JIA 118 (1991) 489-499

J.I.A. **118**, III, 489–499

CHAIN LADDER AND MAXIMUM LIKELIHOOD

By R. J. Verrall, M.A., M.Sc., Ph.D.

(of the City University, London)

ABSTRACT

This paper derives second moments of estimates of the parameters in the chain ladder model. Thus, the so-called link ratios, and proportions of ultimate claims for each development year are considered. This enables confidence statements about the chain ladder parameters to be made with statistical rigour. The methods are illustrated using 6 sets of real data taken from the DTI returns.

KEYWORDS

Chain Ladder; General Insurance; Estimation; Mathematical Models

1. INTRODUCTION

In this paper, the theory of maximum likelihood is utilised to examine the properties of the parameters of the chain ladder model. The exposition concentrates on the development factors, since the prediction of total outstanding claims for each year of business has been considered in a previous paper (Verrall (1989)). The present paper also considers the proportions of ultimate claims which are paid in each development year, being the parameters of a multiplicative model. The paper will consider first the models and then the estimation of the parameters in the models. Finally, the methods will be applied to some actual data taken from the DTI returns.

The method uses the representation of the chain ladder technique formulated by Kremer (1982). A useful discussion of the use of Kremer's formulation from a practitioner's point of view is given by Christofides (1990). A fuller explanation of the relationship between the actuarial technique and its statistical treatment may also be found in that article.

The present paper considers the second moments of the parameters used in the chain ladder technique. This has not been attempted with any statistical rigour elsewhere, and should prove useful in assessing the accuracy of the estimates.

It is assumed in this paper that the incremental claims are positive. The methodology can be extended to cover negative values using the quasi-likelihood approach suggested by Wright (1990).

2. THE MODELS

The chain ladder technique is based on a model which relates the cumulative claims for each year of business recursively throughout the development.

Suppose, without loss of generality, that the data consist of a triangle indexed by i (year of business) and j (development year):

$$\begin{array}{lll} C_{1,1} & C_{1,2} \ldots\ldots\ldots\ldots\ldots\ldots C_{1,n} \\ C_{2,1} & C_{2,2} \ldots\ldots\ldots\ldots C_{2,n} \\ \vdots & \\ \vdots & \\ C_{n,1} & \end{array}$$

It is assumed that the data have been adjusted for exposure and inflation in the usual way (see, for example, Renshaw (1989)). The chain ladder technique is based on the following model:

$$E(C_{ij}) = \lambda_j C_{i,j-1}. \tag{1}$$

We are not concerned with the chain ladder estimation method which is usually applied in conjunction with this model, the structure of the model being the more important part of the chain ladder technique.

There are two alternative ways of expressing the chain ladder model, both of which use the incremental claims rather than the cumulative claims. Thus we define:

$$Z_{i1} = C_{i1} \qquad (\forall i)$$

$$Z_{ij} = C_{ij} - C_{i,j-1} \qquad (\forall i, j \neq 1)$$

and assume that $Z_{ij} > 0$, $\forall i$, $\forall j$.

The first of the alternative representations of the chain ladder model is the multiplicative model, which is given by:

$$E(Z_{ij}) = U_i S_j \tag{2}$$

where:

$$\sum_{j=1}^{n} S_j = 1.$$

In this model, U_i is the expected ultimate claims in year of business i, and S_j is the proportion of the expected ultimate claims which arrive in development year j. 'Ultimate' is used in the sense implied by the chain ladder technique, and does not include any claims beyond the latest development year to have been observed.

By taking logs of the multiplicative model, the second alternative representation of the chain ladder model is obtained, being the familiar two-way analysis of variance model:

$$E(\log Z_{ij}) = \mu + \alpha_i + \beta_j \tag{3}$$

where: $$\alpha_1 = \beta_1 = 0.$$

The right hand side of equation (3) can be obtained from the right hand side of equation (2) by taking logs and reparameterising. Thus:

$$\log (U_i S_j) = \log U_i + \log S_j.$$

The model is now in additive form, and we have to incorporate the restriction:

$$\sum_{j=1}^{n} S_j = 1.$$

The standard method for achieving this is to cast the restriction in a different form. Thus, a different restriction is used in order to obtain a more convenient interpretation of the parameters. This gives equation (3). The resulting estimates of claims are unchanged. It can easily be verified that the number of free parameters has not changed.

Further details of the application of this model can be found in Renshaw (1989) and Renshaw & Verrall (1989).

These three models, given by equations (1) to (3), are equivalent and are reparameterisations of the same structure. The parameters of the first two have physical interpretations, while the statistical analysis of the latter is the more straightforward. For this reason, the analysis is approached via the additive model, given by equation (3), and the results are then related back to the multiplicative and chain ladder models in order to obtain more information on the development factors:

$$\{\lambda_j : j = 2, \ldots, n\}$$

and the proportion of ultimate claims:

$$\{S_j : j = 1, \ldots, n; \sum_{j=1}^{n} S_j = 1\}.$$

The advantage of this approach is that more information can be elicited on these quantities.

The following theorem shows that the chain ladder and multiplicative models are equivalent. The proof is due to Kremer (1982).

Theorem

Suppose that the claims are modelled by equations (1) and (2). Then the following relationships between the parameters hold:

$$S_1 = \frac{1}{\sum_{l=2}^{n} \lambda_l} \tag{4}$$

$$S_j = \frac{\lambda_j - 1}{\sum_{l=j}^{n} \lambda_l} \qquad (j \geq 2) \tag{5}$$

$$U_i = C_{in}. \tag{6}$$

Notes

The proof of this theorem is given in Kremer (1982).

There is a parameter for each business year: this may not be obvious for the chain ladder model.

The equivalence between the multiplicative and additive models was also proved by Kremer (using a slightly different restriction on the parameters of the additive model), and is given by the following theorem.

Theorem

Suppose that the claims are modelled by equations (2) and (3). Then the following relationships between the parameters hold:

$$S_j = \frac{e^{\beta_j}}{\sum_{l=1}^{n} e^{\beta_l}} \qquad (j = 1, \ldots, n) \tag{7}$$

where: $\beta_1 = 0$ by definition,

and:

$$U_i = e^{\mu + \alpha_i} \sum_{j=1}^{n} e^{\beta_j}. \tag{8}$$

Proof

This theorem was proved by Kremer, and the reader is referred to that paper for the details. A few points may help to clarify the relationship between the analysis of variance linear model and the chain ladder technique. It can be seen that:

$$\sum_{j=1}^{n} S_j = 1$$

and that S_j, as given by equation (7), does measure the proportion of claims which emerge in delay year j.

Equation (8) can be obtained simplistically by exponentiating equation (3) and summing over j. It thus represents claims relating to business year i.

These two theorems are all that are needed to relate the three models, but for our purposes the relationship between the parameters of the chain ladder and additive models needs to be made explicit. We thus also need the following theorem.

Theorem

Suppose that the claims are modelled by equations (1) and (3). Then the following relationships between the parameters hold:

$$\lambda_j = 1 + \frac{e^{\beta_j}}{\sum_{l=1}^{j-1} e^{\beta_l}} \tag{9}$$

and

$$C_{in} = e^{\mu + \alpha_i} \sum_{j=1}^{n} e^{\beta_j}. \tag{10}$$

Proof

The expected ultimate claims for each year of business are not the main subject of this paper, having been dealt with in detail in Verrall (1989). The proof of equation (10) is straightforward from equations (6) and (8).

In order to show that equation (9) holds, it is first necessary to establish the following relationship between the parameters of the chain ladder and multiplicative models:

$$\lambda_j = \frac{1 - \sum_{l=j+1}^{n} S_l}{1 - \sum_{l=j}^{n} S_l} \qquad (j \neq n) \tag{11}$$

$$\lambda_n = \frac{1}{1 - S_n}. \tag{12}$$

This can be proved by induction. Consider first λ_n. From equation (5):

$$S_n = \frac{\lambda_n - 1}{\lambda_n}.$$

Thus:

$$\lambda_n - S_n \lambda_n = 1$$

and

$$\lambda_n = \frac{1}{1 - S_n}.$$

Now suppose that equation (11) holds for $\lambda_n, \ldots, \lambda_{j+1}$. From equation (5):

$$S_j = \frac{\lambda_j - 1}{\lambda_j \prod_{l=j+1}^{n} \lambda_l}$$

$$= \frac{\lambda_j - 1}{\lambda_j} \cdot \left(1 - \sum_{l=j+1}^{n} S_l\right)$$

using the assumption of the induction.

Thus:

$$1 - \sum_{l=j+1}^{n} S_l = \lambda_j \left(1 - \sum_{l=j+1}^{n} S_l - S_j\right)$$

and

$$\lambda_j = \frac{1 - \sum_{l=j+1}^{n} S_l}{1 - \sum_{l=j}^{n} S_l}$$

which completes the proof of equation (11) by induction.

Substituting for S_j from equation (7) into equations (11) and (12) gives:

$$\lambda_j = \frac{1 - \sum_{l=j+1}^{n} \left(\frac{e^{\beta_l}}{\sum_{k=1}^{n} e^{\beta_k}} \right)}{1 - \sum_{i=j}^{n} \left(\frac{e^{\beta_l}}{\sum_{k=1}^{n} e^{\beta_k}} \right)}$$

$$= \frac{\sum_{l=1}^{n} e^{\beta_l} - \sum_{l=j+1}^{n} e^{\beta_l}}{\sum_{l=1}^{n} e^{\beta_l} - \sum_{l=j}^{n} e^{\beta_l}}$$

$$= \frac{\sum_{l=1}^{j} e^{\beta_l}}{\sum_{l=1}^{j-1} e^{\beta_l}}$$

$$= 1 + \frac{e^{\beta_j}}{\sum_{l=1}^{j-1} e^{\beta_l}}$$

and

$$\lambda_n = \frac{1}{1 - \frac{e^{\beta_n}}{\sum_{l=1}^{n} e^{\beta_l}}}$$

$$= \frac{\sum_{l=1}^{n} e^{\beta_l}}{\sum_{l=1}^{n-1} e^{\beta_l}}$$

$$= 1 + \frac{e^{\beta_n}}{\sum_{l=1}^{n-1} e^{\beta_l}}$$

which completes the proof of the theorem.

3. MAXIMUM LIKELIHOOD ESTIMATION

The parameters of the additive model can be estimated by least squares, which is equivalent to maximum likelihood estimation if the errors are assumed to be independently, normally distributed. The variance–covariance matrix of the parameter estimates can be obtained from the Fisher information matrix by differentiating the log-likelihood a second time. Further details of the theory of maximum likelihood which is used in this section can be found in Cox & Hinkley (1974).

Since maximum likelihood estimates are invariant under parameter transformations, the maximum likelihood estimates of the development factors and the proportions of ultimate claims can be obtained by substituting the estimates of $\{\beta_j : j = 1, \ldots, n;\ \beta_1 = 0\}$ into equations (7) and (9). In addition to the parameter estimates, it is useful to have standard errors of the parameter estimates which can be obtained by likelihood theory. The particular advantage of using maximum likelihood estimation is that the second moments are relatively straightforward to obtain. Denoting the variance-covariance matrix of $\{\beta_j : j = 1, \ldots, n; \beta_1 = 0\}$ by:

$$V(\boldsymbol{\beta})$$

the variance–covariance matrix of $\{\lambda_j : j = 2, \ldots, n\}$ and $\{S_j : j = 1, \ldots, n;$ $\sum_{j=1}^{n} S_j = 1\}$ are given by:

$$V(\boldsymbol{\lambda}) = \left(\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}}\right)' V(\boldsymbol{\beta}) \left(\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\beta}}\right) \tag{13}$$

and
$$V(S) = \left(\frac{\partial S}{\partial \beta}\right)' V(\beta) \left(\frac{\partial S}{\partial \beta}\right). \tag{14}$$

It is thus necessary to obtain the matrices of the first derivatives of the respective parameter vectors.

The (j,k)th element of $\left(\frac{\partial \lambda}{\partial \beta}\right)$ can be obtained from equation (9) and is given by:

$$\frac{\partial \lambda_j}{\partial \beta_k} = \begin{cases} 0 & (k > j) \\ \dfrac{e^{\beta_j}}{\sum_{l=1}^{j-1} e^{\beta_l}} & (k = j) \\ \dfrac{e^{\beta_j} e^{\beta_k}}{\left(\sum_{l=1}^{j-1} e^{\beta_l}\right)} & (k < j) \end{cases}$$
$$= \begin{cases} 0 & (k > j) \\ \lambda_j - 1 & (k = j) \\ -(\lambda_j - 1)(\lambda_k - 1) & (k < j). \end{cases} \tag{15}$$

Similarly, the (j,k)th element of $\left(\frac{\partial S}{\partial \beta}\right)$ can be obtained from equation (7) and is given by:

$$\frac{\partial S_j}{\partial \beta_k} = \begin{cases} -\dfrac{e^{\beta_j} e^{\beta_k}}{\left(\sum_{l=1}^{n} e^{\beta_l}\right)} & (k \neq j) \\ \dfrac{e^{\beta_j}\left(\sum_{l=1}^{n} e^{\beta_l} - e^{\beta_j}\right)}{\left(\sum_{l=1}^{n} e^{\beta_l}\right)} & (k = j) \end{cases}$$

$$= \begin{cases} -S_j S_k & (k \neq j) \\ S_j(1 - S_j) & (k = j). \end{cases} \tag{16}$$

Estimates of the variance-covariance matrices can be obtained by substituting estimates of the parameters into equations (13) and (14).

A technical note is that the parameter β_1 (which is defined to be zero) has to be included in the matrix of partial derivatives in equation (14) since there are n parameters in the vector $\boldsymbol{S}$. The variance-covariance matrix of the parameters of the additive model, which is obtained from a standard least squares analysis, has to be augmented to include an extra row and column, all of whose entries are zero. This is not necessary for equation (13).

4. EXAMPLE

The method is applied to six sets of employers' liability data which have been obtained from the DTI returns. The names of the companies to which the data apply have been suppressed, and it should be commented that this mathematical analysis is only one part of the process by which reserves are set. In particular, the DTI data are gross of reinsurance. The results here should, therefore, be regarded as a statistical analysis which would give further information to the claims reserver, who would use the other information available.

We now consider the parameter estimates for each of the three models in turn. Beginning with the additive model, given by equation (3), the estimates of the column parameters $\{\beta_j : j = 2, \ldots, n\}$ and their standard errors are given in Table 1:

Table 1.

Company											
1		2		3		4		5		6	
1·796	0·121	1·748	0·148	2·236	0·249	1·846	0·248	1·941	0·201	2·010	0·082
1·848	0·126	1·857	0·155	2·080	0·261	2·260	0·260	2·248	0·211	2·246	0·086
1·669	0·133	1·654	0·163	1·978	0·273	2·159	0·272	2·204	0·221	2·129	0·091
1·413	0·139	1·400	0·171	1·725	0·287	1·986	0·286	1·981	0·232	1·863	0·095
0·994	0·147	1·200	0·180	1·535	0·303	1·535	0·302	1·514	0·245	1·485	0·100
0·615	0·155	0·705	0·190	1·057	0·320	1·235	0·319	0·788	0·259	1·050	0·106
0·415	0·164	0·339	0·201	0·667	0·338	0·644	0·337	0·227	0·274	0·782	0·112
0·038	0·175	0·025	0·215	−0·099	0·360	0·222	0·359	−0·540	0·291	0·234	0·120
−0·812	0·189	−0·407	0·232	−0·300	0·390	0·047	0·388	−0·993	0·315	0·155	0·129
−0·915	0·212	−1·821	0·260	−0·715	0·437	0·382	0·435	−1·311	0·353	−0·324	0·145
−2·513	0·264	−1·492	0·323	−1·708	0·543	−0·896	0·541	−3·206	0·439	−0·304	0·180

Before going on to the parameters which have a physical interpretation, it should be noticed that it is already possible to see some differences between the companies. In particular, the standard errors of the parameters are larger for some companies (3 and 4) than for others (6). This will be mirrored in the parameter estimates and standard errors of the other models.

Next, consider the chain ladder model, given by equation (1). The estimates of the development factors $\{\lambda_j : j = 2, \ldots, n\}$ and their standard errors are given in Table 2:

Table 2.

Company

1		2		3		4		5		6	
7·027	0·727	6·742	0·850	10·356	2·327	7·332	1·569	7·963	1·401	8·466	0·616
1·904	0·101	1·950	0·130	1·773	0·181	2·307	0·300	2·189	0·222	2·117	0·086
1·397	0·041	1·398	0·050	1·394	0·084	1·512	0·109	1·520	0·090	1·469	0·033
1·220	0·022	1·221	0·027	1·219	0·046	1·285	0·059	1·274	0·046	1·245	0·017
1·119	0·012	1·148	0·019	1·149	0·032	1·141	0·030	1·135	0·023	1·135	0·009
1·073	0·008	1·079	0·010	1·080	0·018	1·092	0·020	1·057	0·010	1·077	0·006
1·055	0·006	1·051	0·007	1·050	0·012	1·047	0·011	1·031	0·006	1·055	0·004
1·036	0·005	1·035	0·006	1·022	0·006	1·029	0·008	1·014	0·003	1·030	0·003
1·015	0·002	1·022	0·004	1·018	0·005	1·024	0·007	1·009	0·002	1·027	0·003
1·013	0·002	1·005	0·001	1·012	0·004	1·032	0·011	1·006	0·002	1·016	0·002
1·003	0·001	1·007	0·002	1·004	0·002	1·009	0·004	1·001	0·000	1·016	0·003

Finally, consider the multiplicative model. The estimates of the proportions of ultimate claims in each development year

$$\{S_j : j = 1, \ldots, n; \sum_{j=1}^{n} S_j = 1\}$$

and their standard errors are given in Table 3:

Table 3.

Company

1		2		3		4		5		6	
0·032	0·003	0·032	0·004	0·023	0·005	0·021	0·005	0·023	0·004	0·022	0·002
0·196	0·016	0·184	0·019	0·218	0·036	0·135	0·025	0·162	0·023	0·162	0·010
0·206	0·015	0·205	0·019	0·186	0·029	0·204	0·032	0·220	0·027	0·205	0·011
0·172	0·013	0·167	0·015	0·168	0·026	0·184	0·028	0·211	0·025	0·182	0·009
0·133	0·011	0·130	0·013	0·131	0·021	0·155	0·024	0·169	0·022	0·140	0·007
0·088	0·008	0·106	0·011	0·108	0·019	0·099	0·017	0·106	0·015	0·096	0·006
0·060	0·006	0·065	0·007	0·067	0·013	0·073	0·014	0·051	0·008	0·062	0·004
0·049	0·005	0·045	0·006	0·045	0·010	0·040	0·009	0·029	0·005	0·047	0·003
0·034	0·004	0·033	0·005	0·021	0·005	0·027	0·007	0·014	0·003	0·027	0·002
0·014	0·002	0·021	0·004	0·017	0·005	0·022	0·006	0·009	0·002	0·025	0·002
0·013	0·002	0·005	0·001	0·011	0·004	0·031	0·011	0·006	0·002	0·016	0·002
0·003	0·001	0·007	0·002	0·004	0·002	0·009	0·004	0·001	0·000	0·016	0·002

The run-off patterns of the companies can be compared using Tables 2 and 3. For example, 1 and 2 seem quite similar, and some of the companies have more run-off in later development years than others. The standard errors can also be compared, with the same conclusions as above.

The standard errors of the parameters also affect the predictions of total outstanding claims for each of the companies and the predicted mean square error of the predictions. For these quantities, unbiased estimation is used (see Verrall (1989)).

Table 4.

Company	Predicted total outstanding claims	Root mean square error of prediction	95% upper bound on total outstanding claims
1	36,295,760	4,786,227	44,169,103
2	37,638,112	6,565,736	48,438,748
3	9,912,414	2,582,119	14,160,000
4	21,619,344	5,019,421	29,876,292
5	16,310,471	3,566,213	22,176,891
6	54,559,568	4,550,077	62,044,445

The ratio of the root mean square error to the predicted total outstanding claims mirrors the pattern of the standard errors of the parameters, with 3 the largest at 26·0% and 6 the smallest at 8·3%. These figures should give some indication of how the companies should make provision for the likely variation in total outstanding claims, based on the statistical information in the data (Table 4).

ACKNOWLEDGEMENT

I would like to thank Commercial Union Assurance Company for financial support during the period of writing this paper, and also Mr S. Christofides for useful discussions.

REFERENCES

CHRISTOFIDES, S. (1990). *Regression Models Based on Log-incremental Payments.* Claims Reserving Manual, Institute of Actuaries, London.

COX, D. R. & HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

KREMER, E. (1982). IBNR-Claims and the Two-Way Model of ANOVA. *Scand. Act. J.*, **1**, 47–55.

RENSHAW, A. E. (1989) Chain Ladder and Interactive Modelling. Applications of Mathematics in Insurance, Finance and Actuarial Work; joint seminar of the Institute and the Faculty of Actuaries and the Institute of Mathematics and Its Applications. *J.I.A.* **116**, 559–587.

RENSHAW, A. E. & VERRALL, R. J. (1989). Claims Reserving: Statistical Treatment of the Chain Ladder Technique. Notes for a Short Course at the City University.

VERRALL, R. J. (1989b). On the Unbiased Estimation of Reserves from Loglinear Models. *Insurance: Mathematics and Economics*, **10**, 1, 75–80.

WRIGHT, T. S. (1990). A Stochastic Method for Claims Reserving in General Insurance. *J.I.A.* **117**, 677–731.