhomogeneous model. The data are daily mean-corrected log returns, in percentages, of IBM stock from July 3, 1962 to December 31, 1998, and the threshold is 2.5%. A long financial position is used.
correlation functions of the statistics $z_{ti}$ and $w_{ti}$, defined in Eqs. (7.38) and (7.39), 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 $z_{ti}$ and $w_{ti}$ series have some significant serial correlations. Figures 7.6(a) and (b) show the QQ-plots of $z_{ti}$ and $w_{ti}$ 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 $z_{ti}$ shows some discrepancy.

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

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

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

![Figure 7.6](image-url)
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 $r_{t-1}^o \leq -2.5\%$. Since we focus on holding a long position with threshold 2.5%, an exceedance occurs when the daily price drops over 2.5%. Therefore, $x_{2t}$ is used to capture the possibility of panic selling when the price of IBM stock dropped 2.5% or more on the previous trading day.

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

4. $x_{4t}$: an annual trend defined as $x_{4t} = (\text{year of time } t - 1961)/38$. This variable is used to detect any trend in the behavior of extreme returns of IBM stock.

5. $x_{5t}$: a volatility series based on a Gaussian GARCH($1, 1$) model for the mean-corrected series $r_t^o$. Specifically, $x_{5t} = \sigma_t$, where $\sigma_t^2$ is the conditional variance of the GARCH($1, 1$) model

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

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

These five explanatory variables are all available at time $t-1$. We use two volatility measures ($x_{3t}$ and $x_{5t}$) to study the effect of market volatility on VaR. As shown in Example 7.3 by the fitted AR(2)-GARCH($1, 1$) model, the serial correlations in $r_t^o$ 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 7.5(c) and (d) and Figures 7.6(c) and (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 $z_t$ and $w_t$ are within the asymptotic two standard-error limits. The QQ-plots also show marked improvements as they indicate no model inadequacy. Based on these checking results, the inhomogeneous model seems adequate.

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

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

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Constant</th>
<th>Coef. of $x_{3t}$</th>
<th>Coef. of $x_{4t}$</th>
<th>Coef. of $x_{5t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_t$</td>
<td>0.3202</td>
<td>1.4772</td>
<td>2.1991</td>
<td>(a) Threshold 2.5% with 334 exceedances</td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.3387)</td>
<td>(0.3222)</td>
<td>(0.2450)</td>
<td></td>
</tr>
<tr>
<td>$\ln(\alpha_t)$</td>
<td>-0.8119</td>
<td>0.3305</td>
<td>1.0324</td>
<td></td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.1798)</td>
<td>(0.0826)</td>
<td>(0.2619)</td>
<td></td>
</tr>
<tr>
<td>$k_t$</td>
<td>-0.1805</td>
<td>-0.2118</td>
<td>-0.3551</td>
<td></td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.1290)</td>
<td>(0.0580)</td>
<td>(0.1503)</td>
<td></td>
</tr>
<tr>
<td>$\beta_t$</td>
<td>1.1569</td>
<td></td>
<td>2.1918</td>
<td>(b) Threshold 3.0% with 184 exceedances</td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.4082)</td>
<td></td>
<td>(0.2909)</td>
<td></td>
</tr>
<tr>
<td>$\ln(\alpha_t)$</td>
<td>-0.0316</td>
<td>0.3336</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.1201)</td>
<td>(0.0861)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_t$</td>
<td>-0.6008</td>
<td>-0.2480</td>
<td>0.3175</td>
<td></td>
</tr>
<tr>
<td>(Std.err)</td>
<td>(0.1454)</td>
<td>(0.0731)</td>
<td>(0.0685)</td>
<td></td>
</tr>
</tbody>
</table>

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. The 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

$$k_{9190} = -0.01195, \quad \ln(\alpha_{9190}) = 0.19331, \quad \beta_{9190} = 6.105.$$ 

Assume that the tail probability is 0.05. The VaR quantile shown in Eq. (7.34) gives $\text{VaR} = 3.03756\%$. 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 I 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

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

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

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

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

**EXERCISES**

1. Consider the daily log returns of GE stock from July 3, 1962 to December 31, 1999. The data can be obtained from CRSP or the file “d-geln.dat.” Suppose that you hold a long position on the stock valued at $1 million. Use the tail probability 0.05. Compute the value at risk of your position for 1-day horizon and 15-day horizon using the following methods:
   - (a) The RiskMetrics method.
   - (b) A Gaussian ARMA-GARCH model.
   - (c) An ARMA-GARCH model with a Student-$t$ distribution. You should also estimate the degrees of freedom.
   - (d) The traditional extreme value theory with subperiod length $n = 21$.

2. The file “d-csco9199.dat” contains the daily log returns of Cisco Systems stock from 1991 to 1999 with 2275 observations. 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.

3. Use Hill’s estimator and the data “d-csco9199.dat” to estimate the tail index for daily stock returns of Cisco Systems.

4. The file “d-hwp3dx8099.dat” contains the daily log returns of Hewlett-Packard, CRSP value-weighted index, equal-weighted index, and S&P 500 index from 1980 to 1999. All returns are in percentages and include dividend distributions. 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 2000.

(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 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 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.

REFERENCES


