95% confidence interval for VaR is the range between the 2.5 percentile point and the 97.5 percentile point of the distribution of the VaRs calculated from the data sets.

Suppose, for example, that we have 500 days of data. We could sample with replacement 500,000 times from the data to obtain 1,000 different sets of 500 days of data. We calculate the VaR for each set. We then rank the VaRs. Suppose that the 25th largest VaR is \$5.3 million and the 975th largest VaR is \$8.9 million. The 95% confidence interval for VaR is \$5.3 million to \$8.9 million. Usually, the width of the confidence interval calculated for VaR using the bootstrap method is less than that calculated using the procedure in Section 13.2.

13.4 COMPUTATIONAL ISSUES

Historical simulation involves valuing the whole portfolio of a financial institution many times (500 times in our example). This can be computationally very time consuming. This is particularly true when some of the instruments in the portfolio are valued with Monte Carlo simulation, because there is then a simulation within a simulation problem because each trial of the historical simulation involves a Monte Carlo simulation.

To reduce computation time, financial institutions sometimes use a delta–gamma approximation. This is explained in Chapter 8. Consider an instrument whose price, P, is dependent on a single market variable, S. An approximate estimate of the change, ΔP, in P resulting from a change, ΔS, in S is

$$\Delta P = \delta \Delta S + \frac{1}{2}\gamma(\Delta S)^2 \tag{13.3}$$

where δ and γ are the delta and gamma of P with respect to S. The Greek letters δ and γ are always known because they are calculated when the instrument is marked to market each day. This equation can therefore be used as a fast approximate way of calculating the changes in the value of the transaction for the changes in the value of S that are considered by the historical simulation.

When an instrument depends on several market variables, $S_i (1 \le i \le n)$, equation (13.3) becomes

$$\Delta P = \sum_{i=1}^{n} \delta_i \,\Delta S_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2}\gamma_{ij}\,\Delta S_i \Delta S_j \tag{13.4}$$

where δ_i and γ_{ij} are defined as

$$\delta_i = \frac{\partial P}{\partial S_i} \qquad \gamma_{ij} = \frac{\partial^2 P}{\partial S_i \partial S_j}$$

13.5 EXTREME VALUE THEORY

Section 10.4 introduced the power law and explained that it can be used to estimate the tails of a wide range of distributions. We now provide the theoretical underpinnings for the power law and present estimation procedures more sophisticated than those used in Section 10.4. *Extreme value theory* (EVT) is the term used to describe the science of estimating the tails of a distribution. EVT can be used to improve VaR estimates and to help in situations where analysts want to estimate VaR with a very high confidence level. It is a way of smoothing and extrapolating the tails of an empirical distribution.

The Key Result

The key result in EVT was proved by Gnedenko (1943).[8] It shows that the tails of a wide range of different probability distributions share common properties.

Suppose that $F(v)$ is the cumulative distribution function for a variable v (such as the loss on a portfolio over a certain period of time) and that u is a value of v in the right-hand tail of the distribution. The probability that v lies between u and $u + y$ ($y > 0$) is $F(u + y) - F(u)$. The probability that v is greater than u is $1 - F(u)$. Define $F_u(y)$ as the probability that v lies between u and $u + y$ conditional on $v > u$. This is

RD: above is for the right tail

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

The variable $F_u(y)$ defines the right tail of the probability distribution. It is the cumulative probability distribution for the amount by which v exceeds u given that it does exceed u.

Gnedenko's result states that, for a wide class of distributions $F(v)$, the distribution of $F_u(y)$ converges to a generalized Pareto distribution as the threshold u is increased. The generalized Pareto (cumulative) distribution is

$$G_{\xi,\beta}(y) = 1 - \left[1 + \xi \frac{y}{\beta}\right]^{-1/\xi} \tag{13.5}$$

The distribution has two parameters that have to be estimated from the data. These are ξ and β. The parameter ξ is the shape parameter and determines the heaviness of the tail of the distribution. The parameter β is a scale parameter.

When the underlying variable v has a normal distribution, $\xi = 0$.[9] As the tails of the distribution become heavier, the value of ξ increases. For most financial data, ξ is positive and in the range 0.1 to 0.4.[10]

---

[8] See D. V. Gnedenko, "Sur la distribution limité du terme d'une série aléatoire," *Annals of Mathematics* 44 (1943): 423–453.

[9] When $\xi = 0$, the generalized Pareto distribution becomes

$$G_{\xi,\beta}(y) = 1 - \exp\left(-\frac{y}{\beta}\right)$$

[10] One of the properties of the distribution in equation (13.5) is that the kth moment of v, $E(v^k)$, is infinite for $k \geq 1/\xi$. For a normal distribution, all moments are finite. When $\xi = 0.25$, only the first three moments are finite; when $\xi = 0.5$, only the first moment is finite; and so on.

Estimating ξ and β

The parameters ξ and β can be estimated using maximum likelihood methods (see Section 10.9 for a discussion of maximum likelihood methods). The probability density function, $g_{\xi,\beta}(y)$, of the cumulative distribution in equation (13.5) is calculated by differentiating $G_{\xi,\beta}(y)$ with respect to y. It is

$$g_{\xi,\beta}(y) = \frac{1}{\beta}\left(1 + \frac{\xi y}{\beta}\right)^{-1/\xi-1} \tag{13.6}$$

We first choose a value for u. (A value close to the 95th percentile point of the empirical distribution usually works well.) We then rank the observations on v from the highest to the lowest and focus our attention on those observations for which $v > u$. Suppose there are n_u such observations and they are v_i $(1 \le i \le n_u)$. The likelihood function (assuming that $\xi \neq 0$) is

$$\prod_{i=1}^{n_u} \frac{1}{\beta}\left(1 + \frac{\xi(v_i - u)}{\beta}\right)^{-1/\xi-1}$$

Maximizing this function is the same as maximizing its logarithm:

$$\sum_{i=1}^{n_u} \ln\left[\frac{1}{\beta}\left(1 + \frac{\xi(v_i - u)}{\beta}\right)^{-1/\xi-1}\right] \tag{13.7}$$

Standard numerical procedures can be used to find the values of ξ and β that maximize this expression. Excel's Solver produces good results.

Estimating the Tail of the Distribution

The probability that $v > u + y$ conditional that $v > u$ is $1 - G_{\xi,\beta}(y)$. The probability that $v > u$ is $1 - F(u)$. The unconditional probability that $v > x$ (when $x > u$) is therefore

$$[1 - F(u)][1 - G_{\xi,\beta}(x - u)]$$

If n is the total number of observations, an estimate of $1 - F(u)$, calculated from the empirical data, is n_u/n. The unconditional probability that $v > x$ is therefore

$$\text{Prob}(v > x) = \frac{n_u}{n}[1 - G_{\xi,\beta}(x - u)] = \frac{n_u}{n}\left[1 + \xi\frac{x - u}{\beta}\right]^{-1/\xi} \tag{13.8}$$

Equivalence to the Power Law

If we set $u = \beta/\xi$, equation (13.8) reduces to

$$\text{Prob}(v > x) = \frac{n_u}{n}\left[\frac{\xi x}{\beta}\right]^{-1/\xi}$$

This is

$$Kx^{-\alpha}$$

where

$$K = \frac{n_u}{n}\left[\frac{\xi}{\beta}\right]^{-1/\xi}$$

and $\alpha = 1/\xi$. This shows that equation (13.8) is consistent with the power law introduced in Section 10.4.

The Left Tail

The analysis so far has assumed that we are interested in the right tail of the probability distribution of a variable v. If we are interested in the left tail of the probability distribution, we can work with $-v$ instead of v. Suppose, for example, that an oil company has collected data on daily percentage increases in the price of oil and wants to estimate a VaR that is the one-day percentage decline in the price of oil that has a 99.9% probability of not being exceeded. This is a statistic calculated from the left tail of the probability distribution of oil price increases. The oil company would change the sign of each data item (so that the data was measuring oil price decreases rather than increases) and then use the methodology that has been presented.

Calculation of VaR and ES

To calculate VaR with a confidence level of q, it is necessary to solve the equation

$$F(\text{VaR}) = q$$

Because $F(x) = 1 - \text{Prob}(v > x)$, equation (13.8) gives

$$q = 1 - \frac{n_u}{n}\left[1 + \xi\frac{\text{VaR} - u}{\beta}\right]^{-1/\xi}$$

so that

$$\text{VaR} = u + \frac{\beta}{\xi}\left\{\left[\frac{n}{n_u}(1-q)\right]^{-\xi} - 1\right\} \tag{13.9}$$

The expected shortfall is given by

$$\text{ES} = \frac{\text{VaR} + \beta - \xi u}{1 - \xi} \tag{13.10}$$

13.6 APPLICATIONS OF EVT

Consider again the data in Tables 13.1 to 13.4. When $u = 160$, $n_u = 22$ (that is, there are 22 scenarios where the loss in \$000s is greater than 160). Table 13.10 shows

TABLE 13.10 Extreme Value Theory Calculations for Table 13.4 (the parameter u is 160 and trial values for β and ξ are 40 and 0.3, respectively)

Scenario Number	Loss ($000s)	Rank	$\ln\left[\frac{1}{\beta}\left(1+\frac{\xi(v_i-u)}{\beta}\right)^{-1/\xi-1}\right]$
494	477.841	1	−8.97
339	345.435	2	−7.47
349	282.204	3	−6.51
329	277.041	4	−6.42
487	253.385	5	−5.99
227	217.974	6	−5.25
131	202.256	7	−4.88
238	201.389	8	−4.86
…	…	…	…
…	…	…	…
…	…	…	…
304	160.778	22	−3.71
			−108.37

Trial Estimates of EVT Parameters

ξ	β
0.3	40

calculations for the trial values $\beta = 40$ and $\xi = 0.3$. The value of the log-likelihood function in equation (13.7) is −108.37.

When Excel's Solver is used to search for the values of β and ξ that maximize the log-likelihood function (see worksheet 11 on the website file), it gives

$$\beta = 32.532$$

$$\xi = 0.436$$

and the maximum value of the log-likelihood function is −108.21.

Suppose that we wish to estimate the probability that the portfolio loss between September 25 and September 26, 2008, will be more than $300,000 (or 3% of its value). From equation (13.8) this is

$$\frac{22}{500}\left[1+0.436\frac{300-160}{32.532}\right]^{-1/0.436} = 0.0039$$

This is more accurate than counting observations. The probability that the portfolio loss will be more than $500,000 (or 5% of its value) is similarly 0.00086.

From equation (13.9), the value of VaR with a 99% confidence limit is

$$160+\frac{32.532}{0.436}\left\{\left[\frac{500}{22}(1-0.99)\right]^{-0.436}-1\right\} = 227.8$$

or \$227,800. (In this instance, the VaR estimate is about \$25,000 less than the fifth worst loss.) When the confidence level is increased to 99.9%, VaR becomes

$$160 + \frac{32.532}{0.436}\left\{\left[\frac{500}{22}(1 - 0.999)\right]^{-0.436} - 1\right\} = 474.0$$

or \$474,000. When it is increased further to 99.97%, VaR becomes

$$160 + \frac{32.532}{0.436}\left\{\left[\frac{500}{22}(1 - 0.9997)\right]^{-0.436} - 1\right\} = 742.5$$

or \$742,500.

The formula in equation (13.10) can improve ES estimates and allow the confidence level used for ES estimates to be increased. In our example, when the confidence level is 99%, the estimated ES is

$$\frac{227.8 + 32.532 - 0.436 \times 160}{1 - 0.436} = 337.9$$

or \$337,900. When the confidence level is 99.9%, the estimated ES is

$$\frac{474.0 + 32.532 - 0.436 \times 160}{1 - 0.436} = 774.8$$

or \$774,800.

EVT can also be used in a straightforward way in conjunction with the volatility-updating procedures in Section 13.3 (see Problem 13.11). It can also be used in conjunction with the weighting-of-observations procedure in Section 13.3. In this case, the terms being summed in equation (13.7) must be multiplied by the weights applicable to the underlying observations.

A final calculation can be used to refine the confidence interval for the 99% VaR estimate in Section 13.2. The probability density function evaluated at the VaR level for the probability distribution of the loss, conditional on it being greater than 160, is given by the $g_{\xi,\beta}$ function in equation (13.6). It is

$$\frac{1}{32.532}\left(1 + \frac{0.436 \times (227.8 - 160)}{32.532}\right)^{-1/0.436-1} = 0.0037$$

The unconditional probability density function evaluated at the VaR level is n_u/n = 22/500 times this or 0.00016. Not surprisingly, this is lower than the 0.000284 estimated in Section 13.2 and leads to a wider confidence interval for VaR.

Choice of *u*

A natural question is: "How do the results depend on the choice of *u*?" It is often found that values of ξ and β do depend on *u*, but the estimates of $F(x)$ remain roughly the same. (Problem 13.10 considers what happens when u is changed from 160 to 150 in the example we have been considering.) We want u to be sufficiently high that we are truly investigating the shape of the tail of the distribution, but sufficiently low that the number of data items included in the maximum likelihood calculation is not too low. More data lead to more accuracy in the assessment of the shape of the tail. We have applied the procedure with 500 data items. Ideally, more data would be used.

A rule of thumb is that u should be approximately equal to the 95th percentile of the empirical distribution. (In the case of the data we have been looking at, the 95th percentile of the empirical distribution is 156.5.) In the search for the optimal values of ξ and β, both variables should be constrained to be positive. If the optimizer tries to set ξ negative, it is likely to be a sign that either (a) the tail of the distribution is not heavier than the normal distribution or (b) an inappropriate value of u has been chosen.

SUMMARY

Historical simulation is a very popular approach for estimating VaR or ES. It involves creating a database consisting of the daily movements in all market variables over a period of time. The first simulation trial assumes that the percentage change in each market variable is the same as that on the first day covered by the database, the second simulation trial assumes that the percentage changes are the same as those on the second day, and so on. The change in the portfolio value is calculated for each simulation trial, and VaR is calculated as the appropriate percentile of the probability distribution of this change. The standard error for a VaR that is estimated using historical simulation tends to be quite high. The higher the VaR confidence level required, the higher the standard error. In one extension of the basic historical simulation approach, the weights given to observations decrease exponentially as the observations become older; in another, adjustments are made to historical data to reflect changes in volatility.

Extreme value theory is a way of smoothing the tails of the probability distribution of portfolio daily changes calculated using historical simulation. It leads to estimates of VaR and ES that reflect the whole shape of the tail of the distribution, not just the positions of a few losses in the tails. Extreme value theory can also be used to estimate VaR and ES when the confidence level is very high. For example, even if we have only 500 days of data, it could be used to come up with an estimate of VaR or ES for a confidence level of 99.9%.

FURTHER READING

Boudoukh, J., M. Richardson, and R. Whitelaw. "The Best of Both Worlds." *Risk* (May 1998): 64–67.

Embrechts, P., C. Kluppelberg, and T. Mikosch. *Modeling Extremal Events for Insurance and Finance* (New York: Springer, 1997).

Hendricks, D. "Evaluation of Value-at-Risk Models Using Historical Data,"*Economic Policy Review*, Federal Reserve Bank of New York, vol. 2 (April 1996): 39–69.