Minimum Distance Loss-Reserving

Erhard Kremer

Institut für Mathematische Stochastik
Bundesstrasse 55
20146 Hamburg
Germany

Telephone: +49(40)428 38 4934
Facsimile: +49(40)428 38 4924
E-mail:Kremer@math.uni-hamburg.de

MINIMUM DISTANCE LOSS-RESERVING

ERHARD KREMER, HAMBURG

1. INTRODUCTION

Nowadays loss reserving is a very well developed subfield of nonlife insurance mathematics. In several introductory actuarial textbooks chapters on loss reserving can be found (see e.g. Kremer (1985) and CAS (1993)) and some texts are dedicated to that topic (see e.g. Taylor (1986), Institute of Actuaries (1990)). Many models and methods are related to fields of mathematical statistics, e.g. to (loglinear) regression analysis (see e.g. Kremer (1982), Taylor (1986)), time series analysis (see e.g. Kremer (1984), Verrall (1989)) and Bayes analysis (see e.g. Verrall (1990), Jewell (1989)). Obviously progress is still in the field. Recently the author himself wrote three new papers on the subject (see Kremer (1993a), (1993b), (1996)), giving comparably simple methods that were not yet presented in the already vast literature. In the following the author develops a further approach that surprisingly also was not yet described.

2. THE LOSS RESERVING PROBLEM.

Let X_{ij}, i, j = 1,...,n be nonnegative random variables on a probability space $(\Omega,\mathcal{A},P)$, X_{ij} denoting the total claims amount (or claims size per claim) of a collective of risks. The X_{ij} with $j \leq n-i+1$ are the known past amounts and the X_{ij} with $j \leq n-i+2$ the unknown future amounts. The index i numbers the accident years, the index j the development years. The known triangle

$$X_\Delta = (X_{ij},\ j = 1,\dots,\ n-i+1,\ i = 1,\dots,n)$$

is usually called <u>run-off-triangle</u>. The problem of loss reserving consists in predicting the X_{ij} (with $j \geq n-i+2$) from the run-off triangle X_Δ. Denote the prediction of X_{ij} with $\hat{X}_{ij}$. Then the discounted) loss reserve of accident year i is just:

$$\sum_{j=n-i+2}^{n} (\hat{X}_{ij} - \hat{X}_{i,j-1}) \cdot D_{ij},$$

where D_{ij} is a factor that discounts the j-th development year of the i-th accident year down to the (n-i+1)-th development year.

3. SIMILARITY OF DEVELOPMENTS.

The methods that are presented lateron are based on calculation of the similarity of different developments. Define for i = 1,..., n-j+1, j = 1,...,n the increments:

$$Y_{ij} = (X_{ij}/X_{i,j-1})$$

with the convention that $X_{i0} = 1$ for all i. Consider the situation that one looks at the claims amount $X_{i,n-i+k}$ with $k \geq 2$. One wants to measure the similarity (or reciprocally the dissimilarity) of the development of the i-th accident year with 'relevant' previous ones. The 'relevant' previous ones are those with their accident year no running from 1 to (i-k+1). The similarity one can measure with computing distances between the different developments. It is obvious to take certain (weighted) euklidean distances, more concretely the value:

$$d(i,\ell) = \left(\sum_{j=1}^{n-i+1} W_{ij} \cdot (Y_{ij} - Y_{\ell j})^2 \right)^{1/2}$$

for measuring the distance of the i-th development from the ℓ-th development ($\ell \leq i-k$), where W_{ij} are certain weights. The most simple choice is clearly:

$$W_{i1} = 1, i = n$$

$$W_{i1} = 0, W_{ij} = 1, j = 2,\ldots,n-i+1, i < n,$$

but choices giving for larger j more weight than for smaller ones are also reasonable. Such distances play an important role in the field of multivariate stastical analysis (see e.g. Seber (1984)), e.g. in the field of cluster analysis. Often it will be advantageous to standerdize the Y_{ij}, $j \leq n-i+1$, $i = 1,\ldots,n$ column-wise in advance, before computing the distances. With the help of these distances the author gives in the following certain reasonable new loss-development prediction methods. The author calls these methods minimum distance loss reserving methods.

4. A SIMPLEST MINIMUM DISTANCE METHOD.

Consider the prediction of $X_{i,n-i+k}$ (with $k \geq 2$). One can compare the known development of the i-th accident year with the previous 'relevant' ones, i.e. with those having accident year number smaller or equal to (i-k+1). This means that one determines that accident year with number $\ell_* \leq (i-k+1)$ that minimizes $d(i,\ell)$ in $\ell \leq (i-k+1)$. The resulting accident year is among the 'relevant' ones that one which is most similar to the accident year no. i according to their known developments. One computes the lagfactor:

$$\lambda_{i,n-i+k} = Y_{\ell_*,n-i+k}$$

and uses as predictor of $X_{i,n-i+k}$ the:

$$\hat{X}_{i,n-i+k} = \lambda_{i,n-i+1} \cdot \hat{X}_{i,n-i+k-1}. \tag{4.1}$$

$$\text{(with } \hat{X}_{i,n-i+1} = X_{i,n-i+1}\text{)}$$

Obviously this not yet published procedure is in some sense reasonable and surly quite handy.

5. A MORE REFINED MINIMUM DISTANCE METHOD.

The previous simple procedure lacks a little bit from not averaging when computing the lag-factor. But the method can easily modified such that averaging is used to a certain degree. Instead of determining only the one nearest accident year development, one can determine the two nearest say with accident year number ℓ_1 and ℓ_2. Then one can compute the lagfactor for predicting $X_{i,n-i+k}$ according:

$$(5.1) \qquad \lambda_{i,n-i+k} = w_{1ik} \cdot Y_{\ell_1,n-i+k} + w_{2ik} \cdot Y_{\ell_2,n-i+k}$$

with weights $w_{pik}, p = 1,2$:

$$w_{1ik} + w_{2ik} = 1,$$

e.g.:

$$w_{pik} = \frac{d(i,\ell_p)}{d(i,\ell_1) + d(i,\ell_2)}, \; p = 1,2$$

or simply:

$$w_{pik} = (1/2), \; p = 1,2.$$

One forecasts again with the advice (4.1). For k = i one clearly cannot take (5.1), but can choose:

$$\lambda_{in} = Y_{1n} .$$

This procedure might be preferable to that of section 4. It is straight forward to extend it by taking the m (≥3) nearest accident years instead of only the 2 nearest ones for k ≤ i-m. Taking all 'relevant' accident years into consideration the procedure would give as limiting case the classical link-ratio method that does not destinguish between dissimilar developments.

6. A MOST REFINED MINIMUM DISTANCE METHOD.

The last method type (with e.g. $m = 2$) is in some sense based on making two clusters, one with $(m+1)$ elements and one with remaining $(i\text{-}k\text{-}m)$ 'relevant' ones, when predicting $X_{i,n-i+k}$ (with $k\geq 2$). The size of the first cluster is fixed to $(m+1)$ elements. The idea behind the method is that not all known 'relevant' developments are related to the i-th one, but only the m most similar ones. Prescribing the size of one cluster (to be $m+1$), is in some sense a restriction (that clearly makes the method simpler to apply). A near-lying extension is, not to take a fixed size for the first cluster. What one does then is simply to make a cluster analysis into two clusters, one containing the i-th development and one not containing it. And that, including all 'relevant' developments and for each unknown $X_{i,n-i+k}$ $(k\geq 2)$. For clustering one takes for the j-th development (corresponding to the j-th 'relevant' accident year) the (multivariate) observation:

$$(Y_{j1}), \text{ for } i = n, \text{ and}$$

$$(Y_{j2},\dots, Y_{j,n-i+1}), \text{ for } i < n,$$

when predicting $X_{i,n-i+k}$. Often it is adequate to standardize the Y_{ij} column-wise in advance. Suppose the indices $\ell_1,\dots,\ell_m$ (and i) of the developments (of the accident years) are those of the first cluster, then one computes:

$$(6.1) \qquad \lambda_{i,n-i+k} = \sum_{p=1}^{m} w_{pik} \cdot Y_{\ell_p,n-i+k}$$

with weights w_{pik}, $p = 1,\dots,m$:

$$\sum_{p=1}^{m} w_{pik} = 1$$

e.g.:

$$w_{pik} = d(i,\ell_p) / \sum_{p=1}^{m} d(i,\ell_p),\ p = 1,\ldots,m$$

or simply:

$$w_{pik} = (1/m),\ p = 1,\ldots,m.$$

One again uses the advice (4.1) for prediction. Methods for computing two optimal clusters are wellknown in multivariate statistical analysis (see e.g. Seber (1984)) and can easily be applied to the present context. It can happen that the two resulting clusters are statistically not significantly destinguishable. Then one would sum in (6.1) over the indices of all 'relevant' accident years. The decision about having significantly two or only one cluster can be done with the wellknown Hotelling-test of multivariate statistical analysis (see e.g. Seber (1984)). When applying that test in the given context the author proposes to use instead of the Y_{ij} the transformed values:

$$Y^*_{ij} = \ell n(Y_{ij}).$$

7. OUTLOOK.

Though simple and near-lying, the above procedures seem to be not developed until the appearance of the present paper. The method of section 6 is in some sense a competitor of the author's threshold loss reserving methods (see Kremer (1996)), that also makes a certain clustering with help of a threshold. The practising actuaries are invited to test and compare both new approaches under practical conditions. In practice one should adjust the claims data for inflation in advance. For illustration of the methods:

8. NUMERIAL EXAMPLE.

Consider the run-off triangle X_{Δ}:

31.28	48.98	67.39	79.14	85.43	96.20
60.47	77.53	114.51	154.47	191.31	
33.77	49.39	62.65	82.13		
67.06	95.49	120.54			
29.58	43.22				
30.14					

giving for Y_{ij} the values:

31.28	1.566	1.376	1.174	1.079	1.126
60.47	1.282	1.477	1.349	1.238	
33.77	1.463	1.268	1.311		
67.06	1.424	1.262			
29.58	1.461				
30.14					

First take the method of section 4. One gets the following values for the 'optimal' lagfactors λ_{ij}, $j \geq n\text{-}i+3$:

				1.126
			1.079	1.126
		1.311	1.079	1.126
	1.268	1.311	1.079	1.126
1.461	1.376	1.174	1.079	1.126

giving as completion of the rectangle of all X_{ij}:

				215.41
			88.62	99.79
		158.03	170.51	191.99
	54.80	71.84	77.52	87.29
44.03	60.59	71.13	76.75	86.42

Secondly take the method with m = 2 of section 5.

One gets the following values for the lagfactors λ_{ij}, $j \geq n\text{-}i+3$:

				1.126
			1.158	1.126
		1.242	1.158	1.126
	1.265	1.242	1.158	1.126
1.513	1.426	1.242	1.158	1.126

giving as completion of the rectangle of all X_{ij}:

				215.41
			95.11	107.09
		149.71	173.36	195.20
	54.67	67.90	78.63	88.54
45.60	65.03	80.77	93.53	105.31

Finally take the method of section 6. The author took Wards-clustering method and got the following values for the lagfactors λ_{ij}, $j \geq n\text{-}i+3$:

				1.126
			1.079	1.126
		1.349	1.079	1.126
	1.369	1.349	1.079	1.126
1.390	1.336	1.330	1.079	1.126

giving the completion:

				215.41
			88.62	99.79
		162.61	175.46	192.57
	59.17	79.82	86.13	96.98
41.89	55.96	74.43	80.31	90.43

In all three cases the above given simplest weights were used and for simplicity no standardization of the Y_{ij}, $j \leq n\text{-}i+1$ done in advance. Furthermore the mentioned Hotelling-test was not carried through. The results differ in certain places considerably. The first method seems to give (like expected) the most disappealing results.

REFERENCES.

Casualty Actuarial Society (1993): Foundations of Casualty Actuarial Science. Arlington & New York.

Institute of Actuaries (1990): Claims reserving manual. London.

Jewell, W.S. (1989): Predicting IBNYR events and delays I. Continuous time. Astin Bulletin.

Kremer, E. (1982): IBNR claims and the two-way model of ANOVA. Scandinavian Actuarial Journal.

Kremer, E. (1984): A class of autoregressive models for predicting the final claims amount. Insurance: Mathematics and Economocs.

Kremer, E. (1985): Einführung in die Versicherungsmathematik. Verlag Vandenhoeck und Rubrecht. Göttingen & Zürich.

Kremer, E. (1993a): Random-coefficient autoregressive loss reserving. Blätter der deutschen Gesellschaft für Versicherungsmathematik.

Kremer, E. (1993b): Certain extensions of the chain-ladder technique. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.

Kremer, E. (1996): Threshold loss reserving. Proceeding of the international ASTIN colloquium at Copenhagen

Seber, C.A.F. (1984): Multivariate Observations. John Wiley & Sons, New York.

Taylor, G.C. (1986): Claims reserving in nonlife insurance. North Holland, Amsterdam.

Verrall. R.J. (1989): Modelling claims runoff triangles with two-dimensional time series. Scandinavian Actuarial Journal.

Verrall, R.J. (1990): Bayes and empirical Bayes estimation for the chain ladder model. Astin Bulletin.

ZUSAMMENFASSUNG.

Gewisse noch nicht bislang publizierte, elementare Schadenreservierungsmethoden werden präsentiert. Der grundlegende Ansatz besteht in gewissem Sinne in der Kombination von Link-Ratio- und Cluster-Analyse-Techniken.

SUMMARY.

Certain not yet published, elementary loss reserving methods are presented. The basic approach consists in some sense in combining link-ratio with cluster-analysis techniques.