TESTING GOODNESS-OF-FIT OF AN ESTIMATED RUN-OFF TRIANGLE

G. C. TAYLOR
Sydney, Australia

1. THE RUN-OFF TRIANGLE — ACTUAL AND EXPECTED

By the term *actual run-off triangle* we shall mean the two-way tabulation—according to year of origin and year of payment—of claims paid to date, which has the following form:

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}						

where C_{ij} is the amount paid during development year j in respect of claims whose year of origin is i.

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

Now in seeking to use this triangle as a basis for projection of claims in future development years for each of the years of origin 0, 1, 2, etc., we must recognise that the entries C_{ij} in the above triangle, being random variables, contain random deviations from their expected values μ_{ij}. It is the corresponding triangle of these expected values in which we are interested, and which shall be called the *expected run-off triangle*.

Explicitly, it is:

μ_{00}	μ_{01}	μ_{02}	.	.	.	μ_{0k}
μ_{10}	μ_{11}	μ_{12}	.	.	.	
μ_{20}	μ_{21}	μ_{22}	.	.		
.			.			
.		.				
.	.					
μ_{k0}						

https://doi.org/10.1017/S0515036100006371 Published online by Cambridge University Press

2. The Requirement of a Test of Goodness-of-fit.

One method of projecting future claims is to identify some internal structure within the expected run-off triangle and hence extrapolate outside it. In this respect, a commonly made assumption is the following:

Assumption 1

In the absence of any disturbing influences, e.g. claims cost inflation, changing rate of growth of volume of business etc., the distribution of expected claim delays remains constant over varying years of origin.

We can represent this assumption symbolically. If R_{ij} is the observed proportion of all claim payments in respect of year of origin i made in development year j after removal of the "disturbing influences" referred to above, then $E(R_{ij}) = r_j$ independent of i. Examples of estimation procedures based on this assumption can be found in Beard (1974) and Taylor (1977).

Naturally, if a model based on Assumption 1 is to be used for projection of future claims, it is necessary to check at some stage that this model accords with experience (i.e. that the expected run-off triangle based on the model accords with the actual run-off model) within statistically reasonable limits. Hence the need for a test of goodness-of-fit.

Suppose that the "disturbing influences" in the triangle have been determined so that it is possible to remove them from the data. Let C'_{ij} be the result of adjusting C_{ij} for removal of these influences. Then, according to Assumption 1,

$$\mu'_{ij}(= E(C'_{ij})) = C'_i r_j,$$

where C'_i denotes total claims (some still to be paid) in respect of year of origin i after removal of disturbing influences.

Estimation procedures based on Assumption 1 will produce estimates $\hat{\mu}'_{ij}$ of μ'_{ij}, where $\hat{\mu}'_{ij} = C'_i\,\hat{r}_j$ and $\hat{r}_j$ is an estimate of r_j. It is then necessary to apply a significance test to the deviations $(C'_{ij} - \hat{\mu}'_{ij})$.

One tempting possibility is to set up a contingency table containing the cells as displayed below:

3. A Contingency Table test?

$(0, 0)$	$(0, 1)$	$(0, 2)$	. . .	$(0, k)$	$(0, k+)$
$(1, 0)$	$(1, 1)$	$(1, 2)$	. . .	$(1, k-1)$	$(1, (k-1)+)$
$(2, 0)$	$(2, 1)$	$(2, 2)$	. . .	$(2, (k-2)+)$	
⋮					
$(k, 0)$	$(k, 0+)$				

Here the $(i, (k-i)+)$ cell relates to data for year of origin i and development years $k-i+1$, $k-i+2$, etc. combined. The standard chi-square test might then be applied to this table as in the theorem in Section 30.3 of Cramér (1946, 426-7).

There are, however, several points to be noted in connection with this suggestion.

Firstly, the triangle of previous sections has been augmented with extra cells to form a square. This has been done in conformity with the theorem quoted above which requires that for a given year of origin, the probability of a randomly chosen unit of claim payment being found in some cell of the table should be unity. This augmentation of the triangle can cause difficulties because data may not be available in respect of the extra cells. This point receives further comment in the later section dealing with numerical examples.

Secondly, and more importantly, it is implicit in the theorem quoted above (see both the statement of it on P. 427 and the proof on P. 429) that the marginal distribution of each C'_{ij} is binomial. In the present circumstances this is not true and, in fact, is sufficiently untrue to have important consequences for the contingency table test, as will be dealt with in the next section.

Thirdly, an examination of the theorem stated by Cramér reveals that the chi-square test is strictly applicable only when the expected cell frequencies have been determined by the modified χ^2 minimum method of estimation. When this method has not in fact been used, some consideration should be devoted to the closeness of this and the method actually used. For example, the "separation method" used by Taylor (1977) is not always equivalent to the modified χ^2 minimum method, but is, as shown in Section 6 of that paper, identical in certain cases to the maximum likelihood method which, as pointed out by Cramér (1946, 426), is in turn equivalent to the modified χ^2 minimum method.

4. Modification of the Standard Chi-square Test of a Contingency Table.

The most important of the objections raised against the standard chi-square test is the second which concerns the marginal distributions of the individual cell frequencies. As noted there, the standard test requires that the (i, j) - cell frequency be binomial. The parameters of this binomial distribution would be C'_i and r_j, and hence the variance would be

$$v_{ij} = C'_i \, r_j \, (1 - r_j) = \mu'_{ij} \, (1 - r_j) \tag{1}$$

As also noted in the previous section, the distribution of C'_{ij} will not be binomial in fact. In order to approximate its correct form we make two further assumptions.

Assumption 2

The number of claims pertaining to the (i, j) - cell is a stationary Poisson variable.

Assumption 3

The sizes of the individual claims pertaining to the (i, j) - cell are i. i. d. random variables.

It follows from these two assumptions that C'_{ij} is a compound Poisson variable with variance:

$$\sigma^2_{ij} = \mu'_{ij} \times \frac{\alpha_{2j}}{\alpha_{1j}} \tag{2}$$

where α_{1j}, α_{2j} are the first and second moments (about the origin) respectively of individual claim size in development year j.

It is now evident that in those cases where μ'_{ij} is not too small the compound Poisson distribution of C'_{ij} and the binomial distribution with the same mean and variance (1) will be rather similar except that the former will have a variance greater than that of the latter by a factor of

$$\frac{\sigma^2_{ij}}{v_{ij}} = \frac{\alpha_{2j}}{\alpha_{1j}(1 - r_j)}. \tag{3}$$

Thus, if the standard chi-square statistic,

$$\sum_{\text{all cells}} (C'_{ij} - \mu'_{ij})^2 / \mu'_{ij}$$

is replaced by:

$$\chi^2 = \sum_{all\ cells} (C'_{ij} - \mu'_{ij})^2 v_{ij} / \sigma^2_{ij} \mu'_{ij}$$

$$= \sum_{all\ cells} (1 - r_j) \left(\frac{\alpha_{1j}}{\alpha_{2j}}\right) (C'_{ij} - \mu'_{ij})^2 / \mu'_{ij}, \qquad (4)$$

then χ^2 can be assumed to have an approximate chi-square distribution with an appropriate number of degrees of freedom.

Suppose that it is desired that a significance test be applied to the

Null Hypothesis: $r_j = \hat{r}_j$ for each j.

Then it follows from (4) and the hypothesis that

$$\hat{\chi}^2 = \sum_{all\ cells} (1 - \hat{r}_j) \left(\frac{\alpha_{1j}}{\alpha_{2j}}\right) (C'_{ij} - \hat{\mu}'_{ij})^2 / \hat{\mu}'_{ij} \qquad (5)$$

is a chi-square statistic and can be tested as such for significance.

5. Applying the Modified Test in Practice.

All quantities appearing in statistic (5) are immediately available with the exception of the ratio $(\alpha_{1j}/\alpha_{2j})$. If the investigation is being carried out by an individual company in respect of its own experience, then this ratio can be estimated by means of a cost-band analysis of claims.

On the other hand, if the test is being applied by a supervisory authority, it is unlikely that any cost-band information will be available for estimation of $(\alpha_{1j}/\alpha_{2j})$. The authority will however have returns from each company and may, therefore, consider ways of estimating the ratio from this data.

The slender evidence to which the author had access (a confidential report) suggested that α_{1j}/α_{2j} was *not* independent of company, but that, for a given class of insurance, the coefficient of variation, $w_j = \alpha_{2j}/\alpha^2_{1j}$, varied comparatively little between different companies. This suggests estimating w_j by $\hat{w}_j$, based on data from all companies and replacing $\hat{\chi}^2$ by the alternative statistic:

$$\hat{\hat{\chi}}^2 = \sum_{all\ cells} \left[\frac{C'_{ij} - \hat{\mu}'_{ij}}{\hat{\mu}'_{ij}}\right]^2 \hat{n}_{ij} \frac{1 - \hat{r}_j}{\hat{w}_j} \qquad (6)$$

where n_{ij} is the expected number of claims paid in development year j of year of origin i, and $\hat{n}_{ij}$ estimates n_{ij}.

The difficulty now is, of course, the estimation of w_j. For this purpose, let

r_{jt} denote the value of r_j in the t-th company (for a particular class of insurance);
C'_{ijt} denote the random variable C'_{ij} in the t-th company;
$C'_{i\cdot t}$ denote the constant C'_i in the k-th company.

Let us suppose that, for fixed j, the r_{jt} are realizations of a random variable with mean p_j and variance z_j. Suppose also that $C_{i_1jk_1}$ and $C_{i_2jk_2}$ are stochastically independent whenever $(i_1, k_1) \neq (i_2, k_2)$.

Then it is not difficult to show that, for each i, j,

$$\begin{aligned} Var\,[C'_{ijt} / C'_{i\cdot t}] &= E_{r_{jt}}\,[Var\,[C'_{ijt} / C'_{i\cdot t} \mid r_{jt}]] \\ &\quad + Var_{r_{jt}}\,[E\,[C'_{ijt} / C'_{i\cdot t} \mid r_{jt}]] \\ &= E_{r_{jt}}\,[w_j\, r^2_{ijt}] + Var_{r_{jt}}\,[r_{jt}]. \end{aligned}$$

i.e.

$$w_j = \frac{Var\,[C'_{ijt} / C'_{i\cdot t}] - Var\,[r_{jt}]}{E[r^2_{jt}]}\,. \tag{7}$$

A reasonable estimate $\hat{w}_j$ of w_j can be obtained by replacing each of the three terms on the right of (7) by an estimator. The first term of the numerator can be estimated from the sample variance of the ratios $(C'_{ijt} / C'_{i\cdot t})$ for fixed j. However, the other two terms present difficulties, since the corresponding sample statistics depend upon the observed values of r_{jt} for companies other than the one to which the significance test is being applied. These r_{jt} are neither known nor the subject of our hypothesis.

The simplest way out of the difficulty appears to be as follows:

1. Use some method which is known to be generally fairly reliable to obtain an estimate of r_{jt} for each j and t.
2. Use these estimates to calculate the sample statistics corresponding to the quantities appearing in (7).
3. Use these sample statistics to obtain an estimate of w_j as already described.

A second practical difficulty arises from the appearance of the quantities C'_i in our formulas. These quantities, being total payments after run-off has been completed, are of course unknown for any cohorts not fully developed.

However, this situation is not quite as serious as it might at first appear. Let us consider the impact of the C_i on each of the terms of (6) in turn.

Firstly,

$$\sum_j (C'_{ij} - \hat{\mu}'_{ij}) = C'_i [\sum_j R_{ij} - \sum_j \hat{r}_j] = 0, \tag{8}$$

since both summations yield unity. Thus,

$$C'_{i,(k-i)+} - \hat{\mu}'_{i,(k-i)+} = - \sum_{j=0}^{k-i} (C'_{ij} - \hat{\mu}'_{ij}), \tag{9}$$

and so the terms $[(C'_{ij} - \hat{\mu}'_{ij}) / \hat{\mu}'_{ij}]^2$ are all fully determined.

Secondly, the term $C'_{i \cdot t}$ appears in $\hat{w}_j$ (see (7)). Here it is possible to use equation (8) again and obtain

$$C'_{i \cdot t} = \sum_j C'_{ijt} = \sum_j \hat{\mu}'_{ijt}. \tag{10}$$

Finally the value n_{ij} can be estimated by $\hat{n}_{ij}$, the *actual number of claims* pertaining to the (i, j) - cell.

All of the terms appearing in (6) are now determined.

6. A Practical Simplification of the Test Statistic.

The procedure outlined in the previous section for estimating w_j is complicated and involves lengthy computations. Moreover, no idea of the stability of the estimate of w_j has been obtained.

However, experience indicates that, even in the relatively stable class of business such as private motor insurance, w_j tends to be rarely less than unity. These occasions on which it is < 1 are usually just those on which r_j is relatively large. The result of this is that usually (always ?) we have

$$\frac{1 - r_j}{w_j} < 1. \tag{11}$$

combining (5) and (11) we see that

$$\hat{\chi}^2 < \sum_{\text{all cells}} \left[\frac{C'_{ij} - \hat{\mu}'_{ij}}{\hat{\mu}'_{ij}} \right]^2 \hat{n}_{ij}, \tag{12}$$

and so deduce that treating the right side of (12) as a chi-square statistic amounts to applying a somewhat too stringent test to the hypothesis. The overstringency is not too great, at least for motor portfolios, as typical values of the factor $(1 - r_j) / w_j$ appear to lie in the range 0.3 to 0.7.

7. A Numerical Example.

Let us apply the simplified test developed in Section 6 to the run-off triangle dealt with in Example 1 of Taylor (1977). The actual triangle with each C_{ij} divided by $10^{-3} \times$ numbers of claims for year of origin i, is:

50.4	28.2	9.0	4.8
58.0	29.2	9.7	
59.5	33.2		
66.2			

Multiplying by claim numbers to obtain the C_{ij} gives:

2481	1387	441	237
2899	1463	485	
3126	1744		
3538			

The calculations in Taylor (1975) yield the following C'_{ij}'s:

2481	1217	374	180
2533	1239	368	
2648	1323		
2684			

and the following array of $\hat{\mu}'_{ij}$'s:

2480	1223	368	179		234
2522	1244	374		420	
2647	1306		833		
2695		2178			

There is a certain degree of arbitrariness in the values of $\hat{\mu}_{i,(3-i)+}$ which were not determined by Taylor (1975). These will not affect the result materially, however.

Finally the triangle of $\hat{n}_{ij}$'s is:

30034	13309	960	393		164
30678	12974	1216		458	
31461	15417		1783		
31386		22045			

From these figures we readily obtain:

$$\sum_{\text{all cells}} \left[\frac{C'_{ij} - \hat{\mu}'_{ij}}{\hat{\mu}'_{ij}}\right]^2 \hat{n}_{ij} = 6.25.$$

Now a value of 6.25 for χ^2_3 is not significant at the 5% level and so, recalling that the true χ^2_3 statistic would be appreciably less than 6.25, we should have no hesitation in accepting that the model produced by the separation technique and leading to the above $\hat{\mu}'_{ij}$'s is quite plausible statistically.

8. Acknowledgements.

The author acknowledges the assistance of Macquarie University, Sydney, Australia and Government Actuary's Department, London, U.K. in the preparation of this paper. Several discussions of the subject of the paper with Professor R. E. Beard proved invaluable.

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.

Cramér, H., (1946). *Mathematical methods in statistics*, edition. Princeton University Press.

Taylor, G. C., (1977). Separation of inflation and other effects from the distribution of non-life insurance claim delays. The ASTIN Bulletin, 9, 219-30.