Variance and Covariance Due to Inflation

David R. Clark, FCAS

Abstract

Motivation. This paper looks at the problem of measuring correlation between reserve segments. The research was motivated by the 2005 CAS Working Party on Reserve Variability.

Method. Using a random-walk time series model for inflation, we can estimate the variance of a stream of inflation-sensitive payments. The same calculations can be performed to estimate the covariance between two streams of payments.

Results. Formulas are presented for estimating and calculating the variance in reserves attributable to inflation. All of these calculations are performed analytically, without requiring simulation.

Conclusions. Covariance between reserve segments due to common sensitivity to inflation can be easily modeled. This provides a convenient and intuitive way of calculating dependence between reserve segments in order to estimate variance at a company level.

Availability. Excel spreadsheet examples of the calculations described in this paper are available from the author.

Keywords. Inflation, Reserving, Time-Series, Correlation, Covariance

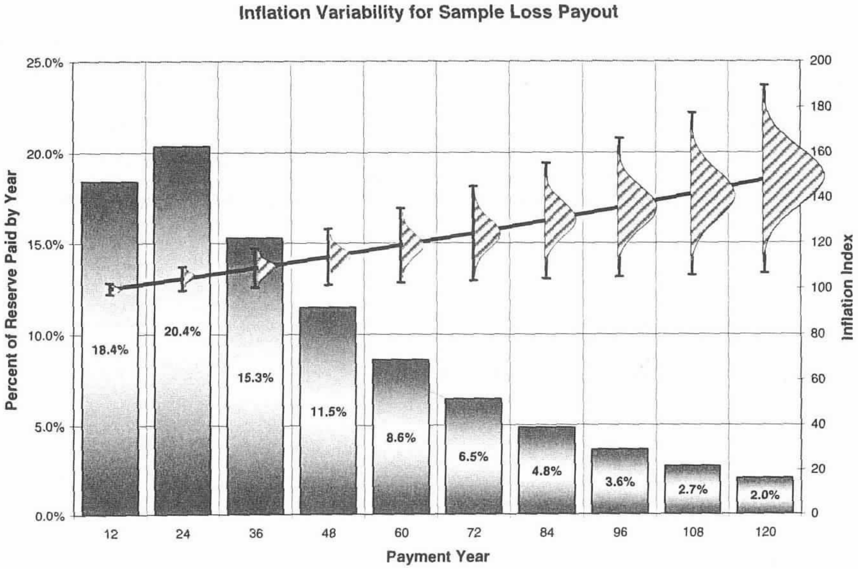
1. INTRODUCTION

This paper addresses the question of how to estimate the correlation between the future payments in two or more different reserve segments.

The motivation for this paper was the Working Party on Reserve Variability [6], which outlined the many current approaches for estimating variability for a single reserve segment – typically based on a single development triangle. An area of research identified by the Working Party was the question of correlation between two or more reserve segments.

The approach that we will follow for evaluating correlation will be based on first principles about one of the underlying causes of correlation. That is, we begin by asking why we think that there is a correlation structure that needs to be considered. From first principles, we know that inflation has an impact on the amount of loss dollars to be paid, and that different reserve segments may be affected by the same inflation index. For example, a medical claim for an injured worker and a bodily injury claim under Auto Liability may both be dependent upon a common medical inflation driver.

This basic concept is illustrated in the graph below. The bars represent a forecast of loss payments over a ten year time horizon; the line represents the "expected" inflation index built into the forecasted payment stream. If we know the variability in the inflation index (represented by the bell curves), then we can calculate the variance of the future loss payments due to inflation[1].



As the bell curves around the inflation index illustrate, the variance due to inflation increases for longer time horizons. The uncertainty in the estimate of a loss payment ten years in the future is greater than the uncertainty in the estimate of a loss payment one year in the future.

The extension to correlation then follows. If we know that two or more reserve segments are affected by the same inflation index, then we know that they will be correlated with each other.

---

[1] This concept is not new: see the papers by Taylor [5], Hodes et al [4], or Brehm [2] listed in the references.

The question then turns to the source of the inflation index used in this variance calculation. The inflation index should ideally be extracted from the insurance loss data itself, but in practice insurance data is rarely stable enough to provide a reliable estimate. A reasonable alternative is to use an external source for the inflation index.

We will follow the inflation model as outlined in the research work commissioned by the Casualty Actuarial Society (see [1]). This research assumes that inflation follows a mean-reverting random walk. Briefly, this means that the inflation rate in one year is dependent on the inflation rate in the prior year, but that it will eventually "revert" to a long-run average inflation rate. More informally, a mean-reverting model allows us to talk about *periods* of high or low inflation rather than just individual years being higher or lower than average.

Because we are limiting the discussion to the variance and covariance due to inflation, we are able to produce closed-form solutions for all of the variance and covariance terms. All of this can alternatively be incorporated into a larger simulation model if that is preferred.

After describing the basic model of inflation variability (section 2) and the formulas for variance and covariance of the reserve segments (section 3), we will look at a method for refining the calculation to include different sensitivities to inflation by reserve segment (section 4), and then finally how to integrate variance due to inflation with variance from other sources (section 5).

2. BASIC MODEL

We assume that loss inflation rates follow a mean-reverting time series model. This is described using an autoregressive AR(1) model.

$$X_t = \mu \cdot (1-r) + X_{t-1} \cdot r + e_t$$

X_t — logarithm of $1+i_t$ (i_t= the inflation rate at time t)

μ — logarithm of the 1+long-term average inflation rate i

r — factor representing the strength of the reversion (or "persistence")
$r = 0$ would be a pure "random draw" model
$r = 1$ would be a pure "random walk" model

e_t — normally distributed error term, with variance σ^2

Because the model can be transformed into a linear relationship, the parameters can be calculated easily with linear regression.

If we select, for example, a component of the consumer price index (CPI), then the variables are:

Independent Variable (X_{t-1}): $\ln\left(\frac{CPI(2)}{CPI(1)}\right), \ln\left(\frac{CPI(3)}{CPI(2)}\right), \cdots, \ln\left(\frac{CPI(n-1)}{CPI(n-2)}\right)$

Dependent Variable (X_t): $\ln\left(\frac{CPI(3)}{CPI(2)}\right), \ln\left(\frac{CPI(4)}{CPI(3)}\right), \cdots, \ln\left(\frac{CPI(n)}{CPI(n-1)}\right)$

The slope of the regression line is the parameter *r*. We can estimate the long-run average inflation rate by the intercept/(1-*r*), though we will see that the magnitude of this average does not affect our variability calculations.

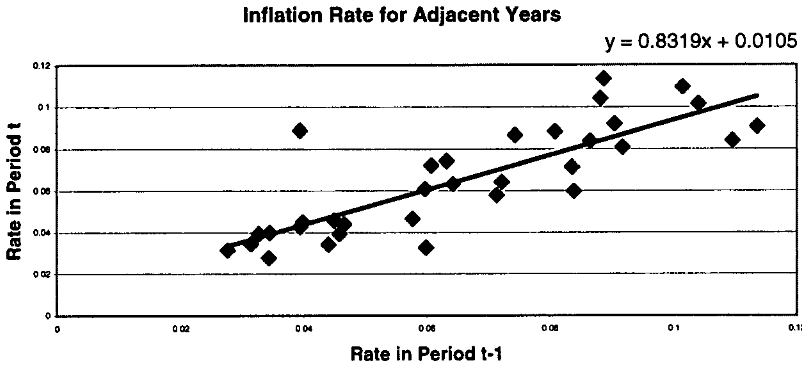
The standard error of the regression (the average deviation of the actual dependent variables from the values predicted by the fitted line) is our estimate of sigma, σ.

We will illustrate this calculation using the medical component of the CPI, though the reserving actuary is free to use any loss-inflation index deemed appropriate. Table 1 below shows this calculation based on data available through the Bureau of Labor Statistics. We calculate the logarithms of changes in the CPI, and then perform a simple linear regression on the X_t and X_{t-1} columns.

This data is, of course, meant purely for illustration and the analyst should decide carefully as to what external inflation index is most representative for the losses to be paid.

Table 1

Year (t)	CPI	Inflation %	X_t	X_{t-1}			
1970	34.0						
1971	36.1	6.18%					
1972	37.3	3.32%	0.0327	0.059932			
1973	38.8	4.02%	0.039427	0.0327			
1974	42.4	9.28%	0.088728	0.039427	X = ln(1+ Inflation %)		
1975	47.5	12.03%	0.113581	0.088728			
1976	52.0	9.47%	0.090514	0.113581			
1977	57.0	9.62%	0.091808	0.090514	Slope	0.831857	r
1978	61.8	8.42%	0.080852	0.091808	Intercept	0.010527	μ*(1-r)
1979	67.5	9.22%	0.088224	0.080852			
1980	74.9	10.96%	0.104026	0.088224	Long-Term	0.062605	μ
1981	82.9	10.68%	0.101481	0.104026			
1982	92.5	11.58%	0.109574	0.101481	Std Error	0.014738	σ
1983	100.6	8.76%	0.083944	0.109574			
1984	106.8	6.16%	0.059806	0.083944			
1985	113.5	6.27%	0.060845	0.059806			
1986	122.0	7.49%	0.072218	0.060845			
1987	130.1	6.64%	0.064282	0.072218			
1988	138.6	6.53%	0.063289	0.064282			
1989	149.3	7.72%	0.074366	0.063289			
1990	162.8	9.04%	0.086565	0.074366			
1991	177.0	8.72%	0.083627	0.086565			
1992	190.1	7.40%	0.071401	0.083627			
1993	201.4	5.94%	0.057743	0.071401			
1994	211.0	4.77%	0.046565	0.057743			
1995	220.5	4.50%	0.04404	0.046565			
1996	228.2	3.49%	0.034325	0.04404			
1997	234.6	2.80%	0.027659	0.034325			
1998	242.1	3.20%	0.031469	0.027659			
1999	250.6	3.51%	0.034507	0.031469			
2000	260.8	4.07%	0.039896	0.034507			
2001	272.8	4.60%	0.044985	0.039896			
2002	285.6	4.69%	0.045853	0.044985			
2003	297.1	4.03%	0.039477	0.045853			
2004	310.1	4.38%	0.042826	0.039477			



3. CALCULATING THE VARIANCE OF PAYMENTS

We proceed by showing the calculation of variance for a single payment and then building the model step-by-step up to the covariance between two streams of payments.

3.1 Calculating the Variance of a Single Payment

A one year inflation factor $(1+i_t)$ is lognormally distributed, which means that a loss payment one year in the future – if unaffected by random factors other than inflation – would also be lognormally distributed.

With no "mean reversion" (r =0), the coefficient of variation, CV, of the loss payment would be $\sqrt{\exp(\sigma^2)-1}$. An inflation factor two years out $CPI(2)=(1+i_1)\cdot(1+i_2)$ would also be lognormally distributed, but the CV would increase to $\sqrt{\exp(2\cdot\sigma^2)-1}$.

The simplicity of this expression is due to the assumption that i_1 and i_2 are independent and identically distributed, and also the fact that the product of two lognormal random variables is also a lognormal random variable.

If we introduce the concept of mean reversion such that $r > 0$, then the formula for the CV of the single year factor does not change, but the two-year inflation index CPI(2) becomes:

$$\text{CPI(2)} = (1+i_1)\cdot\left(\frac{(1+i_1)}{E[1+i_1]}\right)^r\cdot(1+i_2).$$

The $\text{CV}_{n=2}$ increases to become $\sqrt{\exp\{(1+(1+r)^2)\cdot\sigma^2\}-1}$.

The index for subsequent years is created in a similar manner. For n=3, we have

$$\text{CPI(3)} = (1+i_1)\cdot\left(\frac{(1+i_1)}{E[1+i_1]}\right)^r\cdot(1+i_2)\cdot\left(\left(\frac{(1+i_1)}{E[1+i_1]}\right)^r\cdot\frac{(1+i_2)}{E[1+i_2]}\right)^r\cdot(1+i_3).$$

The $\text{CV}_{n=3}$ becomes $\sqrt{\exp\{(1+(1+r)^2+(1+r+r^2)^2)\cdot\sigma^2\}-1}$.

In the special case in which $r=1$, we have a $\text{CV}_{n=3}$ of $\sqrt{\exp\{(1+2^2+3^2)\cdot\sigma^2\}-1}$.

More generally, the CV for *n* years of inflation is given by:

$$CV_n = \sqrt{\exp\{n\cdot\sigma^2\}-1} \qquad \text{for } r = 0.$$

$$CV_n = \sqrt{\exp\left\{\left(\frac{n}{(1-r)^2}-\frac{2\cdot r\cdot(1-r^n)}{(1-r)^3}+\frac{r^2\cdot(1-r^{2n})}{(1-r)^2\cdot(1-r^2)}\right)\cdot\sigma^2\right\}-1} \qquad \text{for } r<1$$

or, alternatively

$$CV_n = \sqrt{\exp\left\{\frac{n\cdot(n+1)\cdot(2n+1)}{6}\cdot\sigma^2\right\}-1} \qquad \text{for } r=1$$

A more detailed derivation of these formulas is given in Appendix A.

We note that when the reversion term r is close to 1, changes in the inflation rate are "persistent," meaning that the inflation level will not return to its long-run average very quickly. In these cases, the variance of a loss payment in the distant future will have a much greater variance than under the "random draw" model with $r = 0$.

The table below shows the CV implied for a single payment at various points in the future using different assumptions about the reversion parameter r.

		Sigma =	0.024996	
	CV_n for Selected Reversion Parameters			
n	r = 0	r = .50	r = .80	r = 1
1	0.0250	0.0250	0.0250	0.0250
2	0.0354	0.0451	0.0515	0.0559
3	0.0433	0.0629	0.0799	0.0937
4	0.0500	0.0785	0.1090	0.1376
5	0.0559	0.0923	0.1380	0.1870

3.2 Calculating the Covariance Between Two Payments

Suppose that we have an inflation factor for a given number of years n, and a second factor for $n+k$. We quickly recognize that there must be a strong correlation since n of the $n+k$ years are common to both factors. Using the same mean reversion model, the correlation coefficient can be written[2]:

---

[2] The term $Cov_{n,k}$ is a "scaled" value which is the dollars of covariance divided by the means of the losses at times n and $n+k$. This is sometimes called the "coefficient of covariation" and is convenient notation because of the parallel to the coefficient of variation (CV) used earlier.

$$\rho_{n.n+k} \quad = \quad \frac{Cov_{n.k}}{CV_n \cdot CV_{n+k}}.$$

The term in the numerator is proportional to the covariance, and is given as follows:

$$Cov_{n.k} \quad = \quad \exp\left\{\left(\frac{n}{(1-r)^2} - \frac{r\cdot(1+r^k)\cdot(1-r^n)}{(1-r)^3} + \frac{r^{k+2}\cdot(1-r^{2n})}{(1-r)^2\cdot(1-r^2)}\right)\cdot\sigma^2\right\} - 1 \qquad \text{for } r < 1$$

or, alternatively

$$Cov_{n.k} \quad = \quad \exp\left\{\frac{n\cdot(n+1)}{2}\left(\frac{2n+1}{3}+k\right)\cdot\sigma^2\right\} - 1 \qquad \text{for } r = 1$$

Note also that $Cov_{n.k} = CV_n^2$ when k =0.

Sigma = 0.025000
Reversion = 0.500000

Matrix of Correlation Coefficients

	1	2	3	4	5
1	1	0.83188775	0.69611104	0.59742763	0.52484632
2	0.83188775	1	0.91052622	0.80678419	0.71882568
3	0.69611104	0.91052622	1	0.94009581	0.85838825
4	0.59742763	0.80678419	0.94009581	1	0.95526523
5	0.52484632	0.71882568	0.85838825	0.95526523	1

3.3 Calculating the Variance of a Stream of Payments

Given these terms, we are able to set up a matrix of correlation coefficients, or covariances, in order to calculate the variance for a sum of payments. The full correlation structure between the individual payments due to inflation is captured in this matrix.

If we have a vector of *N* loss payments, $\vec{P}$, and an *N*-by-*N* matrix of covariance terms such that $M(i,j) = Cov_{i,j-i}$ for $i < j$, then we can calculate the variance for the stream of payments as:

$$Var(P) \quad = \quad \vec{P} \cdot M \cdot \vec{P}^T \qquad\qquad P = \text{sum of all payments in the vector}$$

Or equivalently,

$$Var(P) \quad = \quad \sum_{i=1}^{N}\sum_{j=1}^{N} P(i) \cdot M(i,j) \cdot P(j)$$

3.4 Calculating the Covariance Between Two Streams of Payments

If we have two vectors of loss payments ${}_A\vec{P}$ and ${}_B\vec{P}$, both with *N* elements, then the covariance of the two sums can be calculated in a similar manner.

$$Cov({}_AP, {}_BP) \quad = \quad {}_A\vec{P} \cdot M \cdot {}_B\vec{P}^T.$$

The correlation between the two payouts will be a single number, and generally a number approaching 1.000, indicating a very strong correlation. This is because our model assumes that both payment streams are directly affected by the inflation rate, and that inflation is the only source of variability. In Section 4, we soften the first assumption by allowing different degrees of sensitivity to inflation by line of business. In Section 5, we show how to bring in other sources of variability.

4. MEASURING THE SIGNIFICANCE OF INFLATION BY SEGMENT

As mentioned above, the variance/covariance model assumes that the CPI directly affects the amount of loss payment. This may not be exactly true, and we would want the ability to control the degree to which loss development is dependent on inflation.

The degree of inflation for a given risk class (RC) will be controlled by a parameter ${}_{RC}\gamma$, which is applied as an exponent to the CPI. This parameter could be set equal to zero for the cases in which a risk class is unaffected by inflation.

Adjusted Inflation Index for Risk Class A: $CPI^{{}_{A}\gamma}$

In calculating the time-series parameters for this adjusted index, the reversion parameter r is unchanged regardless of the γ; the sigma will change to become $\sigma \to \gamma \cdot \sigma$. This adjustment is easily incorporated into the CV calculation.

$$CV_n = \sqrt{\exp\left\{\left(\frac{n}{(1-r)^2} - \frac{2 \cdot r \cdot (1-r^n)}{(1-r)^3} + \frac{r^2 \cdot (1-r^{2n})}{(1-r)^2 \cdot (1-r^2)}\right) \cdot \gamma^2 \cdot \sigma^2\right\} - 1}$$

Similarly, the covariance term, when there are two risk classes, A and B, with different degrees of dependence on inflation, is modified as below:

$$Cov_{n,k} = \exp\left\{\left(\frac{n}{(1-r)^2} - \frac{r \cdot (1+r^k) \cdot (1-r^n)}{(1-r)^3} + \frac{r^{k+2} \cdot (1-r^{2n})}{(1-r)^2 \cdot (1-r^2)}\right) \cdot {}_{A}\gamma \cdot {}_{B}\gamma \cdot \sigma^2\right\} - 1$$

We note that this expression is the same as the earlier calculation when ${}_A\gamma = {}_B\gamma = 1$, and the covariance is zero when either ${}_A\gamma$ or ${}_B\gamma$ is zero.

The next question to address is the method for estimating the parameter γ for a given business segment. We begin by defining a simple model for loss payments from a triangle. The formulas below give a model ignoring inflation:

$$c_{y,d} \approx \alpha_y \cdot \beta_d$$

Where $c_{y,d}$ = incremental loss paid in accident year y and development period d. For example, $c_{1999,3}$ would be the amount paid for accident year 1999 between 24 and 36 months.

α_y = a measure of exposure for accident year y, such as onlevel premium. This can be supplied from external sources or be estimated from the triangle itself.

β_d = a parameter representing the amount of development in development period d.

This model is introduced for simplicity only. When we combine this simple two factor (AY and development period) model with an assumption that incremental payments follow an over-dispersed Poisson distribution, then the results match an all-year weighted average chain-ladder calculation.

In order to include an inflation index in this model, we expand the expression with a term including a CPI curve.

$$c^*_{y,d} \approx \alpha_y \cdot \beta_d \cdot CPI(y+d-1)^{\gamma}$$

From this expanded model, we immediately notice that the no-inflation model is a special case when $\gamma = 0$, so that $c_{y,d} = c^*_{y,d}$. If payments are directly proportional to inflation, then we would expect $\gamma = 1$; and if we expect a "leveraged" effect of inflation (say, in excess layers) then $\gamma > 1$.

Given an explicit model, as above, we are then able to estimate the parameter γ that maximizes a likelihood function or minimizes some other error function. We also have available the goodness-of-fit statistics to test the value of including inflation.

To illustrate, we will work with a small triangle of [cumulative] paid data:

AY	1	2	3	4	5	6	7
1998	13,822	26,045	34,915	41,064	45,228	47,942	49,730
1999	13,710	27,104	36,777	43,309	47,266	49,501	
2000	14,409	28,805	38,328	44,772	49,022		
2001	15,120	28,945	38,692	45,169			
2002	13,344	25,970	34,922				
2003	13,506	25,926					
2004	14,765						

The incremental paid losses from this triangle are then given by:

AY	1	2	3	4	5	6	7
1998	13,822	12,223	8,870	6,149	4,164	2,714	1,788
1999	13,710	13,394	9,673	6,532	3,957	2,235	
2000	14,409	14,396	9,523	6,444	4,250		
2001	15,120	13,825	9,747	6,477			
2002	13,344	12,626	8,952				
2003	13,506	12,420					
2004	14,765						

Based on maximum likelihood estimation[3], we have the following fitted parameters:

y	α_y	d	β_d
1998	49,730	1	0.2737
1999	51,347	2	0.2573
2000	53,571	3	0.1814
2001	54,089	4	0.1227
2002	49,018	5	0.0800
2003	48,824	6	0.0490
2004	53,946	7	0.0360

These fitted values are equivalent to calculating the α's as the chain-ladder ultimates. The fitted values from this model, corresponding to the actual incremental payments, are shown in the triangle below.

AY	1	2	3	4	5	6	7
1998	13,611	12,796	9,023	6,099	3,978	2,435	1,788
1999	14,054	13,212	9,316	6,298	4,107	2,514	
2000	14,662	13,784	9,720	6,571	4,285		
2001	14,804	13,917	9,813	6,634			
2002	13,416	12,612	8,893				
2003	13,363	12,563					
2004	14,765						

The model is then expanded for the inflation adjustment.

y	α_y	d	β_d	CPI	Index	γ	Index^{γ}
1998	49,730	1	0.2761	242.1	1.000	1.655	1.000
1999	48,043	2	0.2424	250.6	1.035		1.059
2000	46,709	3	0.1593	260.8	1.077		1.131
2001	43,867	4	0.1003	272.8	1.127		1.218
2002	37,028	5	0.0609	285.6	1.180		1.315
2003	34,448	6	0.0348	297.1	1.227		1.403
2004	35,499	7	0.0239	310.1	1.281		1.506

With the fitted values including this inflation parameter are as follows:

AY	1	2	3	4	5	6	7
1998	13,732	12,764	8,962	6,075	3,981	2,430	1,788
1999	14,046	13,172	9,327	6,331	4,106	2,519	
2000	14,587	13,796	9,783	6,571	4,285		
2001	14,759	13,978	9,808	6,625			
2002	13,440	12,595	8,887				
2003	13,348	12,578					
2004	14,765						

For example, the first development period for AY 2003 has a fitted value equal to:

$$13{,}348 = 34{,}448 \times .2761 \times 1.403.$$

[3] For this calculation, we will assume that each cell follows an Over-Dispersed Poisson (ODP) distribution with a common variance/mean ratio ϕ. Appendix A gives the full details of this model.

The parameter value 1.655 acts as a "leveraging" effect on inflation, meaning that payments increase at a faster rate than the CPI would indicate. However, in this example, as with most real data sets, there is significant uncertainty in the estimate of the γ parameter. The loss development triangle simply is not a sufficient base for estimating it credibly. Informally, the reason for this is that we can pick almost any value for γ and then fit α_y and β_d vectors that reasonably approximate the historical loss development (see Appendix B for further insight as to why this is the case). It is for this reason that we recommend that the γ parameter be selected by the model user rather than via a fitted model.

The example given above shows that the parameter γ, for measuring the sensitivity to inflation, often lacks great predictive value, that is $\hat{c}^*_{y,d}$ is not much better than $\hat{c}_{y,d}$. This suggests that the use of an external inflation index in calculating variability needs to be justified on *a priori* theoretical grounds and not solely on statistical tests. As a starting assumption, $\gamma = 1$ for each risk class is most reasonable.

The difficulty in estimating the parameter γ does not mean that losses are unaffected by inflation, but merely that a triangle format is not a sufficient basis to discern what the relationship to inflation is.

5. COMBINING OTHER SOURCES OF VARIABILITY

The discussion to this point has been limited to the variability strictly due to inflation. Naturally the variability of loss payments is driven by many other sources, and we need to be able to combine these different sources into a single calculation. Some of these other sources would include:

- Changes in an injured person's condition (recovery, deterioration, death)
- Newly reported claims not originally in the triangle ("true" IBNR)
- Legal or regulatory changes impacting the coverage provided in the insurance policy

These types of variability are, arguably, independent of changes in the rate of inflation and can therefore be treated as statistically independent.

The most common method for including all types of variability is through the use of a large simulation model; however, that is not necessary if we are interested just in the means and variances of the payments.

Section 5 will follow the same logic as Section 3, by starting with a single payment and then showing step-by-step how the calculations are generalized to produce a full covariance matrix on payment streams.

5.1 Calculating the Variance of a Single Payment

Suppose that we have a random variable for the payment amount at a specific time t, and denote this expected amount C_t. The timing of the payment is known with certainty, and we have an estimate of its mean $E(C_t)$ and variance $Var(C_t)$ from sources other than inflation. These values may have come from a stochastic reserving model, or may have been simply selected by a reserving actuary.

The next step is to assume that we have an estimate of the inflation index at time t, based on the equations from Sections 3 and 4 above.

$$CV_t = \sqrt{\exp\left\{\left(\frac{t}{(1-r)^2} - \frac{2\cdot r\cdot(1-r^t)}{(1-r)^3} + \frac{r^2\cdot(1-r^{2t})}{(1-r)^2\cdot(1-r^2)}\right)\cdot\gamma^2\cdot\sigma^2\right\} - 1}$$

The inflation index will be represented by a second random variable b_t, with a mean of one $E(b_t)=1$ and a variance of $Var(b_t)=CV_t^2$. We make the further assumption that the inflation index is statistically independent of the other sources of variance in C_t.

The variance of the product of the two random variables is then calculated as follows.

$$Var(b_t\cdot C_t) = Var(b_t)\cdot Var(C_t) + Var(b_t)\cdot E(C_t)^2 + E(b_t)^2\cdot Var(C_t)$$

The derivation of this expression is given in Appendix C.

For the reader familiar with the literature of the Casualty Actuarial Society, the description to this point should not be surprising. In fact, the formulas are identical with what is usually referred to as "mixing" parameters, and the use of the notation "*b*" is a deliberate choice to be consistent with this idea.

The inflation index can be viewed as a "parameter variance" component with the total variance above regrouped as follows.

$$Var(b_t \cdot C_t) = \underbrace{Var(C_t) \cdot E(b_t^2)}_{\text{Process Variance}} + \underbrace{Var(b_t) \cdot E(C_t)^2}_{\text{Parameter Variance}}$$

5.2 Calculating the Covariance of a Two Payments

If we have two payments, taking place at different times, t and $t+k$, then the covariance between these two payments is calculated in a formula that generalizes the variance formula above.

$$\begin{aligned} Cov(b_t \cdot C_t, b_{t+k} \cdot C_{t+k}) &= Cov(b_t, b_{t+k}) \cdot Cov(C_t, C_{t+k}) \\ &+ Cov(b_t, b_{t+k}) \cdot E(C_t) \cdot E(C_{t+k}) + E(b_t) \cdot E(b_{t+k}) \cdot Cov(C_c, C_{t+k}) \end{aligned}$$

For the special case of $k=0$, this expression reduces to the variance formula above.

5.3 Calculating the Variance of a Stream of Payments

The variance of a stream of payments is a linear combination of the variance and covariance terms calculated above.

We again start with a vector of N expected loss payments, $\vec{P} = \{E(C_t)\}_{t=1}^N$. We now assume that we also know the covariance matrix from sources other than inflation, $M_C(i,j) = Cov(C_i, C_j)$.

As in Section 3.3, we also create an N-by-N matrix of covariance terms for the inflation indices corresponding to each loss payment: $M_b(i,j) = Cov(b_i, b_j)$.

The covariance matrix, representing each pair of loss payments in the payment stream $\vec{P}_N$, is calculated by applying the formula from Section 5.2 on an element-by-element basis.

$$M_{b \cdot C}(i,j) = M_b(i,j) \cdot M_C(i,j) + M_b(i,j) \cdot E(C_i) \cdot E(C_j) + M_C(i,j)$$

The variance of the sum of all payments in the stream is then calculated as the sum of all entries in this combined matrix $M_{b \cdot C}$.

Once again, this may be viewed as a combination of a matrix of expected "process variance" and a matrix of "parameter variance" elements.

$$M_{b\cdot C}(i,j) = \underbrace{M_C(i,j)\cdot\{1+M_b(i,j)\}}_{\text{Process Variance}} + \underbrace{M_b(i,j)\cdot E(C_i)\cdot E(C_j)}_{\text{Parameter Variance}}$$

We may also note that the sum of the "parameter variance" elements is identical to what we denoted $Var(P) = \vec{P}\cdot M\cdot \vec{P}^T$ in Section 3.3.

At this point the reader may have a concern about where all of these numbers come from. The matrix of covariances related to inflation M_b is created using the formulas from Section 3, but do we really have all of the covariances from other sources needed for M_C ? It may be that these are not available and a further simplification is needed.

The easiest way to simplify this process is to include an assumption that the ultimate loss C and the variance of the ultimate loss $Var(C)$ are known. We further assume that the payment pattern on a percent basis is fixed and certain. That is, the dollar amount of ultimate loss may vary, the same percent will always be paid in the first year. By this assumption, all of the C_t payments are perfectly correlated and have the same coefficient of variation (standard deviation divided by mean) CV_C. The elements of the M_C matrix are then easily defined as follows.

$$M_C(i,j) = E(C_i)\cdot E(C_j)\cdot CV_C^2$$

The overall covariance matrix then simplifies greatly.

$$M_{b\cdot C}(i,j) = M_b(i,j)\cdot M_C(i,j) + M_b(i,j)\cdot E(C_i)\cdot E(C_j) + M_C(i,j)$$

becomes

$$M_{b\cdot C}(i,j) = \{CV_C^2 + (1+CV_C^2)\cdot M_b(i,j)\}\cdot E(C_i)\cdot E(C_j)$$

5.4 Calculating the Covariance between Two Streams of Payments

The example of how to combine the variance due to inflation with variance from other sources can now be generalized to the discussion of the covariance between two reserve risk classes such as different lines of business.

If we have two risk classes A and B, each with selected payment streams such that we create an NxN matrix of covariance terms between each of the payments. As with the single payment stream example, this can be set up as a matrix.

$$M_{AB}(i,j) = Cov\left({}_AC_i, {}_BC_j\right)$$

To combine this with the variance due to inflation, we then use the following formula.

$$M_{b\cdot AB}(i,j) = M_b(i,j)\cdot M_{AB}(i,j) + M_b(i,j)\cdot E\left({}_AC_i\right)\cdot E\left({}_BC_j\right) + M_{AB}(i,j)$$

If the two reserve segments are not correlated based on any factors other than inflation, then all the elements of this matrix are zero, and no calculations are necessary.

We may also simplify the matrix if, as in the previous section, we introduce the assumption that the percent payment pattern for each risk class is fixed and known. The matrix M_{AB} then becomes a constant amount times the cross-product of the payments. The correlation coefficient ρ_{AB} for sources other than inflation is introduced.

$$M_{AB}(i,j) = E\left({}_AC_i\right)\cdot E\left({}_BC_j\right)\cdot\left\{\rho_{AB}\cdot CV_A\cdot CV_B\right\}$$

This again leads to a simpler version of the covariance matrix.

$$M_{b\cdot AB}(i,j) = \left\{\rho_{AB}\cdot CV_A\cdot CV_B + \left(1+\rho_{AB}\cdot CV_A\cdot CV_B\right)\cdot M_b(i,j)\right\}\cdot E\left({}_AC_i\right)\cdot E\left({}_BC_j\right)$$

The covariance term between the two risk classes is the sum of all of the terms in this matrix.

The correlation coefficient $\rho_{b\cdot AB}$ (including both inflation and other sources) between these two risk classes is then calculated as follows.

$$\rho_{b\cdot AB} = \frac{sum\{M_{b.AB}\}}{\sqrt{sum\{M_{b.A}\}\cdot sum\{M_{b.B}\}}} = \frac{\rho_{AB}\cdot CV_A\cdot CV_B + \left(1+\rho_{AB}\cdot CV_A\cdot CV_B\right)\cdot\Sigma^2_{AB}}{\sqrt{\left\{CV_A^2+\left(1+CV_A^2\right)\cdot\Sigma_A^2\right\}\cdot\left\{CV_B^2+\left(1+CV_B^2\right)\cdot\Sigma_B^2\right\}}}$$

$$\begin{aligned}\text{where }\Sigma_A^2 &= sum\left\{M_b(i,j)\cdot E\left({}_AC_i\right)\cdot E\left({}_AC_j\right)\right\}/E\left({}_AC\right)^2\\ \Sigma_B^2 &= sum\left\{M_b(i,j)\cdot E\left({}_BC_i\right)\cdot E\left({}_BC_j\right)\right\}/E\left({}_BC\right)^2\\ \Sigma_{AB}^2 &= sum\left\{M_b(i,j)\cdot E\left({}_AC_i\right)\cdot E\left({}_BC_j\right)\right\}/\left\{E\left({}_AC\right)\cdot E\left({}_BC\right)\right\}\end{aligned}$$

These expressions can also be written in matrix notation.

$$\begin{aligned}\Sigma_A^2 &= Var\left({}_AP\right) = {}_A\vec{P}\cdot M_b\cdot{}_A\vec{P}^T\\ \Sigma_B^2 &= Var\left({}_BP\right) = {}_B\vec{P}\cdot M_b\cdot{}_B\vec{P}^T\\ \Sigma_{AB}^2 &= Cov\left({}_AP, {}_BP\right) = {}_A\vec{P}\cdot M_b\cdot{}_B\vec{P}^T\end{aligned}$$

In these formulas, we have included the same inflation covariance matrix M_b. However, if we include adjustment factors other than ${}_A\gamma = {}_B\gamma = 1$, then we would need to adjust the matrices as shown in Section 4.

With this formula, we are able to combine the correlation due to inflation with correlation from other sources without having to define all of the inter-dependencies between individual payments. If the user is uncomfortable with assuming that the payout patterns do not vary, then the more general formulas can be run.

6. RESULTS AND DISCUSSION

Having completed a fairly rigorous description of the formulas for calculating covariance due to inflation, it is worthwhile showing a simplified numerical example to illustrate how this can be implemented in practice.

We begin with the inflation model defined in Section 2, in which we calculated:

Reversion parameter r = .831857

Variability Sigma σ = .014738

If both reserve risk classes A and B are directly proportional to this inflation index, such that ${}_A\gamma = {}_B\gamma = 1$, then we have an inflation covariance matrix M_b as show below (each element of the matrix being one calculation of the formula in Section 3.2).

Matrix of Covariance Factors M_b

0.00022	0.00040	0.00055	0.00067	0.00078	0.00086	0.00094	0.00100	0.00105	0.00109
0.00040	0.00095	0.00140	0.00178	0.00210	0.00236	0.00258	0.00276	0.00291	0.00304
0.00055	0.00140	0.00233	0.00311	0.00375	0.00429	0.00473	0.00510	0.00541	0.00567
0.00067	0.00178	0.00311	0.00443	0.00553	0.00644	0.00720	0.00784	0.00837	0.00881
0.00078	0.00210	0.00375	0.00553	0.00722	0.00864	0.00982	0.01080	0.01161	0.01229
0.00086	0.00236	0.00429	0.00644	0.00864	0.01069	0.01240	0.01382	0.01501	0.01600
0.00094	0.00258	0.00473	0.00720	0.00982	0.01240	0.01477	0.01675	0.01840	0.01977
0.00100	0.00276	0.00510	0.00784	0.01080	0.01382	0.01675	0.01941	0.02163	0.02348
0.00105	0.00291	0.00541	0.00837	0.01161	0.01501	0.01840	0.02163	0.02456	0.02699
0.00109	0.00304	0.00567	0.00881	0.01229	0.01600	0.01977	0.02348	0.02699	0.03014

We then introduce two reserve segments, having ten year payment patterns as below.

Year	Risk Class A $_AP$	Risk Class B $_BP$
1	46.40%	15.20%
2	12.10%	11.60%
3	8.40%	10.50%
4	6.80%	10.00%
5	5.70%	9.40%
6	4.90%	9.10%
7	4.50%	8.90%
8	4.00%	8.60%
9	3.70%	8.40%
10	3.50%	8.30%

From this information, we can calculate the CVs from inflation as follows:

$$\Sigma_A^2 = Var({}_AP) = {}_A\bar{P} \cdot M_b \cdot {}_A\bar{P} = .0470^2$$

$$\Sigma_B^2 = Var({}_BP) = {}_B\bar{P} \cdot M_b \cdot {}_B\bar{P} = .0803^2$$

$$\Sigma_{AB}^2 = Cov({}_AP, {}_BP) = {}_A\bar{P} \cdot M_b \cdot {}_B\bar{P} = .0610^2$$

The correlation coefficient from inflation only is then estimated as follows.

$$\frac{\Sigma_{AB}^2}{\sqrt{\Sigma_A^2 \cdot \Sigma_B^2}} = \frac{.0610^2}{.0470 \cdot .0803} = .989$$

This very significant correlation is, again, due to the fact that inflation is the only factor contributing to the variance of either reserve risk class.

We can generalize this by including variability from other sources. We will assume that the risk classes A and B have CVs from sources other than inflation of .100 and .160 respectively, and that these are independent. Further, we will include the simplifying assumption that the ultimate losses are variable but that the percentage payout patterns are fixed. The resulting correlation coefficient, reflecting all sources of variance is given below.

$$\rho_{b \cdot AB} = \frac{\Sigma_{AB}^2}{\sqrt{\left(CV_A^2 + (1+CV_A^2)\cdot\Sigma_A^2\right)\cdot\left(CV_B^2 + (1+CV_B^2)\cdot\Sigma_B^2\right)}}$$

$$= \frac{.0610^2}{\sqrt{\left(.100^2 + (1+.100^2)\cdot .0470^2\right)\cdot\left(.160^2 + (1+.160^2)\cdot .0803^2\right)}} = .188$$

All of these numbers are meant purely for illustration purposes, but they do show that the formulas produce results in reasonable ranges.

The general process for estimating variance and covariance due to inflation can be summarized in the steps below:

- Select an external index, such as a component of the CPI
- Estimate the variance (σ^2) and reversion (r) parameters for the inflation index
- Select a default inflation-sensitivity parameter γ for each risk class
- Estimate the future loss payment stream for each risk class
- Calculate the variance of each risk class due to inflation
- Calculate the covariance between each pair of risk classes

7. CONCLUSIONS

The formulas outlined in this paper provide a very simple method for estimating the sensitivity of losses and reserves to movement in inflation rates. The advantages of this approach may be summarized as below:

1) The basic idea is very easy to explain: loss payments move with inflation
2) Variability due to inflation can be linked to economic forecast models
3) The calculation of variances and correlation can be performed in an Excel spreadsheet in closed form

The chief disadvantage that is identified is that external inflation indices, such as components of the consumer price index (CPI) have not been shown to be significant explanatory variables for movement in insurance loss amounts.

In spite of the difficulty in estimating the sensitivity parameter γ, however, we have a reasonable baseline value of $\gamma = 1$. The model therefore can provide a correlation structure between reserve risk classes based on external knowledge of inflation with a minimal need for arbitrary assumptions.

Appendix A: Derivation of Key Formulas

This appendix provides a more detailed derivation of the key variance and covariance formulas given in the body of the paper.

The formulas in this paper are able to be written in a compact form by capitalizing on a useful property of the lognormal distribution; namely that the product of lognormal random variables is again a lognormal random variable. Analogously, the sum of normal (Gaussian) random variables is again a normal random variable.

The autoregressive model, AR(1), is written in a recursive linear form, after taking the logarithms of the inflation trend factors.

$$X_t = X_{t-1} \cdot r + b + \sigma \cdot e_t$$

X_t logarithm of $1+i_t$ (i_t= the inflation rate at time t)

e_t standard normal random variable, $e_t \propto Normal(0,1)$

The distribution of X_t, conditional upon a known value for X_{t-1}, is then given as

$$X_t \mid X_{t-1} \propto Normal(X_{t-1} \cdot r + b, \sigma).$$

The variance of the conditional random variable is then

$$Var(X_t \mid X_{t-1}) = \sigma^2.$$

The random variable for the logarithm of the inflation rate two or more years out is found by expanding the recursive expression:

$$X_t \mid X_{t-2} = (X_{t-2} \cdot r + b + \sigma \cdot e_{t-1}) \cdot r + b + \sigma \cdot e_t$$

$$X_t \mid X_{t-3} = ((X_{t-3} \cdot r + b + \sigma \cdot e_{t-2}) \cdot r + b + \sigma \cdot e_{t-1}) \cdot r + b + \sigma \cdot e_t$$

This expanding of the recursive formula can be generalized as

$$X_t \mid X_0 = X_0 \cdot r^t + \sum_{i=1}^{t} (b + \sigma \cdot e_i) \cdot r^{t-i}.$$

The variance for this more general form is therefore given as below.

$$X_t \mid X_0 \propto Normal\left(X_0 \cdot r^t + b \cdot \sum_{i=1}^{t} r^{t-i}, \sigma \cdot \sqrt{\sum_{i=1}^{t} r^{2\cdot(t-i)}} \right)$$

The variance for the random variable conditional upon a point "t" years prior is then:

$$Var(X_t \mid X_0) = \sigma^2 \cdot \sum_{i=1}^{t} r^{2\cdot(t-i)} = \sigma^2 \cdot \frac{1-r^{2t}}{1-r^2} \qquad \text{if } r<1$$

or $$Var(X_t \mid X_0) = \sigma^2 \cdot t \qquad \text{if } r=1$$

These and subsequent simplifications are possible based on three fundamental identities.

$$\sum_{k=1}^{n} k = 1+2+3+\cdots+(n-1)+n = \frac{n\cdot(n+1)}{2}$$

$$\sum_{k=1}^{n} k^2 = 1^2+2^2+3^2+\cdots+(n-1)^2+n^2 = \frac{n\cdot(n+1)\cdot(2n+1)}{6}$$

$$\sum_{k=1}^{n} r^{k-1} = 1+r^1+r^2+r^3+\cdots+r^{n-2}+r^{n-1} = \frac{1-r^n}{1-r}$$

The random variable X_t represents the inflation rate "*t*" years in the future, and the expression $Var(X_t \mid X_0)$ is the variance around that rate. For our purposes, we need the variance of the inflation index at this future point; the index includes the variances of all of the annual inflation rates from the base time to the future period.

For this next step, we must remember that the inflation rate at a given point in the future is correlated with the inflation rates at subsequent points. This implies that the normal error terms e_i are included multiple times in the summation below.

$$(X_1+X_2+\cdots+X_n \mid X_0) = \sum_{j=1}^{n}(X_j \mid X_0) = \sum_{j=1}^{n}\left\{X_0\cdot r^j + \sum_{i=1}^{j}(b+\sigma\cdot e_i)\cdot r^{j-i}\right\}$$

If we make the substitution $S_n = (X_1+X_2+\cdots+X_n \mid X_0)$, then the random variable can be written more compactly as below.

$$S_n = E(S_n)+\sigma\cdot\sum_{j=1}^{n}\left\{\sum_{i=1}^{j} e_i\cdot r^{j-i}\right\} = E(S_n)+\sigma\cdot\sum_{j=1}^{n}\left\{e_{n+1-j}\cdot\sum_{i=1}^{j} r^{i-1}\right\}$$

In order to calculate the variance for this summation, we make use of the following relationships.

$$Var(S_n) = E(S_n^2) - E(S_n)^2 \qquad \text{and} \qquad E(e_i) = 0 \quad \forall i$$

The variance for the sum of these annual rates therefore requires the collapsing of the double summation.

$$Var(S_n) = \sigma^2 \cdot \sum_{j=1}^{n} \left(\sum_{i=1}^{j} r^{j-i} \right)^2 = \sigma^2 \cdot \sum_{j=1}^{n} \left(\frac{1-r^j}{1-r} \right)^2$$

This summation can be further simplified as shown below:

$$Var(S_n) = \frac{\sigma^2}{(1-r)^2} \cdot \sum_{j=1}^{n} \left(1 - 2r^j + r^{2j}\right)$$

$$= \frac{\sigma^2}{(1-r)^2} \cdot \left\{ t - 2r \cdot \left(\frac{1-r^n}{1-r} \right) + r^2 \cdot \left(\frac{1-r^{2n}}{1-r^2} \right) \right\}$$

Alternatively, for the special case in which $r = 1$, we can write

$$Var(S_n) = \sigma^2 \cdot \frac{n \cdot (n+1) \cdot (2n+1)}{6}.$$

The final step for the variance calculation is to translate the variance of the normal random variable X_t into the expression for the CV of the lognormal random variable.

We can accomplish this by making note of the following relationship[4].

$$CV\left(e^X\right)^2 = \frac{Var\left(e^X\right)}{E\left(e^X\right)^2} = e^{Var(X)} - 1$$

This provides the translation to all of the formulas given in section 3.1 of the paper.

By analogy, there is an expression for the [standardized] covariance of two random variables.

---

[4] As the reader might expect, this relationship holds when X is a normal random variable, but it is not generally true for other distributions.

$$Cov^*\left(e^X, e^Y\right) = \frac{Cov\left(e^X, e^Y\right)}{E\left(e^X\right)\cdot E\left(e^Y\right)} = e^{Cov(X,Y)} - 1$$

For this covariance expression, we recall that we are looking for the relationship between two sums of random variables $(X_1 + X_2 + \cdots X_n)$ and $(X_1 + X_2 + \cdots + X_n + \cdots + X_{n+k})$, which we may again denote S_n and S_{n+k} for convenience.

$$S_n = E(S_n) + \sigma \cdot \sum_{j=1}^{n}\left\{e_{n+1-j} \cdot \sum_{i=1}^{j} r^{i-1}\right\}$$

$$S_{n+k} = E(S_{n+k}) + \sigma \cdot \sum_{j=1}^{n+k}\left\{e_{n+k+1-j} \cdot \sum_{i=1}^{j} r^{i-1}\right\}$$

$$= E(S_{n+k}) + \sigma \cdot \sum_{j=1}^{n}\left\{e_{n+1-j} \cdot \sum_{i=1}^{j+k} r^{i-1}\right\} + \sum_{j=n+1}^{n+k}\left\{e_j \cdot \sum_{i=1}^{n+k+1-j} r^{i-1}\right\}$$

The logic for calculating the covariance term $Cov(S_n, S_{n+k})$ is similar to that used for the variance above.

$$Cov(S_n, S_{n+k}) = E(S_n \cdot S_{n+k}) - E(S_n)\cdot E(S_{n+k})$$

$$Cov(S_n, S_{n+k}) = \sigma^2 \cdot \sum_{j=1}^{n}\left\{\left(\sum_{i=1}^{j} r^{i-1}\right)\cdot\left(\sum_{i=1}^{j+k} r^{i-1}\right)\right\} = \frac{\sigma^2}{(1-r)^2}\cdot\sum_{j=1}^{n}\left\{(1-r^j)\cdot(1-r^{j+k})\right\}$$

$$= \frac{\sigma^2}{(1-r)^2}\cdot\left\{n - r\cdot\left(\frac{1-r^n}{1-r}\right) - r^{k+1}\cdot\left(\frac{1-r^n}{1-r}\right) + r^{k+2}\cdot\left(\frac{1-r^{2n}}{1-r^2}\right)\right\}$$

$$= \frac{\sigma^2}{(1-r)^2}\cdot\left\{n - r\cdot(1+r^k)\cdot\left(\frac{1-r^n}{1-r}\right) + r^{k+2}\cdot\left(\frac{1-r^{2n}}{1-r^2}\right)\right\}$$

For the special case in which $r = 1$, we can write

$$Cov(S_n, S_{n+k}) = \sigma^2 \cdot \sum_{j=1}^{n}\{j\cdot(j+k)\} = \sigma^2\cdot\sum_{j=1}^{n}\{j^2 + j\cdot k\}$$

$$= \sigma^2 \cdot \left\{ \frac{n \cdot (n+1) \cdot (2n+1)}{6} + \frac{n \cdot (n+1)}{2} \cdot k \right\}.$$

This completes the derivation of the covariance terms given in section 3.2 of the paper.

As a final observation, we may note that the CV and Covariance expressions are dependent upon σ and r (the reversion parameter), but do not involve the intercept b or the starting point X_0. In other words, we can estimate the variance relative to the mean level of the reserves without having to know the current or long-term inflation rates.

Appendix B: Chain-Ladder ODP Model

The over-dispersed Poisson (ODP) model is useful to illustrate the ideas in this paper since it conveniently balances to the well known chain-ladder reserving method.

We define an incremental loss payment in year y and development period d to be distributed as ODP. The distribution is defined as follows:

Probability Function: $$\text{Prob}(c_{y,d}) = \left(\frac{\mu_{y,d}}{\phi}\right)^{c_{y,d}/\phi} \cdot \frac{e^{-\mu_{y,d}/\phi}}{(c_{y,d}/\phi)!}$$

Mean: $$E(c_{y,d}) = \mu_{y,d}$$

Variance: $$Var(c_{y,d}) = \phi \cdot \mu_{y,d}$$

The parameter ϕ is the "dispersion parameter" and represents a constant variance-to-mean ratio. This parameter will be assumed to be fixed and known, and constant for all accident years and development periods. Mathematically it is just a scaling factor that changes a standard Poisson distribution, defined on the integers {0, 1, 2, 3, …} to an ODP distribution, defined on evenly spaced values $\{0, \phi, 2\phi, 3\phi, \ldots\}$.

The mean of each cell in the development triangle will then be defined as:

$$\mu_{y,d} = E(c_{y,d}) = \alpha_y \cdot \beta_d$$

In order to calculate the maximum likelihood estimation (MLE) values for these parameters, we need to evaluate the following expression.

$$LogLikelihood = \sum_{y=1}^{n} \sum_{d=1}^{n-y+1} \left\{ \frac{c_{y,d}}{\phi} \cdot \ln(\alpha_y \cdot \beta_d) - \frac{c_{y,d}}{\phi} \cdot \ln(\phi) - \frac{\alpha_y \cdot \beta_d}{\phi} - \ln((c_{y,d}/\phi)!) \right\}$$

However, since we are assuming that the dispersion parameter is fixed, we do not need to include it in our likelihood calculation. Instead, we use a quasi-likelihood (*QLL*) expression including only the portion of the LogLikelihood that is dependent on α_y and β_d.

$$QLL = \sum_{y=1}^{n} \sum_{d=1}^{n-y+1} \left\{ c_{y,d} \cdot \ln(\alpha_y \cdot \beta_d) - \alpha_y \cdot \beta_d \right\}$$

The derivatives with respect to the two parameters are set to zero.

$$\frac{\partial QLL}{\partial \alpha_y} = \sum_{d=1}^{n-y+1} \left\{ \frac{c_{y,d}}{\alpha_y} - \beta_d \right\} = 0$$

and

$$\frac{\partial QLL}{\partial \beta_d} = \sum_{y=1}^{n-d+1} \left\{ \frac{c_{y,d}}{\beta_d} - \alpha_y \right\} = 0.$$

The derivatives imply that the MLE values satisfy two conditions:

$$\sum_{d=1}^{n-y+1} c_{y,d} = \sum_{d=1}^{n-y+1} \alpha_y \cdot \beta_d \quad \text{and} \quad \sum_{y=1}^{n-d+1} c_{y,d} = \sum_{y=1}^{n-d+1} \alpha_y \cdot \beta_d \qquad \forall\, y,d.$$

That is, the row and column totals of the fitted values must equal the row and column totals of the original incremental triangle. Because these conditions do not result in a unique set of parameters, we can add one more constraint $\sum_{d=1}^{n} \beta_d = 1$, which results in $\alpha_1 = \sum_{d=1}^{n} c_{1,d}$. These constraints then mean that the MLE parameters are equivalent to the values in a standard chain-ladder reserve estimate.

This model can then be expanded to include estimates of trend based on the CPI:

$$\mu_{y,d} = E(c_{y,d}) = \alpha_y \cdot \beta_d \cdot CPI(y+d-1)^{\gamma}$$

$$QLL = \sum_{y=1}^{n} \sum_{d=1}^{n-y+1} \left\{ c_{y,d} \cdot \ln(\alpha_y \cdot \beta_d \cdot CPI(y+d-1)^{\gamma}) - \alpha_y \cdot \beta_d \cdot CPI(y+d-1)^{\gamma} \right\}.$$

We find from this expression that the following conditions must again be met:

$$\sum_{d=1}^{n-y+1} c_{y,d} = \sum_{d=1}^{n-y+1} \mu_{y,d} \quad \text{and} \quad \sum_{y=1}^{n-d+1} c_{y,d} = \sum_{y=1}^{n-d+1} \mu_{y,d} \qquad \forall\, y,d.$$

We must also add the derivative with respect to the CPI curve, γ:

$$\frac{\partial QLL}{\partial \gamma} = \sum_{y=1}^{n}\sum_{d=1}^{n-y+1}\left\{c_{y,d}\cdot\ln(CPI(y+d-1))-\alpha_y\cdot\beta_d\cdot\ln(CPI(y+d-1))\cdot CPI(y+d-1)^{\gamma}\right\} = 0$$

Which is equivalent to

$$\sum_{y=1}^{n}\sum_{d=1}^{n-y+1}\left\{c_{y,d}\cdot\ln(CPI(y+d-1))\right\} = \sum_{y=1}^{n}\sum_{d=1}^{n-y+1}\left\{\mu_{y,d}\cdot\ln(CPI(y+d-1))\right\}.$$

Unfortunately, there is no longer a convenient closed-form solution for calculating the model parameters, though it can be somewhat simplified using the relation below:

$$\beta_d = \frac{\sum_{y=1}^{n-d+1} c_{y,d}}{\sum_{y=1}^{n-d+1}\left\{\alpha_y\cdot CPI(y+d-1)^{\gamma}\right\}}.$$

The parameters in the model including the external CPI values must be estimated via an iterative calculation. This does not create any great difficulty in our model.

What is more interesting, however, is the relatively little improvement in model fit that is seen when the CPI values are introduced. It makes intuitive sense that loss payments should follow inflation, so why does introducing inflation as an explanatory variable add so little to the goodness of fit?

The answer is that a standard chain-ladder or MLE calculation is already estimating many parameters: one for each accident year α_y and one for each of the first n-1 development periods β_d (by constraining these to add to 1.00 we reduce the model by one parameter). This means that in a triangle with n years, we will have $n(n+1)/2$ data points to estimate $2n-1$ parameters; for a 10-year triangle we have 55 incremental payments to estimate 19 parameters. The effects of inflation are "buried" in our otherwise over-parameterized model.

To see this more clearly, we will introduce one more model in which the inflation rate i is assumed to be constant, and is estimated as a parameter of the model.

$$\mu_{y,d} \quad = \quad E(c_{y,d}) \quad = \quad \alpha_y \cdot \beta_d \cdot (1+i)^{y+d}$$

The quasi-likelihood function is given as follows.

$$QLL \quad = \quad \sum_{y=1}^{n} \sum_{d=1}^{n-y+1} \left\{ c_{y,d} \cdot \ln\left(\alpha_y \cdot \beta_d \cdot (1+i)^{y+d}\right) - \alpha_y \cdot \beta_d \cdot (1+i)^{y+d} \right\}$$

Taking the derivative with respect to the inflation rate i, we have

$$\frac{\partial QLL}{\partial i} \quad = \quad \sum_{y=1}^{n} \sum_{d=1}^{n-y+1} \left\{ \frac{c_{y,d} \cdot (y+d)}{(1+i)} - \alpha_y \cdot \beta_d \cdot (y+d) \cdot (1+i)^{y+d-1} \right\} \quad = \quad 0$$

Or equivalently,

$$\frac{\partial QLL}{\partial i} \quad = \quad \sum_{y=1}^{n} \sum_{d=1}^{n-y+1} \left\{ (y+d) \cdot \left(c_{y,d} - \mu_{y,d}\right) \right\} \quad = \quad 0$$

We may note that this condition for the derivative of the loglikelihood with respect to i will automatically be met if we first calculate α_y and β_d via the chain-ladder method (assuming no inflation), and then adjust the numbers as:

$$\alpha_y^* \; = \; \alpha_y \cdot (1+i)^{-y} \qquad\qquad \beta_d^* \; = \; \beta_d \cdot (1+i)^{-d}$$

Such that $\alpha_y^* \cdot \beta_d^* \cdot (1+i)^{y+d} \; = \; \alpha_y \cdot (1+i)^{-y} \cdot \beta_d \cdot (1+i)^{-d} \cdot (1+i)^{y+d} \; = \; \alpha_y \cdot \beta_d$

The MLE for a model with a <u>constant</u> inflation rate is therefore equal to the chain-ladder model with no inflation.

Appendix C: Variance and Covariance of Products of Random Variables

The general form of the variance of a single random variable X, and its covariance with a second random variable Y, are expressed in the following familiar equations.

$$Var(X) = E(X^2) - E(X)^2$$

$$Cov(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

The variance of the product of these two random variables has a somewhat more complex expression:

$$Var(X \cdot Y) = E(X^2 \cdot Y^2) - E(X)^2 \cdot E(Y)^2$$

If X and Y are independent, then this can be re-written as follows.

$$Var(X \cdot Y) = Var(X) \cdot Var(Y) + Var(X) \cdot E(Y)^2 + Var(Y) \cdot E(X)^2$$

Proof:

$$\begin{aligned} Var(X) \cdot Var(Y) &= \{E(X^2) - E(X)^2\} \cdot \{E(Y^2) - E(Y)^2\} \\ &= E(X^2) \cdot E(Y^2) + E(X)^2 \cdot E(Y)^2 \\ &\quad - E(X^2) \cdot E(Y)^2 - E(X)^2 \cdot E(Y^2) \\ &= E(X^2) \cdot E(Y^2) - E(X)^2 \cdot E(Y)^2 \\ &\quad - E(X^2) \cdot E(Y)^2 + E(X)^2 \cdot E(Y)^2 \\ &\quad - E(X)^2 \cdot E(Y^2) + E(X)^2 \cdot E(Y)^2 \\ &= \{E(X^2 \cdot Y^2) - E(X)^2 \cdot E(Y)^2\} \\ &\quad - \{E(X^2) - E(X)^2\} \cdot E(Y)^2 \\ &\quad - \{E(Y^2) - E(Y)^2\} \cdot E(X)^2 \end{aligned}$$

$$Var(X) \cdot Var(Y) = Var(X \cdot Y) - Var(X) \cdot E(Y)^2 - Var(Y) \cdot E(X)^2$$

In a similar fashion, the covariance between two products of random variables can be calculated using the expression below.

$$Cov(X_1 \cdot Y_1, X_2 \cdot Y_2) = E(X_1 \cdot Y_1 \cdot X_2 \cdot Y_2) - E(X_1) \cdot E(Y_1) \cdot E(X_2) \cdot E(Y_2) \backslash$$

Again, if the X 's and Y 's are independent, the covariance formula can be re-written as follows.

$$\begin{aligned} Cov(X_1 \cdot Y_1, X_2 \cdot Y_2) &= Cov(X_1 \cdot Y_1) \cdot Cov(X_2 \cdot Y_2) \\ &+ Cov(X_1 \cdot X_2) \cdot E(Y_1) \cdot E(Y_2) + Cov(Y_1 \cdot Y_2) \cdot E(X_1) \cdot E(X_2) \end{aligned}$$

The proof follows a similar logic as above for the variance calculation.

Proof: $Cov(X_1 \cdot X_2) \cdot Cov(Y_1 \cdot Y_2)$

$$= \{E(X_1 \cdot X_2) - E(X_1) \cdot E(X_2)\} \cdot \{E(Y_1 \cdot Y_2) - E(Y_1) \cdot E(Y_2)\}$$

$$\begin{aligned} &= E(X_1 \cdot X_2) \cdot E(Y_1 \cdot Y_2) + E(X_1) \cdot E(X_2) \cdot E(Y_1) \cdot E(Y_2) \\ &- E(X_1 \cdot X_2) \cdot E(Y_1) \cdot E(Y_2) \\ &- E(Y_1 \cdot Y_2) \cdot E(X_1) \cdot E(X_2) \end{aligned}$$

$$\begin{aligned} &= \{E(X_1 \cdot X_2) \cdot E(Y_1 \cdot Y_2) - E(X_1) \cdot E(X_2) \cdot E(Y_1) \cdot E(Y_2)\} \\ &- \{E(X_1 \cdot X_2) \cdot E(Y_1) \cdot E(Y_2) - E(X_1) \cdot E(X_2) \cdot E(Y_1) \cdot E(Y_2)\} \\ &- \{E(Y_1 \cdot Y_2) \cdot E(X_1) \cdot E(X_2) - E(X_1) \cdot E(X_2) \cdot E(Y_1) \cdot E(Y_2)\} \end{aligned}$$

$$\begin{aligned} &= Cov(X_1 \cdot Y_2, X_2 \cdot Y_2) \\ &- Cov(X_1 \cdot X_2) \cdot E(Y_1) \cdot E(Y_2) \\ &- Cov(Y_1 \cdot Y_2) \cdot E(X_1) \cdot E(X_2) \qquad \text{Q.E.D.} \end{aligned}$$

REFERENCES

[1] Ahlgrim, Kevin C, Stephen P. D'Arcy, Richard W. Gorvett, "Report on Modeling of Economic Series Coordinated with Interest Rate Scenarios," (Section 5) **2001** www.casact.org/research/econ
[2] Brehm, Paul J., "Correlation and the Aggregation of Unpaid Loss Distributions." CAS Forum 2002 (Fall), 1-23.
[3] England, Peter D. and Richard J. Verrall, "A Flexible Framework for Stochastic Claims Reserving," *PCAS* **2001**, Vol. LXXXVIII, 1-38.
[4] Hodes, Douglas M., Sholom Feldblum, Gary Blumsohn, "Workers Compensation Reserve Uncertainty," *PCAS* **1999**, Vol. LXXXVI, 263-392.
[5] Taylor, G.C., "Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claim Delays," ASTIN Bulletin, Volume 9, 1977, 219-230.
[6] Working Party on Quantifying Variability in Reserve Estimates, "The Analysis and Estimation of Loss & ALAE Variability: A Summary Report." CAS Forum 2005 (Fall), 29-146.

Abbreviations and notations:
AR(1), autoregressive time-series model dependent on a single prior point
CPI, Consumer Price Index
CV, coefficient of variation = standard deviation / mean
ODP, Over-Dispersed Poisson

Biography of the Author

Dave Clark is Vice President and Actuary with American Re-Insurance. His paper "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach" received the 2003 Reserves Call Paper Prize.

Acknowledgments

The author thanks Micah Woolstenhulme whose comments during the Working Party on Reserve Variability spurred on this research, and Denise Ambrogio, Chuck Thayer and Roger Hayne whose critiques helped the author improve the rigor of the paper in several ways. Any remaining errors in the final paper are, of course, the fault of the author alone.

Supplementary Material

An Excel spreadsheet including an illustrative example of the variance calculation is available upon request from the author.