ABOUT THE UNCERTAINTY OF PAST INFLATION

A MATHEMATICAL ANSWER TO WHY WE DON'T USE THE DATA OF 80 YEARS AGO

MICHAEL FACKLER

freelance actuary
Klopstockstr. 8-IX
80804 Munich
Germany
michael_fackler@web.de

ABSTRACT

Insurance data, before being used in statistics, often are adjusted for inflation, which is routinely done by application of a suitable inflation index. Whether or not the index exactly matches the inflation of the business remains somewhat uncertain. A model is proposed for this uncertainty, interpreting the gap between the "true" inflation and the applied index series as a random variable. For a special case the mean squared error of the sample mean is calculated. The result is quite different from the traditional model without inflation uncertainty but it looks very much as actuaries would intuitively say it should.

KEYWORDS

Inflation, index series, sample mean, mean squared error, AR(1), Wilkie model

1. INTRODUCTION

In insurance, and in economics in general, the environment is always changing. Sometimes these changes can be accounted for using methods like regression of trends, time series analysis, etc., which attempt to derive all relevant changes out of the data. In the case of inflationary effects (seen here in a very general meaning as all changes affecting the size of insurance losses) this often seems to be impossible and/or unnecessary – a variety of inflation indexes (consumer prices, wages, construction cost, etc.) is available, all well constructed from a lot of data by public bureaus of statistics or by the insurance industry itself. So very often the losses before being input in statistics are adjusted by factors coming from a suitable, carefully selected, inflation index.
This method is fine, the more so as often there is no easy alternative in sight, but we should bear in mind what we are doing: We use sophisticated statistical models that calculate not only expected values but also mean squared errors, confidence intervals and many other features that model uncertainty – and the model input is adjusted data. By doing so it is **implicitly** assumed that the adjustment for inflation is exactly correct, in statistical terms:

The applied index is a **certain** estimation of past inflation.

Of course actuaries are aware that this is not exactly true, and take this into account: They feel more comfortable with very recent data that require no or only slight adjustment than with data that are, say, 20 years old and have to be adjusted by large (and potentially quite incorrect) factors. So one tends to exclude older years from the analysis if possible, i.e. if recent years provide enough data to reduce random effects. However, the decision which years to consider, and which to exclude, is rather judgemental than mathematical.
This is unsatisfactory. It may occur that a drastic change in the environment (like new insurance conditions or laws) makes clear that the years before that change cannot be used since they are not representative at all. However, it is much more common that the environment changes gradually, so that there is no clear separation between recent “good” years and too old data. If in this situation we start with the most recent data and successively add older years to the analysis we increase the quantity of the data, which reduces random effects, but somewhat decrease the quality of the analysis as the older data, although adjusted, is likely to be less representative for the present. It is plausible that at a certain point adding more years does not improve the analysis any more. (Furthermore it comes into mind that it might be a good idea to give less weight to older years, which in fact is sometimes done in practice.)

In other words:
- If we strongly believe in the adequacy of the inflation index for a certain period of time then we should make use of all years of the period, be it 5 or 80.
- If instead we think that there is a deviation between the index and the “true” inflation, then we should try to find a mathematical model for that deviation, and the decision of how many years to take will no longer rely solely on judgement but will come from the model: A mathematical language for what actuarial intuition says, namely the aforementioned trade off (random effects decreasing vs. uncertainty increasing as older years are added to the database) will yield the optimal number of years to be used, and possibly the optimal weights to be given to the single years.

In section 2 we introduce a typical insurance situation where index-adjusted data are the input for the estimators of a risk premium, and propose a model for the gap between the applied index series and the "true" inflation. It is shown how this affects the formula of the mean squared error (mse) of the sample mean of past loss experience.

In section 3 we introduce time series models for the inflation uncertainty, letting us inspire by research from David Wilkie and David Clark.

In section 4 we analyse a special time series where the aforementioned trade off can be verified and the optimal number of years to be taken calculated.

The research described in the following is far from being exhaustive. Therefore rather than giving a a conclusion the paper will end with a lot of questions possibly worth being addressed in the future.

2. THE BASIC MODEL

Let X be the yearly loss of an insurance risk (or of a portfolio of risks) that we have observed for a number of years and that we want to estimate for the present year. We assume that the distribution function of X does not change over the years, apart from inflationary effects.

We count backwards, i.e. X_1 is the loss of past year, X_2 the loss of two years ago, and so on. X_k then is the aggregate loss of year k (= k years ago).

Let Y_k be the loss of year k "as if" in today's money terms, having taken account for inflation. Then the Y_k are an i.i.d. sequence of random variables (represented by Y), and we have

$$Y_k = X_k B_k$$

with $B_k>0$ being the index describing the inflation between year k and today. If we exactly knew the B_k, k>0, then the $X_k B_k$ would be unbiased estimators for the present year $X_0 = Y_0$.

But in practice we often only have an index series $I_k>0$ which is a fair, but not certain, estimation of the inflation between year k and today, and so instead of $X_k B_k$ we usually use the estimators

$$X_k I_k$$

whose observations are the observations of the X_k adjusted for inflation via the index I_k. To study the properties of these estimators we introduce:

Definition 2.1: The ratio $U_k := I_k/B_k$ is called the **past inflation gap**.
$U_k>0$ is unknown and interpreted as a random variable.

As all index series are related to the present year we have $B_0 = I_0 = U_0 = 1$.

Remarks:
- It might appear more natural to define the inflation gap as the inverse B_k/I_k but it will turn out advantageous if we do as above. Anyway inverses are easy to deal with, as all index-like random variables are always greater than zero – in particular as these random variables are commonly modelled as exponentials of real valued r.v. where inversion just means change of sign.

- The restriction to one value per year for all index series might appear too approximate nowadays with continuous time models being well established. Inflation is usually rather seen as a continuous process and not just as one step per year. It shall be noted that when we say years in this analysis we could instead mean quarters or months, etc., according to how the data is organized. However, often loss data are not available (or reliable) in finer grids than annually. Then one arguably best uses one index value per year accordingly, keeping in mind that strictly speaking this is an approximation.

From the definition it follows that $X_k I_k = Y_k U_k$

Hence the $X_k I_k$ can be represented as products of an i.i.d. sequence of r.v. Y_k and a sequence of r.v. U_k. The U_k are independent of the Y_k but as values of an index series they typically depend strongly on each other.

The question of how the U_k (or equivalently the B_k) could be estimated is intriguing but will not be treated in this paper. Here we focus on the mathematical properties of the commonly used estimators $X_k I_k = Y_k U_k$ and of the sample mean. A lot can be said about them without specifying the U_k too much. At the moment we only use the condition $U_k > 0$ and keep in mind that the U_k fluctuate quite closely around the value 1 (if the I_k are carefully selected).

If we recall that for independent r.v. A, B the expected value is multiplicative and generally $\mathrm{Cov}(A,B) = E(AB) - E(A)\,E(B)$ we easily get

Lemma 2.2: For $k>0$, $i>0$ we have

(1) $E(Y_k U_k) = E(Y)\,E(U_k)$

(2) $\mathrm{Var}(Y_k U_k) = \mathrm{Var}(Y)\,E(U_k^2) + E^2(Y)\,\mathrm{Var}(U_k)$

(3) $\mathrm{Cov}(Y_k U_k, Y_i U_i) = E^2(Y)\,\mathrm{Cov}(U_k, U_i)$, $i \neq k$

Even if U_k is very close to the constant 1 we cannot hope that its expectation equals one, think of the classical lognormal example: If the underlying normal distribution has mean zero and a very small variance the lognormal r.v. fluctuates very closely around 1 but its mean is a bit greater than 1. Hence $Y_k U_k$ is likely to be a (slightly) biased estimator of Y.

Let us now look at the **sample mean** of a period of $n>0$ years

$$T_n(Y) := (X_1 I_1 + \ldots + X_n I_n) / n = (Y_1 U_1 + \ldots + Y_n U_n) / n$$

We want to calculate the mean squared error of this most commonly used estimator:

$$\mathrm{mse}T_n(Y) = E(\{T_n(Y) - E(Y)\}^2) = \mathrm{Bias}^2 T_n(Y) + \mathrm{Var}T_n(Y) \quad \text{with} \quad \mathrm{Bias}T_n(Y) = E(T_n(Y)) - E(Y)$$

Definition 2.3: For simplicity of notation we set $\eta := CV(Y) = \mathrm{Sta}(Y)/E(Y)$ and relate the components of the mean squared error to E(Y) as well:

relative Bias: $\mathrm{r.Bias}T_n(Y) := \mathrm{Bias}T_n(Y) / E(Y)$

relative Variance: $\mathrm{r.Var}T_n(Y) := \mathrm{Var}T_n(Y) / E^2(Y)$

relative mse: $\quad r.mseT_n(Y) := mseT_n(Y) / E^2(Y)$

Application of Lemma 2.2 yields (after doing some algebra)

Proposition 2.4: For n>0 we have

$$r.BiasT_n(Y) = \{ \sum_{k=1}^{n} E(U_k) \} / n - 1$$

$$r.VarT_n(Y) = \{ \sum_{k=1}^{n} (Var\, U_k + \eta^2 E(U_k^2)) + 2 \sum_{1\le k<i\le n} Cov(U_k,U_i) \} / n^2$$

Remark: Recall that the traditional model makes the assumption $U_k \equiv 1$. Then

$$r.BiasT_n(Y) = 0 \quad \text{and} \quad r.VarT_n(Y) = \eta^2/n.$$

This is the common result leading to the conclusion that to minimize $r.mseT_n(Y)$ one has to maximize n, i.e. to take as many years as available.

3. TIME SERIES MODELS FOR THE PAST INFLATION GAP

The past inflation gap is derived from time series, and so it seems adequate to use a time series model for it, more specifically the kind of model that is used for index series like consumer prices or wages.
As part of his fundamental “Wilkie model” Wilkie (1995) proposed, among others, the following autoregressive models (see his chapters 2.1, 3.1, 3.3):

Retail (consumer) prices Q_k: $\quad \ln Q_k - \ln Q_{k-1}$ is an AR(1) series
Wages W_k: $\quad \ln W_k - \ln W_{k-1}$ is an AR(1) series

Recall that an AR(1) series is a sequence of random variables R_k with

$$R_k = r\, R_{k-1} + (1-r)\, \mu + \sigma\, E_k$$

with a normally distributed random error E_k having mean 0 and standard deviation 1.

For the ratio W_k/Q_k (real wages) one would get an analogous model but it is plausible and was observed from data that there tends to be a certain connection between W and Q, so W/Q could e.g. be a stationary time series or more specifically a time series as above with $\mu = 0$.

If we compare this to our situation here, with U being the quotient of the inflation of an insurance risk and an inflation (or wages, etc.) index, then it seems both reasonable and flexible to assume that

$$\ln U_k - \ln U_{k-1} \text{ is an AR(1) series.}$$

One could now directly proceed to recalculate Prop. 2.4 for the selected model for the U_n. Before doing so we want to draw the attention of the reader to a surprising analogy between the issue addressed in this paper and a very different problem which has been analysed by Clark (2006). He investigates (see his section 5) a sum

$$C_1b_1 + \ldots + C_nb_n$$

which represents a stream of future payments coming say from a run off portfolio. C_k is the amount to be paid out in k years from now but in today's money terms, which comes e.g. from a stochastic claims reserving calculation. b_k is the inflation (as a whole, not a gap) from now to time k, and the model chosen for this index is exactly as above, i.e.

$$\ln b_k - \ln b_{k-1} \text{ is an AR(1) series}$$

Clark provides formulae for, among others, the variance of such streams of payments, which use the parameters of the AR(1) process and the expected values, variances, and covariances of the C_k.

If we (in an attempt to find good estimators for Y) regard arbitrary linear combinations of the Y_kU_k we get

$$a_0 + a_1Y_1U_1 + \ldots + a_nY_nU_n$$

Although coming from a completely different context the results from Clark can be applied here, just set $b_k = U_k$ and $C_k = a_kY_k$. The C_k are independent in our case, and their expectations and variances can easily be expressed in terms of E(Y) and Var(Y).

The way U_k was defined above and the backward counting of the years pay off: This makes our model look like a model for future inflation and facilitates the application of results from such models.

The generalization of Prop. 2.4. for arbitrary linear combinations of the Y_kU_k and the specification for U_k being a time series as introduced in this chapter are lengthy but straightforward calculations. However, the formulae become intricate with many different parameters. We provide them in Appendix A and now treat a special case which will quickly lead to interesting results.

4. THE SAMPLE MEAN IN THE RANDOM WALK CASE

We go back to the sample mean and investigate a model that should be easy to calculate: Let $\ln U_k$ be a random walk without drift having small variability (to be specified later). This is arguably the simplest non-constant model for the U_k and a stationary time series. In particular the past inflation gap is very low for both recent and older years – an assumption that hardly overstates the impact of inflation uncertainty. Nevertheless this assumption will have a decisive impact, as will be seen soon.

In this model the $\ln U_k - \ln U_{k-1} = \sigma E_k$ are a sequence of independent normally distributed r.v. with mean zero and variance $\sigma^2>0$. If we name their exponentials W_k we get $U_k = W_1 \cdot \ldots \cdot W_k$. The W_k are i.i.d. lognormally distributed (represented by W). Set

$$w := E(W) = \exp(\sigma^2/2) > 1$$

Recall that for a lognormally distributed r.v. L with lnL ~ $N(0,v^{0.5})$ we have

$$E(L) = e^{v/2},\quad E(L^2) = e^{2v} = E^4(L),\quad VarL = E^4(L) - E^2(L)$$

Therefore $\qquad E(W^2) = w^4$ and $Var(W) = w^4 - w^2$

U_k is lognormally distributed, too, with $E(U_k) = w^k$, $E(U_k^{\,2}) = w^{4k}$, $Var(U_k) = w^{4k} - w^{2k}$

In order to calculate $Cov(U_k,U_i)$ for i>k>0 note that U_i is the product of the two independent r.v. U_k and $(W_{k+1}\cdot\ldots\cdot W_i)$, so we get

$$Cov(U_k,U_i) = Cov(U_k,U_k\cdot W_{k+1}\cdot\ldots\cdot W_i) = Var(U_k)\, E(W_{k+1}\cdot\ldots\cdot W_i) = (w^{4k} - w^{2k})\, w^{i-k} = w^{3k+i} - w^{k+i}$$

Now we have all we need to apply Prop. 2.4 to this special model for U_k:

Proposition 4.1: If the inflation gap is the exponential of a random walk without drift then for the sample mean the following formulae hold:

$$r.BiasT_n(Y) = \Big(\sum_{k=1}^{n} w^k\Big) / n - 1 = \{w\,(w^n - 1)\} / \{(w - 1)\, n\} - 1 =: g_1(w)$$

$$r.VarT_n(Y) = \Big\{ \sum_{k=1}^{n} (w^{4k} - w^{2k} + \eta^2 w^{4k}) + 2 \sum_{1\le k<i\le n} (w^{3k+i} - w^{k+i}) \Big\} / n^2 =: g_2(w)$$

with $w = E(U_k/U_{k-1})$.

We want to find out the behaviour of $r.mseT_n(Y)$ for varying n. Recall that without the past inflation gap this sequence tends monotonically to zero as n increases. The intuition tells actuaries that for large n the inflation gap should have some impact and increase the mean squared error. This "feeling" can now be mathematically verified:

As for the bias, it is positive and strictly increasing in n: We have to prove

$$[w+\ldots+w^n] / n < [w+\ldots+w^{n+1}] / (n+1),$$
$$\text{or equivalently } (n+1)\, [w+\ldots+w^n] < n\, [w+\ldots+w^{n+1}],$$
$$\text{or } [w+\ldots+w^n] < n\, w^{n+1},$$

and the latter is clear as w>1.

The sums in the variance formula of Prop. 4.1 could be replaced by application of binomial formulae but the result is intricate. A different approach which will tell us more about the behaviour of the sequence $r.VarT_n(Y)$ for increasing n: To get at first an approximate answer let w be very close to 1, i.e. we assume the variance of the underlying normal distribution is very small. We use the fact that for fixed n

$$g(w) := r.mseT_n(Y)\,(w) = g_1^{\,2}(w) + g_2(w)$$

is an analytical function of w on the real axis and can be developed into a Taylor series around w = 1. If we only regard the terms of first order we should have a fair approximation for w = 1+d with very small d>0:

$$g(1+d) \approx g(1) + d \cdot g'(1) = g_1^{\,2}(1) + d \cdot (g_1^{\,2})'(1) + g_2(1) + d \cdot g_2'(1)$$

From $g_1(1) = 0$ it follows that $(g_1^2)'(1) = 2\, g_1(1)\, g_1'(1) = 0$

Therefore in first order terms the square of the Bias is zero, thus variance and mean squared error coincide. We see at once that $g_2(1) = \eta^2/n$.
The derivative of g_2 at w=1 can easily be calculated element-wise:

First sum: $(w^{4k} – w^{2k} + \eta^2 w^{4k})'(1) = 4k – 2k + 4k\eta^2 = 2k\,(1+2\eta^2)$

The sum over $1 \le k \le n$ of these terms is $n(n+1)(1+2\eta^2)$

Second sum: $(w^{3k+i} – w^{k+i})'(1) = (3k+i) – (k+i) = 2k$

The sum over $1 \le k < i \le n$ of these terms turns out to be $(n^3–n)/3$. Hence

$$g(1+d) \approx g_2(1) + d \cdot g_2'(1) = \eta^2/n + d\,\{n(n+1)(1+2\eta^2) + 2(n^3–n)/3\} / n^2$$

Note that this approximation is also a lower bound for g: The nonnegative g_1^2 is approximated by zero, and if we look again at the two sums in the variance formula of Prop. 4.1 it is not difficult to see that all their derivatives are nonnegative at w = 1:
The two negative terms each have a predecessor with higher exponent. The derivatives of these differences are, as is easily seen by induction, of the form $pw^j – qw^m$ with $p > q \ge 0$ and integers $j > m \ge 0$. By going on taking derivatives the exponents decrease and the negative term vanishes before the positive one.
Therefore any finite Taylor expansion of g_2 is a lower bound for w>1.

Thus by looking at an approximation we have got much more than an approximate result. Recall we were interested in the behaviour of the r.mse for varying n. If we order by powers of n we get

Proposition 4.2: Let the inflation gap be the exponential of a random walk without drift. Then the relative mean squared error of the sample mean is not smaller than

$$\varphi(n) := [(1+2d)\eta^2 + d/3] / n \;+\; [d(1+2\eta^2)] \;+\; [2d/3]\, n$$

If the variance of the underlying normal distribution is very small the relative mse approximately equals this lower bound.

Before interpreting this formula it shall be noted that the approximation appears to be fairly good, as can bee seen from the table in Appendix B which shows both the exact and the approximate calculation of the r.mse for the values used for d and η=CV(Y) in Example 4.4 below. Hence φ(n) is definitely worth further analysis, both as a good approximation and as a lower bound for $\text{r.mse}T_n(Y)$ (w).

First of all note that in case d = 0 the approximation is exact and yields again the model without inflation gap as well as the well known $\text{r.mse}T_n(Y) = \eta^2/n$, decreasing monotonically towards zero as n grows towards infinity. But for any d>0 the sequence φ(n) shows a completely different behaviour:
The first term decreases as above, the second is a constant but the third tends towards infinity with growing n, turning φ(n) into a divergent sequence. Thus $\text{r.mse}T_n(Y)$ is divergent in n, too. This could possibly have been proven more quickly without regarding the Taylor series,

but now we can (approximately) describe the sequence r.mseT_n(Y) for small n as well: If we interpret φ(n) for a moment as a function of positive real numbers and take the second derivative we see at once that φ(n) is convex. It has the shape of a "U" because for $n \to 0$ the limit is infinite as well. Thus φ has a unique local and global minimum on the positive real axis and one (possibly either) of the two natural numbers that are closest to this value is then the minimum of φ as a function of natural numbers. (In case the minimum is a real number less than 1 the sequence φ(n) is increasing with minimum 1).

Coming finally back from curve discussion issues to our model we conclude: The relative mean squared error of the sample mean is first reduced with growing n (as in the traditional model), then it rises again. At this point adding more years to the analysis doesn't improve it any more, which is exactly what actuaries intuitively would have expected. In terms of mean squared error we can state:

Among the sample means there is an optimal one.

Note that this last assertion holds exactly, not only approximately: r.mseT_n(Y) is a divergent sequence, thus has a finite minimum, which for small d will be close to the minimum of φ(n). The latter can be calculated easily: Setting $\frac{\partial}{\partial n}\varphi(n) = 0$ we get $n^2 = 3\eta^2/2d + 3\eta^2 + 1/2$, which approximately equals $3\eta^2/2d$ assuming 2d is somewhat smaller than η^2.

To interpret this result note that $1+d = \exp(\sigma^2/2) \approx 1+\sigma^2/2$. Therefore 2d approximately equals the variance σ^2 of the normal distribution underlying the random walk. The parameter σ^2 reappears in the inflation gap itself: Recall W representing the i.i.d. sequence $W_k = U_k/U_{k-1}$. W can be seen as the **one-year inflation gap**, i.e. the inflation gap between any two subsequent years. We have

$$CV^2(W) = w^2 - 1 = \exp(\sigma^2) - 1 \approx \sigma^2$$

Thus σ approximately equals the coefficient of variation of W. Choosing n_0 as the natural number closest to $\sqrt{3}\,\eta/\sigma$ we have approximately found the minimum of the sequence φ(n):

Corollary 4.3: If the inflation gap is the exponential of a random walk without drift then there is an optimal sample mean using the past n_0 years.
If furthermore the CV of the one-year inflation gap W is very small then n_0 is close to

$$\sqrt{3}\,CV(Y) / CV(W)$$

Thus n_0 is approximately proportionate to the coefficient of variation of Y, the r.v. to be estimated. Double volatility requires twice as many years to be used. Here's an example.

Example 4.4: As insurance risk let us choose a short tail portfolio, say Property lines. (Portfolios as a whole are regarded to assess the overall profitability of a line of business, e.g. for strategic decisions or when a proportional reinsurance treaty is rated.) Assume the portfolio to be stable, having about the same number of policies for a period of several years and no significant changes in the kind of insured risks.
Of course in practice we have never exactly observed Y (the losses perfectly adjusted for inflation), but we may help ourselves through practical experience with arguably similar

values. By looking at e.g. the losses per Total Insured Value or the Loss Ratios (possibly adjusted for premium cycles) we can say that a realistic value for CV(Y) could be in the range of 10%, assuming the exposure to natural hazards is low. However, in case of very small portfolios or high natural hazard exposure we can get 50% or more.
Assumptions about the value of CV(W), or equivalently σ, are more difficult. We might get some orientation from typical parameters of the AR(1) models for the inflation indexes themselves. Clark (2006, section 2) derives a σ of 0.015 from the US CPI index 1970-2004. Wilkie (1995, chapters 2.3, 3.2) takes into account over 70 years of UK history and gets values of σ between 0.04 and 0.06 for retail prices and a bit less for wages. Recall that a big part of the variability of these models comes from the autoregressivity, not from the normal random error.
The random walk U_k models the inflation gap, thus it is plausible that it fluctuates somewhat less than the inflation index itself. However, here the only source of variability is the random error, so it might be reasonable not to assume a too low value for σ. We set CV(W) = 0.04 (which to be exact corresponds to the assumption $\sigma = 0.039984$). Then according to Cor. 4.3 the optimal number of years for the sample mean is about $43.3 \cdot CV(Y)$. Thus in the 10% case we would choose the sample mean of the past 4 years while for a CV of 50% the model suggests to take about 22 years, see Appendix B.

These values are not so much different from what is done in practice, where often 5 to 10 years are taken into account, and sometimes more if a line of business is largely affected by natural, or other, catastrophes. The fact that more than 20 years are rarely used could mean that it is (implicitly) assumed that changes other than inflationary occur, or that the inflation gap is higher, maybe a non-stationary time series.

If we assume a random walk with drift or more complex time series models, as suggested in section 3, we will get intricate formulae (see the basic results in Appendix A) and most likely the optimal number of years will be smaller than here due to a higher impact of the inflation uncertainty.

5. OUTLOOK

Often the loss frequency of the insurance risk changes over the years, due to changing volume or to frequency trends. This is usually accounted for by introduction of a volume v_k into the formulae. $CV(X_k)$ is then usually assumed to be proportionate to $1/v_k^{0.5}$ and the sample mean becomes a weighted mean with weights v_k. This does not complicate formulae too much but one should bear in mind that volumes for insurance risks typically are money terms (e.g. Total Insured Value, Gross Premium Volume), which means that they are affected by inflation. Possibly volume inflation equals loss inflation but usually the two inflation series will be different albeit somehow related, which means that they, and their gaps as well, are correlated in a way to be found out.

One could very generally investigate arbitrary linear combinations of the Y_kU_k, starting with the formulae in Appendix A. The sum of the coefficients does not need to equal one, the more so as the Y_kU_k are likely to have a positive Bias. It is a natural, though complex, optimisation problem to find the linear combination that yields the lowest relative mean squared error, and once it is found the question from the beginning of the paper is answered: Should older years get less weight (probably yes), and what weight?

The paper only looks at the present year. Typically one needs at least an estimation for the following year. This means that we have a combination of the uncertainty of past inflation and the uncertainty in the prediction of the inflation (as a whole) from today to next year.

Long tail business offers further challenges as here many losses require payments over a long period, hence are affected by the inflation of different years. Furthermore one needs to predict inflation for many years in the future. Finally the run off of the losses comes into play, adding the uncertainty of the loss development to the model, having big impact especially on the very recent years.

If one thinks that inflation affects different loss sizes with different factors it seems nearly impossible to find a model at all. An option might be to identify different classes of losses (fire vs. natural perils, property damage vs. personal injuries, etc.) and to assume that inflation applies uniformly within each class.

In any case it would be desirable to have estimations for the past inflation gap from the loss data. Although this seems infeasible in many cases (e.g. because of too few or censored data) there might be situations where it can be done. Then one can use the estimated time series but should bear in mind – and include in the model – that this estimation is not certain.

For the many situations where the data do not allow an estimation of the inflation gap this paper might disclose a pragmatic way to deal with inflation uncertainty: The selection of a simple model for the inflation gap, like the one of section 4, together with plausible parameters. Better a partly judgemental assessment of the inflation gap than completely ignoring it and having to decide without explicit assumptions which data are good and which are too old.

ACKNOWLEDGEMENTS

The author is grateful to Dave Clark, Professor Viktor Sandor, and Erasmus Gerigk for a lot of invaluable suggestions improving the presentation of this paper.

REFERENCES

CLARK, D.R. (2006) Variance and Covariance Due to Inflation, *Casualty Actuarial Society Forum, Fall 2006,* 61-95 (http://www.casact.org/pubs/forum/06fforum/65.pdf)

WILKIE, A.D. (1995) More on a Stochastic Asset Model for Actuarial Use. *B.A.J.* **1**, V, 777-964

APPENDIX

A. FORMULAE FOR THE GENERAL AR(1) CASE

Let us start with arbitrary $U_k>0$ and arbitrary estimators $T(Y) := a_0 + a_1Y_1U_1 + \dots + a_nY_nU_n$. The straightforward generalisation of Prop. 2.4 is

Proposition A.1: For n>0 we have

$$r.BiasT(Y) = a_0 + \sum_{k=1}^{n} a_k E(U_k) - 1$$

$$r.VarT(Y) = \sum_{k=1}^{n} a_k^2 (\mathrm{Var}\, U_k + \eta^2 E(U_k^2)) + 2 \sum_{1\le k<i\le n} a_k a_i \mathrm{Cov}(U_k,U_i)$$

If the U_k are lognormally distributed with underlying $S_k := \ln(U_k)$ we have (see Clark (2006), Appendix A):

$$E_k := E(U_k) = \exp\{ E(S_k) + \mathrm{Var}(S_k)/2 \}$$

$$E(U_k^2) = E_k^2 \exp\{\mathrm{Var}(S_k)\}$$

$$\mathrm{Var}(U_k) = E_k^2 (\exp\{\mathrm{Var}(S_k)\} - 1)$$

$$\mathrm{Cov}(U_k,U_i) = E_k E_i (\exp\{\mathrm{Cov}(S_k,S_i)\} - 1), \quad i\neq k$$

Corollary A.2: If $U_k = \exp(S_k)$ is lognormally distributed then

$$r.BiasT(Y) = a_0 + \sum_{k=1}^{n} a_k E_k - 1 = a_0 + \sum_{k=1}^{n} a_k \exp\{ E(S_k) + \mathrm{Var}(S_k)/2 \} - 1$$

$$r.VarT(Y) = \sum_{k=1}^{n} (a_kE_k)^2 [(1+\eta^2) \exp\{\mathrm{Var}(S_k)\} -1] + 2 \sum_{1\le k<i\le n} a_kE_k a_iE_i (\exp\{\mathrm{Cov}(S_k,S_i)\} -1)$$

Let now R_k be an AR(1) process $R_k = r R_{k-1} + (1-r) \mu + \sigma E_k$ with a standard normal random error E_k. Let S_k be the integrated AR(1) series $S_k = R_1 + \dots + R_k$. Then $U_k = \exp(S_k)$ is the model proposed for the inflation gap in section 3, having the same structure as the models applied by Wilkie (1995) and Clark (2006) to inflation indexes. Thus the U_k are in particular lognormally distributed (strictly speaking after having conditioned on the "starting value" R_0) and we get (see Clark (2006), Appendix A):

Corollary A.3: If $\ln U_k - \ln U_{k-1}$ is an AR(1) time series then (conditionally on R_0) the formulae of Cor. A.2 hold with:

$$E(S_k) = R_0 \sum_{j=1}^{k} r^m + \mu (1-r) \sum_{j=1}^{k} \sum_{m=0}^{j-1} r^m$$

$$\mathrm{Var}(S_k) = \sigma^2 \sum_{j=1}^{k} \left(\sum_{m=0}^{j-1} r^m\right)^2$$

$$\mathrm{Cov}(S_k,S_i) = \sigma^2 \sum_{j=1}^{k} \left(\sum_{m=0}^{j-1} r^m\right)\left(\sum_{m=0}^{j+i-k-1} r^m\right), \quad k<i$$

Appendix B

Sample Mean with Random Walk Inflation Gap

Example 1		CV Y	10%		n_0	4,33
		CV W	4%		d	0,0007997

n	1	2	3	**4**	5	6	7	8	9	10	12	15	20	30	50
r.Bias Tn	0,08%	0,12%	0,16%	0,20%	0,24%	0,28%	0,32%	0,36%	0,40%	0,44%	0,52%	0,64%	0,84%	1,25%	2,07%
1000 r.Var Tn	11,63	7,03	5,85	5,54	5,56	5,76	6,06	6,43	6,83	7,26	8,20	9,69	12,32	17,88	29,69
1000 phi(n)	11,63	7,02	5,84	**5,52**	5,54	5,73	6,02	6,37	6,76	7,18	8,07	9,50	11,99	17,15	27,68
1000 r.mse Tn	11,64	7,03	5,86	**5,54**	5,57	5,77	6,07	6,44	6,85	7,28	8,22	9,73	12,39	18,04	30,11

Example 2		CV Y	20%		n_0	8,66
		CV W	4%		d	0,0007997

n	1	2	3	4	5	6	7	8	**9**	10	11	12	15	20	50
r.Bias Tn	0,08%	0,12%	0,16%	0,20%	0,24%	0,28%	0,32%	0,36%	0,40%	0,44%	0,48%	0,52%	0,64%	0,84%	2,07%
1000 r.Var Tn	41,73	22,10	15,92	13,10	11,62	10,82	10,40	10,23	10,22	10,32	10,50	10,75	11,74	13,87	30,34
1000 phi(n)	41,73	22,10	15,91	13,08	11,60	10,78	10,36	10,17	**10,14**	10,23	10,39	10,62	11,55	13,54	28,33
1000 r.mse Tn	41,73	22,10	15,92	13,10	11,63	10,83	10,41	10,24	**10,23**	10,34	10,52	10,77	11,78	13,95	30,77

Example 3		CV Y	50%		n_0	21,65
		CV W	4%		d	0,0007997

n	1	2	3	4	5	7	10	15	20	21	**22**	23	24	25	50
r.Bias Tn	0,08%	0,12%	0,16%	0,20%	0,24%	0,32%	0,44%	0,64%	0,84%	0,88%	0,92%	0,97%	1,01%	1,05%	2,07%
1000 r.Var Tn	252,40	127,61	86,37	66,02	54,03	40,79	31,69	26,10	24,73	24,70	24,73	24,80	24,92	25,07	34,90
1000 phi(n)	252,40	127,60	86,35	66,00	54,00	40,74	31,60	25,91	24,40	24,33	**24,32**	24,36	24,44	24,55	32,87
1000 r.mse Tn	252,40	127,61	86,37	66,02	54,03	40,80	31,71	26,15	24,81	**24,78**	24,81	24,90	25,02	25,18	35,33