BLOCK 7
RISK MODELS

INTRODUCTION TO BLOCK 7

The last block of this course (**Block 7**) is on Risk Models. This block has three units (Units 20 to 22). **Unit 20** is on 'Theory of Extreme Value'. The unit begins with a brief overview of 'Extreme Value Theory (EVT)'. Under this, concepts like (i) the GEV (generalised extreme value) distribution, (ii) maximum domain of attraction (MDA) and (iii) excess beyond a threshold are explained. The different steps involved in applying the EVT viz. (i) exploratory data analysis, (ii) determination of the 'means excess function', (iii) sampling the maxima and (iv) choice of the threshold are described. Different methods for the estimation of parameters like (i) parametric methods, (ii) semi-parametric methods, (iii) fitting excess over a threshold and (iv) extreme VaR are discussed. A note on the 'limitations of the EVT' is also given in the unit.

Unit 21 is on Credibility Theory. An account of the 'classical credibility' in terms of (i) full credibility for frequency, (ii) full credibility for severity, (iii) full credibility for pure premiums and (iv) partial credibility is first given in this unit. Three different types of credibility measures viz. (i) Bayesian credibility, (ii) Buhlmann credibility and (iii) Buhlmann-Straub credibility are then described. A brief account on the estimation of 'credibility parameters' is then given. A comparative profile of the classical and the Buhlmann credibility is also subsequently presented in the unit. A note on 'maximum aggregate loss and general solution' concludes this unit.

Unit 22 is on Dynamic Financial Analysis. The unit discusses two aspects viz. (i) stochastic simulations and (ii) stochastic variables. Under 'stochastic simulations', the idea of 'efficient frontier' and 'stochastic scenario generator' is discussed. Under 'stochastic variables', concepts like (i) short term interest rate, (ii) term structure and inflation, (iii) stock returns, (iv) non-catastrophe and catastrophe losses and (v) underwriting cycles and payment patterns are explained. A note on 'corporate model' concludes this last unit of the course.

UNIT 20 THEORY OF EXTREME VALUE

Structure

20.0 OBJECTIVES

After reading this unit, you will be able to:

- state the significance of 'extreme value theory' with examples;
- outline the theoretical form of a 'generalised extreme value (GEV)' distribution;
- specify the major theoretical results of 'extreme value theory (EVT)';
- explain the distribution of 'maximum domain attraction (MDA)' with examples;
- discuss the distribution of 'excess beyond a threshold';
- illustrate some practical situations where the models of EVT are applied;

- detail the steps involved in applying the EVT;
- describe the methods for the estimation of parameters in extreme distributions; and
- write a note on the 'limitations of the EVT'.

20.1 INTRODUCTION

There are unusual or rare events such as earthquakes, hurricanes and stock market crashes. They follow no rule of occurrence. Insurance companies have to deal with such extreme events. It is essential for an insurance company to set its premiums at levels high enough to cover losses from such rare but large claims. For instance, a property and casualty insurance company must be able to withstand a loss proportional to the magnitude of the loss caused by a Tsunami of 2004 or Hurricane Andrew of 1992 kind. The occurrence of such events is infrequent but the magnitude of loss is enormous. 'Extreme value theory' (EVT) of statistics deals with such extreme deviations from the median of probability distributions. In general, a method that is used to deal with such situations is to assess the 'type of probability distribution' generated by such a process. We cannot use the usual rule of normal distribution where the distribution of observations fit into the well-known bell-shaped curve. Essentially, one is now looking for extremes, which are in the tails of distributions. These are flatter (heavier) than the normal events. Some examples in this category are: (i) estimation of the probability of extreme insurance claims, (ii) typical sizes of such claims to compute the premium for covering such future claims, (iii) modelling of extreme (individual and simultaneous) price movements for financial assets to build an accurate risk model for a financial portfolio, etc. Thus, an important feature of 'extreme value theory (EVT)' is to provide solutions to problems associated with extreme risks that are extremely rare.

20.2 EXTREME VALUE THEORY (EVT)

There are two significant results of EVT. First is the asymptotic distribution of a series of maxima (or minima) modelled under certain conditions (e.g. the distribution of the standardised maximum of the series). Such distributions are known to converge to either the Gumbel or Frechet or Weibull distribution. These three distributions give the 'generalised extreme value (GEV)' distribution. This result is used to estimate high quantiles (i.e. 0.999 and higher). EVT, thus, shows that the limiting distribution is a 'Generalised Pareto distribution (GPD)'. A method often used is the 'Peaks-Over-Threshold (POT)' method. Here the maxima (or minima) is assumed to follow a GEV distribution. For instance, if we were concerned with monthly peaks in interest rates, we could fit a GEV distribution to the monthly maxima. Excesses over a given high threshold follow a GPD. Hence, if we are interested in the distribution of insurance claims over some high threshold, such excesses would be best modelled by a GPD. Alternatively, if we are concerned about the occurrence times of the losses over some threshold, we can fit a POT model. In a POT model, the number of events during a given time follows a Poisson process and the exceedances over a given high threshold follow a GPD. The number of events being Poisson, the inter-arrival times (time between events) are exponentially distributed. Consequently, by fitting a POT model, (i) we can estimate the average time between events of a given magnitude (threshold), and (ii) we can obtain the distribution of the excess over the threshold.

20.2.1 The GEV Distribution

EVT draws from the limiting distribution of sample extreme (maxima or minima). To appreciate its underlying idea, we take $X = \left(X_{1,}X_{2,}...X_n\right)$ as *a* sequence of 'independent identically distributed (*iid*)' observations with distribution function F. We denote the sample maximum by $M_n = \max\left\{X_1,...X_n\right\}$. Under the assumption of sub-exponential distributions, the tail of the maximum determines the tail of the sum as $n \to \infty$. More generally, the 'generalised extreme value (GEV) distribution', given by $H_\xi\left(x\right)$, describes the 'limit distributions of a normalised maxima'. For instance, we can obtain a standard GEV by writing $\frac{X-\mu}{\sigma}$ for X and obtain a distribution function specified like:

$$H_{\xi,}\left(x\right) = \begin{cases} \exp\left[-\left(1+\xi\frac{X-\mu}{\sigma}\right)^{-1/\xi}\right], & \text{if } \xi \neq 0, 1+\xi\frac{x-\mu}{\sigma} > 0 \\ \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right], & \text{if } \xi = 0 \end{cases} \quad (20.1)$$

where μ, σ and ξ are the location, scale and shape parameters, respectively. Note that μ converges to a scalar value whereas σ is a tendency. The parameter ξ gives the tail index. It indicates the thickness of the tail of distribution i.e. the larger the index, the thicker the tail.

Classical extreme value theory was formulated by Fisher and Tippett (1928). The Fisher-Tippet theorem specifies the form of the 'limit distribution' for the 'centred and normalised maxima (CNM)'. To see its construction, consider $\left\{X_n\right\}$, a sequence of *iid* random variables, with M_n as the $\max\left(X_1,...,X_n\right)$. Now, if there exists constants $c_n > 0$ and a real number $d_n \in R$, then $\left(M_n - d_n\right)/c_n$ yields a 'centred and normalised maximum'. When $\left(M_n - d_n\right)/c_n \xrightarrow{d} H$ (i.e. it converges in distribution to H), where H is some non-degenerate distribution function, then H belongs to one of the three families of extreme value distribution functions (viz. Fréchet, Weibull, and Gumbel). The convergence of $M_n = Max(X_1..., X_n)$ will be to a distribution like the following.

$$H_{(\xi,\mu,\sigma)}(x) = \begin{cases} \exp\left(-\left[1+\xi(X-\mu)/\sigma\right]^{-1/\xi}\right) \text{if } \xi \neq 0, \\ \exp\left(-e^{-(X-\mu)/\sigma}\right) \quad \text{if } \xi = 0 \end{cases} \tag{20.2}$$

where $1+\xi(X-\mu)/\sigma > 0$. When the index is equal to zero, the distribution H corresponds to a Gumbel distribution. When the index is negative, it corresponds to a Weibull distribution. When the index is positive, it corresponds to a Frechet distribution. In particular, the Frechet distribution corresponds to a fat-tailed distribution. Such distributions are found to be the most appropriate for fat-tailed financial data. The asymptotic distribution of the maximum always belongs to one of these three distributions, irrespective of the original distribution. Further, the asymptotic distribution of the maximum can be estimated without making any assumptions on the nature of the original distribution of the observations.

The purpose of tail estimation procedures is to estimate the values of X that lie outside the range of existing data. To do this, we have to employ both extreme events and exceedances from a specified level. The standard approach assumes that the tail of the population follows the select family of distribution as stated above.

20.2.2 Maximum Domain of Attraction (MDA)

The extreme value distributions introduced above represent the 'limit laws' for the normalised maxima of *iid* random variables. We have to consider the conditions under which the distribution function F of the normalised maxima M_n converges to H. In other words, we need to choose constants $c_n > 0$ and d_n such that $(M_n - d_n)/c_n \xrightarrow{d} H$. If this condition is satisfied, then we say that 'the distribution function F belongs to the maximum domain of attraction of H'. In other words, for a random variable X belonging to the 'maximum domain of attraction' of the extreme value distribution H, if for some constants $c_n > 0$ and $d_n \in R$, we have $(M_n - d_n)/c_n \xrightarrow{d} H$, then we write $X \in MDA(H)$. We can take some examples to illustrate the concept of the MDA to find the appropriate constants.

Example 1: Consider a sample $\{X_1, X_2, ..., X_n\}$ from the logistic distribution $F(x) = 1/(1+e^{-x})$. Our first task is to find the constants c_n and d_n that helps achieve normalised distribution. Second, we need to find the limiting distribution of $(M_n - d_n)/c_n$. Third, we have to find the 'centring' constants $\{d_n\}$ close to $E[M_n]$ (i.e. the expected value of M_n). When all these steps are completed, we have $F(X_1),...,F(x_n)$ as an ordered sample from the uniform distribution over (0,1). Then, $F(x_n)$ is the largest of a sample of size n from such a distribution. This gives $E[F(M_n)] = n/(n+1)$. We therefore have:

$$F(E[M_n]) = (1+\exp\{-E[M_n]\})^{-1} = 1-(1+\exp\{E[M_n]\})^{-1}$$

and $E[F(M_n)] = \dfrac{n}{n+1} = 1-(n+1)^{-1}$.

If $F(E[M_n]) \approx E[F(M_n)]$, then $n \approx \exp\{E[M_n]\}$ or $E[M_n] \approx \log n$. This means, a reasonable choice for the centring constants $\{d_n\}$ is the sequence of $\{\log n\}$. To find the limit of the distribution, we proceed as follows:

$$\begin{aligned}
\lim_{n\to\infty} P\left[\frac{M_n - d_n}{c_n} \le y\right] &= \lim_{n\to\infty} P\left[\frac{M_n - \log n}{c_n} \le y\right] \\
&= \lim_{n\to\infty} P[M_n \le c_n y + \log n] \\
&= \lim_{n\to\infty} [F(c_n y + \log n)]^n \\
&= \lim_{n\to\infty} \left(1+e^{-c_n y - \log n}\right)^{-n} \\
&= \lim_{n\to\infty} \left(1+(1/n)e^{-c_n y}\right)^{-n} \\
&= \exp\{-e^{-y}\} \quad \text{for } c_n \equiv 1.
\end{aligned}$$

Consequently, we choose $\{d_n\} = \{\log n\}$ and $\{c_n\} = 1$. The limiting distribution of $(M_n - d_n)/c_n$ is $\exp\{-e^{-y}\}$, which is the Gumbel distribution.

Example 2: Suppose we have a sample from the exponential distribution, $F(x) = 1-e^{x/\beta}$, for $x \in [0,\infty)$.

Then, $E[F(M_n)] = n/(n+1) = 1-1/(n+1) \Rightarrow$

$$F(E[M_n]) = 1-\exp\{-E[M_n]/\beta\} \approx E[F(M_n)].$$

Now, since, $\dfrac{1}{n+1} \approx \exp\{-E[M_n/\beta]\}$, we have $E[M_n] \approx \beta\log(n+1) \approx \beta\log n$. We can, therefore, use $d_n = \beta\log n$. Hence:

$$\lim_{n\to\infty} P\left[\frac{M_n - d_n}{c_n} \le y\right] = \lim_{n\to\infty} P[M_n - \beta\log n \le c_n y]$$

$$= \lim_{n\to\infty} \left[F\left(c_n y + \beta \log n\right) \right]^n$$

$$= \lim_{n\to\infty} \left(1 + e^{-(1/\beta)c_n y - \log n}\right)^n$$

$$= \lim_{n\to\infty} \left(1 - (1/n) e^{-(1/\beta)c_n y}\right)^n$$

$$= \exp\left\{-e^{-y}\right\} \textit{ for } c_n \equiv \beta .$$

Thus, if we choose $\{d_n\} = \{\beta \log n\}$ and $c_n = \beta$, the distribution in the limit of $\left(M_n - d_n\right)/c_n$ is a Gumbel distribution. The conclusion we can draw from these two examples is that the MDA of Gumble distribution can be obtained with the normalised maxima of the logistic and the exponential distributions.

20.2.3 Excess Beyond a Threshold

Let *X* be a random variable with a distribution *F*. Now, if F_u is the distribution of excesses of *X* over a threshold *U*, we have:

$$F_u(x) = P\left(X - u \le x \mid X > u\right),\ x \ge 0 \tag{20.3}$$

If we estimate the threshold with the help of *VaR*, the conditional distribution F_u can be approximated by a 'Generalised Pareto distribution (GPD)'. We can therefore write:

$$F_u(x) \approx G_{\xi,}\beta_{(u)}(x), u \to \infty,\ \ x \ge 0 \tag{20.4}$$

where: $G_{\xi,\beta_{(u)}}(x) = \begin{pmatrix} 1 - \left(1 + \dfrac{\xi x}{\beta}\right)^{-1/\xi} \\ 1 - e^{-x/\beta} \end{pmatrix} if \,|^{\xi \ne 0}_{\xi = 0}$. The Generalised Pareto Distribution (GPD) is the distribution in limit of excesses $Y = \max\{X - u, 0\}$. Over sufficiently high thresholds, *u* offers a good approximation of the tail of *F* for some fixed ξ and β which depends on *u*. Thus, the distribution of *Y* may be thought of as the conditional distribution of *X* given $X > u$. Such a GPD with shape parameter ξ and scale parameter β can be specified as:

$$G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \xi \dfrac{y}{\beta}\right)^{-1/\xi} & if\, \xi \ne 0 \\ 1 - \exp\left(-\dfrac{y}{\beta}\right) & if\, \xi = 0 \end{cases}$$

where $y \in \begin{cases} [0, \infty] & if\, \xi \ge 0 \\ \left[0, -\dfrac{\beta}{\xi}\right] & if\, \xi < 0 \end{cases}$

The sign of the shape parameter ξ determines its tail behaviour and thereby the tail behaviour of the original distribution. If $\xi > 0$, then the tail of the distribution function F of X decays like a power function $X^{-1/\xi}$. In that case, F belongs to a family of heavy-tailed distributions. This, among others, includes the Pareto, log-Gamma, Cauchy and t-distributions. If $\xi = 0$, then the tail of F decreases exponentially and belongs to a class of medium-tailed distributions. These include the normal, exponential, Gamma and log-normal distributions. If $\xi < 0$, the underlying distribution F is characterised by a finite right end point. This class of short-tailed distributions includes the uniform and beta distributions. The mean excess function (expectation) of the GPD is given by:

$$e(u) = E(X - u \mid X > u) = \frac{\beta + \xi_u}{1 - \xi} \quad (20.5)$$

where $\beta = \sigma + \xi(u - \mu)$ and $\max_n y_n$ follows a GEV distribution with parameters ξ, μ, σ. Distributions of this type {i.e. H of MDA} are used to model the behaviour of the maximum of a series. The distributions of the type G (i.e. of GPD) model 'excess beyond a given threshold'. This threshold should be sufficiently large to satisfy the condition of $u \to \infty$.

The GEV describes the limit distributions of normalised maxima. The GPD describes the limit distribution of scaled excesses over high thresholds. While modelling external events, we are often concerned with excesses over some high value. For instance, in a catastrophe bond structure that has a $1 billion attachment point, we need to ascertain if the loss claims are greater than $1 billion. To estimate such claims, we need a definition for the 'excess distribution function'. This is stated as follows.

Let X be a random variable with the distribution function F having a right end-point x_F. If we have fixed $u < x_F$, then $F_u(x) = P(X - u \le x \mid X > u)$, $x \ge 0$, is the 'excess distribution function' of X over the threshold u. The function $e(u) = E[X - u \mid X > u]$ gives the 'mean excess function' of X. This definition says that for any high threshold u, we have the distribution of the amounts over u with its $e(u)$ giving the average excess over u. Thus, when our objective is to find an appropriate value of high threshold, the 'mean excess function' provides a necessary tool. In particular, when the mean excess function of a GPD is linear, we can select a threshold from the region where the empirical mean-excess function is roughly linear.

20.2.4 Applications

In this sub-section, we shall consider three illustrations and note the basic steps for the formulation of POT models.

Danish Fire Insurance Claims: Empirical data on 'mean excess plot' was used [by ABS Research (2000)] to demonstrate the application of the 'mean excess function'. In its Chart, the mean-excess plot was shown to be roughly linear in the range of 10-15. This indicated that a threshold of 10 is appropriate where the claims are in millions of Danish Kroner. On the basis of such a result, it was concluded that this threshold could be used to model the excess distribution. When the number of exceedances of a high threshold is roughly a Poisson process, it is useful to model a 'peaks-over-threshold (POT)' model. Therefore, the study suggested the following:

- the number of exceedances of a high threshold follows a Poisson process;
- excesses over high thresholds can be modelled by a GPD;
- plotting the 'mean-excess function' is useful in determining an appropriate value of the high threshold; and
- the distribution of the maximum of a Poisson number of *iid* excesses over a high threshold is a GEV.

The POT Model: Peaks-over-threshold (POT) models have been used by hydrologists for many years in modelling events like flood levels and dam heights. In the basic steps followed for the formulation of the POT model, the following factors are taken into account.

- Are the occurrence times of the excesses of an *iid* (or stationary) sequence, over a high threshold following a Poisson process?
- Are the corresponding excesses independent and have a GPD? and
- Are the excesses and the occurrence times of the excesses independent of each other?

The POT model can be used: (i) to estimate the excess distribution with respect to a threshold level u and (ii) to estimate the tail shape of the original distribution. The threshold setting of the POT model is data dependent. The model defines a two-dimensional $\left(Y_n, N_u\right)$ space-time point process for $X_n \geq u, n = 1 \ldots N_u$. Here, (i) Y_n and N_u are independent random variables such that Y_n ~ GPD $\left(\xi, b\right)$ and (ii) the number of excesses N_u follows a Poisson process with intensity λ representing the average number of exceedances over the time interval used for sampling process. This is given by:

$$\lambda = \left(1 + \xi \frac{X - \mu}{r}\right)^{-\frac{1}{\xi}} \text{ for } X \geq u \qquad (20.6)$$

The threshold u is usually chosen using the 'mean excess plots'. While assessing the POT models, it will be useful to consider an example such as catastrophe bonds. These instruments become handy because occurrence times of excesses are important. The fitted model helps determine the encounter probabilities of extreme events. Further, we can also determine the non-encounter probabilities of such events. As the 'return period' [i.e. the expected time to an event of a given threshold (from some starting point)] and the 'occurrence times' are independent, the return period is the average time to a given event. Such a feature helps in making statements like: 'a loss of X is expected to occur every four years on average'. The non-encounter probability of an event is indicated by the likelihood of such an event not occurring in a given period (e.g. one year). Thus, the model offers an ideal method for simulation experiments in stress scenarios.

Atlantic Basin Hurricane Model: To appreciate the kind of prediction made through POT model, we can consider the study of 'Parisi and Lund (2000)'. They modelled the Atlantic basin hurricanes making landfall in the United States from 1935 to 1998. From the fitted GPD, they estimated the 'return periods' and 'non-encounter probabilities' for storms of various intensities by simulation. Their three main findings are. (i) A storm with a wind speed of 150 mph is expected to make landfall about once every 10.71 ± 0.104 years with a non-encounter probability of 0.912 ± 0.0028. This means there is a 91% chance that no such storm will make land-fall in any given year. (ii) The complement is therefore that there is about a 9% chance that this event will occur in any given year. In other words, such an event occurs about once every 11 years on average with a 9% chance that it could occur in any year. (iii) As the year-to-year storm counts and intensities are independent, even if a 150 mph storm makes landfall in one year, the probability that it could occur the very next year is about 9%.

Global Reinsurance Industry Application: In another similar analysis, Parisi and Herlihy (1999) modelled the catastrophic losses to the 'global reinsurance industry'. Their paper predicts the 'return periods' and 'non-encounter probabilities' for various levels of global catastrophic losses. An important result from their study is that 'losses of $1 billion or more are expected to occur about once a year'.

We must note that the application of EVT involves a number of considerations. At the outset, it is important for data analysis to determine whether the series has the fat tail needed to apply the EVT results. The parameter estimates of the limit distributions (H and G) depend on the number of extreme observations used. The choice of a threshold should be large enough to satisfy the conditions required to permit its application (i.e. u tends to infinity). At the same time, it should leave sufficient observations for estimation. Different methods of making this choice will be explained in the subsequent section of this unit.

We have, until now, assumed that the extreme observations are '*iid*'. The choice of the method for extracting maxima can be crucial in making this assumption viable. There are some extensions to the theory for estimating the parameters for dependent observations. These are not discussed in the present unit.

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) Indicate the significance of 'Extreme Value Theory'.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) State the important theories on which the EVT is based.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) Specify the approach behind the 'classical extreme value theory'?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

4) Mention the forms of the tail distribution subject to the sign of the shape parameter in the approximated GPD function F.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

5) State the essential difference between the GEV and GPD type of distributions.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

6) Specify the two uses of a POT model.

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

20.3 STEPS IN APPLYING EVT

There are four steps in applying the models of EVT. These are: (i) exploratory data analysis, (ii) determination of 'mean excess function', (iii) sampling the maxima and (iv) choice of the threshold.

20.3.1 Exploratory Data Analysis

For exploring the data, we can use the Q-Q plots. For this, we can start by studying a histogram of the data. In most of the *VaR* methods, the approximation by a normal distribution remains a basic assumption. However, as most financial series are fat-tailed, the graph of the quantiles makes it possible to assess the goodness of fit of the series to a parametric model. In view of this, we can consider a model as follows. Let $X_1 ..., X_n$ be a succession of *iid* random variables. Let $X_{n,n} < ... < X_{1,n}$ be the ordered statistics for the empirical distribution. Note that $F_n(X_k, n) = (n-k+1)/nF$ where F is the estimated parametric distribution for the data. To construct the *Q-Q* plot, a graph of quantiles needs to be drawn. This is defined by the set of the points:

$$\left\{X_{k,n}, F^{-1}\left(\frac{n-k+1}{n}\right), k = 1, ..., n\right\} \tag{20.6}$$

In order to judge that the parametric model fits the data well, we should draw the graph in a linear form. In practice, we need to compare such a graph with various estimated models to select the best. The rule is, more linear the Q-Q plots, the more appropriate the model in terms of goodness of fit. As always, the Q-Q plots can help to detect outliers.

20.3.2 Determination of the Mean Excess Function

For any random variable X and a given threshold x_F, we take:

$$e(u) = E(X - u \mid X > u), \quad 0 \le u < x_F \tag{20.7}$$

The function $e(u)$ is called the 'mean excess function'. This is because $e(u)$ is actually the mean excess over the threshold *u*. To see its specific form, let us take X to follow an exponential distribution with parameter λ. Then the above function becomes $e(u)=\lambda^{-1}$ for any $u>0$. For the GPD, we then have:

$$e(u)=\frac{\beta+\xi u}{1-\xi},\ \beta+\xi u>0 \tag{20.8}$$

The 'mean excess function' for a fat-tailed series is located between the 'constant mean excess' [i.e. $e(u)=\lambda^{-1}$] and the GPD. This is linear and tends to infinity for high thresholds (i.e. as u tends towards infinity). In order to observe the behaviour of the tail, we need to examine the graph. If $X_1...,X_n$ are *iid* with F_n corresponding to the empirical distribution with $\Delta_n(u)=\{i,i=1,...,n,X_i>u\}$, then:

$$e_n(u)=\frac{1}{card(\Delta_n(u))}\sum_{i\in\Delta_n(u)}(X_i-u),\ u\geq 0 \tag{20.9}$$

where '*card*' refers to the 'number of points in the set $\Delta_n(u)$'. In this formulation, the set $\{e_n(X_{k,n}),\ k=1,\ ,,,n\}$ gives the 'mean excess graph'. When the function $e(u)$ tends to infinity, for high-threshold u (i.e. linear shape with positive slope), we will have a fat-tailed distribution.

20.3.3 Sampling the Maxima

There are two approaches considered for building the series of maxima or minima. The first divides the series into non-overlapping blocks of the same length from which we can choose the maximum from each block. The assumption that 'extreme observations are *iid*' is viable in such a case. An intuitive idea on this can also be formed by looking at the data (e.g. financial data). Such data contain periods of high volatility followed by periods of low volatility (clustering). Sampling the maxima using this approach reduces such a feature with increase in the size of the block. However, the risk of losing extreme observations within the same block continues to remain. This makes the choice of the size of the block important.

The second approach chooses a given threshold (high enough) and then considers the extreme observations that exceed this threshold. As regards the choice of the threshold, there is a trade-off between variance and bias. If the number of observations are increased, the series of maxima leads to a lower threshold. To offset this, some observations from the centre of the distribution can be introduced into the series. Consequently, the index of tail becomes more precise owing to smaller variance. However, in the process it also gets biased to an extent. On the other hand, if a high threshold is chosen, it reduces the bias but makes the estimator more volatile due to fewer observations.

20.3.4 Choice of the Threshold

From the preceding discussion, it follows that it is possible to use a graphical tool to choose the threshold. The 'mean excess function' for the GPD being linear (tending towards infinity), for a higher threshold, the excess over the threshold in the series tends to converge to a GPD. It is possible therefore to choose the threshold where an approximation by the GPD is reasonable with the area having a linear shape on the graph.

'Hill graph' serves as another tool for selecting the threshold. If $X_1 > ... > X_n$ are the ordered statistics of *iid* random variables, then a 'hill estimator' of the tail index ξ (using $k+1$ ordered statistics) is defined as:

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{Hi}{H_{k+1}}\right) = \hat{\xi} \qquad (20.10a)$$

In the above equation, the hill graph is defined by the set of points:

$$\left\{\left(k, H_{k,n}^{-1}\right),\ \ 1 \le k \le n-1\right\} \qquad (20.10b)$$

We can select the threshold *u* from such a graph for the stable areas of the tail index. However, this method applies well either for a GPD or close to a GPD type of distribution. The Hill estimator is then the 'maximum likelihood estimator'. Further, since the extreme distribution converges to a GPD over a high threshold *u*, we can justify its use.

20.4 ESTIMATION OF PARAMETERS

The parameters of the extreme distribution needs to be estimated by making certain assumptions. First, we assume that the extreme observations follow exactly the GEV distribution. Second, and more realistically, we assume that the other observations are at least roughly distributed like the GEV distribution. More specifically, the distribution of the observations can be assumed to belong to the 'maximum domain of attraction (MDA)' of H_ξ. With this, the parameters and quantiles can be estimated for the distribution of 'excess over a threshold'. We consider here the following methods for estimation.

20.4.1 Parametric Methods

Here we assume that the extreme observations follow exactly the GEV distribution so that the maximum likelihood estimation (MLE) can be used. As there is no closed form for the parameters, numerical methods may be used to provide good estimates. The *p*-quantile method can be used by defining it as $\hat{x}_p = H_{\xi,\hat{\mu},\hat{\sigma}}{}^{-1}(p)$. Then:

$$In(p) = \left(-\left[1 + \hat{\xi}\left(\widehat{X}_p - \hat{\mu}\right)/\hat{\sigma}\right]^{-1/\hat{\xi}}\right.$$

$$\left(-In(p)\right)=\left[1+\hat{\xi}\left(\widehat{X}_p-\widehat{\mu}\right)/\widehat{\sigma}\right]^{-1/\hat{\xi}}$$

$$\widehat{X}_p=\widehat{\mu}+\frac{\widehat{\sigma}}{\widehat{\xi}}\left[\left(-In(p)\right)^{-\xi}-1\right]$$

$$p=\exp\left(-\left[1+\hat{\xi}\left(\widehat{X}_p-\widehat{\mu}\right)/\widehat{\sigma}\right]^{-1/\hat{\xi}}\right) \quad (20.11)$$

MLE has the advantage of simultaneous estimation of the three parameters. It applies well to the series of maxima per block. Especially, when $\xi>-1/2$, MLE is known to yield good estimates. As the majority of financial series have a positive tail index with $\xi>0$, this becomes a good tool of estimation for such data.

20.4.2 Semi-parametric Methods

The assumption that the extreme observations converge or are distributed exactly as H_ξ is very strong. When we relax this assumption, we may find the observation to be distributed like the GEV. Then, *F*, the distribution of the observations, belongs to the MDA of H_ξ. Thus, the parameters and the quantiles estimation differ from that of the *Parametric methods*. Hence, we consider $X_1...,X_n$ to be *iid* random variables with the distribution $F\in$ MDA$\left(H_\xi\right)$. This is equivalent to saying: $\lim_{n\to\infty}\bar{F}\left(\sigma_n x+\mu_n\right)=-In\left(H_\xi\left(x\right)\right)$. Therefore: $\bar{F}\left(x\right)=x^{-1/\xi}L\left(x\right),\ x>0$ or $\forall\lambda>0,\ \lim_{x\to\infty}\frac{L\left(\lambda x\right)}{L\left(x\right)}=1$. Therefore, for a high *u* such as $u=\sigma_n x+\mu_n$, we have:

$$\bar{F}\left(u\right)=\left[1+\hat{\xi}\left(u-\widehat{\mu}_n\right)/\widehat{\sigma}_n\right]^{-1/\hat{\xi}} \quad (20.12a)$$

The *p*-quantile can now be defined as: $\hat{x}_p=H^{-1}_{\xi,\hat{\mu}_n,\hat{\sigma}_n}\left(p\right)$. Then:

$$\hat{X}_p=\hat{\mu}_n+\frac{\hat{\sigma}_n}{\hat{\xi}}\left[\left(n\left(1-p\right)\right)^{-\xi}-1\right] \quad (20.12b)$$

Usually, we are more interested in very high quantiles beyond the series i.e. out-of-sample estimation. For this purpose, let $u=X_k$, where *u* is a very high threshold with *k/n* as its associated probability. Let *p* be the associated probability of the *p*-quantile x_p. Then:

$$\bar{F}\left(X_k\right)=X_{k,n}{}^{-1/\xi}L\left(X_{k,n}\right)=\frac{k}{n} \quad (20.13_a)$$

$$\bar{F}\left(\hat{X}_p\right)=\hat{X}_p^{-1/\xi}L\left(\hat{X}\right)=1-p \quad (20.13_b)$$

Dividing (20.13_b) by (20.13_a), for $x_p > X_{k,}$, we have $L\left(\hat{X}_p\right)/L\left(X_{k,n}\right) \approx 1$. Therefore:

$$\hat{X}_p = \left(\frac{n}{k}(1-p)\right)^{-\hat{\xi}} X_{k,n} \qquad (20.13c)$$

The tail index ξ can be chosen from a stable area on the Hill graph.

20.4.3 Fitting Excesses Over a Threshold

The GPD estimation involves two steps: (i) the choice of the threshold *u* and (ii) the parameter estimations for ξ and β by the MLE method. For the first, the 'mean excess graph' can be used. *u* can be chosen such that *e*(*x*) is approximately linear for $x > u\left(e(u)\right.$. When the distribution of 'excesses over a threshold' is estimated, an approximation of the parent distribution (that generates the extreme observations) can be had. From this, an estimation of its *p*-quantile can be obtained and used as a proxy for the extreme *VaR*. Thus, if *u* is a threshold and $X_1 ..., X_n$ are random variables exceeding this threshold [following a distribution $F \in MDA\left(H_\xi\right)$], and $Y_1 ..., Y_n$ is the series of exceedances $\left(Y_i = X_i - u\right)$], the distribution of excesses beyond u can be obtained as:

$$F_u(y) = P\left(X - u \le y \mid X > u\right) = P\left(Y \ge y \mid X > u\right), y \ge 0 \qquad (20.14)$$

The distribution, *F*, of the extreme observations, X_i, is then given by:

$$\bar{F}(u+y) = P\left(X \ge u+y\right) = P\left(\left(X \ge u+y \mid X > u\right).P\left(X > u\right)\right)$$

$$F(u+y) = P\left(X - u \ge y \mid X > u\right).\ P\left(X > u\right)$$

$$F(u+y) = \bar{F}_u(y).\ \bar{F}_u(u) \qquad (20.15)$$

Such a result makes it possible to estimate the tail of the original distribution. This can be done by separately estimating *F* and F_u. For a high threshold *u*, we can get this as: $\left(\bar{F}_u(y)\right) \approx \bar{G}\hat{\xi}, \hat{\beta}_{(u)}(y)$. *F*(*u*) can be estimated from the empirical distribution of the observations as:

$$\left(\hat{F}(u)\right) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i > u\}} = \frac{N_u}{n} \qquad (20.16a)$$

and

$$\left(\bar{F}(u+y)\right) = \frac{N_u}{n}\left(1 + \hat{\xi}\frac{y}{\beta}\right)^{-1/\hat{\xi}} \qquad (20.16b)$$

The estimation of the p-quantile for a given threshold u, using this distribution is then:

$$\hat{X}_p = u + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\frac{n}{N}(1-p)\right)^{-\hat{\xi}} - 1\right] \qquad (20.16c)$$

20.4.4 Extreme VaR

By definition, VaR is the p-quantile of the distribution of the log of change in price. EVT helps us to model the empirical distribution of the extreme observations. Here, extreme VaR is defined as the estimated 'p-quantile from the extreme distribution'. Various estimators are available for this. Depending on the estimation method and the assumptions, VaR is estimated by assuming that the extreme observations follow exactly the GEV distribution. This is obtained as:

$$VaR_{extrene} = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-Ln(p)\right)^{-\xi} - 1\right] \qquad (20.17)$$

For non-parametric contexts, VaR is estimated by assuming that the observations follow approximately the GEV distribution. The estimate obtained is:

$$VaR_{extreme\ (in-sample)} = \hat{\mu}_n + \frac{\hat{\sigma}_n}{\hat{\xi}}\left[\left(n(1-p)\right)^{-\xi} - 1\right] \qquad (20.18a)$$

and VaR for out-of-sample is:

$$VaR_{extreme\ (out-of-sample)} = \left(\frac{n}{k}(1-p)\right)^{-\xi} X_{k,n} \qquad (20.18b)$$

The approximation of 'excesses over a threshold' by a GPD leads to the estimator:

$$VaR_{extreme\ GPD} = u + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\frac{n}{N}(1-p)\right)^{-\xi} - 1\right] \qquad (20.18c)$$

In practice, estimation using a GPD is repeated several times to have a graph of quantiles. Various results like quantiles for a stable area of the graph, a quantile corresponding to the peak version, etc. can be estimated. This approach is called 'peak over threshold (POT)' method.

20.4.5 Limitations of the EVT

The approaches described above relate to 'aggregate positions'. The application of the EVT resulting in a multivariate case faces a basic problem i.e. there is no standard definition of 'order' in a vectorial space with dimensions greater than 1. Thus, it is difficult to define the extreme observations for n-dimension vectors $(n > 1)$. To solve this problem, what is proposed is that the estimation of the extreme marginal distribution should use the maxima for the short positions (w_i) and the minima for the long positions. Then, the extreme *VaR* of a portfolio of *n* assets can be calculated as:

$$VaR_{extreme} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}.w_i.w_j.VaR_i.\ VaR_j} \qquad (20.19)$$

The joint distribution of the extreme marginal distributions is not necessarily the distribution of the extremes for the aggregate position. In other words, extreme movements of change in the prices of 'different assets' do not necessarily result in extreme movements for the 'whole portfolio'. This will depend on the composition of the portfolio and on the relations (i.e. dependencies or correlations) between the various assets.

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) State the four steps in applying the EVT models.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) What are the two approaches for building a series of maxima or minima?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) What are the methods by which a threshold level can be decided on the basis of which extreme observations that exceed this threshold can be selected?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

4) State the assumptions generally made for estimating the parameters of the extreme distribution.

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

20.5 LET US SUM UP

The unit has discussed modelling of extreme events. In the field of insurance, it is concerned with situations where maximum or minimum level of some value (rather than the expected for average case) is to be considered. The basic features of extreme value theory (EVT) are introduced as a background required for this. The Fisher – Tippet theorem specifies the form of the limit distribution for centred and normalised maxima. The generalised extreme value distribution (GEV), given by $H_{\xi}(x)$, describes the limit distribution of suitably normalised minima or maxima. Three standard distributions that correspond to different values of ξ are Gumbel, Frichet and Weibull to which H converges. When the conditions under which a distribution function that implies the normalised maxima converges to H are satisfied, we say that the distribution function belongs to the 'maximum domain of attraction' of H. Another distribution that plays an important role in modelling extreme events is the 'generalised Pareto distribution (GPD)'. While GEV describes the limit distributions of normalised maxima, the GPD describes the limit distribution of scaled excesses over length thresholds. Peaks-Over–Threshold (POT) models are introduced to modelling extremes when the occurrence times of the excesses are important.

Steps required in applying EVT include Q–Q plots and distribution of mean excess. While the Q–Q plots help identify the distribution through its histogram, the 'mean excess function' establishes the form of the distribution of mean excess. While building the series of maxima or minima, the series could be divided into non-overlapping blocks. Or, one may choose a given threshold (through the graph of mean excess or Hill graph). To estimate the parameters of extreme distribution, parametric methods like maximum likelihood estimation or p-quantile can be used.

With EVT techniques, it is possible to capture the behaviour of extreme observations. Further, the loss over a very large threshold can be estimated. While these results apply well to the univariate case, the multivariate needs defining the limits. In view of this, the joint distribution of the marginal extreme distributions is not always necessarily an extreme distribution. However, the joint distribution can be used to estimate the probability associated with an extreme scenario. But incorporating this information into the market risk framework remains an issue. The majority of the results are presented assuming that the observations are 'independent and identically distributed (*iid*)'. The sampling of maxima can make this assumption viable.

It is also possible to estimate an index of dependency to incorporate it in the calculation of the extreme *VaR*.

20.6 KEY WORDS

Quantiles : Refers to a set of 'cut off points' that divides a sample data into groups containing nearly equal number of observations.

Fat Tail : Refers to a phenomenon of approximately normal distribution. Fat tail introduces additional risk.

Heavy-Tailed Distribution : Refers to the probability distribution with infinite variance. The simplest heavy-tailed distribution is the Pareto distribution.

Logistic Distribution : This is a continuous probability distribution. This distribution has longer tails than the normal distribution and a higher kurtosis of 1.2 (compared with 0 for the normal distribution).

Degenerate Distribution : This is the probability distribution of a discrete random variable that assigns all the probability (i.e. probability 1) to a single number or to a single point or to just one outcome of a random experiment.

20.7 SUGGESTED BOOKS FOR FURTHER READING

1) Bensalah, Younes (2000). Steps in Applying Extreme Value Theory to Finance: A Review, Bank of Canada Working Paper 2000-20, Bank of Canada, Ottawa (see Internet).

2) Parisi, Fancis (2000). Extreme Value Theory and Standard & Poor's Ratings, Structured Finance Special Report, New York.

3) Parisi, F. and Lund, R. (2000). Seasonality and Return Periods of Land Falling Atlantic Basin.

4) Hurricanes, Australian and New Zealand Journal of Statistics, 42, 271-282.

20.8 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) There are certain extreme events whose occurrence is infrequent but the magnitude of loss is enormous (e.g. Tsunami, hurricane). Insurance companies dealing with occurrences of events with extreme deviations from the median must cover themselves with suitable premiums before

giving coverage to such extreme events. 'Extreme value theory (EVT)' provides solutions to problems associated with extreme risks that are extremely rare. Hence the significance of EVT.

2) One is the 'asymptotic distribution of a series of maxima or minima'. Such distributions are known to converge to one of three well known distributions viz. Gumbel or Frechet or Weibull. These three gives us the 'generalised extreme value' (GEV) distribution. EVT shows that in the case of these three types of limiting distributions, the limiting distribution is a 'Generalised Pareto distribution (GPD). If our interest is to analyse only the peaks i.e. number of times losses over a threshold occurs, we employ the POT (peak over threshold) approach. In this (i.e. POT), the number of events during a given time follows a Poisson process and the exceedances over a given high threshold follow a GPD.

3) It approaches the form of the 'limit distribution' for the 'centred and normalised maxima (CNM)'.

4) If the shape parameter $\xi > 0$, then F belongs to a family of heavy-tailed distributions like Pareto, log-Gamma, Cauchy and t. If $\xi = 0$, then the tail of F decreases exponentially and belongs to a class of medium-tailed distributions like normal, exponential, Gamma and log-normal distributions. If $\xi < 0$, the underlying distribution F is characterised by a finite right short-tail distribution like the uniform or beta distributions.

5) The GEV describes the limit distributions of normalised maxima. The GPD describes the limit distribution of scaled excesses over high thresholds.

6) The POT model can be used to: (i) estimate the excess distribution with respect to a threshold level u and (ii) estimate the tail shape of the original distribution.

Check Your Progress 2

1) (i) exploratory data analysis, (ii) determination of 'mean excess function', (iii) sampling the maxima and (iv) choice of the threshold.

2) The first approach divides the series into non-overlapping blocks of the same length from which we can choose the maximum from each block. This method has the drawback that we may lose extreme observations within the same block. The second approach chooses a given threshold and then considers the extreme observations that exceed this threshold.

3) One, a graphical tool to choose the threshold. 'Hill graphs' is another method. It is defined by the set of points as in Equation (20.10_b).

4) We make two assumptions. First, we assume that the extreme observations follow exactly the GEV distribution. Second, we assume that the other observations are at least roughly distributed like the GEV distribution.

UNIT 21 CREDIBILITY THEORY

Structure

21.0 OBJECTIVES

After reading this unit, you will be able to:

- define 'credibility' specifying the formula for its calculation;
- discuss the concept of 'classical credibility';
- state the meaning of the concept of 'partial credibility' with an empirical example;
- outline the concept of 'Bayesian credibility' with an example;
- explain the difference between 'Bühlmann credibility' and 'Bühlmann-Straub credibility';
- describe the procedure for estimating the credibility parameters;

- write a note on the comparative profile of classical and Buhlmann's credibility indicators; and
- explain the concept of 'maximum aggregate loss and general solution' with an illustration.

21.1 INTRODUCTION

Credibility is a measure of confidence assigned to the data. Credibility theory provide tools to deal with the quality of data used for predicting future losses or costs. For this purpose, we need historical data along with the recent observations. For instance, suppose that the recent experience indicates that skilled workers should be charged a rate of Rs. 5 for workers compensation insurance. Assume that current rate is Rs.10. What should the new rate be? Should it be Rs. 5, Rs. 10 or in between? Credibility theory is used to weigh together these two estimates. The basic expression for calculating credibility weighted estimate is: Z $\times$ [observation] + (1 – *Z*) [other information], $0 \le z \le 1$. Here *Z* is called the 'credibility assigned to the observation' and (1 – *Z*) referred to as the 'complement of credibility'. In the above example, Rs. 10 (current rate) is the other information. Thus, the skilled workers rate of workers compensation insurance is *Z*$\times$ Rs. 5 + (1 – *Z*) Rs. 10. We calculate *Z* for a solution to the issue. The credibility indicator *Z* is thus a function of the 'expected variance of the observations'. This is complemented with the 'selected variance' to be allowed in the first term of the credibility formula i.e. $Z \times$[observation]. In this unit, we study credibility indicators proposed by some contributors like Buhlmann and Straub. We also study the classical and the Bayesian credibility indicators. Buhlmann's credibility is referred to as 'the least squares credibility'. Bayesian credibility combines current observations with prior information to produce a better estimate. Let us begin by studying about 'classical credibility'.

21.2 CLASSICAL CREDIBILITY

In classical credibility, one determines how much more data is needed before assigning 100% credibility to it. Such amount of data is referred to as 'full credibility criterion'. If one has this much of data or more, then the value of *Z* is: z = 1.00. Else, it is between 0 and 1: $0 \le z < 1$. In the latter case, we call it as 'partial credibility'. There are four basic concepts of 'classical probability'. These relate to determining the criterion for full credibility to cover for the three explicit cases of: (i) estimating frequencies, (ii) estimating severities and (iii) estimating pure premiums. The fourth case relates to determining the value that can be assigned for 'partial credibility'. This refers to the ratio or value to be assigned when one has less data than is needed for full credibility. Let us now consider these cases one by one.

21.2.1 Full Credibility for Frequency

Here, we use normal approximation to Poison process. Hence, the probability P that observation X is within $\pm k$ of the mean μ is given by:

$$P = \Pr ob\left[\mu - k\mu \le X \le \mu + k\mu\right]$$

$$= \Pr ob\left[-k\left(\frac{\mu}{\sigma}\right) \le \frac{X-\mu}{\sigma} \le k\left(\frac{\mu}{\sigma}\right)\right] \qquad (21.1)$$

If we denote $u = \frac{X - \mu}{\sigma}$ and assume this to be normally distributed, then for a Poison distribution with expected number of claims n, we have $\mu = n$ and $\sigma = \sqrt{n}$. The probability that the observed number of claims n is within $\pm k$ of the expected number $\mu = n$ is then given by:

$$P = \Pr ob\left[-k\sqrt{n} \le u \le k\sqrt{n}\right] = \Phi\left(k\sqrt{n}\right) - \Phi\left(-k\sqrt{n}\right)$$

$$= \Phi\left(k\sqrt{n}\right) - \left[1 - \Phi\left(k\sqrt{n}\right)\right]$$

$$= 2\Phi\left(k\sqrt{n}\right) - 1 \qquad (21.2)$$

$$\Rightarrow \Phi\left(k\sqrt{n}\right) = \frac{(1+P)}{2} \qquad (2.13)$$

Therefore, to compute the number of expected claims n_0 such that the probability of being within $\pm k$ of the mean is P, we have to set $y = \sqrt{k}$ so that $\Phi(y) = \frac{1+P}{2}$. We can then determine y from the normal table as the one which yields: $n_0 = \frac{y^2}{k^2}$.

21.2.2 Full Credibility for Severity

The classical credibility ideas can also be applied to estimate the claim severity and the average size of a claim. Suppose a sample of N claims $X_1, X_2, X_3, \ldots X_N$ are each independently drawn from a loss distribution having mean μ_s and variance σ_s^2. The severity is then assessed by the mean of the distribution estimated as: $\left(X_1 + X_2 + \ldots + X_N\right)/N$. The variance of the observed severity is then given by:

$$Var\left(\sum_{i-1}^{N}\frac{X_i}{N}\right) = \left(\frac{1}{N^2}\right)\sum_{i=1}^{N} Var\left(X_i\right) = \frac{\sigma_s^2}{N} \qquad (21.4_a)$$

The standard deviation of such a data series is therefore $= \frac{\sigma_s}{\sqrt{N}}$. Hence, the probability that the observed severity S is within $\pm k$ of the mean μ_s is given by: $P = Prob\left[\mu_s - k\ \mu_s \le S \le \mu_s + k\mu_s\right]$. Subtracting throughout μ_s, and dividing by $\frac{\sigma_s}{\sqrt{N}}$ and substituting $u = \frac{S - \mu_s}{\frac{\sigma_s}{\sqrt{N}}}$ we get:

$$P = Prob\left[-k\sqrt{N}\left(\frac{\mu_s}{\sigma_s}\right) \le u \le k\sqrt{N}\left(\frac{\mu_s}{\sigma_s}\right)\right] \qquad (21.4a)$$

As done before, we can define $\Phi(y) = \frac{1+P}{2}$ and set $y = k\sqrt{N}\left(\frac{\mu_s}{\sigma_s}\right)$.

Solving for N we get: $N = \left(\frac{y}{k}\right)^2 \left(\frac{\sigma_s}{\mu_s}\right)^2$. Since $\frac{\sigma_s}{\mu_s} = CVs$ (i.e. the coefficient of variation of the claim size distribution), if n_0 is the 'full credibility standard for frequency', for given p and k, we can calculate $N = n_0 \ CVs^2$. This gives us the 'norm for full credibility for severity'.

21.2.3 Full Credibility for Pure Premiums

Suppose that N claims of sizes $X_1, X_2 \ldots, X_N$ occur during the observation period. The following five quantities are useful in analysing the cost of insuring a risk or group of risks.

i) Aggregate losses $L = \left(X_1 + X_2 + \ldots + X_N\right)$

ii) Pure Premium: $PP = \frac{\left(X_1 + X_2 \ \ldots.. \ + X_N\right)}{Exposures}$

iii) Exposures obtained by dividing the standard of credibility by the no of claims i.e. Loss Ratio: $LR = \frac{\left(X_1 + X_2 + \ldots + X_N\right)}{Earned\ Premium}$

iv) Pure Premium = $\frac{Losses}{Exposures} = \left(\frac{Number\ of\ claims}{Exposures}\right)\left(\frac{Losses}{Number\ of Claims}\right)$

$$= \left(Frequency\right) \ \left(Severity\right)$$

v) When frequency and severity are not independent, process variance of pure premium is equal to:

= (Mean frequency) (Variance of Severity) +

(Mean severity)^2 (Variance of frequency). This can be written as:

$$\sigma_{pp}^2 = \mu_f \ \sigma_s^2 + \mu_s^2 \ \sigma_f^2 \qquad (21.5a)$$

Alternatively, when frequency and severity are not independent, process variance can be obtained by using the variance expression: $V(X) = E\left(X^2\right) - \left\{E(X)\right\}^2$ where X is pure premium. In equation (21.5_a) if we use the Poisson frequency, we then have:

$$\sigma_{PP}^2 = \mu_f\left(\sigma_s^2 + \mu_s^2\right) = \mu_f\left(2nd \quad moment \quad of \quad severity\right) \qquad (21.5b)$$

The subscripts indicate the means and variance of the frequency (*f*) and severity (*s*). Assuming normal approximation, full credibility standards can be calculated. The probability that the observed pure premium *PP* is within $\pm k$ of the mean μ_{pp} is given by:

$$P = \Pr ob\left[\mu_{PP} - k\mu_{PP} \leq PP \leq \mu_{PP} + k\mu_{PP}\right]$$

$$= \Pr ob\left[-k\left(\frac{\mu_{PP}}{\sigma_{PP}}\right) \leq u \leq k\left(\frac{\mu_{PP}}{\sigma_{PP}}\right)\right]$$

where $u = \dfrac{PP - \mu_{PP}}{\sigma_{PP}}$ is a unit normal variable. Now, assuming normal approximation, we can define *y* such that $\Phi\left(y\right) = \dfrac{1+P}{2}$ and $y = k\left(\dfrac{\mu_{PP}}{\sigma_{PP}}\right)$.

If the frequency is a Poisson process and we can assume n_F as the expected number of claims required for full credibility, then: $\mu_f = \sigma_f^2 = n_F$. Further, if the frequency and severity are assumed to be independent, then:

$\mu_{PP} = \mu_f \quad \mu_s = n_F \quad \mu_s$ and $\sigma_{PP}^2 = \mu_f \left(\sigma_s^2 + \mu_s^2\right) = n_F\left(\sigma_s^2 + \mu_s^2\right)$.

Substituting for μ_{pp} and σ_{pp} we get: $y = k\left[\dfrac{n_F \ \mu_s}{\left\{n_F\left(\sigma_s^2 + \mu_s^2\right)\right\}^{\frac{1}{2}}}\right]$.

Solving for n_F we get:

$$n_F = \left(\frac{y}{k}\right)^2 \left[1+\left(\frac{\sigma_s^2}{\mu_s^2}\right)\right] = n_0\left(1+CV_s^2\right) \qquad (21.5c)$$

This is the standard for full credibility of the pure premium with $CV_s = \left(\dfrac{\sigma_s}{\mu_s}\right)$ as the 'coefficient of variation of the severity'. You may note that:

$n_F = n_0 + n_0\, CV_s^2 =$ standard for full credibility of frequency + standard for full credibility of severity.

21.2.4 Partial Credibility

When one has at least the number of claims needed for full credibility, then one assigns 100% credibility to the observations. However, when there is less data than is needed for full credibility, less than 100% credibility is assigned. For this, if *n* is the expected number of claims for the volume of data and n_F is the 'standard for Full credibility', then the 'partial credibility' assigned is

$Z = \sqrt{\frac{n}{n_F}}$. If $n \geq n_F$, then $Z = 1.00$. We use the square root rule for partial credibility for either frequency or severity or pure premium. Let us consider an example to illustrate this. Suppose that the standard for full credibility is 683 claims and one has observed 300 claims. Now the question is how much credibility can be assigned to this data? The answer is: $\sqrt{\frac{300}{683}} = 66.3\%$.

21.2.5 Illustrations

Let us consider some illustrations to help understand the above concepts better.

i) *If the number of claims has a Poisson distribution, compute the probability of being within ±5% of a mean of 100 claims using the normal approximation to the Poison distribution.*

$$2\Phi\left(0\cdot 05\sqrt{100}\right) - 1 = 38\cdot 3\%.$$

ii) *For P = 95% and for k = 5%, what is the number of claims required for full credibility for estimating the frequency?*

$y = 1\cdot 960$. Since $\Phi\left(1\cdot 960\right) = \frac{1+P}{2} = 97.5\%$

$$n_0 = \frac{y^2}{k^2} = \left(\frac{1\cdot 96}{0\cdot 05}\right)^2 = 1537$$

iii) *The coefficient of variation for severity is 3. For P = 95% and k = 5%, what is the number of claims required for estimating the 'full credibility severity'?*

From example *(ii)* above, we have $n_0 = 1537$.

Hence, $N = 1537\cdot(3)^2 = 13{,}833$ claims.

iv) *The number of claims has a Poisson distribution. The mean of the severity distribution is 2000 and the standard deviation is 4000. For P = 90% and k = 5% what is the standard for full credibility of pure premium?*

Note that $n_0 = 1082$ claims and CV = 2.

Hence, $n_F = 1082\left(1+2^2\right) = 5410$ claims.

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) How is credibility defined? What does a 'credibility indicator' convey?

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

2) What is the basic thrust behind 'classical credibility'?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

3) State the four basic concepts of 'classical credibility'.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) How is the 'full credibility' for severity determined in the case of 'classical credibility'?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

5) How is 'partial credibility' empirically defined? Under what condition does the partial credibility criteria changes to that of 'full credibility'?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

21.3 TYPES OF CREDIBILITY MEASURES

In this section, we consider three types of credibility measures viz. (i) Bayesian credibility, (ii) Buhlmann credibility and (iii) Buhlmann-Straub credibility.

21.3.1 Bayesian Credibility

Bayesian analysis is a technique to update a prior hypothesis based on the observations. This theory is based on Baye's theorem. The statement of the theorem, as you may recall, is as follows. If one has a set of mutually disjoint events A_i, then one can write the marginal distribution function *P[B]* in terms of the conditional distribution $P\left[B|A_i\right]$ and the probabilities $P[A_i]$ as:

$$P[B] = \sum_i P(B|A_i)\ P[A_i] \qquad (21.6_a)$$

and $$P[A_i|B] = \frac{P(A_i).\ P[B|A_i]}{\sum_i P[B|A_i]\ .\ P(A_i)} \qquad (21.6b)$$

$P[A_i|B]$ is here the conditional probability of A_i given that B has already occurred.

In order to compute conditional expectation, we take the weighted average over all possible x as:

$$E[X|B] = \sum_x x\ .\ P[X = x|B] \qquad (21.6c)$$

Let us illustrate this with an example. Assume that there are two types of risks each with Bernoulli claim frequencies. One type of risk has 30% probability of a claim and a 70% probability for no claims. The second type has a 50% probability of a claim. Of the universe of risks, ¾ are of the first type with a 30% probability of claim, while ¼ are of second type with a 50% probability of a claim. This data can be arranged in a tabular form as follows:

Type of Risk	Apriori probability that a risk is of the type	Probability of a claim occurring for a risk of type
1	¾	30%
2	¼	50%

Now, if a risk is chosen at random, then the probability of having a claim is given by:

$$\left(\frac{3}{4}\right)(30\%) + \frac{1}{4}(50\%) = 35\%.$$

Thus, the probability of no claims is 65%. Let us now assume that we pick a risk at random and observe no claim. Then, the probability that we are taking a risk of type 1 is given by:

$$P(Type1|n = 0) = P\Big(Type\ \ 1\ \ \text{and}\ \ n = 0\Big) / P(n=0).$$

However, $P\ (Type = 1 \text{ and } n = 0) = P\ (n = 0 \mid \text{Type} = 1).\ P\ (Type = 1)$

$= (0.7)\ (0.75).$

Therefore, $P(Type = 1|n = 0) = \dfrac{P(n = 0|Type1) \cdot P(Type = 1)}{P(n = 0)}$

$$= \frac{(0.7)(0.75)}{0.65} = 0.8077 .$$

This is a special case of Bayes's Theorem: $P(A|B) = P(B|A).\dfrac{P(A)}{P(B)}$. i.e.

$$P(RiskType|Observv) = P(Observv|RiskType) \times P(Risk\ Type) \big/ P(Observv)$$

Let us consider another example. Assume that we pick a risk at random and observe no claim. Now, the probability that we have a risk of Type 2 is given by:

$$P(Type = 2|n = 0) = P(n = 0|Type2).P(Type = 2) \big/ P(n = 0)$$

$$= \frac{(0.5)(0.25)}{0.65} = 0.1923 .$$

21.3.2 Bühlmann Credibility

Bühlmann credibility is also known as least squares credibility or the greatest accuracy credibility. Here, credibility is given by: $Z = N / (N+K)$ where N is the number of observations and K is the Bühlmann credibility parameter. The parameter can be determined by the 'expected value of the process variance (EPV)' and the 'variance of the hypothetical means (VHM)' using analysis of variance. The EPV and the VHM together sum up to the total variance. The Buhlmann credibility parameter is calculated as $K = EPV/VHM$. For N observations, the Bühlmann credibility is given by:

$$Z = \frac{N}{N + k} \tag{21.7}$$

Note that if $N = 1$, then $Z = 1/1+K = \dfrac{VHM}{VHM + EPV} = \dfrac{VHM}{Total\ \text{variance}}$

In practice, one computes EPV and VHM for a single observation and then plugs it into the equation of Bühlmann credibility with the number of observation N. If one is estimating claim frequencies or pure premiums then N is in exposures. If one is estimating claim severities then N is in the number of claims. This means N is in the units of whatever is in the denominator of the quantity one is estimating. There are certain assumptions underlying Buhlmann's Z. These are:

i) the complement of credibility is given to the overall mean,

ii) the credibility is determined as the slope of the weighted least squares line to the Bayesian estimates,

iii) the risk parameters and risk process do not shift over time,

iv) the expected value of the process variance of the sum of N observations increases with N. Therefore, the expected value of the process variance of the average of N observations decreases with $1/N$ and

v) the variance of the hypothetical means of the sum of N observations increases as N^2. Therefore, the variance of the hypothetical means of the average of N observations is independent of N.

21.3.3 Buhlmann-Straub Credibility

The Bühlmann-Straub credibility indicator assumes that the means of the random variables are equal for the selected risks. But the process variances are inversely proportional to the size (i.e. exposure) of the risk during each observation period. This means, when the risk is twice as large, the process variance is halved. These assumptions are summarised in Table 21.1.

Table 21.1: Assumptions of Bühlmann-Starub Credibility

Exposure	Period 1/ m_1	…	Period N / m_N…
Hypothetical mean for risk θ per unit of exposure	$\mu(\theta) = E_{X\|\Theta}\left[X_1,\|\theta\right] =$	…	$= E_{X\|\Theta}\left[X_N\|\theta\right]$…
Process variance for risk θ	$Var_{X\|\Theta}\left[X_1\|\theta\right] = \frac{\sigma^2(\theta)}{m_1}$	…	$Var_{X\|\Theta}\left[X_N\|\theta\right] = \frac{\sigma^2(\theta)}{m_N}$

The random variable X_t represents the number of claims or monetary losses or some other quantity of interest per unit of exposure with m_t as the measure of exposure. Note that a weighted average using the exposure m_t gives a linear estimator for $\mu(\theta)$ with minimum variance. Further, since m is defined as: $m = \sum_{t=1}^{N} m_t$, the weighted average is obtained as: $\bar{X} = \sum_{t=1}^{N}\left(\frac{m_t}{m}\right)X_t$. Hence, the variance of each X_t, given θ, is $\frac{\sigma^2(\theta)}{m_t}$. For a weighted average $\bar{X} = \sum_{t=1}^{N} W_t\ X_t$, the variance of $\bar{X}$ is minimised by choosing the weights as $w_t = \frac{m_t}{m}$. Hence:

$$E_{X|\Theta}\left[\bar{X}|\theta\right] = E_{X|\Theta}\left[\sum_{t=1}^{N}\left(\frac{m_t}{m}\right) X_t|\theta\right]$$

$$= \sum_{t=1}^{N}\left(\frac{m_t}{m}\right) E_{X|\Theta}\left[X_t|\theta\right] = \sum_{t=1}^{N}\left(\frac{m_t}{m}\right)\mu(\theta) = \mu(\theta) \quad (21.8_a)$$

$$Var_{X|\Theta}\left[\bar{X}|\theta\right] = Var_{X|\Theta}\left[\sum_{t=1}^{N}\left(\frac{m_t}{m}\right) X_t|\theta\right]$$

$$= \sum_{t=1}^{N}\left(\frac{m_t}{m}\right)^2 Var_{X|\Theta}\left[X_t|\theta\right]$$

$$= \sum_{t=1}^{N}\left(\frac{m_t}{m}\right)^2 \left[\frac{\sigma^2(\theta)}{m_t}\right] = \frac{\sigma^2(\theta)}{m} \qquad (21.8_b)$$

The *EPV* and *VHM* are now defined as:

$EPV = E_{\Theta}\left[\sigma^2(\Theta)\right]$ and $VHM = Var_{\Theta}\left[\mu(\Theta)\right]$. The expected value is over all risk parameters in the population. This means:

$$E\left(\bar{X}\right) = E_{\Theta}\left[E_{X|\Theta}\left[\bar{X}|\Theta\right]\right] = E_{\Theta}\left[\mu[\Theta]\right] = \mu \qquad (21.9_a)$$

and $Var\left(\bar{X}\right) = Var_{\Theta}\left[E_{X|\Theta}\left[\bar{X}|\Theta\right]\right] + E_{\Theta}\left[Var_{X|\Theta}\left[\bar{X}|\Theta\right]\right]$

$$= Var_{\Theta}\left[\mu(\Theta)\right] + \frac{E_{\Theta}\left[\sigma^2(\Theta)\right]}{m} = VHM + \frac{EPV}{m} \qquad (21.9_b)$$

As in the Bühlmann case, here also the 'credibility assigned' to the estimator $\bar{X}$ of $\mu(\theta)$ is:

$$Z = \frac{VHM}{Total\ \ Variance\ \ of\ \ the\ \ estimator\ \ \bar{X}}.$$

$$= \frac{VHM}{VHM + \frac{EPV}{m}} \qquad (21.9_c)$$

Multiplying the numerator and denominator by $\frac{m}{VHM}$, we get:

$$Z = \frac{m}{m + K} \qquad (21.9_d)$$

The total exposure m replaces N in the Bühlmann formula. The parameter K is defined as:

$$K = \frac{EPV}{VHM} = \frac{E_{\Theta}\left[\sigma^2(\Theta)\right]}{Var_{\Theta}\left[\mu(\Theta)\right]} \qquad (21.9_e)$$

Note that the Bühlmann model is actually a special case of Bühlmann-Straub model with $m_t = 1$, for all t. The credibility weighted estimate is:

$$\hat{\mu}(\theta) = Z\bar{X} + (1-Z)\mu \qquad (21.9_f)$$

21.3.4 Illustration

The actuaries at the Good Health Insurance Company calculates prospective premiums for group insurance policies using a Bühlmann-Straub credibility model. Analysis of Good Health's data leads to the following assumptions for its business: (i) for all policies together, the prospective average annual expected pure premium per insured person is 2,400, (ii) the variance of the hypothetical means that the pure premiums across group plans is 500,000 and (iii) the expected value of the process variance in annual costs per insured person is 250,000,000. One of Good Health's clients had the following experience during a one-year period with costs adjusted to reflect prospective costs.

Groups policy	*Insured persons*	*Cost per insured person*
1	*240*	*3000*

Calculate a credibility weighted pure premium for group policy 1.

Here, $K = \left(\dfrac{250{,}000{,}000}{5{,}00{,}000}\right) = 500$

$$\mu = 2{,}400,\ m = 240,\ \overline{X} = 3000$$

Hence, $Z = \dfrac{240}{240+500} = 0\cdot 3243$

Therefore, estimated pure premium is equal to:

$$= (0\cdot 3243)\ (3000) + (1-0\cdot 3243)\ (2{,}400) = 2{,}594\cdot 58.$$

21.4 ESTIMATORS AND COMPARATIVE PROFILE

The selection of credibility parameters requires a balancing of responsiveness versus stability. Larger credibility weights put more weight on the observations. This means, current data have a larger impact on the estimates i.e. the estimates are more responsive to current data. But this comes at the expense of less stability in the estimates. Credibility parameters are often therefore selected to reflect the actuary's desired balance between responsiveness and stability. In classical credibility, we chose P and k values such that the observation X is within $\pm k$ percent of the mean μ. Usually we keep P = 9% and k = 5%.

21.4.1 Estimators of Credibility Parameters

Suppose that there are M risks in a population and that they are similar in size. Assume that we track the annual frequency year by year for Y years for each of the risks. Then the frequencies are like:

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1Y} \\ X_{21} & X_{22} & \cdots & X_{2Y} \\ \cdots & \cdots & \cdots & \cdots \\ X_{M1} & X_{M2} & \cdots & X_{MY} \end{bmatrix}$$

X_{ij} here represents the frequency of i^{th} risk in j^{th} year. Now, the different estimators are as given below:

Case	Notation	Estimator
Mean frequency for risk i	$\overline{X_i}$	$\frac{1}{Y}\sum_{j=1}^{Y} X_{ij}$
Mean frequency for population	$\overline{X}$	$\frac{1}{M}\sum_{i=1}^{M} \overline{X}_i$
Process variance for risk i	$\hat{\sigma}_i^2$	$\left[\frac{1}{(Y-1)}\right]\sum_{j=1}^{Y}\left(X_{ij} - \overline{X}_i\right)^2$
Expected value of process variance	EPV	$\frac{1}{M}\sum_{i=1}^{M} \hat{\sigma}_i^2$
Variance of the hypothetical means	VHM	$\left[\frac{1}{(M-1)}\right]\sum_{i=1}^{M}\left(\overline{X}_i - \overline{X}\right)^2 - \frac{EPV}{Y}$

Since $K = \frac{EPV}{VHM}$, if sample VHM is zero or negative then the credibility is assigned a value of 0. Further, since a credibility formula of the form $Z = \frac{N}{N+K}$ is usually used to weigh the policyholder's experience, one needs to estimate K by observing which values of K have worked well in the past. The goal is for each policyholder to have the same expected loss ratio after the application of experience rating. Let LR_i be the loss ratio for policyholder i where, in the denominator, we use the premiums after the application of experience rating. Let LR_{AVE} be the average loss ratio for all policyholders. Then, we can define $D(K)$ to be:

$$D(K) = \sum_{all\ i} \left(LR_i - LR_{AVE} \right)^2 \tag{21.10}$$

The sum of squares of the differences is a function of *K*, the credibility parameter that was used in the experience rating. The goal is to find a *K* that minimizes *D*(*K*). This requires re-computing the premium that each policyholder would have been charged under a different K' value. This generates new LR_i's which can then be put into the above equation to compute $D(K')$. Using techniques from numerical analysis, a *K* that minimises *D(K)* can be found out. Another approach is to calculate the credibility parameters by 'linear regression analysis' of a policyholder's current frequency, pure premium, etc. Using historical data for many policyholders we can set up the regression equation as:

Observation in the year *Y* = *m*[observation in year (Y – 1)] + constant (21.11)

The slope *m* from a least squares fit to the data turns out to be the Bühlmann credibility *Z*. The constant term is (1 – *Z*) * (overall average). After we have calculated the parameters using historical data, we can estimate future results using the recent data.

21.4.2 Comparison of Classical and Buhlmann Credibility

Although the equations for classical credibility and Bühlmann credibility look very different, they produce similar results. The most significant difference between the two models is that Bühlmann credibility never reaches *Z* = 1.00, which is an asymptote at the curve. Both models can be effective at improving the stability and accuracy of estimates. Classical credibility and Bühlmann credibility methods will produce approximately the same credibility weights. If the full credibility standard for classical credibility is n_0, it is about 7 to 8 times larger than the Bühlmann credibility parameter *K*.

For a particular application, the actuary can choose the model [out of the three models (classical, Bühlmann and Baysian)] that is appropriate to the goals and data. If the goal is to generate the most accurate insurance rates, with least squares as the measure of fit, then Bühlmann credibility may be the best choice. Bühlmann credibility forms the basis of most experience-rating plans. It is often used to calculate class rates in a classification plans. The use of Bühlmann credibility requires an estimate of the *EPV* and *VHM*.

Classical credibility might be used if estimates for the *EPV* and *VHM* are unknown or difficult to calculate. Classical credibility is often used in the calculation of overall rate increases. Often it is simpler to work with Bayesian analysis. Bayesian analysis can be an option if the actuary has a reasonable estimate of the prior distribution. However, Bayesian analysis too is complicated to apply and is the most difficult of the methods to non actuaries.

21.4.3 Illustrations

Let us consider some empirical illustrations as before.

(i) A sample of 100 claims is distributed as follows:

Size of claim	Number of claim
$1,000	*85*
$5,000	*10*
$10,000	*3*
$25,000	*2*

Estimate the coefficient of variation of the claim severity based on this distribution.

The sample mean and sample standard deviation are:

$$\hat{\mu} = (0.85)(1000) + (0\cdot10)(5000) + 0\cdot03(10,000) + 0\cdot02(25,000) = 2,150.$$

$$\hat{\sigma} = \left[\frac{1}{(100-1)}\left\{85\,(1000-2150)^2 + 10(5000-2,150)^2 + 3(10,000-2,150)^2 + 2(25,0000-2,150)^2\right\}\right]^{\frac{1}{2}} = 3.791.$$

Note that we are dividing by $n - 1$ to calculate an unbiased estimate. Now:

$$CVs = \frac{\hat{\sigma}}{\hat{\mu}} = 1\cdot76$$

$P = 90\%$ & $k = 5\%$.

From the table of standards for full credibility for frequency (claims) $n_0 = 1,082$. So, the full credibility standard for the pure premium is $n_F = n_0(1 + CVs)^2 = 1,082\,(1+1\cdot76^2) = 4,434$.

(ii) There are two auto drivers in a particular rating class. The first driver had the following sequence of claims, in year 1 through 5:2,0,0,1,0 and the second had: 1,1,2,0,2. What are the estimated future claims frequency for the two drivers?

For first driver, $\overline{X_i} = 0\cdot6$ and $\hat{\sigma}_1^2 = 0\cdot80$

For the second driver $\overline{X_2} = 1\cdot2$ and $\hat{\sigma}_2^2 = 0\cdot7$

$$\overline{X} = \frac{\overline{X_1} + \overline{X_2}}{2} = 0\cdot90.$$

$$EPV = \frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}{2} = 0\cdot75.$$

$$VHM = (0\cdot6 - 0\cdot9)^2 + \frac{(1\cdot2-0\cdot9)^2}{1} - \frac{0\cdot75}{5} = 0\cdot03$$

$$K = \frac{EPV}{VHM} = 25$$

$$Z = \frac{5}{5+25} = \frac{1}{6}$$

Thus the estimated future claim frequency for first driver is:

$$\left(\frac{1}{6}\right)(\cdot 6) + \left(\frac{5}{6}\right)(\cdot 9) = 0 \cdot 85$$

and for the second driver is:

$$\left(\frac{1}{6}\right)(1 \cdot 2) + \left(\frac{5}{6}\right)(\cdot 9) = 0 \cdot 95 .$$

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) How is Bühelmann's credibility indicated?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

2) Indicate how the Bühlmann's credibility parameter is estimated.

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

3) What are the significant features of Bühlmann's and Classical credibility measures?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

4) How can a choice between the Bühlmann's and the classical credibility measures be made?

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

………………………………………………………………………………

21.5 MAXIMUM AGGREGATE LOSS AND GENERAL SOLUTION

Aggregate losses, pure premiums and loss ratios depend on both the number of claims and the size of claim. They have more reasons to vary than either frequency or severity. Severity are more difficult to estimate than frequencies. Other things being equal, the standard for full credibility is kept larger. The formulae for the variance of the pure premium, as seen before, are calculated as:

$\sigma_{pp}^2 = \mu_f\ \sigma_S^2 + \mu_S^2\ \sigma_f^2$ (for the general case) and

$\sigma_{pp}^2 = \mu_f \left(\sigma_s^2 + \mu_s^2\right) = \mu_f$ (2[nd] moment of the severity) (for the Poisson frequency).

The subscripts indicate the means and variance of the frequency (*f*) and severity (*S*). Assuming normal approximation, the probability that the observed pure premium *P* is within ±*k* of the mean μ_{PP} is given by:

$$P = \text{Prob}\left[\mu_{PP} - k\mu_{PP} \le PP \le \mu_{PP} + k\mu_{PP}\right]$$

$$= \Pr ob\left[-k\left(\frac{\mu_{PP}}{\sigma_{PP}}\right) \le u \le k\left(\frac{\mu_{PP}}{\sigma_{PP}}\right)\right]$$

where $u = \dfrac{PP - \mu_{PP}}{\sigma_{PP}}$ is a unit normal variable. Consequently, we have the following results:

i) Define *y* such that $\phi(y) = \dfrac{1+P}{2}$. Then, in order to have probability *P* that the observed pure premium will differ from the true pure premium by less than ±k μ_{PP}:

$$y = k\left(\frac{\mu_{PP}}{\sigma_{PP}}\right) \qquad (21.12_a)$$

ii) If the frequency is a Poisson process, and n_F is the expected number of claims required for full credibility of the *p*, given n_F as the expected number of claims, we have $\mu_f = \sigma_f^2 = n_F$. Assuming frequency and severity are independent, we have:

$$\mu_{PP} = \mu_f \quad \mu_s = n_F n_s \qquad (21.12_b)$$

and

$$\sigma_{PP}^2 = \mu_f\left(\sigma_s^2 + \mu_s^2\right) = n_F\left(\sigma_s^2 + \mu_s^2\right) \qquad (21.12_c)$$

Substituting (21.12_b) and (21.12_c) in (21.12_a) we get:

$$y = k\left[\frac{n_F \ \mu_s}{\left\{n_F\left(\sigma_s^2 + \mu_s^2\right)\right\}^{\frac{1}{2}}}\right] \qquad (21.12_d)$$

Solving for n_F we get:

$$n_F = \left(\frac{y}{k}\right)^2 \left[1 + \left(\frac{\sigma_s^2}{\mu_s^2}\right)\right]$$

$$= n_0\left[1 + CVs^2\right] \qquad (21.12_e)$$

This is the standard for full credibility of the pure premium where $n_0 = \left(\frac{y}{k}\right)^2$ is the standard for full credibility for frequency. *CV*s is the coefficient of variation for severity. Equation (21.12e) can also be written as $n_F = \frac{n_0\left(\mu_s^2 + \sigma_s^2\right)}{\mu_s^2}$, where $\left(\mu_s^2 + \sigma_s^2\right)$ is the second moment of the severity distribution. Hence, the standard for full credibility of pure premium is: $n_0\left(1 + CVs^2\right) = \ n_0 + n_0 \ CVs^2$ = standard for full credibility of frequency + standard for full credibility of severity.

Note that if one limits the size of claims, then the coefficient of variation is smaller. Therefore, the criterion for full credibility for 'basic limits losses' is less than that for 'total losses'. It is a common practice in rate making to cap losses so as to increase the credibility assigned to the data. Pure premiums are often approximately normal. Generally, the greater the expected number of claims (or the shorter tailed the frequency) and severity distributions, the better are the normal approximation. It is therefore assumed that one has enough claims that the aggregate losses approximate a normal distribution. Let us consider an illustration. Say, we are given the following information for a group of insured: (i) prior estimate of expected total losses $20,000,000, (ii) observed total losses: $25,000,000, (iii) observed number of claims: 10,000 and (iv) required number of claims for full credibility: 17,500. Let us calculate a credibility weighted estimate of the group's expected total losses.

We have $Z = \sqrt{\frac{10,000}{17,5000}} \ = \ 75 \cdot 6\%$. Therefore, the new estimate is:

$(25\text{million})(0 \cdot 756) + (20\text{million})(1 - 0 \cdot 756) \ = \$ \ 23.78$ million.

21.6 LET US SUM UP

In this unit, we have discussed the tools of credibility theory helpful in predicting events characterised by randomness. Depending upon the information available, three important approaches (viz. classical, Bühlmann and Bayesian) have been discussed. The classical credibility is based on the idea of determining the amount of data needed for assigning 100%

credibility. Bühlmann theory, on the other hand, is built upon the minimum expected variance in the process (i.e. the least squares approach). The Bayesian analysis relies on combining current observations with prior information. Making use of these alternatives, we have seen that the number of claims, the average size of claims (severity), pure premium and aggregate loss could be estimated for actuarial purposes.

21.7 KEY WORDS

Conditional Expectation : This is defined in two cases as follows. For the discrete case this is defined as the conditional expectation or mean values of a joint probability function $g(X, Y)$ given that $Y = y_j$. In other words, this is defined as the expectation of the function $g\left(X, y_j\right)$ given that the conditional distribution of $Y = y_j$. For the continuous case, it is defined as the conditional expectation of $g\left(X, Y\right)$ given that $Y = y$. This means:

$$E\left[g\left(X,Y\right)\middle|Y=y\right]=\int_{-\infty}^{\infty} g\left(x;\; y\right) f_{X|Y}\left(x|y\right)dx$$

Conditional Variance : The conditional variance of X given $Y = y$ is defined as:

$$V\left(X|Y=y\right)=E\left[\left\{X-E\left(X|Y=y\right)\right\}^2|Y=y\right].$$

EPV *(Expected Value of the Process Variance)* : The expected value of the process variance is the average value of the process variance over the entire class of risks.

Hypothetical Mean : Here, each risk in the class is considered to have its own individual mean risk (called the hypothetical mean).

Process Variance : This is the variance of the risk's random experience about its expected value. This is a measure of the variability in an individual risk's loss experience.

VHM: (*Variance of Hypothetical Mean)* : This is defined as the variance of hypothetical mean risks in the class. This is a statistical measure for the homogeneity.

21.8 SUGGESTED BOOKS FOR FURTHER READING

1) Herzog, Thomas N. (1996). An introduction to Credibility, Mad River Books.

2) Hossack I. B., J.H. Pollard and B Zehn Wirth (1983). Introductory Statistics with Applications in General Insurance, Cambridge University Press, New York.

3) Klugman Stuart A Harry H Panjer, Gordon E and Will M (1998). Loss Models: From Data to Decisions, John Wiley and Sons, New York.

21.9 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Credibility is defined as a 'measure of confidence assigned to the data'. In its basic form, it is stated as a weighted estimate like: $Z \times$ [observation based information] + $(1 - Z)$ [other information]. Here Z is called the 'credibility assigned to the observation' and $(1 - Z)$ is referred to as the 'complement of credibility'.

2) In classical credibility, one determines how much more data is needed before assigning 100% credibility to it. Such amount of data is referred to as 'full credibility criterion'. If one has this much of data or more, then the value of Z is: $z = 1.00$.

3) Out of the four, three refer to 'full credibility'. These three refer to the respective cases of estimating: (i) frequencies, (ii) severities and (iii) pure premiums. The fourth case relates to determining the value to be assigned for 'partial credibility'. Evidently, this refers to cases where one has less data than is needed for full credibility.

4) If n_0 is the 'full credibility standard for frequency', and we are given p and k, we calculate N as: $N = n_0 \; CVs^2$ where CV stands for 'coefficient of variation'. N gives us the criteria for a measure of full credibility severity.

5) If n is the expected number of claims for the volume of data and n_F is the 'standard for full credibility', then the 'partial credibility' is defined as $Z = \sqrt{\frac{n}{n_F}}$. When $n \geq n_F$, Z is assigned the value 1.00.

Check Your Progress 2

1) It is indicated as: $Z = N / (N+K)$ where N is the number of observations and K is the Bühlmann credibility parameter.

2) The parameter is determined as the ratio of 'expected value of the process variance (*EPV*)' and the 'variance of the hypothetical means (*VHM*)'. That is: $K = EPV/VHM$. More specifically, for N observations, the Bühlmann credibility is given by: $Z = \frac{N}{N+k}$.

3) (i) Bühlmann credibility being an asymptote at the curve never reaches $Z = 1.00$ (ii) Both measures produce approximately the same credibility weights.

4) If the goal is to generate the most accurate insurance rates, with least squares as the measure of fit, then Bühlmann credibility may be the best choice. However, since it requires the estimates of *EPV* and *VHM*, classical credibility might be used if estimates for the *EPV* and *VHM* are unknown or difficult to calculate.

UNIT 22 DYNAMIC FINANCIAL ANALYSIS

Structure

22.0 OBJECTIVES

After reading this unit, you will be able to:

- outline the concept of 'dynamic financial analysis (DFA)' with the two broad techniques or approaches adopted under it;
- distinguish between the 'asset liability management (ALM)' approach and the approach of 'stochastic simulations';
- state the concept of 'efficient frontier' with its uses in DFA;
- specify the assumptions required, with its applicability, for a 'stochastic scenario generator' in a DFA model;
- write a note on 'stochastic variables' in DFA;
- list the risks that a DFA seeks to address in its scope;
- discuss the impact of 'short term interest rates, term structure and general inflation' on a DFA,
- apply the principle of 'capital asset pricing model (CAPM)' to derive equations for 'stock returns' in a DFA framework;
- describe the application of DFA for the cases of 'non-catastrophe and catastrophe' situations;
- write a note on 'underwriting cycles and payment patterns'; and

- explain the structure and significance of a 'corporate model' under the DFA framework.

22.1 INTRODUCTION

Insurance companies, especially those dealing with non-life insurance, often witness pricing cycles. These cycles are accompanied by volatile insurance profits and increasing catastrophic losses. Under such circumstances, shareholders valuation, as well as solvency positions of companies, are seriously affected. It is, therefore, necessary to evaluate the economic factors behind such cycles. Such evaluations help us identify their nature and interrelationships. Keeping this in view, the discussion in this last unit of this course, centres around the important theme of 'dynamic financial analysis (DFA)'. There are two primary techniques available to analyse the financial effects of different business strategies viz. (i) scenario testing and (ii) stochastic simulation. The latter is what is referred to as 'dynamic financial analysis (DFA)' in financial literature. The 'scenario testing' predicts the business results under 'deterministic scenarios'. This actually refers to situations when risks are associated with a specific case that can be quantified. The other technique, viz. the DFA, overcomes this limitation. It does this by allowing for the full probability distributions of important variables (such as surplus, written premiums or loss ratios) to reveal what can otherwise be not easily grasped.

22.2 STOCHASTIC SIMULATIONS

Traditional Asset-Liability Management (ALM) considers life insurance liabilities as deterministic. This is due to the assumed feature of 'low variability'. This approach cannot be extended to non-life insurance where we are faced with much more volatile liability of cash flows. In non-life insurance, both the date of occurrence and the size of claims are uncertain. Moreover, the claim costs are information sensitive (in contrast to life insurance where they are expressed in nominal terms). Hence, non-life insurance liabilities and assets are dealt with by 'stochastic simulations' or the DFA.

DFA is a part of the financial management of a firm. In view of this, risk control function of DFA are analysed under the management of 'financial stability'. In its first task, it aims at maximising the shareholder valuation of the firm. Its second task seeks to serve maintenance of 'customer value'. Within these two parameters, DFA explains the strategic management decisions. Such decisions relate to aspects like: asset allocation, capital allocation, performance measurement, market strategies, business mix, pricing decisions and product design. The long list of aspects suggests that DFA can be used to answer much broader issues relating to the shareholders and the management. For this, it is necessary to start with a fixed time horizon. On the one hand, we would like to model over a long time period in order to plan for the long-term effects of a particular strategy. Such effects concern the long-tail businesses realised after a lapse of some years. They can hence hardly be recognised in the first few years. On the other hand, simulated values become more unreliable when the projection period is longer. This is due to the accumulation process of 'parameter risks over time'. Therefore, a projection period of five to ten years becomes a reasonable choice. Usually, the time period is split into yearly, quarterly or monthly sub-periods.

22.2.1 Efficient Frontier

The most common framework used in DFA is the 'efficient frontier concept' which is used in 'modern portfolio theory'. In this, we choose a 'return measure' (e.g. expected surplus) and a 'risk measure' (e.g. expected policyholder deficit). Then the measured risk and return of each strategy is plotted as in Figure 22.1. The figure gives for each strategy one spot in the risk-return intersection. A strategy is regarded as 'efficient' if there is no other strategy with a lower risk at the same level of return or higher return at the same level of risk. For each level of risk, there is a maximum return that cannot be exceeded. This gives us an 'efficient frontier'. Such an efficient frontier serves as a tool to compare different strategies through the variables of risk and return. Although efficient frontiers are a good means of communicating the results of DFA, it is criticised on certain grounds. A typical efficient frontier uses risk measures that mixes the systematic risk (non-diversifiable by shareholders) and the non-systematic risks together.

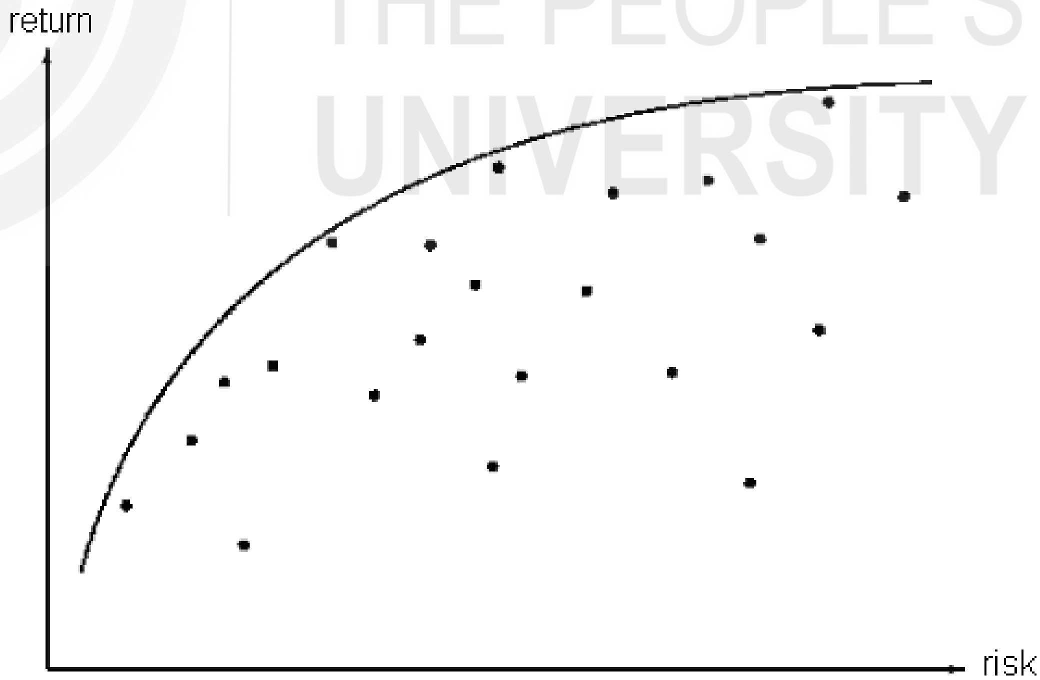

Fig. 22.1: Efficient frontier
Source: Kaufmann et al (2001)

This blurs the shareholder value perspective. Secondly, efficient frontiers give misleading insight if they are used to address investment decisions after the concept of systematic risk has been plugged into the 'decision equation'.

DFA is used as a solvency testing tool. This is done where the financial position of the company is evaluated from the perspective of the customers. It is used to quantify in probabilistic terms whether the company will be able to meet its commitments in the future. In the process, it determines the necessary amount of capital necessary for a given level of risk.

22.2.2 Stochastic Scenario Generator

A 'stochastic scenario generator' yields the main 'structure of a DFA model'. It produces realisations of random variables representing the most important drivers of business results. Such a realisation of a random variable (in the course of simulation) corresponds to fixing a scenario. The data source should consist of company specific inputs like mean severity of losses. This should be per line of business and per accident year. Assumptions required are on: (i) parameters (e.g. long-term mean rate like a mean reverting interest rate model), and (ii) investment strategy. The output generated by the DFA model can then be used to improve the strategy i.e. to make new strategic assumptions.

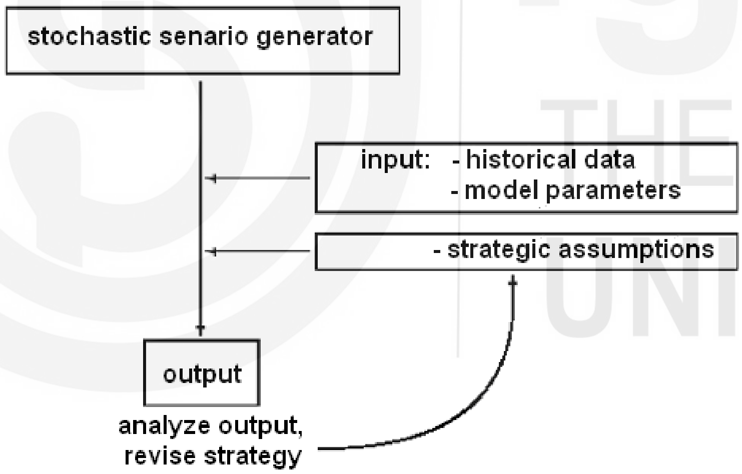

Fig. 25.2: Main structure of a DFA model

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) What is 'dynamic financial analysis (DFA)'? What are the two approaches adopted under it?

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

2) Distinguish between the two approaches of ALM and DFA.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

3) Indicate the scope of DFA in terms of the areas it seeks to cover.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) Define the term 'efficient frontier'. How is it useful?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

5) On what grounds, the concept of 'efficient frontier' is criticised?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

6) How is a 'stochastic scenario generator' useful?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

22.3 STOCHASTIC VARIABLES

While building a stochastic model, it is necessary to identify the key random variables affecting the asset and liability cash flows. Specifically, the risks affecting the financial position of a life/non-life insurer can be categorised in various ways. For instance, we could have groups like pure asset, pure liability and asset/liability risks. A DFA model should at least address the following risks: (i) risk of inadequate premiums, (ii) interest rate risk, (iii) risk of insufficient reserves, (iv) investment risk (associated with volatile investment returns and capital gains) and (v) catastrophes. We may also have to consider credit risks related to reinsurer default, currency risk and exchange rate risks. A critical part of a DFA model is the interdependencies between different risk categories. We therefore have to take exclusive account of risks associated with the asset side and those belonging to liabilities.

While evaluating the above variables, it is critically necessary to evaluate the interest rate risk. This is because non-life insurance companies are often exposed to interest rate behaviour due to large investments in fixed income assets. It is therefore necessary to consider the correlation of interest rates with inflation. Consequently, the inflationary influences of the future changes in 'claim size' and 'claim frequency' also needs considered. Due to the correlation between interest rates and stock returns, investment returns are also affected. The DFA structure chosen should explicitly recognise all these features. On the liability side, four sources of randomness (viz. non-catastrophe losses, catastrophe losses, underwriting cycles and payment patterns) needs to be considered. We need to factor catastrophes separately due to their statistically different behaviour than the non-catastrophe losses. Such a separation helps in having more homogeneous data for non-catastrophe losses. It makes the fitting of data by well-known (right skewed) distributions easier. In case of reinsurance, evaluation of external events with severity of claims needs to be treated separately.

In case of non-life insurance, 'underwritings' reflect market and macroeconomic conditions. It refers to the process of selecting risks that affect the business results and therefore needs to be included in a DFA model. In this, the risks are classified as per the degree of their insurability. Appropriate rates of insurance are then assigned. The process includes the rejection of risks that do not qualify for insurance. Losses influence insurance firms by their size and piecewise payment over time. They increase the uncertainties of the claims process. This gets further compounded by the time value of money which varies with inflation. It is therefore necessary to model claim frequency, severity, inflation, etc. to balance the uncertainties involved in the settlement process. In order to allow for such risks, stochastic payment patterns are used as a means of estimating loss reserves. This is done either on a gross or net basis.

22.3.1 Short Term Interest Rate, Term Structure and Inflation

There are some general features identified on interest rates. These are: (i) volatility of yields varies at different maturities, (ii) interest rates are mean-reverting, (iii) rates at different maturities are positively correlated, (iv) interest rates should not be allowed to become negative and (v) the volatility of interest rates should be proportional to the level of the rate. In addition to these features, some practical issues specified to be considered are that they should be: (i) *flexible* enough to cover most situations arising in practice, (ii) *simple* enough so that one can compute answers in reasonable time, (iii) *well-specified* so that required inputs can be provided or estimated and (iv) *realistic* so that the model will not include unsubstantiated elements. Clearly, an interest rate model meeting all these criteria does not exist. Therefore, most models rely on the one-factor viz. 'Cox–Ingersoll–Ross (CIR)' factor. In this, the instantaneous rate is modelled into DFA as a special case of an Ornstein– Uhlenbeck process. Thus:

$$dr = k(\theta - r)dt + \sigma r^{\gamma} dZ \tag{22.1}$$

In the above, by setting $\gamma = 0.5$, we arrive at CIR, by a process known as the square root process, as:

$$dr_t = a(b - r_t)dt + s\sqrt{r_t}dZ_t \tag{22.2}$$

In (22.2), r_t = instantaneous short-term interest rate, b = long-term mean rate, a = constant that determines the speed of reversion of the interest rate towards its long-run mean b, s = volatility of the interest rate process and Z_i = standard Brownian motion. CIR is a mean-reverting process where the short term rate almost surely stays positive. For simulating the short term rate dynamics over the projection period, we need to discretise the mean reverting model in (22.2). This is done as:

$$r_t = r_{t-1} + a(b - r_{t-1}) + s\sqrt{r_{t-1}}Z_t, \tag{22.3}$$

In (22.3), $Z_t \sim N(0,1), Z_{1,}Z_{2,}, ...$ *i.i.d.* with *a, b, s* being same as in (22.2). Although the short rate process does not have negative values, in its discrete version, the last equation's probability is not zero. To meet such a condition, we change Equation (22.3) as:

$$r_t = r_{t-1} + a(b - r_{t-1}) + s\sqrt{{r_{t-1}}^{+}}Z_t. \tag{22.4}$$

A generalisation of CIR is given by the following equation, where setting $g = .05$ again yields the CIR. That is:

$$r_t = r_{t-1} + a(b - r_{t-1}) + s\left({r_{t-1}}^{+}\right)^{g} Z_t \tag{22.5}$$

The above is the general version. It is flexible in determining the degree of dependence between 'conditional volatility of interest rate changes' and 'the level of interest rates'. It is important to note that to calibrate the equation (22.5): (i) choice of value of *g* is not unique, (ii) the parameters, *a*, *b*, *s* and *g* must be determined to ensure that modelled spot rates (based on the instantaneous rate) correspond to empirical term structures derived from traded financial instruments and (iii) for inflation, a suitable factor based on historical data needs applied. However, it is not certain that the future evolution will follow historical patterns. Therefore, besides historical data, economic reasoning and actuarial judgement needs considered.

Term Structure: Based on equation (22.2), we can calculate the prices $F\left(t,T,\left(r_t\right)\right)$ at time t of zero-coupon bonds. Recall that they would pay 1 monetary unit at time of maturity $t+T$. Thus:

$$F\left(t,T,\left(r_t\right)\right)=\mathrm{E}_Q\left[e^{-\int_0^T r_{t+s}ds}\mid\gamma_t\right]=e^{\log A_T-r_t\ B_T} \qquad (22.6)$$

where, $A_T=\left(\dfrac{2Ge^{(a+G)T/2}}{(a+G)\left(e^{GT-1}\right)+2G}\right)^{2_{ab/s^2}}$ and $B_T=\dfrac{2\left(e^{GT}-1\right)}{(a+G)\left(e^{GT}-1\right)+2G}$ with $G=\sqrt{a^2+2s^2}$. The spot rates $R_{t,T}$ at time t are derived as continuously compounded Equation (22.6). This determines the term structure of zero-coupon yields at time t as:

$$R_{t,T}=-\frac{\log F\left(t,T,\left(r_t\right)\right)}{T}=\frac{r_tB_T-\log A_T}{T} \qquad (22.7)$$

for T as the time to maturity.

For taking in to account the inflation, we can use the annualised short-term interest rate r_t. Then, by using a linear regression model on the short-term interest rate we can estimate:

$$i_t=a^I+b^Ir_t+\sigma^I\varepsilon_t^I \qquad (22.8)$$

where $\varepsilon_t^I\sim N(0,1),\varepsilon_1^I,\varepsilon_2^I,\ldots\ i.i.d.$, and a^I,b^I,σ^I :are parameters to be estimated by using historical data. The index I indicates general inflation. Different 'lines of business' are affected differently due to macro economic decision-making. This affects the costs of claims for specific lines of business (e.g. legislative and court decisions, product liability). In other words, inflation is affected due to exogenous intervention. Therefore, in order to factor change in loss frequency (i.e. δ_t^F or the ratio of number of losses divided by number of written exposure units), the change in loss severity (i.e. δ_t^X) and a combination of both of these, δ_t^p, we take:

$$\delta_t^F=\max\left(a^F+b^Fi_t+\sigma^F\varepsilon_t^F,-1\right) \qquad (22.9)$$

$$\delta_t^X=\max\left(a^X+b^Xi_t+\sigma^X\varepsilon_t^X,-1\right) \qquad (22.10)$$

$$\delta_t^P=\left(1+\delta_t^F\right)\left(1+\delta_t^X\right)-1 \qquad (22.11)$$

where $\varepsilon_t^F\sim N(0,1),\varepsilon_1^F,\varepsilon_2^F,\ldots\ i.i.d.$,
$\varepsilon_t^X\sim N(0,1),\varepsilon_1^X,\varepsilon_2^X,\ldots\ i.i.d.$, $\varepsilon_{t_1}^F,\varepsilon_{t_2}^X$ are independent $\forall t_1,t_2$,
$a^F,b^F,\sigma^F,a^X,b^X,\sigma^X$ are parameters to be estimated based on historical data. The variable δ_t^P in Equation (22.11) represents changes in loss trends triggered by changes in inflation rates. This is applied to premium rates by constructing (22.11) so as to ensure correlation of aggregate loss amounts and premium levels attributed to dynamics of inflation. Note that it is

necessary to impose restriction by setting δ_t^F and δ_t^X to at least –1 so that the number of losses and loss severities have negative values. Since loss frequency changes due to general inflation, when inflation is high policyholders report more claims in certain lines of business. Hence, the corresponding cumulative changes (in $\delta_t^{F,c}$ and $\delta_t^{X,c}$) need to be calculated by:

$$\delta_t^{F,c} = \prod_{s=t_0+1}^{t} \left(1+\delta_s^F\right) \tag{22.12}$$

$$\delta_t^{X,c} = \prod_{s=t_0+1}^{t} \left(1+\delta_s^X\right) \tag{22.13}$$

where $t_0+1 =$ first year to be modelled.

22.3.2 Stock Returns

For stocks, we consider either stock prices or stock returns. We apply the Capital Asset Pricing Model (CAPM) by considering the return on a portfolio. Assuming a significant correlation between stock and bond prices, and taking into account multi-periodicity of a DFA model, we can have the following linear model for the 'stock market return'. The return will be for the projection year *t*, conditional on the one-year spot rate $R_{t,1}$ at time *t*. That is:

$$\mathbb{E}\left[r_t^M \mid R_{t,1}\right] = a^M + b^M\left(e^{R_{t,1}}-1\right) \tag{22.14}$$

where $e^{R_{t,1}}-1 =$ risk-free return, a^M, b^M are parameters to be estimated based on historical data and economic reasoning. Note that r_t^M is not the instantaneous short-term interest rate r_t (as in CIR). Moreover, a negative value of b^M means that increasing interest rates signal falling of expected stock prices. When such considerations are included, we can apply the CAPM to get the conditional expected return on an arbitrary stock S as:

$$\mathbb{E}\left[r_t^S \mid R_{t,1}\right] = \left(e^{R_{t,1}}-1\right) + \beta_t^s\left(\mathbb{E}\left[r_t^M \mid R_{t,1}\right] - \left(e^{R_{t,1}}-1\right)\right) \tag{22.15}$$

where $e^{R_{t,1}}-1$ is the risk-free return, r_t^M = return on the market portfolio and $\beta_t^s = \beta -$ coefficient of stock $S = \dfrac{Cov\left(r_{t,}^S r_t^M\right)}{Var\left(r_t^M\right)}$. Recall that a geometric Brownian motion for the stock price can be projected as a log-normal distribution for $1+r_t^S$. That is:

$$1+r_t^S \sim \text{log-normal}\left(\mu_{t,}\sigma^2\right), r_1^S, r_2^S, \ldots \text{ independent} \tag{22.16}$$

We have to choose μ_t to yield $m_t = e^{\mu_t + \sigma^2/2}$ where $m_t = 1 + \mathbb{E}\left[r_t^s \mid R_{t,1}\right]$ and $\sigma^2 =$ estimated variance of logarithmic historical stock returns.

22.3.3 Non-catastrophe and Catastrophe Losses

In non-catastrophe cases, the loss amounts depend on the age of insurance contracts. The aging phenomenon describes the fact that the loss ratio [i.e. the ratio of (estimated) total loss divided by earned premiums] decreases with increase in the age of the policy. For this reason, the insurance business is divided into three classes viz. (i) new business (superscript 0), (ii) renewal business 1 (i.e. first renewal with superscript 1) and (iii) subsequent renewals (i.e. second and subsequent renewals with superscript 2). Two main stochastic factors affect the total claim amount viz. number of losses and severity of losses. As before, the choice of a specific claim number and claim size distribution depends on the line of business. For non-catastrophe losses, we use negative binomial distribution for 'claim number' and a gamma distribution for 'claim size'. For period *t*, we take the loss numbers as N_t^j, mean loss severities as $X_t^j = \frac{1}{N_t^j}\sum_{i=1}^{N_t^j} X_t^j(i)$ and renewal category as *j* with corresponding mean values $\mu^{F,j}, \mu^{X,j}$ and standard deviations $\sigma^{F,j}, \sigma^{X,j}$. These need to be estimated based on historical data. We account for inflation by estimating the loss frequencies instead of relying on estimates of loss numbers (since the former is more stable than the latter). We consider negative binomial distribution with mean $m_t^{N,j}$ and variance $v_t^{N,j}$ (where the variables *m* and *v* represent mean and variance of different factors for the distribution of N_t^j). These factors are referred to by attaching a superscript (*N,X,Y,...*) to *m* or *v* as follows:

$$N_t^j \sim NB(a, p), j = 0,1,2 \quad (22.17)$$

$N_1^j, N_2^j, \ldots$ are independent. a and p are chosen so as to yield:

$$m_t^{N,j} = \mathbb{E}\left[N_t^j\right] = \frac{a(1-p)}{p} \quad (22.18)$$

Further, note that $v_i^{N,j} = \text{var}\left(N_i^j\right) = \frac{a(1-p)}{p^2}$, $m_t^{N,j} = w_t^j \mu^{F,j} \delta_t^{F,c}$, $v_t^{N,j}, = \left(w_t^j \sigma^{F,j} \delta_t^{F,c}\right)^2$, w_t^j = written exposure units, $\mu^{F,j}$ = estimated frequency, $\sigma^{F,j}$ = estimated standard deviation of frequency and $\delta_t^{F,c}$ = cumulative change in loss frequency. We take up claim size distribution for high frequency and low severity losses. Therefore:

$$X_t^j \sim Gamma(\alpha, \theta), j = 0,1,2, \quad (22.19)$$

$X_{1,}^{j} X_{2,...}^{j}$ are independent with α and θ chosen to yield $m_t^{X,j} = E\left[X_t^j\right] = \alpha\theta$ and $v_t^{X,j} = Var\left(X_t^j\right) = \alpha\theta^2$, $m_t^{X,j} = \mu^{X,j}\delta_t^{X,c}$, $v_t^{X,j} = \left(\sigma^{X,j}\delta_t^{X,c}\right)^2 / \delta_t^{c}$, and, $\delta_t^{X,c}$ = cumulative change in loss severity. When the number of losses is multiplied with the mean severity, the total (non-catastrophic) loss amount in respect of a certain line of business $\sum_{j=0}^{2} N_t^j X_t^j$ can be obtained.

Catastrophes: Let us now consider losses triggered by catastrophic events like windstorm, flood, hurricane, earthquake, etc. Such events are modelled through negative binomial or Poisson, or binomial distributions with mean m^M and variance v^M. We assume that there are no trends in the number of catastrophes i.e. $M_{1,} M_{2,}$ ...*i.i.d.* with m^M = estimated number of catastrophes and v^M = estimated variance (both based on historical data). The total (economic) loss (i.e. not only the part the insurance company has to pay) for each catastrophic event $i \in \{1,..., M_t\}$ can be simulated by taking the GPD (generalised Pareto distribution $G_{\xi,\beta}$). We have seen before that the GPD is the limit distribution of scaled excesses over high thresholds. Therefore, Y_t^i describes the total economic loss caused by catastrophic event $i \in \{1,..., M_t\}$ in the projection period *t*. Hence, if $Y_{t,i}$ ~ *lognormal, Pareto, GPD,* ...(with mean m_t^Y, variance v_t^Y), and $Y_{t,1,} Y_{t,2,}, ...$ *i.i.d.*, $\forall\left(t_1, i_1\right) \neq \left(t_2, i_2\right)$ [with $m_t^Y = \mu^Y \delta_t^{X,c}$, $v_t^Y = \left(\sigma^Y \delta_t^{X,c}\right)^2$, μ^Y = estimated loss severity, σ^Y = estimated standard deviation, and $\delta_t^{X,c}$ = cumulative change in loss severity] then Y_t^i can be generated and 'split up' to reflect the loss portions of different lines of business as:

$$Y_{t,i}^k = a_{t,i}^k Y_{t,i,} \qquad (22.21)$$

where *k* = line of business and *l* = total number of lines considered.

22.3.4 Underwriting Cycles and Payment Patterns

Some of the important factors behind underwriting cycles are: (i) time lag effect of the pricing procedure, (ii) trends, cycles and short-term variations of claims and (iii) fluctuations in interest rate and market values of assets. Short-term interest rates are the main factors affecting all other variables in the DFA. For this, we have to look at the premium cycles by including competitive strategies. To see this, we can take a homogeneous Markov chain model (in discrete time) and assign one of the following states to each line of business for each projection year: (i) weak competition, (ii) average competition and (iii) strong competition. Let state 1 be weak competition. Let the insurance company be assumed to be working under average competition. Let us further assume it demands high premiums since it knows that its market share cannot be increased. Let state 3 be strong competition where the insurance company accepts low premiums (so as to keep up its current market share). If we assume a stable claim environment, high premiums are equivalent to high profit margin over pure premium and low premiums are equal to low profit margin. When change occurs from one state to another, premiums undergo commensurate changes.

Consider the transition probabilities $p_{ij,} i, j \in \{1,\ 2,\ 3\}$, which denote the probability of changing from state i to state j (from one year to the next) which are assumed to be equal for each projection year. This means, the Markov chain is homogeneous. p_{ij}'s form a matrix T like:

$$T = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

The set of transition probabilities $p_{ij}, i, j \in \{1, 2, 3\}$ above depicts more than one possibility. However, it is possible to model the p_{ij}'s so as to depend on current market conditions applicable to each line of business separately. If the company writes l lines of business, this will imply 3^l states. Since the business cycles of different lines of business are strongly correlated, only few of the 3^l states are attainable. As a result, we have to model $L < 3^l$ states, where the transition probabilities $p_{ij,} i, j \in \{1, ..., L\}$ remain constant over time. It is not unreasonable to assume that some of these are zero. This is because there may exist some states that cannot be attained directly from certain other states. Thus, taking the attainable states L, the matrix T can be assigned dimension $L \times L$ as:

$$T = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1L} \\ p_{21} & p_{22} & \cdots & p_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ p_{L1} & p_{L2} & \cdots & p_{LL} \end{pmatrix}.$$

To fix the transition probabilities p_{ij} in any of the above mentioned cases, each state i has to be treated separately and probabilities have to be assigned to the variables $p_{i1}, ..., p_{iL}$ such that $\sum_{j=1}^{L} p_{ij} = 1\ \forall_i$. Then, the stationary probability distribution π has to be so considered that the chosen probability distribution will converge. It is extremely difficult to choose such an estimate of transition probability. Therefore, one has to rely on both historical data and experience based knowledge.

Payment Patterns: In this, the model has to deal with uncertainties of the claim settlement process i.e. to pay at random times. It is possible to see that a whole loss portfolio may belong to a specific line of business. Therefore, we need to consider its aggregate yearly loss payments in different calendar years (or development periods). Then the incremental payment of aggregate losses stemming from the same accident year forms a 'payment pattern'. To form a broad idea on the technique related to these payments, we have to see the portion dealing with the 'loss triangle'. If we consider the yearly loss payments pertaining to a specific accident year t, then the i^{th} development year refers to the calendar year $t+i$. Let us denote accident years by t_1 and development years by t_2. We need to distinguish two cases i.e. the outstanding loss payments pertaining to previous accident years and the loss payments in respect of future accident years. For this purpose, we have to use a 'chain-ladder procedure'. Since a lognormal distribution usually provides a good fit to historical loss factors, we can use it for estimating the outstanding loss payment. Then, for each accident year t_1, we can estimate the ultimate claim amount in each development year t_2 as:

$$\hat{Z}^{ult}_{t_1,t_2} = \prod_{t=t_2+1}^{\tau}\left(1+e^{\mu_t}\right)\sum_{t=0}^{t_2} Z_{t_1,t}, \qquad (22.21)$$

where, μ_t = estimated logarithmic loss development factor for development year t, and $Z_{t_1,t}$ = simulated losses for accident year t_1, to be paid in development year t.

22.4 CORPORATE MODEL

DFAs are meant to facilitate management decisions. Let us first consider one of the decisions viz. maximisation of shareholder's value. This means, we are assuming that the shareholders of the company rely on financial reports in making decisions while valuing the company. Consider the economic surplus U_t, computed as the difference between the 'market value of assets' and the 'market value of liabilities'. This can be derived by discounting loss reserves and the unearned premium reserves of the company. The amount of 'available surplus' reflects the financial strength of an insurance company. This, in turn, serves as a measure of shareholder value. To see the change in the cash flows, we calculate the equation:

$$\Delta U_t = P_t + \left(I_t - I_{t-1}\right) + \left(C_t - C_{t-1}\right) - Z_t - E_t - \left(R_t - R_{t-1}\right) - T_t, \qquad (22.22)$$

where, P_t = earned premiums, I_t = market value of assets (including realised capital gains in year t), C_t = equity capital, Z_t = losses paid in calendar year t, E_t = expenses, R_t = (discounted) loss reserves and T_t = taxes. A company becomes insolvent for $U < 0$. Therefore, in Equation (22.22), the term ($C_t - C_{t-1}$) gives the net additions to capital from the issuance of a new equity capital. To derive the earned premium from written premiums, we take variables such as: (i) line of business, (ii) written premiums P_t^j for renewal class j, (iii) change in loss trends, (iv) the position of the underwriting cycle and (v) the number of written exposures. Putting together all these in a functional form, the written premium $\tilde{P}_t^j$ can be derived as:

$$\tilde{P}_t^j = \left(1+\delta_t^P\right)\left(1+c_{m_{t-1},m_t}\right)\frac{w_t^j}{w_{t-1}^j}\tilde{P}_{t-1}^j,\ j = 0,1,2, \qquad (22.23)$$

where δ_t^P = change in loss trends, m_t = market condition in year *t,* $c_{A,B}$ = constant that describes how premiums develop when changing from market condition A to B ($c_{A,B}$ needs to be estimated from historical data), w_t^0 = written exposure units for new business, w_t^1 = written exposure units for renewal business (first renewal), and w_t^2 = written exposure units for renewal business (for second and subsequent renewals). Studies have revealed that the written premiums given by Equation (22.23) would come close to be adequate if the realisations of all random variables referring to projection year *t* $\left(\delta_t^P, c_{m_{t-1},m_t}, w_t^j\right)$ were known in advance. This is with the assumption of the adequacy of current premiums $\tilde{P}_{t_0}^j$. Premiums to be charged in year *t* needs to be determined prior to the beginning of year *t*. For this, random variables in Equation (22.23) needs to be replaced by estimates in order to factor written premiums P_t^j to be charged in the year *t*. This means:

$$P_t^j = \left(1+\hat{\delta}_t^P\right)\left(1+\hat{c}_{m_{t-1},m_t}\right)\frac{\widehat{w}_t^{\,j}}{w_{t-1}^j}\tilde{P}_{t-1}^j,\ j=0,1,2, \qquad (22.24)$$

where the estimates are to be obtained as expected values. That is:

$$\hat{\delta}_t^P = \left[1+a^X+b^X\left(a^I+b^I\left(ab+(1-a)r_{t-1}\right)\right)\right]\left[1+a^F+b^F\left(a^I+b^I\left(ab+(1-a)r_{t-1}\right)\right)\right]-1.$$

and $$\hat{c}_{mt-1,mt} = \sum_{m=1}^{l(k)} p_{mt-1,m}\, c_{mt-1,m}.$$

Here, $l(k)$ = number of states for line of business *k* and $p_{m_{t-1},m}$ = transition probability. We get $\hat{w}_t$, by estimating the equation $w_t^j = \left(a^j + b^j w_{t-1}^j + \varepsilon_t^j\right)^+,\ j=0,1,2,$ where $\varepsilon_t^j \sim N\left(0,\left(\sigma^j\right)^2\right)$, $\varepsilon_1^j, \varepsilon_2^j, \ldots$ i.i.d., and a^j, b^j, σ^j are parameters to be estimated based on historical data. Thus, $\hat{w}_t^j = a^j + b^j w_{t-1}^j$ in Equation (22.24) gives the expected value of random variable representing the actual written premiums. The time index $t = t_0$ refers to the year prior to the first projection year. By combining (22.23) and (22.24) the initial values $\tilde{P}_{t_0}^j$ can be calculated by using $P_{t_0}^j$ as:

$$\tilde{P}_{t_0}^j = \frac{1+\delta_{t_0}^P}{1+\hat{\delta}_{t_0}^P}\,\frac{1+c_{m_{t_0-1},m_{t_0}}}{1+\hat{c}_{m_{t_0-1},m_{t0}}}\,\frac{w_{t_0}^j}{\hat{w}_{t_0}^j}P_{t_0}^j,\ j=0,1,2, \qquad (22.25)$$

where $P_{t_0}^j$ are the 'written premiums' charged for the last year and still valid just before the start of the first projection year. Taking the written premiums $P_t^j(k)$ from (22.24), the total earned premiums of all lines and renewal classes will be:

$$P_t = \sum_{k=1}^{l}\sum_{j=0}^{2} a_t^j(k)\ P_t^j(k) + \left(1-a_{t-1}^j(k)\right)P_{t-1}^j(k), \qquad (22.26)$$

where $a_t^j(k)$ is the percentage of premiums earned in the year written (estimated on the basis of historical data).

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) List the risks that a DFA seeks to address in its scope.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

2) Why is it particularly crucial to evaluate 'interest rate risk' in DFA? What are its other inter-related risks that gets intertwined in the process?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

3) Mention the four sources of randomness considered by the DFA.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) What is meant by 'stochastic payment patterns'? How are they useful?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

5) State the general features of interest rates identified.

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

6) Mention the factors which affect the 'underwriting cycles'.

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

22.5 LET US SUM UP

In this unit, we have discussed DFA as a tool of financial management in insurance. It helps in decision-making when variables characterised by random occurrence are present as a result of their interrelationships. DFA scores over the deterministic models on this count. If a DFA model is built taking a time period of about 5 to 10 years, its results can capture the movements of the variables better. Being a part of financial management, DFA's framework is developed on an efficiency frontier by taking into account the two factors of risk and return. Its structure consists of stochastic scenario generator, data sources and output. The stochastically modelled variables are selected using the available data sources to evaluate the financial position of a firm, business line selection strategies and economic surplus generation potentials.

22.6 KEY WORDS

Exposure : This is a measure of vulnerability to loss, usually expressed in dollars or units.

Liability : Refers to a legally enforceable obligation. The term is most commonly used in a pecuniary sense.

Ornstein-Uhlenbeck Process : Is a stochastic process used as a theoretical model for Brownian motion.

Solvency : Refers to having sufficient assets-capital, surplus, reserves, etc. and being able to satisfy financial requirements like investments, transact insurance business and meet liabilities.

Underwriting : This is the process of selecting risks for insurance and classifying them according to their degrees of insurability so that appropriate rates could be assigned. The process includes

	rejection of those risks that do not qualify for insurance.
Efficient Frontier :	Refers to all possible asset combination which can be plotted in a risk-return space. The collection of all such possible portfolios defines a region in this space. The line along the upper edge of this region is known as the 'efficient frontier'.
Capital Asset Pricing Model (CAPM) :	CAPM is used theoretically to relate securities to the market as a whole. In a practical sense, it gives us the discount rate to be applied in cash flow situations so as to establish a fair value of an investment.

22.7 SUGGESTED BOOKS FOR FURTHER READING

1) Björk T. (1996). Interest Rate Theory, In *Financial Mathematics* (ed. W. Runggaldier), Lecture Notes in Mathematics 1656, 53-122, Springer, Berlin.

2) Cox J C, Ingersoll J E and Ross S A (1985). A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.

3) Embrechts P, Klüppelberg C and Mikosch T (1997). *Modelling Extreme Events for Insurance and Finance,* Springer, Berlin.

4) Kaufmann R, Andeas Gadmeraved, Ralf Kalf (2001). Introduction to Dynamic Financial Analysis, Asian Bulletin, Vol. 31, No.1, pp.213-249.

5) Lamberton D and Lapeyre B (1996). *Introduction to Stochastic Calculus Applied to Finance,* Chapman & Hall, London.

22.8 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) DFA helps in evaluating the economic factors behind price cycles. Such cycles are also usually accompanied by volatile insurance profits or increasing catastrophic losses. There are two primary techniques available to analyse the financial effects of different business strategies. These are: (i) scenario testing and (ii) stochastic simulation. The 'scenario testing' predicts the business results under 'deterministic

scenarios'. The latter deals with more dynamic situations. For this, it uses full probability distributions of important variables like surplus, written premiums, loss ratios, etc.

2) ALM or asset-liability management refers to life insurance liabilities that are deterministic. It assumes low variability. Such an assumption is not tenable in situations of 'volatile cash flow liability'. Such situations are mainly in non-life insurance where both the 'date of occurrence' and the 'size of claims' are uncertain. Such situations of non-life insurance liabilities and assets are dealt with by 'stochastic simulations' or the DFA.

3) DFA deals with 'strategic management decisions'. Such decisions relate to aspects like: asset allocation, capital allocation, performance measurement, market strategies, business mix, pricing decisions and product design.

4) Efficient frontier is the most common framework used in DFA for dealing with the financial portfolio of a company. In this, the DFA chooses sets of two measures viz. a 'return measure' (e.g. expected surplus) and a 'risk measure' (e.g. expected policyholder deficit). The two measures are then plotted on a graph. Such a graph yields an outer frontier referring to the maximum return for each risk. Such a frontier (or outer boundary) is what is referred to as the 'efficient frontier'. It serves as a tool to compare different strategies through the variables of risk and return.

5) It is criticised on two grounds: (i) it uses the systematic risks with that of unsystematic risks making the value perspective blurred and (ii) it gives misleading insight if used to address investment decisions after the concept of systematic risk has been plugged into the 'decision equation'.

6) It is useful in providing the main 'structure of a DFA model'. It does this by producing realisations of random variables representing the most important drivers of business results.

Check Your Progress 2

1) (i) risk of inadequate premiums, (ii) interest rate risk, (iii) risk of insufficient reserves, (iv) investment risk (associated with volatile investment returns and capital gains), (v) catastrophes and (vi) credit risks (related to reinsurer default, currency risk and exchange rate risks).

2) It is because of the correlation between interest rates and inflation. The inflationary influences result in changes in both 'claim size' and 'claim frequency'. Further, the correlation between interest rates and stock returns, affects investment returns. All these risks are therefore intertwined and needs considered.

3) Non-catastrophe losses, catastrophe losses, underwriting cycles and payment patterns.

4) Losses influence insurance firms in terms of their size, payment pattern, inflation, etc. They increase the uncertainties of the claims process. To allow for such risks, and balance them duly, 'stochastic payment patterns' are used as a means of estimating the loss reserves.

5) The general features of interest rates are: (i) volatility of yields at different maturities, (ii) they are mean-reverting, (iii) rates at different maturities are positively correlated, (iv) are bad for the economy if allowed to become negative and (v) the volatility of interest rates should be proportional to the level of the rate.

6) (i) time lag effect of the pricing procedure, (ii) trends, cycles and short-term variations of claims and (iii) fluctuations in interest rate and market values of assets.