AN ANALYSIS OF CLAIM EXPERIENCE IN PRIVATE HEALTH INSURANCE TO ESTABLISH A RELATION BETWEEN DEDUCTIBLES AND PREMIUM REBATES

G. W. DE WIT and W. M. KASTELIJN

Many studies concerning the frequency of claims by size in health insurance are not generally known *). A possible explanation of this circumstance could be the fact that in most countries this line of insurance has been brought entirely within the ambit of social insurance. Also from the side of the social insurance very few investigations have been published **).

In this paper we will analyse the claim experience (relating to the calendar year 1972) of a private health insurance business. The data have been subdivided according to three levels of coverage (in increasing order of benefits these are: class III, class IIb and class IIa). The claim payments comprise nursing costs, auxiliary costs and the fees for specialist treatment in and out of the hospital.

We will use the following notations:

s_i: claim amount paid for the insured i in one year,
n: number of claims,
v: number of risks (policies insured).

In many instances the premium is simply determined as a level premium. In other words each insured pays the premium p, calculated as follows:

$$p = \frac{\Sigma\, s_i}{v}.$$

---

*) Notably concerning West Germany and Switzerland we refer to some recent articles published in the Blätter der Deutschen Gesellschaft für Versicherungsmathematik and in the Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.

**) See e.g. the analysis made in Finland (Research Institure fot Social Security).

https://doi.org/10.1017/S0515036100011569 Published online by Cambridge University Press

Actually we make the assumption that the claims are normally distributed, the parameters of which can be estimated as follows:

$$\mu = \frac{1}{n} \Sigma s_i$$

$$\sigma^2 = \frac{1}{n} - 1 \Sigma (s_i - \mu)^2$$

which permits the calculation of the premium according to:

$$p = \frac{n}{v} \mu.$$

Plotting the empirical claim distribution on log-normal probability paper suggests however that (like many other distributions in the field of insurance) the log-normal assumption gives a better fit than the normal distribution. Denoting its parameters by μ and σ its mean and variance are:

$$\alpha = \exp \{\mu + \tfrac{1}{2}\sigma^2\} \qquad (1)$$

$$\beta^2 = \exp \{\sigma^2 - 1\} \exp \{2\mu + \sigma^2\}. \qquad (2)$$

The premium can again be found as:

$$p = \frac{n}{v} \alpha. \qquad (3)$$

The parameters of the log-normal distribution can be estimated by means of various methods (Aitchison and Brown: The log-normal distribution). For our purposes we used logarithmic probability paper (absciss: logarithmic; ordinate: probability). This approach has the advantage that besides estimation of the parameters we can test whether the data look like a log-normal distribution.

For our estimations and tests of log-normality we started from the following data:

TABLE I

Claim amount s	Class III Number of claims $\leqslant s$	Class III % claims $\leqslant s$	Class IIb Number of claims $\leqslant s$	Class IIb % claims $\leqslant s$	Class IIa Number of claims $\leqslant s$	Class IIa % claims $\leqslant s$
100	801	19.5	579	18.1	244	18.2
200	1434	34.9	1037	32.5	424	31.6
300	1806	44.0	1336	41.9	527	39.3
400	2113	51.4	1564	49.0	625	46.6
500	2367	57.6	1756	55.0	698	52.0
600	2557	62.2	1899	59.5	754	56.2
700	2675	65.1	2007	62.9	795	59.2
800	2789	67.9	2093	65.6	831	61.9
900	2880	70.1	2162	67.7	866	64.5
1000	2969	72.3	2219	69.5	895	66.7
1500	3282	79.9	2440	76.4	994	74.1
2000	3479	84.7	2589	81.1	1068	79.6
2500	3623	88.2	2686	84.1	1097	81.7
3000	3734	90.9	2768	86.7	1128	84.1
4000	3873	94.3	2882	90.3	1184	88.2
5000	3945	96.0	2968	93.0	1219	90.8
7000	4014	97.7	3069	96.1	1270	94.6
10100	4055	98.7	3135	98.2	1303	97.1
20400	4097	99.7	3183	99.7	1341	99.9
∞	4108	100	3192	100	1342	100

The percentages of claims $\leq s$ are plotted on log-normal probability paper. If the sample points ly approximately on a straight line it is reasonable to assume log-normality. This appears to be the case for each of the three classes (figures 1a, 1b, 1c).

From the graph we can calculate μ and σ. The points s_{50} (the median) and s_{95} can be read from the graph. The two parameters are then determined as follows:

$$\mu = \log s_{50}$$

$$\text{and } \sigma = \log \frac{s_{95}}{s_{50}} \,/\, 1.645.$$

For class III we then find:

$$\mu = \log 400 = 5.99$$

$$\sigma = \log \frac{4210}{400} \,/\, 1.645 = 1.431$$

carrying through the calculations for all possibilities results in the following table:

TABEL 2

	Basic data			Normal assumption		Log-normal assumption							
1	2	3	4	5	6	7	8	9	10	11	12	13	14
Class	Total claims in millions	Number of claims (n)	Number of risks (v)	μ $^2/_3$	p $^2/_4$	50% point	95% point	μ log 7	σ $\frac{\log {}^8/_7}{1.645}$	α	β	$\frac{n}{v}$ $^3/_4$	p 13 × 11
III	4.67	4108	9403	1138	497	400	4210	5.99	1.431	1113	2892	.437	489
IIb	4.48	3192	6264	1403	715	453	5573	6 12	1.526	1451	4413	.510	739
IIa	2.16	1342	2375	1610	910	477	6412	6.17	1.580	1661	5539	.565	939

https://doi.org/10.1017/S0515036100011569 Published online by Cambridge University Press

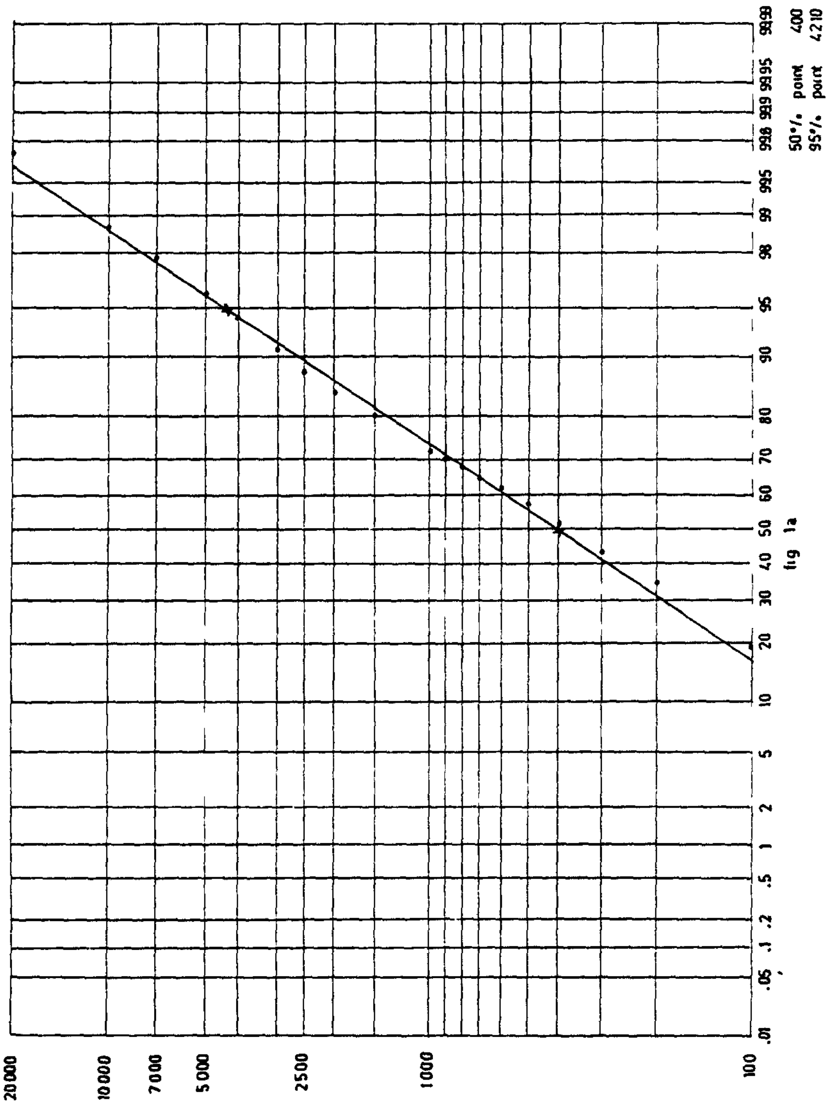
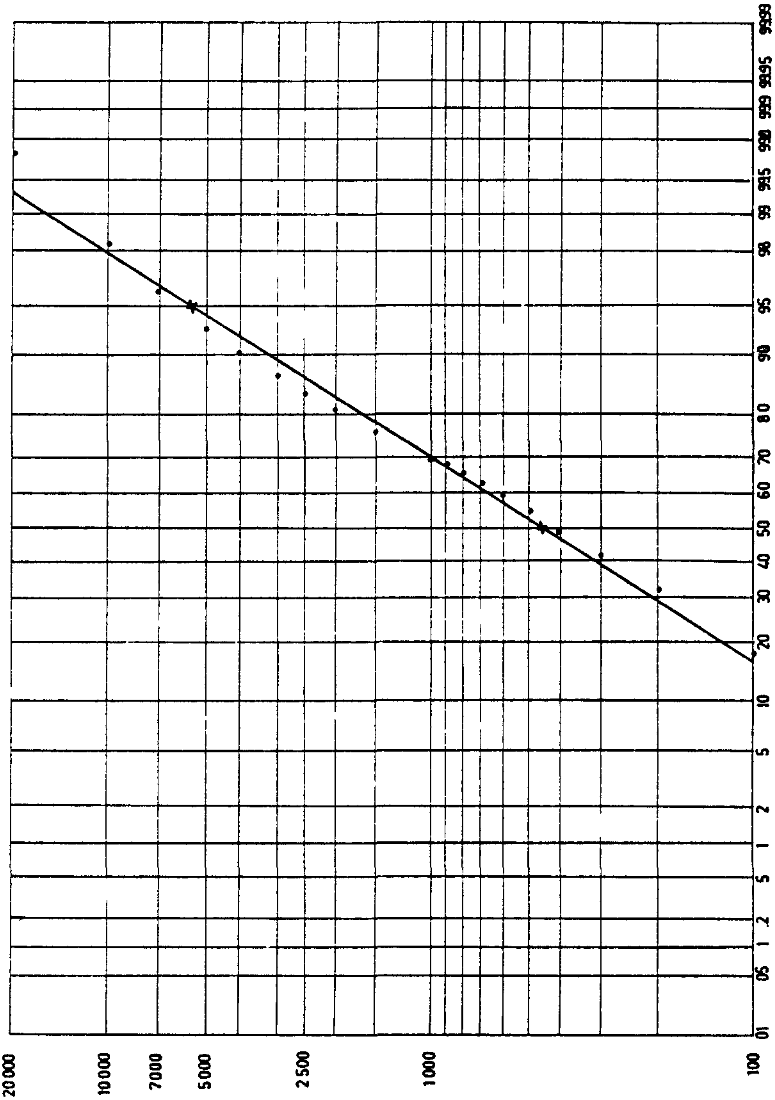


fig 1a

https://doi.org/10.1017/S0515036100011569 Published online by Cambridge University Press



50% point 453
95% point 5573

fig 1 b

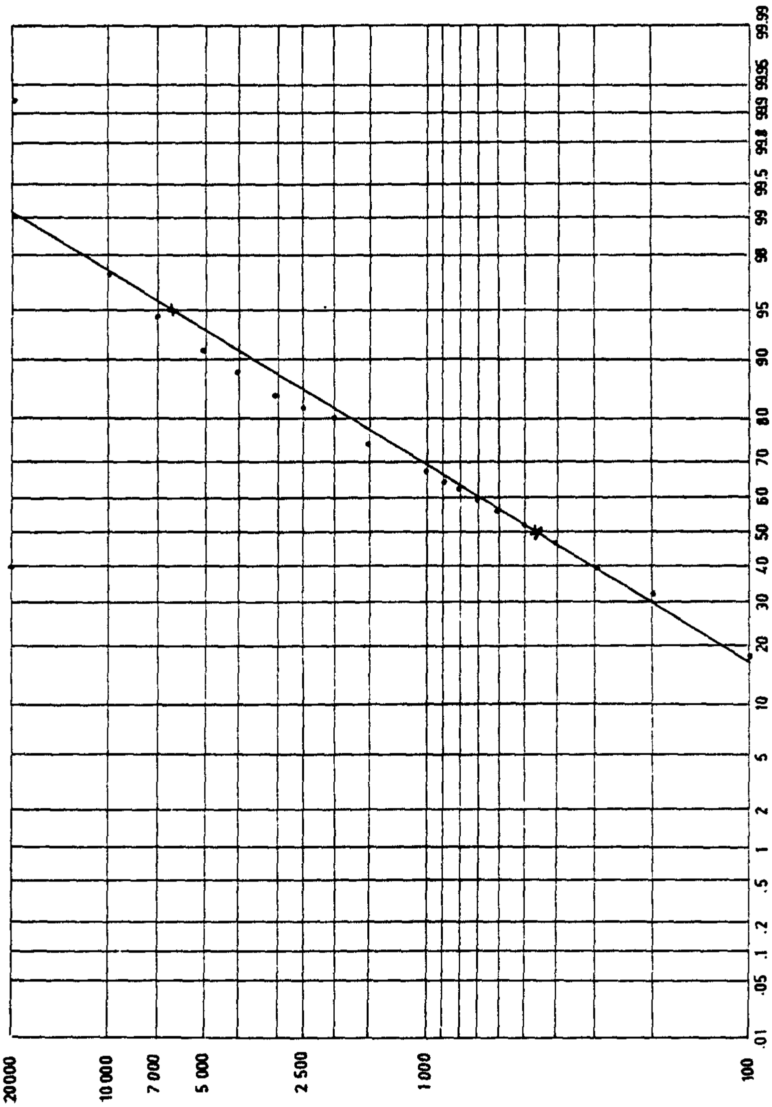


fig. 1c

50% point: 477
95% point: 6412

https://doi.org/10.1017/S0515036100011569 Published online by Cambridge University Press

A "disadvantage" of this method is that the sum of the premiums does not equal the sum of the claims. It seems however questionable whether this is really a disadvantage. If we apply the present premium estimation method to a later year it will give a better quarantee for the adequacy of the rating than the requirement of strict equivalence.

In the foregoing we have considered how the level premium can be derived from the empirical claim distribution. We can also reverse this question: in what manner does this claim distribution depend on the premium.

Knowing the premium is howevei not sufficient to find the claim distribution, because for that purpose we also have to know the variance and $(n/v.)$ It turns out, however, that a relation exists between the quantities p and β on the one hand and between (n/v) and p on the other hand. If we know this relation we are in a position to find β and (n/v) directly from p and α by means of (3).

Figures 2a and 2b show that both relations are linear:

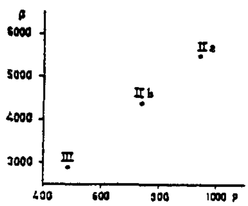


Fig. 2a

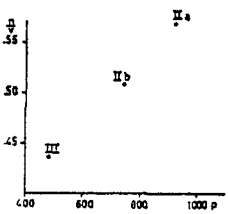


Fig. 2b

The linear relations are:

$$\beta = 5.85\, p + 61.1 \tag{4}$$

$$(n/v) = .000283\, p + .30 \tag{5}$$

(1), (2) and (3) can be written as:

$$\alpha = \frac{p}{\left(\dfrac{n}{v}\right)}$$

$$\mu = \tfrac{1}{2} \log \left(\frac{\alpha^4}{\beta^2 + \alpha^2}\right)$$

$$\sigma^2 = \log \left(\frac{\beta^2 + \alpha^2}{\alpha^2}\right)$$

https://doi.org/10.1017/S0515036100011569 Published online by Cambridge University Press

The last three formulae allow us to calculate β, (n/v), μ and σ^2 successively for given p. We thus have found the distribution we require. The claim distribution as a function of the premium also permits the calculation of the premium rebate for a given deductible. Let $f(s; p)$ be the claim distribution and $\varphi(R, p)$ the rebate factor applicable to the premium as a function of the deductible R and the premium. Then the following relation exists:

$$\varphi(R, p) = \frac{\int_0^R s\, f(s; p)\, ds + R \int_R^\infty f(s; p)\, ds}{\int_0^\infty s\, f(s; p)\, ds.}$$

Actual calculations for various p and R result in the following table for $\varphi(R, p)$:

TABLE 3

R \ p	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400
500	.478	.388	.332	.293	.265	.243	226	.212	.201	.191	.183	.175	.169
1000	.652	.557	.491	.442	.405	.376	.352	.332	.316	.301	.289	.278	.268
1500	.745	.656	.589	.538	.498	.465	.438	.415	.396	.379	.364	.351	.340
2000	.803	.721	.657	.606	.565	.531	.503	.478	.457	.439	.422	.408	.395
2500	.842	.768	.707	.658	.617	.582	.553	.528	.506	.486	.469	.454	.440
3000	.870	.803	.746	.698	.658	.623	.594	.568	.546	.526	.508	.492	.477
3500	.891	.830	.776	.730	.691	.657	.628	.602	.579	.559	.541	.524	.510
4500	.920	.869	.821	.780	.743	.710	.681	.656	.633	.613	.594	.577	.562
5500	.939	.895	.853	.815	.781	.750	.722	.697	.675	.654	.636	.619	.603
10000	.975	.951	.925	.899	.873	.849	.827	.806	.786	.768	.750	.734	.720
15000	.988	.973	.956	.938	.919	.900	.882	.864	.847	.831	.816	.802	.788
20000	.993	.983	.971	.957	.943	.928	.912	.898	.883	.869	.856	.843	.831
30000	.997	.992	.985	.976	.967	.956	.945	.934	.923	.912	.901	.891	.881

Up till now we have assumed throughout that both the level premium and the claim distribution are independent of the age of the insured. This assumption is actually not justified. Usually the claim amount is age dependent as follows:

$$s_x = c_0 \cdot c_1^x.$$

https://doi.org/10.1017/S0515036100011569 Published online by Cambridge University Press

Here c_0 and c_1 are constants. Estimation of these constants from the data available for 1972 produced the following results:

Class	Males		Females	
	c_0	c_1	c_0	c_1
III	62.0	1 034	165.5	1.021
IIb + IIa	54.4	1.045	230.9	1.021

The constant c_1 is as a matter of fact time dependent with respect to the level of medical care and consequently will change only very slowly with time. The constant c_0 on the other hand reflects the price level of medical care of which it is directly dependent.

The calculation of s_x has been carried out however assuming normality. With the log-normal assumption the age dependence of α, β and (n/v) will have to be studied. The extent of the claim data available was not, however, of sufficient size to justify a subdivision by age. Hence, the age dependence of β and (n/v) could not be examined.