Premium Liability Correlations and Systemic Risk

October 29, 2011

Abstract

During the past decade many insurance solvency standards have introduced a requirement to use premium liabilities and diversified risk margins. With the introduction of IFRS 4, most countries' accounting standards augmented unearned premium provisions with premium liabilities. Many countries' accounting standards have also required the inclusion of diversified risk margins in premium liability provisions. Calculating diversified risk margins requires estimates, correlations and variances of the outstanding claims and premium liabilities. This paper shows how the application of a random effects model to claims payments can be used to unify the estimation of outstanding claims provisions and premium liability provisions, including the estimation of correlations between outstanding claims and premium liability. Estimators are proposed for the model parameters. While the model parameters are estimated using aggregated historical data, the model applies to individual claims, because premium liabilities are a subset of each data point in the aggregate data given in claims triangles, and thus an individual claim model is required to obtain the scaling of mean and variance.

Keywords: diversification, systemic risk, premuim liabilities, outstanding claims, random effects, AASB 1023, IFRS 4

Contents

1 Introduction

When an insurance policy is written, the insurer assumes future obligations for the period of cover. To cover such future obligations, insurance companies set aside a provision for future claims expected to occur during the remaining period of insurance. The provision is called the “premium liability" or “unexpired risk." Accounting and regulatory standards usually require unearned premiums, a means of accruing premium income during the contract period. Over time, the insurance company earns the premium and is subject to claims. At the end of the period of insurance the premium liability is extinguished. Premium liabilities do not include provisions for claims which have already occurred. Insurers establish outstanding claims provisions for claims which have already occurred but have not yet been fully paid.

As countries adopt accounting standard IFRS 4 (International Accounting Standards Board, 2008), insurers need to apply a liability adequacy test. This test compares two different estimates of future claims, the unearned premium and the premium liability. Unearned premiums are used for accruing premium across each policy period, and are a function of the insurance premium. Common practice is to earn an insurance premium uniformly during the contract period, resulting in the pro-rata unearned premium u, defined as

$$u = \frac{p^{(I)} \times d_2}{d_1 + d_2}, \tag{1}$$

where:

u is the unearned premium for an insurance policy;

$p^{(I)}$ is the total premium for all policies that have yet to expire;

d_1 is the number of days that the contract has been on risk;

d_2 is the number of days remaining until the contract expires.

Premium liabilities are an estimate of the cost of future claims for insurance policies which have already been written. Thus premium liabilities are prospective estimates, and are not a function of the insurance premium. Insurers apply the liability adequacy test by comparing the premium liability to the unearned premium. When the premium liability exceeds the unearned premium, the insurer must set up a premium deficiency provision in the accounts, causing the profit and loss statement to recognise the shortfall immediately. Australia’s accounting requirements are contained in AASB1023 (Australian Accounting Standards Board, 2007), which contains a liability adequacy test requirement, consistent with IFRS 4.

At the same time, insurance regulators are moving towards the use of premium liabilities to assess the solvency of insurers. In Australia, regulatory requirements for liability provisioning are set out in GPS310 (Australian Prudential Regulatory Authority, 2006), which requires an insurer to assess a premium liability using the expected value of the discounted cash flows plus a risk margin. The total of the expected discounted value and risk margin brings the premium liability to a 75% level of adequacy after allowance for diversification benefits.

Several papers have considered the standard error of outstanding claims liability estimates, including Mack (1993). Braun (2004) extends Mack's approach to apply to a "portfolio of several correlated runoff triangles". Others such as Brehm (2002), Merz and Wuthrich (2009), and Hess et al. (2006) develop models for dependencies within sets of runoff triangles. However, premium liabilities are not discussed. Ohlsson and Lauzeningks (2009) discusses the concepts required to calculate risk margins under Solvency II and IFRS 4 Phase II, including the unexpired risk or premium liability, and discusses the need to allow for diversification with outstanding claims. However Ohlsson's paper does not discuss how diversification allowances may be calculated in practice.

Diversified risk margins cannot be estimated without standard error and correlation estimates for both the outstanding claims and premium liabilities. This paper extends outstanding claims estimation methods to consider the different effects of systemic and non-systemic risk, and the consequences for premium liability estimation. Estimation techniques are proposed for the standard error of the premium liability estimate, and for the correlation of premium liability estimates with outstanding claims estimates.

The proposed estimation techniques are based on the use of a random effects model for claims. The model for individual claim severities can be described as

$$c_{ijk} = \mu + \alpha \cdot \nu_i + \beta \cdot \epsilon_{ijk}, \tag{2}$$

where:

c_{ijk} is a claim payment for a specific claim;

i is the accident period;

j is the development period;

k is an index for the individual claim;

μ is the average claim size;

ν_i is an error term, with unit variance, applying to all claims within accident period i;

ϵ_{ijk} is an error term, with unit variance, specific to individual claims.

Let n accident periods exist in the historical data. The outstanding claims liability consists of future payments belonging to claims from accident periods

$i = 1, 2, ..., n$. Future payments for accident period i come from development periods $j = i+1, i+2, ..., n$. Claim payments for accident period i and development period j come from claims indexed $k = 1, 2, ..., l_{ij}$. All claims from those combinations of accident and development periods are included within the outstanding claims. The outstanding claims amount o is defined as

$$o = \sum_{i=1}^{n} \sum_{j=n-i+1}^{n} \sum_{k=1}^{l_{ij}} c_{ijk}.$$

The value of o is estimated using the method proposed by Mack (1993) to obtain the estimate $\hat{o}$ based on the observed payments. The consistency of Mack's estimate with (2) is considered later in this paper.

The next accident period (accident period $i = n+1$) has an incurred amount c_{n+1} defined as

$$c_{n+1} = \sum_{j} \sum_{k} c_{n+1,jk}. \tag{3}$$

The estimator for incurred claims in the next accident period (denoted by $\hat{c}_{n+1}$) is based on the sample mean of historical claims, separated into the observed claims payments and unobserved claim payments for which the outstanding claims estimate $\hat{o}$ is used. The formula for $\hat{c}_{n+1}$ (which includes $\hat{o}$ in order to give an estimate for $\hat{c}_{n+1}$ that is coherent with $\hat{o}$), is

$$\hat{c}_{n+1} = \frac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n-i} \sum\limits_{k} c_{ijk} + \hat{o}}{n}. \tag{4}$$

The premium liability (denoted by p) represents a subset of the next accident period claims defined by claims from policies having inception dates $\leq n$. Let each development period within the next accident period have l_j claims of the m_j coming from policy periods incepting $\leq n$ then the premium liability amount p is

$$p = \sum_{j} \sum_{k=1}^{m_j} c_{n+1,jk},$$

and the premium liability estimator is

$$\hat{p} = \frac{\sum\limits_{j} m_j}{\sum\limits_{j} l_j} \times \hat{c}_{n+1}.$$

This paper is structured as follows. Section 2 reviews the estimation problem. Section 3 demonstrates how the model described in (2) reproduces empirical characteristics of claim payment behaviour. Section 4 derives equations for the correlation between $\hat{o}$ and $\hat{c}_{n+1}$, and the standard error of $\hat{c}_{n+1}$. Section 5 derives an algebraic solution for the standard error of $\hat{p}$, and the correlation between $\hat{c}_{n+1}$ and $\hat{p}$. Section 6 describes the process for estimating the model parameters. Conclusions are given in the final section, including suggestions for further research directions.

2 Claims Liability Estimation

Better quantitative techniques are required for estimating uncertainty for insurance liabilities. In a survey of Australian actuaries (Gibbs and Hu, 2007), 79% wanted more research on risk margins and 64% of respondents "felt that the actuarial literature on this topic was insufficient". A similar survey in the United Kingdom (Jones et al., 2006) found "approximately half" of the United Kingdom's general insurance actuaries believed the literature on reserving was adequate, while the remainder were either undecided, or believed the literature was inadequate.

Uncertainty estimates have become an essential component of the accounting and regulatory regimes of insurers in many countries. Yet the estimation of uncertainty is not an objective practice, especially for premium liabilities and diversification benefits. Gibbs and Hu found "general reasoning was the most 'popular' method" for deriving diversification benefits, because of "the lack of established quantitative and qualitative analysis that can be used in conjunction with diversification benefits" (Gibbs and Hu, 2007). Much of the published research in this area has covered the topic of estimation of outstanding claims, while research on the topics of premium liabilities and diversification is not as developed.

2.1 The Need for Uncertainty Estimates

Regulators and accounting bodies require insurers to estimate claims liabilities, and are increasingly requiring the estimation of the uncertainty of insurance liabilities. When an insurance contract is written, the insurer accepts a contingent liability to pay any claims occurring during the period of the insurance contract. This liability for potential future claims is known as a premium liability. As the contract ages, the premium liability reduces because less time remains during which claim events may occur. Once a claim occurs, the insurer has a liability to pay for the losses. The insurer's liability for a claim which has already occurred is known as an outstanding claims liability, and this liability is extinguished once the claim is paid. While term life insurance pays a claim amount agreed at the time the contract is underwritten, a general insurer will usually underwrite contracts whereby the claim amount is contingent on the uncertain loss amounts suffered by the policyholder. Since the occurrence and severities of claims are uncertain, an insurer's liabilities are also uncertain. An insurer must use estimates to quantify its liabilities.

IFRS 4 (International Accounting Standards Board, 2008) is the international accounting standard applying to insurers. The standard requires insurers to provision for outstanding claim liabilities, and to accrue premium via the use of an unearned premium provision. However, the two provisions are determined differently. While outstanding claims provisions are set from estimates, unearned premium provisions are determined via a deterministic accrual calculation. Until IFRS 4 came into force, prospective estimates of future claims were not used for accounting purposes. IFRS 4 introduced the liability adequacy test, a comparison between the premium liability (a prospective estimate of future claims) and the unearned premium provision. Whenever the unearned premium provision proves inadequate, the insurer must recognise the premium deficiency. However, the scope of IFRS 4 did not allow for the creation of a detailed accounting regime for insurance contracts. As a result, IFRS 4 does not specify whether cash flows are discounted for the time value of money, or adjusted for risk or uncertainty. IFRS 4 also does not specify whether the liability adequacy test considers the value of embedded options and guarantees.

Nevertheless, individual countries have included guidance in this regard. For example, for the purpose of the liability adequacy test, AASB1023 requires the use of a risk margin "apportioned across portfolios of contracts that are subject to broadly similar risks and are managed together as a single portfolio" (Australian Accounting Standards Board, 2007). As a result of this AASB1023 provision, Australian insurers calculate diversified risk margins for their premium liabilities. The general practice for liability adequacy testing in Australia is to use 75% adequacy diversified risk margins, although varying interpretations of "broadly similar risks" lead to varying practices regarding the grouping of contracts to calculate the diversified risk margins. Many countries interpret IFRS 4 in the same way as Australia. Therefore, to meet existing accounting requirements an insurer in those countries must consider the uncertainty and correlations of its premium liabilities. IFRS 4 allows insurers to hold liability provisions greater than the expected value of the liability, and many insurance companies choose to provision at levels much higher than expected value. For example, the 2009 annual accounts for

Insurance Australia Group indicate the insurer used risk margins of 19.1% to achieve "90%...probability of adequacy for the outstanding claims provision" (Insurance Australia Group, 2009).

IFRS 4 was an interim standard, and a more prescriptive standard is expected in the next few years. The International Accounting Standards Board believes an insurer's liability provision should include "an explicit and unbiased estimate of the margin that market participants require for bearing risk (a risk margin) and for providing other services, if any (a service margin)" (International Accounting Standards Board, 2007). As a result, accounting of insurance liabilities will move in principle towards fair value accounting, although consensus has not yet been reached on the application of value principles to insurance contracts. Nevertheless, one principle of fair value accounting is that the value of a liability or asset should not be owner specific. Therefore, under a fair value accounting regime, an insurance liability would have the same value regardless of which insurer owns the liability. For example, the size of an insurer would not affect the value of the liability, making the risk margin a function of the liability's systemic risk. Such an approach requires an estimate of the systemic component of liability risk, which in turn requires an estimate of both the uncertainty and correlations. The introduction of this new accounting standard will result in most insurers being required to estimate diversified risk margins for their insurance liabilities.

Insurance regulators are primarily concerned with the solvency of insurers. The solvency regulations of many countries have been moving from broad-based capital adequacy measures (such as the ratio of capital to premium) towards risk-based capital measures. Typically risk-based capital measures for general insurance consider the risks from claims costs, exposures on in-force contracts, investment returns and default by creditors. Many regulators require insurers to use diversified risk margins as part of their risk based capital calculations. The claims cost risk is the possibility of an insurer's liability provisions proving inadequate. To calculate a probability of adequacy of provisions, the overall uncertainty of the outcomes must be assessed, after allowing for the benefits of diversification. For example, GPS310 (Australian Prudential Regulatory Authority, 2006) requires insurance liabilities to be valued on a discounted cash flow basis, plus a risk margin set at a level which achieves a 75% adequacy after allowance for diversification between different lines of business across the insurer, and diversification between the outstanding claims and the premium liabilities. To meet this requirement, an insurer assesses both the uncertainties and correlations applying to insurance liabilities, including premium liabilities.

Solvency II (Commission of the European Communities, 2007) will introduce new prudential requirements on European insurers and reinsurers, and will require insurance technical provisions be calculated as the sum of a best estimate and a risk margin, with the risk margin calculated using the "so-called cost-of-capital method" (Commission of the European Communities, 2007). Economic capital must also take account of diversification effects. To meet the new requirements, a European insurer will need to assess the uncertainty and correlations of insurance liabilities, including premium liabilities. Many multinational insurers, such as Swiss Re, Zurich Insurance and Munich Re have their headquarters in Europe, and will report their solvency using the Solvency II framework (note that even though Switzerland is not a member of the European Union, the country has decided to adopt Solvency II). As part of their consolidated reporting, European insurers will require their international offices to report to head office on a Solvency II basis. Solvency II is likely to become a widespread standard for insurer solvency assessment.

2.2 The Statistical Paradigm for Claims Liability Uncertainty Estimation

The estimation of diversified risk margins can be thought of as a statistical estimation task. Diversified risk margins, whether they are determined from the paradigm of probability of adequacy or cost of capital, require the estimation of

1. the variability of claims outcomes versus outstanding claims estimates,
2. the variability of claims outcomes versus premium liability estimates, and
3. diversification effects.

The key statistical values used to quantify the three items are $Var(o)$, $Var(p)$ and $Var(o+p)$ respectively. Let $t = o + p$, then the values to be estimated are $Var(o)$, $Var(p)$ and $Var(t)$. The estimators are denoted as follows.

1. $\hat{\sigma}_o^2$ is the estimator of $Var(o)$. Several methods have been established for determining $\hat{\sigma}_o^2$, including the method proposed by Mack (1993), the method used in this paper.
2. $\hat{\sigma}_p^2$ is the estimator of $Var(p)$. Section 5 of this paper proposes a method for determining $\hat{\sigma}_p^2$.

3. Let $\hat{\sigma}_t^2 = \hat{\sigma}_o^2 + \hat{\sigma}_p^2 + 2\hat{\rho}_t\hat{\sigma}_o\hat{\sigma}_p$. Then $\hat{\rho}_t$ is analogous to the correlation between o and p. Even if o and p are statistically independent, $\hat{o}$ and $\hat{p}$ are not independent, because they are estimated from the same data, and because $\hat{p}$ is a linear function of $\hat{o}$. Likewise, $\hat{\sigma}_o$ and $\hat{\sigma}_p$ are not independent, as will be demonstrated later in this paper. Sections 4 and 5 of this paper develop a method for determining $\hat{\rho}_t$.

2.3 Random Effects Models and Insurance

Random effects models are used when data is grouped, or clustered, in such a manner that the results within a group are dependent, while results between groups are independent. Searle et al. (2006) gives the general form for a random effects model as

$$y_{ij} = \mu + \alpha_i + e_{ij}, \tag{5}$$

where α_i and e_{ij} are random variables. The random effect term is α_i, which applies equally to all y_{ij} having a matching index i. It can be seen that (2) is in the same form as (5), such that the accident year effect is the random effect. Each accident year can be considered a random sample from the population of possible claim environments (e.g. weather conditions), and that claim environment affects the severity of all claims occurring during that period.

This paper is not the first to propose the use of a random effects model for insurance. De Jong and Heller (2008) consider a mixed model to model insurance claims, using area and policy as explanatory variables. Area is regarded as a fixed effect, while policy is a random effect i.e. a cluster is an individual policy with multiple data points, one data point for each year of the policy appears in the data. The proposed model is

$$y = x'\beta + z'\gamma + \epsilon, \tag{6}$$

where $\gamma \sim N(0, G)$ is independently distributed of $\epsilon \sim N(0, \sigma^2)$. While the form of (2) represents a special case of the form of (6), the fixed effects term and random effects terms in (2) are not comparable. In (2) the random effect, or "cluster", is a individual accident year, which includes many individual policies. The fixed effect in (6) is a rating factor, while the fixed effect in (2) is a constant and ignores rating factors. While (6) is being used to model systemic differences in observed data, (2) is being used to extrapolate future claim payments, and to scale the mean and variance of a subset of aggregated data.

2.4 Existing Uncertainty Estimation Techniques

A review of Australian actuarial practices by the Risk Margins Taskforce described a generalised process of four steps which Australian actuaries use for determining risk margins.

> Although there appear to be a wide range of approaches used by Australian actuaries in the assessment of risk margins it is fair to say that most of the differences relate to the analysis and investigations conducted to parameterise a generally adopted risk margin calculation methodology, rather than the calculation methodology itself. The calculation methodology can be generalised as follows:
>
> 1. Coefficients of variation (CoVs) are determined for individual valuation portfolios or groupings of portfolios, where these groupings include insurance classes made up of relatively homogeneous risks.
> 2. A correlation matrix is populated with assumed correlation coefficients reflecting the expected correlations between valuation portfolios or groupings of portfolios.
> 3. CoVs and correlation matrices are determined separately for outstanding claim liabilities and premium liabilities and further assumptions made about the correlation between these two components of the insurance liabilities.
> 4. A statistical distribution is selected and combined with the adopted CoVs and correlation coefficients to determine the aggregate risk margin at a particular probability of adequacy. (Marshall et al., 2008)

The techniques proposed in sections 4, 5 and 6 of this paper are primarily concerned with developing solutions for steps 1 and 3 of the process described above by Marshall et al. (2008). Step 1 involves the separate estimation of $Var\,(o)$ and $Var\,(p)$ for homogenous groups of insurance contracts and claims. The estimation of $Var\,(o)$ is solved using methods such as Mack (1993). However, literature on estimation techniques for $Var\,(p)$ is rare. Step 3 of the process described above involves the estimation of $Var\,(t)$, a problem which has no published solution. This paper proposes techniques to estimate $Var\,(p)$ and $Var\,(t)$.

Two papers, published in 2001, recommend Australian industry benchmarks for risk margins, both undiversified and diversified. Collings and White (2001) and Bateup and Reed (2001) used industry data to determine future correlations. However, historical correlations are not shown, and no method is proposed for measuring historical correlations. The methods for determining the benchmark correlations are not discussed in detail. Due to data constraints, the industry benchmarks were based upon "a significant degree of actuarial judgement" rather than "rigorous statistical analyses" (Bateup and Reed, 2001). The two papers do not show how to use an insurer's internal data to determine suitable correlations / diversification benefits. 85% of respondents to a survey of Australian general insurance actuaries used industry benchmarks, while some "admitted using [industry benchmarks] exclusively without further quantitative or qualitative analysis" (Gibbs and Hu, 2007). According to Marshall et al. (2008), the two industry benchmarks involved formed the basis for the calculation of diversified risk margin estimation within Australia and common industry practice for estimating correlation effects did not include quantitative analysis of past experience.

In recent years papers such as Merz and Wuthrich (2009), Brehm (2002), and De Jong (2009a) developed models which include dependencies for outstanding claims reserving triangles across different lines of business, taking the approach of fitting multivariate statistical models to multiple runoff triangles. De Jong describes a model similar to the model proposed by this paper. Figure 1 is an extract from de Jong's presentation (De Jong, 2009b).

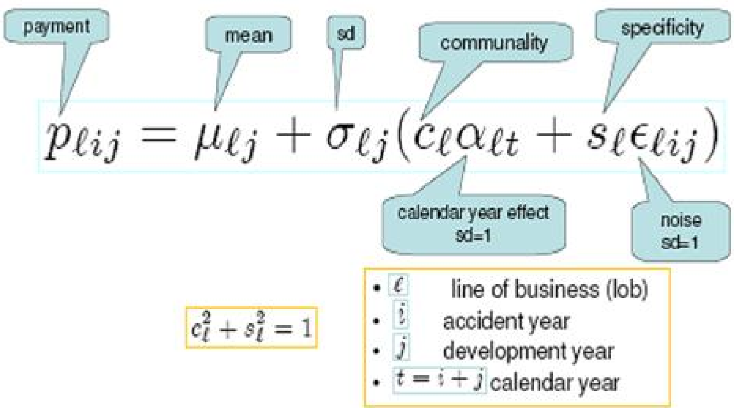

Figure 1: De Jong Presentation: "Simplest Form of Model"

Three differences exist between de Jong's model and the model proposed in this paper. First, in de Jong, multiple lines of business are considered, so the j subscript becomes obsolete. Second, the model proposed in this paper applies to individual claims, whereas de Jong's model is applied to cohorts of claims. Third, de Jong's model is a factor analysis model, using unobserved α_{lt} values, while the random effects model presented in this paper is based upon observed values. The differences in the two models are a result of the differences in intended usage. Multivariate modelling of runoff triangles aligns with steps 1 and 2 described at the beginning of this section, although solely with regards to outstanding claims, while the estimation and correlation of premium liabilities are not considered in the runoff triangle multivariate methods.

A literature search found no peer reviewed papers which model dependencies for premium liabilities or develop models for risk margins on premium liabilities. The lack of papers covering dependency and risk margin models for premium liabilities may be a reflection of the relatively short period during which premium liabilities have been used, especially when compared to the comparatively long and widespread use of outstanding claims provisions. No well-accepted standard method exists for calculating premium liabilities. Commonly, actuaries begin with the unearned premium and then look at historical loss ratios and make adjustments for rate changes, changes in mix, legislative change, and weather patterns. Adjustments are also made after considering the historical incidence of large claims and catastrophes. Premium liability estimates are affected by factors not shown in runoff triangles, e.g. changes to premium rates, changes in mix of business and the magnitude of the unearned premium.

Key stakeholders have "expressed some concern" about existing practices for assessment of risk margins, and the possibility of "significant inconsistencies in the final outcomes" (Marshall et al., 2008). More research is therefore required to develop widely accepted quantitative methods for estimating diversified risk margins.

2.5 Practical Problems With Claims Liability Uncertainty Estimation Techniques

Diversified risk margins require sophisticated models using a large number of parameters, and all the parameters to be estimated. The greater the number of parameters, the greater the amount of data required. Outstanding claim estimates have been an accounting requirement for long enough for runoff triangles to become a standard data output from insurance computer systems. Actuaries can use runoff triangle data to estimate outstanding claims and correlations of outstanding claims. However, stable historical data is required for reliable parameter estimates.

Data requirements for premium liabilities are different to those of outstanding claims. Runoff triangles are summaries of historical claims, with accident periods along the vertical axis, and development periods along the horizontal axis. Grouping of data by accident period allows estimation of outstanding claims liabilities, and separates current year claims incurred amounts from prior year movements. This approach to presenting and analysing insurance data developed at a time when unearned premiums were the sole provision for future claims and premium liabilities were not in use. Figure 2 shows the typical layout of a runoff triangle.

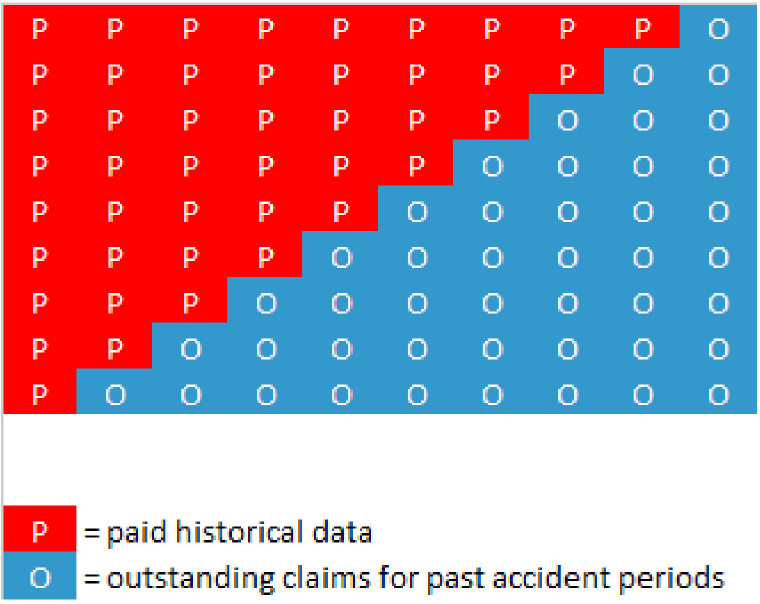

Figure 2: A Typical Runoff Triangle Layout

Two types of cells are shown in Figure 2.

1. cells coloured red and labelled “P" represent the observed historical data, and would be included in a typical triangle analysis.

2. cells coloured blue and labelled “O" represent outstanding claims - future payments on claims occurring in past accident periods - and would not exist in a triangle of historical data. Nevertheless, a typical triangle analysis would attempt to estimate the values of such cells.

All claims in a runoff triangle have already occurred. The amounts in the red P cells are claims which have already been paid. The amounts in the blue O cells are claims which have occurred, although they have not yet been paid. Some of the blue O claims have occurred, although they have not

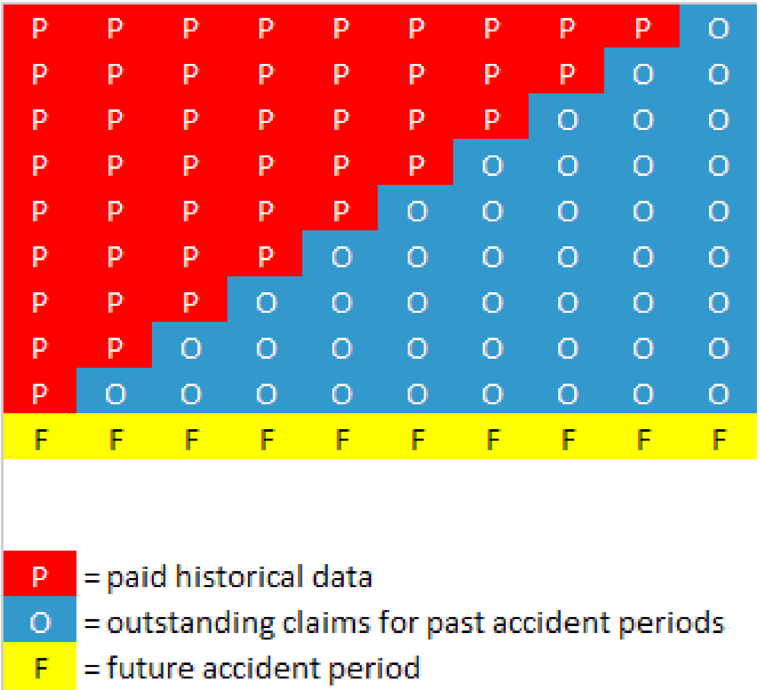

Figure 3: A Runoff Triangle Extended to a Future Accident Period

yet been reported. Figure 3 extends the runoff triangle to include a future accident period.

The yellow cells labelled “F" in Figure 3 do not appear in a traditional triangle analysis. They represent a future accident period: i.e., claims which have not yet occurred. The future accident period cells have claims arising from two different sources or components.

1. a premium liability component: the unexpired periods of existing contracts. For example if the latest accident year finished on 31-Dec-2009 then a one year insurance contract written on 30-Nov-2009 still has 11 months of cover remaining, and the claims occurring during those 11 months would belong to the accident year ending 31-Dec-2010.

2. a future incepting component: insurance contracts written in the future, yet before the end of the relevant accident period. For example, a contract written on 1-Mar-2010 would have 10 months of cover during the year ending 31-Dec-2010 and the claims occurring during those 10 months would belong to the accident year ending 31-Dec-2010.

If the accident period is the same length of time as the contract period, then a single future accident period alone will include claims from the premium liability component. Except in the unusual circumstances of a portfolio closed to new contracts, all future accident periods can contain claims from the future incepting component. The daily amount of premium exposure will reduce throughout the next accident period as policy periods end. If the policies are written uniformly throughout the year and the underlying risk is uniform through the year, then the premium liability exposure relative to the total exposure can be represented graphically as a right-angled triangle within a trapezium. Figure 4 shows the relationship between the premium liability and the future accident period.

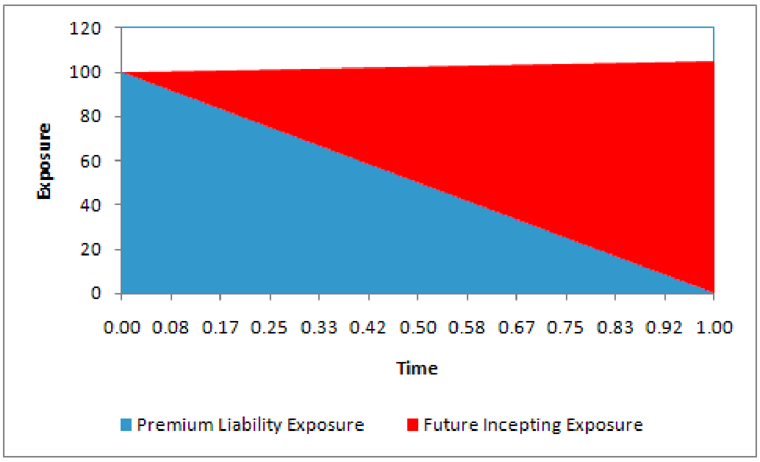

Figure 4: Premium Liability Within a Future Accident Period

In Figure 4 the blue region represents the risk exposure from the premium liability policies from time 0 to time 1. The blue exposure reduces over time as policies expire. The red region represents the exposure from future incepting policies, which increases over time as new policy periods commence. The total exposure will usually increase, as in the graph above, because of factors such as inflation, economic growth and population growth.

Runoff triangle data are inadequate to give information about the split of the premium liability component and the future incepting component in any cell. Similarly this data structure is unable to give information about the split of claims occurring on policies written in the accident period versus claims occurring on policies written before the beginning of the accident period. So the historical data, when presented in the usual triangle format, are not typically designed to allow the estimation of premium liabilities, nor to allow premium liabilities to be linked to historical claims history. Therefore, runoff triangle data are not suited to the estimation of diversified risk margins.

3 Some Important Characteristics of Residuals in Runoff Triangles

Diversified risk margins require estimates of uncertainty. As a result, much can be learned from studying the differences between claims payments and model predictions. As can be demonstrated, the residuals of runoff triangles show the characteristics of heteroscedasticity and systemic variation of correlations, leading to the choice of a random effects model for claim payments.

3.1 Runoff Triangle Residuals

Residuals are the unexplained differences between a model's predicted outcomes and the observed outcomes. The residual at accident period i and development period j is $r_{ij} = c_{ij} - \hat{c}_{ij}$, where $\hat{c}_{ij}$ is the estimate of total claim payments for accident period i and development period j.

A statistical model suitable for estimating uncertainty will explain patterns of the predicted values, and the magnitude and dependencies of residuals. The choice of model will also be determined by the purpose for which the model is being used. Diversified risk margins require a model which can be used for both outstanding claims and premium liabilities. Some models of outstanding claims, e.g. chain ladder, do not give predicted values for the first development period. Since the development period is observed for historical accident periods, the lack of predicted values for the first accident period does not pose a problem for the estimation of outstanding claims. However, having not yet occurred, the first development period for premium liabilities is unknown. As a result, a suitable model for premium liabilities will need to predict values for the first development period. Figure 5 shows the location of the first development period.

The crosshatched area of Figure 5 represents the first development period. None of the blue O cells within Figure 5 are crosshatched. Therefore models predicting outstanding claims need not predict values for the first development period. The left-most yellow F cell within Figure 5 is crosshatched. Therefore premium liability models need to include predictions for the first accident period.

The set of runoff triangle models giving predicted values for the first development period includes models such as payments per claim incurred, payments per claim finalised, payments per active claims and Bornheutter-Ferguson. This section does not propose a particular runoff triangle model. Rather, some general observations are made about the behaviour of residuals, using a variant of the payments per claims finalised model. The model

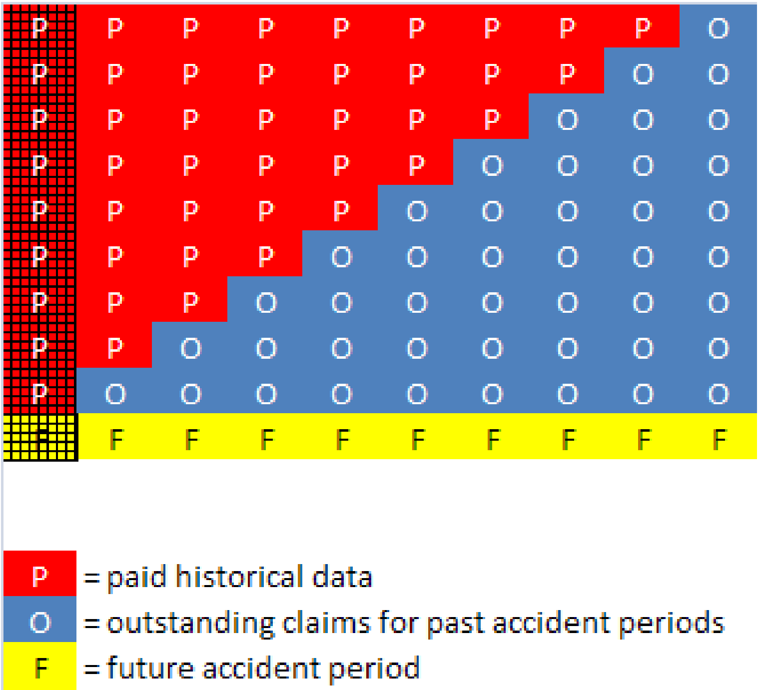

Figure 5: The first development period

explains claim payments for an accident period and development period as a multiple of the number of claims finalised during the accident period and development period. The payments per claim finalised model's generalised form is

$$\hat{c}_{ij} = n_{ij} \times \mu_j, \tag{7}$$

where:

$\hat{c}_{ij}$ is the estimate of total claim payments for accident period i and development period j;

i is the accident period;

j is the development period;

n_{ij} is the number of claims finalised within accident period i and development period j;

μ_j is the average amount of payments per claim finalised for development period j, analogous to an average claim size.

A simpler variation of the payments per claim finalised model will now be considered. First, μ_j is assumed to be the same for all values of j. Second, finalised claim counts for each development period are assumed to be proportional to earned premium. The model is defined as

$$\hat{c}_{ij} = p_i^{(E)} \times \tau_j \times \mu, \tag{8}$$

where:

$\hat{c}_{ij}$ is the estimate of total claim payments for accident period i and development period j;

i is the accident period;

j is the development period;

$p_i^{(E)}$ is the earned premium for accident period i;

τ_j is the ratio of claim count in development period j to earned premium;

μ is the average amount of payments per claim, analogous to an average claim size.

The n_{ij} term is replaced with an exposure based measure because premium liabilities relate to claims which have not yet occurred. Therefore no claim count data are available for premium liabilities. Nevertheless, the unearned premium provision is known. Thus earned premium can be used as data for a model of the premium liabilities.

3.2 Splitting Runoff Triangles By Underwriting Period

Figure 4 showed the premium liability as a subset of the next accident period. Typically, an insurer collects data in sufficient detail to split claims by both accident period and underwriting period. While accident period is required for runoff triangles, underwriting period is required for policy administration purposes and often for reinsurance purposes. Using both accident period and underwriting period together, a runoff triangle can be split into two sub-triangles. One sub-triangle, denoted H, contains claims belonging to contracts commencing before the accident period in which the claim occurred, analogous to the claims belonging to the premium liability at the start of the accident period. The second sub-triangle, denoted F, contains claims belonging to contracts commencing during the accident period during which the claim occurred. The total of the payments for the matching accident period and development period from the two sub-triangles equals the claim payments for the total portfolio i.e. $c_{Hij} + c_{Fij} = c_{ij}$.

Tables 1, 2 and 3, show sample data split into two sub-triangles on the basis of underwriting period. The additive feature of the sub-triangles versus the runoff triangle is demonstrated by comparing $c_{H2,1} + c_{F2,1} = 121686 + 174684 = 296370 = c_{2,1}$.

		Development Period										
Acc Yr	**Earned Premium**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1999	1,749,038	225,633	619,849	660,078	203,140	23,546	20,327	32,575	33,872	120,881	14,297	0
2000	2,358,629	296,370	906,406	427,835	178,646	84,963	101,616	83,971	17,072	30,400	2,722	
2001	2,617,002	303,407	933,581	450,686	249,244	27,045	98,584	7,295	2,334	374		
2002	3,159,899	273,405	1,621,091	1,310,029	910,790	439,714	234,868	140,858	86,572			
2003	3,480,946	285,737	778,189	491,470	186,238	42,922	35,382	25,245				
2004	3,561,378	358,054	1,364,101	1,491,556	1,117,384	802,264	434,305					
2005	3,703,335	349,259	868,717	1,166,016	426,749	400,139						
2006	3,879,841	187,509	879,653	620,731	180,029							
2007	3,857,307	237,752	917,759	1,014,382								
2008	3,451,561	375,578	1,768,252									
2009	3,774,444	993,321										

Table 1: Runoff Triangle: All Claims

		Development Period										
	Earned Premium	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1999	1,059,472	58,768	387,971	584,061	177,981	20,417	12,428	32,575	33,872	115,903	13,618	0
2000	1,311,458	121,686	627,080	266,608	102,941	44,740	44,006	51,850	6,765	13,798	2,722	
2001	1,066,305	132,428	451,024	222,608	49,595	18,643	72,413	2,393	1,299	374		
2002	1,455,441	122,937	949,368	833,736	690,258	411,718	200,098	138,464	81,839			
2003	1,620,151	108,060	544,635	385,158	156,736	24,428	17,758	3,568				
2004	1,563,641	202,774	758,451	922,736	930,366	658,447	342,346					
2005	1,725,857	127,272	372,423	622,170	240,235	233,934						
2006	1,761,483	97,494	560,068	329,758	107,807							
2007	1,607,108	113,435	355,111	542,601								
2008	1,238,449	115,496	859,354									
2009	1,123,013	382,511										

Table 2: Sub-Triangle: Claims for Policies Incepting Before the Accident Period

Residuals can then be calculated for sub-triangles in the same manner as the runoff triangle. As a consequence, $r_{Hij} = c_{Hij} - \hat{c}_{Hij}$ and $r_{Fij} = c_{Fij} - \hat{c}_{Fij}$. This is confirmed by comparing values from tables 4, 5 and 6. For example, $r_{H2,1} + r_{F2,1} = -21496 + 60356 = 38859 = r_{2,1}$.

Splitting a runoff triangle into two sub-triangles defined by underwriting period restructures the data to allow the analysis of the relationship of premium liability claims to the complete accident period to which they belong. Such an analysis allows the construction of a unifying model for outstanding claims and premium liabilities. While splitting a triangle provides valuable data, this paper develops a solution which does not require triangles to be split into sub-triangles.

	Earned Premium	0	1	2	3	4	5	6	7	8	9	10
		Development Period										
1999	689,566	166,865	231,877	76,017	25,159	3,128	7,899	0	0	4,978	679	0
2000	1,047,171	174,684	279,326	161,226	75,704	40,223	57,610	32,121	10,307	16,601	0	
2001	1,550,697	170,979	482,557	228,078	199,648	8,403	26,171	4,902	1,035	0		
2002	1,704,458	150,469	671,723	476,292	220,532	27,996	34,769	2,393	4,734			
2003	1,860,795	177,677	233,554	106,311	29,502	18,494	17,624	21,676				
2004	1,997,737	155,279	605,650	568,820	187,018	143,816	91,959					
2005	1,977,478	221,987	496,295	543,846	186,514	166,205						
2006	2,118,358	90,016	319,585	290,973	72,222							
2007	2,250,199	124,317	562,647	471,781								
2008	2,213,112	260,082	908,897									
2009	2,651,432	610,810										

Table 3: Sub-Triangle: Claims for Policies Incepting During the Accident Period

Acc Yr	0	1	2	3	4	5	6	7	8	9	10
	Development Period										
1999	34,676	34,017	189,466	-43,210	-130,805	-75,261	-5,367	9,126	81,437	7,051	0
2000	38,859	116,395	-206,799	-153,565	-123,183	-27,287	32,805	-16,299	-22,792	-7,051	
2001	17,687	57,029	-253,468	-119,359	-203,902	-44,439	-49,477	-34,692	-58,645		
2002	-71,587	562,698	459,798	465,721	160,856	62,174	72,309	41,865			
2003	-94,307	-387,737	-445,145	-304,050	-264,267	-154,857	-50,269				
2004	-30,772	171,235	533,299	615,767	487,977	239,670					
2005	-55,064	-371,697	169,563	-94,862	73,324						
2006	-236,085	-419,881	-423,214	-366,443							
2007	-183,383	-374,227	-23,500								
2008	-1,257	612,168									
2009	581,233										

Table 4: Residuals for All Policies

3.3 Empirical Correlations of Runoff Triangle Residuals

According to the sample data provided in Tables 1, 2 and 3, runoff triangle residuals exhibit heteroscedasticity. Furthermore, the matching pairs of sub-triangle residuals exhibit positive correlation, and the correlation coefficient of residuals is not constant.

Recall the modified payments per claim finalised model described earlier in this section. $\hat{\tau}_j$ and $\hat{\mu}$ are the estimators for τ_j and μ respectively, calculated by applying the formulae

	Development Period										
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1999	-56,903	33,107	298,989	28,756	-73,080	-45,474	9,592	18,882	92,010	9,229	0
2000	-21,496	187,814	-86,264	-81,776	-70,994	-27,667	23,400	-11,790	-15,778	-2,712	
2001	16,011	93,870	-64,302	-100,593	-75,458	14,138	-20,739	-13,787	-23,673		
2002	-35,966	461,876	442,122	485,260	283,277	120,556	106,891	61,247			
2003	-68,826	1,974	-50,774	-71,461	-118,548	-70,786	-31,578				
2004	32,059	234,717	502,008	710,129	520,458	256,891					
2005	-61,154	-205,645	157,796	-2,850	81,629						
2006	-94,822	-29,932	-144,202	-140,296							
2007	-62,027	-183,181	110,178								
2008	-19,716	444,542									
2009	259,902										

Table 5: Residuals for Policies Incepting Before the Accident Period

	Development Period										
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1999	91,579	910	-109,523	-71,966	-57,725	-29,787	-14,959	-9,756	-10,574	-2,178	0
2000	60,356	-71,419	-120,535	-71,789	-52,189	380	9,405	-4,509	-7,015	-4,339	
2001	1,676	-36,842	-189,167	-18,766	-128,444	-58,577	-28,738	-20,905	-34,971		
2002	-35,621	100,823	17,675	-19,539	-122,421	-58,382	-34,582	-19,382			
2003	-25,482	-389,710	-394,371	-232,589	-145,719	-84,071	-18,691				
2004	-62,830	-63,482	31,291	-94,361	-32,481	-17,221					
2005	6,090	-166,052	11,768	-92,012	-8,305						
2006	-141,263	-389,949	-279,012	-226,147							
2007	-121,356	-191,046	-133,678								
2008	18,458	167,626									
2009	321,331										

Table 6: Residuals for Policies Incepting During the Accident Period

$$\hat{\tau}_j = \frac{\sum_i c_{ij}}{\sum_i p_i^{(E)}} \text{ and } \hat{\mu} = 1, \tag{9}$$

to the runoff triangle data from Table 1.

Table 7 shows the calculation of $\hat{\tau}_j$ estimators from the sample data in Table 1.

These $\hat{\tau}_j$ estimators are applied to the runoff triangle, and also to the two sub-triangles. Calculate the residuals for the runoff triangle and each sub-triangle using the formulae described in the previous subsection. Tables 4, 5 and 6 show the residuals calculated against the data in Tables 1, 2 and 3. Denote $\hat{\sigma}_j$ as the sample standard deviation of the observed residuals for

Development Year	$\sum_i c_{ij}$	$\sum_i p_i^{(E)}$	$\hat{\tau}_j$
1	3,886,025	35,593,380	0.10918
2	10,657,596	31,818,936	0.33495
3	7,632,782	28,367,375	0.26907
4	3,452,219	24,510,068	0.14085
5	1,820,593	20,630,227	0.08825
6	925,082	16,926,892	0.05465
7	289,943	13,365,514	0.02169
8	139,850	9,884,568	0.01415
9	151,654	6,724,669	0.02255
10	17,019	4,107,667	0.00414

Table 7: Calculation of $\hat{\tau}_j$

development period j.

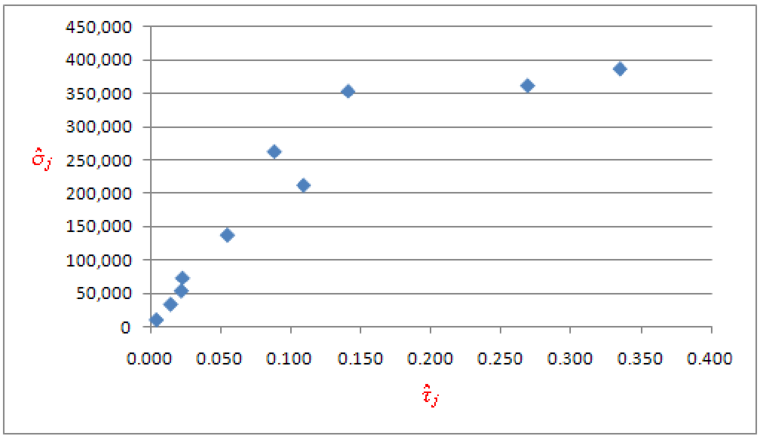

Figure 6: Empirical $\hat{\sigma}_j$ versus $\hat{\tau}_j$

Figure 6 graphs $\hat{\sigma}_j$ against $\hat{r}_j$, and shows the residuals to have significant heteroscedasticity as a function of $\hat{\tau}_j$. The relationship between $\hat{\sigma}_j$ and $\hat{\tau}_j$ appears to be monotonically increasing and concave.

Group matching pairs of sub-triangle residuals $\{r_{Hij}, r_{Fij}\}$ by development period and estimate the correlations of the matching pairs. Let $\hat{\rho}_j$ be the sample correlation coefficient for the paired residuals in development period j. Figure 7 graphs $\hat{\tau}_j$ against $\hat{\rho}_j$.

The correlations are not constant. This is demonstrated in Figure 7 by graphing $\hat{\rho}_j$ against $\hat{\tau}_j$. The relationship between $\hat{\rho}_j$ and $\hat{\tau}_j$ appears to be monotonically increasing and concave. Few values of $\hat{\rho}_j$ are negative.

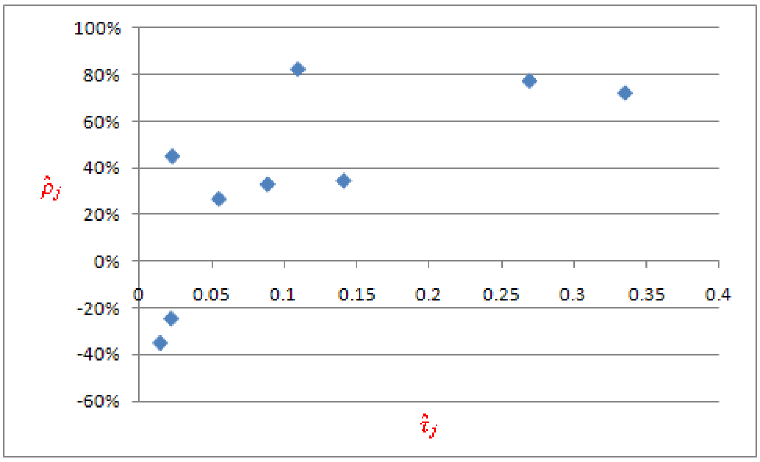

Figure 7: Empirical $\hat{\rho}_j$ versus $\hat{\tau}_j$

Since $\hat{\tau}_j$ is the sum of a number of claim payments in development period j divided by earned premium, higher values of $\hat{\tau}_j$ will occur when a greater number of claims are settled in development period j, everything else being equal. Since c_{ij} is a random variable, $\sum_i c_{ij}$ is also a random variable, with increasing variance as the number of claim payments rises. Predictably, therefore $\hat{\sigma}_j$ is monotonically increasing and concave when mapped against $\hat{\tau}_j$.

3.4 Simulated Correlations of Runoff Triangle Residuals

If the residual characteristics of the sample data are a common feature of runoff triangles, then similar characteristics should be observed from simulations of claim payments. A Monte Carlo simulation was constructed using the following assumptions.

1. Claim sizes are Gamma distributed with mean 1000 and standard deviation of 500.

2. The number of claims in each accident period for each subportfolio is Poisson distributed with mean 100.

3. The number of claims in each accident period for the two subportfolios are linked via a Normal copula with a correlation coefficient of 80%.

4. The development period of payment for each claim is Poisson distributed with mean 2.
5. There is no claim inflation.
6. Claim payments for different accident periods are independent.
7. There are 10 accident periods.

Assumptions 1 to 3 lead to accident period incurred totals with 78% correlation coefficient between the two subportfolios. The results of the simulation were examined for evidence of a concave monotonically increasing relationship between the number of claims and the correlation of residuals within a development period.

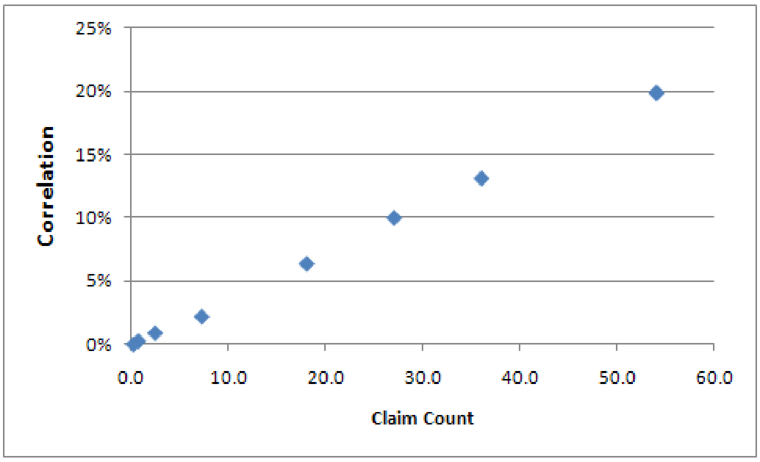

Figure 8: Simulated $\hat{\rho}_j$ versus $\hat{\tau}_j$

The correlations of residuals are not constant, and this is demonstrated in Figure 8. As is the case with the empirical evidence, the relationship between $\hat{\rho}_j$ and $\hat{\tau}_j$ is monotonically increasing. Although Figure 8 does not visually confirm a concave relationship, $\hat{\rho}_j$ is a correlation coefficient and cannot take values greater than 1. Therefore a concave relationship must exist.

Table 8 shows the average correlation coefficient by development period. The maximum correlation coefficient occurs at development periods 2 and 3, coinciding with the maximum average claim counts. While the simulation parameters were chosen to give a correlation coefficient for the total payments in accident periods of approximately 80%, the maximum mean correlation coefficient of the residuals was 20%. As indicated by this important observation, the correlation of residuals will be lower than the correlation of accident

Development Period j	Mean of $\hat{\rho}_j$	Mean Count of Claim Payments
1	10%	27.1
2	20%	54.1
3	20%	54.1
4	13%	36.1
5	6%	18.0
6	3%	7.2
7	1%	2.4
8	0%	0.7
9	0%	0.2

Table 8: Summary of Simulated Correlations

period totals. Since claims liabilities are composed of accident period totals, the correlation coefficient estimates used for claims liabilities will be higher than the correlations of residuals. None of the mean correlations are negative.

Figures 9 to 17 document the changing correlation characteristics across development periods. Correlations for development periods with higher claim counts have smoother and narrower distributions. As the claim counts reduce, the distribution of correlations broadens.

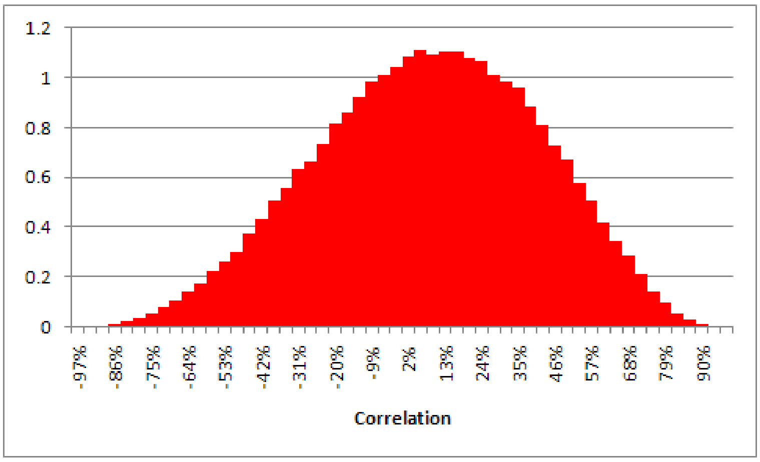

Figure 9: Histogram of $\hat{\rho}_1$

From the simulation results, the histograms of the distributions of $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$, $\hat{\rho}_4$ and $\hat{\rho}_5$ exhibit similar characters. The distributions are smooth, skewed, centred at values greater than zero, and bounded at -1 and 1.

The histogram of the distribution of $\hat{\rho}_6$, while smooth, is not skewed. It is centred at zero and is not as peaked as the distributions of $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$,

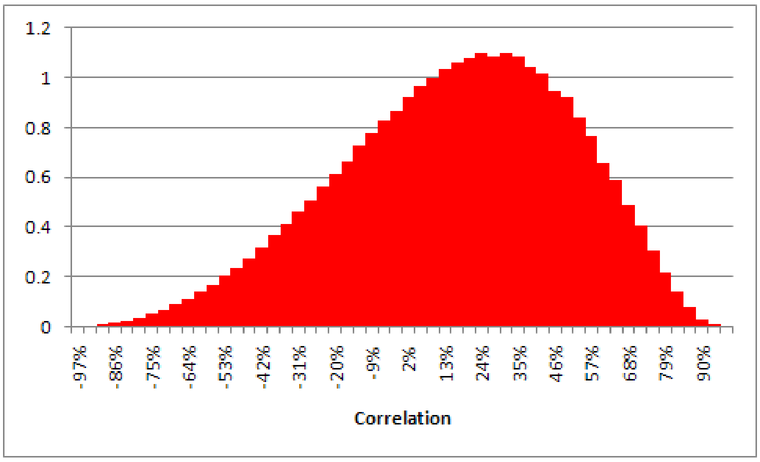

Figure 10: Histogram of $\hat{\rho}_2$

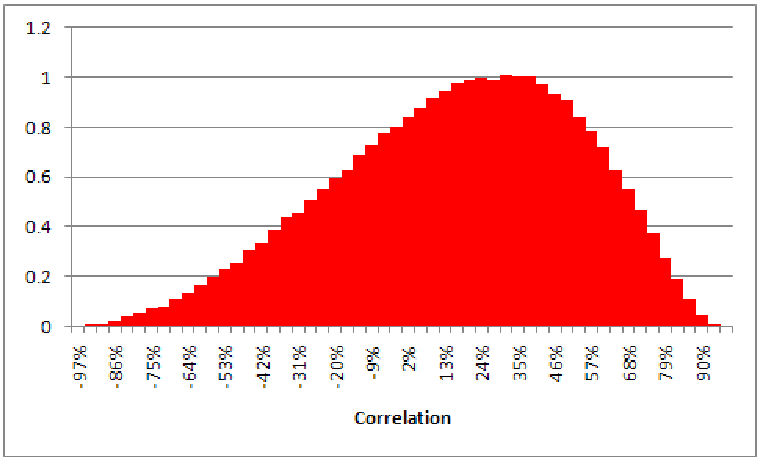

Figure 11: Histogram of $\hat{\rho}_3$

$\hat{\rho}_4$ and $\hat{\rho}_5$. This indicates that the positive correlations observed in earlier development periods do not continue into development period 6.

The histogram of the distribution of $\hat{\rho}_7$ is not smooth. The most significant discontinuity is at $\hat{\rho}_7 = 0$, but there are also discontinuities apparent around $\hat{\rho}_7 = -0.5$ and $\hat{\rho}_7 = 1$. At development period 7, the continuous distributions observed for $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$, $\hat{\rho}_4$, $\hat{\rho}_5$ and $\hat{\rho}_6$ are beginning to collapse towards the discrete distributions that were observed in higher development periods.

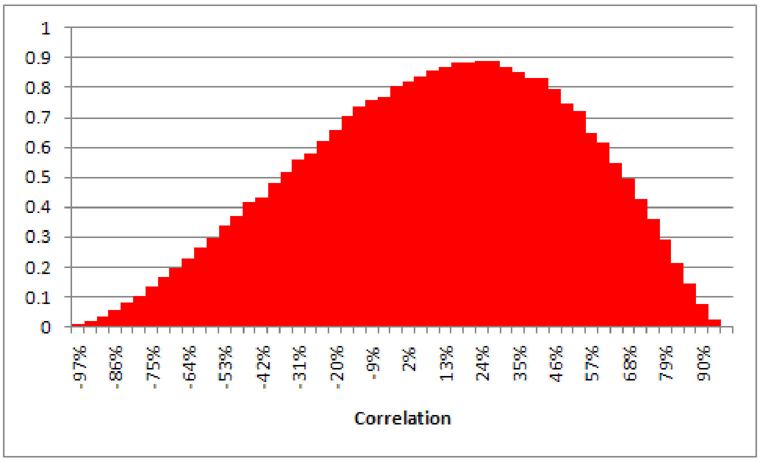

Figure 12: Histogram of $\hat{\rho}_4$

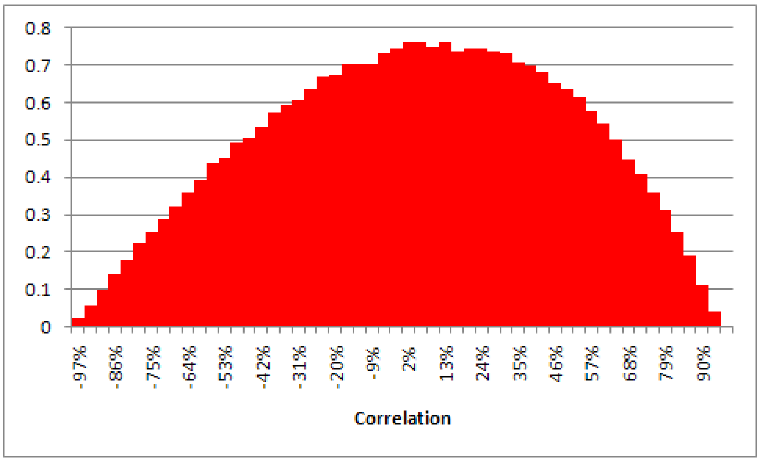

Figure 13: Histogram of $\hat{\rho}_5$

The histogram of the distribution of $\hat{\rho}_8$ is discrete, taking values of -0.5, 0 and 1. These values correspond to the discontinuities observed in development period 7.

The histogram of the distribution of $\hat{\rho}_9$ is also discrete, taking values of -1, 0 and 1.

The simulation results support the initial observation: the relationship between $\hat{\rho}_j$ and $\hat{\tau}_j$ is monotonically increasing and concave. Furthermore, $\hat{\rho}_j$ is less than the underlying correlation of claim totals, and hence is a biased estimator of the correlation of claims liabilities.

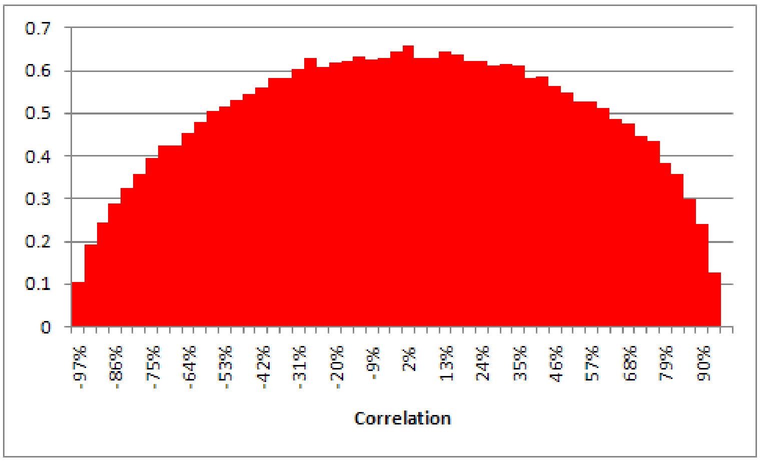

Figure 14: Histogram of $\hat{\rho}_6$

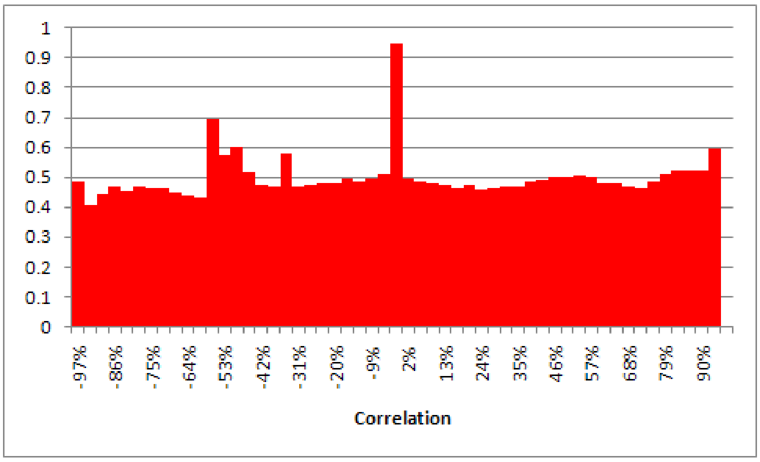

Figure 15: Histogram of $\hat{\rho}_7$

3.5 Systemic Risk

The term "systemic risk" has different meanings in different industries. In the context of the general insurance industry, Bateup and Reed define the term "systemic" as "the element of the total coefficient of variation that is constant across the whole line of business, irrespective of the size of the liability" (Bateup and Reed, 2001). In a similar way, the word "independent" is defined by Bateup and Reed as "the element of the total coefficient of variation that is related to the size of the liability" (Bateup and Reed, 2001). Distinguishing systemic from independent risk is important because

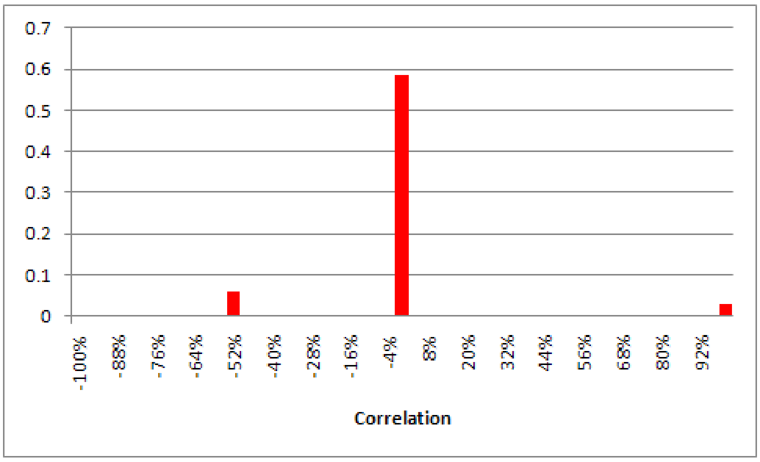

Figure 16: Histogram of $\hat{\rho}_8$

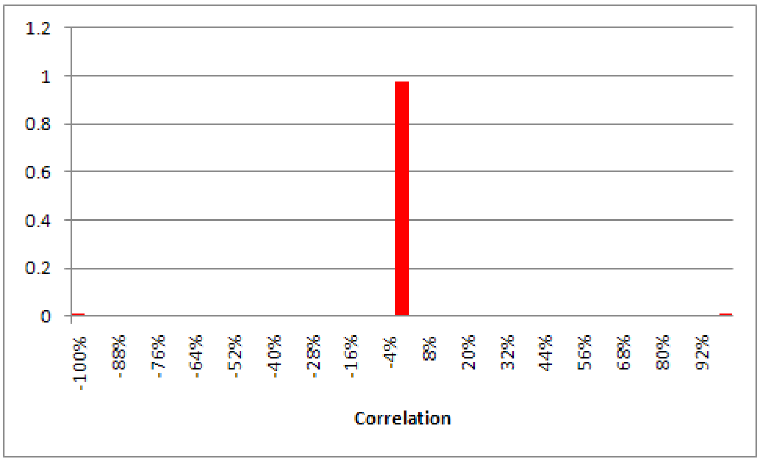

Figure 17: Histogram of $\hat{\rho}_9$

systemic risk is correlated, whereas independent risk is statistically independent. Furthermore, systemic risk is proportional to the size of an insurance portfolio, whereas independent risk is proportional to the square root of the size of a portfolio. Thus, to achieve a suitable description of claims liability correlations, the model must allow for changes in the ratio of systemic to independent risk. In a suitable model, the greater the number of claims, the greater the correlations.

Let s be the systemic variance per claim, and i be the independent variance per claim. The total variance for a portfolio of size n will be

$f(n) = s \times n^2 + i \times n$. A suitable model for claims will produce this form for $f(n)$ when portfolios of claims are totalled together. This form of $f(n)$ results in a monotonically increasing and concave function for the standard deviation, as was observed in the study of the sample data.

Furthermore, as was demonstrated earlier in this paper, a premium liability is a subset of an accident period. If systemic risk is present within an insurance portfolio, then the uncertainty of a premium liability is not proportional to the size of the premium liability versus the full accident period. Thus, systemic risk must be estimated as part of the process of estimating the uncertainty of premium liabilities.

3.6 Correlation and Heteroscedasticity Characteristics of the Random Effects Model

Recall the random effects model for individual claim severities, defined in (2). Within the formula, the $\alpha \times \nu_i$ term models the systemic risk and the $\beta \times \epsilon_{ijk}$ term models the independent risk. Consider the total claim payments for accident period i and development period j, denoted by c_{ij}. Let n_{ij} be the number of claim payments made for accident period i and development period j. Then the variance of the aggregate claims in development period j of accident period i is

$$\begin{aligned} Var(c_{ij}) &= Var\left(\sum_{k=1}^{n_{ij}} c_{ijk}\right) \\ &= Var\left(\sum_{k=1}^{n_{ij}} [\mu + \alpha\nu_i + \beta\epsilon_{ijk}]\right) \\ &= Var\left(\sum_{k=1}^{n_{ij}} \alpha\nu_i\right) + Var\left(\sum_{k=1}^{n_{ij}} \beta\epsilon_{ijk}\right) \\ &= (\alpha n_{ij})^2 \times Var(\nu_i) + \beta^2 n_{ij} Var(\epsilon_{ijk}) \\ &= (\alpha n_{ij})^2 + \beta^2 n_{ij}. \end{aligned} \tag{10}$$

The μ term is a constant and can be dropped from the variance calculation. The ν_i term is common across all of the claims. Since the ϵ_{ijk} term is independently identically distributed, there is no covariance term in the result. This result is in the form $f(n) = s \times n^2 + i \times n$, meeting the requirements of a suitable claims model.

Consider accident period i and development period j for two sub-triangles H and F with counts of claims n_H and n_F respectively, which share the parameters α and β for the systemic and non-systemic risk components respectively. The standard deviations of the claim payments c_{Hij} and c_{Fij} will be $\sigma_H = \sqrt{\alpha^2 n_H^2 + \beta^2 n_H}$ and $\sigma_F = \sqrt{\alpha^2 n_F^2 + \beta^2 n_F}$ respectively. The correlation of c_{Hij} and c_{Fij} is a function of n_H and n_F as follows:

$$\rho_{HF} = \frac{\alpha \times n_H}{\sqrt{\alpha^2 n_H^2 + \beta^2 n_H}} \times \frac{\alpha \times n_F}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}}. \tag{11}$$

Proposition 1 *The correlation of claim payments between two sub-triangles can be calculated as* $\rho_{HF} = \dfrac{\alpha n_H}{\sqrt{\alpha^2 n_H^2 + \beta^2 n_H}} \times \dfrac{\alpha n_F}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}}$.

Proof. From first principles we have

$$\begin{aligned}
\rho_{HF} &= \frac{Cov\left(\sum_{k=1}^{n_H} c_{Hij}, \sum_{k=1}^{n_F} c_{Fij}\right)}{\sqrt{Var\left(\sum_{k=1}^{n_H} c_{Hij}\right) Var\left(\sum_{k=1}^{n_F} c_{Fij}\right)}} \\
&= \frac{Cov\left(\sum_{k=1}^{n_H} \alpha \times \nu_i, \sum_{k=1}^{n_F} \alpha \times \nu_i\right)}{\sqrt{\left(\alpha^2 n_H^2 + \beta^2 n_H\right)\left(\alpha^2 n_F^2 + \beta^2 n_F\right)}} \\
&= \frac{Cov\left(n_H \times \alpha \times \nu_i, n_F \alpha \times \nu_i\right)}{\sqrt{\left(\alpha^2 n_H^2 + \beta^2 n_H\right)\left(\alpha^2 n_F^2 + \beta^2 n_F\right)}} \\
&= \frac{\alpha^2 \times n_H \times n_F Cov\left(\nu_i, \nu_i\right)}{\sqrt{\left(\alpha^2 n_H^2 + \beta^2 n_H\right)\left(\alpha^2 n_F^2 + \beta^2 n_F\right)}} \\
&= \frac{\alpha \times n_H}{\sqrt{\alpha^2 n_H^2 + \beta^2 n_H}} \times \frac{\alpha \times n_F}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}}.
\end{aligned}$$

■

(11) will be used later in this paper for the derivation of the standard deviation and correlation of the premium liability. Furthermore, (11) has the characteristic of being monotonically increasing and concave as a function of the number of claims.

Proposition 2 *The correlation of claim payments between two sub-triangles is a monotonically increasing function of the number of claim payments.*

Proof. From (11),

$$\begin{aligned}\frac{\partial}{\partial n_H}\rho_{HF} &= \frac{\partial}{\partial n_H}\left(\frac{\alpha \times n_H}{\sqrt{\alpha^2 n_H^2+\beta^2 n_H}} \times \frac{\alpha \times n_F}{\sqrt{\alpha^2 n_F^2+\beta^2 n_F}}\right) \\ &= \frac{\frac{1}{2}\alpha^2\beta^2 n_F n_H}{\sqrt{\alpha^2 n_F^2+\beta^2 n_F}\left(\alpha^2 n_H^2+\beta^2 n_H\right)^{\frac{3}{2}}}.\end{aligned}$$

Since n_F, n_H, α, β and $\sqrt{\alpha^2 n_F^2+\beta^2 n_F} > 0$ by definition, then $\frac{\partial}{\partial n_H}\rho_{HF} > 0$.

By symmetry, $\frac{\partial}{\partial n_F}\rho_{HF} > 0$ as well. ■

Proposition 3 *The correlation of claim payments between two sub-triangles is a concave function of the number of claim payments.*

Proof. From the previous proof,

$$\begin{aligned}\frac{\partial^2}{\partial n_H^2}\rho_{HF} &= \frac{\partial}{\partial n_H}\frac{\frac{1}{2}\alpha^2\beta^2 n_F n_H}{\sqrt{\alpha^2 n_F^2+\beta^2 n_F}\left(\alpha^2 n_H^2+\beta^2 n_H\right)^{\frac{3}{2}}} \\ &= \frac{-\alpha^2\beta^2 n_F n_H\left(n_H\alpha^2+\beta^2\right)}{\sqrt{\alpha^2 n_F^2+\beta^2 n_F}\left(\alpha^2 n_H^2+\beta^2 n_H\right)^{\frac{5}{2}}}.\end{aligned}$$

Since n_F, n_H, α, β and $\sqrt{\alpha^2 n_F^2+\beta^2 n_F} > 0$ by definition, then $\alpha^2 n_H^2+\beta^2 n_H > 0$ and $-\alpha^2\beta^2 n_F n_H < 0$ and $n_H\alpha^2+\beta^2 > 0$.

Only one term is negative. Therefore $\frac{\partial^2}{\partial n_H^2}\rho_{HF} < 0$.

By symmetry, $\frac{\partial^2}{\partial n_F^2}\rho_{HF} < 0$ as well. ■

Further to the useful properties of monotonicity and concavity, the random effects model has further useful properties for bounding the range of ρ_{HF}.

Proposition 4 *The correlation of claim payments between two sub-triangles is non-negative i.e.* $\rho_{HF} \geq 0$.

Proof. From (11),

$$\rho_{HF} = \frac{\alpha n_H}{\sqrt{\alpha^2 n_H^2+\beta^2 n_H}} \times \frac{\alpha n_F}{\sqrt{\alpha^2 n_F^2+\beta^2 n_F}}.$$

The denominators of both terms are positive.

α is a standard deviation. Therefore $\alpha \geq 0$.

$n_H > 0$ and $n_F > 0$. Therefore $\alpha n_H \geq 0$ and $\alpha n_F \geq 0$.

The numerator is non-negative and the denominator is positive. Therefore $\rho_{HF} \geq 0$. ■

Proposition 5 *The correlation of claim payments between two sub-triangles is imperfect i.e.* $\rho_{HF} < 1$*.*

Proof. Since $\beta > 0$ and $n_A > 0$, then $\sqrt{\alpha^2 n_A^2 + \beta^2 n_A} > \alpha n_A$.

Similarly, we know $\sqrt{\alpha^2 n_B^2 + \beta^2 n_B} > \alpha n_B$ because $\beta > 0$ and $n_B > 0$.

This means $\frac{\alpha n_A}{\sqrt{\alpha^2 n_A^2 + \beta^2 n_A}} < 1$ and $\frac{\alpha n_B}{\sqrt{\alpha^2 n_B^2 + \beta^2 n_B}} < 1$, and therefore $\frac{\alpha n_A}{\sqrt{\alpha^2 n_A^2 + \beta^2 n_A}} \times \frac{\alpha n_B}{\sqrt{\alpha^2 n_B^2 + \beta^2 n_B}} < 1$. ■

As noted earlier, the level of premium liability exposure reduces throughout the next accident period while the level of future incepting exposure increases during the same period. $E(n_H)$ will reduce throughout the next accident period because the exposure of H is reducing during the period. By contrast, $E(n_F)$ will increase throughout the next accident period because the exposure of F is increasing. At the start of the next accident period $n_F = 0$. Since $\lim \rho_{\ n_B \to 0} = 0$, then at the start of the next accident period the correlation of H and F is zero. Similarly at the end of the next accident period n_H will be zero and the correlation of H and F will be zero. Some time, usually near the middle of the next accident period, a point is reached at which $E(n_H) = E(n_F)$ and the correlation is maximised. Figure 18 below shows an example where $\alpha = 1000$ and $\beta = 5000$ and $n_H = 100$ at the start of the next accident period and $n_F = 100$ at the end of the next accident period.

In this section, the random effects model has been shown to be suitable for modelling claims payments because it produces variances and correlations which are monotonically increasing, concave functions of the number of claim payments.

4 Estimation of Correlations Between Future Claims and Past Claims

While the principal subject of this paper is the link between premium liabilities and outstanding claims, the limitations of runoff triangle data structures

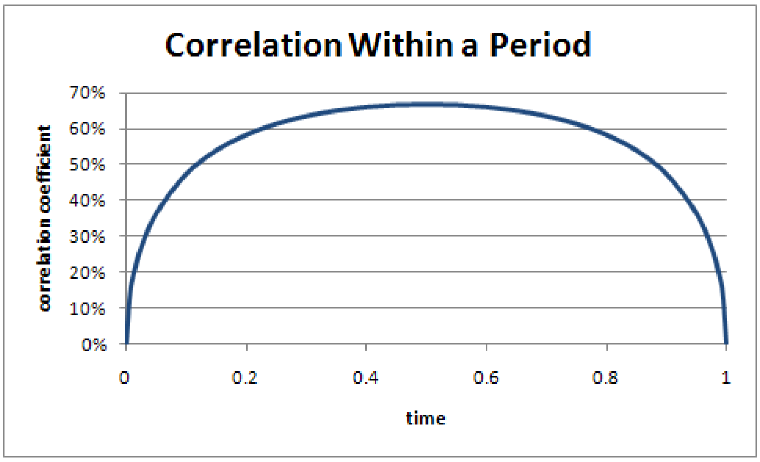

Figure 18: Instantaneous Correlation Within the Next Accident Period

discussed in section 3 mean the historical data can be linked solely to the next accident period rather than directly to the premium liability. To overcome this limitation, this section of the paper discusses the link between past claims and the next accident period, while section 5 examines the link between the next accident period and the premium liability.

4.1 Covariances

The implications of ((2)) are as follows. First, given an accident year and development year, all payments are identically distributed. The accident year effect ν_i is present in all payments falling within a given accident year, irrespective of development year. Accident year effects for different accident years are independent. Hence the mean and covariance structure of all the payments are

$$\mathrm{E}\left(c_{ijk}|v_i\right) = \mu + \alpha\nu_i \ , \qquad \mathrm{E}\left(c_{ijk}\right) = \mu \ , \tag{12}$$

and claim payments are uncorrelated except those falling in the same accident year where

$$\mathrm{cov}(c_{ijk}, c_{i\ell m}) = \begin{cases} \alpha^2 + \beta^2 \ , & j = \ell \ , \ k = m \ , \\ \alpha^2 \ , & \text{otherwise} \ . \end{cases} \tag{13}$$

The above model is a standard variance component or random effects model. In particular the ν_i are “common" random effects, common to all payments falling in a given accident year i. In contrast the ϵ_{ijk} is specific to each payment. The fact that this is a random effects model implies that standard methods can be used to both estimate model parameters as well as predict unobserved components.

Usually a random effects model's parameters are estimated using standard maximum likelihood techniques. However, insurance claims reserving data are typically provided in an aggregated form, in the format of a claims triangle and estimation techniques relying on case based data cannot be applied.

4.2 Correlation of Next Accident Period to Outstanding Claims

Typically, insurance policies will span consecutive accident periods. Some of the variability in claim outcomes relates to heterogeneity of risk between different policies i.e. variation of risk characteristics between different policies. By having some policies in common, the two consecutive accident periods share a source of uncertainty. If there is a change in the mix of risks within a portfolio, then the outcome of the change will be shared by both the outstanding claims and the next accident period. In a similar way, many accident period trends found in run-off triangles should be extrapolated into future accident periods. Examples of accident period trends include weather patterns, change in mix of business, and the insurance cycle.

The next accident period shares payment periods (i.e. diagonals in a run-off triangle) with outstanding claims. Figure 19 shows the first future payment period as a diagonal (indicated using crosshatched cells within the diagram) running from the first development period of the next accident period, down to the last development period of the earliest accident period.

Claim inflation will appear as a diagonal effect in the data. Therefore both the outstanding claims and the next accident period are affected by the same future claim inflation, although they are affected in different magnitudes, since the amount of payments involved differs between the accident years. If the future claim inflation is higher, then both the outstanding claims outcome and the next accident period outcome will be higher, everything else being equal, and the effect of the claim inflation will vary according to the claim payment pattern. As a consequence, the claim inflation effect would result in a positive correlation between the outstanding claims and the next accident period.

Since there is not the scope in this paper to consider accident period dependencies, the particular random effects model used is limited to assume independence between accident periods. The extension of this model to include accident period dependencies has been left for further research.

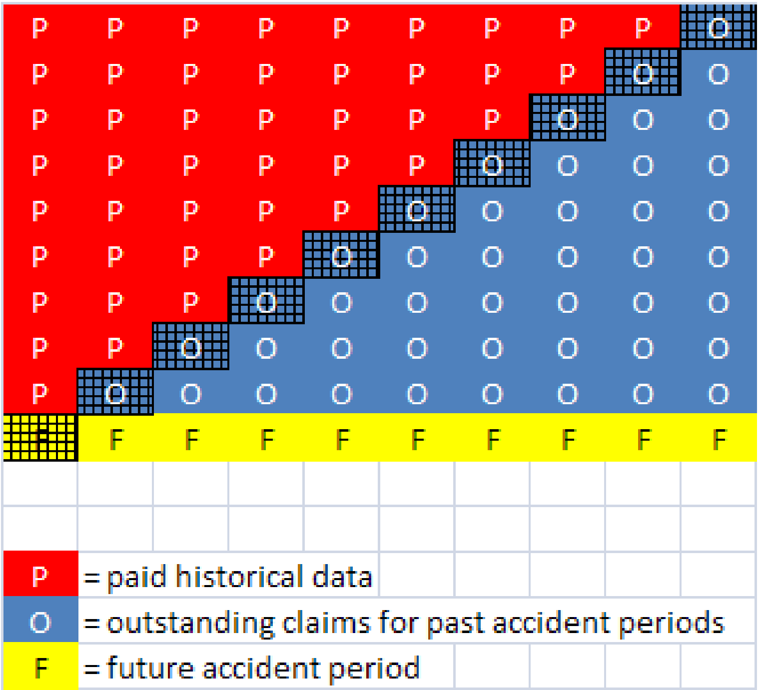

Figure 19: Shared Diagonal for Outstanding Claims and Next Accident Period

4.3 Estimator $\hat{c}_{n+1}$

Since the model defined in (2) assumes independence between accident periods, the total claims cost per accident period is independently identically distributed for each accident period. Thus, the mean of the next accident period equals the mean of the previous accident periods. It follows that

$$\bar{c}. = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k} c_{ijk}}{n}. \tag{14}$$

(14) defines a sample mean for accident year claim totals. Since the sample mean is the uniform minimum variance unbiased estimator of the mean of an independently identically distributed random variable, the sample mean is the starting point for developing an estimator for $\hat{c}_{n+1}$. However, since the claims cost of the past accident periods is not fully observed, $\bar{c}.$ cannot, in practice, be used as an estimator. The data are unbalanced, because earlier accident periods have developed further than later accident periods. To balance the data, the unobserved claims observations are replaced by $\hat{o}$, the

estimator of outstanding claims. (4) substitutes $\hat{o}$ for the unobserved claims observations, defining an estimator for $\hat{c}_{n+1}$ based solely upon observed claim payments.

4.4 Estimator $\hat{\sigma}$

Since the model defined in (2) assumes independence between accident periods, the total claims cost per accident period is independently identically distributed. Thus, the variance of the next accident period equals the variance of the previous accident periods. D_i is the total incurred cost for accident period i, defined as

$$D_i = \sum_{j=1}^{n} \sum_{k} c_{ijk}. \tag{15}$$

(4) is a special case of (3) whereby $\hat{c}_{n+1} = D_{n+1}$. If the D_i values were observable, then the sample standard deviation S^2 could be used for estimator of $\hat{\sigma}_A$, as follows:

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} \left(D_i - \frac{\sum_{i=1}^{n} D_i}{n} \right)^2. \tag{16}$$

However, the claims cost of the past accident periods is not fully observed, and S^2 cannot, in practice, be used as an estimator. When estimating the mean, the unobserved claims observations can be replaced with $\hat{o}$. However, the substitution of a constant value ($\hat{o}$) for variable outcomes would result in the underestimation of the historical variability, as demonstrated by the following thought experiment. Consider a scenario in which each accident period has the same number of claims. In this scenario, a single claim payment C_1 has been observed, and 9 claims remain to be paid. If all of the claims are independently identically distributed with variance σ^2, then $\hat{o} = 9 \times c_1$ and $D_i = c_1$ regardless of i. In consequence $S^2 = 0$ and therefore $S^2 < \sigma^2$.

In a similar way, the substitution of point estimates for D_i within (16) will result in the underestimation of the historical variability. Since S^2 is a convex function of D_i, removing the variability of the D_i values will underestimate the mean value of S^2.

A numerical approach, such as Monte Carlo simulation, could perhaps be used instead to estimate S^2. Consider the sample paths of D_i taken from a Monte Carlo simulation, and shown in Figure 20.

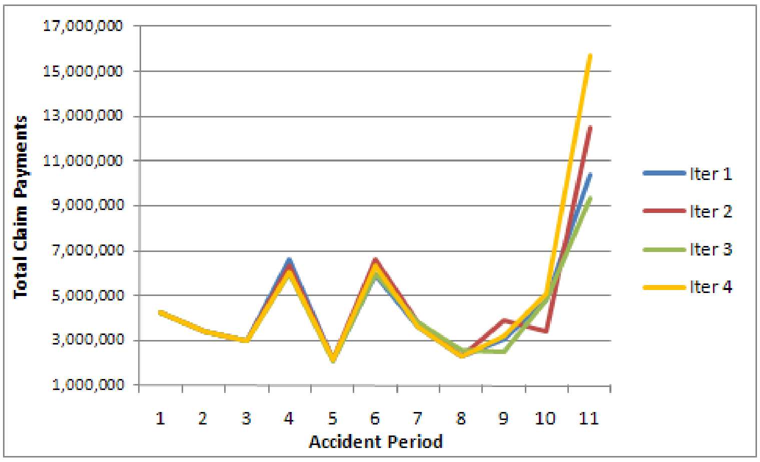

Figure 20: Sample Paths for D_i

In this case, each sample path will lead to a different S^2 value. The Monte Carlo estimator for $\hat{\sigma}^2$, the accident period variance, is the sample mean of S^2 across the iterations of the simulation. However, while they provide an unbiased estimator, Monte Carlo simulations are not the preferred solution because they can be computationally intense and require an assumption for the distribution of each D_i.

A different approach is therefore required to estimate the variance. Let $\hat{D}_i$ be the estimate of the mean of D_i, conditioned upon the observed claims, and let $\hat{\sigma}_i^2$ be the estimate of the variance of D_i, conditioned upon the observed claims. Estimator $\hat{\sigma}^2$, the accident period variance, is derived as a function of $\hat{D}_i$ and $\hat{\sigma}_i^2$ as follows:

From first principles,

$$\sigma^2 = E\left(D^2\right) - \left[E\left(D\right)\right]^2.$$

Using the same form for the estimator,

$$\begin{aligned}\hat{\sigma}^2 &= E\left(D_i^2\right) - \left[E\left(D_i\right)\right]^2 \\ &= \frac{1}{n}\sum_i\left(\left[\hat{D}_i\right]^2 + \hat{\sigma}_i^2\right) - \left(\frac{1}{n}\sum_i \hat{D}_i\right)^2. \end{aligned} \tag{17}$$

(17) is a powerful solution to the problem of estimating accident year variance, because many forms of $\hat{O}$ estimators also provide estimates of $\hat{D}_i$

and $\hat{\sigma}_i^2$ and the equation requires no assumption about the distribution of D_i. For example, the method proposed by Mack (1993) provides closed form estimators for $\hat{D}_i$ and $\hat{\sigma}_i^2$ which are suitable for use in (17).

An estimator for $\hat{D}_i$ can be derived from the model defined in (2) as follows:

$$\begin{aligned}
\hat{D}_i &= E\left(\sum_{j=1}^{n}\sum_{k} c_{ijk}\right) \\
&= E\left(\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right) + E\left(\sum_{j=n-i+1}^{n}\sum_{k} c_{ijk}\right) \\
&= \sum_{j=1}^{n-i}\sum_{k} c_{ijk} + E\left(\sum_{j=n-i+1}^{n}\sum_{k} (\mu + \alpha\cdot\nu_i + \beta\cdot\epsilon_{ijk})\right) \\
&= \sum_{j=1}^{n-i}\sum_{k} c_{ijk} + E\left(\sum_{j=n-i+1}^{n}\sum_{k} \mu\right) + E\left(\sum_{j=n-i+1}^{n}\sum_{k} (\alpha\cdot\nu_i)\right) \\
&= \sum_{j=1}^{n-i}\sum_{k} c_{ijk} + z_i\bar{\mu} + \frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} (c_{ijk} - \bar{\mu}), \qquad (18) \\
\text{where } \bar{\mu} &= \frac{1}{w}\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk} \\
\text{and } x_i &= \sum_{j=1}^{n-i} n_{ij} \text{ and } z_i = \sum_{j=n-i+1}^{n} n_{ij} \text{ and } w = \sum_{i=1}^{n}\sum_{j=1}^{n-i} n_{ij}.
\end{aligned}$$

Similarly, an estimator for $\hat{\sigma}_i^2$ (unconditioned by observations) can be derived from the model defined in (2) as follows:

$$\begin{aligned}
\hat{\sigma}_i^2 &= Var\left(\sum_{j=1}^{n}\sum_{k} c_{ijk}\right) \\
&= Var\left(\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right) + Var\left(\sum_{j=n-i+1}^{n}\sum_{k} c_{ijk}\right) \\
&= Var\left(\sum_{j=n-i+1}^{n}\sum_{k} (\mu + \alpha\cdot\nu_i + \beta\cdot\epsilon_{ijk})\right)
\end{aligned}$$

$$
\begin{aligned}
&= Var\left(\sum_{j=n-i+1}^{n}\sum_{k}(\alpha\cdot\nu_i)\right) + Var\left(\sum_{j=n-i+1}^{n}\sum_{k}(\beta\cdot\epsilon_{ijk})\right) \\
&= (z_i\alpha)^2 + z_i\beta^2. \qquad (19)
\end{aligned}
$$

The estimator $\hat{\sigma}_i^2$, shown in (19), is in the same form as (10). Given this result, the same formula can then be applied to any combination of development periods within an accident period, substituting the total claim counts from the selected development periods.

The estimator $\hat{D}_i$ implies an estimator for $\hat{o}_i$, the outstanding claims for accident period i, derived as follows:

$$
\begin{aligned}
\hat{o}_i &= \left(\hat{D}_i - \sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right) \\
&= \left(\sum_{j=1}^{n-i}\sum_{k} c_{ijk} + z_i\bar{\mu} + \frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k}(c_{ijk}-\bar{\mu}) - \sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right) \text{ from (18)} \\
&= \left(z_i\bar{\mu} + \frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k}(c_{ijk}-\bar{\mu})\right) \\
&= \bar{\mu}z_i + \left(\frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right) - \left(\frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k}\bar{\mu}\right) \\
&= \bar{\mu}z_i + \frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk} - \bar{\mu}z_i \\
&= \frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}. \qquad (20)
\end{aligned}
$$

Estimator $\hat{o}_i$, shown in (20), is in the form of a link ratio, where the link ratio is $\frac{z_i}{x_i}$. This means the claim model defined in (2) is consistent with a chain ladder model, and thus the results generated can be combined with those of existing chain ladder runoff triangle estimators, such as the method proposed by Mack (1993). This is a powerful advantage of the claim model proposed in this paper, as it unifies premium liability estimation techniques with existing outstanding claims estimation techniques. Since $\hat{o}$ is estimated via the chain ladder model, the estimator $\hat{o}$ is

$$\hat{o} = \sum_{i=1}^{n} \frac{z_i}{x_i} \sum_{j=1}^{n-i} \sum_{k} c_{ijk}. \tag{21}$$

4.5 The Variance of $\hat{o}$

In preparation for derivation of the correlation of $\hat{o}$ and $\hat{c}_{n+1}$, the variance of the estimator $\hat{o}$ is

$$\begin{aligned} Var\left(\hat{o}\right) &= Var\left(\sum_{i=1}^{n} \frac{z_i}{x_i} \sum_{j=1}^{n-i} \sum_{k} c_{ijk}\right) \\ &= \sum_{i=1}^{n} \left(\frac{z_i}{x_i}\right)^2 Var\left(\sum_{j=1}^{n-i} \sum_{k} c_{ijk}\right) \\ &= \sum_{i=1}^{n} \left(\frac{z_i}{x_i}\right)^2 \left(\alpha^2 x_i^2 + \beta^2 x_i\right) \\ &= \sum_{i=1}^{n} z_i^2 \left(\alpha^2 + \frac{\beta^2}{x_i}\right). \end{aligned} \tag{22}$$

(22) has two distinct advantages. First, the variance of the estimator increases as the total number of claims increases. Second, the variance of the estimator reduces as the proportion of settled claims increases.

4.6 The Variance of $\hat{c}_{n+1}$

The second step in preparation for derivation of the correlation of $\hat{o}$ and $\hat{c}_{n+1}$, the variance of the estimator $\hat{c}_{n+1}$ is

$$\begin{aligned} Var\left(\hat{c}_{n+1}\right) &= Var\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk} + \hat{o}}{n}\right) \\ &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk} + \sum_{i=1}^{n} \frac{z_i}{x_i} \sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right) \\ &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n}\left(1 + \frac{z_i}{x_i}\right) \sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{i=1}^{n}\left(1+\frac{z_i}{x_i}\right)^2 Var\left(\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{x_i+z_i}{x_i}\right)^2 \left(\alpha^2 x_i^2+\beta^2 x_i\right)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left(x_i+z_i\right)^2 \left(\alpha^2+\frac{\beta^2}{x_i}\right). \qquad (23)
\end{aligned}
$$

Since $\hat{c}_{n+1}$ is a function of $\hat{o}$, (23) has a similar form, and similar advantages, to (22).

4.7 The Covariance of $\hat{o}$ and $\hat{c}_{n+1}$

The final step in preparation for derivation of the correlation of $\hat{o}$ and $\hat{c}_{n+1}$ is the derivation of the covariance of $\hat{o}$ and $\hat{c}_{n+1}$ as follows:

$$
\begin{aligned}
Cov\left(\hat{o},\hat{c}_{n+1}\right) &= Cov\left(\hat{o}, \frac{\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}+\hat{o}}{n}\right)\\
&= \frac{1}{n}Cov\left(\sum_{i=1}^{n}\frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}, \sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}+\sum_{i=1}^{n}\frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right)\\
&= \frac{1}{n}Cov\left(\sum_{i=1}^{n}\frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}, \sum_{i=1}^{n}\left(1+\frac{z_i}{x_i}\right)\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{z_i}{x_i}\left(\frac{x_i+z_i}{x_i}\right) Var\left(\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{z_i}{x_i}\left(\frac{x_i+z_i}{x_i}\right)\left(\alpha^2 x_i^2+\beta^2 x_i\right)\\
&= \frac{1}{n}\sum_{i=1}^{n} z_i\left(x_i+z_i\right)\left(\alpha^2+\frac{\beta^2}{x_i}\right). \qquad (24)
\end{aligned}
$$

(24) shares a similar form to (23) and (22).

4.8 Correlation of $\hat{o}$ and $\hat{c}_{n+1}$

$\hat{o}$ and $\hat{c}_{n+1}$ are related because both $\hat{o}$ and $\hat{c}_{n+1}$ are estimated from the same data. The greater the historical claims payments, the higher the estimates of both $\hat{o}$ and $\hat{c}_{n+1}$.

Proposition 6 $Corr(\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}, \hat{o}) > 0.$

Proof. From (20),

$\hat{o} = \sum_{i=1}^{n}\left(\frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right).$

First, take the partial differential of $\hat{o}$ with respect to c_{ijk}.

$$\begin{aligned}\frac{\partial}{\partial C_{ijk}}\hat{o} &= \frac{\partial}{\partial C_{ijk}}\sum_{i=1}^{n}\left(\frac{z_i}{x_i}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\right)\\ &= \frac{\partial}{\partial C_{ijk}}\frac{z_i}{x_i}c_{ijk}\\ &= \frac{z_i}{x_i}.\end{aligned}$$

Since $z_i > 0$ and $x_i > 0$, it follows that $\frac{z_i}{x_i} > 0$.

Now, take the partial differential of $\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}$ with respect to c_{ijk}.

$$\begin{aligned}\frac{\partial}{\partial C_{ijk}}\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk} &= \frac{\partial}{\partial C_{ijk}}\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}\\ &= \frac{\partial}{\partial C_{ijk}}c_{ijk}\\ &= 1.\end{aligned}$$

Since $\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk}$ and $\hat{o}$ are both increasing functions with respect to c_{ijk}, it follows that they must be positively correlated. ■

In the model defined in (2), accident periods are assumed to be statistically independent. Since o and c_{n+1} are composed of disjoint accident periods, (2) assumes o and c_{n+1} are statistically independent. However, the estimators $\hat{o}$ and $\hat{c}_{n+1}$ are not statistically independent, primarily because (4) defines $\hat{c}_{n+1}$ as a linear function of $\hat{o}$. The correlation of $\hat{o}$ and $\hat{c}_{n+1}$ is derived using the results presented earlier in this section.

$$
\begin{aligned}
Corr\left(\hat{o}, \hat{c}_{n+1}\right) &= \frac{Cov\left(\hat{o}, \hat{c}_{n+1}\right)}{\sqrt{Var\left(\hat{o}\right) Var\left(\hat{c}_{n+1}\right)}} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} z_i\left(x_i+z_i\right)\left(\alpha^2+\frac{\beta^2}{x_i}\right)}{\sqrt{\left(\sum_{i=1}^{n} z_i^2\left(\alpha^2+\frac{\beta^2}{x_i}\right)\right)\times\left(\frac{1}{n^2}\sum_{i=1}^{n}\left(x_i+z_i\right)^2\left(\alpha^2+\frac{\beta^2}{x_i}\right)\right)}} \\
&= \frac{\sum_{i=1}^{n} z_i\left(x_i+z_i\right)\left(\alpha^2+\frac{\beta^2}{x_i}\right)}{\sqrt{\left(\sum_{i=1}^{n} z_i^2\left(\alpha^2+\frac{\beta^2}{x_i}\right)\right)\times\left(\sum_{i=1}^{n}\left(x_i+z_i\right)^2\left(\alpha^2+\frac{\beta^2}{x_i}\right)\right)}} \quad (25)
\end{aligned}
$$

(25) gives the correlation between $\hat{o}$ and $\hat{c}_{n+1}$. Since all of the terms are positive, $\hat{o}$ and $\hat{c}_{n+1}$ are positively correlated. This is important, since the positive correlation result for $\hat{o}$ and $\hat{c}_{n+1}$ reduces the diversification benefits between outstanding claims and premium liability estimates, as will be shown in the next section of this paper.

5 Estimates of the Premium Liability

The previous section of this paper derived the link between outstanding claims and the next accident period. This section discusses the subsequent link between the next accident period and the premium liability. Once the characteristics of the second link have been determined, the two links will be combined to derive the relationship between outstanding claims and premium liabilities.

5.1 Estimator $\hat{p}$

Assume there is no difference in risk characteristics between the premium liability and the future incepting component of the next accident period, then $\hat{p}$ is proportional to $\hat{c}_{n+1}$. Let $\hat{n}_P$ be the number of claims estimated for the premium liability. Let u be the unearned premium. The estimator $\hat{n}_P$ is

$$\hat{n}_P = u\sum_{j=1}^{n} r_j. \quad (26)$$

Similarly, let $\hat{n}_A$ be the number of claims estimated for the next accident period. The estimator $\hat{n}_A$ is

$$\hat{n}_A = \hat{p}^{(E)}_{n+1} \sum_{j=1}^{n} r_j. \tag{27}$$

In (27), the earned premium term $\hat{p}^{(E)}_{n+1}$ is an estimate rather than an observed data point. This follows because $p^{(E)}_{n+1}$ consists of the known value u, plus the earned premium from the unknown amount of future written contracts. In practice $\hat{p}^{(E)}_{n+1}$ is readily available from the budget forecasts of insurance companies. The premium liability estimator $\hat{p}$ is proportional to the claim counts, and proportional to the unearned premium, such that.

$$\hat{p} = \frac{\hat{n}_P}{\hat{n}_A}\hat{c}_{n+1} = \frac{u}{\hat{p}^{(E)}_{n+1}}\hat{c}_{n+1}. \tag{28}$$

Substituting (4) into (28) removes $\hat{c}_{n+1}$ and makes estimator $\hat{p}$ a function of the observed claim payments, $\hat{o}$, $\hat{n}_P$ and $\hat{n}_A$. The estimator is thus

$$\hat{p} = \frac{\hat{n}_P}{\hat{n}_A} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{k} c_{ijk} + \hat{o}}{n}. \tag{29}$$

(29), shows $\hat{p}$ as a linear function of the total observed claim payments, and also as a linear function of $\hat{o}$.

5.2 Accident Period Correlations of Sibling Sub-Triangles

The variance of the premium liability is a function of the variance of the next accident period. Denote the future incepting component as F. From first principles, the relationship between the variance of p and the variance of c_{n+1} is

$$\sigma_A^2 = \sigma_P^2 + \sigma_F^2 + 2\rho_{PF}\sigma_P\sigma_F. \tag{30}$$

As was previously demonstrated, $\rho_{PF} < 1$, and this means the standard deviation of the premium liability is not merely a pro-rata proportion of the standard deviation of the future accident period. In the special case where a portfolio of risks is stable in size, policies are written uniformly during the year and there is no claim inflation, then $E(p) = E(F)$ and $\sigma_P = \sigma_F$ and the equation for the standard deviation of the premium liability reduces to

$$\sigma_P = \frac{\sigma_A}{\sqrt{2 \times (1 + \rho_{PF})}}. \tag{31}$$

Given $0 < \rho < 1$, then the standard deviation of the premium liability is bounded, in this special case, such that

$$\frac{\sigma_{P+F}}{2} < \sigma_P < \frac{\sigma_{P+F}}{\sqrt{2}}. \tag{32}$$

Assuming a situation where there is no difference in risk characteristics between the premium liability and the future incepting component, and where the assumption of uniform size is also absent, then $E(p) \neq E(F)$ and $\sigma_P \neq \sigma_F$. In this case, the uncertainty of the accident period and the correlation are not sufficient to provide a solution. For this to be possible, the relationship between σ_P and σ_F must also be known. The general case, for all possible combinations of P and F, will be solved later in this section.

(11) described the correlation of claim payments, for accident period i and development period j, between sibling sub-triangles H and F. The correlation for the total claim payments across all development periods within an accident period can be obtained in a similar way.

$$\text{Let } m_H = \sum_{j=1}^{n} n_H \text{ and } m_F = \sum_{j=1}^{n} n_F$$

Let $\rho_{HF}^{(i)}$ be the correlation of total claim payments in accident period i

$$\begin{aligned}
\rho_{HF}^{(i)} &= \frac{Cov\left(\sum_{j=1}^{n}\sum_{k=1}^{n_H} c_{Hijk}, \sum_{j=1}^{n}\sum_{k=1}^{n_F} c_{Fijk}\right)}{\sqrt{Var\left(\sum_{j=1}^{n}\sum_{k=1}^{n_H} c_{Hijk}\right) Var\left(\sum_{j=1}^{n}\sum_{k=1}^{n_F} c_{Fijk}\right)}} \\
&= \frac{Cov\left(\sum_{j=1}^{n}\sum_{k=1}^{n_H} \alpha \times \nu_i, \sum_{j=1}^{n}\sum_{k=1}^{n_F} \alpha \times \nu_i\right)}{\sqrt{\left(\alpha^2 m_H^2 + \beta^2 m_H\right)\left(\alpha^2 m_F^2 + \beta^2 m_F\right)}} \\
&= \frac{Cov\left(\alpha \times \nu_i \times \sum_{j=1}^{n} n_H, \alpha \times \nu_i \times \sum_{j=1}^{n} n_F\right)}{\sqrt{\left(\alpha^2 m_H^2 + \beta^2 m_H\right)\left(\alpha^2 m_F^2 + \beta^2 m_F\right)}} \\
&= \frac{\alpha^2 \times m_H \times m_F Cov\left(\nu_i, \nu_i\right)}{\sqrt{\left(\alpha^2 n_H^2 + \beta^2 n_H\right)\left(\alpha^2 n_F^2 + \beta^2 n_F\right)}}
\end{aligned}$$

$$= \frac{\alpha \times m_H}{\sqrt{\alpha^2 m_H^2 + \beta^2 m_H}} \times \frac{\alpha \times m_F}{\sqrt{\alpha^2 m_F^2 + \beta^2 m_F}}. \tag{33}$$

The form of (33) is identical to (11), with substitution of claim counts.

Proposition 7 $\frac{\alpha n_H}{\sqrt{\alpha^2 n_H^2 + \beta^2 n_H}} \frac{\alpha n_F}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}}$ *is increasing for both* n_H *and* n_F*.*

Proof. $\frac{\partial}{\partial n_H}\left(\frac{\alpha n_H}{\sqrt{\alpha^2 n_H^2 + \beta^2 n_H}} \frac{\alpha n_F}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}}\right)$

$$= \frac{\frac{1}{2}\alpha^2 \beta^2 n_F n_H}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}\left(\alpha^2 n_H^2 + \beta^2 n_H\right)^{\frac{3}{2}}}.$$

All of the terms are positive because they are either squared or are claim counts.

Therefore the function is increasing against n_H.

By symmetry, the function is also increasing against n_F. ■

Proposition 8 *The correlation of claim payments for an entire accident period is greater than, or equal to, the correlation of claim payments for a single development period within the same accident period.*

Proof. $m_H \geq n_H$ because n_H is from a subset of the accident period. Similarly $m_F \geq n_F$ because n_F is also from a subset of the accident period.

Since $\frac{\alpha \times n_H}{\sqrt{\alpha^2 n_H^2 + \beta^2 n_H}} \frac{\alpha \times n_F}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}}$ is increasing against both n_H and n_F, then

$$\frac{\alpha \times m_H}{\sqrt{\alpha^2 m_H^2 + \beta^2 m_H}} \frac{\alpha \times m_F}{\sqrt{\alpha^2 m_F^2 + \beta^2 m_F}} \geq \frac{\alpha \times n_H}{\sqrt{\alpha^2 n_H^2 + \beta^2 n_H}} \frac{\alpha \times n_F}{\sqrt{\alpha^2 n_F^2 + \beta^2 n_F}}. \blacksquare$$

Since a premium liability is the sum of claim payments across all development periods within the next accident period, (33) applies to the correlation between the premium liability and the future incepting component of the next accident period. This result allows the calculation of an estimator of the variance of the premium liability.

5.3 Estimator $\hat{\sigma}_P^2$

Using the model of systemic and non-systemic risk, (11) implied the uncertainty of a subportfolio is related to its size relative to its parent portfolio.

In this case, the systemic risk component will be proportional to the relative sizes of the subportfolio and the parent portfolio, and the non-systemic risk will be proportional to the square roots of the relative sizes of the subportfolio and the parent portfolio. In the previous section, (17) gave an estimator for $\hat{\sigma}^2$ which is now applied to obtain $\hat{\sigma}_A^2$, and the results can then be combined to derive an estimator for $\hat{\sigma}_P^2$.

$$\hat{\sigma}_P^2 = \frac{\alpha^2 n_P^2 + \beta^2 n_P}{\alpha^2 n_A^2 + \beta^2 n_A} \left(\frac{1}{n} \sum_i \left(\left[\hat{D}_i \right]^2 + \hat{\sigma}_i^2 \right) - \left(\frac{1}{n} \sum_i \hat{D}_i \right)^2 \right). \tag{34}$$

(34) defines the estimator $\hat{\sigma}_P^2$ as a proportion of $\hat{\sigma}^2$, allowing for the systemic and non-systemic risk characteristics shown in (19). The proportion is not linear, and has the advantages of being non-negative, and asymptotically approaching 1 as n_P approaches n_A.

5.4 Correlations of Portfolio Partitions to a Parent Portfolio

As was demonstration in section 2 of this paper, the next accident period can be partitioned into the premium liability and a future incepting component. (33) gave the correlation of accident year claim payment totals for two partitions of a portfolio. Since the next accident period is the total of the premium liability and the future incepting component, the next accident period is also correlated to the premium liability. From first principles,

$$\begin{aligned}
Corr\left(\hat{p}, \hat{c}_{n+1}\right) &= Corr\left(\sum_{i=n+1} \sum_{j=1}^{n} \sum_{k \in P} c_{ijk}, \sum_{i=n+1} \sum_{j=1}^{n} \sum_{k} c_{ijk} \right) \\
&= \frac{Cov\left(\sum_{i=n+1} \sum_{j=1}^{n} \sum_{k \in P} c_{ijk}, \sum_{i=n+1} \sum_{j=1}^{n} \sum_{k} c_{ijk} \right)}{\sqrt{Var\left(\sum_{i=n+1} \sum_{j=1}^{n} \sum_{k \in P} c_{ijk} \right)} \sqrt{Var\left(\sum_{i=n+1} \sum_{j=1}^{n} \sum_{k} c_{ijk} \right)}} \\
&= \frac{Cov\left(\alpha n_P \nu_i, \alpha n_A \nu_i \right)}{\sqrt{\alpha^2 n_P^2 + \beta^2 n_P} \sqrt{\alpha^2 n_A^2 + \beta^2 n_A}} \\
&= \frac{\alpha^2 n_P n_A Cov\left(\nu_i, \nu_i \right)}{\sqrt{\alpha^2 n_P^2 + \beta^2 n_P} \sqrt{\alpha^2 n_A^2 + \beta^2 n_A}}
\end{aligned}$$

$$= \frac{\alpha^2 n_P n_A}{\sqrt{\alpha^2 n_P^2 + \beta^2 n_P}\sqrt{\alpha^2 n_A^2 + \beta^2 n_A}}. \tag{35}$$

(35) derives the correlation between the premium liability and the next accident period. This equation has the advantage of being non-negative, and asymptotically approaching 1 as n_P approaches n_A.

5.5 Correlation of $\hat{o}$ and $\hat{p}$

(2) has one common error term, operating within, and not between, accident periods, and thus accident periods are assumed to be statistically independent. Therefore the next accident period is statistically independent of historical accident periods, and thus the outstanding claims and premium liability are statistically independent. However, the estimators are not statistically independent. As was demonstrated above, the estimator of the outstanding claims and the estimator of the next accident period are positively correlated. Furthermore, it was demonstrated that the premium liability is a partition of, and positively correlated to, the next accident period.

Insurance companies face the risk their claim provisions will prove inadequate. Since claim provisions are based on estimates, the insurers' risk includes not only the uncertainty of claim outcomes, but also the uncertainty of the estimates upon which the provisions were based. When the claim liability estimates are not independent between the outstanding claims and the premium liability, then the risk of inadequate claims provisions is also not independent. In such a case $Var\left(o + p - (\hat{o} + \hat{p})\right) > Var\left(o + p\right)$ and, when this happens, insurance companies will not receive full diversification benefits between outstanding claims and the premium liability.

The imperfect correlation between the premium liability and the next accident period is due to non-systemic risk, which is not correlated with the outstanding claims. Therefore the correlation of the premium liability estimate with the outstanding claims estimate is the product of the correlations of $\hat{o}$ and $\hat{c}_{n+1}$, and $\hat{c}_{n+1}$ and $\hat{p}$, as follows:

$$\begin{aligned} Corr\left(\hat{o}, \hat{p}\right) &= Corr\left(\hat{o}, \hat{c}_{n+1}\right) \times Corr\left(\hat{c}_{n+1}, \hat{p}\right) \\ &= \frac{\sum_{i=1}^{n} z_i \left(x_i + z_i\right)\left(\alpha^2 + \frac{\beta^2}{x_i}\right)}{\sqrt{\left(\sum_{i=1}^{n} z_i^2 \left(\alpha^2 + \frac{\beta^2}{x_i}\right)\right) \times \left(\sum_{i=1}^{n} \left(x_i + z_i\right)^2 \left(\alpha^2 + \frac{\beta^2}{x_i}\right)\right)}} \end{aligned}$$

$$\times \frac{\alpha^2 n_P n_A}{\sqrt{\alpha^2 n_P^2 + \beta^2 n_P}\sqrt{\alpha^2 n_A^2 + \beta^2 n_A}}. \quad (36)$$

(36) gives the correlation of $\hat{o}$ and $\hat{p}$. The correlation is a function of the claim counts and the systemic and non-systemic risk parameters.

6 Estimation of the Model Parameters

The estimators derived in the previous sections are functions of α and β, the systemic and non-systemic risk parameters. Therefore estimates of α and β are required. This section discusses two techniques for obtaining estimates of α and β.

6.1 Maximum Likelihood Estimates for Normally Distributed Errors

The α and β parameters can be estimated via maximum likelihood methods. Assume c_{ij} is normally distributed and solve for the maximum likelihood, given the variance defined in (10). The maximum likelihood estimates are:

$$\begin{aligned}
L\left(C_{ij}, \alpha, \beta\right) &= \prod_{i=1}^{n} \prod_{j=1}^{n-i} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{c_{ij}-\mu_{ij}}{\sigma}\right)^2}. \\
l\left(\alpha, \beta\right) &= \sum_{i=1}^{n} \sum_{j=1}^{n-i} \left(-\ln\left(\sqrt{2\pi}\right) - \ln\left(\sigma\right) - \frac{1}{2}\left(\frac{c_{ij}-\mu_{ij}}{\sigma}\right)^2 \right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n-i} (-\ln(\sqrt{2\pi}) - \ln(\sqrt{n_{ij}^2\alpha^2 + n_{ij}\beta^2}) - \frac{\left(c_{ij} - n_{ij}\mu\right)^2}{2\left(n_{ij}^2\alpha^2 + n_{ij}\beta^2\right)}). \\
\frac{\partial}{\partial\alpha} l\left(\alpha, \beta\right) &= \sum_{i=1}^{n} \sum_{j=1}^{n-i} \frac{\partial}{\partial\alpha}(-\ln(\sqrt{2\pi}) - \ln(\sqrt{n_{ij}^2\alpha^2 + n_{ij}\beta^2}) - \frac{\left(c_{ij} - n_{ij}\mu\right)^2}{2(n_{ij}^2\alpha^2 + n_{ij}\beta^2)}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n-i} -\frac{\alpha}{\left(n_{ij}\alpha^2 + \beta^2\right)^2}(\alpha^2 n_{ij}^2 + \beta^2 n_{ij} - (c_{ij} - n_{ij}\mu)^2). \quad (37)
\end{aligned}$$

$$\frac{\partial}{\partial\beta}l\left(\alpha,\beta\right)=\sum_{i=1}^{n}\sum_{j=1}^{n-i}\frac{\partial}{\partial\beta}(-\ln(\sqrt{2\pi})-\ln(\sqrt{n_{ij}^2\alpha^2+n_{ij}\beta^2})-\frac{(c_{ij}-n_{ij}\mu)^2}{2(n_{ij}^2\alpha^2+n_{ij}\beta^2)})$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n-i}-\frac{\beta}{n_{ij}\left(n_{ij}\alpha^2+\beta^2\right)^2}(\alpha^2n_{ij}^2+\beta^2n_{ij}-(c_{ij}-n_{ij}\mu)^2). \tag{38}$$

Since (37) and (38) are not linear in α or β, linear methods are not available for solving the maximum likelihood estimate, and numerical techniques are required. Here, Newton-Raphson iteration has been selected to obtain the estimates. While (37) and (38) provide the gradient vector, further calculus is required to obtain the components of the Hessian matrix.

$$\frac{\partial^2}{\partial\alpha^2}l\left(\alpha,\beta\right)=\sum_{i=1}^{n}\sum_{j=1}^{n-i}\frac{\partial}{\partial\alpha}-\frac{\alpha}{\left(n_{ij}\alpha^2+\beta^2\right)^2}\left(\alpha^2n_{ij}^2+\beta^2n_{ij}-(c_{ij}-n_{ij}\mu)^2\right)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n-i}-\frac{1}{\left(n_{ij}\alpha^2+\beta^2\right)^3}(-\alpha^4n_{ij}^3+3\alpha^2\mu^2n_{ij}^3-6\alpha^2\mu c_{ij}n_{ij}^2$$
$$+3\alpha^2c_{ij}^2n_{ij}+\beta^4n_{ij}-\beta^2\mu^2n_{ij}^2+2\beta^2\mu c_{ij}n_{ij}-\beta^2c_{ij}^2). \tag{39}$$

$$\frac{\partial^2}{\partial\beta^2}l\left(\alpha,\beta\right)=\sum_{i=1}^{n}\sum_{j=1}^{n-i}\frac{\partial}{\partial\beta}-\frac{\beta}{n_{ij}\left(n_{ij}\alpha^2+\beta^2\right)^2}\left(\alpha^2n_{ij}^2+\beta^2n_{ij}-(c_{ij}-n_{ij}\mu)^2\right)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n-i}\frac{1}{n_{ij}\left(n_{ij}\alpha^2+\beta^2\right)^3}(-\alpha^4n_{ij}^3+\alpha^2\mu^2n_{ij}^3-2\alpha^2\mu c_{ij}n_{ij}^2$$
$$+\alpha^2c_{ij}^2n_{ij}+\beta^4n_{ij}-3\beta^2\mu^2n_{ij}^2+6\beta^2\mu c_{ij}n_{ij}-3\beta^2c_{ij}^2). \tag{40}$$

$$\frac{\partial^2}{\partial\alpha\partial\beta}l\left(\alpha,\beta\right)=\sum_{i=1}^{n}\sum_{j=1}^{n-i}\frac{\partial}{\partial\beta}-\frac{\alpha}{\left(n_{ij}\alpha^2+\beta^2\right)^2}\left(\alpha^2n_{ij}^2+\beta^2n_{ij}-(c_{ij}-n_{ij}\mu)^2\right)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n-i}\frac{2\alpha\beta}{\left(n_{ij}\alpha^2+\beta^2\right)^3}\left(\alpha^2n_{ij}^2+\beta^2n_{ij}-2\mu^2n_{ij}^2+4\mu c_{ij}n_{ij}-2c_{ij}^2\right). \tag{41}$$

(39) and (40) supply the values for the top left and bottom right cells, respectively, of the Hessian matrix. The remaining two cells in the Hessian matrix both contain the same value from (41).

6.2 Least Squared Estimates

Numerical techniques are not, however,, the preferred solution because they are algebraically complex, computationally intensive and require an assumption about the type of statistical distribution. A simpler estimation method is available, using linear regression and a transformation of the data. As discussed in Section 2, the variance of runoff triangle residuals is a quadratic function of the number of claims, as shown in (10). Using this relationship, a linear relationship exists between the square of the runoff triangle residuals and the number of claims.

$$r_{ij}^2 = \alpha^2 n_{ij}^2 + \beta^2 n_{ij} + \varepsilon_{ij}.$$

Transform α^2 and β^2 to obtain a linear regression form.

$$r_{ij}^2 = a n_{ij}^2 + b n_{ij} + \varepsilon_{ij} \tag{42}$$

where :

$r_{ij}^2 = \left(c_{ij} - \mu_{ij}\right)^2$ is the square of the runoff triangle residual,

$a = \alpha^2,$

$b = \beta^2,$

ε_{ij} is an error term.

(42) is in a form allowing linear least-squared regression, giving well known closed form solutions for a and b. Estimates for α and β are obtained from the inverse transformations $\hat{\alpha} = \sqrt{\hat{a}}$ and $\hat{\beta} = \sqrt{\hat{b}}$ respectively. This result is powerful, allowing α and β to be estimated using straightforward calculations, well within the capabilities of commonly used spreadsheet packages.

7 Conclusion

This paper began by proposing a random effects model as the basis for estimating premium liabilities and correlations between premium liabilities and outstanding claims. Section 3 presented empirical evidence on the features necessary for a suitable model, including variance and correlation being increasing functions of claim count, and residual correlations less than underlying correlations of liabilities. Section 4 then proved the proposed random effects model meets the necessary requirements, and is therefore a suitable candidate. Sections 4, 5, and 6 derived further features to support this random effects model, which was shown to imply a chain ladder estimate for outstanding claims. The model outlined in this paper is therefore consistent with many commonly used estimators, such as the method proposed by Mack. Closed form solutions were derived for variance and correlation estimates, giving the model the advantage of low computational intensity, speed of calculation and easy implementation in standard spreadsheet packages.

Furthermore, the model is practical for actuaries to use, requiring no more data than existing outstanding claims techniques, and using data in the same commonly used runoff triangle format employed for outstanding claims reserving. Using just two parameters, the model is robust, requiring no more historical data stability than commonly used outstanding claim techniques. The two model parameters are estimated using linear regression, which can be implemented in standard spreadsheet packages. Using a unified model, estimates for outstanding claims and premium liabilities are consistent, and estimates of risk margins are coherent.

Finally, the random effects model meets the accounting and regulatory needs of insurers and is objective, measuring historical variances and correlations and estimating variance of correlations for future liabilities. The model is also specific and relevant, using the insurer's own data, and giving estimates which can then be used to calculate diversified risk margins and capital adequacy. Providing estimates of systemic and non-systemic risk, the model represents a way forward for calculating fair value liability values and supplies and has the potential to provide valuable input for strategic planning.

While the proposed random effects model has many advantages, opportunities still exist for further research. In particular, removing or broadening some of the model assumptions would significantly extend the model's applications. While the scope of this paper is restricted to a single line of business, insurers need a model which considers diversification benefits between lines of business. The opportunity therefore exists to unify this model with existing multi-line models. Since the chain ladder model is just one of several commonly used outstanding claims techniques, adapting the model to work with other runoff triangle approaches would also be and advantage, and increase the model's practical value. Finally, the proposed model assumes independence between accident periods, which can be unrealistic for certain lines of business. Broader application is therefore possible once the model is extended to include dependencies across accident periods.

References

Australian Accounting Standards Board, 2007. Insurance Contracts. Disclosure 42, 44.

Australian Prudential Regulatory Authority, 2006. Prudential Standard GPS 310. www.apra.gov.au.

Bateup, R., Reed, I., 2001. Research and Data Analysis Relevant to the Development of Standards and Guidelines on Liability Valuation for General Insurance. Institute of Actuaries of Australia and APRA, Tillinghast Towers-Perrin. Available from www. actuaries. asn. au.

Braun, C., 2004. The prediction error of the chain ladder method applied to correlated run-off triangles. Astin Bulletin 34, 399–424.

Brehm, P., 2002. Correlation and the aggregation of unpaid loss distributions. In: CAS Forum. pp. 1–23.

Collings, S., White, G., 2001. APRA risk margin analysis. In: Institute of Actuaries of Australia XIIIth General Insurance Seminar, Trowbridge Consulting. www.actuaries.asn.au.

Commission of the European Communities, 2007. Proposal for a Directive of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance: Solvency II SEC(2007) 870 SEC(2007) 871 .

De Jong, P., 2009a. Modeling dependence between loss triangles using copulas. In: Institute of Actuaries of Australia Biennial Convention 2009. www.actuaries.asn.au.

De Jong, P., 2009b. Presentation - Report from the Dependency Working Party. In: Institute of Actuaries of Australia Biennial Convention 2009. www.actuaries.asn.au.

De Jong, P., Heller, G., 2008. Generalized linear models for insurance data. Cambridge Univ Pr.

Gibbs, J., Hu, S., 2007. Australian Reserving Practices - An analysis of the 2006 General Insurance Claims Reserving and Risk Margins Survey. In: Institute of Actuaries of Australia. Available from www.actuaries.asn.au.

Hess, K., Schmidt, K., Zocher, M., 2006. Multivariate loss prediction in the multivariate additive model. Insurance: Mathematics and Economics 39 (2), 185–191.

Insurance Australia Group, 2009. Insurance australia group limited preliminary final report 30 June 2009.

International Accounting Standards Board, 2007. Preliminary views on insurance contracts - frequently asked questions: service margins. www.iasb.org.

International Accounting Standards Board, 2008. International financial reporting standard 4: Insurance contracts. www.iasb.org.

Jones, A., Copeman, P., Gibson, E., Line, N., Lowe, J., Martin, P., Matthews, P., Powell, D., 2006. Report of the general insurance reserving issues taskforce GRIT. British Actuarial Journal 12 (3), 435–599.

Mack, T., 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bulletin 23 (2), 213–225.

Marshall, K., Collings, S., Hodson, M., O'Dowd, C., 2008. A Framework for Assessing Risk Margins. In: Institute of Actuaries of Australia XVIth General Insurance Seminar. Available from www.actuaries.asn.au.

Merz, M., Wuthrich, M., 2009. Prediction error of the multivariate additive loss reserving method for dependent lines of business. Variance 3 (1), 131–151.

Ohlsson, E., Lauzeningks, J., 2009. The one-year non-life insurance risk. Insurance: Mathematics and Economics 45 (2), 203–208.

Searle, S., Casella, G., McCulloch, C., 2006. Variance Components. Wiley-Interscience.