5. Quantifying the liability area of risk

5.1 Introduction

This chapter is central to this study. The reason for this centrality is that a fairly new method of quantifying risk is introduced in the form of stochastic simulation[58] and the ideas behind this are then used in a variety of circumstances relevant to quantifying liabilities.

The areas of liability risk to be examined in this chapter are the frequency and severity of occurrences. The severity of occurrences (i.e. claims) is investigated to determine the delay in settlement of claims. The possible differences in parameter sets of the frequency and severity distributions for different classes of business or different homogeneous groups within the same class are also considered. Techniques whereby these parameter sets can be modelled for the different classes are discussed.

5.2 Stochastic simulation

Stochastic simulation is the process of generating a random outcome. Consider a stochastic variable X with a certain probability density function f(x) and cumulative distribution function F(x). Stochastic simulation involves generating a random outcome of X given either f(x) or F(x).

The following theorem proves to be very useful when considering F(x):

If X is continuous with CDF F(x), then U = F(x) ~ UNIF(0,1)[59].

This is extremely useful as an outcome can be randomly generated according to the underlying distribution. The following gives an explanation of the process.

1. Generate a random number between 0 and 1.
2. As F(x) is uniformly distributed (0,1) the randomly generated number can be transformed to a random X with distribution f(x).
3. By generating random numbers and calculating the corresponding values of X several times the distribution of f(x) can be determined.[60]

[58] Stochastic simulation has been mentioned in several publications. DAYKIN, C.D., PENTIKAINEN, T. & PESONEN, M. (1994), *Practical risk theory for actuaries,* is one example. The methodology is, however, fairly new as it relies heavily on extensive numerical techniques which have not been possible in the past due to lack of electronic computing power. Given the developments in the information technology industry this has now changed.
[59] That is to say the cumulative distribution function has a uniform distribution.
[60] This process is, of course, not necessary for a distribution function that can be expressed in closed mathematical form as f(x) can be obtained by differentiating F(x) with reference to x. The process is however very useful to determine f(x̲) where x̲ represents a vector of stochastic variables and F(x) cannot be expressed as a mathematical function.

It can be seen on the graph on the next page that the distribution of X will follow f(x) when F(x) is known. That is to say, very few randomly generated numbers between 0 and 1 will represent the tail values while very many randomly generated numbers will represent the values distributed around the mean.

The dashed line represents a randomly generated number between 0 and 1 which is then used to determine the corresponding x representing P[X<x][61].

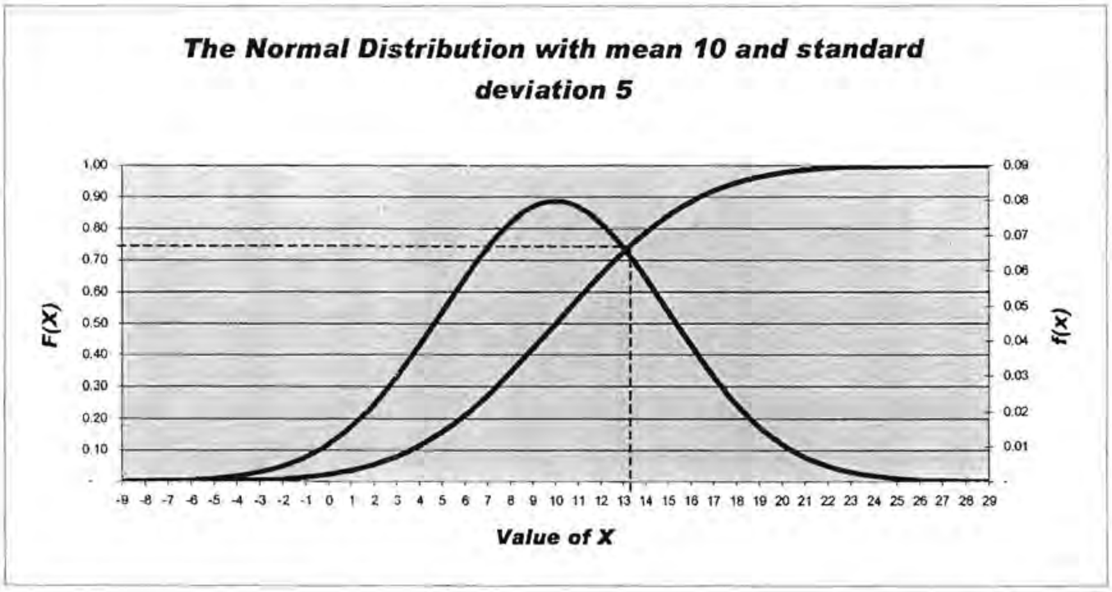


Figure 6

Note that as the normal distribution function is not invertible in mathematical form, this approach can only be presented graphically in the case of the normal distribution.

[61] P[X<x] is the randomly generated number.

In order to generate a normally distributed value with an algorithm a different approach would need to be followed[62].

Stochastic simulation can also be applied to discrete distributions. This will often

$$P[N=0]=e^{-\mu}$$

be the case with frequency distributions as they tend to have a Poisson distribution[63]. Assume that the number of claims N, generated by a certain insurance policy has a Poisson distribution with mean μ. Then

$$P[N=1]=e^{-\mu}\mu$$

$$P[N=2]=\frac{e^{-\mu}\mu^2}{2!}$$

etc.

By considering P[N < n] it is therefore possible to group the number of claims into bands of predetermined probability width. For example, all probabilities in the range [0;exp(-μ)] can be associated with a zero claim outcome.

5.3 Fitting distributions to claims information

As mentioned earlier several aspects regarding claims information have to be considered namely:

---

[62] In one of the workshops presented by the General Insurance Study Group the following code in Visual Basic was presented as an algorithm to generate a normally distributed value with mean 0 and variance 1. This is the work of Andrew Smith:

```
Function  normgen() as double
'Return sample from N(0,1) distribution
Static got_one as Boolean, stored as Double
Dim x1 as Double, x2 as Double, r as Double

If got_one then
        normgen() – stored
        got_one = false
Else 'Generate a point on the unit square
Do
        x1 = 2 * Rnd – 1
        x2 = 2 * Rnd – 1
        r  = x1^2 + x2^2
'reject unless inside unit circle
Loop While r >= 1 or r = 0
r = Sqr(-2 * log(r)/r)
stored = x1 * r
normgen() = x2 * r
got_one = true
End if
End Function
```

[63] The constraints regarding the Poisson distribution were mentioned in the previous chapter.

1. The frequency distribution of claims.
2. The severity distribution of claims.
3. The payment pattern of claims.
4. The correlations between claims and other claims or other economic variables.

It is important first to consider the different methods of curve fitting. To fully understand the process the practitioner should have a sound understanding of loss distributions[64].

5.3.1 The method of moments

This method involves equating the sample moments M' to the moments of the distribution $E[X^j]$ used in order to solve for the parameter set of the distribution to be used. This method therefore aims to find a distribution with the same mean, variance and kurtosis (depending on the number of variables in the parameter set) as that of the sample.

$$M'_j = \frac{\sum_{i=1}^{n} X_i^j}{n} \tag{5.1}$$

denotes the j^{th} moment of the sample and provides estimators for the mean and variance:

[64] For an extensive discussion on loss distributions the reader is referred to HOGG R.V. & KLUGMAN S.A.(1976).

$$\hat{\mu} = \overline{X} \tag{5.2}$$

and

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{X_i^2}{n} - \overline{X}^2 = \sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{n} \tag{5.3}$$

which is closely related to the sample variance S^2 as it converges to the sample variance as n increases,

$$\hat{\sigma}^2 = \left[\frac{n-1}{n}\right] S^2.$$

It is important to note the underlying reasoning behind this method. Considerable emphasis is placed on the mean and variance of the sample. This has to be borne in mind if selecting this approach to curve fitting.

5.3.2 Method of maximum likelihood estimation

This method fits the distribution most likely to reproduce the sample from which the distribution is fitted. Let $L(\underline{\theta}) = f(x_1, \ldots, x_n; \underline{\theta})$ represent the joint density function of n random variables $X_1, \ldots, X_n$ evaluated at the set $x_1, \ldots, x_n$. $L(\underline{\theta})$ then gives the likelihood of that set of values occurring as a function of $\underline{\theta}$. Therefore, by maximising $L(\underline{\theta})$ it will be possible to find the vector set $\underline{\theta}$ most likely to reproduce the sample for a given distribution function.

Maximising $L(\underline{\theta})$ is achieved by considering either the likelihood or the log likelihood function of $L(\underline{\theta})$ and differentiating with reference to $\underline{\theta}$.

$$L(\underline{\theta}) = \prod_{i=1}^{n} f(x_i; \underline{\theta}) \tag{5.4}$$

and then solve

$$\frac{\partial}{\partial \theta_i} \ln L(\underline{\theta}) = 0 \tag{5.5}$$

representing the i^{th} element of the vector $\underline{\theta}$.

Maximum likelihood estimators for large samples have the following asymptotic properties:

1. The estimator exists and is unique;

2. the estimator is a consistent estimator of $\underline{\theta}$, which means that for large samples the difference between the estimator and $\underline{\theta}$ becomes negligible;
3. the estimator is asymptotically normal with asymptotic mean $\underline{\theta}$ and variance

$$1/nE\left[\frac{\partial}{\partial\theta}\ln f(X;\underline{\theta})\right]^2$$

4. the estimator is asymptotically efficient which means that the expected value of $\underline{\theta}$ is equal to the estimator for the specific element of $\underline{\theta}$.

These qualities have lead practitioners to believe that this is, invariably, the best method to use in estimating the underlying distribution. Care must be taken as this method will often produce a distribution with variance smaller than the sample variance due to the fact that the most likely distribution is found. Furthermore the mean found by the maximum likelihood estimation may be vastly different to the sample mean for the same reason.

5.3.3 Method of least squares

The aim of this method is to find the distribution for the dependent variable which minimises the sum of the squared errors between the actual dependent variables and the expected dependent variables. This can be done by considering the probability density function. That is to say, consider the actual number of claims in specific bands compared to the estimated number of claims in specific bands according to the probability density function form that is to be fitted. The parameters of the probability density function can then be adjusted to find the minimum sum of squared errors.

Therefore the following entity is minimised with reference to $\underline{\theta}$:

$$\text{Sum Squared Errors} = \sum_{band=1}^{n}\left(y_{band} - (\hat{y};\underline{\theta})_{band}\right)^2 \tag{5.6}$$

where the first term represents the actual value observed and the second term in the summation represents the expected value of the dependent variable given the parameter set $\underline{\theta}$.

The squared errors may differ in size if the expected number of claims is not the same in each band. To alleviate this problem the errors can be standardised by dividing the squared error by the expected value of the dependent variable[65].

[65] Note that if the number of expected claims within each band has a Poisson distribution then the mean and variance of this distribution is the same.

The standard error divided by the square root of the expected number of claims can therefore be approximated by a normal distribution. This would also be the case for other distributions where the mean and the variance can be approximated by the same value.

$$\text{Sum Standard Squared Errors} = \sum_{band=1}^{n} \frac{\left(y_{band} - (\hat{y};\underline{\theta})_{band}\right)^2}{(\hat{y};\underline{\theta})_{band}} \tag{5.7}$$

5.3.4 Some useful distributions

The aim of this section is to give the reader an insight into some useful distributions which can be applied in the modelling of liability outgo. For every distribution provided below the method of moments and method of maximum likelihood estimators are also given.

5.3.4.1 The Poisson distribution

This distribution is used in the modelling of claim frequencies.

$$f(X;\theta) = \frac{e^{-\theta}\theta^{x}}{x!} \tag{5.8}$$

where $\theta > 0$ and $x = 0,1,2\ldots$

The mean and variance of this distribution are both θ.

The method of moments estimator and the maximum likelihood estimator for θ is the same:

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

where N is the exposure of policies[66] and x_i is the number of claims incurred by the i^{th} policy.

In order to randomly generate values for the Poisson distribution, calculate the probabilities P[X<x]. Generate a random value between 0 and 1. Find the value

---

The sum of the squared errors divided by the expected number of claims therefore has a Chi-square distribution. This is so because a standardised normal variable squared has a Chi-square distribution with one degree of freedom and the sum of Chi-square distributions once again has a Chi-square distribution but the degrees of freedom will depend on the number of variables summed as well as the approximations made in determining each Chi-square distribution.

The least squares estimation method therefore develops into the Chi-square estimation method in this instance and for this specific class of distribution the fit with the smallest Chi-square value will therefore be found.

[66] N is not the number of policies as not all policies may be on risk for the same period of time which could skew the results. It is therefore important to calculate the exposure of each individual policy as explained in chapter 4.

of x for which the randomly generated value is the closest to, but smaller than P[X<x].

The example below should clarify this process:

Assume a claim frequency distribution has a Poisson distribution with mean of 5 claims per period.

P[X=0] =	0.006738	P[X<=0] =	0.006738
P[X=1] =	0.033690	P[X<=1] =	0.040428
P[X=2] =	0.084224	P[X<=2] =	0.124652
P[X=3] =	0.140374	P[X<=3] =	0.265026
P[X=4] =	0.175467	P[X<=4] =	0.440493
P[X=5] =	0.175467	P[X<=5] =	0.615961
P[X=6] =	0.146223	P[X<=6] =	0.762183
P[X=7] =	0.104445	P[X<=7] =	0.866628
P[X=8] =	0.065278	P[X<=8] =	0.931906
P[X=9] =	0.036266	P[X<=9] =	0.968172

If the randomly generated number were to lie between 0.615961 and 0.762183 this would equate to a realisation of 6 claims in the period.

5.3.4.2 The Normal (Gaussian) distribution

This distribution provides an even spread of outcomes and provides the distribution for an individual outcome as the sample size increases to infinity[67]. The probability density function is given by

$$f(X;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \tag{5.9}$$

and $\sigma^2 > 0$.

The mean of this distribution is μ and the variance is σ^2.

The method of moments estimators and the maximum likelihood estimators for μ and σ are approximately the same, namely

$$\hat{\mu}=\tfrac{1}{N}\sum_{i=1}^{N}x_i$$ and

$$\sigma^2=\frac{1}{n-1}\sum(x_i-\hat{\mu})^2 \quad .$$

[67] As per the central limit theorem.

In order to generate a value on the normal distribution the code provided in footnote 62 can be utilised or the distribution function of a spreadsheet can be used. The standard normal distribution can be adjusted to the relevant normal distribution by the linear scaling norm_gen()*σ + μ.

5.3.4.3 The Log-Normal distribution

This distribution is often used to fit claims severity or loss ratio severity. It is positively skewed and hence provides an appropriate starting point when investigating a claim distribution. The probability density function is given by

$$f(X;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2}\left[\frac{\ln(x)-\mu}{\sigma}\right]^2} \quad (5.10)$$

The mean of this distribution is $\exp(\mu + \sigma^2/2)$ and the variance is $\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$.

This distribution can be easily assessed as the natural logarithm of the outcomes has a normal distribution. Hence, to fit this distribution, calculate the natural logarithms of the actual observations and proceed as with the normal distribution described above.

5.3.4.4 The Exponential distribution

This distribution has a thicker tail than the Log-Normal distribution, is integratable in closed form and has only one parameter, hence it is an easy alternative to other distributions. It is also appropriate to model claim severity.

The distribution has the following form:

$$f(X;\theta)=\frac{1}{\theta}e^{\frac{-x}{\theta}} \quad (5.11)$$

The mean of this distribution is θ and the variance is θ^2.

The method of moments estimator and the method of maximum likelihood estimator is the same:

$$\hat{\theta}=\frac{1}{N}\sum_{i=1}^{N}x_i \quad (5.12)$$

$$F(X;\theta)=1-e^{\frac{-x}{\theta}} \quad (5.13)$$

where N is the number of claims.

In order to generate a value on this distribution consider the cumulative distribution function F(X) = P[x<X]. As the cumulative distribution function has a uniform(0,1) distribution a claim value can be generated.

$$x = -\theta \ln(1 - F(X;\theta))$$

5.3.4.5 The Pareto distribution

This distribution can be used to generate a very thick tail and as a result is very suitable for claims experience with wide variability such as liability claims. The Pareto distribution is more flexible than the exponential distribution as it has two parameters as well as being integratable in closed form. As a result it is a very popular distribution to apply in practice and is often found in pricing models.

The distribution has the following form:

$$f(X;\alpha,\lambda) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}} \qquad (5.14)$$

$X > 0$, $\alpha > 0$, $\lambda > 0$.

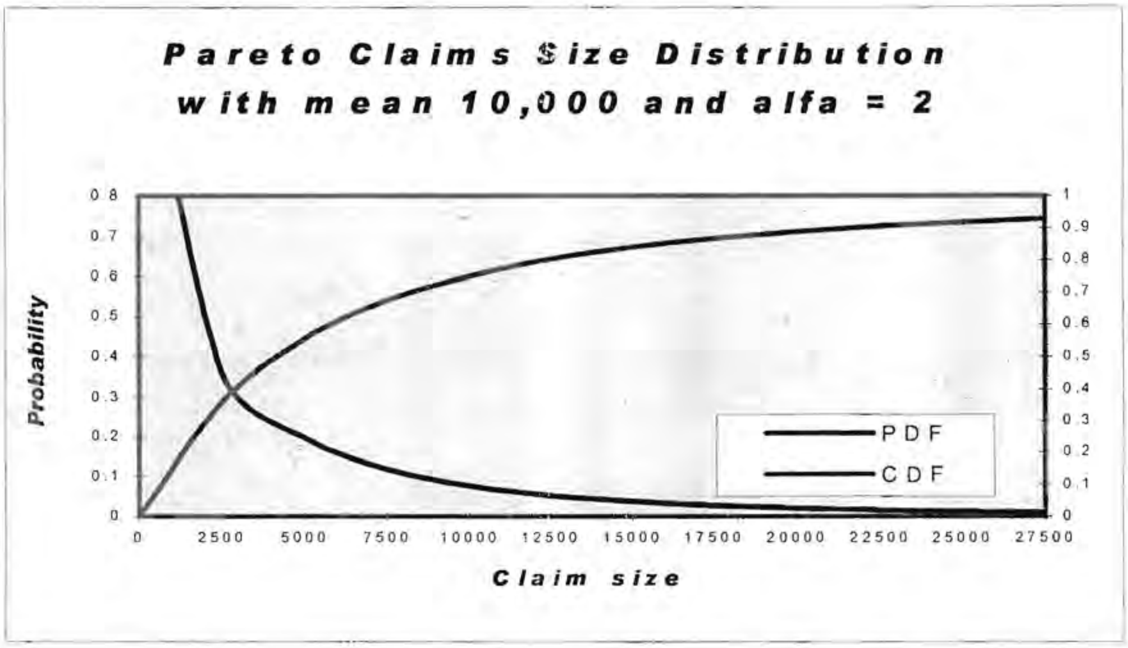
The distribution has a mean of

$$\frac{\lambda}{\alpha - 1}$$

and variance of

$$\frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} .$$

The reader should note from the above that the mean and variance may not be defined for all x. In such instances, the mean and variance over a range of x values can still be obtained.



For the Pareto distribution the method of moments estimators are:

$$\hat{\lambda} = (\hat{\alpha} - 1)\bar{x}$$

where s is the standard deviation of the sample and

$$\hat{\alpha} = \frac{2s^2}{s^2 - \bar{x}^2} \quad .$$

The estimators for the Pareto distribution found by means of maximum likelihood estimation are found by solving the equation $h(\lambda) = 0$,

$$h(\lambda) = \frac{\sum \frac{1}{\lambda + x_i}}{\sum \frac{x_i}{\lambda(\lambda + x_i)}} - \frac{n}{\sum \ln\left(1 + \frac{x_i}{\lambda}\right)} \tag{5.15}$$

The result is then used to solve the equation

$$\hat{\alpha} = \frac{\sum \frac{1}{\hat{\lambda} + x_i}}{\sum \frac{x_i}{\hat{\lambda}(\hat{\lambda} + x_i)}} \tag{5.16}$$

In order to generate a value on this distribution consider the cumulative distribution function again:

$$F(X;\alpha,\lambda)=1-\left(\frac{\lambda}{\lambda+x}\right)^{\alpha} \qquad (5.17)$$

A claim value can therefore be generated using the equation

$$x=\left(\frac{\lambda}{(1-F(X;\alpha,\lambda))^{\frac{1}{\alpha}}}\right)-\lambda \qquad (5.18)$$

5.3.4.6 The Gamma distribution

The Gamma distribution is a very flexible distribution. It is, however, not manipulable in closed form and as a result is best used with a computer package.

The distribution has the following form:

$$f(X;\alpha,\lambda)=\frac{e^{-\lambda x}}{\Gamma(\alpha)}\lambda(\lambda x)^{\alpha-1} \qquad (5.19)$$

$X > 0$, $\alpha > 0$.

It is also known as a Pearson's Type III distribution. If α = 1 the exponential distribution results. If α is a positive integer it is termed an Erlang distribution. If λ = 0.5 and α = 0.5ν then it is termed a chi-squared distribution with ν degrees of freedom.

The distribution has a mean of

$$\frac{\alpha}{\lambda}$$

and variance of

$$\frac{\alpha}{\lambda^2}.$$

The parameter set can therefore easily be determined by the method of moments. Maximum likelihood estimation requires numeric integration.

5.3.5 Testing the fit

Having decided on a distribution to fit to the data, it is subsequently necessary to test the fit. This is normally done first by a visual inspection of the data and subsequently by investigating the sum of standardised squared errors as explained above. This means a Chi-square test is conducted on the fit.

Judgement plays a very important role in testing the suitability of the fit. It has already been alluded to that a fit may not be appropriate even if the fit appears to be good. The reason for this is that the data representing claims may be considered to be a *less than perfect* representation of the actual underlying population. In these instances consultation with experts in the field needs to be considered. This is especially true regarding very large claims as past structures and practices may not be indicative of future experience.

At all times it is therefore necessary to keep the following in mind:

1. Past information is used as a starting point to find a possible trend in future experience.
2. This assumption is flawed in that it assumes that past, present and future structures regarding business operations, claims management, economic scenarios and a variety of other variables remain the same. This includes past trends which are assumed to exhibit themselves in similar fashion in future.
3. As the model builder may be aware of new trends these have to be allowed for as well. The model builder may also make use of experts such as underwriters to determine whether or not the distribution fit is appropriate[68].

Bearing in mind that the aim is to set a model for future experience, which is uncertain, too great a reliance on past data is also problematic as past experience might not necessarily be repeated in future. The model builder will then graduate the distribution function according to all the information at hand to set a best estimate of future expected experience.

An extension of this would be to build a model for each variable in the $\underline{\theta}$ variable set.

5.4 Generalised linear modelling

The crux of liability modelling lies in identifying and quantifying the impact of different risk factors on the parameters used in the model. As an example, the Poisson parameter θ, which is the number of occurrences in a period will depend on the risk characteristics of the pool of policies generating the claims. In estimating θ it is imperative that all appropriate risk factors are taken into account. This ensures that as accurate an assessment as possible of the underlying risk is obtained. This section will focus on the technique of

[68] A good example of such an approach would be to ask the underwriter his or her assessment of the number of claims for a certain type of business in the different bands for a specific size of business. This will enable the model builder to check whether these answers conform to $P[a < X < b]*N$ where a and b will be the lower and upper limit of each band and N will be the size of the business by policy number, for example.

generalised linear modelling in obtaining a function of the form $f(\theta;\underline{\omega})$ where $\underline{\omega}$ is the vector of risk factors having an influence on θ.

Consider a pool of risks such as policyholders with own damage coverage for their motor cars. These policyholders can be grouped into similar classes according to certain risk criteria. For example, the power to weight ratio of the car, the area in which the car is housed, the age of the driver, etc. Consider the frequency of claims arising from such a risk pool. Each one of these pools will have a maximum likelihood estimator, θ, of the frequency of own damage claims. This estimator will be derived by dividing the number of claims by a suitable exposure unit which in this case would be the duration of cover. Each such estimator θ_i, where i refers to the i'th group of policyholders, will depend on the risk characteristics of the i'th group and is assumed to be independent of the other θ_i's. These risk characteristics can be represented by either quantitative factors or qualitative factors. These factors are the explanatory variables used in generalised linear modelling. Quantitative factors tend to be rare. Qualitative risk factors can be presented using dummy variables[69].

The aim of generalised linear modelling is to express some function of θ_i, say $g(\theta_i)$[70] as a linear combination of a set of parameters $\beta_1,\ldots\beta_n$ where n < number of classes.

More precisely the following set of equations must be solved

$$g(\theta_i) = X_i^T \beta + \varepsilon_i \tag{5.20}$$

where ε_i denotes the error term.

This means that when the individual outcome of a claim for any one policy is considered and this outcome can be viewed as the single random variable Y. Y will be from a certain distribution with parameter θ where θ is a function of $\underline{\beta}$, which represents the risk characteristics for the specific policy.

In order to apply generalised linear modelling, the distribution of Y must belong to the family of exponential distributions and must depend on a single parameter. This means that it must be of the following form:

$$f(y;\theta) = \exp[yb(\theta) + c(\theta) + d(y)] \tag{5.21}$$

Given that the distribution of Y can be written in this form it is then possible to solve the coefficient vector $\underline{\beta}$ by maximum likelihood.

[69] For example, consider differentiation by gender. One factor will be required which will be either 0 or 1 depending on the gender.

[70] The function g, termed the link function, must be monotone and differentiable.

Generalised linear modelling is more powerful than ordinary regression as the error structure need not be normal.

In order to facilitate the solving of this problem, the function $g(\theta_i)$ is expressed as η_i. Note that $g(\theta_i)$ is a monotone differentiable function, referred to as the link function. The link function can be identity, $g(\theta_i) = \theta_i$, natural log function, $g(\theta_i) = \ln(\theta_i)$ or another function which satisfies the criteria of monotonicity and differentiability .

The log likelihood function of all the observations $\underline{y}$ can be presented by

$$l(y;\theta) = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i) \quad (5.22)$$

Solving for θ is the equivalent of solving for β. Consider

$$U_j = \frac{\partial l(y;\theta)}{\partial \beta_j} = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_j} \quad (5.23)$$

where $l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i)$ and N is the number of classes of observations

It can be shown that

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \quad (5.24)$$

where $\mu_i = E(Y_i) = -c'(\theta_i)/b'(\theta_i)$ and x_{ij} is the jth risk characteristic of policy group i. In general the equations $U_j = 0$ have to be solved by numerical iteration. If the Newton-Raphson method is used then the nth approximation is given by

$$\underline{b}^{(n)} = \underline{b}^{(n-1)} - \left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right]^{-1}_{\beta = b^{(n-1)}} \underline{U}^{(n-1)} \quad (5.25)$$

where the term in square brackets is the matrix of second derivatives of l evaluated at $\underline{\beta} = \underline{b}^{(n-1)}$ and $\underline{U}^{(n-1)}$ is the vector of first derivatives evaluated at $\underline{\beta} = \underline{b}^{(n-1)}$. It is therefore necessary to start with a randomly or otherwise selected set, $\underline{b}^{(1)}$.

Several commercial applications are available to facilitate the solving of these equations.

5.4.1 Special case of generalised linear modelling

Consider the special case where the individual distribution functions for the frequency of own damage claims for each group of homogeneous policyholders is normally distributed with constant variance but different means.

The likelihood function is given by:

$$l(y,\theta)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2\sigma^2}(y_i-\theta_i)^2} \tag{5.26}$$

The log-likelihood function is given by:

$$=\frac{-N}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i-\underline{x}_i^T\ \underline{\beta}\right)^2$$

$$=\frac{-N}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\left(\underline{y}-\underline{X}\underline{\beta}\right)^T\left(\underline{y}-\underline{X}\underline{\beta}\right)$$

where X is the matrix of explanatory variables. Solving for maximum likelihood estimates yields:

$$\frac{\partial l}{\partial\beta_k}=-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\frac{\partial}{\partial\beta_k}\left(y_i-x_i^T\ \beta\right)^2 \tag{5.27}$$

$$\frac{\partial l}{\partial\beta_k}=-\frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i-x_i^T\ \beta\right)x_{ik}$$

where k has a maximum value equal to the number of explanatory variables plus 1 for intercept.

$$\frac{\partial l}{\partial\underline{\beta}}=-\frac{1}{\sigma^2}X^T\left(\underline{y}-X\underline{\beta}\right)$$

solving this equation for a maximum yields:

$$\underline{\beta}=\left(X^TX\right)^{-1}X^T\underline{y} \tag{5.28}$$

and this is, of course, the standard form of least squares regression. Therefore regression is a special case of generalised linear modelling with the assumptions as set out above.

5.5 Delay modelling

The necessary tools required to model the frequency and severity of liabilities have been discussed above. The severity distribution will, however, refer to the present value of liability outgo. It is possible to extend this idea to incorporate the actual cash-flow of the liability outgo. This extension may be more appropriate since liability structures change. The manner in which claims will occur, be notified, administered and settled and possibly re-opened, will have an impact on both the frequency and severity of claims.

In order to model liability outgo accurately it is important to firstly identify the expected frequency and severity and then subsequently to model the outgo by investigating the actual expected emergence of claims over time as well as the actual payment of those claims. It might very well be that the run-off of claims reported later will be different to the run-off of claims reported at an early stage.

The investigation of the run-off of claims is as follows:

5.5.1 Step 1

Claims are grouped into homogeneous groupings or cohorts such as the period in which the claim arose. This grouping is done for the number of claims in that claim period by considering the delay before a specific claim was actually reported. Subsequent to this the payments made in reference to these claims are also grouped, firstly by the duration at which the claim was reported and secondly by the subsequent period of delay before the claim was settled.

It is not always possible to subdivide payments in such a fashion and therefore in certain instances the grouping is considered for the actual amounts paid by reference to a certain grouping period for the incidence of claims.

It is important to consider paid claims in order to assess actual cash flows. It is important to consider incurred claims in order to assess the incurred cash flow position[71]. When modelling liability outgo the incurred position will normally be reflected. It is important to distinguish between that part of the incurred amount which is held as reserve and that part which is actual payment. The level of reserves has a direct impact on the value of the company as reserves will

[71] Incurred claims are claims paid during the year together with any change in reserves set up for the claims (reserves at end of period less reserves at start of period). The reserving policy of a company can distort the apparent cash-flow position of the company. It is therefore important to consider company policy and identify whether this is a best estimate.

generally secure lower investment returns than other business operations. This is because a matched position in assets is normally held for liabilities. These matched positions are often in cash or fixed interest securities which tend to yield a lower return over long durations of time than investments in other business operations.

5.5.2 Step 2

The grouped information must be expressed in such a fashion as to facilitate projection of the experience. Run-off triangles are well known in actuarial literature. In essence the information will be grouped as follows:

Triangle of reported number of claims.

	Period of reported claim				
Period of loss	In same period	One period later	Two periods later	Three periods later	Four periods later
First period[72]	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	$R_{1,4}$	$R_{1,5}$
Second period	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$	$R_{2,4}$	
Third period	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$		
Fourth period	$R_{4,1}$	$R_{4,2}$			
Fifth period	$R_{5,1}$				

Table 3

The above table assumes five periods are available for investigation. A period can refer to a week, month, quarter year, half year, year etc. The aim is to determine the portion of claims expected to be reported at each duration in the future given the past experience. This can be done by investigating the development of the cumulative number of reported claims and then applying this development to the experience to date. $R_{i,j}$ refers to claims arising in period I and reported in period j.

[72] This refers to the period for which the run-off of losses are investigated. The first period may be the first month or the first quarter or the first year of the period of investigation.

Triangle of cumulative reported claims

	Period of reported claim				
Period of loss	In same period	One period later	Two periods later	Three periods later	Four periods later
First period	$R_{1,1}$	$R_{1,1}+R_{1,2}$	$R_{1,1}+R_{1,2}+R_{1,3}$	$R_{1,1}+R_{1,2}+R_{1,3}+R_{1,4}$	$R_{1,1}+R_{1,2}+R_{1,3}+R_{1,4}+R_{1,5}$
Second period	$R_{2,1}$	$R_{2,1}+R_{2,2}$	$R_{2,1}+R_{2,2}+R_{2,3}$	$R_{2,1}+R_{2,2}+R_{2,3}+R_{2,4}$	
Third period	$R_{3,1}$	$R_{3,1}+R_{3,2}$	$R_{3,1}+R_{3,2}+R_{3,3}$		
Fourth period	$R_{4,1}$	$R_{4,1}+R_{4,2}$			
Fifth period	$R_{5,1}$				

Table 4

The weighted average of the development of the cumulative reported claims at each duration is used to assess a trend. This is an estimate based on the weighted average only. Quantifying the variability within these estimates will be considered at a later stage in this study.

For example, the weighted average of development from the first period after reporting to the second period after reporting will be given by dividing the sum of all the yellow cells in Table 4 by the sum of the blue cells[73]. Such a factor is termed a development factor.

In addition to considering the trend in reported claims, the trend in claim payments needs to be considered if additional delays exist between the time of reporting and the time of actual payment. As mentioned earlier this analysis can also be split for the actual duration of reporting. An illustration is provided assuming the duration at which claim payments are made is different to the duration at which they were reported. Using the notation as in Table 3 the grouping will be as follows:

[73] For an in depth study on triangulation techniques the reader is referred to the core reading published by the Institute and Faculty of Actuaries.

Triangle of claim payments.

	Period of claim payment				
Period of reporting	In same period	One period later	Two periods later	Three periods later	Four periods later
First period ($R_{1,1}$)	$P_{1,1}$	$P_{1,2}$	$P_{1,3}$	$P_{1,4}$	$P_{1,5}$
Second period ($R_{2,1}+R_{1,2}$)	$P_{2,1}$	$P_{2,2}$	$P_{2,3}$	$P_{2,4}$	
Third period ($R_{3,1}$ + $R_{2,2}$ + $R_{1,3}$)	$P_{3,1}$	$P_{3,2}$	$P_{3,3}$		
Fourth period (etc.)	$P_{4,1}$	$P_{4,2}$			
Fifth period (etc.)	$P_{5,1}$				

Table 5

$P_{i,j}$ is the amount reported in financial period i and paid in the subsequent j^{th} period. The grouping is by reporting period and therefore the information can be extended to include payments made for claims having the same period of reporting but an earlier period of loss if it can be assumed that the trend in development has remained stable. If this cannot be assumed such a triagle would be set up for each period in which claims arose.

5.5.3 Step 3

For the two types of development (for reporting and for actual payments) development factors are then determined as described above. These factors represent the expected growth in the cumulative claims reported or paid from one period to the next as the case might be[74]. By considering the cumulative effect of these factors from a specific period to the expected ultimate period of development is very useful.

[74] Reported claims and paid claims are not the only entities which can be investigated in this fashion. Claims incurred are often investigated when substantial reserves are required at initial durations and the actual payment patterns therefore provide poor starting points. As the results will be very sensitive to the initial amounts for later years, this feature is often referred to as a corner sensitivity.

Example: Development factors of claim payments.

	Period of claim payment				
	In same period	One period later	Two periods later	Three periods later	Four periods later
Development factor		$\Sigma C_{2,i} / \Sigma C_{1,i}$ (i = 1 to 4)	$\Sigma C_{3,i} / \Sigma C_{2,i}$ (i = 1 to 3)	$\Sigma C_{4,i} / \Sigma C_{3,i}$ (i = 1 to 2)	$\Sigma C_{5,i} / \Sigma C_{4,i}$ (i = 1)
Denoted by		D_1	D_2	D_3	D_4

Table 6

Where $C_{i,j} = \Sigma_{(t=1 \text{ to } j)} PC_{i,t}$ is the cumulative claims paid to date grouped by period of loss. $PC_{i,t}$ is the amount of claims paid in development period *t* for period of loss *i*. It therefore allows for both the delay in reporting and the subsequent delay in payment.

The development factors denoted by D_i above yields the weighted average of the individual development ratios. The cumulative development factors are then denoted by:

$CD_1 = D_1 * D_2 * D_3 * D_4$

$CD_2 = D_2 * D_3 * D_4$

$CD_3 = D_3 * D_4$

$CD_4 = D_4$

Given that

cumulative amount to date * the appropriate cumulative development factor = expected ultimate amount (or number of reported claims),

it is possible to determine the proportion of expected ultimate claims to be paid at each future development period.

It is important to apply consistent approaches between the modelling of delays and the modelling of claim frequency and severity. The delay pattern identified for the number of claims reported can be applied to spread the claim frequency. Here it is important that a reported claim will therefore be a claim giving rise to a claim, zero claims cannot be used. For the severity of claims the influence of inflation needs to be considered. The severity will be expressed as an inflation adjusted amount to reflect the present value of expected claims. If, however, these payments are subject to further delays, allowance for further expected increases as a result of inflation will be required.

5.6 Bootstrapping

Bootstrapping is a technique involving numerical techniques in the calculation of the variance of a certain set. These techniques are computationally very intensive and require the use of computers. For a detailed discussion the reader is referred to DAVISON and HINKLEY (1997).

A simple example:
Take a set of development ratios[75] for a certain class of business:
Assume these development factors are grouped by year of accident and subsequent development for each accident year.

	Year of development				
Year of accident	$D_{1,0}$	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$	$D_{1,4}$
	$D_{2,0}$	$D_{2,1}$	$D_{2,2}$	$D_{2,3}$	$B_{2,4}$
	$D_{3,0}$	$D_{3,1}$	$D_{3,2}$	$B_{3,3}$	$B_{3,4}$
	$D_{4,0}$	$D_{4,1}$	$B_{4,2}$	$B_{4,3}$	$B_{4,4}$
	$D_{5,0}$	$B_{5,1}$	$B_{5,2}$	$B_{5,3}$	$B_{5,4}$
	?	?	?	?	?

Table 7

$D_{1,0}$ for example is the development of cumulative claims paid in the first accident year from the end of the first period to the end of the second period under investigation.

Past practice has been to investigate such a triangle and to determine development ratios that represent the expected development of cumulative claims between different durations[76]. A development ratio can, however, be viewed as a sample from a certain distribution for the specific duration. Therefore the process of triangulation can be seen as an estimation of the most likely development ratio from a certain unknown distribution of development ratios.

Bootstrapping can be used to gauge the variability of the reserve calculated by means of triangulation[77]. This is done by assuming a certain distribution for the development ratio at each duration. The distribution chosen will depend on the

[75] Development ratios are often used to determine the development of cumulative claims experience based on past data. These ratios are based on the development of cumulative claims from one period to the next for a certain year. The definition of claims by year should be consistent throughout to ensure proper projection.

[76] The last row of Table 7 denotes the development ratios that will be used by the practitioner. In certain instances different development ratios may be used for different years if the practitioner is of the opinion that not all years will exhibit the same development of cumulative claims in future.

[77] The practitioner will need to determine whether or not investment income has to be taken into account.

judgement of the practitioner. A good starting point for a stable[78] book of business will be to assume that the development ratio at each duration has a discrete distribution whereby the development ratio for a certain duration can take any of the ratios shown in Table 7 at the same duration, with equal probability. This assumption can only be made for a book of business that has had a fairly constant size over the period of investigation as a smaller book of business will be subject to a higher level of variability[79].

The calculation process then entails sampling from the triangle to obtain the projected development ratios[80]. This is done by generating a random number and selecting one of the given development ratios for the period of development based on the number generated. A similar process would be followed if a different distribution had been chosen. The selected development ratios are then used to calculate the expected ultimate experience for each year. From this projection all cumulative claims paid to date can be subtracted. The result of this calculation then gives one sample of the approximated distribution of all outstanding experience. Simulating a selected number of these outstanding experiences will give a probability density function of the outstanding experience whereby the variance of future outstanding experience can be approximated.

For a sufficient number of periods of development the result of a bootstrapping exercise can yield a very smooth distribution function. This distribution can be utilised to generate very powerful information. Not only will the practitioner be able to gauge the variability in future experience but the extreme values of the distribution can also be noted.

This method makes allowance for underwriting risk and reserve risk only. Additional allowance will be required for asset risk, credit risk and growth risk through the modelling of all these risks together.

The approach explained above is very simplistic. It is also possible to assume a standard run-off model and then to investigate the error terms generated through a comparison of the model and the actual past experience. These errors can be standardised and used in a distribution as a sampling basis to project future development.

---

[78] The stability of a book of business will depend on the underlying variability of that business. General rules of thumb may be applied (such as a minimum number of 3000 claims per year of accident). Such rules are crude approximations based on a normality assumption. Data sets are, however, often limited. In such instances it will have to be accepted that the results of a bootstrapping exercise will be crude.

[79] Using development ratios based on smaller books of business in the past (for a growing company) will result in the variability of total experience being overestimated. This may be seen as a prudent approach but should be handled with caution.

[80] These are the $B_{x,y}$ values shown in Table 7.

BONNARD R. GREENWOOD M. GREYBE S. (1999) provides a practical illustration of a variety of these techniques appropriate for modelling future run-off of claim payments.

5.7 Model building

The process of stochastic simulation and distribution fitting has been considered. The area of model building is now discussed. In his presentation to the Actuarial Society of South Africa's General Insurance Mini Convention (1999), KARL MURPHY of ENGLISH MATTHEW AND BROCKMAN stated that model building is not only a science but also an art. This author seconds this view.

To put this in perspective, the objective of building a liability model must be borne in mind: The intention is to develop a model that gives as accurate a representation of expected future experience as possible. In order to do this past experience is investigated and assumptions are made as to which characteristics future experience might exhibit. This process relies on a considerable element of judgement or "gut feeling". It is here that the 'art of the science' comes into play.

The following provides a simple approach to model building:

5.7.1 Step 1

Identify the class of business for which the model is to be built. As the risk details differ by class of business it is important to develop different models for different homogeneous groups of business. A balance needs to be struck between the grouping of business and the volume of business as too many subdivisions will lead to spurious accuracy or alternatively too small samples to make any credible conclusions.

5.7.2 Step 2

For the class of business identify the different perils for which a model is required. All classes of peril may be combined or subdivisions may be made. For example, for motor business the subdivision might be for theft, third party, own damage and other.

5.7.3 Step 3

For each peril a frequency distribution will be required and subsequently the delay pattern appropriate to the actual payment pattern will be required.

5.7.3.1 Step 3.1

In order to determine the frequency distribution the first step is to divide the observed number of occurrences by the appropriate measure of exposure, that is the time for which the policies observed were exposed to the possibility of actually incurring a claim.

Now for this frequency distribution, say for example the frequency of own damage claims, the risk factors affecting the distribution need to be identified[81]. As risk factors are not always captured due to their sometimes unquantifiable nature, rating factors are used instead[82].

Rating factors identified might be the age of the driver, the gender of the driver, the duration for which the driver has been insured, the power to weight ratio of the vehicle and whether or not the vehicle is equipped with safety features such as ABS braking systems.

In order to assess the appropriateness of each factor this author proposes an initial one-way analysis of each rating factor. This will involve grouping the entire data set by the one factor and testing whether or not the means of the different groupings are statistically different.

In order to test whether or not two groupings are statistically different the standardised difference in the means can be considered. This is calculated in the following way:

$$Z = \frac{\frac{1}{n_1}\sum_{i=1}^{n_1} X_i - \frac{1}{n_2}\sum_{j=1}^{n_2} X_j}{\sqrt{\frac{1}{n_1}\sum_{i=1}^{n_1}\left(X_i - \bar{X}_1\right) + \frac{1}{n_2}\sum_{j=1}^{n_2}\left(X_j - \bar{X}_2\right)}} \qquad (5.29)$$

where n_1 and n_2[83] are the exposure durations of the first and second group and the X's refer to the number of observations in each group. Z will have a standard normal distribution[84] and therefore be tested on the grounds of the probability of a certain observation actually having a standard normal distribution. Significant differences will normally be assumed for $|Z| > 1,96$.

---

[81] It is important to note that the investigations regarding rating factors are required before any business has been written at the system development stage. Preferably more factors rather than fewer need to be captured as this will facilitate the future analysis of experience. Not all fields captured may have a direct bearing on the risk but their impact needs to be tested over time. Applying a structured thought process to the actual risk is very useful in identifying possible risk factors.

[82] The distinction between risk factors and rating factors is common in the actuarial literature. In essence, risk factors are factors which have a direct impact on the risk to be evaluated. For example the mileage of the vehicle. As the mileage is not always readily available and difficult to verify a rating factor may be used instead, for example, the use of the vehicle (business, leisure, etc.).

[83] In an earlier chapter data considerations were considered as well as the volume of data necessary to form a statistically significant opinion regarding the experience.

[84] For sufficiently large sizes of n_1 and n_2

For example, when considering the gender of the insured, data will be grouped by male or female and a hypothesis test conducted to ascertain whether the means of the two groupings are significantly different.

This one-way analysis will give the practitioner an initial feel as to which risk factors have a significant impact on the risk. In choosing which factors to test, the practitioner should rely on expert opinion in the market or the company.

5.7.3.2 Step 3.3

Once the rating factors to be used have been considered, possible correlation between the rating factors needs to be considered. As an example consider age and gender. The influence of gender on age will not be the same for all ages. Experience of young males will tend to show a greater difference to young females than mature males to mature females. This means that the rating factors are correlated.

In order to allow for this feature, the age and gender should be combined as one factor. That is to say, instead of using a gender differential and a separate age differential, use an age-gender differential. Alternatively a correlation factor could be included.

The correlation between two factors is determined as follows:

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_y}$$

Where $Cov(X,Y)=\frac{1}{n}\sum(x_i-\mu_x)(y_i-\mu_y)$

i.e. the correlation is calculated by dividing the sample covariance by the product of the standard deviation of the two parameters investigated.

This approach can, however, only be applied where sufficient volumes of data exist in order to justify such a grouping.

5.7.3.3 Step 3.4

When the factors affecting the parameter of the frequency function have been determined generalised linear modelling techniques can be applied to obtain a function that will yield the mean value, given the risk characteristics.

The purpose of this exercise is to determine a function of rating factors, based on past experience, such that the mean frequency of a certain homogeneous group is determined accurately by applying the function.

f(risk factors) = mean parameter.

In order to determine the most appropriate set of risk factors it should be borne in mind that the higher the number of risk factors, the higher the likelihood will be of reducing the unexplained variation of the model. The analogy to this is the fitting of a polynomial function to a certain set of points. The higher the rank of the polynomial, the higher the number of points that will lie on the function.

The aim is therefore to obtain the best possible fit with the least number of variables.

This will involve decreasing the unexplained variation

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 ,$$

in comparison to the total variation of the model,

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 .$$

The multiple coefficient of determination, R^2, is then defined as

1 – (Unexplained variation / Total variation) (5.30)

A variety of approaches optimising this problem set are discussed in statistical literature. In essence these techniques entail choosing parameters that contribute the most to declaring the total variation. One method is testing one factor at a time and then choosing the factor with the biggest reduction in variation. Once the first factor is determined the second will be tested and this process will continue until further additional factors have no additional benefit by way of declaring variation. This process is often referred to as Forward Selection. Most statistical applications allow for the automatic calculation of this process. A similar approach could be to start with all parameters and then test the impact of eliminating each factor one at a time and subsequently obtain the model with the highest multiple coefficient of determination and the lowest number of parameters.

5.7.3.4 Step 3.5

Once the model building has been completed a frequency model per peril for a certain class of business will be available to model future expected experience for certain model policies given the risk characteristics of the policies.

5.7.4 Step 4

The next step is to fit a severity function to the claims distribution. As risk factors will have a direct impact on the severity of a claim it will be necessary to apply the same methodology as for frequency calculations to severity calculations. Note that different risk factors will be applicable for the severity distribution as opposed to the frequency distribution.

Severity of claims have a larger spread than frequency of claims and as a result parameter error is more likely for a subdivision by rating factor than would be the case for frequency distributions.

This author proposes that the practitioner first find an appropriate severity distribution for the peril involved using the techniques explained earlier in this chapter. Subsequent to this, the same methodology as used to calculate the average frequency for a certain set of risk characteristics can be used to calculate the average severity and the distribution applied can then be scaled accordingly.

This approach would, however, not allow for changes in the variability of claims per peril by risk factor. For substantial volumes of data proper allowance could be made by fitting a separate distribution for each set of risk characteristics.

The severity calculated per claim will be calculated as the present value of the expected claim. For long tail classes of business this claim can be spread according to the run-off pattern identified for the claim. This process may introduce spurious accuracy for shorter classes of business and may be ignored for such classes.

5.8 Summary

This unit contains a detailed investigation into the methodology required to quantify the liability area of risk. It includes the following:

1. A detailed discussion on the topic of stochastic simulation.
2. Reference to appropriate stochastic distribution for use in a general insurance operation.
3. A discussion on the manner in which to obtain the appropriate distributions.
4. A discussion on generalised linear modelling.
5. A discussion on the determination of cash-flow development trends in claim payments.
6. The assimilation of all concepts in the model building process.

In the next chapter the area of asset risk is considered.

6. Quantifying the asset area of risk

6.1 Introduction

In the previous chapter the liability area of risk was considered for claim outgo only. This chapter considers the modelling of asset proceeds as well as the stability of asset values over time. This is often also referred to as market risk.

Asset proceeds consist of income and capital gain. Each element may be dependent on other factors or may be modelled based only on past data using time series techniques. Asset returns by asset type tend to be more dependent than liability outgo due to the fact that all returns are generally obtained in the same economy. To allow for this feature, asset proceeds are often modelled as a function of a certain base index.

Some of the techniques and processes used for asset modelling will firstly be considered and subsequent to this some of the approaches required for model building will be discussed.

6.2 Stochastic modelling

For the purpose of this study the assets' proceeds will also be modelled stochastically. This will ensure consistency to the modelling of liability outgo. Details regarding the process of stochastic modelling have already been discussed in the previous chapter.

6.3 The modelling of asset proceeds[85]

Research on the modelling of asset proceeds numbers many volumes. Due to the short-term nature of asset returns investigated in this study the results will tend to be more variable than for business of a long term nature such as a pension fund. The result of this is that too much refinement in the investment model is considered spurious and this chapter has been prepared with this in mind[86].

6.3.1 Random walk models

This is a simple model whereby the original asset value can either move up or down by a certain specified amount over a certain specified time period with a certain specified probability. It is, of course, possible to expand the range of outcomes per time period. In the example below Bernoulli trials are used but other probability distributions will be just as appropriate.

[85] For a more detailed discussion the reader is referred to a paper prepared by A. D. Smith concerning a workshop on stochastic Asset Models held at the General Insurance Convention: 18-21 October 1995

[86] Short-term asset movements are driven by sentiment rather than fundamental evaluation of future investment proceeds. As a result the movements can be erratic and extremely unpredictable.

Example:
Let *h* represent time.
Let X be a stochastic variable normally distributed with mean μh and variance $\sigma^2 h$ representing the force of asset growth.
Let Y_t be the asset value at time t, then:

$$Y_{t+h} = Y_t \exp(X) \tag{6.1}$$

Different relationships between Y_{t+h} and Y_t can be tested as well as different distributions for the random variable X. Authors such as S. COUTTS have suggested this model due to its simplicity and ease of application.

The graph given below shows a random walk for an asset portfolio with expected mean return of 10% per annum and standard deviation of 7%. The asset return is expected to be normally distributed.

Figure 7: Simulated asset growth

In order to obtain such a model the following process is required:

1. Obtain the growth for period *h* by considering the ratio of Y_{t+h}/Y_t for all available t.
2. Take the natural logarithm of these ratios in order to determine the force of asset growth of the period *h*, i_{t+h}.
3. Obtain the mean and the standard deviation of all the observed i_{t+h}. This provides the mean growth for the period *h* as well as the standard deviation of growth.
4. Test the results for different periods of duration *h*.

In order to generate the growth over time consider the current value of assets. This value should preferably be smoothed in order to ensure a stable starting point. In order to project the asset value, generate[87] an expected asset growth rate for period h by assuming the asset growth rate distribution estimated as described above.

This distribution will be assumed to be normal with mean and standard deviation as obtained. This assumption can be tested given the information set obtained as described above. Such a test will entail grouping the observations in ranges based on the mean and standard deviation as calculated. For example, observations less than three standard deviations below the mean, observations between two and three standard deviations below the mean and so forth. The actual number of observations can then be compared to the expected, assuming the normal distribution as indicated. A chi-square test can be applied to test the reasonableness of the fit.

The process is repeated over different time periods h in order to simulate the growth of asset proceeds over time. Note that in this instance the proceeds combine both income and capital gains. A specific total time period will be chosen, say one year, for which the asset values will be available at every duration h.

By repeating this projection on a random basis a distribution of the cumulative growth in asset values for the period specified as well as at every duration will be obtainable.

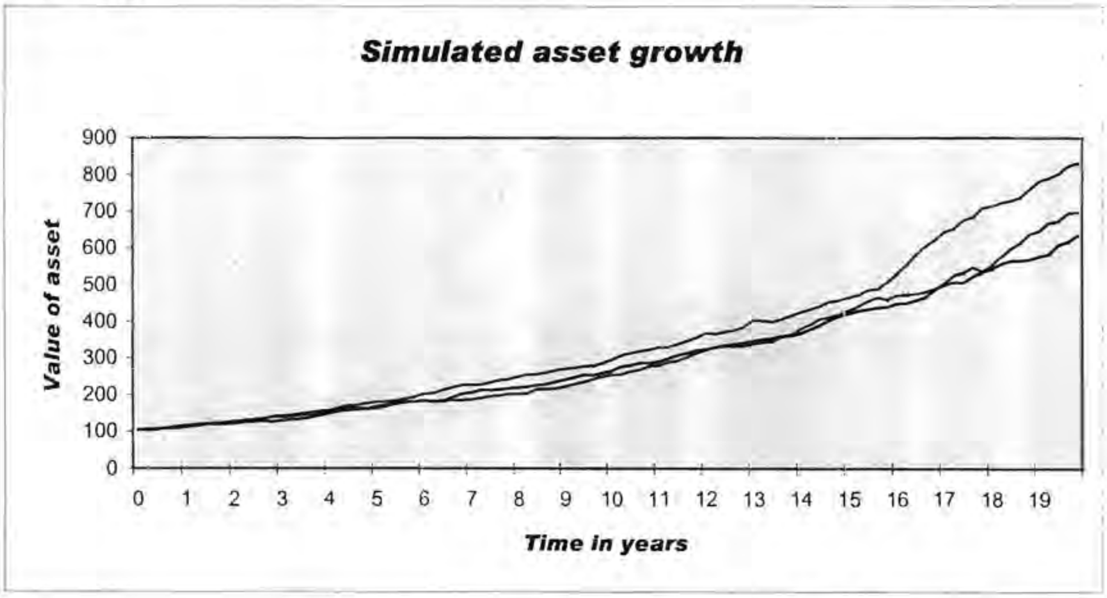


Figure 8: Simulated asset growth examples

[87] Details regarding the generation of normal distribution function values are discussed in the previous chapter.

The graph above shows three additional trials of the model used in the previous graph. This clearly shows the variability in the simulated results. By extending the number of simulated outcomes a distribution of asset growth can be determined at each time period.

6.3.1.1 Example

In order to make this approach more apparent to the reader, the JSE All share index was considered on a monthly basis for the period May 1981 to August 1998. The graph of the index value[88] over this period is shown below:

Figure 9: The JSE All Share index

For the sake of this example the entire data set was used to fit a model. It must be recognised that such a model is only likely to be appropriate if the macro economic environment were to repeat itself in the next 20 year period. This is extremely unlikely and therefore consistency between asset growth and inflation assumptions is critical when setting an appropriate model.

Fitting a normal distribution to the force of interest (i.e. the force on capital growth reflected in the index) the following results were obtained:

[88] The index value was calculated as the average of the highest and lowest values of the index during the month.

Table 8: Annualised returns

Summary Period	Average	Standard Deviation
Monthly	14.5%	18.6%
Quarterly	15.3%	22.5%
Half-Yearly	15.2%	23.7%
Annually	15.9%	21.4%

Table 8 shows the degree of variability inherent in fitting a normal distribution based on monthly, quarterly, half-yearly and annual growth. This reinforces the idea that great care is required to project future expected experience.

The density of monthly returns are shown below:

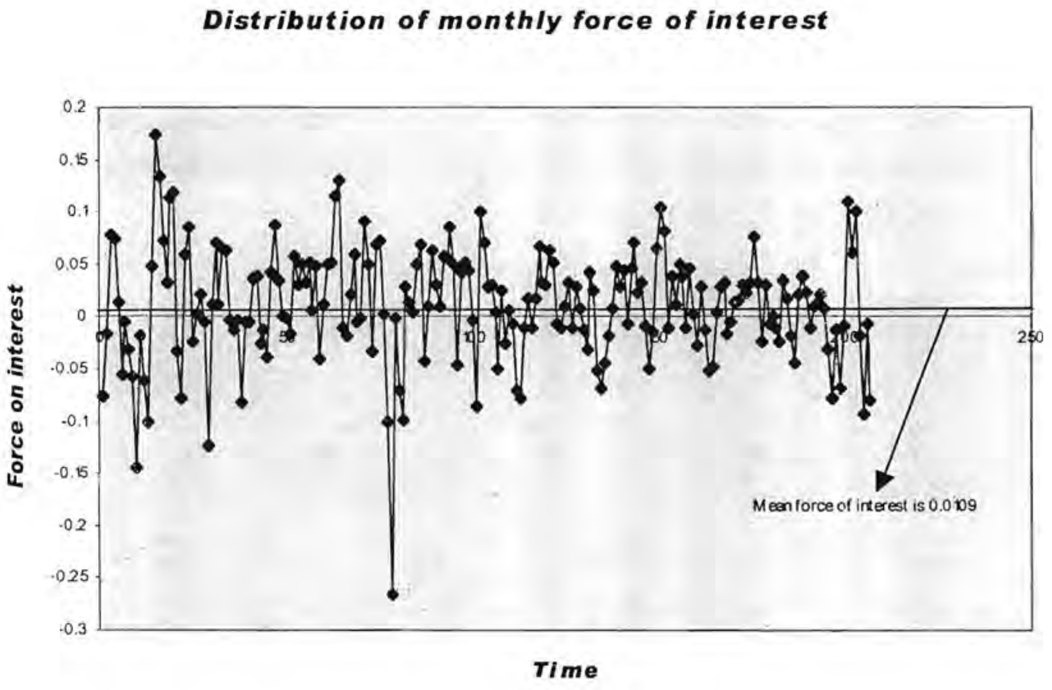


Figure 10: JSE monthly force of interest

In order to apply a random walk model to this data, the monthly force of interest is assumed to have a normal distribution. A random sample is then drawn from this distribution and this is used to project the future capital values of the index. Note that this model does not include the dividend yield which should be allowed for separately.

Comparing the actual density to the expected density yields the following:

Figure 11: Distribution of monthly force of interest

It is clear that assuming a normal distribution is a crude assumption. Nevertheless, this model provides a simple basis and as the tail ends of the expected normal distribution are similar to the actual observations, this model will be appropriate to assess variability in solvency, for example.

6.3.2 Auto-Regressive Models

The term auto-regressive means that a regression is applied to the same data set i.e. regression onto itself. This is appropriate in time series modelling where the future value is a function of some of the previous values in the series. Auto-regression is an additional way in which to set a relationship between a dependent and explanatory variables with the focus on the correlations between the same series at different time intervals. Note that a series can also consist of a multivariate series[89].

These models can take a variety of forms. Some of the models converge to a long-term mean while others are based on the experience of the variable over a certain time period. Some models are therefore assuming that past experience is indicative of future experience while others use this assumption together with the assumption that the variable will converge to a certain mean over time.

[89] Multivariate series are not discussed in this study.

For a detailed discussion on auto-regressive models the reader is referred to the core reading of the INSTITUTE AND FACULTY OF ACTUARIES subject 103.

6.3.2.1 Background to auto-regressive models

The aim of time series modelling is to identify, isolate and remove all possible deterministic trends in a time series through differencing and to end up with a stationary process[90]. Trends can include:

1. Seasonal trends over every year in which case differencing on a yearly basis will be required.
2. Cyclical trends over time. The duration of the cycle would be required before differencing could be applied.
3. A trend over time longer than the period of the study. By differencing this trend can also be eliminated. Consider the growth in an index. By differencing the yield per period can be obtained. This yield may have the same distribution over time thereby allowing the investigation of a stationary process.
4. Random fluctuations. These will not be eliminated through differencing and the aim of the investigation would be to model the randomness in the time series[91].

6.3.2.2 The Box and Jenkins[92] methodology

$\{Y_t\}$ is an Auto-regressive integrated moving average ARIMA (p,d,q) process if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} \qquad (6.2)$$

where $X_t = (1-B)^d Y_t$ and

Y_t is the variable under investigation at time t and

α_i is a correlation factor between X_t and X_{t-i}.

Z_t is the error term at time t which has a standard normal distribution.

β_t is a correlation factor between X_t and Z_t and B is a backward shift operator. therefore (1-B)[93] is a backwards differencing operator which means that the

[90] A stochastic process is said to be stationary if the joint distributions of $X_t, X_{t+1},,,,,X_{t+n}$ and $X_{k+t}, X_{k+t+1},,,,,X_{k+t+n}$ are identical for all t and k+t. This means that the statistical properties of the process remain unchanged as time elapses.

[91] This was already referred to in the previous section where the error distribution was considered.

[92] This methodology is referred to in the Core Reading on applied statistics (Subject 103) issued by the Institute and Faculty of Actuaries.

difference between the t^{th} and the $(t-1)^{th}$ element is taken. The d denotes the number of times the backward shift operator is applied. $\{Y_t\}$ is said to be integrated because $\{X_t\}$ has been obtained through differencing and therefore $\{Y_t\}$ will be obtained through integration (summation).

The autocovariance and the autocorrelation of a time series are important because of the necessity to determine the correlation between future values of a time series and past values.

The covariance of any X_r and X_s from a stationary sequence depends on r and s only through the difference r-s. The autocovariance is defined as follows:

$$\gamma(r-s) = \operatorname{cov}(X_s, X_r) = E(X_s X_r) - E(X_s)E(X_r)$$

The autocorrelation function for elements of the time series with lag n is defined by:

$$\rho_n = \frac{\gamma(n)}{\gamma(0)} \tag{6.3}$$

The autocorrelation is often plotted for different lag periods. Where the autocorrelation tapers off to zero slowly, this indicates a correlation to previous values in the time series and therefore an autoregressive component is required. Where the autocorrelation tapers off to zero quickly, this indicates limited correlation to previous values in the time series and therefore a moving average component would seem more appropriate. For other trends a different structure of autoregressive components as well as moving average components may be required.

In order to set the autoregressive coefficients, the partial autocorrelation is used. The partial autocorrelation is defined by

$$\phi(k) = \frac{\det P_k^*}{\det P_k} \tag{6.4}$$

where P_k is the k x k autocorrelation matrix

[93] (1-B) is often referred to in texts as ∇.

$$P_k = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{pmatrix}$$

and P^*_k is P_k with the last column replaced by $(\rho_1, \rho_2, \ldots, \rho_k)^T$.

Note that the partial autocorrelations provide the solutions to the vector $\underline{\alpha}$ in the equation

$$\underline{\rho} = P_k \underline{\alpha} \text{ [94]}. \qquad (6.5)$$

Therefore, once the structure of the model has been tested the coefficients will be obtained through investigation of the partial autocorrelations.

This model does not necessarily converge to a long-term mean as is evident from the formula given in (6.2) above. In some instances it may, however, be required to construct the model such that the variable under investigation converges to a long term mean. In this instance the model given above will be of the following form:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \left(1 - \sum_{i=1}^{p} \alpha_i\right)\mu + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} \qquad (6.6)$$

where μ is the long-term mean and $\sum_{i=1}^{p} \alpha_i < 1$.

[94] For example, consider an autoregressive process ARIMA(2,d,0) of the form

$$X_n = \alpha_1 X_{n-1} + \alpha_2 X_{n-2} + e_n.$$

Considering the covariance for n = 1 and n = 2, we have the following two equations:

$$\gamma_1 = \alpha_1 \gamma_0 + \alpha_2 \gamma_1$$
$$\gamma_2 = \alpha_1 \gamma_1 + \alpha_2 \gamma_0$$

As the covariance for lag 1 and lag –1 is the same thing. By dividing by γ_0 this yields

$$\rho_1 = \alpha_1 + \alpha_2 \rho_1$$
$$\rho_2 = \alpha_1 \rho_1 + \alpha_2$$

And these equations can be presented in the matrix notation of (6.5) above.

6.3.2.3 Example

The autoregressive model can be estimated using some standard characteristics.

The mean of the stationary model can be estimated by the sample mean

$$\hat{\mu} = \tfrac{1}{T}\sum_{t=1}^{T} x_t$$

The autocovariance function $\gamma(t)$ can be estimated using the sample autocovariances

$$\hat{\gamma}_t = \tfrac{1}{T-t}\sum_{t=t+1}^{T-t}(x_t - \hat{\mu})(x_{t-t} - \hat{\mu})$$

The autocorrelation ρ_t can be estimated using

$$\hat{\rho}_t = \frac{\hat{\gamma}_t}{\hat{\gamma}_0}$$

and the partial autocorrelation $\phi(t)$ can be estimated using

$$\hat{\phi}(t) = \frac{\det \hat{P}_k^{*}}{\det \hat{P}_k}.$$

Recalling Figure 10: JSE monthly force of interest an autoregressive model can be fit to this process, which appears to be stationary.

An investigation into the correlgraph, the autocorrelation at different lags yield the following:

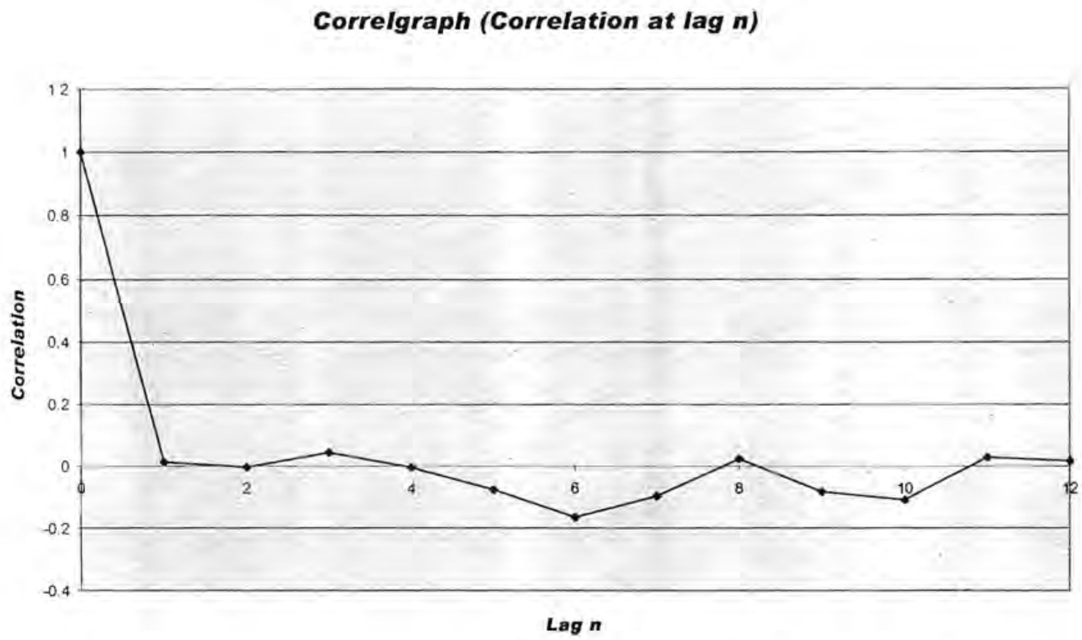


Figure 12: Correlgraph for the force of growth of the JSE All-Share index

The correlgraph shows no clear correlation between the force of growth at the different lag periods. This leads to the conclusion that a random walk model is probably a better model for this process than an autoregressive model.

6.3.2.4 An asset model for use in South Africa

THOMSON (1994) developed an asset model for use in South Africa. The model was originally set up for asset liability matching purposes for the life and pension fund industries but as short-term insurance companies often apply a long-term investment philosophy and this model might therefore sometimes be considered appropriate.

The model considers a variety of asset classes namely,

1. Equity dividend growth and dividend yields. (i.e. the total return on equities).
2. Inflation is also considered as a base for the other asset classes
3. Property unit trust dividend growth and dividend yields.
4. Direct property investment dividend growth and dividend yields.
5. Long-term interest rates.
6. Short-term interest rates.

MAITLAND (1996)[95] provides a review of the Thomson model.

[95] Maitland provides the following summary of the Thomson model (only the equity, inflation and short-term interest rate models are mentioned below):

The model is an integrated autoregressive model and therefore included in this section.

6.3.2.5 Dealing with varying variance in autoregressive functions

The functions considered up to this stage all make use of error structures that have distributions with constant variance. The treatment of changes in the mean value has already been considered through the differencing of a series.

The variables η_t are each independently and identically distributed N(0,1) random variables.

The equity dividend yields and dividend growth rates are modelled as $EQDG_t$ and $EQDY_t$ respectively where $EQDG_t$ is the mean force of equity dividend growth in year t and $EQDY_t$ is the natural logarithm of the All Share Index Dividend Yield at time t per cent.

The equations for equities are as follows:

$$EQDG_t = 0.093 + 0.116\eta_t + 0.076\eta_{t-1}$$

$$EQDY_t = 0.310 + 0.810EQDY_{t-1} + 0.198\eta_t$$

The intended model for inflation is

$$INFL_t \approx 0.008 + 0.899INFL_{t-1} + 0.088EQDG_t - 0.079EQDG_{t-1} + 0.077EQDG_{t-2} - 0.069EQDG_{t-3} + 0.020\eta_t$$

where $INFL_t$ is the mean force of inflation in year t.

The interdependence in the model is clear. Inflation is modelled as a function of dividend growth, which is similar to the growth considered in the share index values.

Short-term interest rates are modelled as the sum of a unit gain function $Z_{MINT,t}$, and an error process $MINTZ_t$. The variable $MINT_t$ is the annual force of interest on money-market instruments as measured by the GINSBERG MALAN & CARSONS Money-market Index.

The equations for short-term interest rates are as follows:

$$Z_{MINT,t} = 0.004 + 0.141INFL_t + 0.85Z_{MINT,t-1}$$

$$MINTZ_t = 0.008 - 0.091EQDG_t + 0.885LINTZ_t + 0.019\eta_t + 0.010\eta_{t-1}$$

$$MINT_t = Z_{MINT,t} + MINTZ_t$$

In order to start the simulation, the one-step-ahead residuals for both MA(q) and ARMA(p,q) processes are needed. They are -0.04587 for *EQDG* and -0.00552 for *MINTZ*. These translate into the following N(0,1) errors which can be used directly in the difference equations:

$\eta_{EQDG,1993} = -0.3954$
$\eta_{MINTZ,1993} = -0.2905$

It is quite likely that the variance of a process will not be constant over time but will change and have periods of very stable variance followed by periods of very unstable variance. This is intuitively obvious as stock markets are known for durations of great variations and durations of fair stability.

ENGLE (1982) developed autoregressive models with conditional heteroscedasticity (ARCH). Later BOLLERSLEV (1986) developed these models into generalised autoregressive models with conditional heteroscedasticity (GARCH).

Normal autoregressive models of the form

$$X_t = \alpha_1 X_{t-1} + Z_t$$

have been considered above. Z_t was considered to have a certain distribution with mean μ and variance σ^2.

The generalised autoregressive function with conditional heteroscedasticy considers the model for σ^2. This function is presented as

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k \left(X_{t-k} - \mu\right)^2 + \sum_{l=1}^{q} \beta_l \sigma_{t-l}^2$$

with constraints: $\alpha_0 > 0$ and the sum of all other coefficients smaller than 1. This ensures that the function is mean reverting. Note that the smaller the other coefficients the sooner the variance will revert back to the mean. This mean reversion process captures the clustering of variability that was mentioned earlier SHAH (2001).

Where the last term is not considered, this is said to be an ARCH(p) process and the general case is referred to as a GARCH(p,q) process.

6.3.3 Fractal models

Consider a graph of asset growth where the size of growth over time is not stated. Some practitioners are of the opinion that it is not possible to discern the actual period of time for which the asset growth pattern was investigated. As a result asset returns can be based on a Poisson process. That is to say, irrespective of the time period considered between consecutive asset values a Poisson process can be applied.

Fractal models are based on the assumption that the same probability distribution will hold irrespective of the time span considered and as such are very similar to a compound claims process with the main difference that downward adjustments should also be allowed for. (For the collective risk model claims are always positive.) Asset changes therefore follow a Poisson process and the severity of the changes can follow a variety of distributions. The actual quantum of the severity will be a function of the time period considered between consecutive asset values.

This study does not consider fractal models in any further detail. The reader is referred to the work of A.D. SMITH (1994) contained in the proceedings of the GISG convention of 1994 for further information regarding fractal models.

6.3.4 Chaotic models

Chaos theory has become a buzzword in the modelling of various enterprises. The reason for this is probably attributable to the interesting graphical patterns generated using chaos theory.

These models are based on non-linear relationships between variables at different times. The models are termed "chaotic" because the interrelationships over time are very complex due to the non-linear relationship between the explanatory and dependent variables. These models have the drawback of producing the same values if the starting set is exactly the same. Depending on the nature of the starting set of variable, however, this may never occur in practice, assuming of course that the model is sufficiently complex.

Due to the complexities and level of judgement required in estimating the interrelationships between the explanatory and dependant variables, these models are not considered further.

6.3.5 Neural networks

Neural networks are designed to replicate the operation of the human brain. That is to say, for a certain parameter set as input the network will generate a certain outcome. Time series modelling and regression analysis can therefore be obtained from neural networks as well.

Neural networks provide more flexibility in that the manner of linking a set of explanatory variables to a dependant variable is based on a weighting system. This allows a variety of combinations of explanatory variables to be used in generating a result based on some function of the explanatory variables. This result can again be applied in subsequent functions until the dependent variable is determined. Software packages are available to generate the best fit. These fits are, however, not always understandable as a result of the complexities of the network.

Neural networks provide an additional tool to practitioners to investigate experience. Given the level of uncertainty in asset returns and the fact that the human brain has not yet identified a suitable relationship between asset growth and explanatory variables, the topic of neural networks is not considered further in this study.

6.4 Fundamentals of Asset Models

An extensive array of models exists to model asset proceeds. The assumptions underlying each group of models as well as their level of complexity may be different but the following factors should also be borne in mind when determining appropriate models (SMITH (1994)):

1. *Practical Issues*

The model must be based on available information. Furthermore, the cost of development should be weighted against the added benefit of a more accurate assessment of the asset proceeds.

2. *Theoretical justification*

The model must conform to sound economic and actuarial principles and should always be checked for reasonableness. This is not a very strict requirement as there are many theories as to how the market behaves. Reasonableness is, however, crucial to avoid non-sensical predictions for the future.

3. *Fit to historic data*

The model must be able to conform to previous data sets. If the model is unable to produce reasonable estimates of past experience it will surely be unable to predict future experience as well. Auto-regressive analysis lends itself perfectly to the testing of models against past data. As a result, a considerable amount of literature is available on auto-regressive asset models.

4. *Market assessment of future trends*

There are also implicit assumptions to be made regarding future asset proceeds available in the market. These assumptions are reflected in the yield curve, market prices as well as in the prices of derivative instruments. The inclusion of parameters relating to these entities will, therefore, prove useful.

6.5 Model building

This author proposes the use of auto-regressive models given the level of flexibility that can be applied, the ease of understanding of these models, the consistency with liability models, the direct link to historical trends and the allowance for future experience. As these methods are also applied extensively in other disciplines, it should be easier for the reader to follow the reasoning below.

The main aim is to find a model which provides an appropriate fit to past experience as well as sensible future projections. The following approach can be applied:

6.5.1.1 Step 1

Consider the economic factor to be modelled. This can be the total return on an asset, the income stream, or the capital appreciation on a certain asset class. Furthermore as the return may depend on other factors, it would be appropriate to model these factors first. R. Thompson (1994) has developed an asset model for use in long-term asset projections such as pension projection. This model, similar to the Wilkie model, is based on a base model[96], in this case real dividend growth, and subsequently inflation and other economic factors can be linked to this model.

This approach is appropriate in allowing for correlations between different economic factors.

6.5.1.2 Step 2

Consider the elements which could have an influence on the projected value of the dependent variable, in particular those elements which can be considered to influence the current value of the dependent variable. This can be either the value of the factor some time ago (or at various durations), a moving average over a certain period of time, a mean base value that is assumed to be appropriate for the period of projection or some other variable such as a forward rate derived from current yield curves.

The elements which are the most important will be identified through autoregressive techniques explained above. The aim is to find the smallest set of explanatory variables that declares the most of the variability.

A variety of factors will need to be tested to ensure that an appropriate model is found.

Considering the Box-Jenkins methodology explained earlier, the model can either be investigated for a specific index or the difference operator can be applied to consider the rate of growth. The difference operator stabilises the index and eliminates any seasonal trends. Furthermore, the Box-Jenkins methodology is applied to series that appear to be stationary.

6.5.1.3 Step 3

Once the appropriate explanatory elements have been found it is important to test the model against past experience and, in particular, to consider the error structure over time. It may be necessary to allow for a GARCH(p,q) model if the error structure does not have constant variability over time. Furthermore, the error structure might not have a normal distribution and this would need to be allowed for.

6.5.1.4 Step 4

The only step remaining is to test the model against randomly selected periods of past data to ensure that the model will yield appropriate results. It must be stressed that this process is part art, part science and that the practitioner will

[96] The base model is the model of an economic variable on which some or all of the other economic variables are built. For example a model on inflation that will affect all other real return assets.

often be confronted with situations where it is extremely difficult to form an opinion as to the validity of the model.

This approach is based on short-term variability and as a result may not recognise the long-term theoretical structure of asset movements but rather aim to replicate the volatility exhibited in the market. This is intentional.

Once a model has been set this can then be applied in similar fashion as was done to liability models in projecting future income.

6.6 Summary

In this chapter asset modelling was considered. The following issues were addressed:

1. The different types of techniques available to model asset proceeds were considered. Considerable focus was placed on random walk models as well as autoregressive models. Other techniques were mentioned and the interested reader is referred to the references given.
2. The considerations required before setting an asset model were considered.
3. The process of fitting an asset model was considered. The techniques are similar to those applied in fitting liability distributions though the basic model may be completely different. This process is visited in further detail in the annexure which provides an example of the application of the ideas set out in this dissertation.

In the next chapter other areas of quantifiable risk will be considered.

7. Quantifying the other areas of risk

7.1 Introduction

The previous two chapters considered the techniques required to model the liability and asset area of risk. This chapter aims to give some background to the other areas of risk mentioned in the introduction. It goes on to consider how all these techniques are used to form a basis to build the model insurer.

7.2 Other areas of risk

7.2.1 Credit risk

Credit risk refers to the possibility that monies receivable from brokers, reinsurers or other counter parties will not be forthcoming as a result of insolvency of the third party. In addition, credit risk can also refer to the risk of default on asset investments.

There is no short way in which to consider this risk. A starting point will be the credit rating of third parties. In order to model this area of risk the different sources of premium income need to be considered and with this the possibility that the provider of the premium will default and that the premium will not be forthcoming or that balances will be lost. Similar evaluations will be required for reinsurance recoverables.

This has been identified as a substantial area of risk given past market experience. As full details regarding third parties are always readily available, this risk must be managed using proper risk management principles e.g. not entering into a relationship with risky partners. In addition, a contingency loading should be included based on past market experience.

The contingency loading can be modelled by assuming a probability that a certain section of premium will not be forthcoming or that a reinsurer will default.

7.2.1.1 A structured approach to credit risk quantification

CREDIT SUISSE (1997) recommended the following structure in order to consider the default of a counter party:

1. Determine the size of the exposure to the counter party.
2. Determine the probability of default.
3. Determine the possible recoveries to be made from the counter party.

The approach is exactly the same as that required to model liability risk and explained previously in this study. Again the frequency of losses is considered and in the event of a loss the size of the loss is determined through the size of the exposure less possible recoveries. Hence the detail explained in the chapter on the modelling of the liability area of risk is relevant here as well.

The model is often applied as either a continuous or discreet approach. In the continuous instance, probability distributions are used whereas in the discreet approach deterministic probabilities and transition matrices are used.

7.2.1.1.1 Determining the size of exposure

This needs to be done on a continuous basis and is as relevant for credit risk as it is for liability risk. The maximum possible loss from any one counter party or any class of business needs to be considered at all times. Banks and financial institutions often place limits and the level of credit exposure that they will allow the institution to carry.

In similar fashion insurance companies have limits on the levels of exposure that they are prepared to take on certain perils for example catastrophes. Credit exposure is much more volatile than insurance exposure and banks are required to monitor these exposures on a daily basis and are also regulated on this basis.

The BASLE accord stipulates the regulatory requirements for banks and relies on the internal models of banks to regulate the industry. Continuous evaluation of exposure is critical in this regard.

7.2.1.1.2 Determining the probability of default

The probability of default has been suggested to be linked to the credit rating of the counter party[97]. It is, however, important to realise that these credit ratings are not static and can change on a regular basis. As a result transition matrices

[97] Recall MAHER (2001):

Table: Quantification of Rating Agencies' Rates

Moody's	S & P	Expected Probability of ruin in 10 years
Aaa	AAA	0.00%
Aa1	AA+	0.10%
Aa2	AA	0.10%
Aa3	AA-	0.20%
A1	A+	0.40%
A2	A	0.70%
A3	A-	1.00%
Baa1	BBB+	1.40%
Baa2	BBB	2.00%
Baa3	BBB-	3.40%
Ba1	BB+	5.20%
Ba2	BB	7.40%
Ba3	BB-	9.70%
B1	B+	12.20%
B2	B	15.00%
B3	B-	19.20%
Caa	CCC+	35.80%

have been developed to model the transition of credit ratings and therefore the probability of default over time. The following provides a transition matrix compiled by STANDARD & POOR'S (1996)

Table 9: Example of credit rating transition matrix

Initial Rating	Rating at year-end							
	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90.81	8.33	0.68	0.06	0.12	0	0	0
AA	0.7	90.65	7.79	0.64	0.06	0.14	0.02	0
A	0.09	2.27	91.05	5.52	0.74	0.26	0.01	0.06
BBB	0.02	0.33	5.95	86.93	5.3	1.17	0.12	0.18
BB	0.03	0.14	0.67	7.73	80.53	8.84	1.00	1.06
B	0	0.11	0.24	0.43	0.48	83.46	4.07	5.20
CCC	0.22	0	0.22	1.3	2.38	11.24	64.86	19.79

Such matrices need to be evaluated on a regular basis. Clearly the basis is not static and would need to allow for other external variables such as economic cycles. MCKINSEY consultants have developed a CREDIT PORTFOLIO VIEW model that allows for such external economic variables.

Credit rating and transition matrices have also come under substantial criticism because of the fact that credit ratings are only issued after substantial time delays due to the investigations required to set a credit rating.

DVORAK (2001) argues that the KMV model provides a better approach. This model is based on a company's share price and argues that the market is aware of credit problems long before the credit agencies. The model uses the share price as a proxy for the market value of its assets and as the solvency of a company deteriorates i.e. as the market value of assets converges to the value of liabilities of the company, the model allocates a probability of default to the company.

DVORAK cites an investigation conducted by KMV and notes the following:

1. The credit markets recognise that not all companies rated the same are equally risky.
2. The range of expected default within a broad rating grade is usually very wide.
3. In an AA- rating group with default probability of 0.2% the default probability calculated by KMV was between 0.02% and 19.7% for companies with the same grading.
4. Therefore in some cases the credit risk was over stated ten times and in others under stated a 100 times.

The approach applied by KMV is also applied by other practitioners. The TY model developed by YAKOUBOV Y.H., TEEGER M.H. and DUVAL D.B (1999) applies a similar approach to the setting of the equity model.

7.2.1.1.3 Determining the possible recoveries to be made from a counter party

The recovery to be made depends on the ranking of the debt to the counter party as well as the net asset value structure of the counter party. It is extremely difficult to model these items appropriately as the detail of all debt issued to the counter party would be required before the recovery could be considered. A prudent approach is therefore preferably in evaluating recoveries i.e. rather recognise recoveries when they are booked than making a projection of expected recoveries. This approach will lead to an overestimation of credit risk but it is more prudent.

Loans can rank in the following manner and this must be borne in mind:

1. Secured loans
2. Unsecured loans
3. Preferential shares
4. Ordinary shares

Credit risk is very important to banks as this constitutes the biggest share of their exposure. General insurance companies have neglected this area of risk and at their own peril as indicated by the research of FINNIS (1995) which indicates that 10% of all insolvencies of US general insurance companies are as a result of credit risk.

7.2.2 Growth risk

Growth risk refers to the risk that volume increases too quickly and as a result capital support for the business is inadequate or, alternatively, the risk that volume reduces and as a result the marginal cost of writing business becomes too high.

The relative sizes of policies are also important as growth in small policies leads to more expenses than growth in large policies as a proportion of the premium income.

Growth risk is modelled by applying sensitivity tests to the model insurer. As these items are seen as management decisions they are not considered to be stochastic variables. Instead, a proper expense investigation is required to assess the costs involved in writing business and hence the impact of a change in policy size or policy volume.

Through the use of financing arrangements, general insurance companies are able to alleviate this risk through the use of quota share reinsurance arrangements[98] as long as the additional business is profitable.

7.2.3 Expense risk

An expense analysis entails the allocation of all expenses to different sources of income as described below.

7.2.3.1 Step 1

Consider all the expenses incurred by the general insurance enterprise. The different departments in the company incur these expenses. This will consist of personnel expenses, accommodation expenses, marketing expenses, investment expenses and other overheads.

Overheads will include all other expenses reflected in the revenue account.

The aim is to allocate the expenses as accurately as possible in order to ensure an equitable charge to the different classes of business. The crucial aspect is, however, that all costs need to be covered.

7.2.3.2 Step 2

Consider each type of expense and allocate the expense to a specific function in the enterprise. The expenses will relate either to business or service functions. The business functions should relate to the business flow and should therefore refer to

1. acquisition expenses
2. policy capturing expenses
3. maintenance expenses
4. claim settlement expenses or
5. investment expenses.

The service function costs need to be allocated in similar fashion. This is required to allocate the different expenses to the different line functions.

It must also be considered whether or not the expenses are directly related to the business sold or are only indirectly linked. Direct expenses will be expected to change as business volume changes whereas indirect expenses will be expected to remain fairly stable.

7.2.3.3 Step 3

The different expenses then need to be allocated to different classes of business. This allocation can be done in a variety of ways depending on the expense.

[98] A company which experiences capital strain will be able to cede a proportion of its business to a reinsurance company. This company will then share in the experience in exactly the same manner as the ceding company. Often the reinsurance company will also provide a profit commission which will aid the ceding company. By taking the business off its books the ceding company is able to protect itself against the risk of excessive growth.

Certain classes of expense are quite clear such as commission which is simply a certain percentage of premium. Computer expenses, which will be service costs as well as the amortisation cost of the computer investment, will be allocated according to computer usage.

Another way of allocating expenses is to conduct a time analysis of the functions of personnel. For example within a certain department a percentage of time will be spent on each element of the line function. This expense can be either direct or indirect. By investigating this time allocation for each department it is possible to split the cost by line function and class of business. This approach can be used to allocate personnel costs. Alternatively, such an exercise might already have been completed in the past or the expertise of external consultants can be obtained in order to set a relative cost for each function. The number of times the function is carried out can then be multiplied by its relative cost in order to obtain weights for each type of function whereby the total cost apportioned to those functions can be split. This is referred to as functional costing.

Rent can be allocated to different departments and can subsequently be included with personnel costs.

7.2.3.4 Step 4

Once the allocation by department, by line function and by type of business has been completed the type of expense can be considered further. The expense can be expressed as

1. a percentage of premium
2. a fixed amount per policy
3. a percentage of sum insured or
4. a combination of the above.

Bear in mind that some expenses are direct and some are not. As a result, fixed expenses should be split further between the direct costs which will increase linearly with an increase in volume and the indirect expenses which will reduce with an increase in volume.

The aim is still to obtain as accurate and equitable an allocation of expenses as possible. Therefore the costs should not be adjusted to accommodate marketing preferences e.g. small policies will have a higher relative cost than larger policies.

7.2.3.5 Step 5

Once this exercise has been completed it is possible to model the expected expenses per line of business and for a certain volume of business. It will also be possible to test the impact of a change in volume and a change in average policy size on the overall profitability of the company.

It is important to apply consistency checks to ensure that the model used to project expenses actually projects current experience

7.2.4 Other risks

Some of the other risks that can be considered are risks involving

1. a change in tax structures,
2. management risk,
3. risk of failure of the owners of capital,
4. political risks,
5. changes in technology,
6. demographic changes,
7. social changes and
8. catastrophe risks.

None of these risks can be easily modelled using stochastic techniques and here a more robust approach will be required. This can be making a rough estimate as to the overall contingency margin that should be allowed for these eventualities. Due to the uncertainty other risks are considered to fall outside the scope of this study.

7.3 Summary

This unit contains a discussion regarding other areas of risk. Not all areas of risk were discussed, as the list provided can never be fully comprehensive. The following issues were addressed:

1. The issues pertaining to credit risk were considered and an approach was outlined whereby credit risk can be assessed.
2. Growth risk was considered and the importance of reinsurance was mentioned for the event of growth risk.
3. An expense analysis was considered and the methodology explained. The expense analysis is a crucial component in any premium rating exercise and enables the company to identify the risk in over or under charging because of an inaccurate allocation of expenses to individual policies.

Other areas of risk were mentioned.

The identification of risk is finalised in the next chapter through a consideration of corporate governance.

8. Corporate governance

8.1 Introduction

This study has considered an appropriate structure to consider risks in a general insurance environment and have paid attention to the quantification of some of the risks. This entire process needs to be supported actively by management in a sound businesses environment. Corporate governance is therefore a critical foundation for proper risk evaluation techniques.

In the banking field of risk management, this risk is evaluated through the evaluation of operational risk. Due to the nature of the risk, a numeric approach is more difficult to apply. A points scoring approach is possible, but should be handled with extreme caution due to possible model risk. An alternative approach to quantifying this risk would be to ensure that all aspects of corporate governance are adhered to.

Without corporate governance, internal control and internal audit, the risk management described in this study would not be sustainable. A stable framework is required wherein the ideas considered so far can be implemented.

The COMMONWEALTH ASSOCIATION FOR CORPORATE GOVERNANCE (1999) cites the following quotes:

□The proper governance of companies will become as crucial to the world economy as the proper governing of countries."

James D. Wolfensohn, President of the World Bank c. 1999

"Capacity should be established in all Commonwealth countries to create or reinforce institutions to promote best practice in corporate governance; in particular, codes of good practice establishing standards of behaviour in the public and private sector should be agreed to secure greater transparency, and to reduce corruption."

Commonwealth Business Forum Resolution, October 1997, endorsed by the Edinburgh Commonwealth Economic Declaration

"Corporate Governance is the system by which companies are directed and controlled."

Cadbury Report c. 1992

The TREADWAY Report (1987) first mentioned the importance of corporate governance in the USA. The CADBURY report (1992) published a "Code of Best Practice" which would lay the foundation for further development in the guidelines for corporate governance. The GREENBURY report (1995) published a report on good practice in determining directors' remuneration.

The HAMPEL Committee was commissioned to evaluate all guidance provided and consolidate this into a "Combined Code". This committee was tasked to

1. Review the Cadbury code to ensure the original purpose was being achieved.
2. Review the role of directors
3. Review the Greenbury recommendations
4. Address the roles of shareholders and auditors in corporate governance.

The TURNBULL report (1999) was issued to give guidance on the implementation of this consolidated code of recommended practice following a request by the Institute of Chartered Accountants in England and Wales and the London Stock Exchange.

The codes were evaluated by the CACG together with codes established in South Africa, Australia, New Zealand and the USA.

8.2 The CACG guidelines for corporate governance

The board should:

Principle 1 – exercise leadership, enterprise, integrity and judgment in directing the corporation so as to achieve continuing prosperity for the corporation and to act in the best interest of the business enterprise in a manner based on transparency, accountability and responsibility.

Principle 2 – ensure that through a managed and effective process board appointments are made that provide a mix of proficient directors, each of whom is able to add value and to bring independent judgment to bear on the decision-making process.

Principle 3 – determine the corporation's purpose and values, determine the strategy to achieve its purpose and to implement its values in order to ensure that it survives and thrives, and ensure that procedures and practices are in place that protect the corporation's assets and reputation.

Principle 4 – monitor and evaluate the implementation of strategies, policies, management performance criteria and business plans.

Principle 5 – ensure that the corporation complies with all relevant laws, regulations and codes of best business practice.

Principle 6 – ensure that the corporation communicates with shareholders and other stakeholders effectively.

Principle 7 – serve the legitimate interests of the shareholders of the corporation and account to them fully.

Principle 8 – identify the corporation's internal and external stakeholders and agree a policy, or policies, determining how the corporation should relate to them.

Principle 9 – ensure that no one person or a block of persons has unfettered power and that there is an appropriate balance of power and authority on the board which is, *inter alia*, usually reflected by separating the roles of the chief executive officer and Chairman, and by having a balance between executive and non-executive directors.

Principle 10 – regularly review processes and procedures to ensure the effectiveness of its internal systems of control, so that its decision-making capability and the accuracy of its reporting and financial results are maintained at a high level at all times.

Principle 11 – regularly assess its performance and effectiveness as a whole, and that of the individual directors, including the chief executive officer.

Principle 12 – appoint the chief executive officer and at least participate in the appointment of senior management, ensure the motivation and protection of intellectual capital intrinsic to the corporation, ensure that there is adequate training in the corporation for management and employees, and a succession plan for senior management.

Principle 13 – ensure that all technology and systems used in the corporation are adequate to properly run the business and for it to remain a meaningful competitor.

Principle 14 – identify key risk areas and key performance indicators of the business enterprise and monitor these factors.

Principle 15 – ensure annually that the corporation will continue as a going concern for its next fiscal year.

8.2.1 An evaluation of the risk evaluation techniques according to the CACG guidelines.

The purpose of this section is to show how the techniques considered thus far tie in with the guidelines set by the COMMONWEALTH ASSOCIATION FOR CORPORATE GOVERNANCE.

Principles 1 and 2 are not addressed.

Principles 3, and 4 are addressed in detail. The structure explained thus far will allow the board to evaluate the appropriateness of certain strategies. In the next chapter the optimisation of the return on capital will be considered and it will be shown how the framework considered assists the board in setting objectives and monitoring those objectives for a general insurance company.

Principle 5 will be adhered to by compliance committees and the structure considered is based on the assumption that the company does comply with all laws.

Principle 6 pertains to communication. Understanding the problems at hand is often difficult. The methodology explained in this study provides a base for quantifying certain problems i.e. quantifying the impact of a risk or problem on the solvency of profitability of the company. Once an understanding of the problem has been laid down by the board, the board will be in a better situation to communicate to all the relevant stakeholders on the problems and their solutions.

Principles 7 to 9 are not addressed in this study.

Principle 10 is concerned with the regular monitoring of procedures and processes. The structure explained in this study is not only relevant for the evaluation of risk at a certain point in time but also for the evaluation of risk over time. The ability to generate expected experience over time allows the company to monitor its own state of health on a regular basis. This provides for a very powerful framework to identify any problem at any point in time and not only with substantial hindsight. The algorithm of COUTS and DEVITT (1997) explained earlier provides an insight into the approach that would be used. Such techniques are often referred to as model office simulations.

Principle 11 is concerned with the monitoring of the performance and effectiveness of the operation. This is similar to principle 10 and therefore this study is also relevant in this instance.

The study does not address principles 12 and 13.

At its core this study meets the full requirements of principle 14. The entire structure set out in this study will be appropriate to identify risk factors and key performance measures, to develop solutions for problems identified through these and to monitor the experience following the implementation of these solutions. This was also explained in the introduction to this study.

Principle 15 will follow naturally after the implementation of the aforementioned principles.

8.3 Other guidance

In South Africa the KING report on corporate governance was issued to assist companies with governing themselves through the board of directors. The KING report is currently (2000) subject to revision and currently contains reference to the following:

1. Significance for planning
2. Background to corporate governance
3. The board of directors
4. Remuneration
5. Nominating committee
6. Stakeholder communications
7. Auditing
8. Internal controls
9. Audit committee
10. Interim report
11. Going concern
12. Illegal acts
13. Internal auditors
14. The company links
15. Affirmative action
16. Ethics

There exists considerable overlap between the principles set out by CACG and the KING report.

Other guidance includes[99]:

1. Australian Investment Manager's Association
2. Stock Exchange of Hong Kong ("SEHK")
3. Summaries of Selected International Codes
4. Business Roundtable
5. GM Board of Directors
6. Report of the NACD Blue Ribbon Commission on Director Professionalism
7. TIAA-CREF Policy Statement

8.4 Enterprise wide risk management

SOUGH (2001) argues that corporate governance, risk management, internal control and internal audit are all part and parcel of enterprise wide risk management.

In order to implement the techniques outlined in this study, this collective thinking pertaining to risk management must be borne in mind.

Recognising the interrelationship between all these aspects leaves some room for confusion. This author suggests that internal audit and corporate governance be viewed as risk mitigating techniques for operational risk. By ensuring regular

[99] All these guidance notes are available on the CACG website.

internal audit, verified by external audit through the adherence to the guidelines set above, it is possible to manage operational risk. The management does not place emphasis on the quantification of these risks but rather ensures that operational risks are negligible and therefore do not need quantification.

HYLTON (2001) argues for the establishment of an independent risk management function solely focused on the implementation of the ideas discussed in this study.

8.5 Summary

This chapter considered the importance of corporate governance in enterprise wide risk management. The link between corporate governance and the ideas set out in this study is considered.

The extent to which the techniques in this study support the principles of corporate governance as outlined by the COMMONWEALTH ASSOCIATION FOR CORPORATE GOVERNANCE is considered and the importance of these techniques stated.

In the next chapter all the ideas set out thus far are consolidated into a framework suitable for implementation as an enterprise wide risk management information system.

9. Enterprise wide risk management

9.1 Introduction

This dissertation has followed the following train of thought:

1. An overview of risk areas, risk evaluation and the uses of risk evaluation were considered in the introduction.
2. The objective was set out two-fold. The first was to consider a framework whereby risk could be evaluated for a general insurance company and the second was to consider how this framework would enable practitioners to optimise the requirements set to them.
3. The framework was discussed based on the income statement and balance sheet of a company. Each item depends on certain factors and correlations exist between these factors.
4. In order to allocate resources in an appropriate manner the most important areas of risk and uncertainty were subsequently considered based on empirical studies. The main areas identified were reserving and pricing risk, asset risk, growth risk and other risks pertaining to management decisions.
5. These risks were then considered in turn. The first two were discussed under the quantification of the liability area of risk. Asset risk quantification was subsequently considered. The importance of a model office was also mentioned whereby all influences can be modelled concurrently. This will enable the practitioner to set limits on the level of growth in business. Expense risk was also considered and the importance of proper quantification highlighted. In order to address the problems that may arise from poor management the issue of corporate governance and the guidelines currently set were considered. The difficulty in trying to quantify this risk was also mentioned. Arguably such quantification is possible through scoring techniques based on corporate governance guidelines.
6. Throughout the dissertation mention was made of a model office and the integration of all these ideas. The model office provides the framework which has been set as an objective at the outset. The optimisation of goals using this framework is now considered in this chapter.

In the introduction, mention was made of different types of business strategies namely:

1. Business mix strategies
2. Managerial strategies
3. Distribution channel strategies
4. Capital allocation strategies
5. Investment strategies
6. Reinsurance strategies

The framework of model office is intended to test these strategies and then to evaluate the outcomes for the objectives set by all the relevant stakeholders. The business strategy to be investigated first will be the current one. This will indicate the solvency of the company as well as the distribution of the expected return for the existing business structure. These results can then be compared against benchmarks or objectives set.

One such an objective is the use of modelled cash flows in order to determine an optimal usage of capital. These ideas are not new to the actuarial profession in South Africa. In the past M.S.Claassen and P.P.Huber (1992) made mention of the use of asset-liability models and these ideas were further discussed by P.W.Ennis and R.E.Dorrington (1994).

N.A.S.Kruger and C.J.Franken (1997) addressed the issue of the probability that the margins on assumptions used in the financial reports of life insurance companies are sufficient. To extend on this idea the concept of probability of ruin is discussed in this dissertation and a suggestion is made to move from a prudent valuation basis to a basis whereby the valuator is, say, 99% confident that reserves will be adequate given the assumptions.

9.2 Capital, business structure and the probability of ruin

All work on risk evaluation is based on three interacting areas:

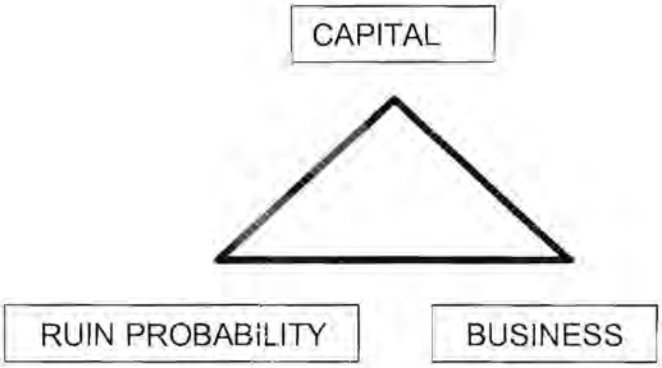


Figure 13: Relation between capital, business and the probability of ruin

The capital block represents the funds of the insurance company. The higher the level of funds available, the lower the probability of ruin will become but at the same time the higher the shareholders' requirements will become for return on capital because of greater amounts invested. As a result, to maintain the same probability of ruin when capital is increased, more business will need to be written which must yield the marginal return required in order to meet shareholders' requirements.

The ruin probability is a measure of risk. It can be limited to a specified probability for example the probability that liabilities will exceed assets in the next year must be less than 2%. This is similar to the value at risk measures commonly found in the banking environment. Alternatively the probability can be associated with a monetary limit as well. For example the probability that liabilities exceed assets by more than 10% and less than 50% must be less than 1%. Alternatively the distribution of the excess of liabilities over assets can be considered. This is similar to the concept of shareholder shortfall found in the banking environment.

The business represents the current business operation, level of business sold and level of profitability. For an increase in business sold given a fixed level of capital allocated to the business environment, the ruin probability will be increased.

Therefore, any change in any one of the elements of the triangle in Figure 13 will mean that at least one of the other elements will change. For example an increase or decrease in capital, given that the business element stays the same will decrease or increase the probability of ruin. Alternatively to decrease the probability of ruin but maintaining the same level of return on capital, more profitable business or more business which is as profitable as the current business must be sold with the available resources.

It is important to note the impact of statutory reserving requirements at this stage. Statutory requirements imposed by regulators aim to protect policyholders and shareholders alike. The requirements also limit the potential increase in risk and as a result *limit the potential return on capital*. The insurance company will not be able to increase risk to enhance return beyond a specific maximum implicitly implied by the statutory requirements.

This framework has drawn the attention of many parties involved in the regulation of the financial services industry. The BASEL II accord for the regulation of banking institutions propose that regulation of these institutions be based on the internal risk management models developed by the bank. This will overcome the problems currently faced by regulators as a result of inappropriate regulatory standards being forced upon the banks as well as the arbitrage applied by banks to overcome the regulatory requirements.

Currently regulators base their requirements on empirical studies for the entire industry, this is not appropriate due to the fact that not one of the banks will have a risk profile similar to that of the entire market. As a result banks may decide to change their business structure to mitigate the penal effect of such regulatory requirements. For example, if the regulatory requirement is a fixed percentage of premium then companies will sell non-proportional insurance covers instead of proportional insurance covers.

This is also the framework that can support the objective mentioned by Alan Greenspan in October 1999 mentioned in the abstract of this dissertation.

9.3 Applying the framework

The framework proposed in chapters 2 and 3 has to be set up. Applying this framework it will be possible to obtain the distribution of the expected earnings based on the current business structure for the next year. This distribution can be compared to the capital base of the company to determine the distribution of the return on capital.

As explained in chapter 5, the capital base of the company also has a distribution, this is as a result of the expected fluctuations in both asset and liability values. In order to ensure that the results are sensible, it is important to evaluate the valuation bases used to set the capital value.

9.3.1 Optimising the reserving policy of the company

Certain minimum reserving requirements may already be laid down either through regulations or through accounting principles. The reserving basis will tend to be either conservative or optimistic. A conservative basis is one where the expected run-off of business is set worse than what would be expected on a best estimate basis. That is to say greater amounts are set aside to meet the outstanding liabilities than what would be considered necessary on a best estimate basis i.e. there is more than a 50% chance that there will be funds available when all the liabilities have been met. An optimistic basis is the reverse of a conservative basis.

When the reserving policy is very conservative the probability of ruin will generally reduce and where the reserving policy is optimistic the reverse will generally hold true. If shareholders realise that part of their capital is tied up in liabilities, the insurance company will therefore wish to hold as little reserve as possible in order to reduce the cost of capital and return profits to the providers of capital as soon as possible. This action will, however, lead to situations where additional capital will be required from shareholders and hence the reserving policy will be optimised by comparing the effect of a change in policy to the subsequent expected rate of return and measure of risk.

This author is of the opinion, and it is a subjective opinion, that shareholders have absolutely no idea whether or not reserves are set at conservative or optimistic bases and as a result do not understand how much capital is tied or may be required in addition. This obviously allows companies to present incorrect results to their shareholders through the overstating of liabilities. If shareholders consider these to be appropriate on a best estimate basis and they are, in fact, conservative then the apparent capital base according to the shareholders is less than what is actually the case and the return that can be secured on this base is higher than what would otherwise be the case. In short, shareholders are living in the illusion that they are obtaining returns on their capital higher than is actually the case.

In order to understand the true capital base, a sensible starting point is the embedded value[100] of the company.

9.3.2 Calculating the embedded value

The embedded value is the expected value of the company on all business written to date. It consists of the following:

The assets of the company (this will be the present value at which the assets can be realised)
less the liabilities of the company

This equals the net asset value of the company

plus any profit on business written to date that has not yet been recognised (margins in reserves) (for general insurance business no renewal of business is expected)
plus any unrecognised balances such as outstanding premium balances

This equals the embedded value of the company

It is important to recognise the possible sources of fluctuation in the embedded value. In particular, great care should be taken in the evaluation of the assets of the company if the basis of valuation is not consistent with the basis used to value liabilities. The reason is simply that if asset values are depressed, say after a fall in the market, then the net asset value might be unrealistically low when it can be reasonably be expected that the market will recover in the foreseeable future. Evaluating assets based on a projected cash flow basis might actually be more suitable in such situations.

The embedded value can be calculated at future durations as well. Using the models of claim and expense outgo and asset income as described, it is possible to calculate the probability distribution of the embedded value of the company.

This goes a long way from using only one set of assumptions in order to calculate the value of an insurance company.

To optimise the reserving policy, the only element of the model that is then changed is the reserving policy and the effect thereof is determined by comparing the probability density function of the embedded value before and

[100] Many practitioners in the market consider the use of the term "Embedded value" to be one only applicable to the life insurance industry. If all bases applied by a general insurance company are best estimate bases i.e. there is a 50% probability of experience being either higher or lower, then this conviction is a correct one. As mentioned earlier, this author is not convinced that this is the case and as a result the concept of an embedded value as described above is also relevant to the general insurance industry.

after the change in the reserving policy. It is important to recognise that different types of investigations will yield different results. For example, a requirement to maintain a minimum level of embedded value over time will yield a different required reserving policy to a requirement to have a minimum embedded value now. This is because the first requirement is subject to more variation.

Another factor to take into account when optimising the reserving policy is the tax requirements of the government. By increasing the reserves at any point in time, this allows for the deferment of tax and subsequently increasing the return on capital. This situation arises because the company is able to secure an income on the funds which would otherwise have been paid to the government as tax.

If the insurance company can invest its reserves in similar investments as that which the investor would otherwise invest in, the investor will prefer to retain the funds in the company as reserves and defer the payment of tax. If the investor, however, has a higher investment requirement, the investor will be prepared to sacrifice some of the return as tax in order to secure the other higher yielding investment opportunities.

When calculating the embedded value it is therefore important to allow for the payment of tax as accurately as possible. The present value of future payments is determined at the risk discount rate. This is the rate of return required by the shareholders. Depending on the shareholders' requirement either a very conservative or very optimistic reserving policy may be suitable.

Using the framework, it is therefore possible to determine the distribution of the return on capital.

9.4 Setting a benchmark

It is important to start this section by evaluating the primary assumptions for the optimal utilisation of capital:

1. Markets are efficient. In particular, the price of capital is not influenced by the actions of any one party in the market.
2. Providers of capital are rational. This means that for the same level of risk a shareholder will prefer the investment with the highest level of return.
3. Market players have full knowledge of the level of return versus risk (risk is assumed to be the variability in return)[101] available in the market and these are available for all market instruments.
4. A specific set of assumptions regarding the business operation equates to an expected return on capital and this can be approximated through the techniques mentioned in this dissertation.

[101] From the variability in return the distribution of the downside risk can be assessed and this is, in essence, more important than total variability *per se*.

These assumptions do not hold in all circumstances. This chapter, however, relies on these assumptions in drawing conclusions on the proper utilisation of capital.
Mention has already been made that in the current economic environment, it is unlikely that assumption 3 above is true. It has to be noted, however, that securities and exchange bourses around the world are driving the speedy delivery of information and the International Accounting Standards Committee and others are driving the process of full disclosure to allow investors to make better decisions. The framework proposed will also assist in this regard.

Assuming the assumptions do hold, then every investment opportunity can be expressed in terms of its risk and return. Portfolio theory dictates that the risk and return on any one investment is irrelevant because the risk will be mitigated through diversification. This can only be done if the risk and return of each element within the portfolio is known as well as the correlation between the returns on the different investments within the portfolio.

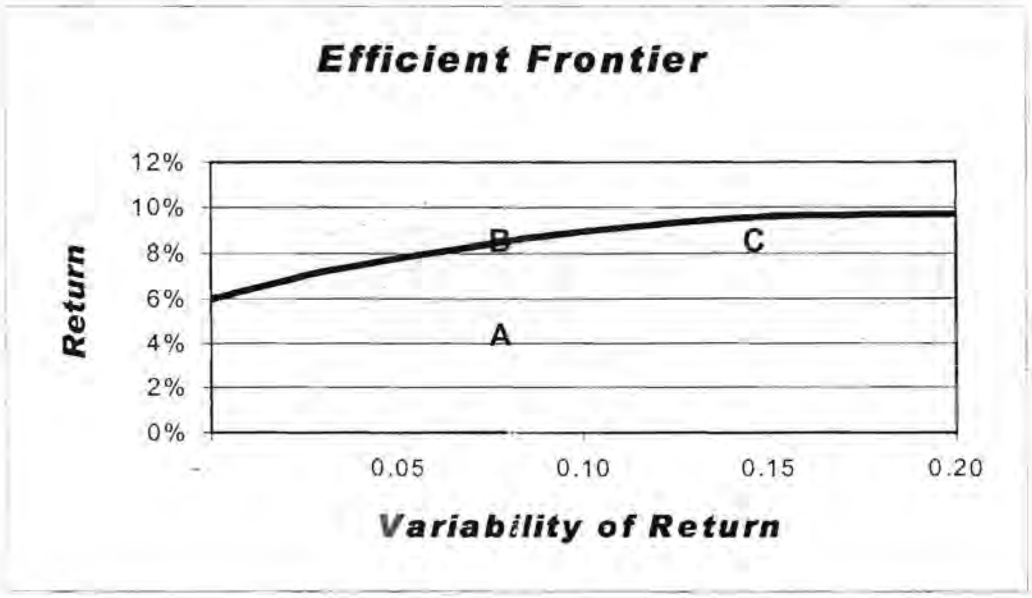
The level of correlation will be determined by the Beta of every investment. That is the correlation of the investment's return to a market index of returns. Investment managers often publish the Beta of shares based on historic experience. This goes beyond the purpose of this dissertation. It suffices to say that if the risk and return of the company are calculated, a portfolio manager will need to consider the Beta of the company and then consolidate the entire portfolio to determine the risk and return of the entire portfolio.

Having said that, portfolio managers are often adamant that the risk and return of an individual investment should be comparable to that of a diversified portfolio, arguing that the investment or business has sufficient diversification within. If this is true, then the risk and return obtained for the company, derived from this proposed framework could be compared to the efficient frontier.

The efficient frontier was introduced in the introduction. It provides the combination of risk versus return for all optimal combinations of investment portfolios within a market. That is to say, based on an investor's appetite for risk, the efficient frontier will indicate the expected return.

Whether or not investors understand the concept of risk is an entirely different matter. It is the author's opinion, following the burst of the e-bubble, that the majority of investors did not understand this concept.

Figure 14: An example of the efficient frontier



Given the investigations conducted using the proposed framework, the company may arrive at the conclusion that the moments of the company's return on equity can be plotted as A on the graph above.

This would indicate that the company is currently not performing optimally. The objective is to change the business process to ensure that the company moves to point B. It is important to note, that if the shareholders were to sell the company and invest in a fully diversified portfolio, they would be able to achieve B. As a result it is reasonable to expect this from the company if the company can also be considered as a fully diversified portfolio. This is never the case and as a result it may be reasonable to set the objective at point C. This would mean that with diversification the risk would reduce to B and the return would still be at the level required.

Depending on the shareholder, the focus might be increasing the return or limiting the risk or a combination of these. Note that there are external constraints as well, for example it is not possible to increase the risk indefinitely. Regulatory requirements may implicitly limit the level of risk that a company can take on. Alternatively, using this framework, the regulator may not accept a level of risk higher than a certain level. The level of risk will, however, not be measured on the variability in return but rather on the probability of insolvency. Note that this can be obtained from the variability in return. The same framework is used to accommodate both types of questions.

It is important to note that the risk measure can change and that the framework will be able to accommodate this. For example, the risk measure might be the probability of not paying a dividend.

Alternatively, the rating scale mentioned in the introduction can be used. Recall that rating agencies provide ratings for listed companies. These ratings can be roughly translated into a probability of ruin in the forthcoming year. This was also elaborated upon in chapter 7 where credit risk was discussed. Using the framework, it is possible to calculate the probability of ruin. This can also be used for comparative purposes using the efficient frontier.

Furthermore, the efficient frontier provides the moments of the return over an indefinite time horizon. When comparing actual returns to expected, this will be allowed for by investigating the confidence interval applicable to the efficient frontier. This will provide an upper and lower bound of return given the level of risk and will appear as an expanding funnel expanding with the increase in risk.

Given the objective, it is possible for management to start testing and subsequently implementing suitable business strategies to achieve the objectives set.

The manner in which this is achieved can be any one of a variety of options:
1. More (profitable) business can be sold with the existing capital base.
2. The existing capital base can be used to purchase additional businesses.
3. The pricing of products can be adjusted to be more representative of the cost of supporting that business.
4. The reserving policy can be changed. This will change the level of risk as well as the level of return. The deferral of the payment of tax must be considered here as well.
5. Those products which are more profitable can be identified through the techniques discussed and the insurance company's market share of those products can be increased.
6. Where problem areas are identified, such as poor claim management, management strategies can be put in place to improve the audit and identify problem areas earlier. For example, the risk rating techniques discussed in chapter 5 can be used to identify fraudulent claims.
7. Similarly, comparison of actual to expected premium will indicate where premium collection might be a problem and this can also be rectified through management strategies.
8. The reinsurance structure can be changed to transfer business with too high levels of risk.
9. The investment policy can be changed to match liability outgo better or to enhance returns further.

9.5 Optimising the pricing of products

Optimising the pricing of products entails calculating as accurately as possible the cost of producing the product and then comparing the cost to the market rates and selling those products which are most profitable.

In order to accurately price production cost the following has to be considered:
1. The cost of claims or risk.
2. The cost of expenses.
3. The cost of capital.
4. Investment returns.

The cost of claims risk was discussed in chapter 5 and the cost of expense risk was discussed in chapter 7. It is possible to allow for investment return, if any, or alternatively to allow for the cost of borrowing funds until the premium is received. This cost of borrowing will be at the same rate as the cost of capital if the funds have to be obtained from shareholders. In addition, the profit criteria of shareholders have to be met. It is therefore necessary to allocate capital to the different classes of business in such a manner that it will result in the equitable rating of the different products. There are different approaches possible.

Various methods have been suggested whereby capital can be adjusted according to risk. Insurance Research Letter 5/96 makes mention that opinions on the cost of capital differ. Some suggestions for splitting capital are mentioned below.

Allowance for investment income has to be taken into account. This will be done by modelling all cash flows. If this is not done the variability for long tail classes of business may be over-estimated and, as a result, the charge for capital provided will be too high. Some practitioners may, however, regard such margins as prudent[102].

It has to be realised that some of the techniques are more appropriate for pricing while others lend themselves to assessing solvency.

9.5.1 Allocating capital by risk premium

Capital can be defined to be total funds allocated to shareholders, the embedded value explained earlier. Shareholders' funds will require a certain return, which has been set in advance. These funds can then be split by the risk premium of business sold. A return on capital for premium rating purposes can then be set according to this allocation. The allocation of capital is dependent on the manner in which the risk premium is calculated[103].

Note that many regulatory capital requirements are based on premium. These capital requirements are based on the assumption that the higher the premium

[102] The aim of this paper is to suggest ideas whereby as many factors as possible are modelled as explicitly as possible thereby eliminating the need for a subjective opinion of prudence and rather putting a figure on how prudent the results are e.g. 9 in 10.

[103] As risk premiums are often based on expected average experience, adequate allowance for variability in experience is not made. As a result, this method is considered to be inappropriate.

the riskier the product. This is generally not the case. It is, however, important to realise that a company with capital close to the statutory minimum requirement will require to meet a cost of capital equivalent to service the statutory requirement as there is no cross subsidisation possible with existing capital balances. This illustrates the advantage of economies of scale.

9.5.2 Allocating capital by statutory requirement

Many practitioners base the charge for capital, for pricing purposes, on the statutory capital requirement for that class of business. In his presentation "Risk Based Capital: Used as an Internal Management Tool" (1996), A. Hitchcox suggests the use of the NAIC RBC rules to allocate capital to different classes of business as this can be viewed as the minimum capital requirement for US domiciled insurers.

The US has the most detailed set of statutory capital requirements. Many statutory requirements can be seen as risk adjusted capital requirements imposed by the regulatory authorities. The aim is to ensure that sufficient capital is held to protect the interests of policyholders. Risk adjusted capital provides an added advantage in that the probability of ruin or the probability of the company not being able to meet its liabilities can be quantified for the specific company. This poses a considerable advantage as opposed to applying fixed rules because allowance is made for the risk specific nature of each individual enterprise.

This method is practical but the danger is that the NAIC RBC rules are fixed and therefore do not adequately allow for an assessment of the risks at hand.

9.5.3 Allocating capital by variability per line of business

Mention has been made of the relationship between the capital available, the probability of ruin and the premium rate given a certain requirement for return on capital. When the purpose of the investigation is to determine appropriate charges for capital, the probability of ruin and the capital available will be set in advance and consequently the premium rate will be determined.

The capital to be allocated by line of business will consist of all shareholders' funds. The reserving policy adopted will have a direct impact on the shareholders' funds in that the accumulation of net retained profit is delayed. The practitioner therefore needs to decide on an appropriate valuation approach for both assets and liabilities consistent with the requirements of shareholders.

Once the capital requirement for the company as a whole has been set, the capital can then be allocated per line of business according to the variability in each line.

In order to calculate the variability per line of business, the techniques discussed in chapter 5 are used. An approach may be to calculate the reserve requirement for each class of business by line of business and subsequently to allocate the capital required proportionally.

It is important to first model the entire business operation otherwise this analysis will not be possible.

The variability per line of business has not been defined. This variability can be measured using a variety of definitions for example using the reserve calculated on a line by line basis. The level of sufficiency of reserve can then be used to allocate capital. For example calculate the reserve requirement for each line of business to ensure a 95% sufficiency of reserve. Subsequently, allocate the capital required proportionally.

The criteria may be changed and will result in different capital allocations. Another example will be to allocate capital according to the expected contribution a class of business can make to a possible loss to the company for a 100-year event. For this type of investigation short tail classes may not have any capital charge due to the stability of their results.

Deciding on a proper indication of the variability of a class of business and subsequently the cost of capital is open to argument. A risk measure equivalent to the level of risk of the entire business operation can be used.

It is reasonable to use the standard deviation of experience. Evaluating the overall capital requirement will also be based on the standard deviation of experience. The overall capital requirement can then be proportionally allocated according to the variance of each class of business assuming independence between the classes of business. If the classes are correlated then the covariance must be allocated in equal proportions and subtracted from the variance.

In other words: The capital requirement is normally derived from the standard deviation which is the square root of the experience.

Std dev = $\text{Var(ALL)}^{0.5} = (\text{Var(A)} + \text{Var(B)})^{0.5}$ assuming independence. Therefore a proportional allocation can be made according to the variability in each class.

Or

Std dev = $\text{Var(ALL)}^{0.5} = (\text{Var(A)} + \text{Var(B)} - \text{Cov(A,B)})^{0.5}$ where independence does not exist. In this instance the proportional allocation can only be made once the covariance has been allocated equally to both A and B.

This exercise will indicate which classes are more capital hungry and evaluate all returns on a level playing field. This will allow for the identification of the most profitable classes of business in return on risk adjusted capital terms. Such analysis can then be used to optimise the mix of business sold.

9.5.4 Allocating capital by marginal capital allocation

Another approach applied by some practitioners is to calculate the marginal capital requirement for new business. This can be done by setting up a model office and calculating the capital requirement before a new policy is written and after a new policy is written[104]. This will lead to very low charges for capital as the company benefits from economies of scale.

9.5.5 A suitable approach

This writer is of the opinion that an equitable allocation of capital can only be made if the statutory requirements and risk characteristics of the business are both taken into account. The cost of capital will therefore consist of a statutory minimum charge together with a charge to those lines of business which are expected to affect the free reserves of the company. Any subsequent changes to this charge will reflect a marketing decision to reflect cross subsidies in the cost of capital.

Note that given recent developments in the approach to setting regulatory capital requirements i.e. the BASEL II accord which encourages internal models and approval of these models by the regulator, the approach to capital allocation will be able to change to a true risk allocation. The regulator will be concerned about the accuracy of the models used and will rely on backdating tests to be carried out on a regular basis. This is the same requirement as the monitoring of experience suggested in the framework. In addition though, the regulator may penalise the company if the models were to exhibit levels of error higher than can reasonably be expected.

9.5.6 Practical Application of Risk Adjusted Capital

9.5.6.1 Variability of experience

In order to properly manage the ideas set forth thus far, a model office based on the framework set out thus far must be set up. This requires substantial initial investment.

The model office will form the base of the calculations. In order to assess the capital required, the variability of experience will be determined. This transforms the information available on the entire business operation into knowledge that can be conveyed to management.

[104] Assuming that the expected experience which will result from writing the business can be assessed.

Knowledge on the variability of experience is a very powerful management tool as it aids the identification of problem areas and the timeous setting up of appropriate, remediable action plans.

9.5.6.2 Internal Solvency

The split of capital according to the variability inherent in each line of business has been considered. Assessing the variability of experience for each line of business means that all the areas of risk need to be appropriately modelled for each line of business.

Once the probability density function of the experience for each line of business can be modelled, the capital required to support that business could be assessed. This capital requirement will be based on the average expected future experience plus a contingency loading. The methodology discussed thus far will enable the practitioner to equate the contingency margin to a specific probability of ruin. This is a powerful result.

It is important to note that the calculation of capital requirements per line of business will over estimate the capital requirement for the whole business operation, as all classes of business are not fully dependent. When assessing internal solvency for the company it will therefore be necessary to model all processes concurrently to allow for correlations.

9.5.6.3 Probability of ruin

General insurance markets tend to follow a cycle of oscillating underwriting results. This means that practitioners will generally not be able to price products based on risk cost plus an appropriate capital charge plus all other loadings but will have to make do with the premium rates available in the market. The insurer who can beat the market cycle, inter alia by better pro-active pricing, stands to be more profitable in the long run. Using the methods of allocating capital according to risk inherent in the business sold, the relationship between capital and premium rates can be used to calculate the probability of ruin for specific premium rates. Management and underwriters can then use this information to decide on the volume of business to be sold as well as the lines of business to be marketed.

It has been found through the author's own experience that the techniques discussed are more appropriately used to test the profitability of market premiums than they are to set premium rates as profit requirements tend to be set at levels that are often not attainable in practice. The techniques set out can therefore be used for risk arbitrage[105].

[105] This means that the practitioner will use the techniques to identify areas of business which are less risky than other areas for the same level of return. By shifting the company's marketing to the less risky business arbitrage is achieved by obtaining the same level of return for a reduced level of risk.

9.5.6.4 Pricing

A number of authors have made mention of the application of risk adjusted capital in the pricing of insurance contracts. This approach is very much dependent on the market to be considered. The capital charge in pricing is based on the capital allocated to each contract as calculated by methods mentioned above.

When applying these ideas to pricing, the run-off of claims must also be considered to ensure that appropriate credit is given for investment income earned on reserves.

The method used, as described above, will be to determine the ratio of capital allocated to premium volume. The overall charge for capital can then be adjusted according to this ratio and subsequently used in the pricing of the business.

9.6 Optimising the investment policy

The previous section considered the optimising of the pricing of insurance products. Prior to that mention was also made of optimising the reserving policy. Apart from pricing and reserving risk, empirical studies cited in chapter 3 i.e. Ryan et.al. (2001) indicated that asset risk is also a great contributor to the risk inherent in the business operation.

In order to evaluate the asset risk, the business operation will be assumed to remain fixed. For the projected business, the distribution of return for a specified asset mix will be considered. Clearly the risk will depend on the extent to which assets and liabilities are matched. For short-term policies, studies may show that the investment policy has to be limited to cash to limit the variability, for example.

The approach will be to consider the different types of investment policies that are possible. For each policy a split of assets will be considered by type, term and currency. For example 20% cash, 50% fixed interest securities, 30% equities. Each of these might be split further for example the equities may be limited to certain industries. For this asset split, the expected experience will then be generated as explained in the framework. The income statements will be consolidated and the variability of return will be modelled for a certain period. A year of projection is normally appropriate as this coincides with the financial period of the company.

For different investment policies the results will be evaluated and based on this analysis an optimal investment structure can be determined based on the shareholders requirements and possible constraints that have been set initially.

An example of this approach will be considered in the appendix.

9.7 Optimising the operation risk of the company

Operational risk is difficult to quantify. The reason is that empirical studies cannot always draw a clear cause and effect line. The operational risk that gave rise to the problem is often overlooked. Instead of attempting to model this directly, a better approach would be to have a scoring technique. This technique would work as follows:

Approach management and senior management and list all the requirements of proper corporate governance as outlined in chapter 8. Ask management to indicate how important they consider the different components to be for the business operation. Rank the requirements based on the opinions of management after careful scrutiny of the results presented.

Subsequently ask management how they perceive the company actually ranks in each of the components introduced in the previous exercise. These outcomes can now be weighted according to the rankings obtained from the previous outcomes. Consideration should then be had as to what level of compliance indicates towards what level of operational risk. Some margin will be required to increase the variability of results as a result of non-compliance to these requirements. The extent of this margin can be established through similar studies across the industry and a subsequent evaluation of the credit ratings of these companies. Similarly cases of companies that went into liquidation can be considered and the rating they would have received in the year prior to going into liquidation.

It is not the purpose of this study to set this framework. The author proposes that management ensures that all the principles cited are adhered to rather than trying to assess the level of compliance. In other words comply fully which will mean that this issue is not a risk anymore.

9.8 The optimisation process

Note that the processes discussed above are iterative in nature. That is to say the optimal investment policy is only optimal while all other factors remain the same. Therefore if the business mix were to change then the investment structure might need to change as well.

In this regard it is important to note the importance of the different areas of risk. These were mentioned earlier. Based on the level of importance, i.e. those risks that have been the highest contributors to company insolvencies, the optimisation should be considered.

Therefore the pricing of business should be considered firstly. This includes the evaluation of the reserving policy. Once the reserving policy and the pricing structures have been set, the business mix can be evaluated. This will be by type of business as well as source of business.

Once the insurance operation has been evaluated, the impact of reinsurance can be considered. This will have a direct impact on the variability of expected returns and is often also viewed as a substitute for capital.

Subsequently the investment policy can be considered. In this instance the different investment policies can be considered.

Allowance will need to be made for credit risk, but this can be done on a deterministic basis. As mentioned previously, not all risks can be properly quantified and where this is the case proper management of the risks provides a suitable solution.

The results can be evaluated for every set of processes and the optimal processes identified. These processes can then be communicated to management and implemented.

9.9 Conclusion

The framework discussed in this dissertation aims to quantify all risks to which the company is exposed. Z. Khorasanee, in one of his editorials in 'The Actuary' magazine, stated that actuaries should translate their knowledge into understandable ratios for their clients. The aim of this study is to further this process by moving away from the "prudent" assumption base to the "99 in a 100" assumption base. The key to this lies in assessing variability based on the framework as discussed.

The current developments on the regulatory side in banking as well as other financial services companies provide this author with the confidence that this framework will eventually be standard practice for all financial institutions.

9.10 Summary

This chapter consolidates the ideas set out in this study. The framework discussed initially is considered again. The methodology used to set up the framework and quantify the different areas of risk and uncertainty is consolidated in the evaluation of the distribution of the return of the business operation.

This return can be expressed as a return on capital.

This return can be used to determine the probability of ruin. I.e. the return required to put the business in an insolvent position can be established as well as the likelihood as well as the distribution of all returns exceeding this return. This links to the value at risk and shareholder shortfall concepts found in the banking arena.

The distribution can then be compared to a benchmark. The efficient frontier is discussed as a benchmark and problem areas current in the market are mentioned.

The idea of applying different business strategies to optimise returns is then considered. A number of different business strategies are mentioned.

The pricing strategy of the company and the proper allocation of capital are considered in detail due to the importance of this risk.

The interlinking between these ideas for internal control and regulatory control is mentioned.

The actuarial profession is in the fortunate position that it has the necessary skills to implement these ideas. The area of application is not limited to general insurance and can be applied to other fields as well.

The opportunity exists to translate complex business structures and valuation bases to a simple probability of ruin. As realistic a basis as possible will allow the practitioner to move away from being simply "prudent" to being "99% confident" given the valuation basis used.