

A STOCHASTIC MODEL TO DETERMINE IBNR-
RESERVES

HANS WASZINK
MICHIEL VAN DER WARDT

1998 GENERAL INSURANCE CONVENTION
AND
ASTIN COLLOQUIUM

GLASGOW, SCOTLAND: 7-10 OCTOBER 1998

A Stochastic Model to determine IBNR-reserves

*Hans Waszink
Michiel van der Wardt*

*William M. Mercer Ten Pas B.V.
Gebouw Metropolitan
Buitenveldertselaan 7
PO BOX 75329
1070 AH Amsterdam
The Netherlands
tel: ++31-205419750
fax: ++31-205419790*

Summary

This paper describes a stochastic model to determine IBNR-reserves based on a Compound Poisson distribution. The number of insureds per accident year and the mean and second moment of payments by accident, reporting and development year are used as input. The claim frequency is estimated by accident and reporting year and the equality of the claim frequencies is tested using a χ^2 goodness-of-fit test. Mean and second moment of the total claim amount are estimated as the sum of payments in development years starting with the reporting year. After that, the model is applied to a fictitious dataset and the arising difficulties are discussed. It is shown that without testing the equality of claim frequencies, the calculated reserve may lead to an underestimation of future liabilities.

Keywords: claim frequency, Poisson distribution, χ^2 goodness-of-fit test.

1 Introduction

In this paper, a stochastic model is described to estimate IBNR-reserves.¹ It is assumed in the model that future claim numbers are Poisson distributed, where the Poisson parameter is proportional to the number of insureds. To estimate the individual claim amounts, the average and second moment of annual payments per claim are estimated from past observations for different accident, reporting years and development years.

In order to establish the validity of past experiences to predict future claims, it can be tested whether there have been significant changes in claim frequencies and amounts during the observed accident years.

Others have used similar ideas before in the construction of IBNR-models. Larsen, Monty and Clemensen (1996) model the number of future claims stochastically and also use the number of insureds as input. Bühlmann, Schnieper and Straub (1980) describe a model that uses a Compound Poisson distribution and the number of insureds. The goal of the model described here is to use all the information available by the insurance company as effectively as possible while maintaining fitness for use.

This paper is organised as follows: in section 2, the model structure is described. In section 3 and 4, the estimation methods for claim numbers and claim amounts are described respectively. In section 5, the model is applied to a fictitious dataset. Section 6 contains the conclusion.

¹ The authors would like to thank prof. dr. W. Schaafsma, University of Groningen, The Netherlands, department of Mathematics and Computer Science, for his advice in the construction of this model.

2 General model assumptions

In this section, the mathematical structure of the model is described. The time frame is defined in the first subsection, the stochastic model structure in the second.

2.1 Time frame

To determine the net present value of future liabilities at the end of a booking period, a distinction is made between accident, reporting and development periods. The accident period is the period in which the event leading to a claim has occurred, the reporting period refers to the period in which the claim is reported and the development periods refer to the periods in which payments, resulting from the claim, are made.

The duration of the periods can be chosen arbitrarily, as long as all periods have the same duration. In the following, the duration of all periods will be one year. Accident years are indicated by calendar year, while reporting and development years are related to accident years and indicated by the number of years passed since the beginning of the accident year. For example, reporting year j is the j^{th} calendar year after the accident year, so that reporting year 0 is the accident year. Claim numbers are observed separately for different accident and reporting years.

The following variables are defined:

n_{ij} = number of claims incurred in year i and reported in year j , the j^{th} calendar year after year i ;

s_{ijk} = net present value at the moment of reporting of the k^{th} claim originating in accident year i , reported in reporting year j , $k=1,2,\dots,n_j$;

$v = 1 / (1 + \text{annual interest rate})$;

i^* = the financial year at the end of which the reserve is determined.

It is assumed that all claims are reported exactly in the middle of the year. Total liabilities, denoted by l , with respect to incurred but unreported claims at the end of year i^* equal:

$$l = \sum_{\substack{i+j \leq i^* \\ i \leq i^*}} \sum_{k=1}^{n_j} s_{ijk} v^{i+j-i^*-1/2}$$

2.2 Description of the stochastic model

For the determination of the IBNR-reserve at the end of year i^* , a stochastic variable N_{ij} is defined with outcome n_{ij} . It is assumed, that N_{ij} is Poisson distributed with parameter λ_{ij} and that λ_{ij} is proportional to the number of insureds x_i in accident year i and an expected claim frequency φ_{ij} , depending on the accident year and the reporting year. The claim frequency F_{ij} with outcome f_{ij} is defined as:

$$F_{ij} = \frac{N_{ij}}{x_i}, \text{ so that } \varphi_{ij} = E F_{ij} = \frac{E N_{ij}}{x_i} \text{ and } E N_{ij} = \lambda_{ij} = x_i \varphi_{ij}.$$

The assumption of Poisson distributed claim numbers is made because the Poisson distribution is an adequate distribution to model stochastic events occurring with a small probability.

Furthermore, stochastic variables S_{ijk} with outcomes s_{ijk} are defined. L_{ij} represents the net present value of all future liabilities to be taken into account at the end of year i^* resulting from claims incurred in year i , with $i \leq i^*$, and reported in reporting year j . For the IBNR-reserve, only claims that have incurred in year i , with $i \leq i^*$, and will be reported in year j , with $i+j > i^*$, have to be taken into account. L_{ij} is defined as:

$$L_{ij} = \sum_{k=1}^{N_{ij}} S_{ijk} v^{i+j-i^*-1/2}.$$

The net present value at the end of year i^* of the total of future liabilities is defined by L and equals:

$$L = \sum_{\substack{i+j > i^* \\ i \leq i^*}} L_{ij}.$$

It is assumed that all the S_{ijk} and N_{ij} are mutually independent. Mean and variance of L_{ij} equal:

$$E[L_{ij}] = E[N_{ij}]E[S_{ijk}] v^{i+j-i^*-1/2},$$

$$\text{VAR}[L_{ij}] = E[N_{ij}]E[S_{ijk}^2] (v^{i+j-i^*-1/2})^2.$$

Mean and variance of L equal:

$$E[L] = \sum_{\substack{i+j \geq t^* \\ i \leq t^*}} E[L_y];$$

$$VAR[L] = \sum_{\substack{i+j \geq t^* \\ i \leq t^*}} VAR[L_y].$$

In the next two sections, estimation methods for the moments of N and S are described. Based on the model structure described in this section, estimations for the variance, expected value and all higher moments of future liabilities can be derived.

3 Estimation of claim frequencies and the expected number of claims

In this section, methods to estimate the claim frequencies are described. This is done using an example in which an IBNR reserve has to be computed at the end of 1997 for a product that was first given coverage in 1994. The frequency of claims is estimated by accident and reporting year as described in the previous section.

Numbers of insureds and claims in the example are represented in the table below:

Table 1: numbers of insureds and claims by accident and reporting year

x_i = number of insureds in accident year i .

n_{ij} = number of claims incurred in year i , reported in year j .

		Reporting year										
Accident year	Number of insureds	0	1	2	3	4	5	6	7	8	9	10
		Number of claims:										
1994	x_{93}	$n_{93,0}$	$n_{93,1}$	$n_{93,2}$	$n_{93,3}$							
1995	x_{94}	$n_{94,0}$	$n_{94,1}$	$n_{94,2}$								
1996	x_{95}	$n_{95,0}$	$n_{95,1}$									
1997	x_{96}	$n_{96,0}$										

It is assumed that no claims will be reported more than ten years after the accident year. This is a more or less arbitrary assumption, which can be adapted depending on the expected claim behaviour.

The claim frequencies f_{ij} can be observed when $i \leq i^*$ and $i+j \leq i^*$. For $i \leq i^*$ and $i+j > i^*$, the f_{ij} will be predicted under the assumption that $\varphi_{ij} = \varphi_j$, where φ_j denotes the expected claim frequency in reporting year j , equal for all accident years i . The hypothesis that $\varphi_{ij} = \varphi_j$ can be tested for given j and $i+j \leq i^*$. This hypothesis can also be tested for all j together. See appendix I for the description of the tests. If the claim frequency changes considerably through the years, the hypothesis that $\varphi_{ij} = \varphi_j$ will possibly be rejected. In this case, one can remove the earliest accident year from the test and perform the test again. This process can be repeated until the hypothesis will not be rejected any more. Also if in one particular accident year, other than the earliest observable one, claim frequencies differ greatly from the other accident years for a known reason, one could remove this accident year from the test. One should not, however, remove an accident year from the test only because it differs greatly from all the other years, since by doing so the level of the test would be influenced.

Given $\varphi_i = \varphi_j$ for all i and given j , the UMVU-estimator of φ_j is given by :

$$\hat{\varphi}_j = \frac{\sum_i n_{ij}}{\sum_i x_i}.$$

See appendix II for a proof.

The expected numbers of claims for future reporting years are predicted by

$\hat{\lambda}_y = \hat{\varphi}_j x_i$, with $i \leq i^*$ and $i+j > i^*$. In the example shown in table 1, no observations are available for $j \geq 4$. Additional assumptions are therefore necessary in this case to estimate φ_j . These assumptions cannot be tested due to the absence of data. Instead, they will have to be based on more general experience of the actuary regarding late claims.

4 Estimation of the distribution of liabilities per claim

In the first subsection of this section, the estimation of the individual claim amounts is discussed. In the second subsection, some comments are made with respect to testing the equality of claim distributions for different accident years.

4.1 Estimation of claim amounts

To estimate the net present value of the liabilities resulting from an individual claim, registrations of benefits paid and additional costs are available. Based on those registrations, expected value, variance and possibly higher moments of the total costs per claim can be estimated. The problem arises that at the instant the reserve has to be determined, many claims have not been fully settled yet. Therefore, annual payments per claim are estimated separately as a function of accident, reporting and development year. The claim amount is estimated as the sum of the estimated payments in different development years.

B_{ijk} is defined as the total of payments in the l^{th} development year after the accident year resulting from the k^{th} claim reported in the j^{th} reporting year, incurred in accident year i . In order to take inflation into account, payments done in the past can be adjusted to the level at the end of year i^* . Future inflation can be taken into account by assuming an inflation rate for the future and adjusting all payments B_{ijk} for which $i+l > i^*$ accordingly. Note that in case no payments are made in development year l , B_{ijk} equals zero, so that B_{ijk} is defined for all $l \geq 0$.

All payments are assumed to be made exactly in the middle of the year, so that the delay between reporting and payment is exactly $l-j$ years. Therefore,

$$S_{ijk} = \sum_{l \geq j} B_{ijk} v^{l-j}.$$

The first and second moment of the S_{ijk} can be written in terms of the B_{ijk} as:

$$\begin{aligned} E[S_{ijk}] &= \sum_{l \geq j} E[B_{ijk}] v^{l-j} \text{ and} \\ E[S_{ijk}^2] &= \sum_{l \geq j} E[B_{ijk}^2] v^{2(l-j)} + E\left[\sum_{\substack{l \geq j \\ m \geq j \\ m \neq l}} B_{ijk} B_{ijk} \right] v^{2(l-j)}. \end{aligned}$$

To estimate $E[S_{jk}]$ and $E[S_{jk}^2]$ it is assumed that:

1. $E[B_{jkl}]$ and $E[B_{jkl}^2]$ are equal for all k , so that they can be written as $E[B_{j,l}]$ and $E[B_{j,l}^2]$ instead;
2. B_{jkl} and B_{jkm} are uncorrelated for $l \neq m$, so that $E[B_{jkl}B_{jkm}] = E[B_{jkl}]E[B_{jkm}]$.

The first assumption implies that all claims incurred in the same accident year and reported in the same reporting year give rise to the same claim amount distribution. Assuming this, any seasonal fluctuations that may be present in the claim pattern are disregarded.

The second assumption is based on the assumption that the magnitude of a payment in one development year possibly provides an indication of the variance of payments in other development years but not of the expected value.

As a result, $E[S_{jk}^2]$ can be rewritten as:

$$E[S_{jk}^2] = \sum_{l \geq j} E[B_{j,l}^2] v^{2(l-j)} + \sum_{\substack{l \geq j \\ m \geq j \\ m \neq l}} E[B_{j,l}] E[B_{j,m}] v^{2(l-j)},$$

so that unbiased estimates for $E[S_{jk}]$ and $E[S_{jk}^2]$ can be derived from unbiased estimates for $E[B_{j,l}]$ and $E[B_{j,l}^2]$.

For future development years, $E B_{j,l}$ and $E B_{j,l}^2$ are estimated by $\hat{E} B_{j,l}$ and $\hat{E} B_{j,l}^2$:

$$\hat{E} B_{j,l} = \frac{1}{T} \sum_{\substack{m+j \leq l \\ k}} b_{mjkl} ;$$

$$\hat{E} B_{j,l}^2 = \frac{1}{T} \sum_{\substack{m+j \leq l \\ k}} b_{mjkl}^2 ,$$

where the b_{mjkl} denote the outcomes of the B_{mjkl} observed in the past and T is the number of observations b_{mjkl} over which the summation is taken.

4.2 Testing equality of annual payments for different accident years

It would be interesting to test the hypothesis that annual expected payments or claim amounts after correction for inflation are equal for all observed accident years. For given j and l , a test statistic can be derived for certain distribution functions of the annual payments. It will not be possible, however, to derive a test for all j and l simultaneously at any given level without making assumptions about the dependence structure of the B_{ijt} . Therefore, testing of the equality of moments of the claim amounts and annual payments is left for further research.

5 Calculation of the IBNR-reserve for an existing portfolio

In this section, the theory described in the previous sections is applied to an example with fictitious data. In the first subsection, claim frequencies are estimated. In the second subsection, the individual claim amounts are estimated and in the third subsection, the effect of changes in the claim frequencies is investigated.

5.1 Estimation of the future number of claims

Future claim frequencies are estimated at the end of 1997 for a portfolio, that has come into existence in 1988. The following numbers of claims and insureds for accident years 1988 through 1997 are given:

Table 2: Number of insureds and claims by accident and reporting year

		Reporting year									
		0	1	2	3	4	5	6	7	8	9
Accident year	Number of insureds	Number of claims									
1988	3,520	214	189	46	28	21	10	9	2	3	0
1989	3,740	198	191	49	38	29	18	10	5	4	
1990	4,070	239	183	57	39	28	16	16	3		
1991	4,280	223	201	51	36	30	21	12			
1992	4,160	202	232	55	28	38	19				
1993	4,340	207	221	59	39	29					
1994	4,020	206	205	53	37						
1995	4,010	212	184	52							
1996	4,250	252	190								
1997	4,880	263									

For each reporting year separately and for all reporting years together, the hypothesis is tested that expected claim frequencies depend only on the reporting year and not on the accident year. See appendix I for the calculation of the test-values. The results are as follows:

Table 3: Test values and critical values by reporting year and for all reporting years together

Reporting year	0	1	2	3	4	5	6	7	8	total
Test value	13.40	10.39	0.80	3.56	3.27	2.55	1.59	1.37	0.09	37.02
Critical value ² for										
$\alpha = 10\%$	14.68	13.36	12.02	10.64	9.24	7.78	6.25	4.61	2.71	57.5
$\alpha = 5\%$	16.92	15.50	14.06	12.59	11.07	9.49	7.81	5.99	3.84	61.7
$\alpha = 1\%$	21.67	20.09	18.48	16.81	15.09	13.28	11.34	9.21	6.63	70.0

The hypothesis is not rejected for any reporting year nor for all reporting years together (total) at any reasonable test level. It can be discussed whether reporting year 0, the accident year, should be involved in the test, since the IBNR-reserve at the end of any financial year will not provide for claims reported during the accident year. On the other hand, it is logical to assume that claim numbers in different reporting years resulting from the same accident year are always interrelated. Therefore, reporting year 0 is also involved in the test and the test results are shown in the 'total'-column.

The predictions of the claim frequencies in reporting years 1 through 8 are calculated as the sum of claims in the reporting year divided by the number of insureds in the corresponding accident years. In reporting year 9 and 10, respectively 2/3 and 1/3 of the predicted claim frequency in reporting year 8 are taken as prediction values.

The following predictions for claim frequencies are found:

Table 4: Predicted claim frequencies by reporting year

Reporting year	Claim frequency
0	0.0537
1	0.0494
2	0.0131
3	0.00871
4	0.00726
5	0.00425
6	0.00301
7	0.000883
8	0.000964
9	0.000643
10	0.000321

Table 2 can now be extended with expected numbers of future claims:

² The critical values are based on χ^2 -distributions with df equal to 9 minus the reporting year. For the 'total' column, df equals 45, the sum of the df for the separate reporting years.

Table 5: Numbers of insureds and (expected) claim numbers by accident and reporting year

Table 3: Numbers of insureds and (expected) claim numbers by accident year													
Accident year	Number of insureds	Reporting year											
		0	1	2	3	4	5	6	7	8	9	10	
Number of claims													
1988	3,520	214	189	46	28	21	10	9	2	3	0	1.13	
1989	3,740	198	191	49	38	29	18	10	5	4	2.40	1.20	
1990	4,070	239	183	57	39	28	16	16	3	3.92	2.62	1.31	
1991	4,280	223	201	51	36	30	21	12	3.78	4.13	2.75	1.38	
1992	4,160	202	232	55	28	38	19	12.53	3.67	4.01	2.67	1.34	
1993	4,340	207	221	59	39	29	18.44	13.07	3.83	4.18	2.79	1.39	
1994	4,020	206	205	53	37	29.18	17.08	12.10	3.55	3.88	2.58	1.29	
1995	4,010	212	184	52	34.93	29.11	17.04	12.07	3.54	3.87	2.58	1.29	
1996	4,250	252	190	55.80	37.02	30.85	18.06	12.80	3.75	4.10	2.73	1.37	
1997	4,880	263	240.85	64.07	42.50	35.42	20.73	14.69	4.31	4.71	3.14	1.57	

5.2 Estimation of total costs per claim

Total costs per claim are estimated by summing over the estimated annual payments in corresponding development years and dividing by the number of claims in the reporting year. For the 10th development year, for which no observations are available, an expected payment of 50% of the average payment in the 9th development year and equal reporting year is assumed. For later development years, it is assumed that no further payments will take place. The following estimates for the mean and second moment of total liabilities per claim are found:

Table 6: Estimated mean and second moment of claims by reporting year

Reporting year	Mean	Second moment
1	561	320,539
2	540	296,777
3	552	308,617
4	558	316,086
5	591	355,302
6	608	373,774
7	606	376,518
8	802	651,267

Because of the small number of claims in reporting years 6, 7 and 8, one average and one variance of total costs per claim are computed for those years together. These are used to predict total costs per claim in reporting year 6,7,8,9, and 10. The interest and inflation rate are assumed to be zero. The results are as follows:

Table 7: Expected value and standard deviation of future liabilities by reporting year

Reporting year	Expected value	Standard deviation
1	135,158	8,786
2	64,744	5,965
3	63,149	5,943
4	69,543	6,275
5	56,706	6,012
6	47,959	5,529
7	16,404	3,234
8	20,356	3,602
9	15,063	3,099
10	8,234	2,291
total	497,316	17,102

5.3 Effects of changes in the claim frequency

In the next example, claim numbers in 1993 through 1997 are increased by 20%, while the moments of the costs per claim remain unchanged. This adjustment is made to show how the method deals with claim frequencies that differ significantly through the years. If no test is performed on the frequencies and the method shown above is applied, the following results are found.

Table 8: *Expected value and standard deviation based on all observations of claims*

Reporting year	Expected value	Standard deviation
1	147,124	9,167
2	69,500	6,180
3	66,758	6,110
4	71,530	6,364
5	56,706	6,012
6	47,959	5,529
7	16,404	3,234
8	20,356	3,602
9	15,063	3,099
10	8,234	2,291
total	519,633	17,466

Given the fact that claim frequencies are increased by 20% in the last 5 accident years, the increase of the expected value with only 4.5% compared to table 7 is an underestimation of the real increase of expected future liabilities.

The test of equality of claim frequencies by accident year gives the following results:

Table 9: *Test results for the claim frequencies*

Reporting year	0	1	2	3	4	5	6	7	8	total
Test value	28.52	18.89	4.71	6.33	3.04	2.55	1.59	1.37	0.09	67.09
Critical value ³ for										
$\alpha = 10\%$:	14.68	13.36	12.02	10.64	9.24	7.78	6.25	4.61	2.71	57.5
$\alpha = 5\%$:	16.92	15.5	14.06	12.59	11.07	9.49	7.81	5.99	3.84	61.7
$\alpha = 1\%$:	21.67	20.09	18.48	16.81	15.09	13.28	11.34	9.21	6.63	70.0

The test rejects the hypothesis for reporting year 0 at 1% and for reporting year 1 and for all reporting years together (total) at level 5%.

³ The critical values are based on χ^2 -distributions with df equal to 9 minus the reporting year. For the 'total' column, df equals 45, the sum of the df for the separate reporting years.

The earliest accident years will now be removed from the test one by one until the hypothesis of equal expected claim frequencies in the remaining accident years is not rejected any more at level 5%. This yields the following results:

Table 10: The effect of removing early accident years from the test

Last removed accident year	Test value in reporting year 0	Critical value for $\alpha=5\%$ in reporting year 0	Test value in reporting year 1	Critical value for $\alpha=5\%$ in reporting year 1	Critical value for all reporting years together	Test value for all reporting years together	Critical value for $\alpha=5\%$ for all reporting years together
1988	34.51	15.55	23.21	14.06	61.39	51.00	
1989	30.71	14.06	22.15	12.59	56.02	41.34	
1990	30.24	12.59	13.05	11.07	46.54	32.67	
1991	22.85	11.07	3.93	9.49	30.36	25.00	
1992	5.79	9.49	2.84	7.81	10.48	18.31	

The test for all reporting years together does not reject once all accident years prior to 1993 have been removed. Therefore, the IBNR-reserve is computed with different claim frequencies for 1988 through 1992 and 1993 through 1997. For the total costs per claim, estimates based on the entire period from 1988 through 1997 are used.

The results for accident years 1993 through 1997 cannot be given without additional assumptions. Since the claim frequencies in the first five accident years differ significantly from those in the last five years, there are no unbiased estimators available for the latter. The additional assumptions that will have to be made cannot be tested due to a lack of observations. They will have to be based on more general assumptions about late claims.

One such assumption is that all the claim frequencies with respect to accident years 1993 through 1997 will increase by a constant factor compared to accident years 1988 through 1992. This factor can be estimated by comparing average claim frequencies in 1988 through 1992 with those in 1993 through 1997. In the following, it is assumed that claim numbers in the last five accident years will respectively be 20% and 50% higher than in the first five accident years.

The results for this approach are given in the two tables below.

Table 11: Expected value and standard deviation of total future liabilities with 20% higher claim numbers for the last five accident years

<i>Reporting year</i>	<i>Expected value</i>	<i>Standard deviation</i>
1	158,018	9,500
2	77,732	6,536
3	78,056	6,607
4	75,059	6,519
5	68,047	6,586
6	55,996	5,974
7	18,760	3,458
8	22,929	3,823
9	16,779	3,270
10	9,092	2,407
total	580,467	18,473

Table 12: Expected value and standard deviation of total future liabilities with 50% higher claim numbers for the last five accident years

Reporting year	Expected value	Standard deviation
1	158,018	9,500
2	77,732	6,536
3	78,056	6,607
4	75,059	6,519
5	85,059	7,363
6	68,051	6,586
7	22,293	3,770
8	26,790	4,132
9	19,352	3,512
10	10,378	2,572
total	620,789	19,156

As one can see, the rejection of the hypothesis of equal claim frequencies for all accident years leads to a considerably higher reserve. By the choice of test level, the actuary's belief in the hypothesis has a direct effect on the reserve that will be determined.

6 Conclusion

This model provides an estimation of expected value, variance and possibly higher moments of the total of future liabilities resulting from incurred but not yet reported claims. By way of statistical testing of additional assumptions, in particular shifts in claim frequencies, the model provides extra insight into the nature of risks undertaken. Statistical testing of the assumptions made leads to a prediction that is more directly related to the company's claim experience.

In the last section, it is shown that an increase of claim frequencies several years after the earliest observable accident year may lead to a serious underestimation of future liabilities in case all past accident years are used to determine the IBNR-reserve. By testing the equality of claim frequencies, this underestimation can be detected. To correct it, however, no explicit procedure can be prescribed.

The test results do not provide strict guidelines to determine an IBNR-reserve, but can only be seen as a way of providing extra insight to the nature of risks. The interpretation of the test results and the way estimations of future payments are made in the absence of past observations is still largely dependent on specific choices to be made by the actuary.

References

- Bühlmann, Schnieper, Straub (1980).
Claims Reserves in Casualty Insurance based on a Probabilistic Model;
Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker
p.21-46.
- Dudewicz, Mishra (1988).
Modern Mathematical Statistics, Wiley & Sons.
- Van Eeghen (1981).
Loss Reserving Methods.
- Larsen, Monty, Clemmensen (1996).
Dynamic loss reserving in a collective model.
XXVII ASTIN Colloquium Kopenhagen p.403.
- Mack (1994).
Which Stochastic Model is Underlying the Chain Ladder Method?
IME 15 (1994), 133-138.
- Mack (1996).
Stochastic Analysis of the Chain Ladder Method.
Handout at ASTIN Meeting Amsterdam December 1996.
- Mood, Graybill, Boes (1974).
Introduction to the Theory of Statistics, McGraw Hill International.

Appendix I

1 Testing the equality of claim frequencies for each reporting year separately

Define $H_0: \varphi_{ij} = \varphi_j$ for all i with given j .

$H_A: \exists k, l$ such that $\varphi_{ij} \neq \varphi_j$.

Given H_0 , we have: $N_{ij} \sim \text{Poisson}(x_i \varphi_j)$ and also $\sum_{i=1}^m N_{ij} \sim \text{Poisson}(\varphi_j \sum_{i=1}^m x_i)$

The conditional distribution of the N_{ij} given $\sum_{i=1}^m N_{ij}$ is multinomially distributed given H_0 :

$$P[N_{1j} = n_{1j}, \dots, N_{mj} = n_{mj}] = \begin{cases} \frac{(\sum_i N_{ij})!}{n_{1j}! \dots n_{mj}!} k_1^{n_{1j}} \dots k_m^{n_{mj}} & \text{when } n_{ij} \in \{0, 1, \dots, \sum_i N_{ij}\} \text{ and } \sum_i n_{ij} = \sum_i N_{ij} \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{with } k_1 = \frac{x_1}{\sum x_i}, k_2 = \frac{x_2}{\sum x_i}, \dots, k_m = \frac{x_m}{\sum x_i}.$$

This assumption can be tested with a goodness-of-fit test.

$$\text{Under the null-hypothesis we have: } E[N_{ij} | \sum_{k=1}^m N_{kj}] = x_i \frac{\sum_{k=1}^m N_{kj}}{\sum_{k=1}^m x_k}.$$

$$\text{Define } v_{ij} = x_i \frac{\sum_{k=1}^m N_{kj}}{\sum_{k=1}^m x_k}$$

The test value T_j , with:

$$T_j = \sum_{i=1}^m \frac{(n_{ij} - v_{ij})^2}{v_{ij}}.$$

is approximately χ^2 -distributed with $m-1$ degrees of freedom given H_0 .

2 Testing the equality of expected claim frequencies for all reporting years together

Null hypothesis and alternative are now as follows:

H_0 : for all j we have: $\varphi_{ij} = \varphi_j$ for all i with given j .

H_A : $\exists j$ for which $\exists k, l$ for which $\varphi_{kj} \neq \varphi_{lj}$.

If H_0 holds, then all T_j are χ^2 -distributed. Also the sum of the T_j is then χ^2 -distributed. The number of degrees of freedom of the sum, which is used as test value, now equals the sum of the degrees of freedom of the separate T_j .

Appendix II

Proof, that $\hat{\varphi}_j$ is the UMVU-estimator of φ_j .

It will be shown, that $\hat{\varphi}_j$ satisfies the next three conditions:

1. $\hat{\varphi}_j$ is an unbiased estimator of φ_j ;
2. $\hat{\varphi}_j$ is a sufficient statistic;
3. $\hat{\varphi}_j$ is a complete statistic.

According to the theorems of Rao-Blackwell and Lehmann-Scheffé, an estimator is the UMVU-estimator if these conditions are satisfied.

To verify the first condition, the expected value of $\hat{\varphi}_j$ is computed:

$$\begin{aligned}
 E[\hat{\varphi}_j] &= \\
 E\left[\frac{\sum_i N_{ij}}{\sum_i x_i}\right] &= \\
 \frac{\sum_i E[N_{ij}]}{\sum_i x_i} &= \\
 \frac{\sum_i x_i \varphi_j}{\sum_i x_i} &= \\
 \varphi_j .
 \end{aligned}$$

To show that $\hat{\varphi}_j$ is a sufficient statistic, it has to be shown that the distribution of the N_{ij} conditionally given $\sum_i N_{ij}$ does not depend on φ_j . Given $\sum_i N_{ij} = n$, the N_{ij} are multinomially distributed:

$$P[N_{1j} = n_{1j}, \dots, N_{mj} = n_{mj}] = \begin{cases} \frac{n!}{n_{1j}! \dots n_{mj}!} k_1^{n_{1j}} \dots k_m^{n_{mj}} & \text{when } n_{ij} = 0, 1, \dots, \sum_i N_{ij} \text{ and } \sum_i n_{ij} = \sum_i N_{ij} \\ 0 & \text{otherwise.} \end{cases}$$

with m the number of observable accident years and

$$k_1 = x_1 / \sum x_i, \quad k_2 = x_2 / \sum x_i, \dots, \quad k_m = x_m / \sum x_i.$$

φ_j is not a parameter in this distribution and so $\hat{\varphi}_j$ is a sufficient statistic.

$\hat{\varphi}_j$ is a complete statistic if the following condition is satisfied:

If for all values of φ_j and for all functions f we have that $E[f(\hat{\varphi}_j)] = 0$ then we also have that $f(x) = 0$ for all non-negative integers x .

This condition holds, which is now shown.

The condition that $E[f(\hat{\varphi}_j)] = 0$ holds for all functions f is equivalent with the condition that $E[g(\sum_i N_{ij})] = 0$ holds for all functions g . Since the N_{ij} are Poisson distributed, $\sum_i N_{ij}$ is also Poisson distributed. Define θ as the parameter of this distribution. $E[g(\sum_i N_{ij})]$ can be written as:

$$E[g(\sum_i N_{ij})] = \sum_{k=0}^{\infty} \frac{e^{-\theta} \theta^k}{k!} g(k).$$

The condition $\sum_{k=0}^{\infty} \frac{e^{-\theta} \theta^k}{k!} g(k) = 0$ is equivalent with

$$\sum_{k=0}^{\infty} \frac{\theta^k}{k!} g(k) = 0.$$

$\sum_{k=0}^{\infty} \frac{\theta^k}{k!} g(k)$ is a power series in θ with coefficients $\frac{g(k)}{k!}$. Two power series in θ have equal values in an interval if and only if all coefficients are equal. Therefore, we have $\sum_{k=0}^{\infty} \frac{\theta^k}{k!} g(k) = 0$ if and only if $g(k) = 0$ for all $k=0, 1, 2, \dots$. This proves, that

$\hat{\varphi}_j$ is a complete statistic.