Correlations

This was felt to be a fundamental part of the work. Very little work has been done on this in the past although its importance had been recognised in several Institute papers.

As with both the asset and the liability models it was also recognised that there was no right answer. What was important for one project might well be unimportant for another. For many insurance operations inflation might be a factor affecting both sides of the balance sheet. However for some classes of business it might not be important such as LMX spiral business or even certain types of latent claims where the key uncertainty is a legal dispute. However it was felt important to examine all the correlations so that these could be surfaced.

The following is the sub working party report. Chairman - Richard Bulmer.

The Appendix has been supplied by Patrick Byrne.

CORRELATIONS BETWEEN ASSET VARIABLES AND LIABILITY VARIABLES

SYNOPSIS

This paper has been produced for the October 1996 General Insurance Convention in Stratford-on-Avon by a sub-group of the Stochastic Asset Liability Modelling Working Party. The members of this sub-group are:

Richard Bulmer (Chairman)
Patrick Byrne
Peter Green
Richard Rodriguez

The paper describes the correlations between asset variables and liability variables within a general insurance company and outlines a theoretical framework for the modelling of such correlations. The paper also suggests areas for further work.

CORRELATIONS BETWEEN ASSET VARIABLES AND LIABILITY VARIABLES

CONTENTS

1 INTRODUCTION

1.1 Stochastic asset liability modelling involves the application of a large number of simulations to a financial planning model of the future experience of a general insurance company. An important part of this process is to determine probability distributions which define the mean value and variability of the principal variables affecting the company. Although many of these variables are either independent or weakly correlated, some variables are strongly correlated. Any stochastic asset liability modelling exercise which ignores these correlations is likely to produce misleading results.

1.2 The purpose of this paper is to describe the correlations between asset variables and liability variables within a general insurance company, and to outline a theoretical framework for the modelling of such correlations. The paper has been produced by a sub-group of the general insurance Stochastic Asset Liability Modelling Working Party. We have not considered correlations between asset variables or correlations between liability variables, as these are being considered by other sub-groups.

1.3 As far as we are aware, very little material has been published on the subject of correlations within a general insurance company. Consequently, the paper starts from first principles and only provides an introduction to the subject. A substantial amount of further work remains to be done, and we have suggested areas where we consider future research could usefully be undertaken.

1.4 A central theme of the paper is that it is essential to adopt a pragmatic approach to the treatment of correlations and not get too bogged down in the details of the mathematics, although it is important for modelling work to have a sound theoretical basis. It is also important, when estimating correlations for a general insurance company, to ensure that the selected correlations are real features of the underlying business rather than the result of coincidence. Actuaries need to have confidence in the assumed degree of correlation, before basing any decisions on the results of an asset liability modelling exercise.

1.5 This paper will be the subject of a workshop at the 1996 General Insurance Convention.

2 PRACTICAL USES OF A CORRELATIONS ANALYSIS

2.1 An understanding of the nature and quantum of correlations is important not just within a stochastic asset liability model, but also in the context of other methodologies, such as risk-based capital and dynamic solvency testing, which seek to answer the question "How much capital do I need to run a general insurance business?" If correlations are ignored, the calculated required amount of capital is likely to be substantially different from the calculated value taking correlations into account.

2.2 However, the use of techniques such as stochastic asset liability modelling is much wider than the determination of capital requirements. Such a model can be used to understand and manage the diverse risks affecting a general insurance company. A general insurance company risk manager should be on the look-out for low, zero or negative correlations, which enable him to diversify his portfolio of risks. This might include:

- Diversifying his asset portfolio by asset class and currency.
- Writing a wider range of classes of business.
- Balancing the risks on the asset and liability sides of the balance sheet. This might involve, for example:
  - avoiding combinations which amplify risk such as a substantial investment in commercial property and writing a book of mortgage indemnity business.
  - investing in assets whose value tends to increase with inflation if the company has a particular exposure to claims inflation.

3 PRINCIPAL VARIABLES AFFECTING A GENERAL INSURANCE COMPANY

3.1 For the purposes of this paper, consideration has been limited to the variables which we consider to be the most important for a general insurance company, and these are listed below:

- Price inflation
- Earnings inflation
- Economic growth
- Economic cycle
- Taxation
- Solvency margin
- Short-term interest rates
- Bond yields
- Equity market performance
- Property market performance
- Premium rates
- Claim frequency
- Average claim amounts
- Loss reserves
- Business volumes
- Expenses
- Commissions
- Currency exposure
- Catastrophes
- Reinsurance bad debts

This list is not exhaustive and other representations are possible.

3.2 Many of the variables listed in 3.1 only apply to either the asset or the liability side of the balance sheet. However, a number of variables, in particular price inflation, earnings inflation, economic growth and economic cycle, impact both sides of the balance sheet.

3.3 Inflation is the most important link between the asset and liability sides of the balance sheet. This has some interesting implications for the choice of asset model. Some asset models do not incorporate inflation as a variable, and it would be necessary in such cases for inflation to be included before stochastic asset liability modelling work can be undertaken effectively. However, the inclusion of inflation as a stochastic variable may change the pattern of simulations produced by the asset component of the overall model.

4 DESCRIPTION OF CORRELATIONS BETWEEN VARIABLES

4.1 Members of the correlations sub-group completed the grid shown on the next page, to stimulate discussion regarding the nature of correlations between asset variables and liability variables for a general insurance company. The variables in the grid are those listed in section (3). Each combination of variables was assigned a value on a scale from -3 to +3 with:

+3 representing a strong correlation.
0 representing no correlation.
-3 representing a strong anti-correlation.

We suggest that readers of this paper complete the grid for themselves before attending the workshop organised by the correlations sub-group.

4.2 Members of the sub-group found the grid surprisingly difficult to complete. Also, the values selected by different sub-group members in many cases had a relatively low correlation coefficient! We found the discussion of these apparently conflicting views to be a valuable exercise.

General comments

4.3 The sub-group came to the following preliminary general conclusions regarding correlations between asset variables and liability variables for a general insurance company:

- Correlations have a variety of response times.
  For example:
  - A movement in variable A may cause an instantaneous movement in variable B.
  - A movement in variable A may cause a lagged movement in variable B.
  - Variables A and B may follow cycles, with cycle B lagging cycle A.

  It follows that any theoretical framework needs to be able to model these different types of responses.

	Reinsurance bad debts	Catastrophes	Currency exposure	Commissions	Expenses	Business volumes	Loss reserves	Average claim amounts	Claim frequency	Premium rates
Price inflation										
Earnings inflation										
Economic growth										
Economic cycle										
Taxation										
Solvency margin										
Short-term interest rates										
Bond yields										
Equity market performance										
Property market performance										

- Some correlations are direct (eg. inflation and interest rates) whereas other correlations are indirect in the sense that variable A is only correlated with variable D because A is related to B is related to C is related to D. We believe that in any practical correlations analysis it is important to identify and then focus on the direct correlations which really matter. It is also important to look for causality, to ensure that the selected correlations are real features of the underlying business rather than the result of coincidence.
- Although many correlations apply reasonably uniformly across most classes of business, there are also numerous examples of correlations which apply only to an individual class. For example, a decline in commercial property values may coincide with a surge in mortgage indemnity claims.
- It is important to define variables carefully. For example, "premium rates" should be defined more precisely as an index of rating levels.
- Variables have different characteristics. Some are truly stochastic whereas others, for example taxation, are simply the mathematical result of a combination of other variables.
- Some "correlations" would be dealt with most efficiently by decision rules within the asset liability model.

Specific comments

4.4 The following paragraphs contain a general discussion of correlations between asset variables and liability variables for a general insurance company. Each liability variable is discussed in turn regarding its correlation with each asset variable. Paragraphs 4.5 to 4.30 are summarised in the table in 4.31, for the benefit of readers who prefer to skip the discussion.

Premium rates

4.5 "Premium rates" may be defined more precisely as an index of rating levels.

4.6 Premium rates are considered to be positively correlated with price inflation and earnings inflation. It is worth mentioning, in passing, that, in our view, inflation is the strongest link between asset variables and liability variables in a stochastic asset liability model for many (but not all) classes of business. It is essential therefore that inflation should be included in any asset model used for this purpose, and it is desirable that economic growth should also be included.

4.7 Premium rates can be considered to be either positively or negatively correlated with economic growth, which we define as *real* economic growth. The alternative arguments are:

- Economic growth leads to higher insured values and hence higher premium rates.
- Economic growth leads to higher stock market values, higher insurance capital and hence (in due course) lower premium rates.

In each of these cases, significant time lags may be involved. We prefer the first, more direct argument.

4.8 We decided that taxation is not a true stochastic variable, and so we did not consider correlations between taxation and liability variables such as premium rates.

4.9 Premium rates can be considered to be either positively or negatively correlated with the actual solvency margin.

The alternative arguments are:

- A high solvency margin, which may be associated with excess capital in other companies, may lead (in due course) to lower premium rates.
- A relatively high level of premium rates is likely to lead to a higher solvency margin.

The relationship in both of these cases is likely to be lagged. In any event, we concluded that solvency margin is not a true stochastic variable. It is the mathematical result of a combination of other variables. Also, we consider that the causal link between solvency margins and premium rates may be dealt with more satisfactorily by means of a decision rule within the asset liability model rather than by using correlations.

4.10 Premium rates can be considered to be either positively or negatively correlated with short-term interest rates, bond yields and equity market performance. The alternative arguments are:

- High interest rates allow a company to reduce their premium rates to reflect the time value of money.
- High interest rates are likely to be associated with high inflation rates which, as already discussed, are correlated with higher premium rates.

We prefer the latter argument, although this correlation is dealt with most efficiently by a link between inflation and premium rates.

4.11 We consider that there is a positive correlation between premium rates and property market performance, via the associated increase in insured values.

Claim frequency

4.12 We believe there is no strong correlation between claim frequency and either price inflation and earnings inflation.

4.13 We consider that there is a negative correlation between claim frequency and (real) economic growth for many (but not all) classes of business. However, the relationship is not necessarily symmetric, and is complex in nature. Claim numbers tend to increase during a recession, but there may not be a corresponding reduction in claim frequency during a period of strong economic growth. For some classes of business, claim frequency is more closely related to the rate of change in the rate of economic growth than the rate itself.

4.14 We consider that there is a negative correlation between claim frequency and property market performance for certain classes of business such as mortgage indemnity and surveyors' professional indemnity. However, the relationship is again not symmetric.

4.15 We consider there are no other significant correlations between claim frequency and asset variables.

Average claim amounts

4.16 We consider that there is a strong correlation between average claim amounts and inflation, with the strength of this relationship varying by class of business.

4.17 We consider there is a (weaker) correlation between average claim amounts and economic growth, via higher insured values.

4.18 There is a weak correlation between average claim amounts and each of interest rates, bond yields, equity market performance and property market performance. However, this is considered to be an indirect relationship, the more direct relationship being with inflation. It may be worth mentioning, in passing, that investment market performance is often a lead indicator of trends in inflation rates.

Loss reserves

4.19 Loss reserves are defined for this purpose to be the total of future claim payments, with no allowance being made for the time value of money; rather than the loss reserves companies actually hold.

4.20 We consider that loss reserves are strongly correlated with price inflation and earnings inflation for many (but not all) classes of business.

4.21 We consider there is a negative correlation between loss reserves and property market performance for certain classes of business such as mortgage indemnity and surveyors' professional indemnity. However, the relationship is again probably not symmetric.

Business volumes

4.22 Business volumes are defined for this purpose to be exposures or the underlying real demand for insurance.

4.23 There is a strong correlation between business volumes and real economic growth. The sub-group was undecided whether there was a weak relationship with inflation. However, it was agreed that economic growth is the key driver. We also felt that the relationship between business volumes and economic growth would break down at a certain point because companies would not wish to acquire unprofitable business or to reduce solvency ratios to levels which the market would find unacceptable. This relationship may again be dealt with within the model by means of a decision rule.

4.24 We consider that business volumes are negatively correlated with short-term interest rates and bond yields, and positively correlated with equity market performance and property market performance. However, it appears again that the key driver is real economic growth.

Expenses

4.25 Expenses are considered to be strongly correlated with inflation and less strongly with economic growth.

Commissions

4.26 Rates of commission as a percentage of premiums are also considered to be positively correlated with inflation and economic growth, but to a lesser extent than expenses. The effect is likely to be relatively insignificant.

Currency exposure

4.27 Currency exposure is an important issue for many general insurance companies. Currency exposure is not considered to be primarily a stochastic variable, but is more often the consequence of management action. Such an exposure can arise both within both the balance sheet and revenue account. In practice, a separate model is likely to be established for each currency, with currency exposure being modelled at the end of the process. Some argue that currency rates are related to differential inflation rates over the long term. However, the timescale involved is likely to be longer than the normal time horizon for a general insurance asset liability model.

Catastrophes

4.28 We consider there is a positive correlation between the amount of catastrophe claims and economic growth, because economic growth leads to an increase in the frequency of man-made catastrophes and an increase in the average amount of a catastrophe, through the growth of insured values within the area affected by the catastrophe.

4.29 If the catastrophe is large enough, for example a Californian or Tokyo earthquake, there is likely to be an adverse effect on world stock markets.

Reinsurance bad debts

4.30 We consider reinsurance bad debts are positively correlated with inflation, interest rates and bond yields, and negatively correlated with solvency margins.

Summary

4.31 The discussion in paragraphs 4.5 to 4.30 may be summarised in the following table:

	Price inflation	Earnings inflation	Economic growth	Economic cycle	Taxation	Solvency margin	Short-term interest rates	Bond yields	Equity market performance	Property market performance
Premium rates	+	+	+-	+-			+-	+-	+-	+
Claim frequency			-	-						
Average claim amounts	++	++	+	+						
Loss reserves	++	++								
Business volumes			++	++			-	-	+	+
Expenses	++	+								
Commissions	+									
Currency exposure										
Catastrophes		+							-	-
Reinsurance bad debts	+	+					+	+		

++ strong positive correlation.
\+ weak positive correlation.
\- weak negative correlation.
-- strong negative correlation.
+- argument could be put forward for either positive or negative correlations.

5 PRACTICAL CONSIDERATIONS

5.1 The following paragraphs outline some of the practical considerations which apply to the treatment of correlations within a general insurance stochastic asset liability model.

5.2 We consider that it is essential to adopt a pragmatic approach to the treatment of correlations and not get too bogged down in the details of the mathematics, although it is important for modelling work to have a sound theoretical basis. A general insurance company model is likely to be complex, even before consideration of correlations. There are then potentially hundreds of cross-correlations between variables which could be specified in the model. However, the results from a model which incorporates such a large number of cross-correlations are likely to be difficult, if not impossible, to understand and interpret. In addition, such a detailed approach would, we believe, be spuriously accurate, given the uncertainty surrounding the probability distributions for many general insurance company variables.

5.3 Consequently, we consider that a correlations analysis should seek to:

- Identify and then focus on the relatively small number of correlations which really matter. This may mean concentrating on as few as 5 correlations.

  Establish the approximate magnitude of these key correlations, rather than producing precise values representing a spurious level of accuracy. We consider it may be sufficient in most cases to classify correlations into one of the following categories:

- strong positive correlation.
- weak positive correlation.
- no correlation.
- weak negative correlation.
- strong negative correlation.

A possible counter-argument to this approach is to ask what a correlation coefficient of say 0.3 to 0.8 really tells us about the link between the variables.

5.4 Although the mathematics of correlations is more tractable if variables are assumed to be normally distributed, in practice this convenient simplification is likely to be inappropriate because many general insurance variables have a skew distribution.

5.5 As already stated, when estimating correlations for a general insurance company, it is important to look for causality, to ensure that the selected correlations are real features of the underlying business rather than the result of coincidence.

5.6 It would probably be instructive to test the sensitivity of the results emerging from the asset liability model to different values of the correlations between key variables.

6 POSSIBLE AREAS FOR FURTHER WORK

6.1 As stated earlier, this paper is only an introduction to the subject of correlations between asset variables and liability variables for a general insurance company, and a large amount of further work could usefully be undertaken including:

- The development of methods of measuring correlations between variables for general insurance companies based on historic data.
- The analysis of data for individual companies and the whole market, for the purpose of estimating correlations between variables.
- The reconciliation of this analysis with the real underlying features of the company's business or the market. In other words, to establish causal as well as mathematical relationships.

6.2 We would welcome feedback on this and any other aspects of this paper at the 1996 General Insurance Convention.

The purpose of this Appendix is to take a short sprint through some of the mathematical footwork involved in obtaining the correlation structure of a multivariate set of observations.

As in the univariate case, a good place to start is with the first two moments of each of our p variables so as to derive their respective means and variances. We then need a measure of the relation between each pair of variables which we call covariances. Although mathematically useful, covariances do not lend themselves readily to useful interpretations as descriptive statistics; where a linear relationship between two variables is measured by their covariance, the sign of the covariance is easily interpreted, but the size of the coefficient is not - it is not necessarily dimensionless, being dependent on the units in which the two variables are measured.

We can standardize each of the covariances by dividing them through by the product of the standard deviations of the two relevant variables to derive the correlation coefficients.

Having derived the dispersion (variance/covariance) matrix and the sample correlation matrix, we can then test the significance of the coefficients with the help of some handy distributional assumptions.

Some Formulae:

(For brevity we will use vector notation)

The **mean vector** (for the continuous case)

$$\boldsymbol{\mu}^T = [\mu_1, \ldots, \mu_p]$$

is such that

$$\mu_i = E(X_i) = \int_{-\infty}^{\infty} x.f(x)\,dx$$

is the mean of the *ith* component of the p-dimensional random vector $\mathbf{X}$, where

$$\mathbf{X}^T = [X_1, \ldots, X_p]$$

The variance of the ith component of X is given by

$$Var(X_i) = E[(X_i - \mu_i)^2] = E(X_i^2) - \mu_i^2$$

and is usually denoted σ_{ii} in the multivariate case.

The covariance of X_i and X_j is defined by

$$Cov[X_i, X_j] = E[(X_i - \mu_i)(X_j - \mu_j)]$$

and denoted by σ_{ij}.

For p variables there will be ½$p(p$-1) **covariances**, and these can be represented in a dispersion matrix as:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

where the *(i,j)th* element is σ_{ij}.

The leading diagonal elements are simply the variances. The correlation coefficients are now obtained by dividing the elements of the dispersion matrix Σ by the product of the standard deviations of the two variables in question. Thus

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

The correlation coefficient must lie between +1 and -1. The correlation coefficient measures the linear association between two variables. Values close to +1 (or -1) may indicate a strong positive (or negative) linear association between the two variables. If two variables are statistically independent then their covariance (and correlation) will be zero. But it is not necessarily true that if the correlation is zero, the two variables will be independent. Two variables may have zero correlation, but be highly dependent on one another (e.g. in a non-linear fashion). However, if we are entitled to make the distributional assumption that the two variables follow a bivariate normal distribution, then a zero correlation implies independence.

The correlation matrix is denoted by **P**, where **P** is

$$\mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}$$

We can link **P** and Σ by means of the diagonal matrix **D** where

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_p \end{bmatrix}$$

such that

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{DPD} \\ \text{or} \quad \mathbf{P} &= \mathbf{D}^{-1}\mathbf{\Sigma}\mathbf{D}^{-1} \end{aligned}$$

in the usual matrix notation. We also point out at this stage that $\mathbf{\Sigma}$ and $\mathbf{P}$ are both positive semidefinite matrices.

In any analysis of the data, we are unlikely to be considering anything approaching the complete population. So we are more concerned with the sample dispersion matrix **S** and the sample correlation matrix **R**. These are simply the sample versions of $\mathbf{\Sigma}$ and $\mathbf{P}$ with standard adjustments so as to derive unbiased sample variance/covariance estimates of the true population variances/covariances.

S and **R** are given by

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \cdot & \cdot & & \cdot \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \cdot & \cdot & & \cdot \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix}$$

and the sample correlation coefficient is given by

$$r_{ij} = \frac{s_{ij}}{s_i s_j}$$

Given the symmetrical nature of **R**, we could simply represent it as

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ & 1 & \dots & r_{2p} \\ & & & \cdot \\ & & & \vdots \\ & & & 1 \end{bmatrix}$$

Similarly we can define a diagonal matrix so as to relate **S** and **R**, thus

$$\tilde{\mathbf{D}} = \begin{bmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & s_p \end{bmatrix}$$

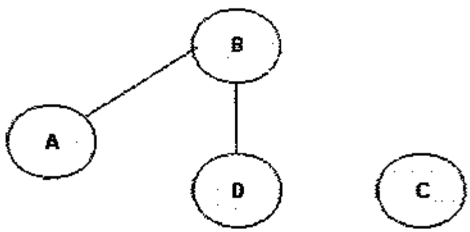
and we can calculate **R** from **S** using

$$\mathbf{R} = \tilde{\mathbf{D}}^{-1}\mathbf{S}\,\tilde{\mathbf{D}}^{-1}$$

We are now at a stage where we can interpret the sample correlation matrix. One approach is to assess how much of the sum of the total corrected sum of squares of say, y is explained by the linear regression of y on x.

Another interesting approach is to use Graphical Modelling to illustrate the correlation structure. A very brief explanation of such an approach follows.

Consider the graphical model of four variables A, B, C, D represented as



We can give the following interpretation to the above graphical model: A is statistically dependent on B (and therefore correlated) and B is statistically dependent on D. Conditional on knowing B, A is statistically dependent on D (and therefore correlated). It may be the case that not having B means that A is independent of D. C is statistically independent of A, B, D, and therefore has zero correlation with A, B, D. In the graphical model we could find ourselves at any node (A, B, C, or D) and update the conditional dependence structure given any new piece of information.