STATISTICAL REVIEW OF A MOTOR INSURANCE PORTFOLIO

P. D. JOHNSON and G. B. HEY

1. *Introduction*

Motor insurers in the United Kingdom are not subject to Government control over the rating structures they use or the levels of premiums they charge. The market is highly competitive, and each insurer therefore needs to make the best estimate he can of the premium required for each category of risk, to produce a given level of profit. He also needs to estimate the extent to which departures from such premium levels can be justified, for example in order to have a simple rating structure or to meet competition.

The purpose of this note is to give some examples of the statistical tabulations being produced in one insurance office in the United Kingdom. The statistical system which has been developed covers many aspects of motor insurance management, but in this note we shall direct attention to just two, namely

(i) continuous review of the variations in the claims experience from one risk category to another, as a guide to the relative premiums required for the different categories; and

(ii) continuous review of the changing composition of the portfolio and of the movements in and out, to try to assess the extent to which gains and losses of business can be attributed to pricing differences, marketing strategies, etc.

2. *Reviewing the relative claims experience of different risk categories*

The system provides for the separate study of claim frequencies and amounts of claims. The study of amounts of claims is used to produce assumed average amounts of claim which are then associated with the claim frequencies.

An important feature of the method is the use of a model, with a standard set of parameters, to represent the expected claims experience. The parameters used in the model are based on earlier research on the same portfolio analysed by all the main factors simultaneously; a form of least squares fitting was used, and the

resulting parameters were adjusted after studying the experience of a number of successive periods. The actual experience is compared with that expected, quarter by quarter and year by year. An additive model is used for the claim frequency and a similar additive model is used for the average amount of claim. In each case the model provides for parameters for each level of up to nine factors, and also for a constant parameter since although this is not strictly necessary it is convenient to have one. The standard table currently used for claim frequencies has, in addition to the constant parameter, fifty-one different parameters related to eight of the nine factors.

The collection of the data for statistical analysis forms part of a data processing system based on magnetic tape files and incorporating the preparation of policy schedules, of endorsements giving details of changes (e.g. change of car), of renewal notices (premium billing) and accounting documents associated with premium and claim payments, and the automatic calculation of premiums by reference to the factors on which they are based. Numerous checks are incorporated in the system to try to ensure that the data are complete and accurate.

Because of the time which elapses between the date on which insurance begins for a new policy and the date on which the new record appears on the file, and because of the similar kinds of delay which occur with lapses, claims and changes generally, it is impossible for a statistical system to be both up to date and accurate. (It is all too easy to be out of date and inaccurate.) In the system described, no attempt is made to study the experience until at least three months after the end of the period under review; this allows time for most of the changes, including claims, relating to the period to be notified to the file.

Table 1 shows an extract from one of the tables which bring together the results of the analyses of claim frequencies and average amounts of claims. The table refers to private cars insured comprehensively and the period covered is the year 1969. The actual tabulations include other columns, but those shown in the table are sufficient to illustrate the method. The big advantage of using a model to represent the expected experience is that it enables us to look at the experience by one factor at a time, and make reasonable

allowance for the effect of associations between that factor and those in the standard table. However, there is no reason why a factor should not be defined so as to incorporate a simultaneous analysis by two or more of the basic factors, and this is done in a few cases.

TABLE I

Factor level (1)	Exposure in years (2)	Actual no. of claims (3)	Expected no. of claims (4)	1000 × Actual ÷ Expected claims (5)	Actual claim frequency per 1000 (6)	Adjusted claim frequency per 1000 (7)	Assumed claim cost £00 (8)	Premiums £00 (9)	Claim ratio per 1000 (10)
I	15581	1339	1246	1075	86	137	886	1912	463
2	194393	26851	26905	998	138	136	22065	32340	682
.	...	...	...	...	...	...	...	...	...
.	...	...	...	...	...	...	...	...	...
All	470416	68012	68012	1000	145	...	57417	83781	685

Notes to Table 1

Col. 2. is based on quarterly censuses of the policies in force; each census is taken 3 months after the end of the quarter.

Col. 3 is based on the claims which have accident dates in the quarter and have been notified to the file when the census for the end of the quarter is taken. A further count is made 3 months later to test the proportion of late notifications.

Col. 4 is obtained by calculating an expected annual claim frequency for each policy in force on each census date, and averaging the totals from successive censuses in the same way as for the exposures in col. 2. The expected annual claim frequencies are calculated from a set of standard parameters stored in the computer, using an additive model. The expected numbers of claims are multiplied by a scaling factor to make the total expected number of claims equal to the total actual number of claims for each period, for comprehensive and third party policies separately. Thus the figure on the total line in col. 4 is the same as that in col. 3.

Col. 5 = 1000 × col. 3 ÷ col. 4.

Col. 6 = 1000 × col. 3 ÷ col. 2.

Col. 7 is calculated by adding to col. 6 an adjustment to allow for the composition of the portfolio for the particular level of the

factor being considered. It is intended to show the variation in claim frequency which can be attributed to the factor being considered, after allowing for the associations which exist between that factor and those (or those others) to which the standard parameters relate. In the example given, factor levels 1 and 2 refer to catogories of cars which have similar characteristics, but the average age of the cars in level 1 is much higher. Comparison of cols. 6 and 7 suggests that the difference in the actual claim frequency between levels 1 and 2 can be attributed to factors other than the one being considered— particularly the age of car.

Col. 8 is obtained by applying to each of the actual claims counted in col. 3 an expected average claim amount, using a set of standard parameters and an additive model similar to that used for the expected claim frequency. The parameters for expected average claim amounts are based on a separate analysis of amounts of claims.

Col. 9 is analogous to col. 4. It is obtained by calculating, for each policy in force on each census date, the annual rate of premium corresponding to the status of the policy on the census date. The totals from successive censuses are then averaged in the same way as for the exposures and expected numbers of claims. A single premium scale has been applied at each of the censuses used in the above table, although there was a change in premium scale during 1969.

Col. 10 = 1000 × col. 8 ÷ col. 9.

3. *The study of amounts of claims*

There are two main difficulties in studying average amounts of claims. One is the fact that of the claims which occurred in any recent period a proportion will still be outstanding, and the uncertainty regarding the amount for which they will ultimately be settled can represent a considerable proportion of the total claims cost. The other difficulty concerns the treatment of large claims. Some adjustment is clearly needed to compensate for the fact that certain risk categories have by chance attracted more or less than their usual share of large claims; the difficulty arises when we try to determine what their "usual share" is.

Table 2 refers to claims incurred in 1967 and closed by the end of 1969 on private car comprehensive policies.

TABLE 2

Closed claims

Factor level	0	0—	5—	10—	15—	20—	.	1250—	3000—	No	Amount £ 000	Average inc 0	Average exc 0	P Mil zero
1	509	87	207	213	143	287		7	0	3194	222	69 4	82 6	159
2	2073	460	981	838	628	1024		26	5	12698	932	73 4	87 7	163
3								41	3	17393	1252	72 0	85 9	162
4								26	3	17479	1202	68 8	84 2	183
5								17	3	8991	640	71 2	88 0	191
All	10315	2255	3881	3755	2827	4993		117	14	59755	4248	71 1	85 9	173

Notes to Table 2

1. These tables are produced for nine main categories corresponding to three classes of vehicle (private cars, motor cycles and commercial vehicles) and three categories of cover (comprehensive, third party with fire and theft, and third party only)

2. For each of the nine categories, tables are produced for 9 rating factors (10 in the case of private cars with comprehensive cover).

3. The tables are produced cumulatively every 3 months to include all closed claims processed up to the end of the quarter

4 To save space, some of the columns have been omitted There are 16 columns used in the distribution of claim amounts

After 2 years, tables similar to Table 2 are produced each year to include all closed claims to date and all claims still outstanding with estimates of their ultimate cost Table 3 is an example corresponding to Table 2

TABLE 3

Closed and outstanding claims

Factor level	0	0—	5—	10—	15—	20—	1250—	3000—	No	Amount £ 000	Average inc 0	Average exc 0	P Mil zero
1	510	88	209	213	144	291	12	6	3239	271	83 5	99 2	157
2	2074	463	985	846	635	1041	45	17	12919	1077	83 4	99 3	161
3							66	29	17737	1546	87 1	103 6	159
4							52	18	17773	1365	76 8	93 7	180
5							26	8	9170	713	77 8	95 8	188
All	10332	2272	3900	3780	2861	5078	201	78	60838	4972	81 8	98 4	170

The study of tables of the kinds illustrated in Tables 2 and 3 has shown the need for some process to counteract the disturbing effects of large estimates and large settlements. Experiments have been carried out using different methods of smoothing out fluctuations attributable to large claims. Table 4 shows the results obtained using 6 methods of smoothing.

TABLE 4

Modified average amounts of claim

Factor level	No. of claims	Actual average claim	M_1	M_2	M_3	M_4	M_5	M_6
				Modified average claims				
1	3239	83.5	82.3	83.0	83.4	82.3	83.2	83.4
2	12919	83.4	84.3	84.2	83.2	84.4	83.1	83.3
3	17737	87.1	83.6	82.9	82.4	83.9	83.2	82.6
4	17773	76.8	78.5	79.1	80.0	78.1	79.4	79.8
5	9170	77.8	81.0	81.3	81.7	80.5	81.0	81.6
Total	60838	81.8	81.8	81.8	81.8	81.8	81.8	81.8

Note to Table 4

The smoothing formulae used in Table 4 are as follows (where C is the actual claim cost and S is the modified claim cost):—

M_1 $0 \leqslant C \leqslant 500 : S = C$ $C > 500 : S = 500$

M_2 $0 \leqslant C \leqslant 250 : S = C$ $C > 250 : S = 250$

M_3 $0 \leqslant C \leqslant 100 : S = C$ $C > 100 : S = 100 \log_{10}\left(\dfrac{C}{10}\right)$

M_4 $0 \leqslant C \leqslant 500 : S = C$ $C > 500 : S = 500 \log_{10}\left(\dfrac{C}{10}\right)$

M_5 $0 \leqslant C \leqslant 50 : S = C$ $C > 50 : S = 50 \left(\dfrac{C}{50}\right)^K$

M_6 $0 \leqslant C \leqslant 100 : S = C$ $C > 100 : S = 100 \left(\dfrac{C}{100}\right)^L$

where $K = \log_5 (2.5)$
and $L = \log_3 (1.5)$

Methods M_3 to M_6 apply weights which decrease as C increases. The result is that the actual amount of each large claim has some effect on the mean for the group, whereas this is not true in M_1 and M_2 for claims above the cut-off point.

The smoothed means have been multiplied by a "grossing-up factor" equal to the total actual amount of the claims divided by the total of the smoothed amounts, taken over all levels of all factors. It

is becoming clear that the appropriate grossing-up factors vary between some levels of some of the factors (notably the age of the policyholder), but further research, preferably on a national scale, will be needed to determine the extent of the variation.

Of the methods tried so far, M_6 seems to be the most satisfactory. We consider that the final column of Table 4 can be taken as a fair indication that there is no important difference in average claim between the various levels of the factor concerned.

4. *Reviewing the business gained and lost, and the changing pattern of business in force*

The quarterly censuses used in the study of the claims experience show how the characteristics of the business in force are changing. Further analyses are prepared each month for policies falling due for renewal in the following month, for new policies issued during the past month and for lapses reported during the past month. (For this purpose, a lapse is a policy for which renewal has been invited but not accepted. Policies which are cancelled part-way through the policy year are treated separately.) Table 5 is an example of one such tabulation.

TABLE 5

Month	Factor level					
	1	2	3	4	. . . .	All
10/70	72	258	322	253	. . . .	34870
9/70	72	258	329	252	. . . .	35415
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
10/69	69	253	321	257	. . . .	33947
All (numbers) 33398	127007		. . .	. . .	. . . .	506170
Average	66	251	325	260	. . . .	39352
Trend 3-month	3	3	—2	—2	. . . .	—5012
6-month	2	1	—1	—2	. . . .	—2817
12-month	0	1	1	—1	. . . .	+1022

Notes to Table 5

1. Tables of this kind are produced each month for renewals, lapses and new business separately, in each case for 13 different factors including (for private cars) rating district, car group, age of policyholder and age of car.

2. The entries in the main part of the table are the numbers of cases expressed per 1000 of the total in each line. Thus, on the first line the entry 322 for factor level 3 means that this comprises $\frac{322}{1000} \times$ 34870 cases.

3. The month referred to in the first column is the month for which renewals were being prepared when the figures on the line were extracted. In the above table, produced at the beginning of September 1970, the renewals falling due in October 1970 have been processed and the file has been updated for new business and lapses notified to the computer during August 1970.

4. Data are retained on summary cards and next month a new line (11/70) will be added and 10/69 will be omitted.

5. The trends for the last 3, 6 and 12 months are calculated by the computer by a simple regression method. There are some pronounced seasonal variations in the percentages, although they are much less than the variations in the monthly totals.

Tables of the kind illustrated in Table 6 below are prepared each month for new business and lapses to show both the flows of business in and out and the delays between the effective months and the months of processing.

TABLE 6

"Processing month"	Renewal month							
	9/70	8/70	7/70	6/70	5/70	. . . .	9/69	8/69
10/69	...	...	...	...	...	. . . .	32	3885
11/69	...	...	...	...	...	. . . .	3379	3123
.	...	...	...	...	...	. . . .	...	...
.	...	...	...	...	...	. . . .	...	...
.	...	...	...	...	31	. . . .	...	...
.	...	...	...	14	3486	. . . .	7	7
8/70	...	...	35	4239	5334	. . . .	10	9
9/70	...	16	2446	1778	496	. . . .	1	2
10/70	24	3289	2942	644	433	. . . .	4	6
All	24	3305	5423	6675	9780	. . . .	7844	8431
Renewals	47928	50817	68863	63795	68498	. . . .	47106	50136
Per mille	1	65	79	105	143	. . . .	167	168

Similar tables to the above are prepared in order to adjust for seasonal variation by expressing the figures in the main part of each

table as a proportion of the number of renewals at the foot of each column. An example corresponding to Table 6 is given below in Table 7.

TABLE 7

"Processing month"	Renewal month										
	9/70	8/70	7/70	6/70	5/70	4/70	3/70	2/70	1/70	12/69	....
10/69	...	...	...	...	...	...	...	...	...	...	....
.	...	...	...	...	...	...	...	...	...	...	....
.	...	...	...	...	...	...	...	...	...	...	....
.	...	...	...	...	...	...	...	...	...	...	....
2/70	...	...	...	...	...	...	...	...	...	38	....
3/70	...	...	...	...	...	...	...	...	50	95	....
4/70	...	...	...	...	...	...	...	54	70	18	....
5/70	...	...	...	...	...	...	38	73	17	7	....
6/70	...	...	...	...	...	49	85	23	8	4	....
7/70	...	...	...	...	51	66	13	6	3	1	....
8/70	...	...	1	66	78	15	8	4	1	1	....
9/70	...	...	36	28	7	3	2	1	...	...	....
10/70	1	65	43	10	6	3	1	1	...	...	....
Per mille	1	65	79	105	143	136	148	161	149	165	...

Note to Table 7

The figures still to come in any column can, with some reservations, be estimated from those in the last line in columns to the right.

A further series of tables gives adjusted rates of lapse and of new business by each of the rating factors.

The number of lapses is expressed as a proportion of the renewals to which the lapses can be considered to relate. The number of renewals counted for this purpose is based on the latest available table showing the pattern of delay, the assumption being made that the delays are the same for all rating groups. It has been found useful to relate the new business to the renewals in a similar way, partly for comparison with the lapse rates and partly as a seasonal adjustment since the seasonal variation in new business is likely to be broadly similar to the variation in renewals.

Table 8 is an example of one of these tables. The rates are shown per 1000.

TABLE 8

"Processing month"	Factor level					All
	1	2	3	4	. . . .	
10/70	75	67	63	59	. . . .	65
9/70	58	50	50	48	. . . .	50
8/70	93	82	86	88	. . . .	87
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
10/69	92	79	80	73	. . . .	79

Because of the variations in the flow of new business and lapses, both as regards the actual numbers notified and the delays between notification and recording on the file, it has been found useful to produce a second series of tables similar to Table 8 but expressing the figures as a percentage of the total for the line. Table 9, which corresponds to Table 8, is an example of such a table.

TABLE 9

"Processing month"	Factor level					All
	1	2	3	4	. . . .	
10/70	115	103	97	91	. . . .	65
9/70	116	100	100	96	. . . .	50
8/70	107	94	99	101	. . . .	87
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
.	. . .	. . .	. . .	. . .	. . . .	. . .
10/69	116	100	101	92	. . . .	79

5. *General comments*

(i) Much of the work of devising a statistical system for a motor insurance portfolio needs to be concerned with the accuracy of the data. This requires careful definition of terms and the provision of frequent checks to ensure that the definitions are being consistently applied. Particular care is needed to avoid bias in the collection of data from a constantly changing file.

(ii) In the conduct of motor insurance, so many factors are relevant and so many questions may need to be answered that it is essential to devise a system which will present the important features of the results in a concise and easily understood form. The concept of

comparing the actual results with those expected on a standard basis is extremely helpful.

(iii) It is not essential that the model used to represent the standard basis should match the observed experience extremely closely. Any one of several models might be adequate as a basis for standardisation. If the observed experience diverges from that assumed in the model, for any of the factors on which the model is based, this will become apparent from the comparisons between actual and expected, and the model can be refined if necessary.

(iv) Whilst the separate study of claim frequencies and average amounts of claim is essential, we must try to ensure that when the results are combined the two sets of figures are comparable. This is difficult since there will inevitably be much uncertainty regarding the eventual average claim costs for the claims on which the latest frequencies are based.

(v) We have confined our attention to giving examples of tables which are being used, and have not referred to tests of significance of the results. Such tests may sometimes be useful, but they need to be applied with care since in the circumstances found in motor insurance the conditions of randomness on which the tests are based may not be satisfied. It appears that over the last few years there have been considerable changes of a non-random nature in the shapes of the claim distributions in the United Kingdom.