SEPARATION OF INFLATION AND OTHER EFFECTS FROM THE DISTRIBUTION OF NON-LIFE INSURANCE CLAIM DELAYS
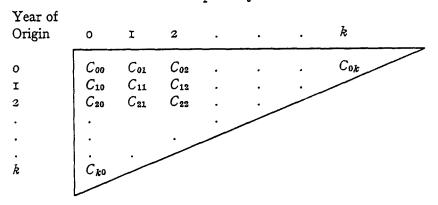
G. C. TAYLOR

Macquarie University, Sydney, Australia
and Government Actuary's Department, London, U.K.

1. THE RUN-OFF TRIANGLE

In recent years, as a result of more concentrated research together with the ravages wrought upon some insurers by inflation, the fundamental significance of the so-called run-off triangle in the calculation of provisions for outstanding claims has been increasingly recognised. The run-off triangle, which is a two-way tabulation—according to year of origin and year of payment—of claims paid to date, has the following form, where C_{ij} is the amount paid by the end of development year j in respect of claims whose year of origin is i, i.e. C_{ij} is the total amount paid in year of origin i and the following j years.

Development year

Year of Origin	0	1	2	.	.	.	k
0	C_{00}	C_{01}	C_{02}	.	.	.	C_{0k}
1	C_{10}	C_{11}	C_{12}	.	.	.	
2	C_{20}	C_{21}	C_{22}	.	.		
.	.		.				
.	.		.				
.	.	.					
k	C_{k0}						

The information relating to the area below this triangle is unknown since it represents the future development of the various cohorts of claims.

2. The "Chain-ladder Method" for Outstanding Claims Provision

Consider the problem of estimating $C_{i\infty}$ for $i = 0, 1, 2, \ldots, k$, given the above run-off triangle. The various methods of tackling this problem exploit the fact (Beard, 1974; Clarke 1974) that, in the absence of exogeneous influences such as monetary inflation, changing rate of growth of a fund, changing mix of business in a fund etc., the distribution of delays *) between the incident giving rise to a claim and the payment of that claim remains relatively stable in time. In this case the columns (or rows) of the run-off triangle are, apart from random fluctuation, proportional to one another.

One method which is based upon this assumption, and the further assumption that the "exogeneous influences" referred to above are not too great, is the so-called *chain ladder method*. According to this method we calculate the ratios

$$\hat{M}_j = \left(\prod_{h=j}^{k-1} \hat{m}_h \right) \hat{M}_k, \tag{1}$$

where $\hat{M}_j$ is an estimate of $C_{i\infty}/C_{ij}$ and $\hat{m}_h$, an estimate of $C_{i,h+1}/C_{ih}$, is calculated as:

$$\hat{m}_h = \sum_{i=0}^{k-i-1} C_{i,h+1} / \sum_{i=0}^{k-i-1} C_{ih}. \tag{2}$$

$\hat{M}_k$ needs to be calculated from (inter alia) an *estimate* of outstanding claims at the end of development year k. Although an important issue, this does not affect the reasoning of this paper and so does not receive detailed comment at this point. The factors $\hat{M}_j$ can now be used to calculate outstanding claims provisions. The outstanding claims provision in respect of year of origin i is:

$$C_{i,k-i}(\hat{M}_{k-i} - 1).$$

3. Difficulties arising from the Chain-ladder Method

It is crucial to the logic underlying the chain-ladder method that the "exogeneous influences" should not be too great. If this

---

*) These "delays" do not refer to any deliberate delaying on the part of the insurer, but to delays in notification of the claim by the insured and further delays caused by litigation, etc.

assumption does not hold, then the conclusion, that the columns of the run-off triangle are proportional goes awry too, and the chain-ladder method can give misleading results. This criticism has been made and illustrated by Clarke (1974), who demonstrated the effects of a large rate of growth and large and volatile rate of inflation.

One possible method of overcoming this weakness of the chain-ladder is to recognise the variation (with i) of the ratios $C_{i,h+1}/C_{ih}$, to seek trends in these rations and project these trends. This modification too has a serious drawback in that the trend may be almost entirely due to monetary inflation, and if rates of inflation have fluctuated in the past, there will not exist any smooth trend. Furthermore, if the rate of inflation is thought likely to fall (say) during the next few years, then it is not clear how this trend should be reflected in the sequence (over i) of ratios $C_{i,h+1}/C_{ih}$.
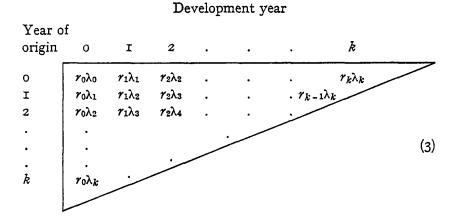
4. The "Separation Method"

Clearly, it would be preferable to separate, if possible, the basic stationary claim delay distribution from the exogeneous influences which are upsetting the stationarity. This can be done as shown below.

We assume that, if the conditions affecting individual claim sizes remained always constant, then the ratios of average claim amount paid in development year j per claim with year of origin i to the average amount paid to the end of development year k per claim with year of origin i would have an expected value r_j which is stationary, i.e. independent of i.

We further assume that claims cost of a particular development year is proportional to some index which relates to the year of payment rather than the year of origin. This is particularly appropriate when claims cost is dominated by high rates of inflation. It is not so appropriate in respect of influences such as changing mix of business within a risk group, which is related rather to policy year. This point receives further comment later in Section 7.

According to the assumptions made above, the expected claims cost of development year j per claim with year of origin i is $r_j \lambda_{i+j}$ where λ_k is exogeneity index—that is an index of the effect of exogeneous influences—appropriate to year of payment k. These

expected values then form the following run-off triangle (but note that claim amounts in this triangle are *not* cumulative for each year of origin).

<div align="center">Development year</div>

Year of
origin 0 I 2 . . . k

$$
\begin{array}{l|cccccccc}
0 & r_0\lambda_0 & r_1\lambda_1 & r_2\lambda_2 & \cdot & \cdot & \cdot & & r_k\lambda_k \\
\text{I} & r_0\lambda_1 & r_1\lambda_2 & r_2\lambda_3 & \cdot & \cdot & & r_{k-1}\lambda_k & \\
2 & r_0\lambda_2 & r_1\lambda_3 & r_2\lambda_4 & \cdot & \cdot & \cdot & & \\
\cdot & & \cdot & & & & & & \\
\cdot & & \cdot & & & \cdot & & & \\
\cdot & & \cdot & & \cdot & & & & \\
\cdot & & & & & & & & \\
k & r_0\lambda_k & & \cdot & & & & &
\end{array}
$$

<div align="right">(3)</div>

The problem now is to separate the values $r_0, r_1, \ldots, r_k$ from $\lambda_0, \lambda_1, \ldots, \lambda_k$ using only the corresponding triangle of observed values

$$s_{ij} = (C_{ij} - C_{i,j-1})/n_i,$$

where n_i is the number of claims with the year of origin i.

This number n_i can be a little problematic. In practice, the total number of claims for year of origin i will not be known until a much later development year than the one just completed. Therefore, it will be necessary to take n_i to be the sum of reported claims and outstanding claims. But at which development year? It may at first seem logical to take both of these figures as at the end of the latest development year available. However, this latest development year decreases as year of origin increases. If, as sometimes happens, a company tends to overestimate (say) the number of outstanding claims in the early development years, then, even if $\lambda_0 = \lambda_1 = \ldots = \lambda_k$, the triangle of s_{ij}'s will tend to increase as one move down the columns. The result would be underestimation of the λ_h's and hence of the provisions for outstanding claims. Thus, to ensure consistency down columns of the s_{ij} triangle it seems necessary to take.

n_i = number of claims settled in development year 0 + estimated number of claims outstanding at end of development year 0 (both in respect of year of origin i).

5. HEURISTIC SOLUTION OF THE SEPARATION PROBLEM

First note that, by definition,

$$\sum_{j=0}^{k} r_j = 1. \tag{4}$$

Hence if we sum along the diagonal involving λ_k, we obtain:

$$d_k = \lambda_k(r_0 + r_1 + \ldots + r_k) = \lambda_k.$$

Thus our estimate of λ_k is:

$$\hat{\lambda}_k = d_k.$$

If the next diagonal up is summed, the result is:

$$d_{k-1} = \lambda_{k-1}(r_0 + r_1 + \ldots + r_{k-1}) = \lambda_{k-1}(1 - r_k).$$

Thus λ_{k-1} could be estimated if only we knew r_k. But an obvious estimate of r_k is:

$$\hat{r}_k = v_k/\hat{\lambda}_k,$$

where v_k is the sum of the column of the triangle involving r_k.

Now,

$$\hat{\lambda}_{k-1} = d_{k-1}/(1 - \hat{r}_k).$$

This procedure can be repeated, leading to the general solution:

$$\hat{\lambda}_h = d_h/(1 - \hat{r}_k - \hat{r}_{k-1} - \ldots - \hat{r}_{h+1}); \tag{5}$$

$$\hat{r}_j = v_j/(\hat{\lambda}_j + \hat{\lambda}_{j+1} + \hat{\lambda}_k), \tag{6}$$

where d_h is the sum along the $(h + 1)$-th diagonal and v_k is the sum down the $(h + 1)$-th row.

6. RELATION TO VERBEEK'S PROBLEM

Verbeek (1972) considered a similar problem in which s_{ij} was number of claims reported in development year j in respect of year of origin i. He assumed the triangle of expected values of s_{ij}'s to have the same structure as that displayed in (3) and, as in our case, sought estimates of the r_j's and λ_h's. He assumed

further that the total number of claims relating to any one year of origin has a Poisson distribution. Then, employing the method of maximum likelihood estimation, he obtained (5) and (6) as estimates of λ_h and r_j respectively.

Verbeek's analysis can be generalised slightly so as to make it appropriate to claim amounts rather than claim numbers. In particular, if in the model of Section 4, we denote $E[s_{ij}]$ by μ_{ij} and if the likelihood of individual claim size can be represented approximately by a function of the form:

$$f(s_{ij} \mid \mu_{ij}) = g(s_{ij}) \, \mu_{ij}^{s_{ij}} \exp\,[-\mu_{ij}],\ s_{ij} > 0,$$

then all of the working goes through once again to produce estimates (5) and (6).

This observation provides ground for expecting (5) and (6) by reasonable estimators from a statistical viewpoint. Conversely, the development of Section 5 provides a readily understood heuristic basis for Verbeek's statistical analysis.

7. AN EXTENDED SEPARATION MODEL

It was mentioned in Section 4 that there are some influences at work which tend to make claim sizes vary by year of origin as well as by year of payment. We could construct a model to acknowledge this by representing the (i, j)-element of triangle (3) by the form:

$$q_i r_j \lambda_{i+j},$$

with the q_i's normalised so that

$$\sum_{i=0}^{k} q_i = 1.$$

However, this not only produces computational difficulties, but also reduces the number of degrees of freedom from $\frac{1}{2}k(k-1)$ to $\frac{1}{2}k(k-3)$. Thus even with a 5×5 triangle containing 15 entries, the number of degrees of freedom in the estimation is only 2.

For these reasons it seems that the extended model is inappropriate and that the model described in Section 4 should be used as being closer to reality.

8. APPLICATION OF THE SEPARATION METHOD

It is now necessary to consider the application of the estimates $\hat{\lambda}_h$, $\hat{r}_j$ to the calculation of provisions for outstanding claims. They

can be applied immediately to complete each row up to and including development year k.

Later development years cause some difficulty. Suppose we write

$$s_{ik+} = \sum_{j=k+1}^{\infty} s_{ij}.$$

Then

$$E[s_{ik+}] = \sum_{j=k+1}^{\infty} E[s_{ij}] = \sum_{j=k+1}^{\infty} r_j \lambda_{i+j}.$$

Since we have no information in respect of the development years involved here except that included in any estimate of total claims outstanding as at the end of the latest development year, it is not possible to separate the r_j's and the λ_h's precisely. This is a verbal expression of the fact that

$$\frac{E[s_{ik+}]}{E[s_{gk+}]} = \frac{\sum\limits_{j=k+1}^{\infty} r_j \lambda_{i+j}}{\sum\limits_{j=k+1}^{\infty} r_j \lambda_{g+j}} \tag{7}$$

does not in general simplify. It is useful to note, however, that if it is assumed that $\lambda_h = \text{const.} \times (1+K)^h$ for the next few years into the future, then (7) simplifies to

$$\frac{E[s_{ik+}]}{E[s_{0k+}]} = (1+K)^i,$$

and so s_{ik+} is estimated by

$$\hat{s}_{ik+} = \hat{s}_{0k+}(1+K)^i.$$

In case variable inflation rates are required for future years, it will usually be sufficiently accurate, unless the claim delay distribution has an extremely long tail, to take

$$\hat{s}_{ik+} = \hat{s}_{0k+}(\lambda_{k+i+1}/\lambda_{k+1}), \tag{8}$$

particularly in view of the uncertainty of the values of λ_h in future years.

It is still necessary to obtain $\hat{s}_{0k+}$, an estimate of s_{0k+}. This can be done by simply setting

$$\hat{s}_{0k+} = s_{0k+}. \tag{9}$$

15

It might be objected that this makes no use of the company's estimate of outstanding claim account in respect of years of origin later than o and that s_{ik+} should first be estimated by:

$$\hat{\hat{s}}_{ik+} = s_{i,k-i+} - (\hat{s}_{i,k-i+1} + \ldots + \hat{s}_{i,k}),$$

and then s_{0k+} estimated by some (possibly weighted) average of the values of $\hat{\hat{s}}_{ik+}(\lambda_{k+1}/\lambda_{k+i+1})$.

However, although this method makes use of more information than does method (9), is also has a couple of drawbacks. Firstly, $\hat{\hat{s}}_{i,k+}$ is dependent upon the values of λ_h for future years, and is therefore suspect to the extent that the λ_h's used explicitly in the calculations are inconsistent with those implicit in the claims adjuster's estimates of outstanding liabilities. This can be particularly important if its effect is to produce estimates $\hat{\hat{s}}_{ik+}$ which are biased on the low side, for this means that the resulting estimate of s_{0k+} will also be low and hence all the estimates $\hat{s}_{ik+}$ will be too low.

For these reasons it may often (for a supervisory authority, always) be advisable to use formula (9) in conjunction with (8).

Having calculated the matrix:

$$
\begin{array}{cccccc}
\hat{s}_{00} & \hat{s}_{01} & \hat{s}_{02} & \ldots & \hat{s}_{0k} & \hat{s}_{0k+} \\
\hat{s}_{10} & \hat{s}_{11} & \hat{s}_{12} & \ldots & \hat{s}_{1k} & \hat{s}_{1k+} \\
\cdot & \cdot & \cdot & & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot \\
\hat{s}_{k0} & \hat{s}_{k1} & \hat{s}_{k2} & \ldots & \hat{s}_{kk} & \hat{s}_{kk+},
\end{array}
$$

we are in a position to calculate factors which correspond to the chain ladder factors. We calculate

$$M_{ij} = (\hat{s}_{i0} + \ldots + \hat{s}_{ik} + \hat{s}_{ik+})/(\hat{s}_{i0} + \ldots + \hat{s}_{ij}).$$

Note that, in principle, there is a different sequence of such factors, $\hat{M}_{i0}$, $\hat{M}_{i1}$, etc., for each year of origin i. In fact, however, we require only one of these factors for each year of origin, and estimate the outstanding claims provision in respect of year of origin i by:

$$C_{i,k-i}(M_{i,k-i} - \mathrm{I}).$$

9. COMPARISON WITH OTHER METHODS

Section 3 dealt with a couple of difficulties arising out of use of the chain-ladder method. These difficulties concerned that method's characteristic of not making past experienced and future expected exogeneity factors explicit. The separation method overcomes this major objection by calculating estimates of these factors from past data (in the λ_h's) and allowing flexibility in the choice of future exogeneity factors.

However, once the λ_h's have been estimated, the method becomes essentially similar to the chain-ladder method in the calculation of the $\hat{M}$ factors and their use in estimating appropriate outstanding claim provisions. Hence, it is reasonable to regard the separation method as simply a variant of the chain ladder method with provision for explicit recognition of exogeneous influences.

It was already noted in Section 3 that the chain ladder technique had been strongly criticised by Clarke (1974), and it is, therefore, of some interest to compare the methods recommended by him with the separation method. Indeed, an examination of Clarke's methods (1974; Clarke and Harland 1974) shows that they are based on principles very similar to those of the separation method. There are two main differences. Firstly, Clarke deals with monthly data, rather than the annual data used here. This is not an essential difference, the choice of frequency of data collection being dictated by practical considerations. Clearly monthly figures are preferable but, for a supervisory authority such as the UK Department of Trade, not possible.

The second main difference is perhaps in favour of the separation method. It consists in the fact that the estimation of past rates of inflation (as part of the λ_h's) from past data is integrated into the whole estimation procedure, whereas it is not entirely clear whence Clarke obtains them. Moreover, the "exogeneity factors" employed here incorporate not only inflation but *all* influences on the distribution of claims delays.

10. NUMERICAL RESULTS

The method developed here was applied to a number of cases which had proved difficult to handle by other methods. In nearly

all cases, satisfying results were obtained. Two examples are given below—one in which results were satisfactory, and one in which they were unsatisfactory.

Example 1: A Motor Account

The run-off triangle is:

	0	1	2	3
0	50.4	28.2	9.0	4.8
1	58.0	29.2	9.7	
2	59.5	33.2		
3	66.2			

This yields:
$$v_0 = 234.1; \qquad d_0 = 50.4;$$
$$v_1 = 90.6; \qquad d_1 = 86.2;$$
$$v_2 = 18.7; \qquad d_2 = 97.7;$$
$$v_3 = 4.8; \qquad d_3 = 113.9.$$

Hence,
$$r_0 = 0.5835; \qquad \lambda_0 = 86.4;$$
$$r_1 = 0.2878; \qquad \lambda_1 = 98.9;$$
$$r_2 = 0.0866; \qquad \lambda_2 = 102.0;$$
$$r_3 = 0.0421; \qquad \lambda_3 = 113.9.$$

The "fitted run-off triangle" based on these 8 parameters is:

	0	1	2	3
0	50.4	28.5	8.8	4.8
1	57.7	29.4	9.9	
2	59.5	32.8		
3	66.5			

This fits the original triangle well, which is reassuring. On the other hand, however, it must be remembered that there are only 3 degrees of freedom in the fitting process and so the fit is forced to a considerable extent.

Perhaps just as important as the goodness of fit is the requirement that the r's and λ's produced from the triangle which includes only the first 3 rows and first 3 columns of the above 4×4 triangle should be consistent with the r's and λ's already calculated. This 3×3 triangle produces

$$r_0 = 0.6101; \qquad \lambda_0 = 82{,}6;$$
$$r_1 = 0.2980; \qquad \lambda_1 = 94.9;$$
$$r_2 = 0.0921; \qquad \lambda_2 = 97.7.$$

Now these values do not agree immediately with those already calculated. However, this is principally due to the constraint

$$\sum_{j=0}^{k} r_j = 1,$$

which means that $r_0 + r_1 + r_2 = 1$ for the 3 × 3 triangle, whereas $r_0 + r_1 + r_2 = 1 - 0.0421$ for the 4 × 4 triangle. We can multiply all of our r's by some constant, and provided we divide all λ's by the same constant, the scaled results will be equivalent to the unscaled ones. Choosing this constant to be $(1 - 0.0421)$, we rescale the last set of r's and λ's to obtain:

$$r_0 = 0.5844; \qquad \lambda_0 = 86.2;$$
$$r_1 = 0.2855; \qquad \lambda_1 = 99.1;$$
$$r_2 = 0.0882; \qquad \lambda_2 = 102.0.$$

These figures agree very well with those calculated previously.

If it is assumed that λ will increase in future at a rate of 10% per annum, then

$$\lambda_4 = 125.3, \ \lambda_5 = 137.8, \ \lambda_6 = 151.6, \ \lambda_7 = 166.8.$$

The procedure described in Section 8 may now be applied and the rectangle

	0	1	2	3	3 +
0	50.4	28.5	8.8	4.8	7.6
1	57.7	29.4	9.9	5.3	8.4
2	59.5	32.8	10.8	5.8	9.2
3	66.5	36.1	11.9	6.4	10.2

obtained,

$$\hat{M}_{0,3} = 1.082$$
$$\hat{M}_{1,2} = 1.141$$
$$\hat{M}_{2,1} = 1.281$$
$$\hat{M}_{3,0} = 1.971$$

Example 2: A Pecuniary Loss Account

	0	1	2	3
0	231.1	336.6	237.3	975.1
1	9435.3	3902.2	89.9	
2	70.8	234.6		
3	82.5			

This yields: $r_0 = 0.1866$; $\cdot\lambda_0 = 1238.5$;
              $r_1 = 0.0870$; $\lambda_1 = 35716.0$;
              $r_2 = 0.0209$; $\lambda_2 = 14296.4$;
              $r_3 = 0.7055$; $\lambda_3 = 1382.1$,

which leads to the following fitted triangle:

	0	1	2	3
0	231.1	3107.3	298.8	975.1
1	6664.6	1243.8	28.9	
2	2667.7	120.2		
3	257.9			

This does not agree well with the actual run-off figures, the reason being that, under the assumption of r_j's being unrelated to year of origin, line 1 of the actual run-off triangle is grossly inconsistent with lines 2 and 3.

11. ACKNOWLEDGEMENTS

REFERENCES

BEARD, R. E. (1974). Some historical, theoretical and practical aspects. Appears in *Claims provisions for non-life insurance business*. Institute of Mathematics and its Applications.

CLARKE, T. G. (1974). An actuary looks at claims provisions in general insurance. Appears in *Claims provisions for non-life insurance business*. Institute of Mathematics and its Applications.

CLARKE, T. G. and HARLAND, N. (1974). A practical statistical method of estimating claims liability and claims cash flow. *Astin Bulletin*, 8, 26-37.

VERBEEK, H. G. (1972). An approach to the analysis of claims experience in motor liability excess of loss reassurance. *Astin Bulletin*, 6, 195-202.